

ANAIS



CSBC 2016

XXXVI Congresso da Sociedade Brasileira de Computação
PUCRS | PORTO ALEGRE/RS

Sociedade Brasileira de Computação | SBC

Organizadores

Avelino Francisco Zorzo

Alexandre Agustini

Márcia de Borba Campos

Realização

Faculdade de Informática | FACIN

Pontifícia Universidade Católica do Rio Grande do Sul | PUCRS

Sociedade Brasileira de Computação | SBC

Copyright © 2016 da Sociedade Brasileira de Computação Todos os direitos reservados

Av. Bento Gonçalves, 9500 – Setor 4 – Prédio 43.412 – Sala 219
Bairro Agronomia – CEP 91.509-900 – Porto Alegre – RS
Fone: (51) 3308-6835
E-mail: sbc@sbc.org.br

Dados Internacionais de Catalogação na Publicação (CIP)

S678 Sociedade Brasileira de Computação. Congresso. (36. :
2016 : Porto Alegre, RS)
Anais do XXXVI congresso da sociedade brasileira de
computação : #Computaçãoeinterdisciplinaridade
[recurso eletrônico] / organizadores: Alexandre Agustini,
Avelino Francisco Zorzo, Marcia de Borba Campos. –
Dados eletrônicos. – Porto Alegre : Sociedade Brasileira
de Computação, 2016.

Modo de acesso: <http://www.pucrs.br/edipucrs/>
ISSN 2175-2761
ISBN 978-85-7669-322-2

1. Informática – Congressos. 2. Interdisciplinaridade.
I. Agustini, Alexandre. II. Zorzo, Avelino Francisco.
III. Campos, Marcia de Borba.

CDD 23. ed. 004.06081

Salete Maria Sartori CRB 10/1363
Setor de Tratamento da Informação da BC-PUCRS

**Alexandre Agustini
Avelino Francisco Zorzo
Marcia de Borba Campos
Organizadores**



CSBC 2016

XXXVI Congresso da Sociedade Brasileira de Computação

#ComputaçãoInterdisciplinaridade

ANAIS



PORTO ALEGRE - 2016

Diretoria da Sociedade Brasileira de Computação (SBC)

Presidente

Lisandro Zambenedetti Granville (UFRGS)

Vice-Presidente

Thais Vasconcelos Batista (UFRN)

Diretora Administrativa

Renata de Matos Galante (UFRGS)

Diretor de Finanças

Carlos André Guimarães Ferraz (UFPE)

Diretor de Eventos e Comissões Especiais

Antônio Jorge Gomes Abelém (UFPA)

Diretor de Educação

Avelino Francisco Zorzo (PUCRS)

Diretor de Publicações

José Viterbo Filho (UFF)

Diretora de Planejamento e Programas Especiais

Claudia Lage Rebello da Motta (UFRJ)

Diretor de Secretarias Regionais

Marcelo Duduchi Feitosa (CEETEPS)

Diretora de Divulgação e Marketing

Eliana Almeida (UFAL)

Diretor de Relações Profissionais

Roberto da Silva Bigonha (UFMG)

Diretor de Competições Científicas

Ricardo de Oliveira Anido (UNICAMP)

Diretor de Cooperação com Sociedades Científicas

Raimundo José de Araújo Macêdo (UFBA)

Diretor de Articulação com Empresas

Sergio Soares (UFPE)

Comitê de Organização

Coordenação Geral do CSBC 2016

Avelino Francisco Zorzo (PUCRS)

Comitê Gestor

Antônio Jorge Gomes Abelém (UFPA)

Avelino Francisco Zorzo (PUCRS)

Carlos André Guimarães Ferraz (UFPE)

Elisangela Magario (SBC)

Fernanda dos Santos Jorge (SBC)

Lisandro Zambenedetti Granville (UFRGS)

Luciano Silva (Mackenzie)

Nizam Omar (Mackenzie)

Renata de Matos Galante (UFRGS)

Sergio Castelo Branco Soares (UFPE)

Thais Vasconcelos Batista (UFRN)

Equipe Organizadora

Alexandre Agustini (PUCRS)

Ana Paula Terra Bacelo (PUCRS)

Cristina Moreira Nunes (PUCRS)

Daniel Antonio Callegari (PUCRS)

Marcelo Hideki Yamaguti (PUCRS)

Márcia de Borba Campos (PUCRS)

Coordenação dos eventos

BraSNAM – 5º Brazilian Work on Social Network Analysis and Mining: Geral: Jonice de Oliveira Sampaio (UFRJ), Giseli Rabello Lopes (UFRJ) - Local: Isabel Manssour (PUCRS)

BreSci – 10º Brazilian e-Science Workshop: Geral: Kary A. D. C. S. Ocaña (LNCC), Regina Braga (UFJF) - Local: Avelino Francisco Zorzo (PUCRS)

Cloudscape - 1º Cloudscape Brazil: Geral: Priscila Solis Barreto (UnB), Marco Vieira (Universidade de Coimbra), Sara Pittonet (IT Trust Services) - Local: Tiago Ferreto (PUCRS)

COMPUTECC – 18º Seminário de Computação e Mercado: Geral: Rafael Prikladnicki (PUCRS) - Local: Ricardo Melo Bastos (PUCRS)

CQ – 17º Curso de Qualidade: Geral: Ronaldo Celso Messias Correia (UNESP) - Local: Marcelo Hideki Yamaguti (PUCRS)

CTD – 29º Concurso de Teses e Dissertações: Geral: André Soares (UFPI) - Local: Rodrigo Barros (PUCRS)

CTIC – 35º Concurso de Trabalhos de Iniciação Científica: Geral: Carina Friedrich Dorneles (UFSC)

DesafIE! – 4º Workshop de Desafios da Computação Aplicada à Educação: Geral: Dra. Patrícia Jaques (UNISINOS) - Local: Liane Tarouco (UFRGS), Marcia de Borba Campos (PUCRS)

ENCompIF – 3º Encontro Nacional de Computação dos Institutos Federais: Geral: Carlos Roberto de Oliveira Junior (IFRJ) - Local: Regio Michelin (IFRS-Restinga)

ETC – 1º Encontro de Teoria da Computação: Geral: Cláudia Linhares Sales (UFC), Rosiane de Freitas (UFAM) - Local: Alfio Martini (PUCRS)

FAILCON - 5º Failcon Brazil: Geral: Rafael Matone Chanin (PUCRS), Flávio Steffens de Castro (Bicharia)

FIBRE - 1º Workshop do testbed FIBRE: Geral: Leandro Neumann Ciuffo (RNP) - Local: Tiago Coelho Ferreto (PUCRS)

JAI – 35º Jornadas de Atualização em Informática: Geral: José Carlos Maldonado (USP), José Viterbo Filho (UFF), Marcio Eduardo Delamaro (USP) - Local: Sabrina Marczak (PUCRS)

SBCUP – 8º Simpósio Brasileiro de Computação Ubíqua e Pervasiva: Geral: Adenauer Yamin (UCPEL/UFPEL) - Local: César Marcon (PUCRS)

SBSC – 13º Simpósio Brasileiro de Sistemas Colaborativos: Geral: Sabrina Marczak (PUCRS), Josiane Kroll (PUCRS)

SECOMU – 42º Seminário de Computação na Universidade: Geral: Raimundo José de Araújo Macêdo (UFBA/UFSE)

SEMISH – 43º Seminário Integrado de Software e Hardware: Geral: Luciano Paschoal Gaspary (UFRGS) - Local: César Augusto FonticIELha De Rose (PUCRS)

WCAMA – 7º Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais: Geral: Diana Francisca Adamatti (FURG)

WCN - 1º Workshop on Cloud Networks: Geral: Priscila Solis Barreto (UnB), Marco Vieira (Universidade de Coimbra) - Local: Tiago Ferreto (PUCRS)

WEI – 24º Workshop sobre Educação em Computação: Geral: Ecivaldo Matos (UFBA), Leandro Galvão (UFAM) - Local: Lucia Giraffa (PUCRS)

WIM – 16º Workshop de Informática Médica: Geral: Débora Christina Muchaluat Saade (UFF)

WIT – 10º Women in Information Technology: Geral: Priscila Solis (UnB), Juliana Freitag Borin (UNICAMP) - Local: Ana Cristina Benso da Silva (PUCRS)

WPerformance – 15º Workshop em Desempenho de Sistemas Computacionais e de Comunicação: Geral: Daniel de Oliveira (UFF), Wagner Meira Jr. (UFMG) - Local: Afonso Salles (PUCRS)

WPIETFIRTF - 3º Workshop Pré IETF/IRTF: Geral: Leandro Melo de Sales (UFAL) - Local: Alexandre Agustini (PUCRS)

Sumário

Mensagem da Coordenação Geral	8
BraSNAM - 5º Brazilian Workshop on Social Network Analysis and Mining	11
BreSci - 10º Brazilian e-Science Workshop	223
Cloudscape - 1º Cloudscape Brazil	314
COMPUTEC - 18º Seminário de Computação e Mercado	319
CQ - 17º Curso de Qualidade.....	321
CTD - 29º Concurso de Teses e Dissertações	345
CTIC - 35º Concurso de Trabalhos de Iniciação Científica	447
DesafIE! - 4º Workshop de Desafios da Computação aplicada à Educação.....	551
ENCompIF - 3º Encontro Nacional de Computação dos Institutos Federais.....	684
ETC - 1º Encontro de Teoria da Computação	764
Failcon - 5º Failcon Brazil	907
FIBRE - 1º Workshop do testbed FIBRE	910
JAI - 35º Jornadas de Atualização em Informática	911
SBCUP - 8º Simpósio Brasileiro de Computação Ubíqua e Pervasiva.....	916
SBSC - 13º Simpósio Brasileiro de Sistemas Colaborativos.....	1236
SECOMU - 42º Seminário de Computação na Universidade	1607
SEMISH - 43º Seminário Integrado de Software e Hardware	1613
WCAMA - 7º Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais.....	1842
WCN - 1º Workshop on Cloud Networks	1917
WEI - 24º Workshop sobre Educação em Computação	1984
WIM - 16º Workshop de Informática Médica	2452
WIT - 10º Women in Information Technology	2629
Wperformance - 15º Workshop em Desempenho de Sistemas Computacionais e de Comunicação.....	2751
WPIETFIRTF - 3º Workshop Pré-IETF.....	2863

Mensagem da Coordenação Geral

A Computação permeia muitas atividades da sociedade, seja controlando partes do carro que utilizamos para nos deslocar todos os dias, seja para nos auxiliar quando precisamos encontrar algum serviço utilizando nosso celular, ou mesmo nos ajudando quando temos um problema de saúde e algum equipamento é utilizado para encontrá-lo. Estes são somente alguns exemplos de atividades que a Computação tem ajudado nos últimos anos, mas podemos incluir qualquer área do conhecimento, como Engenharia (cálculos e projetos de prédios/casa, ou mesmo de carros, robôs, equipamentos para extração de petróleo, ...), Direito (busca por processos, ...), Odontologia (projeto de próteses, ...), Medicina (análise de imagens, descoberta de curas para doenças como câncer, ...), Biologia (descoberta de novas espécies, mapeamento genético, funcionamento do corpo humano, ...), Sociologia (análise do comportamento humano, ...), Química (descoberta de novos produtos, ...), Letras (facilidade de escrita de livros, aprendizagem de línguas estrangeiras, ...), ou seja, pense em algo e é bem provável que, de alguma forma, a Computação esteja auxiliando.

Em 2016, a Sociedade Brasileira de Computação (SBC) e a Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) realizaram, entre 3 e 7 de julho, no Centro de Eventos da PUCRS, o principal congresso da área de Computação no Brasil, o CSBC, cujo tópico principal do evento é “Computação e Interdisciplinaridade”.

Durante o CSBC, 33 atividades, entre eventos e reuniões, houve discussões relacionadas com o futuro da Computação no Brasil e como a Computação está permeando diversas atividades da sociedade. Entre os destaques podemos citar:

Seminário de Computação na Universidade (SECOMU) com 3 painéis que tiveram como foco: a) discutir o papel das associações científicas na interdisciplinaridade, com participação da Sociedade Brasileira de Informática em Saúde, Sociedade Brasileira de Matemática e da Federação Internacional para Processamento da Informação; b) debater sobre o papel do raciocínio computacional como instrumento para a interdisciplinaridade, com participação de professores do MIT (EUA), Unicamp, UFRGS e UFSB; c) discutir as questões sobre a governança da Internet, no qual o Marco Civil da Internet tem papel fundamental. Nesta discussão participaram o Conselheiro do Comitê Gestor da Internet no Brasil, Demi Getschko, o ex-diretor do ICANN (responsável pelos nomes dos endereços na Internet no mundo), Fadi Chehadé, e o gerente geral para as diferentes funções da governança da Internet na América Latina e Caribe, Andres Piazza.

Concurso de Teses e Dissertações (CTD), que reuniu as principais teses de doutorado e dissertações de mestrado na área de Computação no Brasil. Este é um dos prêmios mais valorizados na área acadêmica da Computação no Brasil.

Diversas palestras e painéis com pesquisadores nacionais e internacionais, com profissionais de empresas, entre as quais podemos ressaltar: a) palestra, durante o Workshop de Ensino de Computação (WEI), do professor Jeffrey Forbes, da Universidade de Duke (EUA), onde foi discutido o programa “Ciência da Computação para Todos”. Este programa é um esforço dos EUA anunciado pelo Escritório de Políticas de Ciência e Tecnologia (OSTP), e coordenado pela Fundação Nacional de Ciências (NSF) e pelo Departamento de Educação dos EUA; b) painel, durante o Workshop de Ensino de Computação (WEI), sobre o ensino de Computação na Educação Básica e como Computação pode ou deve ser inserido na Base Nacional Comum Curricular, assim como é já feito em diversos países no mundo; c) palestra, durante o Simpósio Brasileiro de Sistemas Colaborativos (SBSC) em cooperação com o DesafIE e com o WEI, da professora Geraldine Fitzpatrick, da Universidade de Tecnologia de Viena, Áustria, sobre suporte tecnológico para aprendizado de habilidades sócio emocionais para interação em trabalho cooperativo auxiliado por computador, no qual o foco é habilitar pessoas a colaborar e se comunicar utilizando os avanços da computação móvel e tecnologias de redes sociais; c) palestras realizadas no Seminário de Computação e Mercado (COMPUTEC) sobre Empreendedorismo, inovação, e criatividade e interdisciplinaridade, com participação de empresas como HP Inc., HPE, ThoughtWorks, Giga, SAP, Upy e DBServer.

Minicursos durante as Jornadas de Atualização em Informática (JAI) sobre Computação Urbana, Cidades Inteligentes, Agentes Inteligentes Conversacionais, Interface Cérebro-Computador e Teoria da Computação, com professores da UFMG, UFRGS, USP, Unicamp, UCS, UFPE e UFRJ.

Curso de Qualidade, voltado exclusivamente para coordenadores de curso de graduação na área de Computação, no qual foram discutidos temas como o papel da interdisciplinaridade nos cursos de graduação em Computação, oficinas sobre a gestão de conflitos e a mudança do ensino presencial para a EaD (Educação a Distância).

Evento (WIT - Women in Technology) voltado exclusivamente para discutir as questões de gênero na área de Computação, apresentando histórias de sucesso, políticas de incentivo e formas de engajamento e atração de jovens, especialmente mulheres, para as carreiras associadas à Computação.

Este ano, tivemos 940 trabalhos submetidos para os eventos que compõem o CSBC, com 303 trabalhos aceitos para serem apresentados como trabalho completo. Para que o CSBC tivesse este sucesso, foi necessário o trabalho de algumas centenas de pessoas, sejam coordenando os eventos ou participando dos comitês de programas. A eles, o muito obrigado da organização geral do CSBC.

O CSBC é também um evento que discute o futuro da SBC por meio de reuniões da Diretoria e Conselho com os Coordenadores de Comissões Especiais, com os Secretários Regionais e com os associados durante a Assembleia Geral.

Além de todas estas atividades, o CSBC é uma oportunidade única para montar uma rede de relacionamentos com cerca de 1.500 pessoas da área de Computação no Brasil, de praticamente todos os estados brasileiros, entre acadêmicos das principais universidades brasileiras e de algumas universidades internacionais, e profissionais de empresas.

Agradecemos aos nossos patrocinadores CGI.Br, Dell, EMC2, GE, ADP, DBServer, Kobe e EUBrasilCloudForum, aos órgãos de fomento CAPES, CNPq e FAPERGS, e as diversas instituições apoiadoras do CSBC. Agradecemos também à SBC e a Faculdade de Informática da PUCRS pela oportunidade e apoio para organizar um evento do porte do CSBC. E, por último, mas não menos importante, agradecemos a todos os associados da SBC.

Nestes Anais, vocês encontrarão uma apresentação dos eventos que participaram do CSBC, bem como dos trabalhos que foram apresentados durante o evento. Aproveitem.

Avelino Francisco Zorzo

Em nome da Equipe Organizadora do CSBC 2016

BraSNAM - V Brazilian Workshop on Social Network Analysis and Mining

Apresentação

É um prazer anunciar o V Brazilian Workshop on Social Network Analysis and Mining (BraSNAM). O workshop acontecerá juntamente com o 36º Congresso da Sociedade Brasileira de Computação (CSBC) o qual é o evento oficial da Sociedade Brasileira de Computação. O objetivo do encontro é prover uma oportunidade valiosa para grupos multidisciplinares encontrarem-se e engajarem-se em discussões sobre análise de redes sociais. A conferência será realizada em Porto Alegre (RS), nos dias 5 e 6 de julho de 2016.

O estudo sobre redes sociais tem suas origens em comunidades sociais, educacionais e empresariais. O interesse acadêmico nessa área tem crescido desde a metade do século 20. O recente aumento no número de usuários Web estimula a interação entre pessoas, disseminação de dados, troca de informação e também aumenta significativamente os dados disponíveis, providos por pessoas, suas interações, logs e servidores Web. Com grandes conjuntos de dados (big data), as análises podem ser mais acuradas e trazem também a oportunidade de avaliar e desenvolver novas técnicas de análise de redes sociais e identificação de dados e mineração. Isso tem despertado o interesse de um vasto conjunto de áreas - tais como academia, política, segurança, negócios, marketing, ciência - na análise de redes sociais.

Em sua quinta edição, o BraSNAM recebeu um número significativo de submissões, totalizando 66 artigos submetidos, dos quais foram aceitos 16 artigos completos e 3 artigos resumidos. O processo de revisão foi *double blind review*, tendo 3 revisores por artigo. Os melhores artigos serão publicados no Journal of the Brazilian Computer Society (JBCS).

O BraSNAM oferece uma palestra intitulada “Socially Intelligent Machines” proferida por Claudio Pinhanez (IBM). Além disso, teremos o painel “Legado digital e dados sensíveis - O que podemos e não podemos usar?” composto por Prof. Cristiano Maciel (UFMT, Moderador), Prof. Arthur Ziviani (LNCC), Karin Breitman (EMC) e Prof. Regina Ruaro (PUCRS). Neste ano, promoveremos uma maior discussão e troca de ideias. Com isto, inauguramos o “Bate Papo”, no qual um pesquisador convidado apresentará sua visão sobre uma determinada área e como os estudos das redes sociais podem auxiliá-la ou influenciá-la. Os Bate Papos ocorrerão antes das sessões técnicas e discutem os seguintes temas: Cidades Inteligentes – Prof. Lisandro Granville (UFRGS), Privacidade – Prof. Luciano Antonio Digiampietri (USP/Leste), Educação – Prof. José Palazzo M. de Oliveira (UFRGS) e Ecossistemas de Software – Prof. Rodrigo Pereira dos Santos (UNIRIO).

O V Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2016) reunirá pesquisadores e profissionais interessados na análise de rede social e áreas afins, e promoverá a colaboração, bem como o intercâmbio de ideias e experiências.

Nós gostaríamos de agradecer aos organizadores do CSBC 2016, ao diretor de publicações da SBC, Prof. José Viterbo Filho, à Springer, aos periódicos i-Sys e JIDM, aos membros do Comitê de Programa, aos revisores, aos membros do Comitê Diretivo, aos palestrantes, autores e participantes do evento. Tenham um excelente evento!

Jonice de Oliveira Sampaio (UFRJ)

Giseli Rabello Lopes (UFRJ)

Comitê de Organização

Coordenação Geral / Coordenação do Comitê de Programa

Jonice de Oliveira Sampaio (UFRJ)
Giseli Rabello Lopes (UFRJ)

Coordenação Local

Isabel Manssour (PUCRS)

Comitê Diretivo

Daniel Figueiredo (UFRJ)
Jano Moreira de Souza (UFRJ)
Jonice Oliveira (UFRJ)
Juliana Valério (UFRJ)
Mirella Moro (UFMG)
Renata Galante (UFRGS)
Roberto Imbuzeiro Oliveira (IMPA)
Fabrício Benevenuto (UFMG)
Li Weigang (UnB)

Comitê de Programa

Adriano César Pereira (UFMG)
Adriano Veloso (UFMG)
Alexandre Duarte (UFPB)
Altigran Soares da Silva (UFAM)
Ana Paula Appel (IBM Research - Brazil)
Ana Paula Couto da Silva (UFMG)
Anderson Ferreira (UFOP)
Bernardo Pereira Nunes (PUC-Rio)
Cassio Prazeres (UFBA)
Claudio de Farias (UFRJ)
Daniel Figueiredo (UFRJ)
Daniel Menasche (UFRJ)
Eduardo Borges (FURG)
Elizeu Santos-Neto (Google)
Fabrício Benevenuto (UFMG)
Genaina Rodrigues (UnB)
Gisele Pappa (UFMG)
Giseli Lopes (UFRJ)

Gustavo Paiva Guedes (CEFET/RJ)
Hernane Pereira (SENAI CIMATEC)
Humberto Marques (PUC Minas)
Isabela Gasparini (UDESC)
Jairo Souza (UFJF)
Jonice Oliveira (UFRJ)
José Palazzo Moreira de Oliveira (UFRGS)
Juliana Valério (UFRJ)
Jussara Almeida (UFMG)
Leandro Wives (UFRGS)
Leila Weitzel (UFF)
Li Weigang (UnB)
Luciana Romani (Embrapa Informática Agropecuária)
Luciano Digiampietri (USP)
Luiz André Paes Leme (UFF)
Luiz Henrique Merschmann (UFOP)
Marco Antonio Casanova (PUC-Rio)
Mirella Moro (UFMG)
Nazareno Andrade (UFCG)
Pedro Olmo Vaz de Melo (UFMG)
Renata Galante (UFRGS)
Ricardo Prudêncio (UFPE)
Sabrina Marczak (PUCRS)
Sean Siqueira (UNIRIO)
Sergio Manuel Serra da Cruz (UFRRJ)
Tatiana Tavares (UFPE)
Valeria Bastos (UFRJ)

Revisores Externos

Eduardo Nakamura (UFAM)
Karina S. Machado (FURG)
Ricardo Ferreira (UFV)

Painéis / Palestras

Socially Intelligent Machines

Claudio Pinhanez (IBM)

Nesta palestra, serão exploradas as novas fronteiras da Computação Cognitiva, onde máquinas tentam exibir uma inteligência social. Serão explorados as tecnologias e desafios em fazer máquinas capazes de participar em interações do tipo “multi-agent:person conversations”, em minerar e contextualizar o conhecimento disseminado nas mídias online e também em construir modelos de grupos e seus relacionamentos, organizações e sociedade. O desafio de prover habilidades sociais a computadores será discutido no contexto de algumas aplicações reais, envolvendo gestão financeira, análise de imagens e robótica.

Legado digital e dados sensíveis - O que podemos e não podemos usar?

Cristiano Maciel (UFMT) - *moderador*

Artur Ziviani (LNCC)

Karin Breitman (EMC)

Regina Ruaro (PUCRS)

Dados sensíveis são informações pessoais que podem causar discriminação ou estigmatização ao seu titular quando acessadas, inferidas e divulgadas sem autorização, tais como informações que revelem a origem racial ou étnica do titular, suas convicções religiosas, filosóficas ou morais, orientação sexual, doenças, rotas frequentes, dentre outras. Muitos destes dados podem ser obtidos e extraídos das mídias sociais, especialmente quando analisamos o relacionamento entre as pessoas e como elas interagem. Neste painel discutiremos o estado atual, principais desafios e problemas do tratamento e uso de dados sensíveis (em vida e após a morte do usuário) no Brasil.

Trabalhos aceitos

Trabalhos Completos

Análise da Relação entre Obtenção de Bolsas de Produtividade do CNPq e Medidas Bibliométricas e de Análise de Redes Sociais

Felipe P. C. Fonseca (USP)

Luciano A. Digiampietri (USP)

Análise e Ranqueamento da Rede de Advogados induzida por Processos Judiciais Trabalhistas

Leonardo Filipe Rodrigues Ribeiro (UFRJ)

Daniel R. Figueiredo (UFRJ)

Paulo Roberto Nascimento (UFRJ)

Avaliação Automática da Utilidade de Reviews usando Redes Neurais Artificiais no Corpus do Steam

Jardeson L. N. Barbosa (UFPI)

Raimundo S. Moura (UFPI)

Caracterização dos Perfis Comerciais na Rede Social Instagram

Gabriela Enes Campos (UFOP)

Helen Costa (UFOP)

Caracterizando Departamentos e Programas de Computação Utilizando Análise de Redes Sociais e Bibliometria

Luciano Digiampietri (USP)

Ricardo Linden (FSMA)

Lenin Barbosa (USP)

Co-authorship Prediction in Academic Social Network

William Takahiro Maruyama (USP)

Luciano Antonio Digiampietri (USP)

Comparing and Recommending Conferences

Grettel Monteagudo García (PUC-Rio)

Bernardo Pereira Nunes (PUC-Rio)

Giseli Rabello Lopes (UFRJ)

Marco Antonio Casanova (PUC-Rio)

Luiz André Paes Leme (UFF)

Desenvolvimento de um Modelo Semântico para Recomendação Baseado em Grafos

Hugo Guércio Fernandes (UFJF)

Victor Ströele (UFJF)

Effects of Social Ties in Knowledge Diffusion: Case Study on PLOS ONE

Felipe Eltermann (Unicamp)

Alan Godoy (Unicamp)

Fernando J. Von Zuben (Unicamp)

Extração de Métricas e Análise de Sentimentos em Comentários Web no Domínio de Hotéis

Roney L. de S. Santos (UFPI)

Raimundo S. Moura (UFPI)

Formal Information Flows Among Top Authorities of the Brazilian Federal Government based on Co-word Analysis of Data Published in the Official Gazette

Rafael Henrique Santos Soares (UnB)

Jorge H. C. Fernandes (UnB)

Ricardo Sampaio (UnB)

How Congressmen Connect: Analyzing Voting and Donation Networks in the Brazilian Congress

Victor S. Bursztyn (UFRJ)

Marcelo Granja Nunes (UFRJ)

Daniel R. Figueiredo (UFRJ)

Predição de Relacionamentos em Redes Sociais, uma Revisão Sistemática

William Takahiro Maruyama (USP)

Luciano Antonio Digiampietri (USP)

Revealing User Influence in an Online Newspaper

Gilberto Flores Pochet (UFABC)

Carlos Kamienski (UFABC)

Uma Análise da Polaridade Expressa nas Manchetes de Notícias Brasileiras

Pedro Ramos (UFMG)

Julio Reis (UFMG)

Fabício Benevenuto (UFMG)

Using Topological Properties to Measure the Strength of Co-authorship Ties

Michele A. Brandão (UFMG)

Matheus A. Diniz (UFMG)

Mirella M. Moro (UFMG)

Resumos

Detecção de Casos de Violência Patrimonial a partir do Twitter

João Paulo Clarindo dos Santos (UFAL)

Fábio Coutinho (UFAL)

André Lage Freitas (UFAL)

Inferindo o Sexo de Usuários de Redes Sociais utilizando o LIWC em Português do Brasil

Luiz Antonio da Ponte Junior (CEFET/RJ)

Gustavo Paiva Guedes (CEFET/RJ)

Eduardo Bezerra (CEFET/RJ)

Análise da relação entre obtenção de bolsas de produtividade do CNPq e medidas bibliométricas e de análise de redes sociais

Felipe P. C. Fonseca¹, Luciano A. Digiampietri¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
São Paulo – SP – Brazil

felipe.fonseca@usp.br, digiampietri@usp.br

Abstract. *There are different ways of encouraging scientific work. One of them is the research productivity grants given by CNPq. However, with the scarcity of resources, analyze the performance of researchers and select those who will receive the grants are becoming even more complex and challenging activities. Thus, this paper aims to build a classifier by using the Lattes Platform as a data source that can identify the grant level of a given researcher and help researchers to contextualize themselves in relation to other scholars, considering social network and bibliometric analyses.*

Resumo. *Existem diferentes formas de incentivo ao trabalho científico. Uma delas é a concessão de bolsas de produtividade em pesquisa por parte do CNPq. Porém, com a escassez de recursos, a análise do desempenho dos pesquisadores e seleção daqueles que receberão as bolsas são atividades cada vez mais complexas e desafiadoras. Desta forma, este artigo visa a construir um classificador, usando currículos da Plataforma Lattes como fonte de dados, que consiga identificar qual o nível da bolsa de um dado pesquisador e que o auxilie a se contextualizar em relação aos bolsistas considerando medidas bibliométricas e métricas oriundas da análise de redes sociais.*

1. Introdução e Motivação

Atualmente, o governo brasileiro incentiva a área acadêmica e a pesquisa por meio de instituições de fomento, tais como a CAPES e o CNPq. Dentre as diversas formas de se promover esse incentivo existe a modalidade de bolsas e auxílios, que tem como público alvo tanto alunos do ensino médio, alunos de graduação e de pós quanto doutores-pesquisadores. O foco desse artigo será nos doutores-pesquisadores.

A pequena disponibilidade de bolsas combinada com um cenário econômico desfavorável e certa subjetividade inerente ao processo torna a análise da produtividade acadêmica dos pesquisadores e a seleção dos que serão contemplados com as bolsas de produtividade em pesquisa atividades cada vez mais importantes e desafiadoras. Portanto, conseguir identificar pesquisadores com potencial para receber essas bolsas é uma tarefa relevante, pois pode ajudar o governo a distribuir de forma mais clara as bolsas fazendo com que o planejamento desse programa seja mais eficaz. Adicionalmente, não existem muitas informações para auxiliar um pesquisador que está pensando em pleitear uma dessas bolsas a identificar o contexto em que se encontra em relação aos demais pesquisadores e, em particular, àqueles possuidores de bolsa de produtividade em pesquisa.

Contudo, são poucas as pesquisas na área de identificação de potenciais bolsistas. Assim sendo, este projeto visa a construir um classificador usando os dados dos doutores na área de Ciências da Computação disponíveis a partir da plataforma Lattes¹ para identificar o nível da bolsa de um dado pesquisador, bem como permitir que um pesquisador consiga ter uma visão contextualizada de algumas de suas métricas em relação às métricas dos bolsistas.

2. Trabalhos Relacionados

Esta seção descreve trabalhos correlatos ao trabalho atual que focam tanto na avaliação de instituições de ensino e pesquisa como na avaliação de pesquisadores.

Uma análise da produtividade de uma instituição como um todo pode ser conferida em [Tess et al. 2009]. Esse artigo faz a avaliação da produção do Instituto do Coração, uma das 13 instituições da escola de medicina da Universidade de São Paulo. Destaca-se a abordagem usada, que se baseia em indicadores bibliométricos para fazer uma avaliação de toda a produção científica da instituição no período de janeiro de 2000 até dezembro de 2003.

Lima et al. (2015) fizeram a avaliação de bolsas produtividade e seu impacto sobre o desempenho dos bolsistas. Nesse trabalho é apresentada uma explicação detalhada de como um pesquisador aplica para uma bolsa e como ele é avaliado. Basicamente existem seis principais classes de bolsas produtividade em pesquisa fornecidas pelo CNPq: Sênior, 1A, 1B, 1C, 1D e 2.

Com o intuito de classificar um pesquisador candidato, cada comitê de avaliação analisa o projeto de pesquisa e o currículo do proponente. Tal currículo possui: suas publicações (artigos em revistas, artigos em conferências, livros, etc.); supervisão de estudantes (incluindo mestrado e doutorado); contribuições para a ciência, tecnologia e inovação (incluindo patentes); coordenação e participação de projetos de pesquisa; inserção internacional; participação como avaliador de conferências; e outras atividades relacionadas à ciência e à academia [Lima et al. 2015]. Contudo, os autores informam como índices de produtividade ou qualidade (por exemplo, o JCR e o Qualis usados no presente artigo) devem ser avaliados, respeitando as diversas áreas de atuação do pesquisador, a fim de ajudar a produzir uma análise da produtividade de uma forma mais imparcial e contextualizada.

Uma análise dos pesquisadores que possuem bolsas produtividade utilizando-se de uma rede social de coautorias para o cálculo das métricas já foi feita anteriormente [Andrade 2015]. Andrade e Régo mostram que métricas estruturais de redes sociais (de centralidade) têm uma correlação positiva com o nível do desempenho dos bolsistas. Conforme apresentado pelos autores “*as métricas de centralidade de grau, centralidade de proximidade, centralidade de intermediação e centralidade de autovetor são importantes atributos estruturais na rede social e estão relacionadas com a eficiência, liderança e satisfação*” [Andrade 2015].

Digiampietri et al. (2014) apresentam uma análise da participação dos orientandos na produção dos seus orientadores por meio de uma rede de coautoria (da mesma forma que o artigo citado anteriormente). O mesmo autor e seus colaboradores analisam

¹<http://lattes.cnpq.br/>

redes sociais de pesquisadores por meio de subredes regionais de forma a se identificar as características de relacionamentos entre os doutores [Digiampietri et al. 2014a].

3. Metodologia

A metodologia usada se baseia na proposta por Digiampietri et al. (2013). No presente trabalho, algumas tarefas foram executadas (na forma de fases) a fim de se obter os dados necessários para a classificação dos pesquisadores. A primeira tarefa foi a obtenção de dados propriamente dita e a transformação dos mesmos (por meio de um *parser*) para torná-los mais fáceis de se manejar. Nessa fase as medidas bibliométricas de cada pesquisador foram obtidas.

Após isso, uma rede social foi construída (representada como um grafo) utilizando-se as relações de coautorias como arestas. Tal rede foi utilizada para a obtenção das medidas de centralidade. Por meio dos dados obtidos nas fases anteriores alguns testes, usando vários classificadores disponíveis no Weka, foram feitos para se avaliar o desempenho da classificação considerando o conjunto de medidas extraídas ou calculadas.

3.1. Obtenção e tratamento de dados

Foi escolhido como amostra o conjunto formado por todos os doutores que atuam em Ciência da Computação e possuem currículo Lattes². Este conjunto é composto de 7.469 doutores. A fase de obtenção de dados é aquela responsável por transformar os dados brutos em formato XML dos Currículos Lattes dos doutores da amostra em um arquivo de texto tabulado, que posteriormente foi utilizado para montar o conjunto de dados. Note que nesse primeiro momento, as informações obtidas são referentes a cada pesquisador, sendo elas de carácter pessoal (nome, número de identificação do Lattes, entre outros) ou de carácter bibliométrico (número de publicações, número de orientações, etc.). Além disso, dados sobre as produções científicas foram obtidos (revistas e congressos em que cada artigo foi publicado, e os coautores) para serem usadas tanto na obtenção de índices de impacto referentes aos veículos de publicação quanto na construção da rede social de coautoria.

Foram analisados os dados de produção e a rede de coautoria considerando as publicações e orientações no período de 2010 e 2014. A relação de bolsistas produtividade foi extraída do site do CNPq, considerando os pesquisadores que possuíam bolsa vigente na área de Ciência da Computação em 2014.

3.2. Construção da rede social usando coautorias como arestas

Como citado no seção 2, a abordagem de se construir uma rede social usando coautorias como as arestas não é nova [Digiampietri et al. 2014a, Andrade 2015]. Nesta abordagem, a rede social é representada por um grafo não direcionado em que cada nó representa um pesquisador e cada aresta representa uma coautoria entre os pesquisadores (coautoria em publicações científicas).

Os algoritmos para resolução de entidades e construção da rede de coautoria utilizados neste trabalho foram aqueles disponibilizados por Digiampietri et al. (2014b). Já os algoritmos para o cálculo das métricas estudadas foram os disponibilizados por Digiampietri et al. (2016).

²<http://lattes.cnpq.br/>

3.3. Construção do conjunto e classificadores testados

Partindo-se dos dados coletados e/ou calculados nas fases anteriores, um conjunto de dados foi construído utilizando as medidas bibliométricas e de centralidade (considerando a rede social acadêmica [Digiampietri et al. 2012b]) de cada pesquisador [Wasserman and Galaskiewicz 1994, Lemieux and Ouimet 2008, Poblacion et al. 2009, Prell 2012]. Os atributos do conjunto podem ser consultados na tabela 1, junto das siglas que serão utilizadas no decorrer deste artigo.

Tabela 1. Atributos extraídos ou calculados.

Atributo	Descrição
Bolsista (BOL)	Adota valores entre 0 (o pesquisador não possui bolsa) ou 1 (ele possui).
Nível (NIV)	Adota valores entre 0 e 6. Além de representar se o pesquisador é bolsista ou não, denota o nível da bolsa do pesquisador (sendo 0 atribuído ao pesquisador sem bolsa; 1 ao pesquisador com bolsa nível 2; até 5 ao pesquisador 1A e 6 ao bolsista sênior.)
Dissertações de mestrado concluídas (DMC)	Representa o número de orientações em dissertações de mestrado concluídas.
Iniciações científicas concluídas (ICC)	Representa o número de orientações em iniciações científicas completas.
Monografias de conclusão de curso em aperfeiçoamento e especialização concluídas (MAC)	Representa o número de orientações em monografias de aperfeiçoamento e especialização concluídas.
Trabalhos de conclusão de curso concluídos (TCCC)	Número de orientações em TCCs concluídos.
Teses de doutorado completos (TDC)	Número de orientações em teses de doutorado concluídas.
Orientações de outra natureza concluídas (ONC)	Representa o número de orientações de qualquer natureza diferentes das citadas anteriormente, concluídas.
Supervisões de Pós-Doutorado concluídas (SPD)	Número de supervisões de pós-doutorado concluídas.
Iniciações científicas em andamento (ICA)	Número de orientações em iniciações científicas em andamento.
Trabalhos de conclusão de curso em andamento (TCCA)	Número de orientações em TCCs em andamento.
Monografias de conclusão de curso em aperfeiçoamento e especialização em andamento (MAA)	Representa o número de orientações em monografias de aperfeiçoamento e especialização em andamento.
Orientações de outra natureza em andamento (ONA)	Representa o número de orientações de qualquer natureza diferentes das citadas anteriormente em andamento.
Degree (DGR)	Medida de centralidade de grau da rede social construída na seção 3.2.
Betweenness (BTW)	Medida de centralidade de intermediação da rede social construída na seção 3.2
Closeness (CIS)	Medida de centralidade de proximidade da rede social construída na seção 3.2
Strength (STR)	Medida da força que o nó representado pelo pesquisador possui na rede social.
PageRank (PRK)	Valor do PageRank que o nó do pesquisador possui na rede social.
Número de Eventos (NDE)	Número de artigos completos publicados em anais de eventos pelo pesquisador.
Número de Periódicos (NDP)	Número de artigos publicados em periódicos pelo pesquisador.
JCR Periódicos (JCR)	Valor do JCR para os periódicos nos quais pesquisador publicou.
Qualis Periódicos (QP)	Valor do Qualis para os periódicos nos quais o pesquisador publicou.
Qualis Conferências (QC)	Valor do Qualis para as conferências nas quais o pesquisador publicou.
Qualis Total (QT)	Valor do Qualis total para o pesquisador.

O próximo passo foi avaliar (por meio de um estudo exploratório) o desempenho de vários classificadores usando o conjunto de dados construído. Os classificadores avaliados foram os seguintes: *Naive Bayes*, *Árvore de Decisão*, *Random Tree*, *Rede Bayesiana*, *Adaboost* e *Decision Table*. Estes classificadores foram escolhidos por representarem classificadores de diferentes tipos/famílias.

4. Análise dos Resultados

Os resultados dos classificadores foram obtidos utilizando o arcabouço Weka³ ou por meio de sua interface gráfica ou usando sua API num aplicativo desenvolvido em Java.

³<http://www.cs.waikato.ac.nz/ml/weka/>

É importante mencionar que dois teste foram feitos inicialmente sobre o conjunto. O primeiro avaliou o desempenho dos algoritmos usando como classe o nível da bolsa, e o outro avaliou se o pesquisador é bolsista ou não, sendo este o principal objetivo do trabalho.

Para o segundo caso, pelo fato do conjunto de dados ser desbalanceado (apenas 4,9% dos pesquisadores da amostra são bolsistas) foram realizados testes não balanceando e balanceando o conjunto de treinamento (utilizando a estratégia de *resampling*). Por esse motivo a API do Weka foi necessária, por não existir nenhuma implementação de *resampling* disponível no Weka da forma necessária ao presente trabalho. Tal aplicação computa a revocação (*recall*), Medida-F (*f-measure*) e precisão (*precision*) de cada classe para verificar se, mesmo com uma eventual perda de acurácia, o desempenho dos algoritmos melhora nesses outros indicadores. Todos os resultados apresentados nesta seção constituem a média dos resultados utilizando a estratégia de validação cruzada com 10 subconjuntos (*10-fold cross validation*).

Com o intuito de estruturar melhor os resultados, esta seção está organizada em quatro subseções: (i) uma que diz respeito aos resultados acerca do nível da bolsa; (ii) uma com resultados com respeito se o pesquisador é bolsista ou não mantendo o conjunto de dados original; (iii) uma que faz a mesma análise, mas utilizando-se da técnica de *resampling* para tentar obter melhores resultados; e (iv) uma avaliando o impacto que os atributos do conjunto de dados têm sobre os resultados dos classificadores, por meio do cálculo das correlações entre os mesmos e da análise de componentes principais (*Principal Component Analysis*).

4.1. Resultados - Nível da Bolsa

Os algoritmos testados nessa fase foram os seguintes: *Naive Bayes*, Árvore de Decisão, *Random Tree*, Rede Bayesiana e Adaboost. Em cada um deles as medidas de *acurácia*, *revocação*, *precisão* e *Medida-F* foram cálculos. Os resultados podem ser vistos nas subseções a seguir.

4.1.1. Árvore de Decisão

Os resultados para a árvore de decisão podem ser observados na tabela 2. Note que a acurácia geral deste classificador foi de 94,2964%, o que não é muito alta, visto que a classe dominante (não bolsistas) representa 95,06% dos dados. Desta forma, classificar todos os doutores como não bolsistas teria um desempenho acima deste classificador (considerando apenas esta métrica).

4.1.2. *Naive Bayes*

Na tabela 3 são apresentados os resultados para o classificador *Naive Bayes*. A acurácia geral foi de 84,362%, sendo o pior resultado entre os classificadores. Porém, mesmo com a acurácia baixa, vemos a revocação para a classe Bolsa Tipo 2 consideravelmente alta, visto que não foi realizada nenhuma abordagem de balanceamento para esse teste.

Tabela 2. Resultados dos testes usando a técnica de validação cruzada com 10 subconjuntos (Árvore de Decisão).

classe	precisão	revocação	Medida-F
Não Bolsista	0,97	0,982	0,976
Bolsa Tipo 2	0,322	0,243	0,277
Bolsa Tipo 1D	0,098	0,08	0,088
Bolsa Tipo 1C	0,057	0,056	0,056
Bolsa Tipo 1B	0,182	0,083	0,114
Bolsa Tipo 1A	0,214	0,13	0,162
Bolsa Tipo SR	0	0	0

Tabela 3. Resultados dos testes usando a técnica de validação cruzada com 10 subconjuntos (Naive Bayes).

classe	precisão	revocação	Medida-F
Não Bolsista	0,987	0,894	0,939
Bolsa Tipo 2	0,212	0,545	0,305
Bolsa Tipo 1D	0,161	0,3	0,21
Bolsa Tipo 1C	0,114	0,278	0,161
Bolsa Tipo 1B	0,08	0,167	0,108
Bolsa Tipo 1A	0,026	0,043	0,033
Bolsa Tipo SR	0	0	0

4.1.3. Random Forest

A acurácia geral do algoritmo *Random Forest* foi de 95,1533%, o que é um pouco melhor do que os 95,06% da classe dominante (não bolsistas). Uma limitação desse algoritmo e de praticamente todos avaliados nessa seção é que a revocação para as classes não dominantes é consideravelmente baixa (a melhor delas é de 16,6%).

Tabela 4. Resultados dos testes usando a técnica de validação cruzada em 10 subconjuntos (Random Forest).

classe	precisão	revocação	Medida-F
Não Bolsista	0,961	0,994	0,978
Bolsa Tipo 2	0,419	0,166	0,238
Bolsa Tipo 1D	0,278	0,1	0,147
Bolsa Tipo 1C	0	0	0
Bolsa Tipo 1B	0,286	0,083	0,129
Bolsa Tipo 1A	0,167	0,043	0,069
Bolsa Tipo SR	0	0	0

4.1.4. Rede Bayesiana

Assim como o Naive Bayes, a outra abordagem avaliada que utiliza o Teorema de Bayes como base para a classificação (Rede Bayesiana), não teve uma acurácia muito alta: 87,0933%. Porém, assim como com o Naive Bayes, observa-se a revocação das outras classes aumentar consideravelmente, sendo o melhor caso, para a classe Bolsa Tipo 2, com um valor de 62,6%.

Tabela 5. Resultados dos testes usando a técnica de validação cruzada em 10 subconjuntos (Rede Bayesiana).

classe	precisão	revocação	Medida-F
Não Bolsista	0,993	0,892	0,94
Bolsa Tipo 2	0,169	0,626	0,266
Bolsa Tipo 1D	0,16	0,32	0,213
Bolsa Tipo 1C	0,029	0,028	0,028
Bolsa Tipo 1B	0,042	0,083	0,056
Bolsa Tipo 1A	0,133	0,174	0,151
Bolsa Tipo SR	0	0	0

4.1.5. Adaboost

De todos os algoritmos avaliados, o que obteve o pior desempenho geral foi o Adaboost que classificou todos os pesquisadores como não bolsistas.

Tabela 6. Resultados dos testes usando a técnica de validação cruzada em 10-subconjuntos (Adaboost).

classe	precisão	revocação	Medida-F
Não Bolsista	0,951	1	0,975
Bolsa Tipo 2	0	0	0
Bolsa Tipo 1D	0	0	0
Bolsa Tipo 1C	0	0	0
Bolsa Tipo 1B	0	0	0
Bolsa Tipo 1A	0	0	0
Bolsa Tipo SR	0	0	0

4.2. Resultados - Bolsista ou não bolsista (sem *Resampling*)

Os resultados desta seção foram obtidos por meio de um aplicativo Java desenvolvido neste trabalho usando a API do Weka para acesso aos classificadores. Para os experimentos apresentados nesta seção não foi utilizada nenhuma estratégia de balanceamento do conjunto de treinamento, assim não são esperados bons resultados de revocação (*recall*) da classe bolsista, visto que as classes são desbalanceadas. Os resultados podem ser vistos na tabela 7.

O classificador que obteve a maior acurácia foi o *Rotation Forest* (96,18%), obtendo também a maior precisão para a classe positiva (65,55%). Em termos de revocação, destaca-se a Rede Bayesiana (87,80%). Por fim, o algoritmo que obteve a maior medida-F da classe positiva, medida que tenta ponderar a precisão e a revocação, foi o Adaboost, com 0,5544 (valor muito próximo ao obtido pelo classificador *Rotation Forest*).

4.3. Resultados - Bolsista ou não bolsista (com *resampling*)

Nesta seção são apresentados os resultados da classificação usando *resampling* como estratégia de balanceamento do conjunto de treinamento. Os valores de *recall*, *f-measure*, *precision* (por classe) e acurácia global foram calculados, usando como base os falsos positivos e negativos e os verdadeiros positivos e negativos. Pode-se conferir os resultados na tabela 8.

Tabela 7. Resultados de todos os classificadores usando o fato de ser bolsista ou não como classe (sem *resampling*).

		Recall	Precision	F-Measure	Acurácia
Árvore de Decisão	F	0,9803	0,97261	0,97643	0,9550
	T	0,4688	0,5527	0,5073	
Random Forest	F	0,9869	0,9733	0,9801	0,9618
	T	0,4797	0,6555	0,5540	
Naive Bayes	F	0,9290	0,9858	0,9565	0,9198
	T	0,7425	0,3521	0,4778	
Rede Bayesiana	F	0,9015	0,9930	0,9451	0,9004
	T	0,8780	0,3167	0,4655	
Adaboost	F	0,9667	0,9804	0,9735	0,9501
	T	0,6287	0,4957	0,5544	
Decision Table	F	0,9801	0,9727	0,9764	0,9550
	T	0,4715	0,5523	0,5087	

Tabela 8. Resultados de todos os classificadores usando o fato de ser bolsista ou não como classe (com *resampling*).

		Recall	Precision	F-Measure	Acurácia
Árvore de Decisão	F	0,9446	0,9832	0,9636	0,9321
	T	0,6910	0,3935	0,5015	
Random Forest	F	0,9613	0,9873	0,9741	0,9514
	T	0,7615	0,5054	0,6076	
Naive Bayes	F	0,9117	0,9871	0,9479	0,9048
	T	0,7723	0,3125	0,4450	
Rede Bayesiana	F	0,8880	0,9943	0,9381	0,8887
	T	0,9024	0,2952	0,4449	
Adaboost	F	0,8939	0,9953	0,9419	0,8952
	T	0,9187	0,3104	0,4641	
Decision Table	F	0,8666	0,9935	0,9257	0,8679
	T	0,8915	0,2578	0,3999	

Conforme esperado, o balanceamento do conjunto de treinamento utilizando *re-sampling* acarretou num aumento na revocação da classe positiva (isto é, bolsistas), porém com uma diminuição de precisão da classe positiva e também da acurácia geral. Destaca-se a revocação da classe positiva de 91,87% obtida pelo classificador Adaboost. Este tipo de abordagem é útil quando se deseja aumentar a revocação (com eventual perda de precisão) para, por exemplo, maximizar as chances dos indivíduos da classe minoritária serem efetivamente classificados como tal.

4.4. Análise dos Atributos

Para realizar uma análise detalhada sobre os atributos e suas influências sobre o resultado dos classificadores, serão usadas as correlação entre os atributos e também será realizada a Análise de Componentes Principais (*Principal Component Analysis*).

A correlação entre todos os atributos usados na classificação pode ser vista na tabela 9. Observe que os dois primeiros atributos são, respectivamente, o fato de o pesquisador ser bolsista ou não (0 para caso não seja bolsista e 1 caso seja) e o nível da bolsa (segundo a mesma representação da seção 4.1).

Como pode ser observado na primeira linha da tabela 9, as medidas de centrali-

Tabela 9. Correlação entre os atributos

	BOL	NIV	DMC	ICC	MAC	TCCC	TDC	ONC	SPD	ICA	TCCA	MAA	ONA	DGR	BTW	CLS	STR	PRK	DNE	NDP	JCR	QP	QC	QT
ROI	1,000	0,813	0,302	-0,031	0,011	0,095	0,404	-0,016	0,139	0,044	-0,014	-0,015	-0,005	0,418	0,381	0,196	0,418	0,401	0,333	0,213	0,100	0,532	0,398	0,526
NIV	0,813	1,000	0,244	-0,030	-0,013	0,065	0,400	-0,015	0,125	0,022	-0,019	-0,014	-0,004	0,378	0,342	0,160	0,378	0,365	0,290	0,226	0,110	0,552	0,351	0,507
DMC	0,302	0,244	1,000	0,054	0,187	0,302	0,551	0,062	0,166	0,148	0,033	0,010	0,013	0,431	0,376	0,285	0,431	0,474	0,647	0,377	0,087	0,331	0,422	0,438
ICC	-0,031	-0,030	0,054	1,000	0,157	0,055	-0,020	0,047	-0,010	0,025	0,076	0,177	-0,005	-0,016	-0,009	-0,014	-0,016	-0,011	0,072	0,043	-0,023	-0,042	-0,034	-0,043
MAC	0,011	-0,013	0,187	0,157	1,000	0,393	0,033	0,280	0,005	0,121	0,223	0,040	-0,010	0,079	0,070	0,076	0,079	0,097	0,206	0,098	-0,021	0,001	0,058	0,037
TCCC	0,095	0,065	0,302	0,055	0,393	1,000	0,143	0,371	0,075	0,356	0,100	0,022	-0,010	0,155	0,128	0,144	0,155	0,193	0,340	0,216	0,040	0,118	0,152	0,157
TDC	0,404	0,400	0,551	-0,020	0,033	0,143	1,000	0,013	0,343	0,070	-0,041	-0,021	-0,009	0,443	0,435	0,215	0,443	0,494	0,549	0,454	0,189	0,462	0,431	0,511
ONC	0,016	0,015	0,062	0,047	0,280	0,371	0,013	1,000	0,018	0,120	0,084	0,028	0,005	0,024	0,023	0,018	0,024	0,031	0,106	0,073	0,009	0,011	0,027	0,012
SPD	0,139	0,125	0,166	-0,010	0,005	0,075	0,343	0,018	1,000	0,043	-0,022	-0,009	-0,003	0,124	0,126	0,059	0,124	0,157	0,163	0,282	0,228	0,184	0,100	0,158
ICA	0,044	0,022	0,148	0,025	0,121	0,356	0,070	0,120	0,043	1,000	0,148	0,030	-0,009	0,074	0,056	0,092	0,074	0,095	0,180	0,118	0,043	0,064	0,082	0,085
TCCA	-0,014	-0,019	0,033	0,076	0,223	0,100	-0,041	0,084	-0,022	0,148	1,000	0,084	-0,006	0,035	0,019	0,062	0,035	0,042	0,071	0,013	-0,020	-0,016	0,009	-0,002
MAA	-0,015	-0,014	-0,010	0,177	0,040	0,022	-0,021	0,028	-0,009	0,030	0,084	1,000	-0,002	-0,016	-0,014	-0,006	-0,016	-0,014	0,006	0,005	-0,011	-0,021	-0,017	-0,022
ONA	-0,005	-0,004	-0,013	-0,005	-0,010	-0,010	-0,009	-0,005	-0,003	-0,009	-0,006	-0,002	1,000	-0,013	-0,007	-0,024	-0,013	-0,018	-0,014	-0,009	-0,003	-0,008	-0,009	-0,010
DGR	0,418	0,378	0,431	-0,016	0,079	0,155	0,443	0,024	0,124	0,074	0,035	-0,016	-0,013	1,000	0,805	0,546	1,000	0,910	0,575	0,268	0,083	0,498	0,573	0,619
BTW	0,381	0,342	0,376	-0,009	0,070	0,128	0,435	0,023	0,126	0,056	0,019	-0,014	-0,007	0,805	1,000	0,316	0,805	0,750	0,523	0,250	0,072	0,444	0,510	0,552
CLS	0,196	0,160	0,285	-0,014	0,076	0,144	0,215	0,018	0,059	0,092	0,062	-0,006	-0,024	0,546	0,316	1,000	0,546	0,644	0,369	0,155	0,029	0,255	0,322	0,336
STR	0,418	0,378	0,431	-0,016	0,079	0,155	0,443	0,024	0,124	0,074	0,035	-0,016	-0,013	1,000	0,805	0,546	1,000	0,910	0,575	0,268	0,083	0,498	0,573	0,619
PRK	0,401	0,365	0,474	-0,011	0,097	0,193	0,494	0,031	0,157	0,095	0,042	-0,014	-0,018	0,910	0,750	0,644	0,910	1,000	0,615	0,334	0,111	0,487	0,548	0,598
DNE	0,333	0,290	0,647	0,072	0,206	0,340	0,549	0,106	0,163	0,180	0,071	0,006	-0,014	0,575	0,523	0,369	0,575	0,615	1,000	0,361	0,057	0,414	0,638	0,617
NDP	0,213	0,226	0,377	0,043	0,098	0,216	0,454	0,073	0,282	0,118	0,013	0,005	-0,009	0,268	0,250	0,155	0,268	0,334	0,381	1,000	0,756	0,576	0,227	0,440
JCR	0,100	0,110	0,087	-0,023	-0,021	0,040	0,189	-0,009	0,228	0,043	-0,020	-0,011	-0,003	0,083	0,072	0,029	0,083	0,111	0,057	0,756	1,000	0,508	0,070	0,305
QP	0,532	0,552	0,331	-0,042	0,001	0,118	0,462	-0,011	0,184	0,064	-0,016	-0,021	-0,008	0,498	0,444	0,255	0,498	0,487	0,414	0,576	0,508	1,000	0,510	0,838
QC	0,398	0,351	0,422	-0,034	0,058	0,152	0,431	0,027	0,100	0,082	0,009	-0,017	-0,009	0,573	0,510	0,322	0,573	0,548	0,638	0,227	0,070	0,510	1,000	0,897
QT	0,526	0,507	0,438	0,043	0,037	0,157	0,511	0,012	0,158	0,085	0,002	0,022	0,010	0,619	0,552	0,336	0,619	0,598	0,617	0,440	0,305	0,838	0,897	1,000

dade da rede social (especialmente, grau, intermediação, *strength* e *PageRank*) junto do número de orientações concluídas de doutorado e as medidas Qualis são aquelas com maior correlação com os dois atributos usados na classificação. Em particular, a maior correlação foi identificada com o atributo Qualis Periódicos (QP). Assim sendo, estes atributos são aqueles que potencialmente têm grande influência sobre a classificação.

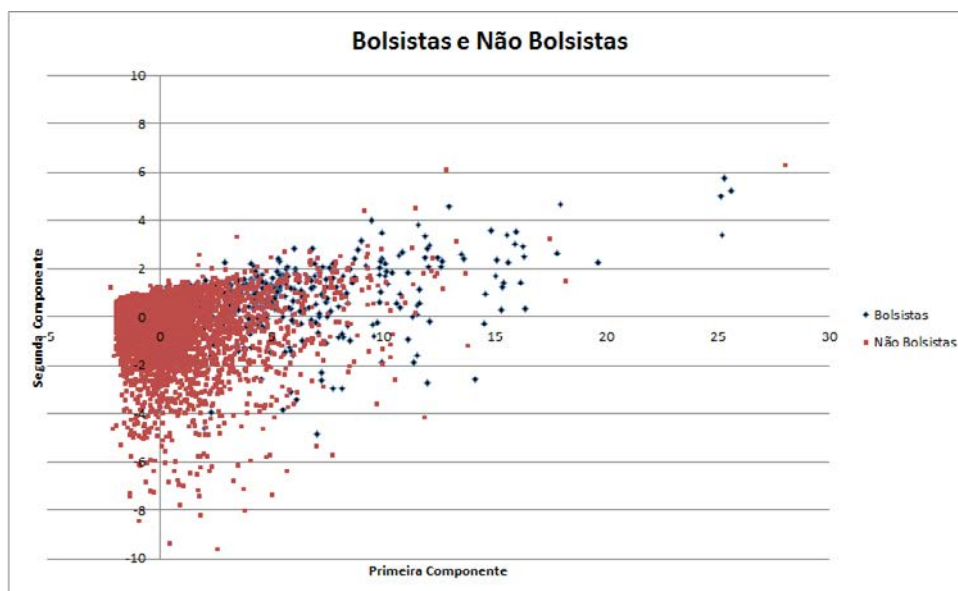


Figura 1. Plotagem das duas primeiras componentes resultantes do PCA

O próximo resultado corresponde ao gerado pela PCA (*Principal Component Analysis*). Usando os valores das duas primeiras componentes resultantes dessa abordagem como entradas para um gráfico e colorindo bolsistas e não bolsistas com uma cor diferente, tem-se as figuras 1, 2 e 3. O interessante de tais imagens é que pode-se obser-

var que os doutores não bolsistas concentram-se perto da origem do gráfico, enquanto que os bolsistas tendem a ficar mais espalhados e, em especial, com valores positivos para o eixo x do gráfico (que corresponde à componente principal resultantes da PCA).

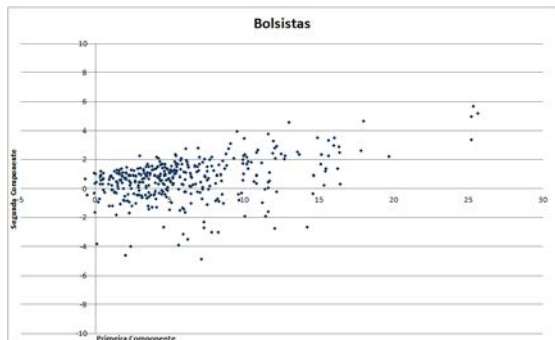


Figura 2. Plotagem das duas primeiras componentes resultantes do PCA - Bolsistas

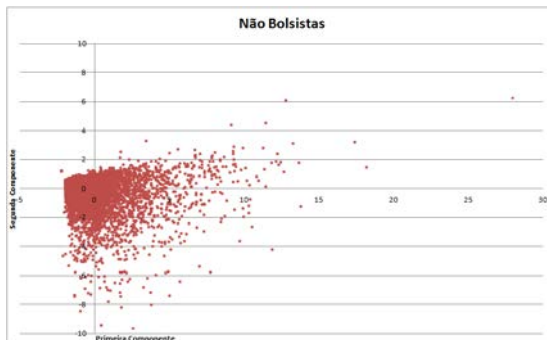


Figura 3. Plotagem das duas primeiras componentes resultantes do PCA - Não Bolsistas

Além disso, para auxiliar um dado pesquisador a conseguir se contextualizar mais facilmente em relação aos outros é gerado um gráfico posicionando as métricas desse pesquisador (aqui exemplificado por um pesquisador chamado de Pesquisador X) em relação aos demais bolsistas (note que o Pesquisador X pode ser qualquer pesquisador da área de Ciências da Computação). Essa análise é feita sobre os seus atributos e consegue mostrar o quão eficiente ou não este está em relação aos seus semelhantes, para isto, cada atributo (ou métrica) é normalizado para valores entre zero e um e a mediana e a média para cada atributo dos bolsistas produtividade são calculadas. Um exemplo desta contextualização é apresentado na figura 4.

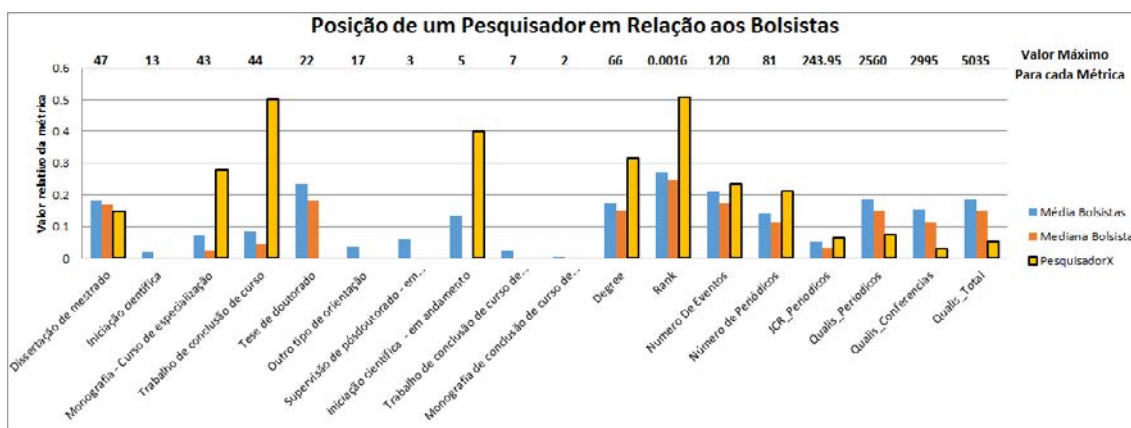


Figura 4. Contextualização das métricas de um pesquisador em relação às métricas dos bolsistas

5. Conclusões e Trabalhos Futuros

O processo de seleção de bolsistas de produtividade em pesquisa do CNPq é um processo complexo que envolve características objetivas e subjetivas que, atualmente, só é possível com o auxílio de milhares de pareceristas *ad-hoc*.

Como pôde ser observado ao se utilizar apenas algumas características objetivas (calculadas a partir de medidas bibliométricas e da análise de redes sociais), a maioria dos classificadores não obteve um desempenho melhor (em questão de acurácia) do que classificar todos os pesquisadores como não bolsistas, ou seja, nenhum deles conseguiu ultrapassar de forma consistente o valor de 95,06% para a acurácia. Porém, é importante destacar que há outra medida importante para a classificação de bolsistas que é a revocação (*recall*).

Observa-se que na classificação entre bolsistas e não bolsistas as medidas de revocação da classe positiva (bolsistas) e precisão para esta mesma classe possuem resultados complementares. Enquanto o melhor resultado de precisão para esta classe foi de 65,5% (Rotation Forest) com revocação de pouco menos de 48%, tem-se no outro extremo os resultados da Rede Bayesiana com revocação da classe positiva superior a 87%, mas precisão pouco acima de 31%. Isto é, o algoritmo foi capaz de identificar mais de 87% dos bolsistas, porém do total de indivíduos que o algoritmo indicou como bolsistas, menos de 32% deles realmente eram. Por isso, entre os algoritmos testados, o classificador que utiliza a Rede Bayesiana foi aquele que apresentou a pior acurácia geral para o caso de identificação entre bolsista e não bolsista sem *resampling* (90,04%).

Uma estratégia comumente utilizada para tentar aumentar a revocação das classes minoritárias em conjunto de dados não balanceados é o balanceamento do conjunto de treinamento. Para isto, foram realizados os testes com *resampling*. Nestes, a revocação das classes minoritárias para todos os algoritmos subiu consideravelmente, enquanto o valor de acurácia diminuiu (fato que já era esperado). Todos os resultados obtiveram uma revocação de pelo menos 69% para a classe minoritária, mostrando que fazer o *resampling* funcionou de maneira satisfatória em relação a esta medida. A maior revocação foi obtida pelo algoritmo que utiliza Adaboost, chegando a um valor de 91,87%, com precisão para a classe positiva de 31% e acurácia geral de 89,52%

O intuito desse trabalho foi criar uma abordagem que conseguisse classificar pesquisadores da área de Ciências da Computação em bolsistas e não bolsistas além de facilitar a contextualização de um pesquisador em relação aos bolsistas e, conforme apresentado, considera-se que os objetivos foram atingidos de maneira satisfatória. Utilizando-se do Adaboost com *resampling*, por exemplo, foi obtida uma acurácia de 89,52% e uma revocação para ambas as classes de aproximadamente 90%. Este tipo de classificação, mais do que verificar se efetivamente um doutor é ou não bolsista de produtividade, pode ser utilizada para indicar a um doutor que, de acordo com as métricas estudadas, ele possui melhores ou piores chances de obter uma bolsa de produtividade em pesquisa.

Como trabalhos futuros pretende-se explorar outras medidas bibliométricas e estruturais de redes sociais visando a obter resultados ainda melhores.

Agradecimentos

O trabalho apresentado neste artigo foi parcialmente financiado pela CAPES e pelo CNPq (processos 306046/2013-0 e 477246/2013-3).

Referências

Andrade, R. L.; Rego, L. C. (2015). A influência da rede de coautoria no nível das bolsas de produtividade da Área de Engenharia de Produção. In *IV Brazilian Workshop on*

Social Network Analysis and Mining (BraSNAM 2015).

- Digiampietri, L., Linden, R., and Barbosa, L. (2016). Caracterizando departamentos e programas de computação utilizando análise de redes sociais e bibliometria. In *V Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2016)*.
- Digiampietri, L., Mena-Chalco, J., Pérez-Alcázar, J. J., Tuesta, E. F., Delgado, K., and Mugnaini, R. (2012a). Minerando e caracterizando dados de currículos Lattes. In *III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2014), Anais do XXXIV Congresso da Sociedade Brasileira de Computação (CSBC2014)*.
- Digiampietri, L. A., Alves, C. M., Trucolo, C. C., and Oliveira, R. A. C. (2014a). Análise da rede dos doutores que atuam em computação no Brasil. In *III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2014), Anais do XXXIV Congresso da Sociedade Brasileira de Computação (CSBC2014)*.
- Digiampietri, L. A., Mena-Chalco, J., Silva, G. S., Oliveira, L., Malheiro, A., and Meira, D. (2012b). Dinâmica das relações de coautoria nos programas de pós-graduação em computação no Brasil. In *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2012) - Anais do XXXII Congresso da Sociedade Brasileira de Computação (CSBC 2012)*, page 12, Curitiba, PR, Brasil.
- Digiampietri, L. A., Mena-Chalco, J. P., Vaz de Melo, P. O. S., Malheiro, A. P. R., Meira, D. N. O., Franco, L. F., and Oliveira, L. B. (2014b). BraX-Ray: An X-Ray of the Brazilian Computer Science Graduate Programs. *PLoS ONE*, 9(4):e94541.
- Digiampietri, L. A., Mugnaini, R., and Alves, C. M. (2013). Análise da participação dos orientandos na produção dos orientadores: um estudo de caso em ciência da computação. In *II Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2013) - Anais do XXXIII Congresso da Sociedade Brasileira de Computação (CSBC 2013)*.
- Lemieux, V. and Ouimet, M. (2008). *Análise Estrutural das Redes Sociais*. Instituto Piaget.
- Lima, H., Silva, T. H., Moro, M. M., Santos, R. L., Meira, Jr, W., and Laender, A. H. (2015). Assessing the profile of top Brazilian computer science researchers. *Scientometrics*, 103(3):879–896.
- Poblacion, D., Mugnaini, R., and Ramos, L. (2009). *Redes sociais e colaborativas em informação científica*. Angellara Editoras, Sao Paulo, 1st edition.
- Prell, C. (2012). *Social network analysis history, theory & methodology*. Los Angeles London SAGE.
- Tess, B. H., Furuie, S. S., Castro, R. C. F., Barreto, M. d. C. C., and Nobre, M. R. C. (2009). Assessing the scientific research productivity of a Brazilian healthcare institution: a case study at the heart institute of São Paulo, Brazil. *Clinics*, 64:571 – 576.
- Wasserman, S. and Galaskiewicz, J. (1994). *Advances in social network analysis research in the social and behavioral sciences*. SAGE.

Análise e Ranqueamento da Rede de Advogados induzida por Processos Judiciais Trabalhistas

Leonardo Filipe Rodrigues Ribeiro¹, Daniel R. Figueiredo¹,
Paulo Roberto Nascimento¹

¹Programa de Engenharia de Sistemas e Computação (PESC/COPPE)
Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro – RJ

leoribeiro@cos.ufrj.br, daniel@land.ufrj.br, prmn@cos.ufrj.br

Abstract. *Who are the most important lawyers in the labor court of Rio de Janeiro state? Using data from tens of thousands of cases, we created a network of lawyers, directed and weighted, in order to answer that and some other questions. An empirical evaluation indicates that the structure of the network has very commonly features founded in other real social networks. Metrics as closeness, betweenness, vertex strength and pagerank were used to perform ranking of its vertices and identifying the main lawyers of the network. Lastly, we found information about how lawyers are related to each other in the different trials in which they worked. The main objective of this work is to understand how labor lawyers are connected and what their importance is in the community in which they work.*

Resumo. *Quais são os advogados mais influentes da justiça do trabalho do estado do Rio de Janeiro? Utilizando dados de dezenas de milhares de processos, construímos uma rede de advogados, direcionada e com pesos, com o objetivo de responder a essa e algumas outras perguntas. Uma avaliação empírica indica que a estrutura da mesma apresenta características muito comumente encontradas em outras redes sociais reais. As métricas closeness, betweenness, força do vértice e pagerank foram utilizadas para realizar o ranqueamento dos seus vértices e identificar os principais advogados da rede. Por fim, identificamos como os advogados estão relacionados nos diferentes processos onde atuaram. O principal objetivo desse trabalho é entender como os advogados trabalhistas estão conectados e qual sua importância na comunidade em que trabalham.*

1. Introdução

A solução judicial de conflitos decorrentes das relações de trabalho, segundo a Constituição, é de competência da Justiça do Trabalho, a qual está estruturada em um Tribunal Superior do Trabalho (TST), com competência em todo o território nacional, e 24 Tribunais Regionais do Trabalho (TRT) distribuídos pelo país. O Tribunal Regional do Trabalho - 1ª Região (TRT1) tem competência para julgar os processos trabalhistas no âmbito do Estado do Rio de Janeiro.

Os processos de natureza trabalhista são norteados pelos princípios e normas do Direito Processual do Trabalho. Um processo possui um autor, também chamado de reclamante na reclamação trabalhista, e um réu, também chamado de reclamado. Na

justiça do trabalho reclamante é aquele que aciona o judiciário, geralmente o empregado, ao passo que reclamado é aquele que é chamado a se defender perante o juízo, em geral é a empresa ou empregador.

Atualmente, todos os novos processos iniciados no TRT1 são eletrônicos. Para isso, utiliza-se o Processo Judicial Eletrônico da Justiça do Trabalho (PJe-JT), que é um sistema desenvolvido pelo Conselho Superior da Justiça do Trabalho (CSJT) a partir de um *branch* do Processo Judicial Eletrônico (PJe), o qual é desenvolvido pelo Conselho Nacional de Justiça (CNJ) em parceria com tribunais e tendo a participação da Ordem dos Advogados do Brasil. O PJe-JT foi implantado no TRT1 em junho de 2012 e hoje já conta com mais de 713 mil processos na 1ª instância e mais de 57 mil processos na 2ª instância. Todos os atos e documentos de processos PJe-JT são eletronicamente cadastrados. Os dados são públicos, salvo aqueles em segredo de justiça, e podem ser acessados através do portal do tribunal.

Este trabalho utiliza as informações dos processos judiciais trabalhistas oriundas da base de dados do sistema PJe-JT, em particular, informações das partes reclamante e reclamada. Essas partes possuem um ou mais advogados. Apesar do portal público, estas informações foram obtidas diretamente do banco de dados do PJe-JT e os dados foram anonimizados, de forma que os nomes dos advogados foram transformados em números aleatórios a fim de preservar a privacidade dos advogados.

Iremos construir uma rede de advogados onde os vértices representam os advogados e uma aresta direcionada de um advogado A para um advogado B indica que A “perdeu” um ou mais processos para B . As arestas possuem pesos, então se o advogado A perdeu k processos para o advogado B , existirá uma aresta direcionada de A para B com peso k . A ideia é que um advogado, ao perder um processo para outro, está conferindo importância ao segundo. Logo, um advogado que possui muitas arestas de entrada e poucas de saída terá um maior prestígio na rede de advogados, por ter vencido muitos processos e perdido poucos.

A Figura 1 mostra um pequeno pedaço da rede de advogados, mais especificamente advogados que atuaram na 2ª Vara de Macaé. Os advogados **38846** e **681953** ganharam dois e cinco processos, respectivamente, do advogado **492270**. Os advogados **111275** e **44401** perderam e ganharam entre si um processo, e o mesmo ocorreu entre os advogados **44401** e **492270**. Repare as intensidades (pesos) das arestas, e a possibilidade de reciprocidade entre um par de advogados (A perde para B , e vice-versa).

As principais contribuições deste trabalho são a construção, caracterização e ranqueamento de uma rede empírica de advogados a partir de processos judiciais. A caracterização desta rede indica que a mesma possui propriedades tipicamente encontradas em redes sociais, tais como baixa densidade, distâncias curtas e distribuição de grau com cauda pesada. A partir da rede, realizamos um ranqueamento utilizando métricas distintas (*pagerank* e força do vértice, entre outras) com o objetivo de identificar os advogados mais influentes. Ao final, comparamos o ranqueamento com o valor monetário dos processos envolvendo os respectivos advogados, que indica uma correlação com o ranqueamento.

O restante desse artigo está organizado da seguinte forma: na seção 2 apresentamos os trabalhos relacionados; uma descrição dos detalhes do *dataset* e de como foi

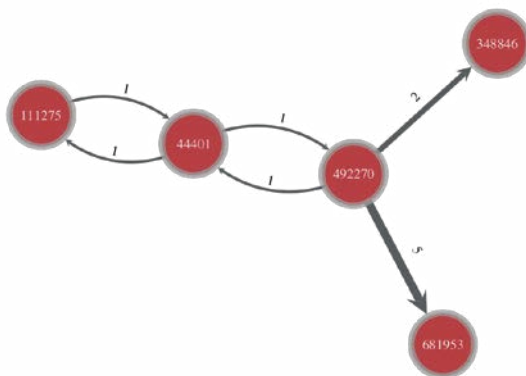


Figura 1. Rede de advogados com pesos nas arestas que indicam vitórias em processos julgados.

criada a rede são exibidos na seção 3; na seção 4 é apresentada a análise da rede; o ranqueamento dos advogados é relatado na seção 5 e as conclusões são traçadas na seção 6.

2. Trabalhos Relacionados

Em [16] é feita uma análise estrutural de uma rede de tenistas criada com base nos dados da Associação de Tenistas Profissionais (ATP). É proposto um ranking de tenistas utilizando um algoritmo análogo ao *pagerank*. Em [8] são produzidos ranqueamentos em redes sociais esportivas, levando em conta o aspecto temporal inerentemente presente nos esportes para realizar o cálculo dos pesos das arestas da rede, capturando a noção de que resultados mais antigos não são tão importantes quanto resultados mais recentes, para a criação de um ranqueamento atual.

Em [6] foi criada a “Rede da Lei”, uma rede com várias citações feitas em casos, estatutos ou por autoridades na área da jurídica nos EUA. O objetivo é caracterizar e entender como o sistema legal americano tem se comportado. Essa rede possui várias semelhanças estruturais com a Web.

Newman [12][13] caracteriza redes de colaborações científicas onde os vértices são cientistas e haverá uma aresta entre dois cientistas se eles foram co-autores em um ou mais artigos científicos. O estudo mostra diversas propriedades da rede criada como curtas distâncias entre cientistas, a presença de agrupamentos, além de diferentes padrões encontrados dependendo do campo de colaboração estudado.

Em [7] é explorado como padrões presentes na rede de interações entre pessoas de uma organização afetam a percepção que os empregados têm uns dos outros. Isso acontece através da influência social baseada na proximidade das pessoas e do poder exercido baseado na centralidade da rede. Os resultados sugerem que algumas características da rede explicam melhor as percepções que as pessoas têm umas das outras, do que atributos individuais ou posições formais na organização.

3. Dataset e Rede

Os dados dos processos utilizados na criação da rede de advogados trabalhistas do estado do Rio de Janeiro foram capturados do sistema PJe-JT. Selecionamos processos

eletrônicos distribuídos de junho de 2012 à outubro de 2015 e dentre estes selecionamos somente processos pertencentes às seguintes classes processuais: Ação Trabalhista - Rito Sumário (Alçada), Ação Trabalhista - Rito Sumaríssimo e Ação Trabalhista - Rito Ordinário.

O Rito Sumário, também chamado de alçada, é regulado pela lei 5584/70, e abrange as causas de pequeno valor que não excedam a dois salários mínimos. Ele foi criado no intuito de acelerar a resolução dos processos individuais do trabalho, não permitindo recursos e limitado a três testemunhas por parte. Devido ao valor muito baixo dos processos que se sujeitam ao Rito Sumário, o mesmo se revelou pouco eficiente nas regiões sul e sudeste do Brasil, regiões em que as demandas processuais trabalhistas superam em sua esmagadora maioria o valor de dois salários mínimos [1].

O Rito Sumaríssimo é previsto no artigo 852-A e seguintes da Consolidação das Leis Trabalhistas (CLT), e versa normalmente sobre matérias não complexas. Para um processo se enquadrar neste rito é necessário que o valor da causa não exceda a 40 salários mínimos. Algumas características básicas desse tipo de processo é que os pedidos devem ser apresentados de forma líquida, os incidentes processuais devem ser necessariamente resolvidos em audiência e o número de testemunhas por parte deve ser duas.

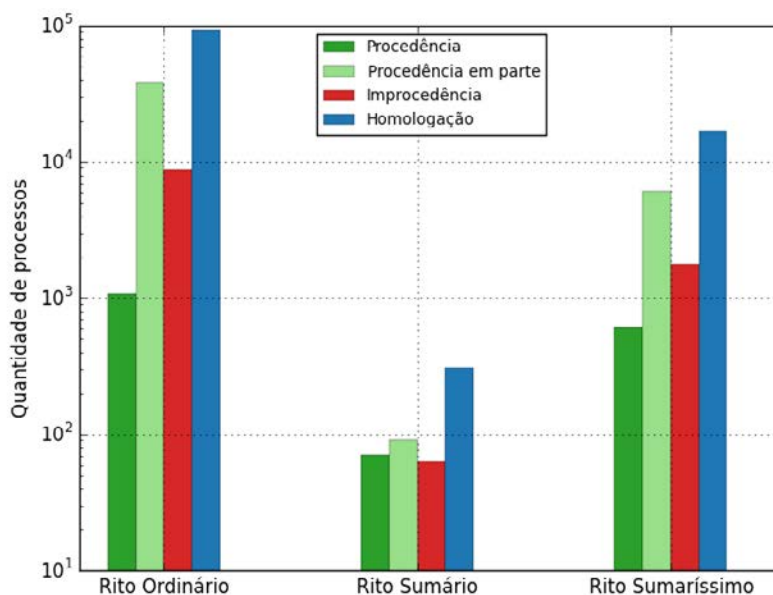


Figura 2. Quantidade de processos divididos por classes processuais. Para cada classe é mostrada a quantidade de processos procedentes (verde), procedentes em parte (verde claro), improcedentes (vermelho) ou processos com homologação de transação (azul).

No Rito Ordinário são enquadrados os processos que não se enquadram no Rito Sumário ou Sumaríssimo. Ele tem como características básicas: três testemunhas por parte (salvo o inquérito para apuração de falta grave que exige seis testemunhas), os pedidos podem ser ilíquidos e na sentença há exigência de relatório. É utilizado normalmente para causas mais complexas ou contra os entes públicos.

Além das três classes processuais, iremos focar em quatro resultados que podem ocorrer nesses tipos de processo: procedência, procedência em parte, improcedente e homologação de transação. A procedência ocorre quando o juiz julga procedente o pedido, aceita o pedido do autor da ação e diz que ele tem razão, ou seja, o autor ganha a ação. A procedência em parte ocorre quando o juiz profere uma decisão dando ganho de causa parcial ao autor, ou seja, ele ganhou a ação, mas não o valor total solicitado. A improcedência ocorre quando o julgamento da causa é desfavorável ao autor, significa que o juiz deu ganho de causa para o réu e não para o autor, ou seja, o pedido do autor foi improcedente. Quando as duas partes entram em um acordo e o juiz homologa, estamos diante da homologação de transação. [11]

Por fim, iremos considerar apenas os processos que foram transitados em julgado, isso significa que chegaram ao fim. A decisão do juiz ou desembargador é definitiva e não pode mais ser modificada, portanto não é possível apresentar mais nenhum recurso. A figura 2 mostra todos os 166183 processos presentes na base de dados avaliada, o número de processos das diferentes classes e os diferentes resultados por classe. Repare que a classe Rito Sumário possui uma minoria dos processos (menos de 300) em qualquer resultado, enquanto o Rito Ordinário possui aproximadamente 10 mil processos (ou mais) em três resultados.

Os vértices da rede a ser construída a partir dos dados dos processos são advogados e uma aresta direcionada de um advogado A para um B indica que o advogado A perdeu um processo para B . Portanto precisamos definir o significado de perder ou ganhar e para isso utilizaremos os resultados dos processos.

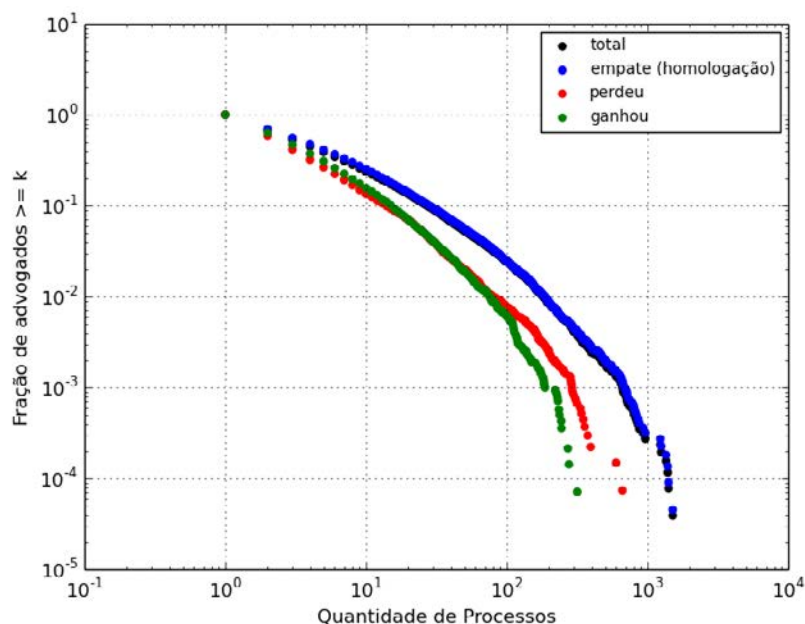


Figura 3. Distribuição Complementar Cumulativa Empírica (CCDF) de advogados que atuaram (preto), empataram (azul), ganharam (verde) ou perderam (vermelho) em k ou mais processos.

Nos processos que foram procedentes ou procedentes em parte consideramos que

os advogados do autor (ou o advogado do autor, se existir apenas um) ganharam dos advogados do réu (ou do advogado do réu, se existir apenas um), dando origem a uma aresta direcionada de cada advogado do réu para cada advogado do autor. No caso de um processo improcedente consideramos o que os advogados do autor (ou o advogado do autor, se existir apenas um) perderam para os advogados do réu (ou do advogado do réu, se existir apenas um), dando origem a uma aresta direcionada do(s) advogado(s) do autor para o(s) advogado(s) do réu.

Nos casos de processos em que ocorreu a homologação de transação, as partes do processo chegaram a um acordo. Esses processos são a maioria, conforme ilustrado na figura 2, e esse comportamento é esperado pois a CLT torna obrigatória a proposta de conciliação, assim o acordo, na Justiça do Trabalho, tem prioridade absoluta [9]. Como houve um acordo no processo, não existiu um ganhador, por isso não iremos considerar processos homologados na rede pois eles não contribuem para a importância relativa dos advogados.

A figura 3 apresenta a fração de advogados da rede que atuaram em k ou mais processos. É possível perceber que a maioria dos advogados da rede participaram de apenas um processo na justiça trabalhista, levando em conta a amostra de dados coletados para nossa rede. Pouco menos de 10% dos advogados ganharam ou perderam mais de 10 processos e este número cai para menos de 0.5% se tratando de 100 processos. Por outro lado, temos advogados que atuaram em mais de mil processos, indicando que a distribuição possui cauda pesada.

Outro ponto mostrado na figura é a grande quantidade de homologações entre advogados na rede, ou seja, ocorreu um acordo entre as partes e o juiz realizou uma homologação de transação. A cauda da distribuição de advogados que perderam processos é mais pesada do que a dos que ganharam, ou seja, é mais “fácil” (comum) perder do que ganhar. Além disso, a cauda da distribuição de advogados que homologaram processos é ainda mais pesada, ou seja, é mais “fácil” homologar um processo do que perdê-lo.

Tabela 1. Informações da rede

Vértices	17575	Arestas	113990
Diâmetro	9	Densidade	3.69×10^{-4}
Grau médio (entrada + saída)	12.97	Distância média	2.24
Grau mínimo de entrada	0	Grau mínimo de saída	0
Grau máximo de entrada	326	Grau máximo de saída	408
Componentes conexas	177	Tamanho da maior componente	17173(97.71%)
Componentes fortemente conexas	9464	Tamanho da maior componente fortemente conexa	8080(45.97%)
Coefficiente de clusterização local	1.31×10^{-2}	Coefficiente de clusterização global	7.71×10^{-3}
Reciprocidade	0.15		

4. Análise da Rede

A tabela 1 apresenta diversas informações sobre a estrutura da rede de advogados, que possui 17575 vértices (advogados) e 113990 arestas direcionadas. Nota-se ainda que a rede possui diversas características encontradas em outras redes reais [5]:

- **Distância e diâmetro baixos:** A rede possui uma distância média muito pequena em relação a quantidade de vértices: a distância média entre 17575 vértices é de apenas 2.24. O diâmetro da rede também é bem pequeno em relação aos vértices: 9, ou seja, a maior distância entre dois vértices são 9 passos.
- **Esparsa e conectada:** Muitas redes reais são extremamente esparsas, exibindo uma densidade muito baixa (ordens de grandeza menor que 1), porém são extremamente “conectadas”, com quase todos os vértices pertencendo à maior componente conexa. A rede de advogados possui uma densidade bem baixa (3.69×10^{-4}) e sua maior componente conexa possui 97.71% dos vértices da rede.
- **Alto coeficiente de clusterização:** Redes sociais estão cheias de triângulos, ou seja, se um advogado A está relacionado aos advogados B e C , é bem provável que os advogados B e C também estejam relacionados. O coeficiente de clusterização de muitas redes reais, que mede a tendência da rede a exibir triângulos, é ordens de magnitude maior que a densidade da rede. A rede de advogados possui o coeficiente de clusterização (7.71×10^{-3}) uma ordem de grandeza maior que a densidade da rede (3.69×10^{-4}). É importante notar que um único processo não gera nenhum triângulo na rede, pois um processo gera um grafo bipartido completo entre advogados das duas partes do processo.
- **Distribuição de grau com cauda pesada:** Em muitas redes reais, os graus dos vértices são bem desiguais, vértices com graus ordens de grandeza maior que a média são encontrados com probabilidade não desprezível, exibindo uma distribuição de grau com cauda pesada. A rede de advogados possui cauda pesada pois, como é possível ver na figura 4a, podemos encontrar vértices com graus bem maiores que o grau médio da rede (12.97), com probabilidade não desprezível, tanto para o grau de entrada quanto para o grau de saída.
- **Distribuição da força dos vértices:** A força de um vértice é dada pela soma dos pesos de suas arestas de entrada ou saída, respectivamente. Na figura 4b, que mostra a distribuição das forças dos vértices, é possível ver que menos de 1% dos advogados trabalharam em centenas de processos. É interessante notar que a distribuição da força de entrada dos vértices possui a cauda menor que a distribuição da força de saída. Essa diferença é percebida para advogados que possuem mais de 100 processos, indicando ser mais fácil perder do que ganhar um processo.

A rede possui uma reciprocidade relativamente alta (0.15), significando que muitas vezes (15%) os advogados, que se enfrentam, perdem um do outro mutuamente, em diferentes processos. Ou seja, não há uma preponderância significativa em sempre vencer (ou sempre perder) quando consideramos um par de advogados que se enfrentam.

5. Ranqueamento de advogados

A partir da estrutura da rede, queremos identificar os advogados mais influentes da justiça trabalhista do estado do Rio de Janeiro. Para isso utilizamos algumas métricas de centralidade para definir a importâncias relativa desses advogados.

A forma mais simples de medir a “importância” de um vértice em uma rede é considerar seu grau, ou seja, a quantidade de arestas de entrada no vértice, essa métrica é chamada de centralidade de grau. Porém, também é importante levar em conta o peso das arestas que apontam para o vértice, pois o peso de uma aresta direcionada, do vértice

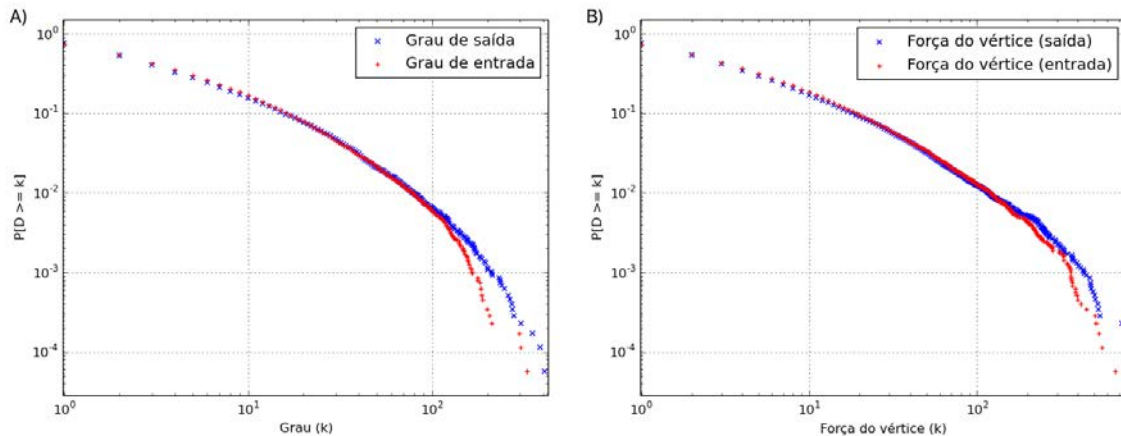


Figura 4. a) Distribuição Complementar Cumulativa Empírica (CCDF) dos graus de entrada e saída da rede. b) Distribuição Complementar Cumulativa Empírica (CCDF) das forças dos vértices (entrada e saída).

A para o vértice B , é a quantidade de vezes que B ganhou de A e para isso utilizamos a *força do vértice* [2], que é dada por:

$$S_v = \sum_{u \in \Gamma^-(v)} w_{u \rightarrow v}$$

Onde $\Gamma^-(v)$ são todos os vértices vizinhos de v que têm arestas apontando para v e $w_{u \rightarrow v}$ é o peso da aresta direcionada de u para v .

Na tabela 2 é possível ver os 10 vértices com maior força e seus respectivos graus, de entrada e saída. Repare que nem sempre o vértice com maior grau possui maior força, devido aos vértices que ganharam ou perderam várias vezes dos mesmos vértices, fazendo com que a aresta entre eles tenha um peso maior. É possível notar ainda os vértices com maior força de saída, ou seja, aqueles que mais perderam processos.

O *pagerank* [14], algoritmo desenvolvido pelos fundadores do Google, vem sendo utilizado para calcular a importância em redes direcionadas, levando em conta a qualidade e a quantidade de arestas que apontam para os vértices. O valor de *pagerank* de um vértice é definido recursivamente através da seguinte equação:

$$PR(v) = \frac{(1-d)}{n} + d \sum_{u \in \Gamma^-(v)} \frac{PR(u)w_{u \rightarrow v}}{d^+(u)}$$

Onde d é o fator de amortecimento, que foi definido como 0.85, valor tradicional utilizado em diversos artigos; $\frac{(1-d)}{n}$ é a importância mínima de um vértice na rede com n vértices, que pode ser interpretado como sua aptidão externa; $\Gamma^-(v)$ são todos os vértices vizinhos de v que têm arestas apontando para v ; $w_{u \rightarrow v}$ é o peso da aresta direcionada do vizinho u de v para v e $d^+(u)$ é igual a soma dos pesos de todas as arestas direcionadas que saem de u . O valor de *pagerank* pode ser calculado através de um algoritmo iterativo até a convergência dos valores.

5.1. Ranqueamento dos vértices

A tabela 3 apresenta os 10 primeiros advogados quando utilizamos diferentes métricas para realizar o ranqueamento. Como é possível notar, muitos advogados aparecem nas primeiras colocações no ranqueamento de mais de uma métrica. O advogado **16604** figura em primeiro na métrica de *pagerank*, isso mostra que ele tem grande importância no âmbito de processos trabalhistas. Ele também está em primeiro na métrica de *betweenness*, isso quer dizer que ele está no meio da rede e possui influência política. De fato, se trata de um advogado com mais de 20 anos de experiência que fundou seu escritório na década de 90. Processos vinculados a ele estão em vários tribunais do país, mostrando assim sua influência em nível nacional. Além do mais, recebeu em 2015 o título de cidadão honorário da cidade de São Paulo.

O advogado **13444** está entre os melhores nas métricas *pagerank* e *betweenness*. Ele também é um advogado importante na justiça trabalhista. **13444** é um advogado de renome com 36 anos de prática jurídica trabalhista na cidade do Rio de Janeiro.

Tabela 2. 10 maiores forças de advogados, tanto de entrada quanto de saída, e seus respectivos graus.

Entrada			Saída		
Advogado	Grau do vértice	Força do vértice	Advogado	Grau do vértice	Força do vértice
914	105	666	1772	408	1290
4674	149	555	16604	351	928
890	61	532	15928	84	908
17522	117	512	9409	239	730
14876	143	506	13155	247	541
800	298	447	2471	192	529
16496	149	417	9430	386	525
17508	188	399	6942	235	508
59	326	392	6538	234	502
244	184	386	4048	116	492

O advogado **9430** figura como o primeiro no ranqueamento de *closeness* e em sexto no ranqueamento de *betweenness*. Isso significa que esse advogado tende a conhecer muitos outros advogados, pois está próximo a outros advogados da rede, exercendo grande influência em seu círculo profissional. A revista Análise Advocacia 500 [4] traz os escritórios e advogados mais admirados no Brasil em cada ano. São vários critérios para definição dos advogados mais admirados, sendo o mais comum uma pesquisa de satisfação feita em grandes empresas ao redor do país. Na edição de 2014, encontramos o advogado **9430** entre os melhores advogados do estado do Rio de Janeiro. No entanto, **9430** figura entre os 10 nos ranqueamentos de centralidade de grau e força de saída do vértice, sendo um dos advogados da rede com uma grande quantidade de processos perdidos.

O advogado **59** começou suas atividades em 1989 e está entre os 10 nos ranqueamentos de *pagerank*, *betweenness* e força do vértice. Também está em primeiro

no ranking de centralidade de grau de entrada, ou seja, é o advogado que mais ganhou processos de advogados diferentes. É um advogado que atua em prol dos direitos dos trabalhadores nas cidades do Rio de Janeiro, Niterói e Nova Iguaçu, tendo atuado em um grande número de processos.

O advogado **15928**, segundo no ranqueamento de *pagerank*, trabalha num dos maiores escritórios de advocacia do país. Esta empresa é reconhecida por analistas do mercado jurídico e por publicações realizadas na área, tendo ganhado inúmeros prêmios nos últimos anos, sendo um dos melhores escritórios de advocacia do Brasil em 12 áreas de atuação pela revista Análise Advocacia 500 2015 [4], e um dos melhores escritórios de advocacia do Brasil em 22 áreas de atuação pela Chambers Latin America 2016 [3].

O advogado **9873** está presente nos ranqueamentos de *pagerank*, *betweenness* e *closeness*. Ele é co-autor de um livro na área de direito empresarial, trabalha em um grande escritório de advocacia do país e foi indicado um dos melhores advogados trabalhistas da América Latina em 2015 e 2016 pela LACCA no guia Latin Lawyer 250 [10].

Tabela 3. Top 10 advogados em cada métrica de centralidade.

<i>Pagerank</i>	Força do vértice	Centralidade de grau	Closeness	Betweenness
16604	914	59	9430	16604
15928	4674	943	1042	59
59	890	800	7498	943
6942	17522	3117	37	6942
13155	14876	17134	4034	9873
943	800	328	4662	9430
13444	16496	1790	3570	13444
17134	17508	17508	4853	13155
9873	59	11278	4178	8284
9409	244	244	7284	7927

Quase a totalidade de advogados listados entre os 10 melhores em cada métrica possui anos de experiência na área de direito trabalhista. Esta observação indica que a rede de advogados construída pode de fato ser utilizada para identificar advogados influentes a partir de métricas de ranqueamento de vértices bem objetivas. Outro ponto interessante é que a boa parte dos advogados bem ranqueados atuam para funcionários e sindicatos. Muitos advogados acreditam que a legislação trabalhista no Brasil, muitas vezes considerada excessiva, favorece o trabalhador. Desta forma, ao dar uma proteção excessiva ao trabalhador, a Justiça acaba desprestigiando o trabalho e o emprego. Eles acreditam que esse cuidado concedido ao empregado é resultado de sua fraqueza em relação ao empregador. Outro ponto é que os altos índices de condenação das empresas são resultado da dificuldade que os empregadores têm em cumprir o excesso de normas que compõem a legislação trabalhista. [15]

A tabela 4 mostra a soma de todos os valores pedidos nos processos, divididos em homologados, procedentes e procedentes em parte e improcedentes, para cada advogado na métrica de força de entrada do vértice. É possível perceber que os maiores valores se encontram nos processos homologados. Os valores monetários possuem uma correlação com o ranqueamento, pois todos os 10 advogados melhores ranqueados possuem valores

muito mais altos em seus processos procedentes do que em seus processos improcedentes. Os valores monetários dos processos não foram utilizados em nosso trabalho porque não representam quanto efetivamente foi ganho ou perdido em um processo, pois são valores pedidos no início do processo e não o valor real definido em sentença. A base de dados de onde foram coletados os dados não ofereceu este valor de forma fácil.

Tabela 4. Top 10 advogados na métrica de força de entrada do vértice e a soma dos valores monetários pedidos nas causas que foram homologadas, procedentes e improcedentes.

	Força de entrada do vértice	Homologação por transação	Procedentes e Proc. em parte	Improcedentes
1	914	1.929.663, 78	5.283.287, 24	632.000, 00
2	4674	2.354.408, 22	4.835.868, 50	682.748, 05
3	890	1.999.900, 00	8.165.000, 00	1.227.500, 00
4	17522	8.027.831, 50	8.381.018, 26	2.757.052, 12
5	14876	2.284.865, 14	4.426.888, 49	622.748, 05
6	800	11.719.880, 60	10.175.541, 85	1.277.629, 69
7	16496	5.700.311, 72	2.101.000, 00	724.534, 80
8	17508	21.865.543, 78	4.415.000, 00	942.736, 91
9	59	14.965.705, 43	7.555.000, 00	2.622.000, 00
10	244	21.123.000, 00	4.175.000, 00	832.736, 91

6. Conclusão

Uma vez que muitas redes reais possam ser representadas a partir de dados coletados, utilizar métricas e técnicas que levem em conta a estrutura da rede para entender seu comportamento tem se tornado bem útil. Neste trabalho criamos uma rede de advogados trabalhistas do estado do Rio de Janeiro a partir de dados públicos sobre processos judiciais. Esta rede tem como objetivo capturar a importância relativa dos advogados, pois uma aresta direcionada de um advogado para outro é criada se o primeiro perdeu em um processo para o segundo. Uma avaliação empírica indicou que esta rede possui muitas características presentes em outras redes reais estudadas. A semelhança estrutural com essas redes permitiu utilizar as métricas apresentadas, como a força do vértice e o *pagerank*, para caracterizar o relacionamento entre os advogados.

Nossa abordagem para identificar os advogados mais influentes da rede identificou advogados com tradição na justiça do trabalho e com participação em muitos processos, inclusive com um deles sendo citado em uma revista especializada, como estando entre os melhores advogados do estado do Rio de Janeiro. Desta forma, ilustramos como a rede de advogados induzidos por processos judiciais trabalhistas pode ser utilizada para revelar aspectos importantes, como a importância relativa.

Como trabalhos futuros, propõe-se utilizar na representação da rede o valor ganho ou perdido em um processo e com isso utilizar outras métricas de centralidade para ranquear os advogados.

Referências

- [1] Patrícia Ribeiro de Azevedo e Kalina Valéria Bastos Pedroza. *Procedimento Sumaríssimo: aspectos polêmicos*. Acessado em 11/03/2016. URL: <http://www.webartigos.com/artigos/procedimento-sumarissimo-trabalhista-aspectos-polemicos/90877/>.
- [2] A. Barrat et al. “The architecture of complex weighted networks”. Em: *Proc. Natl. Acad. Sci.* 101 (2004), pp. 3747–3752.
- [3] Chambers e Partners. *Chambers Latin America 2016*. Acessado em 18/03/2016. URL: <http://www.chambersandpartners.com/guide/latin-america/9>.
- [4] Análise Editora. *Advocacia 500 - 2014*. Acessado em 16/12/2015. URL: <http://www.analise.com/site/publicacoes/exibe/2/analise-advocacia-500>.
- [5] Daniel Ratton Figueiredo. “Introdução a Redes Complexas”. Em: *Atualizações em Informática 2011*. Ed. por Alberto Ferreira de Souza e Wagner Meira Jr. PUC-Rio, 2011. Cap. 7, pp. 303–358.
- [6] James H. Fowler et al. *Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court*. 2007.
- [7] Steven B. Andrews Herminia Ibarra. “Power, Social Influence, and Sense Making: Effects of Network Centrality and Proximity on Employee Perceptions”. Em: *Administrative Science Quarterly* 38.2 (1993), pp. 277–303. ISSN: 00018392.
- [8] Péterson Sampaio Procópio Júnior et al. “Time-aware Ranking in Sport Social Networks”. Em: *Journal of Information and Data Management* (2012).
- [9] Conselho Nacional de Justiça. *Resolução Nº 125 de 29/11/2010*. Acessado em 11/03/2016. URL: <http://www.cnj.jus.br/busca-atos-adm?documento=2579>.
- [10] LACCA. *Latin Lawyer 250*. Acessado em 18/03/2016. URL: <http://latinlawyer.com/11250/>.
- [11] Jocélia Lopes. *Termos Judiciais*. Acessado em 16/12/2015. URL: <http://www.jocelialopesadvocacia.com.br/termos.html>.
- [12] M. E. J. Newman. “The structure of scientific collaboration networks”. Em: *Proceedings of the National Academy of Sciences* 98.2 (2001), pp. 404–409.
- [13] Mark E.J. Newman. “Complex Networks”. Em: ed. por Eli Ben-Naim, Hans Frauenfelder e Zoltan Toroczkai. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. Cap. Who Is the Best Connected Scientist? A Study of Scientific Coauthorship Networks, pp. 337–370.
- [14] Lawrence Page et al. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab, nov. de 1999.
- [15] *Quem é protecionista é a legislação, não a Justiça do Trabalho, dizem advogados*. <http://www.conjur.com.br/2015-set-28/protetora-trabalhador-legislacao-nao-justica>. Acessado em 16/12/2015.
- [16] Filippo Radicchi. “Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis”. Em: *PLoS ONE* 6.2 (fev. de 2011), e17249.

Avaliação automática da utilidade de *reviews* usando Redes Neurais Artificiais no *corpus* do *Steam*

Jardeson L. N. Barbosa¹ Raimundo S. Moura¹

¹Departamento de Computação – Universidade Federal do Piauí (UFPI)
64.049-550 – Teresina – PI – Brasil

jardesonbarbosa@gmail.com, rsm@ufpi.edu.br

Abstract. *Through online reviews, consumers can communicate with product suppliers and influence the buying decision of other consumers over the Internet. However, thanks to the high number of published reviews daily, it is difficult to identify which texts to read. As a solution for that problem, some websites use a review evaluation system based on the users votes, which, while useful, is not always ideal. This paper proposes an automatic method to analyze the helpfulness of online user reviews from Steam using Multilayer Perceptron Artificial Neural Network. We found out that certain features of reviews affect the perception of helpfulness and we discuss applications and future researches.*

Resumo. *Através de reviews, consumidores podem se comunicar com os fornecedores de produtos e serviços e influenciar a decisão de compra de outros consumidores na Internet. Porém, com o alto número de reviews publicados diariamente, é difícil identificar quais textos devem ser lidos. Como uma solução para esse problema, alguns sites utilizam um sistema de avaliação de reviews baseado no voto dos usuários que embora útil, nem sempre é ideal. Este trabalho propõe um modelo de análise automática da utilidade de reviews online de usuários do Steam, usando Rede Neural Artificial Perceptron Multicamadas. Descobriu-se que certas características de reviews afetam a percepção de utilidade e discutimos aplicações e pesquisas futuras.*

1. Introdução

Com números que ultrapassam 12 milhões de lojas mundialmente, o *e-commerce* é, atualmente, um dos principais serviços da Internet¹. Porém, graças à enorme quantidade de páginas e produtos disponíveis, essa pode ser uma atividade desgastante, porque, durante o processo de busca de informação, consumidores são geralmente submetidos a múltiplas escolhas [Vohs et al. 2014].

Para auxiliar o processo de decisão dos consumidores, algumas empresas de *e-commerce* introduziram alguns sistemas de recomendação automáticos que oferecem recomendações personalizadas aos consumidores. Entretanto, apesar do bom desempenho, esses sistemas ainda são menos eficientes do que recomendações de pessoas conhecidas, como familiares e amigos [Sinha and Swearingen 2001].

[Goldenberg et al. 2001] demonstraram que o processo de tomada de decisão dos consumidores é fortemente influenciado pela comunicação “boca-a-boca” (WOM, do

¹<http://blog.lemonstand.com/just-how-big-is-the-ecommerce-market-youll-never-guess>

inglês “word of mouth”). Na Internet, redes sociais *online* concentram a comunicação WOM, uma vez que permitem que consumidores compartilhem suas experiências e recomendações através de opiniões, ou *reviews* – seja escrevendo, votando, comentando ou compartilhando [Kim and Srivastava 2007].

Consumidores que leem *reviews* online se atentam não somente às notas dos *reviews*, mas também às demais informações textuais e não-textuais [Hu et al. 2008]. Entretanto, dada a enorme quantidade de *spam*, textos de baixa qualidade e informações falsas na Internet, consumidores podem necessitar de uma grande quantidade de tempo para que encontrem *reviews* que afetem a decisão de compra.

Uma solução encontrada pela maioria dos *sites* de comércio eletrônico foi o estabelecimento de um padrão chamado de “utilidade do *review*”. Em *sites* como *Amazon*², usuários podem votar em *reviews* que eles consideram úteis ou inúteis no processo de decisão de compra. Esses votos são utilizados para ordenar os *reviews* em função dos mais úteis. Dessa forma, os próprios usuários filtram os textos que devem ser lidos. Um dos problemas desse método é que *reviews* precisarão de muitos votos positivos para serem lidos e, conseqüentemente, postagens mais recentes deixarão de ser lidas por não possuírem votos suficientes, mesmo que sejam mais relevantes que as postagens mais votadas [Li et al. 2013].

Neste trabalho, estuda-se a percepção de utilidade de *reviews online* através da análise das descrições textuais e informações quantitativas presentes no texto. Este trabalho propõe um modelo capaz de prever o grau de percepção de utilidade de *reviews online* de forma automática utilizando uma Rede Neural Artificial *Multilayer Perceptron* (RNA MLP). Esta investigação é baseada em textos coletados da loja *Steam*³, que distribui jogos para computadores pessoais. O site *Steam* foi escolhido por oferecer uma loja virtual e rede social com mais de 125 milhões de usuários ativos⁴ e por permitir que seus usuários publiquem, leiam, compartilhem e votem em *reviews online*. Além disso, ainda são poucos os trabalhos que exploram a mineração de opiniões do site *Steam*.

O restante deste artigo é organizado da seguinte forma: a Seção 2 trata da utilidade de *reviews online* e apresenta uma revisão literária realizada na área. A Seção 3 apresenta o nosso modelo proposto. A Seção 4 apresenta os resultados dos experimentos. A Seção 5 apresenta as considerações finais e trabalhos futuros.

2. Revisão literária

2.1. Percepção de utilidade de *reviews online*

Em um sistema de votos como o proposto pelo site *Amazon*, a utilidade de um *review* pode ser definida como

$$H = \frac{n_p}{n_p + n_n} \quad (1)$$

onde n_p representa o número de votos positivos e n_n representa o número de votos negativos [Kim et al. 2006]. Entretanto, essa abordagem apresenta algumas tendências desvantajosas como, por exemplo, um cenário em que *reviews* novos ou com poucos votos

²<https://www.amazon.com.br/>

³<http://store.steampowered.com>

⁴<http://store.steampowered.com/news/16000/?l=portuguese>

nunca sejam lidos e, conseqüentemente, votados [Li et al. 2013]. Além disso, nem todo mundo que lê um *review* necessariamente vota ao final [Kim et al. 2006].

Outro problema dessa abordagem é que ela não leva em consideração a quantidade de votos recebidos por cada *review*. É impossível, por exemplo, diferenciar dois *reviews* com 80% de aprovação, mesmo que um deles possua 10000 votos e outro apenas 10. Por se tratar de uma distribuição Binomial, uma solução eficiente para contornar esse último problema é o uso do limite inferior do intervalo de confiança de Wilson para uma Binomial [Agresti and Coull 1998] apresentada em (2)

$$H = \left(\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{[\hat{p}(1 - \hat{p}) + z_{\alpha/2}^2/4n]/n} \right) / (1 + z_{\alpha/2}^2/n) \quad (2)$$

onde \hat{p} representa a porcentagem de votos positivos observados, n é o total de votos recebidos e z é o quantil da distribuição normal padronizada para o intervalo de confiança escolhido.

Essa equação considera a quantidade de votos de um *review* como uma amostra de uma população hipotética em que todos os *reviews* são votados em igual quantidade. Assumindo um determinado nível de confiança (95% neste trabalho), é determinada uma nova probabilidade de sucesso para a distribuição analisada. Esse mesmo cálculo é utilizado em alguns *sites* no processo de ordenação de *reviews*, tópicos e comentários, como *Reddit*⁵ e *Yelp*⁶.

Neste trabalho, a medida de utilidade de *reviews online* usada será a porcentagem de votos positivos do *review* com base na equação (2). Formalmente, entende-se a medida de utilidade de *reviews* de produtos *online* como a medida na qual consumidores percebem um *review* de produto como sendo capaz de facilitar o julgamento ou decisão de compra [Li et al. 2013].

2.2. Características formativas da percepção de utilidade de *reviews online*

Evidências suportam a conceitualização teórica de utilidade de *reviews* como uma construção formativa [Li et al. 2013]. Assim, entende-se que existe uma série de características que influenciam a percepção de utilidade de *reviews*.

Alguns trabalhos identificaram a influência direta do conteúdo do *review* na percepção de sua utilidade. Com a ajuda de um modelo de aprendizado de máquina supervisionado, [Kim et al. 2006] descobriram que o tamanho do texto, suas unigramas e a classificação do produto são características importantes na definição da percepção de utilidade de *reviews*. [Danescu-Niculescu-Mizil et al. 2009] realizaram um estudo sobre o *corpus* do *site Amazon* e identificaram que um *review* é avaliado como mais útil quando a sua nota (ou avaliação) é próxima ao consenso das demais notas para o mesmo produto. [Schindler and Bickart 2012] examinaram características do conteúdo do *review* e descobriram que o tamanho do texto e a quantidade de informações expressas influenciam positivamente a percepção de utilidade de *reviews*, mas somente até certo ponto. [Lee and Choeh 2014] descobriram que o tamanho do texto e o número de palavras com apenas uma letra são bons indicativos da percepção de utilidade do *review*.

⁵<http://www.redditblog.com/2009/10/reddits-new-comment-sorting-system.html>

⁶<http://officialblog.yelp.com/2011/02/the-most-romantic-city-on-yelp-is.html>

Alguns pesquisadores também exploraram a influência da declaração de autoria na percepção de utilidade de *reviews*. [Connors et al. 2011] examinaram alguns fatores básicos associados com a utilidade de *reviews*. Eles perceberam que a *expertise* do autor, ou seja, o conhecimento do autor a respeito do domínio do produto, possui influência positiva na percepção de utilidade de *reviews*. [Forman et al. 2008] descobriram que a presença de informações descritivas sobre a identidade do autor de *reviews online* possui impacto positivo na percepção de utilidade de *reviews*.

Este trabalho apresenta como diferencial o uso de RNA, a análise da utilidade de *reviews* escritos em Português Brasileiro e a exploração de características de *reviews* pouco estudadas em trabalhos anteriores, como o histórico de publicações dos autores de *reviews*, descrições sobre o uso do produto avaliado e a quantidade de relacionamentos dos autores dentro de uma rede social *online*. Essas características puderam ser facilmente extraídas graças ao modelo de comunidade virtual proposto pelo *Steam*.

3. Modelo proposto

Para um dado *review*, o objetivo deste trabalho é encontrar o grau de percepção de utilidade H , definido pela equação (2), onde H é um número real na faixa de 0 a 1. Para desenvolvermos um modelo que prevê essa taxa, estudou-se as características formativas da percepção de utilidade de *reviews*. Com base no nosso estudo, hipotetiza-se que existam três importantes grupos de características de *reviews* que modelam a sua percepção de utilidade: características a respeito do autor do *review*, características textuais e características relativas aos metadados do *review*. O modelo proposto é dado por uma função de regressão que tem como entrada um vetor X contendo essas características e como saída um valor escalar H , que é a medida de percepção de utilidade. Escolheu-se uma RNA MLP para aproximar essa função. Uma RNA MLP para aproximação de funções pode ser descrita da seguinte forma:

$$y(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n_1} \lambda_i g_j(u_i) \quad (3)$$

$$u_i = \sum_{j=1}^n W_{ij} x_j - \theta_i \quad (4)$$

onde a função y a ser mapeada pelo MLP será constituída pela superposição de funções de ativação do tipo logística, representadas pelos termos $g_j(u_i)$, as quais são ponderadas pelos fatores λ_i [da Silva et al. 2010]. A função logística é definida como

$$g(u) = \frac{1}{1 + e^{-\beta * u}} \quad (5)$$

onde β é uma constante real associada ao nível de inclinação da função logística frente ao seu ponto de inflexão.

As RNA são adequadas para esse problema porque são aproximadores de funções universais. O teorema da aproximação universal aplicado ao MLP fornece as bases necessárias para definir as configurações estruturais dessas redes com a finalidade de mapear funções algébricas [da Silva et al. 2010]. Além disso, o modelo de RNA permite

que qualquer forma funcional que relacione as variáveis independentes à variável dependente não precise ser imposto pelo modelo, ou seja, é possível capturar relacionamentos não-lineares nos dados sem a necessidade de especificação prévia [Lee and Choeh 2014].

Para definição das características no modelo proposto, analisou-se dados de *reviews* sobre jogos digitais do *site Steam* descritos nas subseções a seguir.

3.1. Coleta dos dados

Diferente de um grande número de *sites* de *e-commerce*, o *Steam* se comporta como uma comunidade online. Usuários do *Steam* possuem perfis com foto; podem adicionar amigos; trocar mensagens privadas e públicas entre si; publicar imagens, textos e vídeos, e podem participar de grupos.

O *site Steam* possui um detalhado sistema de *reviews* que inclui, além do texto do usuário, diversas outras informações. A Figura 1 mostra um típico *review* do *Steam*. A avaliação final do usuário é apresentada em destaque acima do texto (Recomendado ou Não Recomendado). Em um *review* também se encontram informações sobre o autor (nome, quantidade de produtos na conta e quantidade de análises), a quantidade de horas gastas com o produto, a quantidade de pessoas que consideraram a análise útil, além do sistema de votos que inclui a opção “engraçado”, que não interfere no voto de utilidade.



Figura 1. Exemplo de *review* do *Steam* com 245 votos positivos.

A coleta dos dados para o treinamento e teste do modelo foi realizada de forma automática através de um *scraper*⁷ escrito em *Python* utilizando a biblioteca *Scrapy*⁸. Definiu-se como a população de interesse, todos os *reviews* de jogos digitais do *site Steam* escritos em Português Brasileiro, com pelo menos 1 voto de utilidade e postados por usuários que possuem perfis públicos. Como os *reviews* do *Steam* são ordenados de forma que *reviews* com poucos votos ou com nota baixa sejam dificilmente acessíveis, a coleta concentrou-se apenas nos 100 jogos mais jogados no *Steam* no dia 26 de fevereiro. Essa lista é atualizada diariamente no próprio *site*⁹.

Dos 100 jogos mais jogados no dia 26 de fevereiro, 5 não possuíam *reviews* em português brasileiro com pelo menos 1 voto. Inicialmente coletou-se uma amostra de 10.168 *reviews* retroativamente publicados entre 25 de novembro 2013 e 26 de fevereiro de 2016. Optou-se por eliminar *reviews* com menos de três votos para evitar assimetria e tendência nos dados, já que a maioria dos *reviews* possui poucos votos.

⁷Técnica computacional para a extração de informações de *websites*

⁸<http://scrapy.org/>

⁹<http://store.steampowered.com/stats/>

Como, neste trabalho, objetiva-se uma análise automática, optou-se por eliminar todos os textos com grau de corretude abaixo de um *threshold* definido arbitrariamente a fim de remover ruídos dos dados (como textos em outros idiomas). Modelou-se a corretude do texto como o número de palavras corretas dividido pelo número total de palavras em um texto. Para determinar se a palavra pertence ou não ao Português, usou-se o dicionário *br.ispell* da USP¹⁰. Eliminando os *reviews* com menos de 3 votos e os *reviews* com corretude abaixo do *threshold*, definido como a média da amostra inicial (0.76), obteve-se uma amostra final contendo 5823 *reviews*.

3.2. Características de *reviews* analisadas

3.2.1. Características relativas à autoria

Em relação ao autor, destacam-se duas características: a reputação e a *expertise*. Estudos sugerem que *reviews* submetidos por autores com histórico positivo são vistos como mais úteis [Wathen and Burkell 2002]. A reputação também incorpora elementos referentes à credibilidade do autor. A *expertise* indica o nível de conhecimento de um autor a respeito do tópico abordado. Modelou-se a reputação de um autor a partir de três variáveis: a média de votos de *reviews* pela quantidade de *reviews* publicados pelo usuário, a razão dos votos positivos pela quantidade total de votos do usuário; e o número de amigos que o usuário possui na comunidade *Steam*. Por *expertise*, pode-se explorar a quantidade de horas que o autor do *review* dedicou ao jogo analisado. Um autor que jogou apenas 2 horas de um determinado jogo pode ter menos proficiência no tópico que foi analisado do que aquele que jogou 20 horas.

3.2.2. Características relativas ao conteúdo textual do *review*

Fundamentalmente, pesquisas anteriores se dedicaram a dois tipos de análise textual em *reviews online*: análise das características semânticas e análise das características estilísticas do texto [Kim et al. 2006, Cao et al. 2011]. Ao explorar o conteúdo textual de um *review*, pode-se atentar às opiniões e sentimentos expressos no texto. De acordo com o dicionário Oxford (tradução livre), opinião é uma visão ou julgamento formada a respeito de algo, não necessariamente baseada em fatos ou conhecimento.

Hipotetiza-se que *reviews online* com uma grande quantidade de opiniões são mais informativos. Nesta abordagem, opiniões devem ser modeladas de forma numérica. Logo, é interessante saber apenas a quantidade de opiniões expressas em cada *review*. Para tal, realiza-se um processo de extração de opiniões segundo uma adaptação do modelo proposto por [Sousa et al. 2015].

Na extração de opiniões, os textos são analisados de acordo com a estrutura frasal das sentenças. Adotou-se o padrão "sujeito + verbo + predicativo do sujeito", onde o núcleo do sujeito representa a característica qualificada e o núcleo do predicado representa a palavra qualificadora. Por exemplo, "esse jogo é divertido" é extraído como (*jogo, divertido*). Esses padrões linguísticos foram definidos através de uma análise manual realizada sobre uma outra amostra de 385 *reviews* coletadas no mesmo *site* (nível de confiança 95% e erro 5%). O resultado é apresentado na Tabela 1.

¹⁰<http://www.ime.usp.br/ueda/br.ispell/>

Tabela 1. Padrões linguísticos
 SUBST ADV? V? (ADV? ADJ — ADV V)*
 ADJ SUBST
 ADV ADJ

Foram considerados como elementos estilísticos a legibilidade do texto e o tamanho do texto. Legibilidade é a facilidade em que um texto pode ser compreendido. A legibilidade de um texto em inglês pode ser analisada utilizando o teste de legibilidade de Flesch-Kincaid, um método matemático que avalia o quão legível é um texto por meio da sua média de palavras por sentenças e média de sílabas por total de palavras [Kincaid et al. 1975]. Uma adaptação para o Português foi feita pelo jornalista Alberto Dines [Squarisi and Salvador 2005]

$$\left(\frac{n_w}{n_s} + n_p \right) * 0.4 \quad (6)$$

onde n_w é o número total de palavras do texto, n_s é o número total de sentenças e n_p é o número total de polissílabos. A constante 0,4 representa a média de letras da palavra na frase do Português. Quanto maior o *score*, menos legível é o texto. Um texto com *score* 1 pode ser facilmente lido por qualquer pessoa.

Baseado em descobertas de [Kim et al. 2006] e [Lee and Choeh 2014], modelou-se o tamanho do texto a partir do número de palavras, número de sentenças e a quantidade de monossílabos do texto, já que o Português não possui grande quantidade de palavras semanticamente relevantes com apenas uma letra.

3.2.3. Metadados do *review*

Consideram-se como metadados a avaliação final do *review* e a data de postagem do *review* em dias. Avaliação final refere-se à avaliação de produto binária presente em cada *review*, que pode ser “recomendado” ou “não recomendado”. É importante notar que leitores podem ser influenciados pela média das avaliações dos produto. Essa ideia foi explorada por [Danescu-Niculescu-Mizil et al. 2009].

Considerando uma avaliação de *review* como sendo 1, para “recomendado” (ou positivo), e 0, para “não recomendado” (ou negativa), modelou-se a avaliação como uma expressão, onde p é porcentagem de *reviews* com avaliações positivas (em representação decimal) para um produto e x é a avaliação individual do *review* para o mesmo produto:

$$xp + (1 - x) * (1 - p) \quad (7)$$

Partindo das descobertas de [Otterbacher 2009], modelou-se o tempo de postagem de um *review* como a diferença entre a data de lançamento do produto e a data de postagem do *review* em dias.

3.3. Validação

A RNA MLP para aproximação de funções é composta de três camadas, sendo uma camada de entrada, apenas uma camada escondida e uma camada de saída. A camada de

entrada mapeia cada uma das variáveis de entrada estudadas. A camada de saída possui apenas um neurônio que mapeia a variável de saída H , ou utilidade do *review*. A validação cruzada permite definir a quantidade mais adequada de neurônios na camada escondida. Utilizou-se a função logística como ativação para os neurônios da camada escondida (5). Já para a camada de saída, utilizou-se uma função de ativação linear (3), já que o neurônio de saída realiza apenas uma combinação linear das funções de ativação logística implementadas nos neurônios da camada escondida [da Silva et al. 2010]. Portanto, após o processo de treinamento da rede MLP, a matriz de pesos referentes ao neurônio de saída corresponderá aos próprios parâmetros λ_i da equação (3), isto é $\lambda_i = W_{1,i}$. A ilustração dessa rede é apresentada na Figura 2.

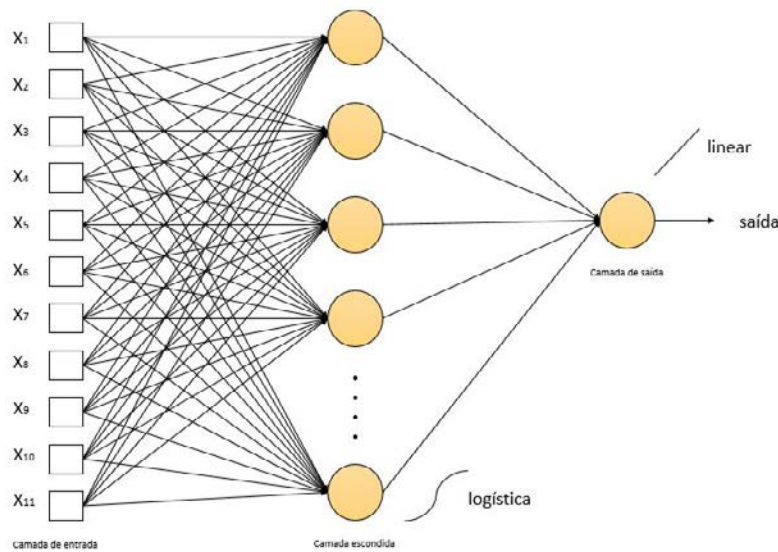


Figura 2. Ilustração da rede MLP aplicada em aproximação funcional

As variáveis de entrada, definidas na especificação de hipóteses são: i) a quantidade de votos de *reviews* pela quantidade total de *reviews* de um usuário; ii) a razão dos votos positivos pela quantidade total de votos de um usuário; iii) o número de amigos do usuário na comunidade *Steam*; iv) a quantidade de horas (em horas) que o autor dedicou ao jogo analisado; v) a quantidade de opiniões expressas no texto; vi) a legibilidade do texto; vii) o tamanho do texto em palavras; viii) o número de sentenças no texto; ix) o número de palavras monossílabas no texto; x) a diferença entre a avaliação média do produto e a avaliação do usuário; xi) a diferença entre a data do lançamento do produto e a data da publicação do *review* em dias. A variável de saída esperada é a utilidade do *review*.

Pode-se calcular o relacionamento entre cada variável de entrada e a variável de saída utilizando o cálculo de *Relative Strength* (RS) [Lee and Choeh 2014] [Yoon et al. 1994]:

$$RS_{ji} = \frac{\sum_{k=0}^n (W_{ki} W_{jk})}{\sum_{i=0}^m \sum_{k=0}^n |W_{ki} W_{jk}|} \quad (8)$$

onde W_{ki} indica o peso entre a k -ésima unidade escondida e a i -ésima unidade de entrada. W_{jk} denota o peso entre a j -ésima unidade de saída e a k -ésima unidade escondida. RS_{ji} calcula o peso relativo entre a i -ésima variável de entrada e a j -ésima variável de saída. O numerador do cálculo de RS_{ji} calcula a relação entre a i -ésima variável de entrada e a j -ésima variável de saída e pode ser tanto positivo quanto negativo, dependendo dos pesos. Já o denominador calcula o relacionamento total entre todas as variáveis de entrada e saída.

4. Resultados

A especificação da topologia de rede MLP mais apropriada para mapear um problema específico é usualmente efetuada de forma empírica [da Silva et al. 2010]. O processo de seleção de topologias candidatas geralmente se dá por tentativa e erro [Lee and Choeh 2014]. Para a seleção da melhor topologia dentre as candidatas, optou-se pela validação cruzada k -fold, ou validação cruzada k -partições. Nesse processo, a amostra total é dividida em k partições, sendo $(k - 1)$ delas usadas para compor o subconjunto de treinamento enquanto a partição restante constitui o subconjunto de teste [da Silva et al. 2010]. O processo de aprendizado se repete k vezes até que todas as partições tenham sido utilizadas como subconjunto de teste. Neste trabalho, definiu-se k como 10. O desempenho global de cada topologia candidata é obtido em função da média entre os desempenhos individuais observados quando da aplicação das k partições.

A RNA implementada neste estudo tem valores randômicos como pesos iniciais, taxa de aprendizado 0.1 e *momentum* 0.3. Optou-se por valores pequenos para garantir que a convergência ocorra, mesmo que o processo necessite de mais épocas. Para evitar *overfitting*, cada topologia foi treinada até que fosse encontrado a menor raiz quadrada do erro quadrático médio (RMSE) local ou atingisse 10^4 épocas, o que ocorrer primeiro.

A Figura 3 apresenta um gráfico com os RMSE para cada uma das topologias testadas. Iniciamos com uma topologia de 1 neurônio na camada escondida e aumentamos essa camada em 1 neurônio a cada nova topologia. Observamos que o erro aumenta a cada novo neurônio incluído na camada escondida, mas passa a diminuir assim que atinge 5 neurônios. Escolheu-se como melhor topologia, aquela que apresenta o menor número de neurônios na camada escondida e o menor RMSE. O modelo final de RNA MLP escolhido é uma rede de três camadas que consiste em 11 nós de entrada, 5 nós na camada escondida e 1 nó de saída. Dada a complexidade de realizar previsões baseadas no comportamento humano, o RMSE de 0.1929 foi considerado aceitável.

A Tabela 2 apresenta o peso relativo dos determinantes da utilidade dos *reviews* utilizando a equação (8). Valores acima de 0.02 são bons indicativos de impacto na percepção de utilidade do *review* [Lee and Choeh 2014]. Os melhores índices foram da média de votos por *review* do usuário, nota média do usuário, quantidade de horas jogadas e tamanho do texto em palavras.

Os resultados desta análise corroboram pesquisas anteriormente realizadas por [Danescu-Niculescu-Mizil et al. 2009], [Schindler and Bickart 2012] e [Kim et al. 2006]. Notou-se que a avaliação do usuário e o tamanho do texto são características que impactam positivamente a percepção de utilidade de *reviews*. Já a data de postagem dos *reviews* aparenta não ter forte impacto na percepção de utilidade em comunidades *online* como o Steam, o que contraria descobertas realizadas por [Otterbacher 2009]. Uma das possíveis

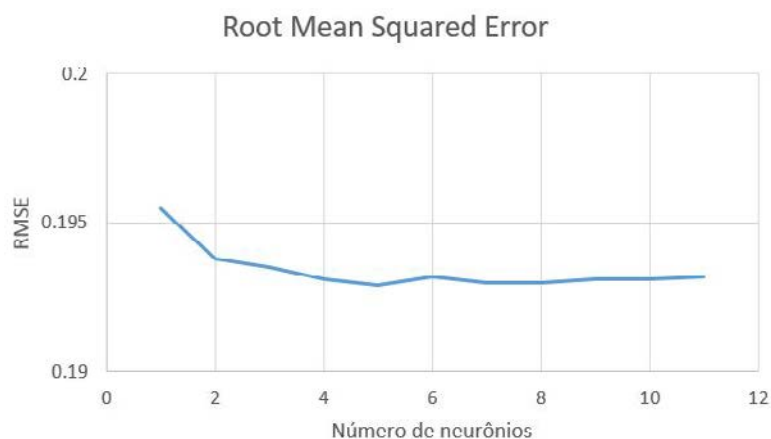


Figura 3. Variação do RMSE a medida que o número de neurônios na camada escondida aumenta.

Tabela 2. Peso relativo de cada variável.

Variáveis	Força
média de votos por review do usuário	0.14272
nota média do usuário	0.06112
quantidade de horas jogadas	0.02507
quantidade de amigos	0.01743
padrões linguísticos	-0.00442
legibilidade	0.00789
quantidade de palavras	0.02794
quantidade de sentenças	0.01224
diferença entre avaliação do usuário e média do produto	0.01302
quantidade de palavras monossílabas	-0.00962
diferença de dias entre a lançamento do produto e postagem	0.00122

justificativas poderia ser o fato de o *Steam* realizar promoções regulares que movimentam as páginas de produtos velhos.

Entre as novas descobertas, tem-se que a reputação do autor é fator importante na percepção de utilidade de *reviews*. Descobriu-se que autores com histórico positivo de votos recebidos têm *reviews* melhor avaliados. A *expertise* do autor também aparenta ser uma característica importante. No caso do *Steam*, a quantidade de horas jogadas por um autor no momento da publicação do *review* tem impacto positivo sobre a percepção de utilidade. A quantidade de amigos do usuário na comunidade também possui algum impacto positivo sobre a percepção de utilidade de *reviews*.

5. Conclusões e trabalhos futuros

Este trabalho teve como objetivo estudar formas de quantificar e prever a percepção de utilidade de *reviews* de usuários *online*. Os resultados deste estudo foram positivos e dão indícios sobre quais características formam a percepção de utilidade de *reviews*. Para o setor de *e-commerce*, estes resultados dão pistas sobre quais estruturas de *website* podem ser utilizadas para que os melhores *reviews* sejam lidos pelos consumidores e, consequentemente, maximizar as vendas.

Para trabalhos futuros, considera-se substituir o uso de padrões linguísticos

por técnicas de PLN mais bem elaboradas como LSA, uma técnica de extração e representação de uso contextual de palavras utilizando computações estatísticas para um grande *corpus* de texto. O princípio da LSA é que o agregado de todos os contextos de palavras em que uma dada palavra aparece ou não aparece provém um conjunto de restrições mútuas que determina, em grande parte, a similaridade do significado de palavras e conjuntos de palavras [Landauer et al. 1998]. Espera-se que com técnicas mais robustas, seja possível identificar situações que não puderam ser analisadas em nosso modelo, como a ocorrência de sujeito oculto e o humor em textos.

Por fim, pretende-se explorar o impacto do humor na percepção de utilidade de *reviews*. O próprio *Steam* permite que usuários votem em *reviews* engraçados, opção pouco utilizada atualmente, mas que pode servir como base para estudos futuros.

Referências

- [Agresti and Coull 1998] Agresti, A. and Coull, B. A. (1998). Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, 52(2):119–126.
- [Cao et al. 2011] Cao, Q., Duan, W., and Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, 50(2):511 – 521.
- [Connors et al. 2011] Connors, L., Mudambi, S. M., and Schuff, D. (2011). Is it the review or the reviewer? a multimethod approach to determine the antecedents of online review helpfulness. In *In 44th Hawaii International Conference on System Sciences*.
- [da Silva et al. 2010] da Silva, I. N., Spatti, D. H., and Flauzino, R. A. (2010). *Redes Neurais Artificiais: para engenharia e ciências aplicadas*, chapter Redes Perceptron Multicamadas. Artliber.
- [Danescu-Niculescu-Mizil et al. 2009] Danescu-Niculescu-Mizil, C., Kossinets, G., Kleinberg, J., and Lee, L. (2009). How opinions are received by online communities: A case study on amazon.com helpfulness votes. *18th International Conference on World Wide Web*,.
- [Forman et al. 2008] Forman, C., Ghose, A., and Wiesenfeld, B. (2008). *Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets*. INFORMS.
- [Goldenberg et al. 2001] Goldenberg, J., Libai, B., and Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters, Volume 12, Issue 3*, pages 211–223.
- [Hu et al. 2008] Hu, N., Liu, L., and Zhang, J. (2008). Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management Vol. 9 No. 3*, pages 201–214.
- [Kim et al. 2006] Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In Jurafsky, D. and Gaussier, r., editors, *EMNLP*, pages 423–430. ACL.

- [Kim and Srivastava 2007] Kim, Y. A. and Srivastava, J. (2007). Impact of social influence in e-commerce decision making. In *Proceedings of the Ninth International Conference on Electronic Commerce, ICEC '07*, pages 293–302, New York, NY, USA. ACM.
- [Kincaid et al. 1975] Kincaid, J. P., Fishburne, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Technical report.
- [Landauer et al. 1998] Landauer, T., Foltz, P., and Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25:259–284.
- [Lee and Choeh 2014] Lee, S. and Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Syst. Appl.*, 41(6):3041–3046.
- [Li et al. 2013] Li, M., Huang, L., Tan, C.-H., and Wei, K.-K. (2013). Helpfulness of online product reviews as seen by consumers: Source and content features. *International Journal of Electronic Commerce*, 17:101–136.
- [Otterbacher 2009] Otterbacher, J. (2009). 'helpfulness' in online communities: A measure of message quality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09*, pages 955–964. ACM.
- [Schindler and Bickart 2012] Schindler, R. and Bickart, B. (2012). Perceived helpfulness of online consumer reviews: The role of message content and style. *Journal of Consumer Behaviour*, 11:234–243.
- [Sinha and Swearingen 2001] Sinha, R. R. and Swearingen, K. (2001). Comparing recommendations made by online systems and friends. In *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*.
- [Sousa et al. 2015] Sousa, R. F. d., Rabelo, R. A. L., and Moura, R. S. (2015). A fuzzy system-based approach to estimate the importance of online customer reviews. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.
- [Squarisi and Salvador 2005] Squarisi, D. and Salvador, A. (2005). A arte de escrever bem: um guia para jornalistas e profissionais do texto. pages 50–52. Editora Contexto.
- [Vohs et al. 2014] Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., and Tice, D. M. (2014). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Motivation Science, Vol 1(S)*, pages 19–42.
- [Wathen and Burkell 2002] Wathen, C. N. and Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*, 53(2):134–144.
- [Yoon et al. 1994] Yoon, Y., Guimaraes, T., and Swales, G. (1994). Integrating artificial neural networks with rule-based expert systems. *Decis. Support Syst.*, 11(5):497–507.

Caracterização dos perfis comerciais na rede social Instagram

Gabriela Enes Campos¹, Helen Costa¹

¹Departamento de Computação e Sistemas – Universidade Federal de Ouro Preto (UFOP)
João Monlevade – MG – Brasil

`gabriela_enes@hotmail.com, helen@decsi.ufop.br`

Abstract. *This paper presents a resulting study of users data analysis in Instagram, identifying the different existing profiles. Instagram is an application for sharring photos and videos that became popular in recent years and because of this, businesses are increasingly investing in dissemination of advertisement through it. This shows that brands are interested in disseminating their products in a relaxed atmosphere, so that their customers can have a closer relationship with the company. Thus, this research aims to characterize Instagram users, identifying and differentiating ordinary users from commercial ones. We leveraged our characterization study towards a classification approach able to differentiate these users with high accuracy.*

Resumo. *Neste artigo é apresentado um estudo resultante da análise de dados dos usuários na rede social Instagram, identificando os diferentes perfis existentes. O Instagram é um aplicativo para divulgação de fotos que se tornou popular nos últimos anos e devido a isto, empresas e comerciantes estão cada vez mais investindo na divulgação através dele. Isto mostra que as marcas estão interessadas em divulgar seus produtos de forma descontraída, fazendo com que seus clientes tenham um relacionamento mais próximo com a empresa. Sendo assim, esta pesquisa visa caracterizar as contas do Instagram, identificando e diferenciando usuários comuns de usuários comerciais. Em seguida, o estudo de caracterização é utilizado em uma abordagem de classificação que foi capaz de diferenciar estes usuários com alta precisão.*

1. Introdução

As redes sociais online (Online Social Network - OSN) têm sido um fenômeno na Internet, atraindo usuários cada vez mais. Isto é devido ao grande número de pessoas que utilizam computadores, smartphones e tablets com o acesso a Internet. Atualmente na Web existem diversos tipos de OSNs que propõem diversas funcionalidades, são redes de profissionais (ex., LinkedIn), rede de amizades (ex., MySpace, Facebook) e redes para o compartilhamento de conteúdos específicos, tais como mensagens curtas (ex., Twitter), fotos (ex., Flickr, Instagram), vídeos (ex., YouTube), entre outras OSNs [Benevenuto, 2010].

O estudo das OSNs tem sido alvo de grandes pesquisadores, pois envolve fatores de interesse no mundo inteiro, como: comunicação, segurança, comércio e privacidade. Neste trabalho, será enfatizado o fator comercial, que está relacionado à divulgação de produtos e serviços, que é facilitado por OSNs, melhorando a interação com clientes e fornecedores de diversas localidades. Com isso, os usuários de redes sociais online

proporcionam uma melhor divulgação quando curtem, comentam, compartilham as informações sobre marcas, produtos e promoções, exercendo influência na sua rede social [Leskovec et al., 2007].

O Instagram é uma rede social utilizada para divulgação de fotos e vídeos curtos que se tornou muito popular nos últimos anos. Devido a isto, empresas e comerciantes estão cada vez mais investindo na divulgação através dele. Isto mostra que as marcas estão interessadas em divulgar seus produtos/serviços de forma descontraída, fazendo com que seus clientes tenham um relacionamento mais próximo da empresa. Além de empresas, usuários que são figuras públicas ou celebridades também fazem uso do Instagram para divulgação de sua própria imagem através de contas comerciais.

O Instagram ainda traz um grande desafio para estes tipos de usuários, pois não possui uma ferramenta de publicação direcionada para perfis comerciais, o que torna o processo de divulgação muito manual. Para contornar esta situação, o Instagram disponibiliza um manual para orientar empresas que querem divulgar suas marcas no aplicativo ¹. Este manual contém informações que vão desde a criação de uma conta com perfil mais comercial até o processo de divulgação dos produtos.

Apesar do Instagram possuir um manual para criação de conta diferenciada, nem todos tem o conhecimento sobre a sua disponibilidade. Com isso, nota-se que a própria empresa reconhece a necessidade de diferenciação entre perfis de usuários, que podem ter objetivos diferentes na rede. Além disso, a criação de um mecanismo automático de detecção de contas comerciais permitiria que o próprio Instagram pudesse disponibilizar aos seus usuários uma ferramenta de identificação desses tipos de usuários num mecanismo de busca dentro do próprio sistema, por exemplo. Assim, erros como o caso de um designer Andrés Iniesta (usuário comum), que teve sua conta excluída por ter o mesmo nome de um jogador de futebol famoso, não precisariam acontecer [Estado de Minas].

Neste sentido, este trabalho visa identificar quais características são eficientes para diferenciar usuários comerciais de usuários comuns no Instagram. Adicionalmente, é proposto um método de detecção de usuários comerciais utilizando a tarefa de classificação da mineração de dados.

Desta forma, o objetivo geral deste trabalho é a descoberta de padrões dos usuários do Instagram, através da criação de uma base de dados coletada utilizando a API do Instagram. Ela contém informações relativas aos dados públicos disponibilizadas nos perfis dos usuários. A partir de tais informações é possível encontrar as características que proporcionem a diferenciação dos tipos de usuários. Nesse sentido, foi proposto um método de detecção de usuários comerciais.

O restante do artigo está organizado como a seguir. A Seção 2 apresenta trabalhos relacionados com este assunto. A Seção 3 descreve como foi feita a coleta de dados e a estratégia usada para rotular usuários comerciais e comuns. A Seção 4 investiga o conjunto de atributos gerados e sua capacidade de discriminar usuários comerciais de usuários comuns. A Seção 5 descreve e avalia a estratégia usada para detectar os diferentes tipos de usuários. Finalmente, a Seção 6 oferece conclusões e direções para trabalhos futuros.

¹Instagram: <https://help.instagram.com/454502981253053/>

2. Trabalhos Relacionados

Khosla et al. [2014] analisaram os principais fatores que tornam uma imagem popular. Os dados foram coletados do Flickr, que é uma rede social que permite a seus usuários criarem álbuns para armazenarem suas fotografias. Para prever a popularidade da imagem através do conteúdo de imagem, foram utilizados duas metodologias: recursos de interpretação humana de imagem, como cor e variações de intensidade e; recursos de visão computacional de baixo nível. Para analisar cor e recurso simples de imagem, foi feita uma relação de pixels de diferentes canais de cores, mostrando que este tipo de recurso tem uma correlação muito pequena com popularidade. Nesta abordagem também foi utilizado um histograma de cores, que consistiu de um padrão de 50 cores distintas, que diagnosticou a importância das cores na previsão de popularidade. Utilizando os recursos de visão computacional, foram investigados cinco atributos de imagem:

- *Gist*: analisa a essência da imagem;
- *Texture*: analisa as texturas das imagens;
- *Color Patches*: analisa as imagens de acordo com uma paleta de cores;
- *Gradient*: analisa o gradiente das imagens, utilizando algoritmo de visão computacional HOG - Histograma Orientado a Gradiente;
- *Deep Learning*: analisou as imagens à partir de um vetor treinado, contendo um conjunto de palavras, para identificar a relação de popularidade quando esses objetos estavam presentes nas imagens.

Outro contexto analisado foi sobre a previsão da popularidade da imagem através do conteúdo social, onde foram analisados: média de visualizações das imagens publicadas, número de fotos públicas, número de grupos, número de membros dos grupos, há quanto tempo é membro, *IsPro* (conta especial). Dessa maneira, os resultados obtidos utilizando os recursos de interpretação humana, verificou-se que imagens com cores fortes (vermelho/laranja) tem mais importância levando a um maior número de visualizações. Com as análises de visão computacional, observou que objetos como pessoas contribuem positivamente para a popularidade da foto. Os cinco atributos de imagem analisadas mostraram que o conteúdo da imagem faz muita diferença na previsão da popularidade. Este trabalho se diferencia do descrito, pois é focado somente nos usuários. Porém, são analisados os atributos da imagem do perfil, utilizando a ferramenta desenvolvida por esses autores.

Silva et al. [2013] propõem um estudo do comportamento dos usuários do Instagram na identificação de pontos de interesse de uma cidade, através do acesso das informações geográficas disponibilizadas pela RSP (Rede de Sensoriamento Participativo). O Instagram possibilita que fotos publicadas na sua rede sejam integradas ao Twitter. Desta forma, os autores coletaram sua base de dados através do Twitter, sendo 2.272.556 *tweets* contendo fotos georreferenciadas, postadas por 482.629 usuários distintos. Com isso, as análises foram feitas em regiões proporcionalmente distribuídas, para que a probabilidade de uma área aleatória ser sensoreada em um horário aleatório seja bem baixo, uma vez que os usuários compartilham fotografias em diferentes escalas de tempo. Os resultados mostraram que a aplicação criada conseguiu identificar os PDIs (Pontos de Interesses) em um contexto espaço-temporal, o que é fundamental, uma vez que os PDIs são dinâmicos e mudam ao longo do tempo. Diferentemente deste trabalho, não será feita uma caracterização de usuários utilizando informações de geolocalização, mas serão utilizadas outras características relacionadas ao comportamento do usuário.

Com o objetivo de adquirir uma compreensão inicial sobre o tipo de fotos compartilhadas por indivíduos no Instagram, Hu et al. [2014] coletaram informações sobre perfil, as 20 primeiras fotos postadas, legendas e tags associadas com fotos, seguidos e seguidores de 50 usuários escolhidos aleatoriamente. As análises consistiram de técnicas de visão computacional e humanas, onde as imagens foram separadas em 8 categorias. Assim, eles identificaram que 46,6 % representam imagens das categorias *selfie* (24,2%) e amigos (22,4%). As categorias animais e moda foram as menos populares, compreendendo apenas 5%. Outras categorias como, alimentação, tecnologia e fotos com texto representaram um pouco mais de 10% do número total de imagens. Eles também identificaram os tipos de usuários do Instagram e categorizaram como: usuários que gostam de comida; usuários que postam fotos com texto; usuários que praticam atividades físicas; usuários que amam tirar *selfies* e; usuários que postam *selfies*, mas também postam fotos com amigos. Tanto esse trabalho quanto o presente estudam o comportamento dos usuários na identificação de padrões que possam ser categorizados, sendo ambos de extrema importância para o entendimento dos diversos interesses dos usuários na rede.

3. Coleta de dados

Para analisar o comportamento de usuários no Instagram, foi necessário primeiro criar uma base de dados através de informações de usuários que estivessem disponíveis publicamente e pudessem ser coletadas através da API². Essa API utiliza o protocolo OAuth 2.0³, que fornece uma forma padronizada de acessar os dados protegidos. Ele proporciona autorização específica para várias aplicações. O aplicativo desenvolvido seguiu a política e os termos de uso de dados da API do Instagram, garantindo a privacidade dos usuários e respeitando a conduta desta OSN.

Para criar a base de dados proposta, primeiramente foi criada uma lista com 50 palavras que estão relacionadas as marcas empresariais, com base em uma lista disponível na página da revista Exame⁴. Utilizou-se esta lista para dar início a coleta. Em seguida, foi desenvolvido um *crawler* que, para cada termo da lista, recuperou usuários no Instagram relacionados com o termo, utilizando o método de busca da OSN. Para cada termo pesquisado, foram escolhidos os cinco primeiros usuários listados. E para cada usuário escolhido, foram coletados também os cinco primeiros seguidores e os cinco primeiros seguidos deste usuário.

A coleta foi feita no período de 23 a 29 de março de 2015 e o *crawler* desenvolvido coletou **1389** usuários distintos. Adicionalmente, também foram coletadas informações públicas disponíveis nos perfis destes usuários. Os atributos extraídos inicialmente foram:

- Username: define o nome de login do usuário;
- Bio: Descreve um “*status*” ou uma mensagem para descrição da conta;
- Website: Permite que o usuário insira uma url de contato;
- Profile picture: Fornece o caminho para acesso a imagem do perfil do usuário;
- Full name: Nome de apresentação do usuário;
- Media: Define a quantidade de fotografias que o usuário publicou;

²Instagram Desenvolvedor: <https://instagram.com/developer/>

³OAuth 2.0: <http://tools.ietf.org/html/draft-ietf-oauth-v2-12>

⁴Lista das 50 marcas mais valiosas do mundo: <http://exame.abril.com.br/marketing/noticias/as-50-marcas-mais-valiosas-do-mundo-em-2013>

Tabela 1. Quantidades de dados coletados separados em usuários comercial e comuns

Classe	Número de usuários	Porcentagem
Comum	761	55%
Comercial	628	45%
Total	1389	100%

- Followed by: Mostra a quantidade de seguidores que o usuário tem;
- Follows: Mostra a quantidade de seguidos que o usuário tem.

Através das informações coletadas do perfil dos usuários, foi feita uma análise manual de cada usuário, classificando-o entre usuário comum ou comercial. Uma sumarização dos usuários classificados pode ser vista na Tabela 1. A partir desta classificação manual, foi possível identificar que a base de dados obtida é composta tanto por usuários comerciais como por usuários comuns, mesmo a coleta tendo partido de uma busca por usuários relacionados com conteúdo comercial.

4. Caracterização de usuários

Os usuários do Instagram tem diferentes objetivos ao utilizar o sistema, e portanto, comportam-se de maneira diferente para alcançá-los. Sendo assim, nesta seção serão analisados os vários atributos que refletem o comportamento do usuário na rede, para identificar as características que distinguem usuários comuns de usuários comerciais. Foram considerados três grupos de atributos: atributos de conteúdo, atributos de imagem e atributos de usuário. Esse grupos serão discutidos a seguir.

4.1. Atributos de conteúdo

Atributos de conteúdo são propriedades do texto contido na descrição de perfil feita por cada usuário através do atributo bio. Os seguintes atributos foram gerados para avaliação de cada usuário.

- número de palavras ou expressões que estão contidas em uma lista popular de conteúdo *spam*;
- número de caracteres maiúsculos;
- número de caracteres numéricos;
- número de URL's no texto;
- número de endereços de e-mail;
- número de telefones;
- número de informações de contato no texto (que representa a soma do número de endereços de e-mail, telefone, e palavras Facebook e Twitter);
- número de palavras;
- número de palavras maiúsculas.

Para calcular o **número de palavras ou expressões que representam *spam***, foi utilizado uma lista com termos *spam* identificada no trabalho de Costa et al. [2013]. Esta lista está disponível em uma página do GitHub⁵, que apesar de apresentar termos relacionados a *spam*, percebeu-se a semelhanças nessas palavras ao tentar atrair consumidores/clientes. Isso permitiu avaliar este atributo para a identificação desses termos na bio dos usuários.

⁵Lista de palavras ofensivas: github.com/spam-detection/badwords-pt-br

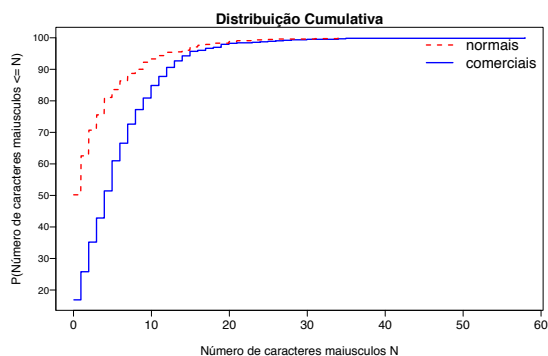


Figura 1. Número de Caracteres Maiúsculos na Bio

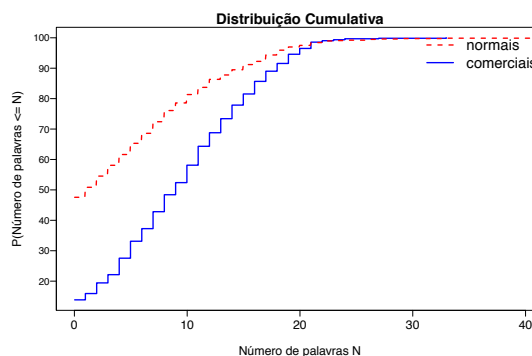


Figura 2. Número de Palavras na Bio

Para ilustrar o poder discriminativo dos atributos gerados, na Figura 1 temos a CDF (*Cumulative Distribution Function*) do atributo número de caracteres maiúsculos. Pode-se notar que os perfis comerciais possuem mais caracteres maiúsculos nas informações descritas na bio. Este fato acontece porque esta classe busca chamar mais a atenção dos outros usuários. Desta forma, percebe-se que aproximadamente 80.0% dos usuários comerciais tem até dez caracteres maiúsculos presentes nas informações da bio, enquanto que na classe de usuários comuns, cerca de 50.0% não possui nenhum carácter maiúsculo na bio.

Outra descoberta relacionada com os atributos do conteúdo é que usuários comuns utilizam menos os recursos disponíveis para a divulgação de suas informações. A Figura 2 mostra a CDF do atributo número de palavras. Observa-se que quase 50% dos usuários comuns não possuem nenhuma palavra na bio, enquanto que apenas 10.0% aproximadamente dos usuários comerciais não possuem nenhuma palavra na bio.

4.2. Atributos de usuário

O segundo grupo de atributos consiste de propriedades específicas do comportamento do usuário na rede. Sendo assim, foram considerados os seguintes atributos de usuário:

- número de fotos postadas;
- número de seguidos;
- número de seguidores;
- fração do número de seguidores por seguidos;
- se possui informação na bio, sendo 1 para sim e 0 para não;
- se possui informação no website, sendo 1 para sim e 0 para não.

O número de seguidores se mostrou muito importante, como observado na Figura 3. Os usuários pertencentes a classe comercial têm mais seguidores do que a classe de usuários comuns. Isso acontece porque é comum os perfis comerciais serem mais expostos dos que os perfis comuns. Desta forma, nota-se que 60.0% da classe comercial tem até 10^{10} seguidores, enquanto que 60.0% da classe comum possui até 10^5 seguidores. Nesse sentido, ao invés de simplesmente medir o número de seguidores e seguidos, também foi calculada a fração de seguidores por seguidos, onde usuários comerciais mostraram um valor menor da fração. Isto mostra que os perfis comerciais produzem mais interações no

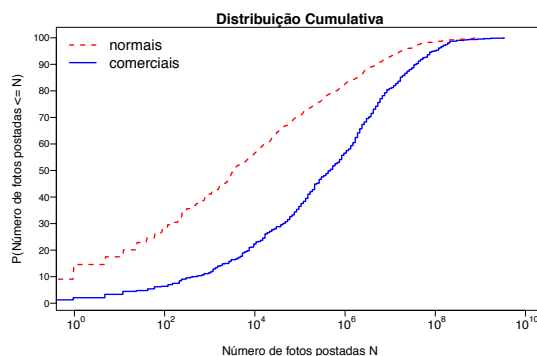
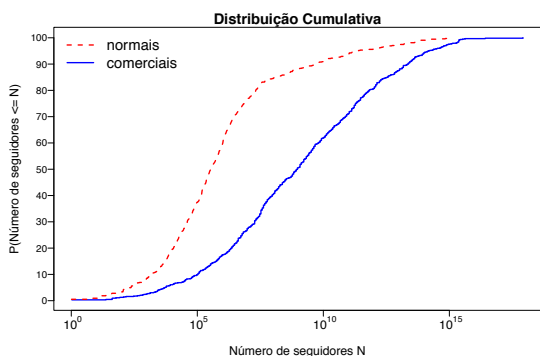


Figura 3. Seguidores

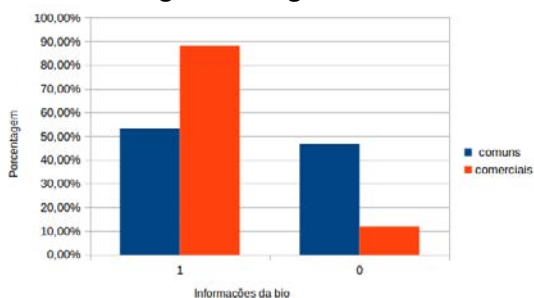


Figura 4. Fotos Publicadas

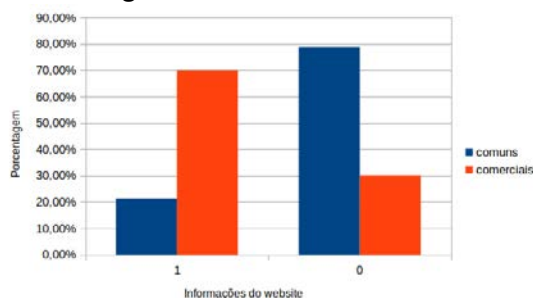


Figura 5. Informações da Bio

Figura 6. Informações da Website

Instagram, seguindo um número de usuários maior até do que seu número de seguidores. E esta estratégia faz com que eles sejam mais efetivos em conquistar os outros usuários.

A Figura 4 mostra a CDF do número de fotos postadas pelo usuário, onde é verificado que 60% dos usuários comuns postaram aproximadamente até 10^4 fotos, enquanto que apenas 22% dos usuários comerciais postaram a mesma quantidade de foto.

A Figura 5 mostra a distribuição do atributo de informações de bio. Observa-se que quase 90.0% dos usuários pertencentes a classe comercial apresentam textos na bio e apenas um pouco mais de 50.0% da classe comum expõem informações na bio.

Website é um recurso disponível ao usuário do Instagram, para que possam ser inseridas informações de sites ou outras redes sociais para um possível contato. Pode-se observar na Figura 6 que aproximadamente 80.0.% dos usuários comuns não divulgam outros sites ou redes sociais, diferentemente dos usuários comerciais, em que aproximadamente 70.0% destes inserem este tipo de informação no Instagram. Isso mostra que o os usuários comerciais utilizam o máximo possível dos recursos do Instagram para divulgar o seu produto/marca.

4.3. Atributos das imagens

O terceiro tipo de atributo calculado faz referência a popularidade da imagem disponibilizada no perfil do usuário:

- score de popularidade da imagem do perfil.

Para calcular este atributo, utilizou-se um aplicativo que foi desenvolvido por Khosla et al. [2014] com base no estudo proposto por eles. Esse aplicativo está disponível⁶ na Web e seu desenvolvimento utilizou recursos de algoritmos de visão computacional, onde foram analisados:

- Essência: analisa a ideia central da imagem;
- Textura: analisa as características visuais da imagem;
- Paleta de cores: analisa as cores das imagens, de acordo com uma paleta de 50 cores pré-definidas;
- Gradiente: analisa a intensidade dos pixels da imagem;
- Aprendizagem profunda: treinou-se um vetor de regressão linear com diferentes imagens relacionadas com objetos populares e assim, realizou as análises das imagens.

O score prevê o número de visualizações que a imagem irá receber, por exemplo, se o score é 5, espera-se que sejam obtidas $2^5=32$ visualizações por dia desta imagem. O mais importante é o score relativo. Se A tem score maior que B, provavelmente A terá maior tendência a receber mais curtidas.

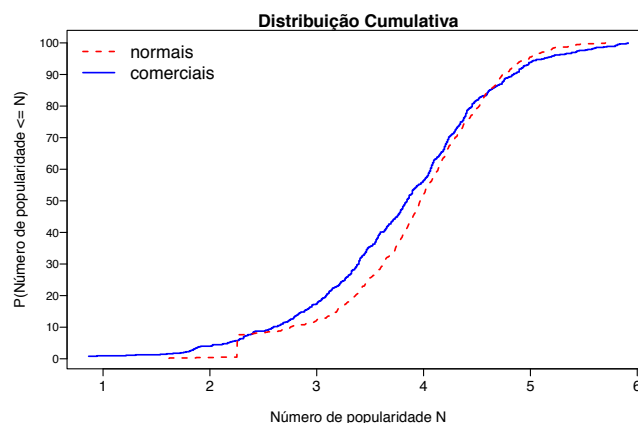


Figura 7. CDF da popularidade das imagens

Analisando a Figura 7, observa-se que os valores mais frequentes estão entre os scores 2.5 e 4, e neste mesmo intervalo é onde há uma maior diferenciação entre as distribuições, com usuários comuns tendo um score de popularidade das imagens de perfil um pouco maior que usuários comerciais.

4.4. Importância dos atributos

Para avaliar a relevância dos 16 atributos selecionados em discriminar usuários comuns de usuários comerciais, aplicou-se uma métrica de avaliação bem conhecido na literatura: Ganho de Informação [Quinlan, 1986]. Para esse experimento, utilizou-se a implementação do método fornecida pelo Weka, chamada *infoGain*. A Tabela 2 apresenta o ranqueamento dos atributos, mostrando a posição referente a pontuação dos atributos que proporcionam o maior ganho de informação de acordo com o conjunto de cada categoria (conteúdo, usuário e imagem). Dos 10 atributos mais importantes, pode-se notar que 5 pertencem a categoria de atributos de usuários.

⁶Ferramenta de popularidade da imagem: <http://popularity.csail.mit.edu/>

Tabela 2. Ranking dos atributos

Categoria	Ranking Ganho de Informação	Descrição
Conteúdo 9 atributos	5	Número de caracteres maiúsculos
	7	Número de palavras
	9	Número de informações de contato no texto
	10	Número de caracteres numéricos
	12	Número de endereços de e-mail
	13	Número de URL's no texto
	14	Número de telefones
	15	Número de palavras maiúsculas'
	16	Número termos publicitários
Usuário 6 atributos	1	Número de seguidores
	2	Informações da bio
	3	Fração do número de seguidores por seguidos
	4	Número de fotos postadas
	6	Informações de website
	11	Número de seguidos
Imagem 1 atributo	8	score de popularidade da imagem

5. Detectando diferentes tipos de usuários

Nesta seção, será avaliada a aplicação de um algoritmo de mineração de dados, a fim de diferenciar perfil dos usuários comerciais e comuns. A execução do algoritmo utiliza os atributos descritos na seção anterior, construindo um modelo de classificação por meio da análise do conjunto de instâncias de treinamento (usuários) representadas por um vetor de valores de atributos e um rótulo de classe. Em uma segunda etapa, o modelo de classificação é utilizado para classificar instâncias de teste (usuários) entre as classes: comercial e comum.

O trabalho proposto consiste em um problema de classificação plana, que é a maneira mais simples de lidar com problemas de classificação. Neste tipo de abordagem, um único classificador é treinado a partir da base de dados de treinamento contendo instâncias associadas a uma classe. Então, dada uma nova instância a ser classificada, o classificador atribui a ela uma classe de treinamento, que neste caso são: comum e comercial.

5.1. Métricas de avaliação

Medidas de qualidade de classificação são calculadas a partir de uma matriz que armazena as instâncias que foram classificadas corretamente e incorretamente para cada classe, denominada matriz de confusão. Assim, utilizou-se a seguinte matriz de confusão, mostrada na Figura 8.

		Classe predita	
		Comercial	Comum
Classe verdadeira	Comercial	<i>a</i>	<i>b</i>
	Comum	<i>c</i>	<i>d</i>

Figura 8. Matriz de confusão

Analisando a primeira linha da matriz, *a* representa a porcentagem de usuários comerciais que foram classificadas corretamente e *b* representa a porcentagem de usuários comerciais que foram classificados incorretamente como usuários comuns. Na segunda linha, *c* representa a porcentagem de usuários comuns que foram classificados incorretamente como usuários comerciais, e *d* representa a porcentagem de usuários comuns que foram classificados corretamente.

Para avaliar o desempenho do método de classificação proposto, foram consideradas as seguintes métricas, comumente utilizadas nas áreas Aprendizagem de Máquina e Recuperação de Informação [Baeza-Yates and Ribeiro-Neto, 1999]: acurácia, *recall*, Precisão e *F-measure*. *Recall* (R) e Precisão (P) são definidas como:

$$R_i = \frac{TP_i}{TP_i + FN_i}, P_i = \frac{TP_i}{TP_i + FP_i}, \quad (1)$$

onde TP_i é o percentual de instâncias corretamente classificadas como a classe i , FN_i é o percentual de instâncias que pertencem à classe i , mas que não foram classificadas como i e FP_i é o percentual de instâncias que não pertencem a classe i , mas foram classificadas incorretamente como i pelo classificador.

Os valores da diagonal principal da matriz de confusão mostrada na Figura 8 representam o *recall* de cada classe.

A métrica *F-measure* é a maneira padrão de sumarizar *recall* e precisão, definida como $F_i = 2 \times \frac{P_i \times R_i}{P_i + R_i}$. Essa métrica infere que quanto mais próximo de 1, melhor é o resultado e, quanto mais próximo de 0, pior é o resultado.

Acurácia é o acerto do classificador considerando a proporção de instâncias corretamente classificadas no total dos registros classificados: $\frac{a+d}{a+b+c+d}$.

5.2. Configuração experimental

O algoritmo de classificação escolhido para nossos experimentos foi o Random Forest (RF), que é o estado da arte em técnicas de classificação. O RF é um classificador formado por uma coleção de árvores de classificação, cada qual construída a partir de um subconjunto aleatório de instâncias da base de treinamento. Para determinar a classe de uma instância, o método combina o resultado de várias árvores de decisão por meio de um mecanismo de votação, ou seja o método percorre cada um das árvores de decisão da floresta, e em seguida cada árvore dá uma classificação (votos) para a instância [Breiman, 2001]. A predição do RF é dada pela classe que recebeu a maioria dos votos entre todas as árvores da floresta.

Com o intuito de encontrar o melhor conjunto de parâmetros do algoritmo RF, foi executado o algoritmo de otimização de parâmetros *GridSearch* para os seguintes parâmetros do RF: *numFeatures* (usado na seleção aleatória de atributos) e *numTrees* (número de árvores a serem geradas). Como resultado, encontrou-se os valores *numFeatures* = 2 e *numTrees* = 144, que foram adotados nos experimentos.

O desempenho preditivo foi medido usando o método de validação cruzada (CV-*Cross-Validation*) 10-fold. O conjunto de dados foi dividido em 10 partes igualmente distribuídas e a cada experimento, um dos subconjuntos de 10 é utilizado como conjunto de teste e os nove restantes são usados como dados de treinamento. Este processo é então repetido 10 vezes. Sendo assim, nossos resultados são médias de 10 experimentos, onde cada parte do todo foi usada uma vez como teste do classificador treinado com as outras partes.

5.3. Resultados experimentais

Devido ao fato de ter-se gerado e analisado atributos de imagem do perfil do usuário apenas ao final do projeto, optou-se por dividir os experimentos e a análise dos resultados

obtidos em dois testes: 1) conjunto de dados que não levam em consideração o atributo de imagem e 2) conjunto de dados em que é considerado o atributo de imagem. As Figuras 9 e 10 mostram as matrizes de confusão obtidas como resultado de nossos experimentos para a abordagem de classificação usando o classificador RF.

		Classe predita	
		Comercial	Comum
Classe verdadeira	Comercial	73,2%	26,8%
	Comum	21%	79%

Figura 9. Classificação plana usando *Random Forest* sem atributo imagem

A Figura 9 mostra os resultados em que não foi considerado o atributo de imagem. O *recall* das classes estão em negrito e indicam que 73,2% dos usuários comerciais e 79% dos usuários comuns foram corretamente classificados pelo RF. Pode-se observar que os resultados obtidos foram bons (*recall* > 70%), porém ainda é necessário melhorar os resultados encontrados, já que 26,8% dos usuários comerciais foram classificados incorretamente como usuários comuns. Para este teste, obteve-se *F-measure* de 73,3% para a classe comercial e 78,6% para a classe comum. A acurácia encontrada para este teste foi de 76,4%.

		Classe predita	
		Comercial	Comum
Classe verdadeira	Comercial	82,6%	17,4%
	Comum	19,2%	80,8%

Figura 10. Classificação plana usando *Random Forest* com atributo imagem

Com o intuito de obter melhores resultados, foram executados testes considerando o atributo de imagem. Os resultados são mostrados na Figura 10, onde 82,6% dos usuários comerciais e 80,8% dos usuários comuns foram corretamente classificados pelo RF. Desta vez pode-se observar que os resultados obtidos para ambas as classes alcançaram *recall* > 80%. Os valores da *F-measure* foram 82,8% para a classe comum e 80,3% para a classe comercial. O atributo de imagem se mostrou muito importante devido a melhoria significativa dos resultados.

Quando compara-se os resultados desses dois experimentos, pode-se verificar que a Precisão foi melhor no teste com atributo de imagem, sendo 78% para usuários comerciais e 84,9% para usuários comuns, em comparação aos valores de 74,2% para usuários comerciais e 78,2% para usuários comuns do teste feito sem o atributo imagem. E comparando a acurácia encontrada no primeiro teste, este segundo teste teve um melhor desempenho, alcançando 81,64% de acurácia. Desta forma, a abordagem que utilizou o atributo de imagem gerou um classificador que obteve o melhor desempenho.

6. Conclusão e trabalhos futuros

O trabalho apresentou um método para identificar perfis comerciais em uma rede especializada em publicação de fotos. Através de uma base de dados coleta do Instagram, fez-se uma rotulação manual de seus usuários em duas classes distintas: comum e comercial. Utilizando a base de dados rotulada, fez-se uma caracterização, revelando os aspectos comportamentais que diferenciam essas duas classes.

Utilizou-se a tarefa de mineração de dados, denominada classificação plana, que foi capaz de distinguir de forma eficaz usuários comuns e comerciais. Com a classificação, identificou-se corretamente 82,6% de usuários comuns e 80,8% de usuários comerciais. A imagem de perfil do usuário foi capaz de melhorar o desempenho do nosso classificador gerado em 5,25% em termos de acurácia (sendo 81,64% o melhor resultado).

É esperado que a identificação, caracterização e diferenciação de tipos de usuários apresentadas neste estudo possam ser aplicadas em outras OSNs. Além disso, o próprio Instagram poderia disponibilizar aos seus usuários uma ferramenta de identificação de usuários comerciais num mecanismo de busca dentro do próprio sistema.

Quando foram feitos testes envolvendo o atributo de imagem, foi observado que houve uma melhoria na classificação realizada, mostrando que a popularidade das imagens comerciais tende a ser menor. Pode-se então concluir que a imagem do perfil do usuário é um fator crucial para a diferenciação de ambas as classes.

Como trabalhos futuros, pretende-se aumentar nossa base de dados de usuários comerciais e além disso, aplicar algoritmos de mineração de dados para a tarefa de agrupamento com intuito de identificar diferentes tipos de usuários comerciais existentes.

Referências

- R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.
- F. Benevenuto. Redes sociais online: Técnicas de coleta, abordagens de medição e desafios futuros. *Tópicos em Sistemas Colaborativos, Interativos, Multimidi, Web e Banco de Dados*, pages 41–70, 2010.
- Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- Helen Costa, Fabricio Benevenuto, and Luiz HC Merschmann. Detecting tip spam in location-based social networks. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 724–729. ACM, 2013.
- Estado de Minas. Designer andrés iniesta acusa instagram de “roubar” sua conta e dá-la para jogador homônimo. <http://bit.ly/1oKcThD>. Acessado em abril de 2016.
- Yuheng Hu, Lydia Manikonda, Subbarao Kambhampati, et al. What we instagram: A first analysis of instagram photo content and user types. *Proceedings of ICWSM. AAAI*, 2014.
- Aditya Khosla, Atish Das Sarma, and Raffay Hamid. What makes an image popular? In *Proceedings of the 23rd international conference on World wide web*, pages 867–876. International World Wide Web Conferences Steering Committee, 2014.
- Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007.
- J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.
- Thiago H Silva, Pedro OS Vaz de Melo, Jussara M Almeida, and Antonio AF Loureiro. Uma fotografia do instagram: Caracterização e aplicação. *Proc. of XXXII SBRC'13*, 2013.

Caracterizando Departamentos e Programas de Computação Utilizando Análise de Redes Sociais e Bibliometria

Luciano Digiampietri¹, Ricardo Linden², Lenin Barbosa¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (EACH-USP)

²CEPEL/Faculdade Salesiana Maria Auxiliadora (FSMA)

digiapietri@usp.br, rlinden@pobox.com, lenin.ferreira@gmail.com

Abstract. *In this paper we compare the graduate programs in Computer Science of levels 6 and 7 (which CAPES assume to be equivalent to the international best) and those programs best ranked in three international rankings. We used both bibliometric metrics and social networks analysis and found out that there is a great difference between the results achieved by the national and international programs. The results from a principal component analysis show that the Brazilian programs are a class of their own and do not share the characteristics of the best international graduate programs.*

Resumo. *Neste artigo apresentamos uma comparação entre os programas de pós-graduação em computação de nível 6 e 7 (os chamados de nível internacional pela CAPES¹) e os programas melhores colocados em três rankings internacionais. Foram usados métricas bibliométricas e métricas de análise de redes sociais e pôde-se verificar uma grande diferença de resultados entre os programas nacionais e os internacionais. O resultado da análise de componentes principais mostra que os programas brasileiros residem em uma classe própria, não possuindo as características dos melhores programas internacionais.*

1. Introdução

A avaliação da qualidade de departamentos ou programas de pós-graduação (muitas vezes relacionada a análises bibliométricas) tem se tornado cada vez mais relevante, uma vez que os recursos de fomento são escassos em comparação com os pesquisadores que os desejam. Assim, necessita-se de uma análise mais abrangente e objetiva, tanto para melhor alocar os recursos quanto para elaborar políticas científicas eficazes.

Entender as características que levam um grupo a obter reconhecimento internacional pode servir de base para o desenvolvimento de outros grupos. Porém, avaliar programas ou departamentos é uma tarefa complexa, envolvendo a análise de diferentes métricas de características quantitativas e qualitativas [Digiampietri et al. 2014], sobre as quais não há consenso. Além da complexidade e possível subjetividade, os trabalhos da literatura usam índices diversos ou fontes de dados diferentes para a análise bibliométrica.

Na avaliação de grupos acadêmicos e, especialmente quando se foca em departamentos ou programas de pós-graduação, há algumas métricas mais utilizadas [Laender et al. 2008], muitas das quais usaremos neste trabalho. Além de análises bibliométricas, a

¹https://www.capes.gov.br/images/stories/download/avaliacao/COMPUTACAO_05mar10.pdf

análise de redes sociais vem sendo utilizada para tentar medir a influência ou importância de um grupo em um dado contexto, bem como identificar padrões de relacionamentos entre os membros de cada grupo. No Brasil, para a avaliação dos programas de pós-graduação a CAPES² (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) atribui aos programas de pós-graduação notas de três a sete, sendo que os programas com notas 6 ou 7 são considerados os programas de nível internacional.

Este trabalho visa a caracterizar 28 departamentos ou programas de pós-graduação em ciência da computação (oito brasileiros e vinte internacionais) considerando medidas bibliométricas e oriundas da análise de redes sociais, analisar a correlação entre estas medidas e o ranqueamento provido por três fontes internacionais e avaliar se os programas brasileiros apresentam padrões diferentes dos grupos internacionais avaliados.

O restante deste artigo está organizado da seguinte maneira. A Seção 2 contém a descrição de alguns conceitos básicos e trabalhos correlatos. A Seção 3 apresenta a metodologia utilizada. Na Seção 4 os resultados são apresentados e analisados. Por fim, a Seção 5 contém as conclusões e trabalhos futuros.

2. Conceitos Básicos e Trabalhos Correlatos

Os principais conceitos envolvidos no presente trabalho estão ligados a análises bibliométricas e de redes sociais e são descritos nesta seção. Existem diversos termos relacionados à análise quantitativa da produção científica, e dois dos mais usados são **Bibliometria** (*bibliometrics*) e **Cientometria** (*scientometrics*), que estudam os aspectos quantitativos da produção e da ciência, respectivamente.

2.1. Análise de Redes Sociais

Rede social é uma estrutura composta por indivíduos (pessoas ou organizações) que são conectadas por um ou mais tipos de relações, como amizade, crença ou trabalho. O conceito de redes sociais surgiu há séculos nas áreas de humanas, porém, apenas nas últimas décadas, os estudos em redes sociais se intensificaram analisando tanto as características individuais de seus componentes, como também as características estruturais da rede [Lemieux and Ouimet 2008], baseando-se em dados que contemplam milhares ou milhões de indivíduos. Um dos métodos mais comuns para a representação computacional destas redes, também usado aqui, é utilizando grafos, nos quais cada indivíduo da rede social é um nó (ou vértice), e cada relação entre indivíduos é uma aresta, direcionada ou não.

2.2. Análise de grupos acadêmicos

Tipicamente, a análise de grupos acadêmicos é realizada em duas vertentes: análises bibliométricas e de redes sociais. Os grupos a serem analisados podem ser formados de diferentes maneiras, como por exemplo, de acordo com a formação/titulação dos indivíduos, sua localização geográfica, com suas áreas de atuação ou com o local/instituição em que trabalha (universidade, departamento ou programa de pós-graduação).

A colaboração científica é influenciada por diversos fatores como proximidade geográfica e os relacionamentos anteriores dos pesquisadores (como uma relação de orientação) [Menezes et al. 2009]. Para descobrir redes sociais acadêmicas e avaliar

²<http://www.capes.gov.br/>

grupos de pesquisa precisa-se de informação tipicamente dispersa e de diferentes naturezas. Muitos desses tipos de informação são agrupados nos currículos dos pesquisadores, o que ajuda a enfrentar alguns dos desafios encontrados ao se utilizar apenas bases de citações bibliográficas, tais como: homonímia e polissemia, dispersão da produção bibliográfica em diferentes fontes, afiliação do pesquisador, relações de orientação, áreas de interesse, identificação das colaborações, etc.

Ao se pensar na avaliação de toda a rede social acadêmica de um país, o primeiro obstáculo (e muitas vezes um dos maiores) é a obtenção do grande volume de informação relacionada, normalmente distribuída em muitas fontes (como páginas Web, relatórios institucionais, bibliotecas digitais internacionais, etc). No entanto, esta atividade é simplificada no Brasil pela existência da Plataforma Lattes [Digiampietri et al. 2012]. Por outro lado, para a análise de grupos da área de ciência da computação utilizando dados internacionais, a principal fonte utilizada tem sido o projeto DBLP³ [Laender et al. 2008, Freire and Figueiredo 2011].

2.3. Trabalhos correlatos

A relevância desta área pode ser vista no grande número de trabalhos correlatos existentes, entre os quais podemos destacar alguns, descritos a seguir. Por questões de limitação de espaço, não foi possível descrever todos os artigos relevantes da área.

Os estudos acadêmicos mais relacionados ao presente artigo são os trabalhos de caracterização das redes sociais acadêmicas por meio do cálculo de métricas utilizando, por exemplo, teoria dos grafos/teoria de redes [Ulrik and Erlebach 2005]. Estas pesquisas incluem a tentativa de oferecer uma visão geral sobre a produção científica brasileira [Leite et al. 2011] ou grupos de pesquisa específicos [Wainer and Vieira 2013].

Entre os trabalhos com enfoque bibliométrico, algumas das principais medidas bibliométricas utilizadas foram número de publicações, índice-h e número de citações [Hirsch 2005], sendo que alguns trabalhos ponderam a participação dos autores nas publicações de acordo com a posição do pesquisador na lista de autorias [Duffy et al. 2011]. Além da análise de grupos, alguns destes trabalhos também analisam se há influência/correlação entre gênero e as demais medidas.

Já Menezes et al. 2009 combinaram a análise bibliométrica com a de redes sociais para analisar a área de ciência da computação e suas subáreas em várias regiões do mundo (Estados Unidos, Europa e Brasil) no período de 1994 a 2006. Os autores destacaram os aspectos comuns e as diferenças observadas entre as redes de cada região.

O trabalho de Franceschet 2011 usou dados do DBLP a fim de analisar as redes de coautorias na área de ciência da computação e compará-las com dados de redes de outras áreas. Eles observaram que, em comparação com outras áreas, o nível de colaboração (em termos de números de coautores nas publicações) em ciência da computação é baixo.

Em uma análise da produtividade dos departamentos de computação brasileiros, Laender et al. 2008 usaram dados do DBLP e compararam indicadores de programas brasileiros considerados de nível internacional com alguns programas da América do Norte e Europa, concluindo que estes programas brasileiros atingiram a maturidade.

³<http://dblp.uni-trier.de/db/>

Também utilizando dados do DBLP, Freire and Figueiredo 2011 analisaram a rede social acadêmica brasileira na área de ciência da computação. Os autores encontraram *super peers*, isto é, alguns indivíduos com grau muito acima dos demais da rede. Os autores também propuseram uma métrica para a avaliação de redes de colaboração que considera a importância de um nó na conexão de indivíduos de diferentes grupos.

3. Metodologia

A metodologia deste trabalho foi dividida em quatro atividades: seleção da amostra; obtenção dos dados; cálculo das métricas; e análise dos resultados. Ao se tratar de avaliação internacional, muitas vezes são considerados departamentos, unidades ou universidades enquanto que na avaliação nacional, tipicamente são considerados os programas de pós-graduação (que podem, inclusive, conter docentes de diferentes universidades). Ao longo deste texto o termo departamento será utilizado, por simplicidade, tanto para se referir a departamentos ou unidades quanto a programas de pós-graduação.

Para o presente trabalho optou-se por **selecionar** vinte departamentos internacionais com destaque na área de ciência da computação. Para isto foram utilizados três ranqueamentos internacionais de 2016, referentes a dados do ano de 2015: Shanghai⁴ (considerando os 200 melhores departamentos do mundo), Top Universities⁵ e The Guardian⁶ (considerando os 100 melhores departamentos do mundo). Adicionalmente, foram inseridos os oito programas de pós-graduação em ciência da computação no Brasil considerados de nível internacional (isto é, com nota 6 ou 7 atribuída pela CAPES). Devido ao objetivo do presente trabalho que visa a caracterizar os principais departamentos da área de computação do mundo e os principais programas de pós-graduação em computação no Brasil, além dos departamentos melhor ranqueados também foram feitas algumas escolhas para manter a diversidade da amostra e para incluir departamentos com ranqueamentos mais parecidos com os programas brasileiros. A Tabela 3 apresenta os 28 departamentos ou programas analisados e suas posições nos três ranqueamentos utilizados. Estes departamentos estão ordenados pelo ranqueamento médio obtido.

Para cada um dos departamentos internacionais, a lista de seus pesquisadores foi obtida do site do respectivo departamento. Para os programas de pós-graduação brasileiros, a lista dos pesquisadores foi obtida do sistema Sucupira⁷. Ao todo, 1.232 nomes de pesquisadores foram identificados.

Neste trabalho foram **obtidos dados** de duas fontes principais. Os dados sobre as publicações dos pesquisadores selecionados foram obtidos do projeto DBLP (*Digital Bibliography & Library Project*). Já os dados sobre as citações das publicações foram obtidos do site *Microsoft Academic Search*⁸, conforme será apresentado a seguir.

Os mantenedores do projeto DBLP realizam um processo de desambiguação de nomes para evitar o cadastro repetido de um mesmo autor e a base conta com mais de um milhão e meio de registros de autores. Para realizar o casamento entre a lista de nomes de pesquisadores de cada departamento e os registros do DBLP adotou-se a estratégia para

⁴Academic Ranking of World Universities: <http://www.shanghairanking.com/>

⁵Worldwide university rankings: <http://www.topuniversities.com/>

⁶The Guardian: <http://www.theguardian.com/education/universityguide>

⁷<https://sucupira.capes.gov.br/>

⁸<http://academic.research.microsoft.com/Microsoft Academic Search>

Tabela 1. Departamentos ou Programas de Computação Analisados

Universidade	Abrev. usada	País	Shanghai	Top Universities	The Guardian
MIT - Electrical Eng. and Comp. Science Dep.	MIT EECS	Estados Unidos	2	1	1
Stanford University	Stanford	Estados Unidos	1	2	2
Harvard - School of Eng. and Applied Sciences	Harvard SEAS	Estados Unidos	4	4	4
Carnegie Mellon University	CMU	Estados Unidos	6	5	4
University of California - Berkeley	UCBerkeley	Estados Unidos	3	7	6
Princeton University	Princeton	Estados Unidos	5	10	9
Swiss Federal Institute of Technology in Zurich	ETH Zurich	Suíça	17	8	9
University of Cambridge	Cam	Reino Unido	23	6	8
University of Toronto	Toronto	Canada	11	11	16
University of Oxford	Oxford	Reino Unido	34	3	3
University of California - Los Angeles	UCLA	Estados Unidos	9	16	17
Hong Kong University of Science and Tech.	HKUST	China	39	14	8
The University of Texas - Austin	UTexas	Estados Unidos	7	34	22
Cornell University	Cornell	Estados Unidos	8	26	31
California Institute of Technology	Caltech	Estados Unidos	11	28	27
University of Illinois	Illinois	Estados Unidos	13	30	30
Chinese University of Hong Kong	CUHK	China	37	22	18
Ecole Polytechnique Federale de Lausanne	EPFL	Suíça	35	18	29
University of Washington	UW	Estados Unidos	51-75	28	20
Universidade Estadual de Campinas	UNICAMP	Brasil	101-150	51-100	51-100
Universidade de São Paulo - São Carlos	USP São Carlos	Brasil	151-200	51-100	51-100
Universidade de São Paulo - São Paulo	USP São Paulo	Brasil			
Boston University	BU	Estados Unidos	101-150	101-150	
Universidade Federal do Rio Grande do Sul	UFRGS	Brasil		151-200	
Universidade Federal do Rio de Janeiro	UFRJ	Brasil		151-200	
Universidade Federal de Minas Gerais	UFMG	Brasil		151-200	
Universidade Federal de Pernambuco	UFPE	Brasil		201-250	
Pontifícia Universidade Católica (RJ)	PUC-Rio	Brasil		201-250	

resolução de nomes apresentada por Mugnaini et al. (2012), que utiliza um sistema de pontuação para identificar o melhor casamento entre dois nomes considerando que eles podem estar no formato utilizado em citações bibliográficas (isto é, com a ausência de alguns “nomes do meio” e/ou com abreviações). Por meio deste processo, foram encontrados 1.170 registros de autores. Estes foram utilizados para identificar 73.232 registros de diferentes artigos (25.360 artigos em periódicos e 47.872 artigos em eventos).

Cada registro de publicação possui a lista de autores representados da maneira que estes autores são cadastrados na base do DBLP. Assim, a partir da lista de publicações é bastante simples criar as redes de coautoria dos pesquisadores envolvidos. Neste projeto foram calculadas métricas de 29 redes sociais de coautoria. A primeira rede possui como nó cada um dos departamentos e as arestas representam relações de coautorias entre pesquisadores de dois departamentos (Figura 1). A legenda de cores usada em todos os grafos apresentados neste artigo é apresentada na Figura 2.

A Figura 1 contém um grafo organizado automaticamente de acordo com as ligações. Neste grafo podemos ver que todos os programas brasileiros estão à esquerda do grafo e apresentam poucas ligações tanto entre si quanto com instituições estrangeiras, indicando um nível mais baixo de pesquisas conjuntas. A densidade de arestas em cada nó representando um departamento internacional é, na média, muito maior do que as nacionais, indicando que estes trabalham mais em cooperação com outras instituições, possivelmente aproveitando sinergias e “importando” ideias novas.

Adicionalmente foi gerada uma rede para cada departamento, na qual cada pesquisador é um nó e cada aresta representa relações de coautoria entre pesquisadores. Apenas

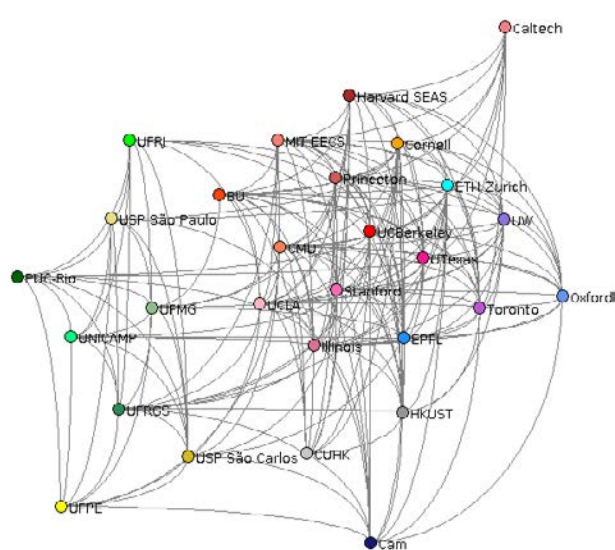


Figura 1. Rede de coautorias

- University of California - Berkeley
- Boston University
- California Institute of Technology
- Carnegie Mellon University
- Cornell University
- MIT - Electrical Engineering and Computer Science Department
- Harvard - School of Engineering and Applied Sciences
- Princeton University
- Stanford University
- The University of Texas - Austin
- University of California - Los Angeles
- University of Illinois
- University of Toronto
- University of Washington
- University of Oxford
- University of Cambridge
- Ecole Polytechnique Federale de Lausanne
- Swiss Federal Institute of Technology in Zurich
- Chinese University of Hong Kong
- Hong Kong University of Science and Technology
- Pontificia Universidade Católica do Rio de Janeiro
- Universidade Federal de Minas Gerais
- Universidade Federal do Rio Grande do Sul
- Universidade Federal do Rio de Janeiro
- Universidade Estadual de Campinas
- Universidade de São Paulo - São Paulo
- Universidade de São Paulo - São Carlos
- Universidade Federal de Pernambuco

Figura 2. Legenda de cores

para ilustrar estas relações entre os pesquisadores estudados, as Figuras 3 e 4 apresentam a rede de coautorias incluindo todos os pesquisadores estudados. Na Figura 3 são apresentadas as coautorias entre pesquisadores de diferentes departamentos (arestas em cinza) e entre pesquisadores do mesmo departamento (arestas coloridas). Já a Figura 4 contém apenas as relações de coautoria entre pesquisadores do mesmo departamento.

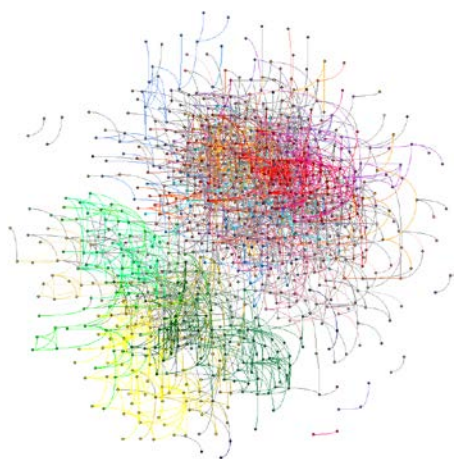


Figura 3. Rede de coautorias - todas as arestas

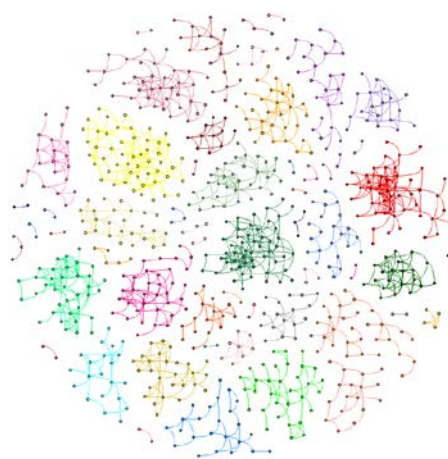


Figura 4. Rede de coautorias - arestas internas

O título de cada um dos artigos identificados foi buscado no site *Microsoft Academic Search*, e criamos um programa que verifica se cada artigo foi encontrado na busca (usando casamento aproximado de *strings* e permitindo uma distância de edição de no máximo 10% do tamanho do título do artigo) e, em caso afirmativo, o programa salva o número de citações recebido pelo artigo. Ao todos, os 73.232 artigos receberam 1.197.416 citações (658.721 para artigos em periódicos e 538.695 para artigos em eventos).

A escolha das métricas aplicadas ao estudo de cada rede social está diretamente relacionada com objetivos específicos do estudo a ser realizado [Scott 2009]. No caso

deste trabalho, com base nos dados obtidos, **dois tipos de métricas foram calculados:** bibliométricas e de análise de redes sociais. As medidas bibliométricas calculadas estão relacionadas ao número de artigos publicados e às citações recebidas. Oito das 16 medidas bibliométricas calculadas são apresentadas na Tabela 3, as demais oito medidas são as mesmas, só que calculadas em uma janela de tempo de cinco anos (de 2010 a 2014).

Tabela 2. Medidas Bibliométricas Utilizadas

Abrev. Utilizada	Descrição
Pubs	Quantidade total de publicações em periódicos ou anais de eventos dos pesquisadores de um dado departamento.
PubsCo	Quantidade total de publicações em periódicos ou anais de eventos em coautoria com pesquisadores de outro departamento do estudo.
Cits	Quantidade total das citações às publicações dos artigos do departamento.
CitsCo	Quantidade total das citações às publicações dos artigos do departamento envolvidas em coautoria com pesquisadores de outro departamento.
PubsPP	Taxa de publicações por pesquisador.
PubsCoPP	Taxa de publicações em coautoria por pesquisador.
CitsPP	Taxa de citações por pesquisador.
CitsCoPP	Taxa de citações por pesquisador envolvendo as publicações em coautoria com pesquisadores de outro departamento.

Para a análise de redes sociais, dois tipos de redes de coautoria foram construídas: uma em que cada nó corresponde a um departamento e uma aresta corresponde à relação de coautoria entre pesquisadores dos dois departamentos ligados pela aresta; e uma rede para cada departamento, na qual cada pesquisador corresponde a um nó e as arestas representam relações de coautoria entre pesquisadores de um departamento. Para o primeiro tipo de rede foram calculadas cinco medidas de centralidade individuais (ou locais) de cada nó (departamento), já para o segundo tipo foram calculadas nove métricas globais da rede de cada departamento. A Tabela 3 apresenta as métricas utilizadas.

Tabela 3. Métricas da Análise de Redes Sociais Utilizadas

Abrev. Utilizada	Descrição
Medidas Locais	
Betweenness _l	Centralidade de intermediação (betweeness) - baseada na frequência em que um dado vértice aparece entre todos os caminhos mínimos da rede.
Closeness _l	Centralidade de proximidade (closeness)
Degree _l	Centralidade de grau - baseada no grau do vértice.
Eigen _l	Centralidade de autovalor (eigenvalue) - baseada na importância dos vértices que estão ligados ao vértice em análise.
Page-Rank _l	Centralidade PageRak - baseada na quantidade e qualidade das ligações existentes envolvendo o vértice em análise.
Medidas Globais	
Clique _g	Tamanho da clique máxima - tamanho da maior clique na rede.
MCC _g	Média dos caminhos mínimos - valor médio do caminho mínimo entre todos os pares de vértices da rede.
C.Clusterização _g	Coefficiente de clusterização - medida da transitividade das relações da rede.
Degree _g	Centralização de grau - medida do quão importante o vértice mais importante da rede é em relação aos demais (em relação ao grau).
Closeness _g	Centralização de proximidade - medida do quão importante o vértice mais importante da rede é em relação aos demais (em relação à proximidade).
Eigen _g	Centralização de autovalor - medida do quão importante o vértice mais importante da rede é em relação aos demais (em relação ao autovalor).
Assort.Grau	Assortatividade de grau - medida da tendência de nós de mesmo grau estarem ligados.
Diâmetro _g	Diâmetro - diâmetro da rede (distância mínima entre os dois nós mais afastados na rede).
Densidade _g	Densidade da rede - relação entre o número de arestas da rede e o número máximo de arestas possíveis na rede.

A **análise dos resultados** apresentada a seguir baseia-se na importância das métricas dos programas, considerando-se as análises de componentes principais e da correlação de postos entre características calculadas e ranqueamento dos departamentos.

4. Resultados

Os 28 departamentos (ou programas de pós-graduação) foram caracterizados por 34 métricas: o número de pesquisadores em cada departamento (Pesq), as 16 medidas bibliométricas apresentadas, as cinco medidas de centralidade de cada departamento na rede envolvendo os departamentos, nove medidas globais extraídas das redes de cada departamento, e as posições desses departamentos nos três ranqueamentos utilizados.

Para cada uma das medidas os departamentos foram ranqueados da 1ª posição (maior valor para a métrica) até a 28ª (menor valor para a métrica), conforme apresentado na Figura 5. Os fundos das células estão coloridos de acordo com o ranqueamento: quanto mais próximo de 1 a célula tem seu fundo com um verde mais intenso, por outro lado, células com valores próximos a 28 são coloridas com fundo vermelho. Destaca-se que o departamento com maior valor para uma dada métrica é ranqueado em 1º lugar, mas valores altos de uma métrica não necessariamente é uma característica positiva.

Departamento	Pesq	PubS	PubCo	Cits	CitsCo	PubCPP	PubCoPP	CitsPP	CitsCoPP	PubS_5anos	PubCo_5anos	Cits_5anos	CitsCo_5anos	PubCPP_5anos	PubCoPP_5anos	CitsPP_5anos	CitsCoPP_5anos	Betweenness_I	Closeness_I	Degree_I	PageRank_I	Eigen_I	Clque_g	MCC_g	C.Custeizacao_g	Degree_g	Closeness_g	Eigen_g	Assort_Gru	Diametro_g	Densidade_g	Shanghai	TopUniversities	TheGuardian		
MIT EECS	17	11	6	5	5	4	4	2	3	13	13	8	19	6	11	5	18	17	11	11	13	10	10	13	27	18	24	9	18	12	24	2	1	2	1	
Stanford	11	12	3	6	3	9	3	4	1	12	6	5	5	9	7	6	7	5	3	3	4	3	4	7	18	6	19	6	20	3	25	1	2	2	2	
Oxford	19	13	19	13	14	7	21	11	14	9	24	4	6	4	21	2	5	26	18	18	20	17	10	11	26	20	11	7	6	15	15	3	1	3		
Harvard SEAS	24	20	11	16	8	11	2	10	4	21	10	21	10	8	2	15	6	21	11	11	12	11	23	20	20	14	12	20	23	22	14	4	4	4		
CMU	2	5	2	4	4	13	7	9	7	6	2	2	18	13	8	8	19	3	3	3	3	4	10	5	26	16	23	4	14	3	28	6	5	4	4	
UCBerkeley	4	1	1	1	1	5	1	1	2	3	1	6	4	7	3	9	9	2	1	1	2	1	1	14	9	15	5	15	6	6	23	3	7	6	6	
Cam	27	26	27	18	26	15	26	17	22	27	27	20	8	17	24	10	2	13	18	18	18	18	23	26	23	2	14	14	24	26	7	14	6	7	7	
ETH Zurich	4	9	16	11	13	16	22	18	17	18	15	11	28	15	19	14	26	10	7	7	7	8	10	12	24	7	13	12	25	12	18	13	8	9		
Princeton	3	10	4	3	2	21	16	12	5	10	3	19	20	21	15	28	23	5	5	5	6	5	23	2	25	26	18	21	1	26	3	3	3	9	9	
HKUST	12	8	17	15	10	6	18	19	19	5	21	15	17	3	22	19	17	22	14	14	15	16	10	14	20	15	13	1	6	13	18	11	7	3	9	
Toronto	17	15	14	14	15	17	13	15	15	18	14	9	22	19	12	7	21	16	9	9	10	9	10	1	17	12	16	7	22	1	16	11	10	11	11	
UCLA	20	2	9	9	10	1	6	3	10	1	3	7	11	1	3	10	4	5	5	5	7	10	25	16	4	27	2	13	25	20	9	12	12	1	12	
CUHK	23	26	24	23	20	25	19	22	20	25	26	27	26	26	27	26	25	27	22	22	27	20	23	27	14	11	21	16	27	26	1	17	14	13	15	
Cornell	6	6	5	2	6	12	10	7	9	7	8	10	9	14	13	13	15	11	7	7	8	6	10	3	22	24	25	3	5	3	27	8	15	19	6	
UTexas	9	7	7	7	7	8	9	6	8	11	4	16	13	12	4	20	13	8	11	11	11	15	4	6	12	22	11	10	15	6	22	7	19	15	15	
FPTI	13	4	12	8	12	2	11	5	12	4	16	13	15	2	18	16	12	7	9	9	9	13	21	4	19	13	18	5	17	6	19	16	13	17	17	
UW	21	16	10	12	9	10	5	8	6	17	11	23	21	10	5	22	20	25	14	14	17	12	10	15	21	8	22	19	26	22	17	19	16	14	14	
Caltech	28	28	28	19	24	19	25	16	18	28	28	24	27	16	23	11	27	28	28	28	28	21	28	28	28	28	28	1	28	26	5	10	16	16	16	16
Illinois	8	3	8	10	11	3	12	14	13	2	12	14	12	5	17	21	16	1	1	1	1	2	10	8	13	23	17	18	9	6	21	12	18	18	18	
UNICAMP	10	18	15	22	19	26	15	24	21	16	8	25	14	22	9	24	14	23	26	22	23	26	1	21	2	9	3	25	3	19	8	20	20	20	20	
USP São Paulo	14	27	20	26	21	28	23	25	25	24	20	28	23	28	20	28	24	15	22	22	24	25	4	19	1	17	10	26	12	12	3	22	20	20	20	
USP São Carlos	21	24	21	24	22	25	23	24	20	23	24	22	17	1	3	18	16	1	3	14	21	21	21	24	4	23	3	10	6	17	19	22	20	20	20	
BU	25	23	22	17	17	18	17	11	11	26	23	18	24	23	14	12	22	19	14	14	16	14	10	22	10	5	9	23	4	19	6	20	23	23	23	
UFMG	15	21	13	20	16	23	14	21	16	19	7	3	1	21	6	4	1	12	20	20	19	22	3	16	5	3	2	21	11	12	11	24	24	23	23	23
UFRRS	7	14	25	24	25	20	27	26	26	14	19	12	2	20	26	18	4	9	17	17	14	19	4	18	7	21	8	17	2	12	9	24	24	23	23	
UFRRJ	15	19	21	27	28	22	24	27	28	20	25	26	27	25	26	27	18	22	22	25	23	10	9	8	27	7	24	10	12	12	24	24	23	23	23	
UFPE	2	17	26	28	27	27	28	28	27	15	21	17	7	24	28	23	11	24	27	27	26	28	10	17	6	19	4	28	3	12	10	24	27	23	23	
PUC-Rio	25	22	18	21	22	14	8	20	23	23	17	22	16	11	10	17	8	20	22	22	22	22	27	4	24	4	1	1	22	16	22	2	24	27	23	23

Figura 5. Ranqueamento dos programas de acordo com as métricas utilizadas

A Figura 6 contém a correlação de postos de Spearman calculada para todos os pares de métricas utilizados no presente estudo. As três últimas linhas e colunas contêm as correlações entre cada medida e os três ranqueamentos internacionais utilizados. Destaca-se que os três ranqueamentos possuem uma correlação maior ou igual a 0,76 entre si. Ao se comparar os valores das métricas calculadas com os valores dos ranqueamentos, correlações positivas significam que quanto maior a medida melhor será a posição do departamento no ranqueamento (ou seja, o departamento será melhor ranqueado), de maneira simétrica, correlações negativas significam que quanto maior o valor da métrica pior será a posição do departamento no ranqueamento (ele será pior ranqueado).

As medidas com maiores correlações positivas com os ranqueamentos são todas bibliométricas, apresentadas conforme seus valores de correlação média com os três ranqueamentos: CitsCoPP, Cits, CitsPP e CitsCo. Observa-se que as citações recebidas por pesquisador em trabalhos em coautoria com pesquisadores de outro departamento foi a medida mais correlacionada com os ranqueamentos (seguida pelo total de citações recebidas pelos artigos publicados do departamento). Isto é compatível com o conceito que se

A língua não é uma limitação ao número de citações, posto que os dados das publicações foram obtidos do DBLP, que indexa principalmente artigos em revistas e conferências internacionais. Logo, este fator não afetou de forma relevante esta diferença significativa.

Das medidas de análise de redes sociais, a que apresentou a maior correlação positiva com os ranqueamentos foi a centralidade de autovalor ($Eigen_l$). Todas as medidas de centralidade analisadas (considerando a rede em que cada departamento foi representado como um nó) possuem correlações positivas com o ranqueamento.

Ao se analisar as correlações negativas com os ranqueamentos destacam-se três medidas de redes (com correlações, na média, abaixo de -0,5). A primeira é o coeficiente de clusterização ($C.Clusterização_g$) com correlações entre 0,6 e 0,67 com os ranqueamentos internacionais. Isto indica que quanto maior a transitividade entre as relações de um programa pior o seu ranqueamento.

A segunda medida de destaque é a correlação negativa entre os ranqueamentos e a centralização de proximidade ($Closeness_g$), indicando que uma maior importância do pesquisador mais central de um dos departamentos em relação aos demais (considerando-se a proximidade como medida de importância) leva a uma tendência do departamento ser pior ranqueado. Isto é compatível com a ideia de que a diversidade de pensamento leva a resultados melhores da pesquisa e, por conseguinte, do ranqueamento.

Por fim, destaca-se a densidade da rede de cada um dos departamentos ($Densidade_g$). A correlação indica que quanto mais densa a rede é, pior a posição do departamento, indicando que departamentos que tem muitas arestas internas tendem a ter um ranqueamento pior. Isto é compatível com o conceito de que a colaboração com pessoas “de fora” traz novas ideias e perspectivas, permitindo novas soluções e avanços.

Para estes três índices, a maioria dos programas brasileiros encontram na zona verde-verde amarelada, perto dos 10 primeiros lugares. Na métrica $C.Clusterização_g$ os brasileiros estão nas 8 primeiras posições. Na $Closeness_g$, 10 primeiras. Para a medida $Densidade_g$, sete dos oito programas estão nas 12 primeiras posições. Estas medidas indicam um relacionamento mais intenso entre pesquisadores brasileiros de um mesmo programa e a dependência de cada programa de seu pesquisador mais central maior do que os melhores programas internacionais.

Foi utilizada a técnica de Análise de Componentes Principais para mapear as 31 métricas dos departamentos (todas as métricas menos as classificações nos ranqueamentos internacionais) para mapear os departamentos em um espaço bidimensional. A Figura 7 apresenta os 28 departamentos mapeados no espaço das duas componentes principais.

A primeira componente (eixo das abscissas na Figura 7) representa 51,4% da variância dos dados originais e está mais relacionada a medidas de importância dos programas (citações recebidas e as centralidades $PageRank_l$, $Degree_l$ e $Closeness_l$). Já a segunda componente (eixo das ordenadas) representa cerca de 13% da variância dos dados e está mais relacionada ao modo de interação dos pesquisadores de um departamento. As três medidas mais intensamente representadas na segunda componente são Assortatividade de Grau, Coeficiente de Clusterização e Tamanho da Clique Máxima. A componente principal teve correlação maior ou igual a 0,59 com todos os ranqueamentos internacionais utilizados, destacando-se a correlação de 0,81 com o ranqueamento Shanghai.

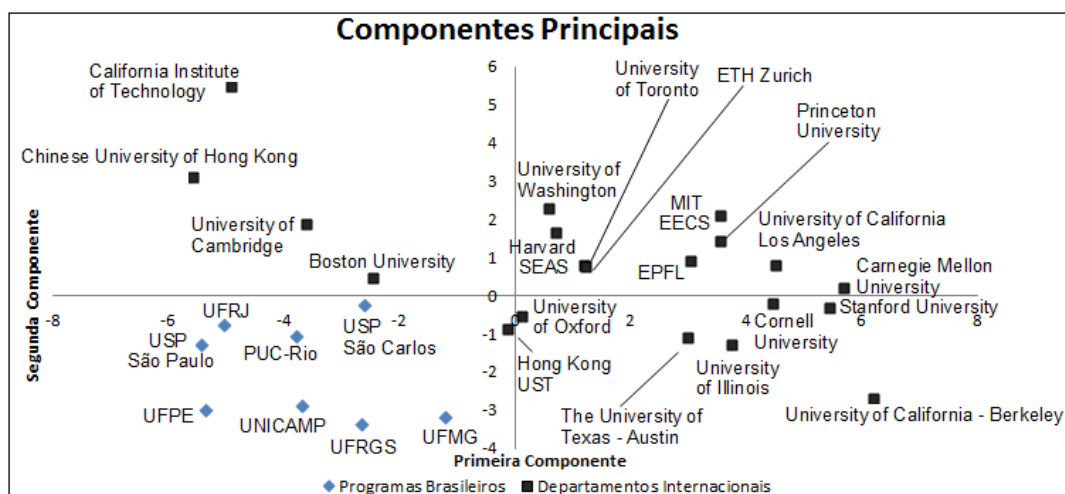


Figura 7. Plotagem das duas componentes principais

Destaca-se que nove dos dez melhor ranqueados encontram-se no lado direito da figura (valores positivos no eixo das abscissas). Por outro lado, dos dez piores ranqueados, nove se encontram do lado esquerdo da figura. Chama a atenção na figura que os oito programas de pós-graduação brasileiros analisados encontram-se na região esquerda inferior da figura (e são os únicos programas/departamentos nessa região), sugerindo que eles possuem características parecidas entre si e bastante diferente dos demais departamentos.

5. Conclusões

Neste trabalho, caracterizamos 20 departamentos internacionais da área de computação e oito programas de pós-graduação utilizando-se diferentes medidas bibliométricas e da análise de redes sociais. Foram calculadas as correlações entre todas as medidas e entre três ranqueamentos internacionais de universidades/departamentos.

A análise dos grafos gerados mostra que os programas brasileiros têm um nível de cooperação entre si e com departamentos internacionais abaixo do que é comum nestes últimos. Pode-se observar, por meio da análise de componentes principais, que os programas brasileiros não apresentam características similares àqueles que são considerados os melhores programas mundiais. Não é possível ainda estabelecer a razão desta diferença, nem se ela é positiva ou negativa, mas este trabalho indica uma desarmonia entre os critérios adotados pela CAPES e aqueles utilizados pelos ranqueamentos internacionais. Nesta análise também se vê que os oito programas de pós-graduação brasileiros analisados se diferenciam dos demais departamentos e se encontram muito próximos entre si. Assim, é possível inferir que existe uma “ciência à brasileira”, cujos critérios de qualidade parecem ser distintos daqueles adotados mundialmente.

Observou-se também que medidas bibliométricas estão fortemente correlacionadas com a posição dos programas nesses ranqueamentos, indicando a ausência de um “preconceito anti-periferia”. As posições dos programas são corretas dentro do modelo de valoração de pesquisa usados pelos mesmos (e que aproximadamente coincidem nos três ranqueamentos internacionais apresentados).

Como trabalho futuro pretende-se estender a presente pesquisa analisando uma quantidade maior de grupos e utilizando métricas adicionais.

Agradecimentos

Este trabalho foi financiado pelo CNPq (processos 306046/2013-0 e 477246/2013-3).

Referências

- Digiampietri, L. A., Mena-Chalco, J., Perez-Alcazar, J. J., Tuesta, E. F., Delgado, K., and Mugnaini, R. (2012). Minerando e caracterizando dados de currículos Lattes. In *BraSNAM 2012*, page 12, Curitiba, PR, Brasil.
- Digiampietri, L. A., Mena-Chalco, J. P., Vaz de Melo, P. O. S., Malheiro, A. P. R., Meira, D. N. O., Franco, L. F., and Oliveira, L. B. (2014). BraX-Ray: An X-Ray of the Brazilian Computer Science Graduate Programs. *PLoS ONE*, 9(4):e94541.
- Duffy, R. D., Jadidian, A., Webster, G. D., and Sandell, K. J. (2011). The research productivity of academic psychologists: assessment, trends, and best practice recommendations. *Scientometrics*, 89(1):207–227.
- Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology*, 62(10):1992–2012.
- Freire, V. and Figueiredo, D. (2011). Ranking in collaboration networks using a group based metric. *Journal of the Brazilian Computer Society*, pages 1–12.
- Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16569–16572.
- Laender, A., de Lucena, C., Maldonado, J., de Souza e Silva, E., and Ziviani, N. (2008). Assessing the research and education quality of the top Brazilian Computer Science graduate programs. *ACM Special Interest Group on Computer Science Education*, 40(2):135–145.
- Leite, P., Mugnaini, R., and Leta, J. (2011). A new indicator for international visibility: exploring Brazilian scientific community. *Scientometrics*, 88:311–319.
- Lemieux, V. and Ouimet, M. (2008). *Análise Estrutural das Redes Sociais*. Instituto Piaget.
- Menezes, G. V., Ziviani, N., Laender, A. H. F., and Almeida, V. (2009). A geographical analysis of knowledge production. In *Computer Science In Proceedings of the 18th international conference on World Wide Web*, pages 1041–1050.
- Mugnaini, R., Digiampietri, L. A., de Oliveira, L. C., and Ferreira, S. M. S. P. (2012). Normalização de nomes de autores em fontes de informação institucionais: proposta de um método automático de verificação de erros. *Em Questão*, 18(3):263–279.
- Scott, J. (2009). *Social network analysis: a handbook*. SAGE, 2 edition.
- Ulrik, B. and Erlebach, T. (2005). *Network Analysis: Methodological Foundations*. Springer-Verlag.
- Wainer, J. and Vieira, P. (2013). Correlations between bibliometrics and peer evaluation for all disciplines: the evaluation of Brazilian scientists. *Scientometrics*, 96(2):395–410.

Co-authorship prediction in academic social network

William Takahiro Maruyama¹, Luciano Antonio Digiampietri¹

¹Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP)
Av. Arlindo Bétio, Ermelino Matarazzo – 03828-000 – São Paulo – SP – Brasil

Abstract. *The prediction of relationships in a social network is a complex and extremely useful task to enhance or maximize collaborations by indicating the most promising partnerships. In academic social networks, prediction of relationships is typically used to try to identify potential partners in the development of a project and/or co-authors for publishing papers. This paper presents an approach to predict coauthorships combining artificial intelligence techniques with the state-of-the-art metrics for link predicting in social networks.*

1. Introduction

The growth of the Web has enabled the creation of many tools for users' interaction, such as online social networks (Facebook, Twitter, LinkedIn, etc.). They are increasingly present in the day lives, and are an abstraction of a (real) social network and can directly or indirectly reflect the interactions of people in the real world [Viswanath et al. 2009]. This happens because these networks provide information exchange and communication between users virtually. Due to its popularity and consequently the enormous amount of data produced, the analysis of social networks has attracted considerable attention from the scientific community, in order to better understanding the behavior of human interactions. In a social network, people or groups of people are represented by the vertices (nodes), and the relationships, interactions or links between them, the edges of a graph [Wasserman and Faust 1994, Newman 2001, Hasan and Zaki 2011, Newman 2010]. These links between people in a social network may be intangible and can have many meanings as friendships, family, professional, interaction, among others.

One type of social network is the scientific collaboration network or academic network. In these networks, the vertices represent researchers, and the edges scientific collaboration. Therefore, two researchers are connected if they are coauthors in one or more publications. In this context, you can explicitly build an academic social network from information about publications [Newman 2010, Newman 2001, Barabási et al. 2002]. According to Newman (2010), the scientific collaboration network is genuinely a network of people, because it has a significant amount of people and the collaborations are a direct reflection of the documentation of the authors collaborations.

In the social network analysis field, the data mining techniques that are focused on links is called Link Mining [Getoor and Diehl 2005, Maruyama and Digiampietri 2016]. One of its tasks is the link prediction. This task aims to predict the formation of links between network entities. The prediction of relationships within a social network is a task that has gained attention in recent years, because it may helps from finding friends who were not connected in an online social network [Vasuki et al. 2010, Tian et al. 2010, Perez et al. 2012, Fire et al. 2011, Zhong et al. 2013, Quercia and Capra 2009], to enhance the performance of work in companies or in universities [Hsieh et al. 2013,

Dong et al. 2012, de Sa and Prudencio 2011]. In the academic social network context, the link prediction is mainly used for predicting coauthorships, whose purpose is to indicate a pair of researchers that potentially will collaborate in the publication of papers [Guo and Guo 2010, Makrehchi 2011, Dong et al. 2011, Gao et al. 2012, Lin et al. 2012, Digiampietri et al. 2013, Digiampietri et al. 2015].

Scientific collaboration can improve the production of scientific works [Pavlov 2007], aggregating different expertise and perspectives, and with a possible division of efforts. However, finding suitable researchers for each team is not a quick and easy task. Thus, link prediction can enhance team building, by recommending researchers to work together. One way of performing the link prediction is addressing it as a classification problem, i.e., using a supervised learning strategy. Therefore, finding a set of characteristics is a very important step [Hasan et al. 2006].

This paper presents a link prediction system which combines domain specific attributes with network structural ones, in order to perform a binary classification. The classification will indicate if a pair of researchers will or not become collaborators. Two types of information are considered: the ones extracted from the application domain; and the ones from the network topology.

The rest of this paper is organized as follows. Section 2 summarizes the related work. Section 3 presents the methodology. Section 4 contains the results, and, finally, Section 5 presents the conclusions.

2. Related Work

A supervised learning strategy to predict links was introduced by Liben-Nowell and Kleinberg [Liben-Nowell and Kleinberg 2003]. The authors derive a set of similarity measures from the graph topology. Five sections of the Physics arXiv e-Print were used to build five co-authorship networks. They defined a three-year time window. The period from 1994 to 1996 was used for training and 1997 to 1999 for testing. They compared their results with the prediction using a random method. The results obtained with *Katz* and its variants showed good performance in most of the data sets. Moreover, according to the authors, *Common Neighbors* and *Adamic-Adar* attributes did not achieve bad results.

Bartal et al. [Bartal et al. 2009] combined social network analysis with text mining to predict co-authorships in academic journals using thirteen attributes (one related to text mining and the others related to social network analysis) and achieved a hit rate of 91%, using data from DBLP¹.

Hasan and Zaki [Hasan et al. 2006] conducted an empirical research on link prediction in a co-authorship network, using data from BIO-BASE² and DBLP. For BIOBASE a time window from 1998 to 2002 was used, where the first 4 years were used as training and the last one as test. For DBLP, the years from 1990 to 2004 were used, the first 11 years used as training and the last 4 for testing. They extracted nine attributes considering aggregating functions, proximity and topology. The SVM classifier with RBF kernel obtained the best results, with a square error of 0.0945 and 0.1760

¹Digital Bibliography & Library Project: <http://dblp.uni-trier.de/xml/>

²Bibliographic Database, Elsevier BIOBASE - Current Awareness in Biological Sciences (CABS)

in BIOBASE and DBLP, respectively. According to the authors, the best attributes for linking prediction were: shortest path in BIOBASE and keywords in common in DBLP.

In order to predict new links and treating this problem as a time series, Soares et al. [da Silva Soares and Bastos Cavalcante Prudencio 2012] used two subsets from arXiv: theoretical high energy physics (data from 1991 to 2010) and high energy physics (1993-2010). The basic idea is to construct time series for each pair of nodes not connected by using a similarity score calculated using a topological metric. A predictive model is used to predict the next value in the series. This value will be used for the link prediction methods, tested using both supervised and unsupervised approaches. According to the authors, the supervised approach was better in all predictions models.

Lu et al. [Lu et al. 2010] developed a different approach, but also used arXiv data (from high energy physics, 1992-2003), with data from CiteSeer (1993-2003) and from the Society of Industrial and Applied Mathematics Publications (1999-2004). They proposed an approach to predict coauthoring relationships, citations and references. They used a supervised approach, multiple data sources and historical network observations. According to the authors, the experimental results confirm the accuracy of the approach.

Sun et al. [Sun et al. 2012] developed an algorithm that besides trying to predict links, it also tries to identify when these new links will occur. In another study, Sun et al. [Sun et al. 2011] focused on the prediction of links in bibliographic networks using different structural metrics. In tests conducted using DBLP data, the maximum hit rate was about 75%.

In their review, Hasan et al. [Hasan et al. 2006] identified three different models for supervised link prediction: binary classification, probabilistic model and linear algebraic model. The binary classification uses a set of features that can be extracted from the graph topology and from properties of the vertices and edges. The topological attributes can be subdivided in: node based and neighborhood based. Lu and Zhou [Lü and Zhou 2010] performed a review, in which link prediction methods were classified in three groups: similarity based, maximum likelihood based, and probabilistic models. The first group uses network structural information (topological), the features calculated can be subdivided into local and global.

3. Methodology

The methodology was composed of four activities: data gathering; features extraction/calculation; classification; and analysis of the results. Figure 1 outlines a schematic representation of activities performed.

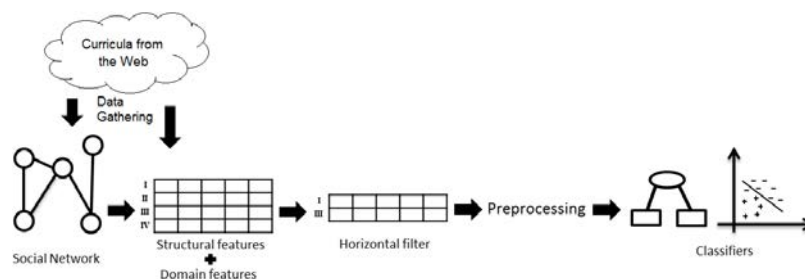


Figure 1. Schematic representation of the work process

3.1. Data gathering

All data used is public available on the Lattes Platform. This is a system created and maintained by CNPq (The Brazilian National Council for Scientific and Technological Development), which records the science curricula of the researchers, mostly Brazilians. It is possible to access information about research projects, publications, academic formation, areas of interest, and advising.

The sample selected correspond to all the 657 researchers that were professors in Brazilian Computer Science Graduate Programs during the period 2004-2009. We collected data in a time window from 1971 to 2015. For the training step, data from 1971 to 2000 was considered *past*; from 2001 to 2005 *present*; and from 2006 to 2010 was considered *future* (i.e., the coauthorships the system should predict). For the test step, data from 1976 to 2005 was considered *past*; from 2006 to 2010 *present* and the system tried to predict the coauthorships that occurred from 2011 to 2015.

The researchers curricula were downloaded from the Internet in XML format directly from the Lattes Platform. From these XML files a relational database was created using the methodology presented in [Digiampietri et al. 2012a, Digiampietri et al. 2012b], among the activities performed, we highlight the publication disambiguation (in order to detect collaborative publications), and researchers' names disambiguation (in order to identify the advisor-advise relationships). An academic social network was built based on the coauthorship relationships.

3.2. Features extraction/calculation

From the curricula information 32 attributes/features were extracted or calculated (Table 1). The first one is the class (which indicates if the pair of researchers will or not become coauthors); there are 16 domain specific attributes, and 15 structural (topological) attributes. The structural attributes were selected from the related literature (see Section 2). These attributes were calculated for all the pairs of researchers in the sample.

3.3. Classification

All the features extracted (or calculated) from the data were used as input data for the classifiers. In this work, several classifiers from Weka project³ were used.

Two types of tests were performed: with and without the balancing of the training data. The balancing strategy used was the oversampling of the minority class.

The solution of two different link prediction problems were tested. The new link prediction (i.e., the prediction of links that do not exist in the *present*), and the general link prediction problem (the prediction of link independently of their existence in the *present*).

The combinations of the 657 researchers in pairs produces 215.496 pairs, most of them will not become coauthors in the future. Before the execution of the classifiers, an horizontal filter was performed to exclude the pairs with low potential to become coauthors. The criteria used was to exclude all the pairs, which the ten first domain specific features had values equal to zero. This process excluded more than 200.000 pairs (see Section 4), which were pre-classified as "will not become coauthors". Only one pair was

³<http://www.cs.waikato.ac.nz/ml/weka/>

Table 1. Attributes used in the proposed system

Feature	Description
class	This attribute assumes the value <i>true</i> if the pair of researchers will become coauthors, and <i>false</i> otherwise.
past journal papers	Number of jointly journal papers published by the pair of researchers in the <i>past</i> .
past conference papers	Number of jointly conference papers published by the pair of researchers in the <i>past</i> .
present journal papers	Number of jointly journal papers published by the pair of researchers in the <i>present</i> .
present conference papers	Number of jointly conference papers published by the pair of researchers in the <i>present</i> .
past advising	This attribute assumes the value 1 if the one of the researchers in the pair was the advisor of the other in the <i>past</i> .
present advising	This attribute assumes the value 1 if the one of the researchers in the pair was the advisor of the other in the <i>present</i> .
on going advising	This attribute assumes the value 1 if the one of the researchers in the pair was the advisor of the other in an ongoing advising.
common advisors	Number of common advisors that the pair of researchers shared.
common advisees	Number of common advisees that the pair of researchers shared.
common graduate program	This attribute indicates if the pair of researchers work in the same graduate program.
journal papers1	Number of journal papers published in the present by the first researcher in the pair.
conference papers 1	Number of conference papers published in the present by the first researcher in the pair.
journal papers 2	Number of journal papers published in the present by the second researcher in the pair.
conference papers 2	Number of conference papers published in the present by the second researcher in the pair.
geographical distance	Geographical distance between the professional address of the two researchers in the pair.
common subareas	Number of common interesting research areas.
past and present CN	Number of common neighbors in the social academic network built using data from <i>past</i> and <i>present</i> .
present CN	Number of common neighbors in the social academic network built using data only from the <i>present</i> .
SAL	Salton Index - measures the co-occurrence of two elements divided by the square root of the multiplication of the occurrence of each element. On social networks can be used to measure the relationship between the number of neighbors that two people have in common divided by the square root of the number of neighbors multiplication of each.
JAC	Jaccard's coefficient - measures the similarity between two sets by dividing the number of elements of the intersection of the sets by the number of union members (e.g. number of common neighbors divided by the union of the neighbors of two people).
AA	Adamic-Adar - index that assigns weight in the relationship of two people favoring the relationships between people who have few relationships.
RA	Resource Allocation - index that assigns weight in the relationship of two people favoring the relationships between people who have few relationships.
SOR	Sorensen Index - index calculated as being twice the intersection between two sets divided by the sum of each set of elements (e.g., number of common neighbors divided by the the number of people in the union of the neighbors).
HPI	Hub Promoted Index - index calculated by dividing the number of intersection of two sets of elements by the minimum number of elements of the sets (e.g., number of neighbors in common of two persons divided by the minimum number of neighbors of these persons).
HDI	Depressed hub Index - index calculated by dividing the number of intersection of two sets of elements divided by the maximum number of elements of these two sets (e.g., number of common neighbors of two people divided by the maximum number of neighbors of these people).
LHM	Leicht-Newman-Holme Index - index calculated by dividing the number of elements of intersection of two sets by the multiplication of the number of elements of each set (e.g., number of common neighbors divided by multiplication of the number of neighbors of each person).
PA	Preferential Attachment - index calculated multiplying the number of elements from two sets (for example, multiplying the number of neighbors).
KATZ0.05 KATZ0.005 KATZ0.0005	Katz is an index calculated iteratively to estimate the influence of a pair of nodes in a network considering the existing paths between nodes. For this calculation, it is necessary to define a constant <i>Beta</i> . The values used were: 0.05; 0.005; and 0.0005.
SP	Shortest Path - graph shortest path between the two researchers.

incorrectly classified using this approach. The results presented in Section 4 will only deal with the remaining of the pairs.

3.4. Analysis of the results

The results were analyzed considering the metrics: accuracy, recall and area under curve (AUC). These results will be presented in the next section.

4. Results

This section presents the results for the link prediction general problem (Subsection 4.1) and prediction of new links (Subsection 4.2). In each subsection, we first present the results considering the unbalanced training set and, after these results, the ones using a balanced training set.

4.1. Link prediction general problem

The dataset (after the horizontal filtering) for the experiments presented in this section is composed of 26,979 pairs. From these, 11,833 pairs belong to the training set (10,955 negative instances and 878 positive ones) and 15,146 pairs belong to the test set (14,425 negative instances and 721 positive ones). It is worth to mentioning that a classifier which classify all the instance in the test set as negative would have an accuracy of 95.24%. This value will be considered the baseline for the accuracy.

Table 2 presents the top 3 accuracy results. The best result was achieved by the *AttributeSelectedClassifier* (96.091%), it corresponds to an improvement lower then 1% when compared with the baseline, but for the link prediction problem it is a relevant improvement. Moreover, it was verified that no individual feature was able to achieve a better improvement. The *AttributeSelectedClassifier* was able to correctly classify 41.3% of the positive instances (recall).

Table 2. Top 3 accuracy results

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
Attribute Selected Classifier	F	0.778	0.98	0.971	0.988	0.587	96.091
	T	0.778	0.502	0.638	0.413	0.012	
	Avg	0.778	0.957	0.955	0.961	0.559	
BFTree	F	0.865	0.98	0.965	0.995	0.73	96.065
	T	0.865	0.396	0.736	0.27	0.005	
	Avg	0.865	0.952	0.954	0.961	0.695	
ADTree	F	0.782	0.98	0.964	0.996	0.742	96.045
	T	0.782	0.383	0.744	0.258	0.004	
	Avg	0.782	0.951	0.954	0.96	0.707	

Table 3 presents the top 3 recall results. The *VFI* classifier was able to correctly classify 88.2% of the positive instances, but achieved a general accuracy of only 54.998%. In other words, in order to achieve high values of recall, this classifier incorrectly classified almost half of the test instances.

Table 3. Top 3 recall results

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
VFI	F	0.829	0.693	0.989	0.533	0.118	54.998
	T	0.829	0.157	0.086	0.882	0.467	
	Avg	0.829	0.668	0.946	0.55	0.134	
BayesNet	F	0.881	0.921	0.986	0.864	0.24	85.924
	T	0.881	0.34	0.219	0.76	0.136	
	Avg	0.881	0.894	0.95	0.859	0.235	
Naive Bayes Updateable	F	0.874	0.943	0.985	0.905	0.28	89.601
	T	0.87	0.397	0.274	0.72	0.095	
	Avg	0.873	0.917	0.951	0.896	0.271	

Table 4 presents the top 3 AUC results. This metric is typically used because presents a good tradeoff between the accuracy and recall. *DMNBtext* achieved the best result (0.886), with a accuracy of 95.444% and a recall of 45.5%.

Table 4. Top 3 AUC results

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
DMNBtext	F	0.886	0.976	0.973	0.979	0.545	95.444
	T	0.886	0.487	0.525	0.455	0.021	
	Avg	0.886	0.953	0.952	0.954	0.52	
Bayes Net	F	0.881	0.921	0.986	0.864	0.24	85.924
	T	0.881	0.34	0.219	0.76	0.136	
	Avg	0.881	0.894	0.95	0.859	0.235	
Logit Boost	F	0.88	0.979	0.971	0.986	0.589	95.88
	T	0.88	0.487	0.598	0.411	0.014	
	Avg	0.88	0.955	0.953	0.959	0.562	

For the next results presented in this section, the training set was balanced using the *oversampling* technique.

Table 5 presents the top 3 accuracy results. The best result (95.629%), achieved by the *RandomCommittee* classifier is worse than the one achieved without the training set balancing (96.091%), what was expected for this type of problem. But, even the recall was also worse (39.5%).

Table 5. Top 3 accuracy results - balanced training set

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
Random Committee	F	0.822	0.977	0.97	0.984	0.605	95.629
	T	0.822	0.463	0.558	0.395	0.016	
	Avg	0.822	0.953	0.951	0.956	0.577	
Rotation Forest	F	0.829	0.975	0.973	0.977	0.534	95.286
	T	0.829	0.485	0.505	0.466	0.023	
	Avg	0.829	0.952	0.951	0.953	0.51	
ZeroR	F	0.5	0.976	0.952	1	1	95.24
	T	0.5	0	0	0	0	
	Avg	0.5	0.929	0.907	0.952	0.952	

Table 6 presents the top 3 recall results. The two best results for this metric corresponds to classifiers which classified all the instances as positive (achieving an accuracy of 4.76%). The *ClassificationViaClustering* was able to recall 98.1% of the positive instances, but with an accuracy of 9.97%.

Table 6. Top 3 recall results - balanced training set

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
StackingC	F	0.5	0	0	0	0	4.76
	T	0.5	0.091	0.048	1	1	
	Avg	0.5	0.004	0.002	0.048	0.048	
DMNBtext	F	0.744	0	0	0	0	4.76
	T	0.744	0.091	0.048	1	1	
	Avg	0.744	0.004	0.002	0.048	0.048	
Classification Via Clustering	F	0.518	0.105	0.983	0.056	0.019	9.97
	T	0.518	0.094	0.049	0.981	0.944	
	Avg	0.518	0.105	0.938	0.1	0.063	

Table 7 presents the top 3 AUC results. The *ADTree* classifier was able to achieve an AUC of 0.88, with an accuracy of 88.34% and a recall of 74.1% of the positive instances. This result was one of the best tradeoff achieved between accuracy and recall (an useful result for the users that needs a high recall value).

4.2. Prediction of new links

The dataset for the experiments presented in this subsection is composed of 25,088 pairs, and corresponds to the same dataset used in the previous subsection, excluding the pairs

Table 7. Top 3 AUC results - balanced training set

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
ADTree	F	0.88	0.936	0.986	0.891	0.259	88.34
	T	0.88	0.377	0.253	0.741	0.109	
	Avg	0.88	0.909	0.951	0.883	0.252	
Naive Bayes Simple	F	0.876	0.941	0.985	0.901	0.279	89.245
	T	0.871	0.39	0.267	0.721	0.099	
	Avg	0.876	0.915	0.951	0.892	0.27	
Threshold Selector	F	0.874	0.927	0.986	0.875	0.251	86.927
	T	0.874	0.353	0.231	0.749	0.125	
	Avg	0.874	0.9	0.95	0.869	0.245	

that are coauthors in the *present*. From these, 10,976 pairs belong to the training set (10,537 negative instances and 439 positive ones) and 14,112 pairs belong to the test set (13,838 negative instances and 274 positive ones). A classifier which classifies all the instance in the test set as negative would have an accuracy of 98.05839%. This value will be considered the baseline for the accuracy.

Table 8 presents the top 3 accuracy results. *BayesianLogisticRegression* presented the best result of accuracy (98.065%), which was slightly higher than the baseline. But, it presented poor recall (0.4%). The other accuracy results presented in this table correspond to classifiers that classified all the instances as negative.

Table 8. New links - top 3 accuracy results

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
Bayesian Logistic Regression	F	0.502	0.99	0.981	1	0.996	98.065
	T	0.502	0.007	1	0.004	0	
	Avg	0.502	0.971	0.981	0.981	0.977	
NBTree	F	0.622	0.99	0.981	1	1	98.058
	T	0.622	0	0	0	0	
	Avg	0.622	0.971	0.962	0.981	0.981	
Decision Stump	F	0.572	0.99	0.981	1	1	98.058
	T	0.572	0	0	0	0	
	Avg	0.572	0.971	0.962	0.981	0.981	

Table 9 presents the top 3 recall results. The *VFI* classifier was again the best classifier regarding recall. It was able to correctly classify 85.8% of the positive instances, but achieved a general accuracy of only 37.791%. The *ClassificationViaClustering* achieved the second best recall (46.7%) with an accuracy of 81.094%.

Table 9. New links - top 3 recall results

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
VFI	F	0.71	0.537	0.992	0.368	0.142	37.791
	T	0.71	0.051	0.026	0.858	0.632	
	Avg	0.71	0.528	0.974	0.378	0.152	
Classification Via Clustering	F	0.642	0.895	0.987	0.818	0.533	81.094
	T	0.642	0.088	0.048	0.467	0.182	
	Avg	0.642	0.879	0.969	0.811	0.526	
Bayes Net	F	0.734	0.914	0.987	0.85	0.58	84.212
	T	0.734	0.094	0.053	0.42	0.15	
	Avg	0.734	0.898	0.969	0.842	0.572	

Table 10 presents the top 3 AUC results. The best result was achieved by *ClassificationViaRegression* (0.752), but with poor recall (0.007). We can highlight the results from the DMNBtext, which presented approximately same AUC, but with accuracy and recall of the positive class slightly higher than the first classifier.

Table 10. New links - top 3 AUC results

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
Classification Via Regression	F	0.752	0.99	0.981	1	0.993	98.037
	T	0.752	0.014	0.286	0.007	0	
	Avg	0.752	0.971	0.967	0.98	0.973	
DMNBtext	F	0.751	0.99	0.981	1	0.985	98.058
	T	0.751	0.028	0.5	0.015	0	
	Avg	0.751	0.972	0.972	0.981	0.966	
Naive Bayes Updateable	F	0.742	0.949	0.986	0.914	0.675	90.271
	T	0.742	0.115	0.07	0.325	0.086	
	Avg	0.742	0.932	0.968	0.903	0.664	

For the next results presented in this section, the training set was balanced using the *oversampling* technique. Table 11 presents the top 3 accuracy results. As occurred with the unbalanced, the best results are the ones that classified all instances as negative.

Table 11. New links - top 3 accuracy results - balanced training set

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
ZeroR	F	0.5	0.99	0.981	1	1	98.058
	T	0.5	0	0	0	0	
	Avg	0.5	0.971	0.962	0.981	0.981	
Vote	F	0.5	0.99	0.981	1	1	98.058
	T	0.5	0	0	0	0	
	Avg	0.5	0.971	0.962	0.981	0.981	
Stacking	F	0.5	0.99	0.981	1	1	98.058
	T	0.5	0	0	0	0	
	Avg	0.5	0.971	0.962	0.981	0.981	

Table 12 presents the top 3 recall results. The two best results for this metric corresponds to classifiers which classified all the instances as positive (achieving an accuracy of 1.942%). *ConjunctiveRule* achieved a recall of 87.2% with an accuracy of 38.91%.

Table 12. New links - top 3 recall results - balanced training set

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
StackingC	F	0.5	0	0	0	0	1.942
	T	0.5	0.038	0.019	1	1	
	Avg	0.5	0.001	0	0.019	0.019	
DMNBtext	F	0.449	0	0	0	0	1.942
	T	0.449	0.038	0.019	1	1	
	Avg	0.449	0.001	0	0.019	0.019	
Conjunctive Rule	F	0.626	0.549	0.993	0.38	0.128	38.91
	T	0.626	0.053	0.027	0.872	0.62	
	Avg	0.626	0.54	0.975	0.389	0.137	

Table 13 presents the top 3 AUC results. *NaiveBayesUpdateable* and *NaiveBayes* achieved the same results: an AUC of 0.742, with an accuracy of 87.252%, and a recall of 39.8%. *ThresholdSelector* achieved an AUC of 0.738, with a little lower accuracy (85.636%) and a higher recall (46.0%).

5. Conclusions

This paper presented a study about the prediction of links in academic social networks. The problem of link prediction was treated as an classification problem, and 32 attributes/features (domain specific and structural) were extracted from the curricula data to be used by different classifiers. Among these attributes are the ones that achieved the best results in the related literature.

Table 13. New links - top 3 AUC results - balanced training set

Classifier		AUC	F-Measure	Precision	Recall	FP rate	%Accuracy
Naive Bayes Updateable	F	0.742	0.931	0.987	0.882	0.602	87.252
	T	0.742	0.108	0.063	0.398	0.118	
	Avg	0.742	0.915	0.969	0.873	0.593	
Naive Bayes	F	0.742	0.931	0.987	0.882	0.602	87.252
	T	0.742	0.108	0.063	0.398	0.118	
	Avg	0.742	0.915	0.969	0.873	0.593	
Threshold Selector	F	0.738	0.922	0.988	0.864	0.54	85.636
	T	0.738	0.111	0.063	0.46	0.136	
	Avg	0.738	0.906	0.97	0.856	0.532	

In order to deal with the link prediction problem in academic social networks it is necessary to analyze the tradeoffs between the recall of the positive class and the general accuracy. Tests considering different metrics were performed and the oversampling balancing strategy was used. Most of the tested classifiers was not able to produce results more accurate than the ones produced by a simple classification strategy which considers all the instances as negative ones. Some promising results were achieved in the general link prediction problem. But no classifier was able to achieve a satisfactory accuracy for the prediction of new links (links between researchers that were not related).

Acknowledgments

The work presented in this paper was funded by CAPES and CNPq (processes 306046/2013-0 and 477246/2013-3).

References

- Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614.
- Bartal, A., Sasson, E., and Ravid, G. (2009). Predicting links in social networks using text mining and sna. In *Social Network Analysis and Mining, 2009. ASONAM '09. Int. Conf. on Advances in*, pages 131–136.
- da Silva Soares, P. and Bastos Cavalcante Prudencio, R. (2012). Time series based link prediction. In *The 2012 Int. Joint Conference on Neural Networks (IJCNN)*, pages 1–7.
- de Sa, H. and Prudencio, R. (2011). Supervised link prediction in weighted networks. In *The 2011 Int. Joint Conference on Neural Networks (IJCNN)*, pages 2281–2288.
- Digiampietri, L., Mena-Chalco, J., de Jesús Pérez-Alcázar, J., Tuesta, E. F., Delgado, K., and Mugnaini, R. (2012a). Minerando e caracterizando dados de currículos lattes. In *CSBC 2012 - BraSNAM*.
- Digiampietri, L., Mena-Chalco, J., Silva, G. S., Oliveira, L., Malheiro, A., and Meira, D. (2012b). Dinâmica das relações de coautoria nos programas de pós-graduação em computação no brasil. In *CSBC 2012 - BraSNAM*.
- Digiampietri, L., Santiago, C., and Alves, C. (2013). Predição de coautorias em redes sociais acadêmicas: um estudo exploratório em ciência da computação. In *CSBC 2013 - BraSNAM*.

- Digiampietri, L. A., Maruyama, W. T., Santiago, C. R. N., and da Silva Lima, J. J. (2015). Um sistema de predição de relacionamentos em redes sociais. In *XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015)*, pages 139–146.
- Dong, Y., Ke, Q., Rao, J., Wang, B., and Wu, B. (2011). Random walk based resource allocation: Predicting and recommending links in cross-operator mobile communication networks. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th Int. Conf. on*, pages 358–365.
- Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N., Rao, J., and Cao, H. (2012). Link prediction and recommendation across heterogeneous social networks. In *Data Mining (ICDM), 2012 IEEE 12th Int. Conf. on*, pages 181–190.
- Fire, M., Tenenboim, L., Lesser, O., Puzis, R., Rokach, L., and Elovici, Y. (2011). Link prediction in social networks using computationally efficient topological features. In *Privacy, security, risk and trust (passat), 2011 IEEE third Int. conference on and 2011 IEEE third Int. conference on social computing (socialcom)*, pages 73–80.
- Gao, S., Denoyer, L., and Gallinari, P. (2012). Link prediction via latent factor block-model. In *Proceedings of the 21st Int. Conf. Companion on World Wide Web, WWW '12 Companion*, pages 507–508, New York, NY, USA. ACM.
- Getoor, L. and Diehl, C. P. (2005). Link mining: A survey. *SIGKDD Explor. Newsl.*, 7(2):3–12.
- Guo, J. and Guo, H. (2010). Multi-features link prediction based on matrix. In *Computer Design and Applications (ICCD), 2010 Int. Conf. on*, volume 1, pages V1–357–V1–361.
- Hasan, M. and Zaki, M. (2011). A survey of link prediction in social networks. In Aggarwal, C. C., editor, *Social Network Data Analytics*, pages 243–275. Springer US.
- Hasan, M. A., Chaoji, V., Salem, S., and Zaki, M. (2006). Link prediction using supervised learning. In *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*.
- Hsieh, C.-J., Tiwari, M., Agarwal, D., Huang, X. L., and Shah, S. (2013). Organizational overlap on social networks and its applications. In *Proceedings of the 22Nd Int. Conf. on World Wide Web, WWW '13*, pages 571–582, Republic and Canton of Geneva, Switzerland. Int. World Wide Web Conferences Steering Committee.
- Liben-Nowell, D. and Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the Twelfth Int. Conf. on Information and Knowledge Management, CIKM '03*, pages 556–559, New York, NY, USA. ACM.
- Lin, Z., Yun, X., and Zhu, Y. (2012). Link prediction using benefitranks in weighted networks. In *Proceedings of the The 2012 IEEE/WIC/ACM Int. Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT '12*, pages 423–430, Washington, DC, USA. IEEE Computer Society.
- Lü, L. and Zhou, T. (2010). Link prediction in complex networks: A survey. *Physica A*, abs/1010.0725(6):1150–1170.

- Lu, Z., Savas, B., Tang, W., and Dhillon, I. (2010). Supervised link prediction using multiple sources. In *Data Mining (ICDM), 2010 IEEE 10th Int. Conf. on*, pages 923–928.
- Makrehchi, M. (2011). Social link recommendation by learning hidden topics. In *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11*, pages 189–196, New York, NY, USA. ACM.
- Maruyama, W. and Digiampietri, L. (2016). Predição de relacionamentos em redes sociais, uma revisão sistemática. In *CSBC2016 - BraSNAM*.
- Newman, M. (2010). *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA.
- Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404–9.
- Pavlov, M. (2007). Finding experts by link prediction in co-authorship networks. *CEUR Workshop Proceedings*, 290:42–55.
- Perez, C., Birregah, B., and Lemercier, M. (2012). The multi-layer imbrication for data leakage prevention from mobile devices. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th Int. Conf. on*, pages 813–819.
- Quercia, D. and Capra, L. (2009). Friendsensing: Recommending friends using mobile phones. In *Proceedings of the Third ACM Conference on Recommender Systems, RecSys '09*, pages 273–276, New York, NY, USA. ACM.
- Sun, Y., Barber, R., Gupta, M., Aggarwal, C. C., and Han, J. (2011). Co-author relationship prediction in heterogeneous bibliographic networks. In *IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, pages 121–128.
- Sun, Y., Han, J., Aggarwal, C. C., and Chawla, N. V. (2012). When will it happen?: relationship prediction in heterogeneous information networks. In *Proceedings of the fifth ACM Int. Conf. on Web Search and Data Mining, WSDM '12*, pages 663–672, New York, NY, USA. ACM.
- Tian, Y., He, Q., Zhao, Q., Liu, X., and Lee, W. (2010). Boosting social network connectivity with link revival. In *Proceedings of the 19th ACM Int. Conf. on Information and Knowledge Management, CIKM '10*, pages 589–598, New York, NY, USA. ACM.
- Vasuki, V., Natarajan, N., Lu, Z., and Dhillon, I. S. (2010). Affiliation recommendation using auxiliary networks. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, pages 103–110, New York, NY, USA. ACM.
- Viswanath, B., Mislove, A., Cha, M., and Gummadi, K. P. (2009). On the evolution of user interaction in facebook. *Proceedings of the 2nd ACM workshop on Online social networks - WOSN '09*, page 37.
- Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.
- Zhong, E., Fan, W., Zhu, Y., and Yang, Q. (2013). Modeling the dynamics of composite social networks. In *Proceedings of the 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '13*, pages 937–945, New York, NY, USA. ACM.

Comparing and Recommending Conferences

Gretel Monteagudo García¹, Bernardo Pereira Nunes^{1,2},
Giseli Rabello Lopes³, Marco Antonio Casanova¹, Luiz André P. Paes Leme⁴

¹ Pontifical Catholic University of Rio de Janeiro, RJ, Brazil

² Federal University of the State of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

³ Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

⁴ Federal Fluminense University, Niterói, RJ, Brazil

{ggarcia, bnunes, casanova}@inf.puc-rio.br, giseli@dcc.ufrj.br,
lapaesleme@ic.uff.br

Abstract. *This paper first introduces techniques for comparing conferences that use familiar similarity measures and a new measure based on co-authorship communities. Then, it focuses on two families of techniques for conference recommendation, the first one based on the similarity measures and the second on the idea of finding the most related authors in the co-authorship network. The experiments suggest that the best performing techniques are: the technique for comparing conferences that uses the new similarity measure based on co-authorship communities; and the conference recommendation technique that explores the most related authors in the co-authorship network.*

Resumo. *Este trabalho introduz técnicas para comparar conferências a partir de medidas de similaridade clássicas e uma nova medida baseada em comunidades de coautores. Em seguida, o trabalho foca em dois grupos de técnicas para recomendação de conferências: o primeiro se baseia nas medidas de similaridade entre conferências e o segundo na estratégia de encontrar os autores mais relacionados na rede de coautores. Os experimentos sugerem que as melhores técnicas são: a técnica de comparação de conferências que utiliza a nova medida de similaridade baseada em comunidades de coautores; e a técnica para recomendação de conferências que explora os autores mais relacionados na rede de coautores.*

1. Introduction

In this work, we propose, implement and evaluate techniques to automatically compare and recommend conferences. The techniques for comparing conferences adopt familiar similarity measures, such as the Jaccard similarity coefficient, the Pearson correlation similarity and the cosine similarity, and a new similarity measure, the *co-authorship network community similarity index*.

We proceed to define two families of conference recommendation techniques. The first family of techniques adopts collaborative filtering based on the conference similarity measures investigated in the first part of the paper. The second family includes two techniques based on the idea of finding, for a given author, the strongest re-

lated authors in the co-authorship network and recommending the conferences that his co-authors usually publish in. The first recommendation technique uses the *weighted semantic connectivity score* – *WSCS* [Nunes et al. 2013], an index for measuring relatedness of actors. However, since this index proved to be accurate, but quite costly for large co-authorship networks, we define a second recommendation technique that adopts an approximation of the *WSCS*, but which is much faster to compute and as accurate as the *WSCS*.

The paper concludes with a description of experiments to test and compare the techniques with data extracted from a triplified version of the DBLP repository, which stores Computer Science bibliographic data for more than 4,500 conferences and 1,500 journals (as of early 2016). The experiments were performed using a Web-based application that enables users to interactively analyze and compare a set of conferences. The experiments indicate that the best performing techniques are: (1) the technique for comparing conferences that uses the new co-authorship network community similarity index; and (2) the conference recommendation technique that explores the co-authorship network and adopts an approximation of the *WSCS*. These two techniques are therefore the major contributions of this paper.

This paper is structured as follows. Section 2 summarizes related work. Section 3 presents the techniques for comparing conferences. Section 4 covers the conference recommendation techniques. Section 5 briefly presents an application that implements the techniques and covers the evaluation of the techniques. Finally, Section 6 contains the conclusions and proposes future work.

2. Related Work

Henry et al. (2007) analyzed a group of the four major conferences in the field of Human-Computer Interaction (HCI). The authors discovered many global and local patterns using only article metadata, such as authors, keywords and year. Blanchard (2012) presented a ten-year analysis of the paper production in Intelligent Tutoring Systems (ITS) and Artificial Intelligence in Education (AIED) conferences and shows that Western, Educated, Industrialized, Rich, and Democratic bias observed in psychology may be influencing AIED research. Chen, Zhang and Vogeley (2009) proposed an extension of the contemporary co-citation network analysis to identify co-citation clusters of cited references. Intuitively, the authors synthesize thematic contexts in which these clusters are cited and trace how the research focus evolves over time. Gasparini, Kimura and Pimenta (2013) present a visual exploration of the field of Human Computer Interaction in Brazil from a fifteen-year analysis of paper production in the Brazilian Symposium on Human Factors in Computing Systems (IHC). Chen, Song and Zhu (2007) opened a wide range of opportunities for research agendas and trends in ER conferences.

Zervas et al. (2014) applied social network analysis (SNA) metrics to analyzing the co-authorship network of the Educational Technology & Society (ETS) Journal. Procópio, Laender and Moro (2011) did a similar analysis for the Databases field. Cheong and Corbitt (2009a, 2009b) analyzed the Pacific Asia Conference on Information Systems and the Australasian Conference on Information Systems.

Recently, Lopes et al. (2015, 2016) carried out an extensive analysis of the WEBIST conferences, involving authors, publications, conference impact, topics cover-

age, community analysis and other aspects. Linked Data principles to publish conference data were also used by [Batista and Lóscio 2013; Lopes et al. 2015, 2016].

Contrasting with the above references, in this paper we propose, implement and evaluate several techniques to compare conferences in general, and not a specific conference series. The current implementation works with the triplified version of the DBLP repository, which covers the vast majority of Computer Science conferences.

Turning to conference recommendation, Luong et al. (2012) proposed and compared three recommendation methods for conferences. To define the methods, they recursively collected the co-authors of the co-authors, until a network of 3 levels deep was created, in a set of the more important co-authors. The best performing recommendation method, which we will refer simply as the *most frequent conference*, weights the contributions of each co-author by the number of papers they have co-authored with the main author. It is defined as:

$$coauthor_CONF_i = \sum_{m=1}^N coauthors_w_{i,m} \quad (1)$$

where N is main author(s) of the test paper and $coauthors_w_{i,m}$ is the co-authors' conference weight between the main author m and her co-authors in the network and is defined as:

$$coauthors_w_{i,m} = \sum_{k=1}^{CoA} (nfreq_CONF_{i,m} + nfreq_CONF_{i,k}) * w_CoA_{k,m} \quad (2)$$

where CoA is a co-author(s) of the main author m who have published respectively at conference i , $w_CoA_{k,m}$ is the number of times a main author m co-authored papers with another member k in the network, and $nfreq_CONF_{i,m}$ is the probability of the author m to publish in conference i .

In this paper, we propose two conference recommendation techniques based on the social network analysis of the co-authorship network, but we adopt a measure of the strength of the connections between the authors in the network which is computed differently from Luong's method. We first propose to estimate the relatedness of actors in a social network by using a semantic connectivity score [Nunes et al. 2013], denoted SCS, which is in turn based on the Katz index [Katz 1953]. This score takes into account the number of paths between two nodes of the network and the accumulated weights of these paths. Then, we propose a second score that approximates the SCS score and that uses the shortest path between two nodes. In addition to these two strategies, we also propose to construct a utility matrix and to implement recommendation techniques based on collaborative filtering using a utility matrix.

3. Comparing Conferences

In what follows, we use the following notation:

- C is a set of conferences
- A is a set of authors
- P is a set of papers

- $p: A \rightarrow 2^P$ is a function that assigns to each author $i \in A$ the set of papers $p(i) \subseteq P$ that author i published (in any conference)
- $pc: A \times C \rightarrow 2^P$ is a function that assigns to each author $i \in A$ and each conference $x \in C$ the set of papers $pc(i, x) \subseteq P$ that author i published in conference x
- A_x and A_y are the set of authors that published in conferences x and y , that is, $A_x = \{i \in A / |pc(i, x)| > 0\}$ and, likewise, $A_y = \{i \in A / |pc(i, y)| > 0\}$.
- $A_{x,y}$ is the set of authors that published in both conferences x and y , that is, $A_{x,y} = \{i \in A / |pc(i, x)| > 0 \wedge |pc(i, y)| > 0\}$
- $G_x = (N_x, E_x)$, the *co-authorship network* of conference x , is an undirected and un-weighted graph where $i \in N_x$ indicates that author i published in conference x and $\{i, j\} \in E_x$ represents that authors i and j co-authored one or more papers published in conference x

In what follows, we adapt familiar similarity measures to conferences or authors and introduce a new measure called *communities similarity*.

The *Jaccard similarity coefficient* for conferences x and y is defined as

$$jaccard_sim(x, y) = (A_x \cap A_y) / (A_x \cup A_y) \quad (3)$$

The utility matrix expresses the preferences of an author for a conference to publish his research. More formally, the utility matrix $[r_{x,i}]$ is such that the lines represent conferences and the columns represent authors and is defined as:

$$r_{x,i} = \frac{|pc(i, x)|}{|p(i)|} \quad (4)$$

Based on the utility matrix $[r_{x,i}]$, we define the Pearson's correlation coefficient similarity between conferences x and y as follows:

$$pearson_sim(x, y) = \frac{\sum_{i \in A_{x,y}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in A_{x,y}} (r_{x,i} - \bar{r}_x)^2 \sum_{i \in A_{x,y}} (r_{y,i} - \bar{r}_y)^2}} \quad (5)$$

where \bar{r}_x is the average of the elements of line x of the utility matrix (and likewise for \bar{r}_y).

Again based on the utility matrix $[r_{x,i}]$, we define the cosine similarity between conferences x and y as follows:

$$cos_sim(x, y) = \frac{\sum_{i \in A_{x,y}} (r_{x,i} r_{y,i})}{\sqrt{\sum_{i \in A_{x,y}} (r_{x,i})^2 \sum_{i \in A_{x,y}} (r_{y,i})^2}} \quad (6)$$

We introduce a new similarity measure between conferences based on communities defined over the co-authorship network of the conferences.

Given the co-authorship network $G_x = (N_x, E_x)$ of conference x , we define an *author community* c_x of x as the net of nodes of a connected component of G_x . Let c_x and c_y be author communities in the co-authorship networks of conferences x and y ,

respectively. We say that c_x and c_y are *equivalent* w.r.t. a similarity measure sim and a threshold level α iff $sim(c_x, c_y) \geq \alpha$. For example, sim may be defined using Jaccard similarity coefficient between pairs of conferences introduced above.

Let C_x and C_y be the sets of communities of conferences x and y , respectively. Let $EQ[sim, \alpha](x, y)$ be the set of communities in the co-authorship network of conference x that have an equivalent community in the co-authorship network of conference y (and symmetrically $EQ[sim, \alpha](y, x)$).

The *co-authorship network communities similarity* (based on a similarity measure sim and a threshold level α) between conferences x and y is then defined as:

$$c_sim[sim, \alpha](x, y) = \frac{|EQ[sim, \alpha](x, y)|}{\min\{|C_x|, |C_y|\}} \quad (7)$$

Note that $C_x > 0$ and $C_y > 0$ since G_x and G_y must have at least one node each and therefore at least one connected component each.

4. Recommending Conferences

4.1. Conference Recommendation Techniques based on Classical Similarity Measures

As defined in [Leskovec, Rajaraman and Ullman 2014], in a recommendation system, there are two classes of entities – *users* and *items*. Users have preferences for certain items, which must be extracted from the data. The data itself is represented as a utility matrix giving, for each user-item pair, a value that represents what is known about the degree of preference or *rating* of that user for that item. An unknown rating implies that there is no explicit information about the user’s preference for the item. The goal of a recommendation system is to predict the unknown ratings in the utility matrix.

In our context, we recall from Section 3 that the utility matrix $[r_{x,i}]$ is such that $r_{x,i}$ expresses the preference (i.e., rating) of an author i for a conference x to publish his research. To predict an unknown rating, we compute the similarity between conferences and detect their nearest neighbors or most similar conferences. With this information, the rating of conference x for author i is defined as follows:

$$CF(x, i) = \frac{\sum_{y \in S_x} (r_{y,i}) sim(x, y)}{\sum_{y \in S_x} sim(x, y)} \quad (8)$$

where S_x is the set of conferences most similar to x and $r_{y,i}$ is the rating of conference y for author i .

Therefore, we may immediately define a family of conference recommendation techniques based on the utility matrix and the classical similarity measures introduced in Section 3, that we call *CF-Jaccard*, *CF-Pearson*, *CF-Cosine* and *CF-Communities*, according to the similarity measure adopted. Section 5 discusses how they perform in detail.

4.2. Conference Recommendation Techniques based on the Weighted Authorship Network

Recall from Section 3 that $p: A \rightarrow 2^P$ is the function that assigns to each author $i \in A$ the set of papers $p(i) \subseteq P$ that author i published (in any conference). The *weighted co-authorship network* based on p is the edge-weighted undirected graph $G = (N, E, w)$, where $i \in N$ represents an author, $\{i, j\} \in E$ indicates that i and j are co-authors, that is, $\{i, j\} \in E$ iff $p(i) \cap p(j) \neq \emptyset$, and $w(\{i, j\})$ assigns a weight to the co-authorship relationship between i and j and is defined as:

$$w(\{i, j\}) = \frac{|p(i) \cap p(j)|}{|p(i) \cup p(j)|} \quad (9)$$

Hence, the larger $w(\{i, j\})$ is, the stronger the co-authorship relationship will be: if authors i and j co-authored all papers they published, then $w(\{i, j\}) = 1$; and if they have not co-authored any papers, then the edge $\{i, j\}$ does not exist.

The second family of conference recommendation techniques explores the weighted co-authorship network and adopts two scores: the *weighted semantic connectivity score* – $WSCS$ and the *modified weighted semantic connectivity score* – $MWSCS$. Hence, these techniques are called *WSCS-based* and *MWSCS-based recommendation techniques*.

The *weighted semantic connectivity score*, $WSCS_e$, is defined by modifying the semantic connectivity score SCS_e to take into account the weight of the paths between two authors i and j , computed as the sum of the weights of the edges in the path:

$$WSCS_e(i, j) = \sum_{w=1}^T \beta^w \cdot |paths_{\langle i, j \rangle}^{\leq w}| \quad (10)$$

where $|paths_{\langle i, j \rangle}^{\leq w}|$ is the number of paths of weight equal to w between i and j and T is the maximum weight of the paths and $0 < \beta \leq 1$ is a positive damping factor.

The conference recommendation technique based on $WSCS_e$ works as follows. Given an author i , it starts by computing $WSCS_e(i, j)$, the score between i and any other author j in the weighted co-authorship network. Then, it sorts authors in decreasing order of $WSCS_e$, since authors that are better related to author i will have a higher $WSCS_e(i, j)$ value. For better performance, the technique considers only the first n authors in the list ordered by $WSCS_e$. Call this set F_i . For each author j in F_i , the technique selects the conference with the highest rank, denoted $MaxC_j$. The rank of conference x for author i is defined as follows:

$$rank(x, i) = \sum_{j \in F_i \text{ and } MaxC_j = x} WSCS_e(i, j) \quad (11)$$

Since computing the $WSCS_e$ score can be very slow for large graphs, we propose to compute only the shortest paths from author i to other authors using Dijkstra's algorithm. We then redefine the score as follows:

$$MWSCS_e(i, j) = \beta^w \quad (12)$$

where w is a length of the shortest path from author i to author j . The recommendation technique remains basically the same, except that it uses the $MWSCS$ score.

5. Evaluation and Results

5.1. Application Architecture

Figure 1 summarizes the architecture of the application developed to run the experiments. The *Conferences Data Service* handles queries to the triple store with conference data. The *Co-authorship Network Service* receives data from the *Conferences Data Service* and handles queries to the *Neo4j* database. When an analysis is executed, the system stores the results for future reuse; the *Previous Calculation Service* manages these functions. All experiments that follow were executed in an Intel Core Quad 3.00GHz, with 6 GB RAM, running Windows 7.

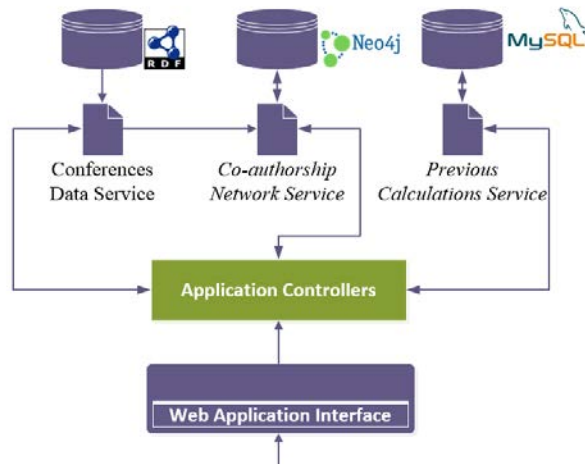


Figure 1. Web Application Architecture.

5.2. Experiments with the Conference Similarity Techniques

We evaluated the conference similarity techniques assuming that the most similar conferences should fall into the same category. We selected as benchmark the List of Computer Science Conferences defined in Wikipedia¹, which contains 248 academic computer science conferences, classified into 13 categories. That is, the categories define a partition \mathbf{P} of the set of conferences. Then, we applied the same clustering algorithm to this set of conferences, but using each of the conference similarity measures. Finally, we compared the clusters thus obtained with \mathbf{P} . The best conference similarity measure would therefore be that which results in conference clusters that best match \mathbf{P} .

We adopted the hierarchical agglomerative clustering algorithm, which treats each conference as a singleton cluster at the outset and then successively merges (or agglomerates) pairs of clusters, using similarity measures, until achieving the desired number of clusters. To determine how similar clusters are, and agglomerate them, a linkage criterion was used. The shortest value of these links that remains at each step causes the fusion of the two clusters whose elements are involved.

Let $d(a, b)$ denote the distance between two elements a and b . Familiar linkage criteria between two sets of elements A and B are:

¹ https://en.wikipedia.org/wiki/List_of_computer_science_conferences

- Complete-linkage: the distance $D(A, B)$ between two clusters A and B equals the distance between the two elements (one in each cluster) that are farthest away from each other:

$$D(A, B) = \max \{d(a, b) / a \in A, b \in B\} \quad (13)$$

- Single-linkage clustering: the distance $D(A, B)$ between two clusters A and B equals the distance between the two elements (one in each cluster) that are closest to each other:

$$D(A, B) = \min \{d(a, b) / a \in A, b \in B\} \quad (14)$$

- Average linkage clustering: the distance $D(A, B)$ between two clusters A and B is taken as the average of the distances between all pairs of objects:

$$D(A, B) = \frac{\sum_{a \in A} \sum_{b \in B} d(a, b)}{|A||B|} \quad (15)$$

Before explaining the measures used to compare how well different data clustering algorithms perform on a set of data, we need the following definitions. Given a set of n elements S and two partitions X and Y of S , where X is the correct partition and Y is the computed partition, we define:

- *TP* (True Positive) is the number of pairs of elements in S that are in the same set in X and in the same set in Y
- *TN* (True Negative) is the number of pairs of elements in S that are in different sets in X and in different sets in Y
- *FN* (False Negative) is the number of pairs of elements in S that are in the same set in X and in different sets in Y
- *FP* (False Positive) is the number of pairs of elements in S that are in different sets in X and in the same set in Y

The measures to evaluate the performance of the clustering algorithms using the proposed similarity functions are:

- *Rand Index*: measures the percentage of correct decisions made by the algorithm:

$$RI = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

- *F-measure*: balances the contribution of false negatives by weighting the recall through a parameter $\beta > 0$:

$$F = \frac{(\beta^2 + 1)P \cdot R}{(\beta^2 P) + R} \quad (17)$$

where $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$

Figure 2 shows the Rand index obtained by executing the hierarchical agglomerative clustering algorithm with different linkages criteria, using the Jaccard, Pearson, cosine and communities similarity measures. Note that, in general, the algorithm based on communities similarity had the best performance, followed by the Jaccard similarity. In this case, the cosine similarity had the worst behavior.

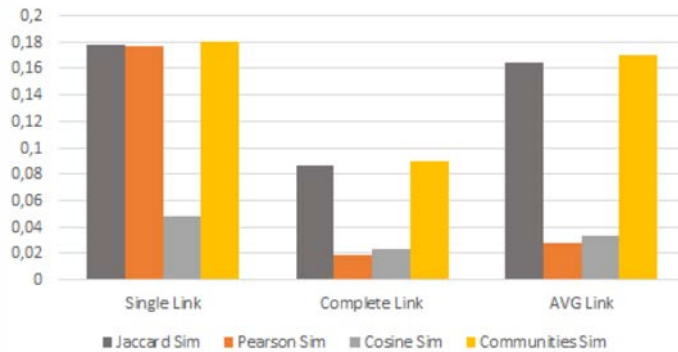


Figure 2. Rand Index of the clustering algorithms.

Figure 3 shows the F-measure obtained by executing the same algorithms. Analyzing the results presented in Figure 3, we observe that the best performances were also obtained using the communities similarity and the Jaccard similarity measures. The worst performance was obtained using the Pearson similarity measure. The algorithm using the cosine similarity measure achieved the worst performance only with the single link linkage criterion.

Therefore, these experiments suggest that the best performing algorithm is that which adopts the communities similarity measure.

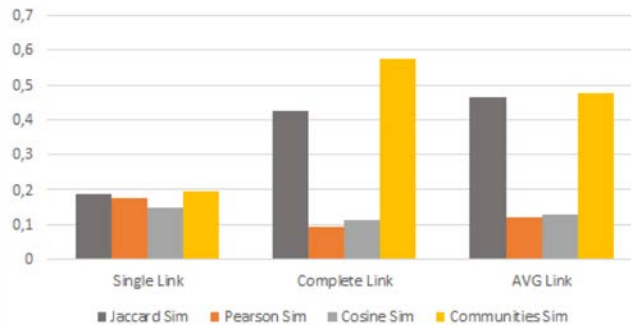


Figure 3. F-measure with $\beta=1$ of clustering algorithms.

5.3. Experiments with the Conference Recommendation Techniques

Recall that we proposed two families of recommendation techniques. One family is based on classical similarity measures – Jaccard, Pearson, and cosine similarity – and a new similarity measure, the communities similarity. These techniques are respectively called CF-Jaccard, CF-Pearson, CF-Cosine and CF-Communities. The second family includes two techniques based on the weighted and the modified weighted semantic connectivity, called WSCS-based and MWSCS-based recommendation techniques.

We evaluated the conference recommendation techniques using the same dataset as in Section 5.2, with the 248 academic computer science conferences, and selected 243 random authors to predict their conferences ranking, for that we delete all publica-

tions of the author on the conferences that we want to rank. We adopted Luong’s most frequent conference technique as the benchmark (see Section 2).

Also recall that the *mean average precision* measures how good a recommendation ranking function is. Intuitively, let a be an author and \mathbf{C}_a be a ranked list of conferences recommended for a . Let \mathcal{S}_a be a *gold standard* for a , that is, the set of conferences considered to be the best ones to recommend for a . Then, we have:

- $Prec@k(\mathbf{C}_a)$, the *precision at position k* of \mathbf{C}_a , is the number of conferences in \mathcal{S}_a that occur in \mathbf{C}_a until position k , divided by k
- $AveP(\mathbf{C}_a)$, the *average precision of \mathbf{C}_a* , is defined as the sum of $Prec@k(\mathbf{C}_a)$ for each position k in the ranking \mathbf{C}_a in which a relevant conference for a occurs, divided by the cardinality of \mathcal{S}_a :

$$AveP(\mathbf{C}_a) = \frac{\sum_k Prec@k(\mathbf{C}_a)}{|\mathcal{S}_a|} \quad (18)$$

- MAP , the *Mean Average Precision* of a rank score function over all the authors used in these experiments (represented by set \mathbf{A}) is then defined as follows:

$$MAP = average \{AveP(\mathbf{C}_a) / a \in \mathbf{A}\} \quad (19)$$

Consider first the two conference recommendation techniques based on the co-authorship network, the WSCS-based and MWSCS-based recommendation techniques. To compare them, we performed experiments that measured their runtime, accuracy and average precision of the Top-10 conferences of an author (thus, in this situation the maximum $|\mathcal{S}_a|$ value used in $AveP$ calculation is 10). Figure 4 shows the runtime results of the algorithms that implement these recommendation techniques. Note that the MWSCS-based algorithm is far more efficient than the WSCS-based algorithm.

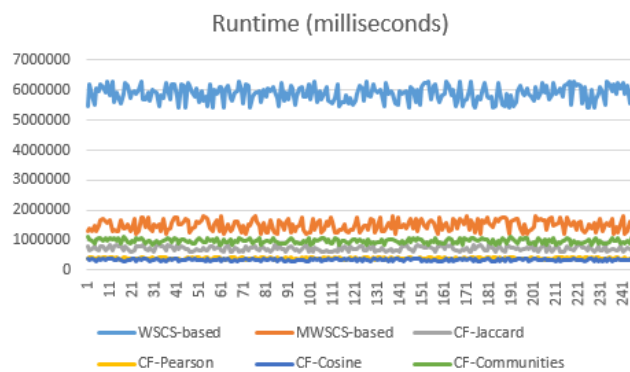


Figure 4. Runtime of the recommendation algorithms.

Table 1 shows the accuracy and MAP of the seven conference recommendation techniques. The two proposed techniques (first two rows of Table 1) have very similar accuracy. In fact, of the 243 authors that we tested, the balance of the correct predictions was 201 against 197. Based on these results, we may conclude that the MWSCS-based technique is much more efficient and maintains acceptable accuracy level and MAP , when compared with the WSCS-based technique.

Table 1 also indicates that the WSCS-based and the MWSCS-based techniques have better accuracy and *MAP* than the benchmark. The CF-Jaccard and the CF-Communities techniques have very acceptable results and very close to the benchmark, but less than the WSCS-based and the MWSCS-based techniques. The CF-Pearson and CF-Cosine techniques have poor accuracy.

Table 1. Comparison of the Accuracy and *MAP* of the recommendation techniques.

Technique	Accuracy	MAP
WSCS-based	82.72%	80.93%
MWSCS-based	81.07%	80.01%
CF-Jaccard	78.19%	77.73%
CF-Pearson	55.56%	50.21%
CF-Cosine	56.79%	51.89%
CF-Communities	79.02%	77.93%
Benchmark	79.84%	77.88%

6. Conclusions

In this work, we presented techniques to compare and recommend conferences. The techniques to compare conferences are based on some classical similarity measures and on a new similarity measure based on the co-authorship network communities of two conferences. The experiments suggest that the best performance is obtained using the new communities similarity measure.

We introduced two families of conference recommendation techniques, following the collaborative filtering strategy. The first family is based on similarity measures proposed to compare conferences. The second family is based on the relatedness of two authors in the co-authorship network, using the weighted and the modified weighted semantic connectivity score. The experiments suggest that the techniques of the second family perform better than the benchmark and better than the techniques based on similarity measures. Furthermore, the technique based on the new modified weighted semantic connectivity score is much faster than the technique based on the original weighted semantic connectivity score.

As for future work, we plan to make the tool and the test datasets openly available and to expand the scope of the work to other publications datasets.

Acknowledgments

This work was partly funded by CNPq under grant 444976/2014-0, 303332/2013-1, 442338/2014-7 and 248743/2013-9 and by FAPERJ under grant E-26/201.337/2014.

References

- Batista, M.G.R., Lóscio, B.F. 2013. “OpenSBBD: Usando Linked Data Para Publicação de Dados Abertos Sobre O SBBD” Proc. 28th Brazilian Symposium on Databases, Recife, Brazil, Sept. 30 – Oct. 03, 2013.

- Blanchard, E.G. 2012. “On the WEIRD Nature of ITS/AIED Conferences” Proc. 11th Int’l. Conf. on Intelligent Tutoring Systems, Chania, Greece, June 14-18, pp. 280–285.
- Chen, C., Song, Il-Y., Zhu, W. 2007. “Trends in Conceptual Modeling: Citation Analysis of the ER Conference Papers (1979-2005)” Proc. 11th Int’l. Conf. of the International Society for Scientometrics and Informetrics, Madrid, Spain, June 25-27, pp. 189–200.
- Chen, C., Zhang, J., Vogeley, M.S. 2009. “Visual Analysis of Scientific Discoveries and Knowledge Diffusion” Proc. 12th Int’l. Conf. on Scientometrics and Informetrics - ISSI 2009, Rio de Janeiro, Brazil, July 14-17, 2009.
- Cheong, F., Corbitt, B.J. 2009a. “A Social Network Analysis of the Co-Authorship Network of the Australasian Conference of Information Systems from 1990 to 2006” Proc. 17th European Conf. on Info. Systems, Verona, Italy, June 8-10, 2009.
- Cheong, F., Corbitt, B.J. 2009b. “A Social Network Analysis of the Co-Authorship Network of the Pacific Asia Conference on Information Systems from 1993 to 2008” Proc. Pacific Asia Conference on Information Systems 2009, Paper 23.
- Gasparini, I., Kimura, M.H., Pimenta, M.S. 2013. “Visualizando 15 Anos de IHC” Proc. 12th Brazilian Symposium on Human Factors in Computing Systems, Manaus, Brazil, pp. 238–47.
- Henry, N., Goodell, H., Elmqvist, N., Fekete, J-D. 2007. “20 Years of Four HCI Conferences: A Visual Exploration” *Int’l. J. of Human-Comp. Inter.* 23(3): 239–85.
- Katz, L. 1953. “A New Status Index Derived from Sociometric Analysis” *Psychometrika* 18(1): 39–43.
- Leskovec, J., Rajaraman, A., Ullman, J.D. 2014. *Mining of Massive Datasets*. Cambridge University Press.
- Lopes, G.R. et al. 2015. “Knowing the Past to Plan for the Future - An In-Depth Analysis of the First 10 Editions of the WEBIST Conference” Proc. 11th Int’l. Conf. on Web Information Systems and Technologies, Lisbon, Portugal, May 20-22, 2015, pp. 431–42.
- Lopes, G.R. et al. 2016. “A comprehensive analysis of the first ten editions of the WEBIST conference” *Lect. Notes in Business Information Processing* 246:252–274.
- Luong, H. et al. 2012. “Publication Venue Recommendation Using Author Network’s Publication History” Proc. 4th Asian Conf. on Intelligent Information and Database Systems - ACIIDS 2012, Kaohsiung, Taiwan, March 19-21, 2012, pp. 426–35.
- Nunes, B.P. et al. 2013. “Interlinking Documents Based on Semantic Graphs” *Procedia Computer Science* 22: 231–40.
- Procópio, P. S., Laender, A.H.F. and Moro, M.M. 2011. “Análise da Rede de Coautoria do Simpósio Brasileiro de Bancos de Dados” Proc. 26th Brazilian Symposium on Databases, Florianópolis, Brazil, Oct. 3-6, 2011.
- Zervas, P. et al. 2014. “Studying Research Collaboration Patterns via Co-Authorship Analysis in the Field of TeL: The Case of Educational Technology & Society Journal” *Journal of Educational Technology & Society* 17(4): 1–16.

Desenvolvimento de um Modelo Semântico para Recomendação Baseado em Grafos

Hugo Guércio Fernandes¹, Victor Ströele¹

¹Departamento de Ciência da Computação
Universidade Federal de Juiz de Fora
Juiz de Fora – MG – Brazil

{hugo.guercio,victor.stroele}@ice.ufjf.br

Abstract. *Recommender Systems aims to help a user or groups of users in identifying the most relevant items based on their needs. The items may have different characteristics and can be services, products or miscellaneous information. With the growth of data amount, recommender systems are being increasingly studied because it is increasingly difficult to find the desired information due to the available alternatives. This paper proposes and develop a model for a recommender system based on bipartite graphs, formed by semantic information of users and items. This recommendation model was evaluated based on two Proof of Concepts (PoC).*

Resumo. *Os sistemas de recomendação tem como objetivo auxiliar um usuário, ou grupos de usuários, na identificação de diferentes itens relevantes às suas necessidades. Com a crescente quantidade de dados, os sistemas de recomendação estão sendo muito estudados, pois é cada vez mais difícil encontrar a informação desejada devido a grande quantidade de alternativas disponíveis. Neste trabalho é proposto e desenvolvido um modelo para um sistema de recomendações baseado em grafos bipartidos, composto por informações semânticas de usuários e itens. Este modelo de recomendação foi avaliado com base em duas provas de conceitos (PoC).*

1. Introdução

O desenvolvimento e uso de tecnologias, que dão apoio à interação social, cresce rapidamente. Com isso novas formas de comunicação, trabalho e entretenimento são usadas por bilhões de pessoas gerando redes complexas que representam essas interações entre elas.

Com o desenvolvimento das tecnologias e aumento do volume de dados é necessário auxiliar os usuários a encontrarem o que, de fato, é do interesse deles. Assim, os sistemas de recomendação tentam identificar os recursos que sejam aderentes aos interesses dos seus usuários em diferentes cenários, como: análise de mercado, recomendação de recursos educacionais e de apoio à pesquisa, recomendação de produtos, controle de produção ou retenção de consumidores [Han and Kamber 2006].

As recomendações, na maioria das vezes, são feitas a partir do conhecimento extraído de dados históricos dos usuários e itens. Casos bem conhecidos de recomendações podem ser exemplificados pela sugestão de itens na Amazon.com como analisado em [Oestreicher-Singer and Sundararajan 2010], e em filmes na Netflix.com.

O objetivo deste trabalho é propor e desenvolver um modelo de recomendação de itens a pesquisadores baseado em grafos bipartidos. Esses grafos são compostos por pesquisadores, vídeos e informações semânticas extraídas através da análise dos pesquisadores e dos vídeos.

Como resultado do modelo proposto foram geradas recomendações de vídeos para alguns pesquisadores com o intuito de avaliar a proposta. Para isso, foram extraídas informações de uma rede social científica [Ströele et al. 2013], composta pelos relacionamentos de coautoria entre os pesquisadores presentes na DBLP (Bibliografia de Ciência da Computação).

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta a metodologia adotada neste trabalho; a Seção 3 tem como objetivo fundamentar os conceitos utilizados; a Seção 4 apresenta alguns trabalhos relacionados; na Seção 5 é apresentado o modelo proposto para o problema; já na Seção 6 são apresentados os resultados encontrados seguidos pela Seção 7 que apresenta uma avaliação do modelo proposto. Por fim, são apresentadas as considerações finais na Seção 8.

2. Metodologia

A metodologia que este trabalho utilizou possui: (1) pesquisa bibliográfica para identificar problemas e abordagens da área de recomendação, identificação dos tipos de recursos a serem recomendados e identificação de trabalhos relacionados; (2) proposta e desenvolvimento de um modelo para recomendação através de informações semânticas; e (3) avaliação do modelo desenvolvido através de duas provas de conceitos.

A partir da pesquisa bibliográfica foi possível identificar os aspectos mais relevantes das soluções relacionadas e utilizar o resultado dessa pesquisa para fundamentar a construção do modelo. Para validação do resultado foi utilizada uma rede de coautoria extraída do repositório DBLP¹. Assim, os usuários são pesquisadores de uma rede social científica e os itens recomendados foram vídeos extraídos do Youtube. Para avaliação do modelo desenvolvido foram realizadas duas provas de conceitos com alguns pesquisadores dessa rede social.

3. Recomendação de conteúdo com base em redes sociais

Os Sistemas de Recomendação representam uma área de pesquisa importante devido a abundância de aplicações práticas, que ajudam os usuários a lidar com a grande quantidade de dados presentes em diversas plataformas. De acordo com [Adomavicius and Tuzhilin 2005] os sistemas de recomendação vem sendo, em geral, separados em três categorias: baseada em conteúdo, colaborativa e híbrida. Neste trabalho foi utilizada uma abordagem baseada em conteúdo.

A abordagem de filtragem baseada em conteúdo tem suas raízes nos processos de recuperação da informação [Baeza-Yates et al. 1999]. Neste tipo de filtragem é realizada uma descrição dos usuários e itens com base em suas características. Após essa caracterização é realizada uma comparação entre as descrições, verificando as relações entre elas. Nos casos onde usuários e itens possuem características semelhantes, infere-se que o item é relevante ao usuário [Balabanović and Shoham 1997].

¹<http://dblp.uni-trier.de/>

O grande desafio dos sistemas de recomendação é identificar corretamente as necessidades dos usuários. Por isso, as redes sociais têm sido cada vez mais utilizadas no âmbito dos sistemas de recomendação, pois, através delas, é possível extrair diversos tipos de informação dos usuários e definir o perfil dos mesmos.

O assunto de rede sociais pode ser compreendido como o estudo das relações entre indivíduos e as implicações dessas relações [Wasserman 1994]. Os indivíduos são conhecidos como atores, esses atores podem representar organizações, conjuntos de pessoas ou um único indivíduo e estão conectados entre si através de laços relacionais.

As redes sociais podem ser definidas como um conjunto finito de atores, ou grupos de atores, e as relações definidas entre eles. Diferentes tipos de redes sociais podem ser observadas com finalidades diversas. O uso de redes sociais, em conjunto com os sistemas de recomendação, pode acrescentar informações na forma de relacionamentos entre os usuários. Em geral o relacionamento entre os usuários pode ser implícito ou explícito.

Neste trabalho foi utilizada uma rede social científica. Este tipo de rede social refere-se a redes nas quais os atores são representações de pesquisadores e as ligações são estabelecidas por relacionamentos de co-autoria entre os autores [Goldenberg et al. 2010, Ströele et al. 2013]. Essas redes podem ser estruturadas e representadas através de Grafos. Os Grafos são modelos versáteis que possibilitam a análise de diferentes problemas onde os pontos e suas conexões possuem alguma interpretação, seja ela física ou conceitual. Neste trabalho, os grafos bipartidos foram utilizados para modelar a rede social semântica utilizada como base para realizar recomendações.

4. Trabalhos Relacionados

Como verificado em [Shi et al. 2014] as abordagens baseadas em grafos vem sendo bem estudadas e muito utilizadas, principalmente na predição de links em redes sociais. Diversos trabalhos atacam o problema de recomendações a partir de uma abordagem que utiliza grafos bipartidos utilizado para diferentes propósitos. Em [Kunegis et al. 2010] o objetivo é especializar algoritmos de recomendação para o caso de grafos bipartidos. De acordo com os estudos, dentre os métodos de predição avaliados, os modelos de recomendação algébricos, que levam todo o grafo em consideração, podem ser aplicados.

Outros trabalhos utilizam a ideia de arestas com peso em grafos bipartidos, em [Sawant 2013] as arestas possuem peso referente à qualificação. Neste trabalho uma rede de alocação dos recursos foi utilizada para explicitar a similaridade entre pares de usuários e pares de objetos, que são utilizados para predições de recomendações e qualificações. Eles concluíram que as predições de recomendação são afetadas de maneira considerável pela acurácia da função de similaridade entre as entidades. Em [Ting et al. 2013] o conceito de peso nas arestas para demonstrar o interesse entre usuários e objetos também é utilizado. O objetivo é aumentar a viabilidade do uso de algoritmos para recomendação em grafos bipartidos.

A abordagem de redes semânticas também pode ser vista em [Maehara et al. 2012] onde o objetivo é recomendar itens de exibição em museus. A rede semântica é construída utilizando apenas as características dos itens, possibilitando os usuários a entender o contexto que engloba o item, despertando o interesse dos mesmos. Em [Shieh et al. 2008], uma abordagem colaborativa é escolhida para

recomendação de imagens a partir de um grafo semântico. O grafo semântico é construído a partir da redução de grafos bipartidos entre autores e termos. Nesta abordagem, o grafo bipartido é reduzido a um grafo unidimensional onde as arestas possuem peso e os nós representam os autores.

Diferentemente de [Kunegis et al. 2010, Shi et al. 2014] a abordagem deste trabalho utiliza grafos bipartidos com pesos em suas arestas para auxiliar o processo de recomendação. Em [Sawant 2013, Ting et al. 2013] também é adotado o peso nas arestas, mas esse peso não é definido com base nas informações semânticas dos autores e dos objetos.

Neste trabalho as arestas possuem um peso que representa o interesse que um autor possui pelo objeto com qual ele está conectado. Diferentemente de [Maehara et al. 2012], esse interesse é definido pela análise semântica tanto dos autores quanto dos objetos. Ao final tem-se um grafo semântico bipartido utilizado para recomendação de objetos aos autores, o que torna essa abordagem diferente de [Shieh et al. 2008].

5. Modelo de Recomendação baseado em Grafo Bipartido

Nesta seção serão apresentadas as etapas que compõem o modelo para resolução do problema de recomendação com filtragem baseada em conteúdo.

Na Figura 1 é apresentada uma visão geral do *workflow* que representa a solução proposta neste trabalho. No primeiro passo foi realizada a coleta dos dados nas fontes de informação selecionadas. Após a extração e tratamento dos dados, eles são transmitidos ao analisador de conteúdo que irá fazer a análise semântica dos mesmos.

Depois da análise do conteúdo ter sido realizada as informações semânticas são extraídas para que o grafo semântico seja construído pelo quantificador de relacionamentos. Esse módulo é responsável por dar pesos aos relacionamentos entre os autores e itens a partir das informações semânticas identificadas anteriormente. Com base no grafo obtido pela relação entre autores e itens é possível realizar as recomendações. De posse do grafo semântico o recomendador irá ordenar os itens, segundo os interesses do autor, e apresentar a lista de itens recomendados a ele. As próximas seções apresentam os detalhes do desenvolvimento de cada uma dessas etapas.

5.1. Fontes de Informação

As fontes de informação são repositórios que contém dados para extração de conteúdos. Esses repositórios devem possuir dados que contemplam as necessidades de informação de um usuário ou grupos de usuários. Neste trabalho serão usadas como fontes de informações a DBLP e o YouTube.

5.1.1. Base de Pesquisadores

A DBLP possui uma vasta coleção de dados com diversos tipos de informação relativas às publicações da área de ciência da computação, como artigos, *journals*, livros e outros. Os dados estão disponíveis no formato XML e foram mapeados para um banco de dados relacional. Esse mapeamento foi feito com auxílio da biblioteca *JAVA SAXParser* permitindo a análise das diferentes *tags* encontradas para realizar a identificação da estrutura do documento.

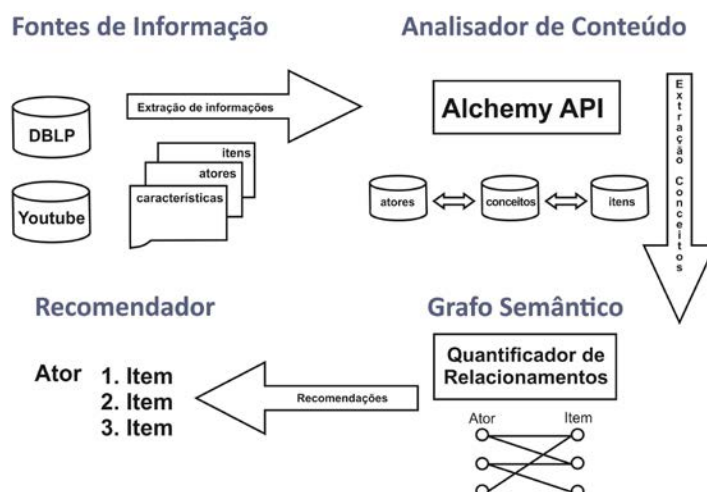


Figura 1. Workflow desenvolvido

Em uma etapa posterior ao processamento das informações do arquivo XML foi realizada a carga em um banco de dados *PostgreSQL*. Após a consolidação dos dados foi possível selecionar as informações dos autores e dos artigos publicados por eles para o módulo analisador de conteúdo.

5.1.2. Base de Vídeos

Conforme dito anteriormente, os objetos a serem recomendados pelo sistema de recomendação são os vídeos. Para isso, é necessário definir um repositório com esses objetos e, neste trabalho, será utilizado o YouTube. Esse repositório foi escolhido devido à grande quantidade e diversidade dos vídeos na plataforma. A variedade dos vídeos possibilita a recomendação para diferentes usuários em diferentes contextos.

O conteúdo do YouTube pode ser coletado a partir de consultas no site ou utilizando a *YouTube Data API*². Neste trabalho a coleta dos dados foi realizada, exclusivamente, através da API citada anteriormente. A linguagem JAVA foi utilizada para realizar as chamadas a API para capturar dados de vídeos, *playlists* e canais.

Com a API pronta para receber as requisições foi possível iniciar o processo de extração dos vídeos, as respostas da API são transmitidas no formato JSON e convertidas para classes JAVA. Para extração dos dados foi realizada uma chamada no método *search* retornando 50 resultados. Este método teve como parâmetro os conceitos identificados nos autores. A partir do retorno obtido foi possível realizar a coleta dos dados pertinentes de cada vídeo. Os dados coletados foram transcritos para o banco de dados *PostgreSQL*.

5.2. Análise do Conteúdo

Para a análise do conteúdo e extração de informações semânticas foi utilizada a *AlchemyAPI*³. A API realiza uma análise de textos, estruturando-os e tornando-os legíveis aos computadores.

²<https://developers.google.com/youtube/v3/>

³<http://www.alchemyapi.com/>

Essa API foi aplicada para a extração de informações semânticas relacionadas aos vídeos e pesquisadores. Ela realiza análises semânticas de texto, permitindo a extração de conjuntos de conceitos relevantes ao texto analisado. Os conceitos identificados são as ideias discutidas no texto, independente se elas são mencionadas explicitamente ou não. Somente os textos da língua inglesa foram avaliados pela API.

Além dos conceitos de cada vídeo e autor, um coeficiente de relevância é gerado por essa API indicando a importância de cada conceito para um dado vídeo ou autor. De acordo com estudos realizados por [Saif et al. 2012] foi observado que os resultados da extração de conceitos apresentados pela plataforma são bons quando comparados a outras soluções, fundamentando o uso neste trabalho.

Para os vídeos, a extração de conceitos ocorreu a partir da análise do título e descrição. A especificidade de um conceito para um dado vídeo também foi determinada pela *AlchemyAPI*. Desta forma, foram armazenados os pares compostos por conceitos e vídeos, assim como sua relevância.

A identificação dos conceitos associados aos autores ocorreu através da análise de todos os títulos dos trabalhos publicados por esses autores. Assim como foi feito para os vídeos, também foram armazenados os pares de conceitos e autores com suas respectivas relevâncias.

Após a identificação dos conceitos foram construídos dois grafos bipartidos. Ambos possuem dois tipos de nós, no primeiro estão representados os autores e os conceitos extraídos, enquanto no segundo grafo são representados os vídeos e os seus conceitos. A aresta dos grafos bipartidos gerados, neste momento, tem como peso o valor da relevância entre as suas extremidades, essa medida foi extraída pelo extrator de conteúdo e representa a importância do conceito aos dados analisados pela API.

5.3. Construção do Grafo Semântico

O grafo semântico foi construído a partir da união dos dois grafos bipartidos gerados a partir do módulo de *Análise do Conteúdo* (descrito anteriormente), conforme representado na Figura 2. Os nós do grafo semântico representam os autores (pesquisadores) e itens (vídeos) e as arestas que os relacionam possuem um peso que representa a relevância dos vídeos para os pesquisadores. O peso dessas relações foi calculado com a Equação 1.

$$R = \frac{\sum_{i \in C} (CP_i + CV_i)}{1 + \sum_{j \in CV} CV_j} \quad (1)$$

Onde CP e CV são os conjuntos das relevâncias dos conceitos extraídos dos pesquisadores e dos vídeos, respectivamente, e $C = \{CP \cap CV\}$ é o conjunto das relevâncias dos conceitos comuns aos pesquisadores e aos vídeos. Assim, R representa o quão relevante um vídeo é para determinado pesquisador.

Nesta equação, no numerador são somadas as relevâncias referentes aos conceitos presentes nos dois nós (pesquisadores e vídeos). O denominador é a soma de todos os conceitos que apenas o vídeo possui, este denominador tem como objetivo penalizar a relevância do relacionamento para os conceitos que não são comuns aos dois nós. Como as

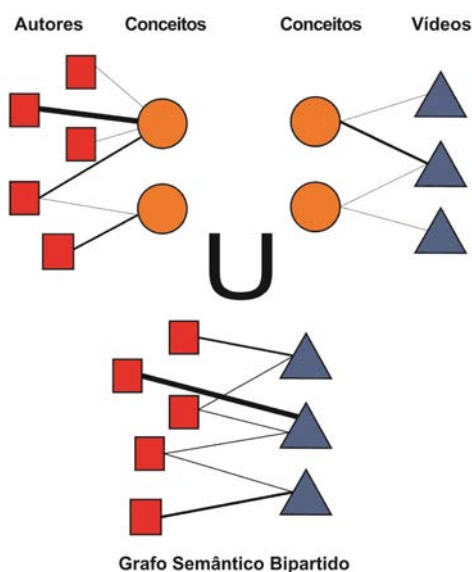


Figura 2. Representação do grafo semântico.

relevâncias fornecidas pela API variam entre 0 e 1, onde os valores próximos a 1 indicam maior importância, o intervalo da equação é $[0, 2n]$ onde n é a quantidade de conceitos comuns. Desta forma, os valores fornecidos pela equação que estão mais próximos a 0 indicam uma menor aderência entre os nós e valores mais distantes de 0 indicam maior aderência entre os nós.

Somente os conceitos presentes nos dois grafos gerados foram considerados. Desta forma somente os conceitos que aparecem nos dois grafos podem ser utilizados para ajustar a relevância dos relacionamentos.

5.4. Recomendação

Foi realizada uma análise semântica dos dados extraídos com a finalidade de representar o perfil dos autores e itens, que, neste trabalho, são pesquisadores e vídeos, respectivamente. A Equação 1 indica a aderência existente entre os pesquisadores e os vídeos que serão recomendados. Assim, listas de recomendação foram geradas analisando os valores das arestas do grafo semântico ordenadas de forma decrescente, visto que os maiores valores indicam a maior aderência entre um autor e um vídeo.

6. Resultados

Após o desenvolvimento e execução da solução proposta foram identificados os dois grafos gerados no processo descrito na Subseção 5.2. Foram avaliadas 89.973 instâncias com um total de 22.428 conceitos identificados gerando um total de 626.445 relacionamentos.

Para a construção do grafo semântico, nem todos os autores possuíam conceitos relacionados aos vídeos. Mesmo assim, a quantidade de autores e vídeos que compõem o grafo contemplou a maioria dos autores e itens classificados. Aproximadamente 78% dos autores e 72% dos vídeos estão presentes no grafo semântico. Esses nós se relacionaram através de 1598 conceitos.

A Tabela 1 mostra os conceitos que são compartilhados pelo maior número de autores e vídeos. Pode ser observado que a quantidade dos conceitos mais identificados

para os autores foi bem maior, isto aconteceu pois o conjunto de instâncias de autores analisados foi superior aos vídeos.

Tipo	Nome do conceito	Quantidade
Vídeo	Vector space	215
	Mathematics	202
	Group	144
Autor	Mathematics	8448
	Algorithm	8413
	Logic	4215

Tabela 1. Conceitos mais comuns.

Autor	Conceito	Relevância
22	Logic	0.778654
	Reasoning	0.719578
	Problem solving	0.689492

Tabela 2. Conceitos identificados para o autor 22.

A seguir seguem dados coletados de um autor presente no grafo semântico obtido através do *workflow* desenvolvido neste trabalho. A Tabela 2 mostra os conceitos identificados para o autor 22. Os conceitos apresentados são apenas aqueles em que existe algum vídeo com o mesmo conceito.

Na Tabela 3 podemos ver os conceitos associados a alguns vídeos que se relacionam com o autor 22. Com as informações referentes à relevância dos conceitos para os autores o grafo bipartido é construído. Da mesma forma, é construído um grafo bipartido que representa os vídeos e seus conceitos. Nos dois grafos construídos o peso das arestas representa a relevância encontrada pelo Analisador de Conteúdo.

A Tabela 4 representa a lista de recomendação para o autor 22. Essas listas de recomendação foram concebidas a partir da construção do grafo semântico, que representa a união dos dois grafos bipartidos construídos.

Vídeo	Conceito	Relevância
DIKkujAieTY	Logic	0.988
	Reasoning	0.638043
	Inductive reasoning	0.618041
	Fallacy	0.617589
	Deductive reasoning	0.608903
	Inference	0.596085
	Analogy	0.593086
	Abductive reasoning	0.55106
ervHbKa7R5g	Problem solving	0.92607
hq1bUM2tyg0	Problem solving	0.865606
kpqSeXpiM6k	Problem solving	0.895888
SpDIXJ2I2D4	Problem solving	0.954089
	Problem	0.470192
Zyq6TmQVBxk	Problem solving	0.895888

Tabela 3. Conceitos identificados para os vídeos relacionados ao autor 22.

Vídeo	Relevância
ervHbKa7R5g	1,615562
kpqSeXpiM6k	1,585380
Zyq6TmQVBxk	1,585380
hq1bUM2tyg0	1,555098
SpDIXJ2I2D4	1,117936
DIKkujAieTY	0,681447

Tabela 4. Listas de recomendação para o autor 22.

7. Avaliação

Este trabalho foi avaliado com base em duas provas de conceitos (PoC) desenvolvidas segundo o método GQM [Basili 1992]. As provas de conceito foram realizadas com

usuários reais, professores do departamento de ciência da computação. A avaliação foi feita através de um questionário onde esses professores avaliaram diferentes questões sobre a abordagem proposta neste trabalho.

7.1. Primeira PoC: definição de perfil por uma rede social científica

Nesta primeira prova de conceito é analisada a definição do perfil dos professores para avaliar a viabilidade de definição de perfil utilizando uma rede social científica. O objetivo é verificar se as informações extraídas estão corretas, levando em consideração a opinião do usuário.

Através dos conceitos extraídos para cada um dos cinco professores foi possível identificar o interesse de pesquisa deles. Dos cinco professores convidados a participar desta avaliação quatro deles desenvolvem seus trabalhos em linhas de pesquisa distintas, embora todos sejam da grande área da computação. Em média foram extraídos sete interesses de pesquisa (conceitos) de cada professor.

Após identificar esses interesses os professores foram convidados a avaliá-los respondendo à seguinte pergunta: *As informações extraídas da DBLP representam de fato seus interesses de pesquisa?*

Quando são analisadas as avaliações em (a) na Figura 3 verifica-se que quatro professores concordam que os interesses identificados pelo processo de extração são aderentes aos seus perfis de pesquisa. Nenhum professor discordou do perfil identificado tenha relação com seus interesses de pesquisa, mas houve um professor que indicou ser neutro em relação à esse questionamento.

Esta primeira prova de conceito mostra a viabilidade de extrair informações relacionadas aos interesses de pesquisa dos professores através de uma rede social científica. Têm-se indícios de que o uso das redes sociais científicas é viável no processo de definição dos interesses de pesquisa dos professores e que, tais interesses, são compatíveis com a realidade dos professores.

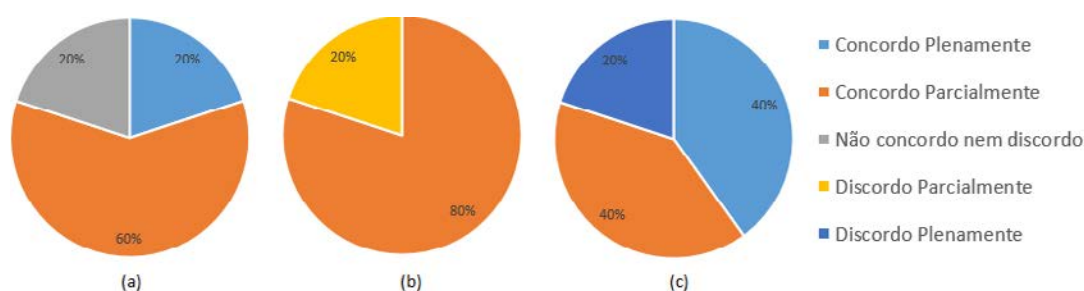


Figura 3. Gráficos de resposta das Perguntas (a), (b) e (c)

7.2. Segunda PoC: aderência das recomendações aos interesses dos professores

O objetivo da segunda prova de conceito é analisar a viabilidade da geração de recomendações e avaliar a qualidade das mesmas com base na aderência aos interesses de pesquisa dos professores considerando a opinião do usuário.

Para a realização desta prova de conceito foi solicitado ao usuário que ele avaliasse uma lista de vídeos recomendados a ele. Essa lista foi ordenada segundo a relevância

dos termos extraídos para cada autor, calculada de acordo com a Equação 1. Foram recomendados entre 7 e 10 vídeos para cada autor.

Os cinco professores foram convidados a avaliar as recomendações geradas para os conteúdos extraídos para cada um deles. Para realizar a avaliação eles responderam às seguintes perguntas: (b) O tema dos vídeos recomendados a você está coerente com as informações extraídas da DBLP? (c) Você compartilharia o vídeo recomendado para algum aluno ou professor da mesma área?

O gráfico (b) da Figura 3 apresenta as avaliações dos professores com relação a primeira pergunta. Pode-se observar que quatro, dentre os cinco professores, concordam parcialmente que as recomendações são coerentes com as informações extraídas da DBLP. Um professor discordou parcialmente que haja uma relação entre as recomendações e os conceitos extraídos.

Também foi solicitado aos professores que eles deixassem seus comentários e críticas para que o sistema de recomendação pudesse ser aprimorado. Nesses comentários foi possível compreender melhor as avaliações feitas por eles. Em geral, sempre há algum vídeo que não é condizente com os interesses de pesquisa dos professores e, por isso, nenhum deles concordou plenamente com as recomendações. O professor que discordou parcialmente recebeu recomendação de vídeos de um canal de humor que abordava, de alguma maneira, os conceitos extraídos da DBLP. Em um trecho dos seus comentários o professor diz que:

“...acredito que possam incluir algum tipo de filtro para selecionar canais relacionados a educação/pesquisa, já que o foco é identificar vídeos associados ao interesse de pesquisa do indivíduo.”

O gráfico (c) da Figura 3 consolida as respostas dos professores à segunda pergunta. O objetivo dessa pergunta é identificar se existe alguma recomendação que o professor considera relevante ao ponto de repassar o vídeo sugerido. Observa-se que quatro professores consideram os vídeos interessantes e recomendariam os mesmos para outras pessoas. Um professor, o mesmo que discordou parcialmente na questão anterior, discorda plenamente, ou seja, ele não recomendaria os vídeos a outras pessoas. Este professor considerou os vídeos superficiais aos seus temas de pesquisa e, por isso, não os recomendaria. Os resultados da primeira prova de conceito indicam que a estratégia de definição do perfil de pesquisadores a partir de uma rede social científica foi bem avaliada pelos professores. Dentre os professores que avaliaram essa estratégia apenas um não concordou com o perfil definido para ele.

Da mesma forma, os resultados da segunda prova de conceito também trouxeram indicativos de que as recomendações feitas aos professores foram, de certa forma, coerentes aos perfis identificados. Como era esperado, o professor que não concordou com a definição de seu perfil também não concordou com os vídeos recomendados.

Assim sendo, os resultados observados mostram a viabilidade das estratégias de extração de características em redes sociais científicas e de recomendação de vídeos.

7.3. Ameaças a validade

Existem algumas ameaças à validade deste estudo. O estudo simula que os professores estavam precisando dos vídeos que foram recomendados. Portanto, é importante que

esse estudo seja expandido para que os professores possam acessá-lo quando realmente estiverem necessitando de alguma recomendação.

A qualidade dos termos extraídos para identificar o perfil dos professores também é uma ameaça a validade. Há um risco de que, quando outros professores forem selecionados, termos muito genéricos ou ambíguos sejam identificados, impactando diretamente na qualidade das recomendações.

Outra ameaça é a quantidade de professores convidados a fazerem a avaliação das estratégias apresentadas neste trabalho. Embora tenham sido selecionados professores de diferentes áreas de pesquisa, sabe-se que é necessário avaliar a proposta com um maior quantitativo de pessoas.

8. Considerações Finais

O objetivo deste trabalho foi desenvolver um modelo de recomendação baseado em informações semânticas e, a partir do modelo proposto, gerar recomendações. Para isso, foi realizada uma pesquisa bibliográfica a fim de fundamentar a abordagem da solução proposta e identificar os trabalhos relacionados.

Após a modelagem, foi realizada a extração dos dados e, de posse dos dados, foram extraídos conceitos com o objetivo de descrever os itens e atores utilizados no sistema de recomendação. De posse dos atores e itens, com seus respectivos perfis, foi possível quantificar a relação entre eles e criar um grafo semântico bipartido. A partir do grafo semântico construído foi possível extrair uma lista de recomendações ordenada pela relevância entre os nós.

Com o intuito de avaliar a viabilidade da solução proposta foram realizadas duas provas de conceitos. Um questionário foi desenvolvido e cinco professores o responderam. Com base nas respostas obtidas verificou-se que o modelo de recomendação baseado em grafos semânticos é viável. Além disso, avaliando as sugestões desses professores, algumas evoluções podem ser aplicadas ao modelo para aprimorar as recomendações.

Como trabalhos futuros tem-se a modificação do modelo proposto adicionando o *feedback* fornecido pelo usuário após a recomendação para auxiliar as próximas recomendações. A partir do *feedback* pode ser utilizada uma abordagem de filtragem híbrida para fazer recomendações utilizando, além das informações semânticas, a opinião dos usuários com perfil semelhantes. Outra oportunidade para trabalhos futuros é abordar o uso de ontologias na etapa de análise do conteúdo afim de agrupar os conceitos dentro de um domínio, podendo assim modificar a qualificação entre os relacionamentos e alterar a lista de recomendação.

9. Agradecimentos

Agradecemos a CAPES, CNPQ, FAPEMIG e UFJF.

Referências

- [Adomavicius and Tuzhilin 2005] Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749.

- [Baeza-Yates et al. 1999] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- [Balabanović and Shoham 1997] Balabanović, M. and Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72.
- [Basili 1992] Basili, V. R. (1992). Software modeling and measurement: the goal/question/metric paradigm.
- [Goldenberg et al. 2010] Goldenberg, J., Libai, B., Muller, E., and Stremersch, S. (2010). Database submission-the evolving social network of marketing scholars. *Marketing Science*, 29(3):561–567.
- [Han and Kamber 2006] Han, J. and Kamber, M. (2006). *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann.
- [Kunegis et al. 2010] Kunegis, J., De Luca, E. W., and Albayrak, S. (2010). The link prediction problem in bipartite networks. In *Computational intelligence for knowledge-based systems design*, pages 380–389. Springer.
- [Maehara et al. 2012] Maehara, C., Yatsugi, K., Kim, D., and Ushiyama, T. (2012). An exhibit recommendation system based on semantic networks for museum. In *Innovations in Intelligent Machines-2*, pages 131–141. Springer.
- [Oestreicher-Singer and Sundararajan 2010] Oestreicher-Singer, G. and Sundararajan, A. (2010). Recommendation networks and the long tail of electronic commerce. *Available at SSRN 1324064*.
- [Saif et al. 2012] Saif, H., He, Y., and Alani, H. (2012). Semantic sentiment analysis of twitter. In *The Semantic Web-ISWC 2012*, pages 508–524. Springer.
- [Sawant 2013] Sawant, S. (2013). Collaborative filtering using weighted bipartite graph projection: a recommendation system for yelp. *CS224W: Social and Information Network Analysis (December 10, 2013)*.
- [Shi et al. 2014] Shi, Y., Larson, M., and Hanjalic, A. (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1):3.
- [Shieh et al. 2008] Shieh, J.-R., Yeh, Y.-T., Lin, C.-H., Lin, C.-Y., and Wu, J.-L. (2008). Collaborative knowledge semantic graph image search. In *Proceedings of the 17th international conference on World Wide Web*, pages 1055–1056. ACM.
- [Ströele et al. 2013] Ströele, V., Zimbrão, G., and Souza, J. M. (2013). Group and link analysis of multi-relational scientific social networks. *The Journal of Systems and Software*, 86:1819–1830.
- [Ting et al. 2013] Ting, Y., Yan, C., and Xiang-wei, M. (2013). Personalized recommendation system based on web log mining and weighted bipartite graph. In *Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on*, pages 587–590. IEEE.
- [Wasserman 1994] Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Effects of Social Ties in Knowledge Diffusion: case study on PLOS ONE

Felipe Eltermann^{1,2}, Alan Godoy^{1,3}, Fernando J. Von Zuben¹

¹School of Electrical and Computer Engineering – University of Campinas
Campinas, São Paulo, Brazil

²Elabora Consultoria
Campinas, São Paulo, Brazil

³CPqD Foundation
Campinas, São Paulo, Brazil

{eltermann, godoy, vonzuben}@dca.fee.unicamp.br

Abstract. *In order to capture the effects of social ties in knowledge diffusion, this paper examines the publication network that emerges from the collaboration of researchers, using citation information as means to estimate knowledge flow. For this purpose, we analyzed the papers published in the PLOS ONE journal finding strong evidence to support that the closer two authors are in the co-authorship network, the larger the probability that knowledge flow will occur between them. Moreover, we also found that when it comes to knowledge diffusion, strong co-authorship proximity is more determinant than geographic proximity.*

Keywords: knowledge diffusion, co-authorship network, multiplex network

1. Introduction

It is of primary interest of organizations and society in general to promote knowledge diffusion. This interest is often twofold: first, knowledge acquisition fosters best problem solving techniques and allows better decision making; on the other hand, a knowledge producer may want to position itself as source (or holder) of a piece of knowledge. The literature highlights two knowledge-intensive activities: academic research¹ [Lieberman and Wolf 1997, Barabási et al. 2002, Newman 2004] and industrial R&D [Jaffe et al. 1993, Singh 2005] and their interactions [Trajtenberg et al. 1992, Cohen et al. 2002, Sorenson and Singh 2006].

Based on their capability to represent flow, exchange and communication, networks are used in literature as a functional framework for the study of organizational arrangements between economic actors [Powell 1990]. Considering its production-consumption structure, the study of knowledge diffusion commonly uses network-based models as a foundation [Barabási et al. 2002, Newman 2004, Singh 2005].

¹Public research, scientific research and academic research are commonly interchangeable and in this paper, except when explicitly put otherwise, represents research conducted in universities and governmental laboratories.

Published documents (scientific papers and patents) are considered to be a recorded track that reflects the actual occurrence of knowledge flow between its producers and consumers [Jaffe et al. 1993]. Specifically, it has been considered that knowledge flow occurs when authors cite previous works of others [Jaffe et al. 1993, Singh 2005, Sorenson and Singh 2006] or when they collaborate as joint co-authors [Sidone et al. 2016] of the same document.

In this paper, we explore the methodology used by [Singh 2005] to study how collaboration ties in a patent network predict the occurrence of knowledge flow between teams, applying it to a network of scientific papers. A multiplex network is considered, in which papers are represented by nodes and two layers are present: the citation layer – in which oriented edges represent the citations between papers – and the co-authorship layer in which edges represent the co-authorship relationships between papers.

We observed that social proximity is a strong determinant for the occurrence of knowledge diffusion in scientific research.

This paper is organized as follows: Section 2 is devoted to describing the preprocessing stages necessary to achieve the ready-to-use dataset, followed by a brief explanation of how to build the multiplex network directly from the analysis of the dataset. Section 3 deals with the methodology adopted to extract knowledge from the multiplex network. The main results are outlined at Section 4, followed by a discussion in Section 5. Concluding remarks and further steps of the research are presented in Section 6.

2. Methods

The implementation of the methods described below is fully available online².

2.1. Citation Network and Paper-centric Co-authorship Network

The methodology to relate social ties with knowledge flow was strongly based on [Singh 2005], which argues that knowledge flow occurs when one team cites work authored by another. This definition provides a reasonable intuition for detecting the actual occurrence of knowledge flow that takes place within the scientific community. The original definition requires that, to consider a citation as corresponding to a flow of knowledge, the authors of the citing paper must be all different from the authors of the cited work. We also tested a relaxed version of this definition, considering that knowledge flow occurs provided that the citing team has at least one different author from the cited team. Our preliminary analysis showed that the difference considering both definitions was not significant, so we only present results based on the original definition.

While it is possible to map the occurrence of knowledge flow in a structure of teams or authors, greater granularity is achieved when allowing the concept of flow to be applied between two individual papers. Knowledge actually flows between people, but we consider that the concrete output of a particular occurrence of flow is embedded in the citation between two papers (provided the constraint mentioned in the previous paragraph).

Therefore, differently from the usual *author-centric* co-authorship network analysis [Barabási et al. 2002, Newman 2004], it is also possible to represent the co-authorship

²<https://bitbucket.org/eltermann/plosone-analysis>

relationship as follows: papers are nodes and there is an edge between any given pair of papers if, and only if, there is at least one person that is author of both papers. This *paper-centric* definition reflects the social ties in the sense that the network evolves according to the pattern of collaboration between authors.

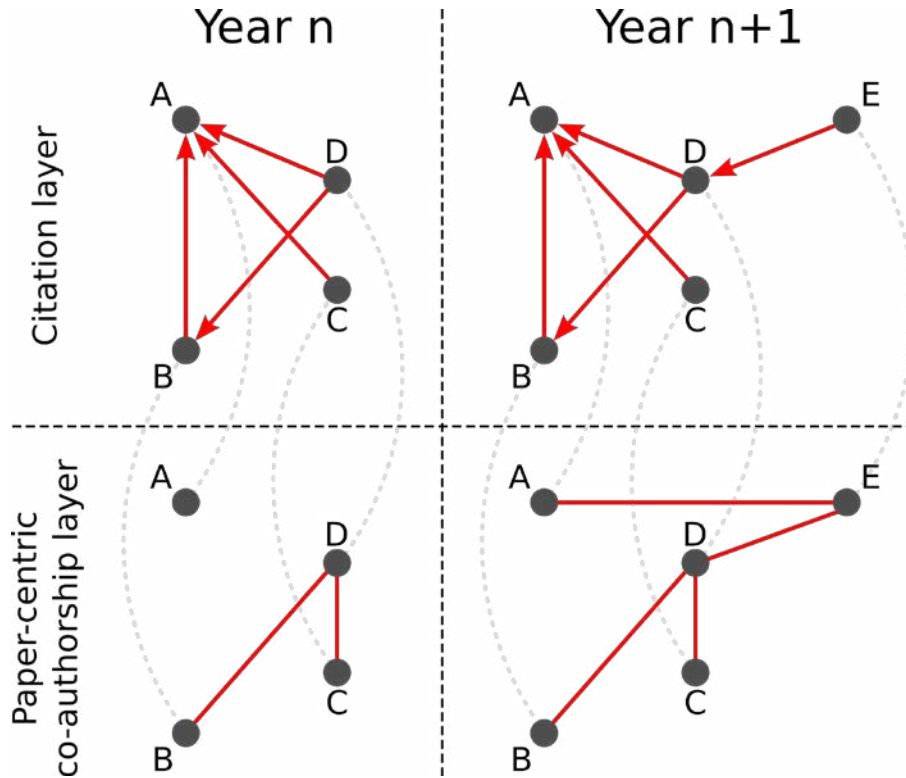


Figure 1. Let A , B , C , D and E be scientific papers. Also, let $a_A = \{1, 2\}$, $a_B = \{3\}$, $a_C = \{5\}$, $a_D = \{3, 4, 5\}$, and $a_E = \{1, 4\}$ be the sets of authors of each paper. Citations are drawn as follows: A is cited by B , C and D ; B is cited by D ; and D is cited by E . Finally, A , B , C and D are published in year n and E is published in year $n+1$. Given this setting, the top half of the figure presents the citation layer of our paper-centric network: papers are represented as nodes and there is an oriented edge from one paper to another if, and only if, the former cites the latter. The bottom half presents the co-authorship layer of the network, in which there is an edge between two papers if, and only if, they share at least one common author. The social distance between two papers is defined as the number of nodes in the shortest path between them in the co-authorship layer. Given $d_n(X, Y)$ as the social distance between papers X and Y in year n , we pick the following examples: $d_n(B, D) = 0$, $d_{n+1}(B, E) = 1$, $d_n(A, B) = \infty$, and $d_{n+1}(A, B) = 2$. Now, the occurrence of knowledge flow between two papers $y(X, Y)$ is defined as a Boolean variable that is set to 1 if, and only if, one paper cites the other and there is no common author between them. Considering this definition, we have: $y(B, D) = 0$, $y(B, E) = 0$, and $y(A, B) = 1$.

From the two-layered multiplex network represented in Figure 1, it is possible to observe that the social distance between two papers can assume an infinite value (when the papers are in different connected components in the co-authorship layer). It is also possible to observe that the distance between two papers may vary with time, directly affecting the design of the experiment: each pair of papers must be considered in its proper period (this is detailed next).

Algorithm 1 describes how to build the paper-centric co-authorship layer of the network and to collect data of social distance and knowledge flow for later analysis. It is important to keep in mind that only the paper-centric co-authorship layer is indeed built, while information from the citation layer is implicitly used to compute the occurrence of knowledge flow.

In order to ensure that the social distances are computed correctly, the co-authorship layer expands gradually in a discrete year-by-year fashion. For each year, the graph is incremented with papers published that year and for each of those papers, pairs are formed by visiting all other nodes in the graph. By doing so, we only compute those social distances that are related to at least one paper that was published in the most recent year available in the graph. To shorten our notation, we define $d(X, Y)$ (without the subscript) as the aforementioned contextualized distance between papers X and Y . Even though it would be more precise to consider the passage of time in a more granular fashion, the use of years is imposed by the data available. This limitation is argued to be negligible because of the nature of the application: while recent papers are more likely to receive new citations [Redner 2005], both research and publishing take time, imposing a delay of at least a few months between the publishing of a paper and the first appearance of its citation.

Algorithm 1 Extracting social distances and occurrence of knowledge flow.

```

1:  $G \leftarrow \text{empty\_graph}$ 
2:  $d \leftarrow \{\}$ 
3:  $y \leftarrow \{\}$ 
4: for each available year  $n$  taken in order do
5:    $G.\text{vertices} \leftarrow G.\text{vertices} \cup \{\text{paper with year} == n\}$ 
6:   for each paper  $A$  with  $\text{year} = n$  do:
7:     for each paper  $B$  from  $G.\text{vertices}$  with  $B \neq A$  do:
8:       if  $A$  and  $B$  share at least one common author then
9:          $G.\text{edges} \leftarrow G.\text{edges} \cup \{(A, B)\}$ 
10:      end if
11:    end for
12:  end for
13:  for each paper  $A$  with  $\text{year} = n$  do:
14:    for each paper  $B$  from  $G.\text{vertices}$  with  $B \neq A$  do:
15:       $d(A, B) \leftarrow \text{num\_intermediate\_nodes\_in\_shortest\_path}(A, B)$ 
16:      if  $A$  cites  $B$  and there is no person being author of both  $A$  and  $B$  then
17:         $y(A, B) \leftarrow 1$ 
18:      else
19:         $y(A, B) \leftarrow 0$ 
20:      end if
21:    end for
22:  end for
23: end for

```

2.2. Data collection and pre-processing

We decided to use data from PLOS ONE³ for the following reasons: first, PLOS ONE content is published under the permissive Creative Commons license which allows its usage for any purpose⁴; second, its data is made available in XML files making it easy to be gathered and processed; third, it comprises publications from different fields providing a broad (yet limited) view on patterns of scientific publication; and finally, its large corpus of more than 150,000 papers allows the usage of tools that exploit statistical characteristics of data.

A preprocessing step is performed to transform the original XML data into a database, in which it is included: all PLOS ONE papers (and their authors), each of the cited papers (and their authors) – which included several non-PLOS ONE papers –, and each citation made by PLOS ONE papers (information on citations made by non-PLOS ONE papers were not available). Data was collected on February 2016.

2.3. Entities Disambiguation

Precise connections are essential in our analysis: we want to be able to properly assign authors to papers and citations between papers. Unfortunately, no identifier is available for authors and the DOI⁵ was present for only a fraction of 15% of all papers. This led to the necessity of dealing with pure text information to establish an approximate canonical representation of those entities.

When DOI is not available, matching of papers is solely based on the title information. Since it is rare that two different papers have the same title, our effort was to deal with a single paper title being spelled in slightly different ways. The method used is described as follows: non-alphabetical characters were removed, all letters were changed to uppercase and exceeding blank spaces were removed. We manually checked a random sample of one hundred papers matched by title and the result was satisfactory (at least 95% of the matches occurred as expected).

Author names were already provided as separate parts: surname and given names. In one hand, since it is common that two different authors have similar names, it is possible to assign two different authors names to the same entity (considering the task of name matching as a classification, this case would be a false positive). On the other hand, depending on the abbreviation used, two different representations of one author's name may lead to duplicate entries in the database (that being a false negative). Our approach here was the one adopted by [Barabási et al. 2002]: surnames were changed to uppercase, and given names were represented as initials (all characters but capital letters were removed). The parts were concatenated into what we call canonical author name. We manually checked a random sample of one hundred authors matched by name and the strategy was satisfactory (at least 85% of the matches occurred as expected).

Also, we are aware of more robust disambiguation methods that could be used.

³<http://journals.plos.org/plosone>

⁴PLOS ONE also stands up for Open Access Research, a publishing format that does not charge users for having access to the full outputs of published research. For more information, see <https://www.plos.org/open-access>.

⁵The digital object identifier (DOI) is a code that uniquely identifies electronic documents, as scientific papers.

Nevertheless, we believe that the chosen method is simple enough to be easily replicated and that its false negatives introduce only a negligible noise on the effect of social ties to be analyzed.

3. Extraction and Analysis Methodology

The following aspects of the methodology are also implemented in the aforementioned source code we made available.

3.1. Geographic Localization of Knowledge Flow

Taking the geographic dimension into consideration when studying knowledge diffusion is relevant to broaden the understanding of its patterns and to provide richer information for strategic decisions of public and private institutions.

A widely accepted pattern is that knowledge flow is geographically localized: it occurs mostly between teams geographically close to each other [Singh 2005]. However, considering our definition of co-authorship distance (or social distance) between papers, it is reasonable to state that geographically closer papers tend to be socially closer too (for instance, in Brazil, the increase of 100 km between two researchers reduces the probability of collaboration by 16% on average [Sidone et al. 2016]). Therefore, there is space to explore the joint effects of social and geographic distances in the occurrence of knowledge flow.

Fortunately, geographic information is available in our dataset: information of country and region⁶ is available in 99% of PLOS ONE papers and 83% of the others. To be precise, the country and region of each paper were inferred from the countries and regions of its authors. In order to decide which country and region to assign for each paper, we tested two approaches: taking the country and region of the first author and taking the majority of the countries and regions between the authors. The experiment revealed insignificant difference between these approaches and therefore the former one was chosen for further discussions.

When comparing a pair of papers with geographic assignments, we can then define two Boolean variables: *is_same_country*, which is 1 if, and only if, the two papers are from the same country; and *is_same_region*, which is 1 if, and only if, the two papers are from the same region. Even though these variables do not accurately represent the distance between the two addresses, it provides a good approximation.

3.2. Variables and Representation

Considering a pair of papers (X, Y) , we have so far defined three categories of variables: occurrence of knowledge flow $y(X, Y)$, co-authorship distance $d(X, Y)$, and geographic assignments $is_same_country(X, Y)$ and $is_same_region(X, Y)$. Taking into consideration that the co-authorship distances, when finite, increase much slower than the number of papers in the network [Watts and Strogatz 1998], the distance variable was represented as a one-hot vector, in which position i is 1 only if the distance between the pair of papers is equal to i and all other positions are equal to 0.

⁶Considering the hierarchy in an address line, the region is the highest entity after the country. In our dataset, it is commonly a state or a city.

In order to study the impact of each of these variables in the probability of knowledge flow, we trained Logistic Regression [Bishop 2006] models and analyzed their coefficients. The higher a coefficient, the most relevant the variable is considered in determining the occurrence of knowledge flow.

3.3. Sampling Strategy and Weighted Regression Model

It is possible to note that the number of possible pairs of papers increases quadratically as more papers are added to the network. In fact, the straightforward approach of using all data to train regression is unnecessarily expensive, since a large portion of this data corresponds to pairs for which there is no occurrence of knowledge flow. This issue is addressed next.

In a large dataset, the occurrence of knowledge flow is a rare event, i.e. the probability of randomly choosing a pair of papers in which knowledge flow is present is very low. In other words, it was observed that the occurrence of knowledge flow ($y(X, Y) = 1$) is a highly infrequent event in comparison with the volume of all pairs of papers. Therefore, an efficient method is to use all registers with $y(X, Y) = 1$ and just a small sample taken from the $y(X, Y) = 0$ cases:

$$num_samples_{y(X,Y)=1} = \alpha * num_records_{y(X,Y)=1}$$

$$num_samples_{y(X,Y)=0} = \beta * num_records_{y(X,Y)=0},$$

in which $num_samples$ is the number of samples used to train the regression, $num_records$ is the total number of samples in our database and α and β are coefficients indicating the proportion of the database to be sampled.

Aiming at avoiding the introduction of bias because of this endogenous stratified sampling, we associated each sample with a weight inversely proportional to the probability of it being sampled:

$$weight_{y(X,Y)=1} = 1/\alpha$$

$$weight_{y(X,Y)=0} = 1/\beta$$

In other words, the unbalance between $num_records_{y(X,Y)=1}$ and $num_records_{y(X,Y)=0}$ is equiposed by the weighted regression. In our approach, α is 1 and β is taken empirically: we take β so that $num_samples_{y(X,Y)=1} / num_samples_{y(X,Y)=0} = \beta$ and, thus, $\beta = number_of_citations / number_of_papers^2$.

4. Results and Discussion

On average, each of the 152,406 PLOS ONE papers was authored by 6.8 people and cited 43.3 papers. Given that 92% of these papers are from the major subject Biology and Life Sciences, the high number of authors per paper is expected [Newman 2004]. The database resulting from the preliminary step consists of 3,373,968 papers (including papers that are cited by PLOS ONE papers), 2,435,359 authors and 6,294,099 citations.

Collected PLOS ONE papers range from 2006 to 2015. Given that computing the shortest paths in a large network is a very time-consuming task, we only managed to build the graph up to 2009, encompassing 8,340 PLOS ONE papers (we used the *igraph*⁷ library and it took 7 days of computation on a DELL PowerEdge R430 server with Intel® Xeon® E5-2609 processor and 96GB of RAM). Even though this corresponds to just 5% of the total PLOS ONE papers, we were able to observe some interesting phenomena. The resulting collaboration network consisted of 2,353,314 nodes (papers) and 313,507,251 edges.

The coefficients obtained for the logistic regression are displayed in Figure 3. They evidence that low social distances lead to a high frequency of knowledge flow, confirming the initial expectations. Despite it is possible to note a clear decrease in the effect of distance, the tendency smooths at some point. Interestingly, this point (around 3 and 4) is consonant with [Christakis and Fowler 2013], which states that the effects of social ties (e.g. in diffusion of behaviors) fades until a distance of approximately 3 (beyond which it becomes inconspicuous). It is important to keep in mind that this comparison was drawn between our results from a network of produced papers and others directly from social networks. A systematic exploration of the effects of an author-centric network in the corresponding document-centric network (one can name it: scientific papers, patents, software projects) would be useful to support this kind of comparison and to deepen the understanding of social collaboration.

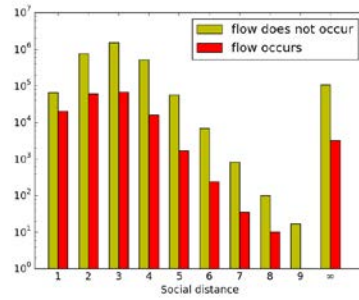
We also fit a second set of logistic regressions, but using the binary variables *is_same_country* and *is_same_region* in addition to the variables indicating social distance, in order to compare the effects of geographic and social distances. Figure 4 shows the coefficients of the regressions. It is possible to see that the coefficients related to the social distance are extremely close to those observed in the previous regressions. Also, it is noticeable that coefficients related to the geographic variables are less relevant than those related to short social distances, which indicates that being in the same geographic region matter less than being socially close to other teams for the occurrence of knowledge flow.

5. Concluding Remarks

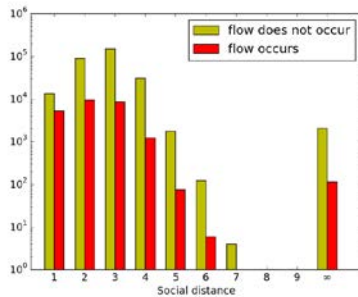
Comparing our results to those presented by [Singh 2005] related to patent data, geographic location is markedly less relevant in the present work. Such finding may indicate either that academic research is globally more integrated than industrial R&D or that citation practices are significantly different between both communities.

Though restricted to the context of scientific production, it is possible to generalize the main consequences of the obtained results to other types of organizations. The results are suggesting that the promotion of information flow is not only influenced by the geographical proximity of key players, but also by the density of the social interactions established. Information technology is certainly a key instrument to mitigate the effects of geographical distance, together with initiatives to establish offices and representatives in strategic locations. Additionally, any organization devoted to maximizing the information flow among its peers should strongly encourage the creation of solid professional links involving external actors. This is particularly interesting for developing countries,

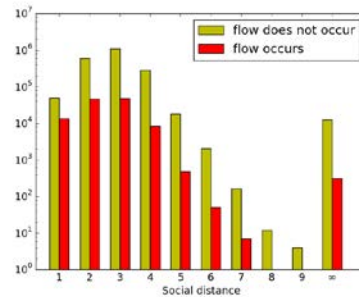
⁷<http://igraph.org>



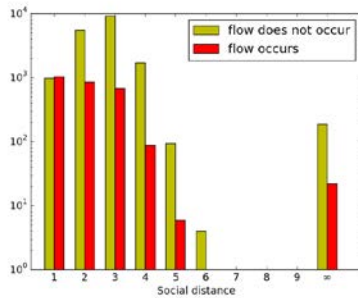
(a) All pairs



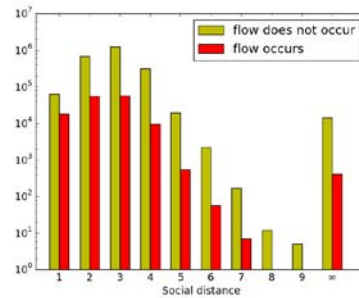
(b) Same country



(c) Different countries



(d) Same region



(e) Different regions

Figure 2. Histograms for 5 cohorts: (a) all sampled pairs, (b) pairs of the same country, (c) pairs of different countries, (d) pairs of the same region and (e) pairs of different regions. For each plot, the horizontal axis presents each variable of the one-hot distance vector $d(X, Y)$. The reason we do not consider the case where $d(X, Y) = 0$ (both here and in the regression) is that by construction, these cases always imply $y(X, Y) = 0$. The largest observed finite social distance between two papers was 9, which reflects the small-world effect. It is also possible to note the effect of absence of geographic assignment in many papers (for instance, the sum of pairs of the same country (b) with pairs of different countries (c) is lower than the unrestricted total (a)).

like Brazil, as it indicates that strategic foreign social connections are a valuable path for economic and scientific growth as a result of the tapping of knowledge from abroad.

5.1. Future work

A clear point for improvement in this experiment would be to enrich the database. Since the information of citations was not available in any of the non-PLOS ONE papers, there

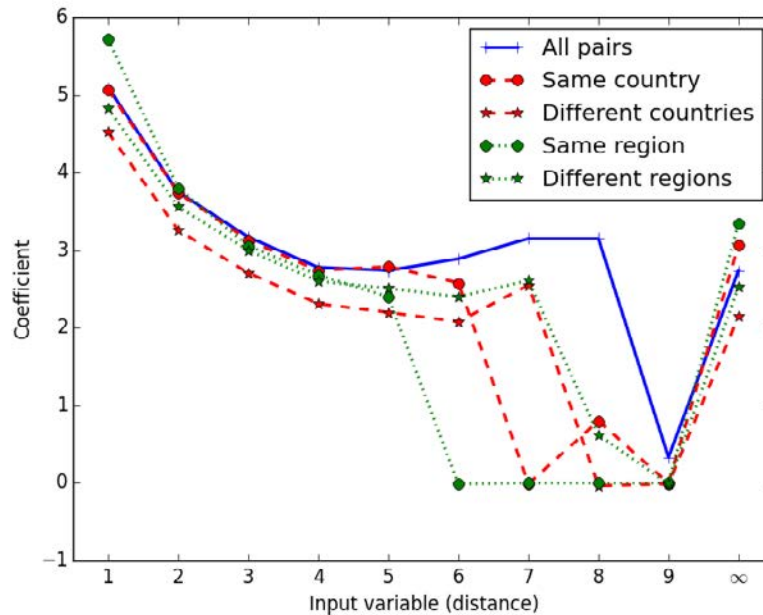


Figure 3. Each set of data in Figure 2 was used to fit the regression: the x-axis represents the binary variables of the one-hot input vectors and the y-axis corresponds to the coefficients of the regression in each case. It is possible to observe a clear decrease in relevance to predict knowledge flow as the distance increases (specially until a distance of 4 is reached).

is space to improve the approximation of the occurrence of knowledge flow. In that sense, co-authorship also represents only a fraction of the actual social ties between authors, and other sources of interactions could be incorporated.

Another useful analysis could be drawn considering the subject of papers, which would allow the comparison between general fields.

Finally, in order to speed-up the time for computing the distances in the co-authorship network, we plan to incorporate the usage of a GPU-based implementation to leverage the possibility of calculating the shortest paths in parallel.

6. Acknowledgements

We thank Elaborata Consultoria for providing the computational structure to collect the data and run the experiments. We also thank Leonardo Maia for the helpful suggestions on working with data provided by PLOS ONE.

References

- Barabási, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, a., and Vicsek, T. (2002). Evolution of the social network of scientific collaboration. *Physica A*, 311(March 2016):590–614.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

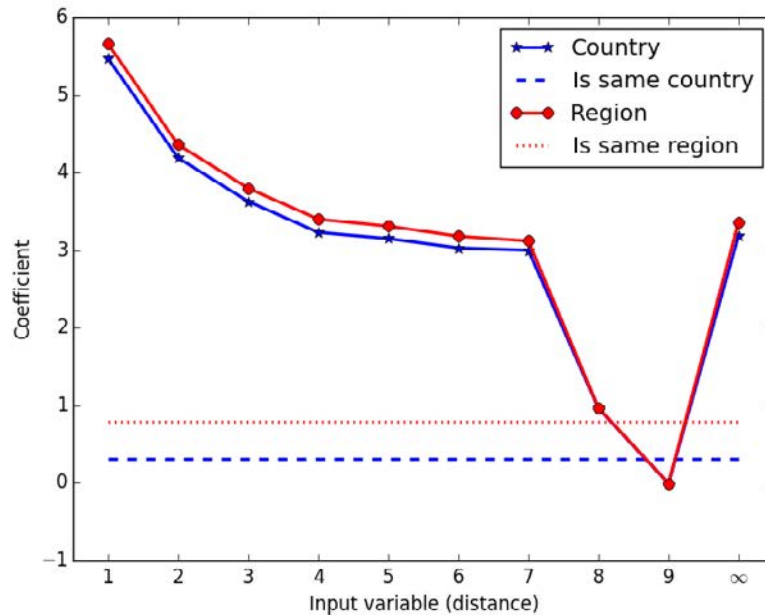


Figure 4. Data of cohorts (b) and (c) of Figure 2 were used to generate the aforementioned Boolean variable *is_same_country* and jointly fit the regression. The resulting coefficients are displayed with cross marks and the coefficient for variable *is_same_country* is presented as the dashed horizontal rule. The same process was performed to data of cohorts (d) and (e) of Figure 2 and the coefficients are presented with bullet marks. The coefficient related to variable *is_same_region* is presented as the dotted horizontal rule.

- Christakis, N. a. and Fowler, J. H. (2013). Social contagion theory: Examining dynamic social networks and humanbehavior. *Statistics in Medicine*, 32(4):556–577.
- Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2002). Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science*, 48(June 2015):1–23.
- Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3):577–598.
- Liberman, S. and Wolf, K. B. (1997). The flow of knowledge: Scientific contacts in formal meetings. *Social Networks*, 19(3):271–283.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101 Suppl:5200–5.
- Powell, W. W. (1990). Neither Market nor Hierarchy: Network Forms of Organization.
- Redner, S. (2005). Citation Statistics from 110 Years of Physical Review. *Physics Today*, 58(June):49–54.
- Sidone, O. J. G., Haddad, E. A., and Mena-Chalco, J. P. (2016). Scholarly publication and collaboration in Brazil: The role of geography. *Journal of the Association for Information Science and Technology*, 14(4):n/a–n/a.

- Singh, J. (2005). Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, 51(5):756–770.
- Sorenson, O. and Singh, J. (2006). Science, Social Networks and Spillovers. *SSRN Electronic Journal*, (9):11.
- Trajtenberg, M., Henderson, R., and Jaffe, A. B. (1992). Ivory Tower Versus Corporate Lab: An Empirical Study of Basic Research and Appropriability. (4146).
- Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2.

Extração de Métricas e Análise de Sentimentos em Comentários *Web* no Domínio de Hotéis

Roney L. de S. Santos¹, Raimundo S. Moura¹

¹Departamento de Computação – Universidade Federal do Piauí (UFPI)
64.049-550 – Teresina – PI – Brasil

roneylira@hotmail.com, rsm@ufpi.edu.br

Abstract. *The significant growth of the Web has made from it a rich source for the evaluate of public opinion on a specific entity. Consequently, the number of opinions available makes the decision-making process impossible if it was necessary to read and to analyse all reviews. This paper presents a Web application prototype where from a review are returned the feeling (positive, negative or neutral), its features and other analysis metrics using Natural Language Processing and Sentiment Analysis. Experiments show efficacy in the precision of reviews with negative polarity and recall of reviews with positive polarity in 84.93% and 94.33% respectively.*

Resumo. *O significativo crescimento da Web tem a tornado uma rica fonte para a avaliação da opinião pública sobre uma entidade específica. Consequentemente, o número de opiniões disponíveis torna impossível uma tomada de decisão se for necessário ler e analisar todas as opiniões. Este trabalho apresenta um protótipo de uma aplicação Web que a partir de um comentário seja retornado o sentimento (positivo, negativo ou neutro), suas características e outras métricas de análise utilizando técnicas de Processamento de Linguagem Natural e Análise de Sentimentos. Experimentos mostram eficácia nos resultados de precisão de comentários com polaridade negativa e cobertura de comentários positivos em 84.93% e 94.33% respectivamente.*

1. Introdução

Quando um consumidor tem interesse por um produto ou serviço é comum que ele procure referências ou opiniões. Empresas que vendem produtos e disponibilizam serviços também são motivadas a ter conhecimento das opiniões dos consumidores, tendo que procurar formas de analisar essas informações para conduzir ações de *marketing* e tomada de decisão.

Pesquisadores da área de Processamento de Linguagem Natural (PLN) têm buscado extrair informações úteis de dados não estruturados. O termo *big data* se refere ao massivo conjunto de informações digitais coletadas. Todos os dias são criados 2.5 quintilhões de *bytes* de dados, tanto que 90% dos dados do mundo hoje foram criados apenas nos últimos dois anos¹. Essa grande quantidade de dados faz com que a análise manual se torne uma tarefa impossível, sendo necessária a criação de métodos automáticos para analisar os dados [Liu 2010].

¹http://www.isaca.org/groups/professional-english/big-data/groupdocuments/big_data_top_ten_v1.pdf

Dessa forma, para auxiliar no processo de análise e formação de opinião do consumidor, este trabalho propõe um protótipo de uma aplicação *Web* baseada no domínio de hotéis onde a entrada é um comentário e como saída são retornadas a polaridade do comentário (positivo, negativo ou neutro), as características encontradas no comentário, número de tuplas <característica, palavra opinativa> e a corretude das palavras do comentário. Para construir as saídas da aplicação foram utilizadas técnicas de PLN e Análise de Sentimentos.

As saídas são métricas que serão utilizadas em sistemas para inferir os comentários mais úteis. Em alguns sites de hotéis, usuários podem votar em comentários que eles consideram úteis ou inúteis quando estão pesquisando sobre o hotel. Entretanto, nem sempre apenas informações de polaridade do comentário [Kim et al. 2006] são suficientes pois outros problemas, como por exemplo, comentários mais novos que nunca foram votados nunca serão lidos e votados [Li et al. 2013] podem acontecer.

Os experimentos mostraram que o protótipo tem grande cobertura dos comentários positivos e uma boa precisão dos comentários negativos. Os comentários considerados neutros podem ter sua polaridade alterada caso uma análise de contexto fosse realizada. O protótipo da aplicação também mostrou as características encontradas e suas polaridades, para uma análise mais fiel do usuário de hotéis.

O restante do artigo está organizado como: A Seção 2 relata a Fundamentação Teórica deste trabalho. A Seção 3 destaca os trabalhos relacionados. A Seção 4 discorre sobre o funcionamento do protótipo da aplicação *Web*. A Seção 5 mostra os experimentos realizados e os resultados. Por fim, a Seção 6 conclui o artigo e discute os trabalhos futuros.

2. Análise de Sentimentos e Mineração de Opinião

Análise de Sentimentos ou Mineração de Opinião é o estudo computacional de opiniões, avaliações, atitudes e emoções das pessoas referenciando entidades, indivíduos, questões, eventos, tópicos e seus atributos. O estudo é realizado sobre a emoção do usuário em relação ao item, classificando a emoção em positiva, negativa ou neutra, sendo determinada por técnicas de PLN. É tecnicamente desafiador e muito útil.

[Liu 2012] define uma opinião sendo uma quintupla $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ onde:

- e_i é o nome de uma entidade;
- a_{ij} é um aspecto de uma entidade (opcional);
- s_{ijkl} é o sentimento sobre um aspecto a_{ij} da entidade e_i , emitido pela pessoa h_k no tempo t_l ;
- h_k é o detentor da opinião (*opinion holder*);
- t_l é o instante no qual a opinião foi expressada por h_k .

Um típico sistema de Análise de Sentimentos envolve três etapas: i) extração de opiniões e suas características; ii) identificação da orientação semântica da opinião; e iii) sumarização dos resultados. A detecção do sentimento de um texto pode ocorrer em diferentes níveis: documento, sentença e entidade ou aspecto. Pesquisas que contém mineração de opinião e métricas para sumarização da informação são encontradas em [Khan et al. 2009], [Seerat and Azam 2012] e [Vinodhini and Chandrasekaran 2012].

3. Trabalhos Relacionados

Nesta seção são discutidos os conceitos utilizados neste artigo através de trabalhos relacionados, como aplicações *Web* usando Análise de Sentimentos e utilização de ontologias no âmbito do PLN.

São vários os estudos onde a extração de sentimento de características em comentários de consumidores é realizada através de Mineração de Opinião abrangendo técnicas de PLN [Li et al. 2009, Hu and Liu 2004, Htay and Lynn 2013]. [Aravindan and Ekbal 2014] propõem técnicas baseadas em mineração de dados e aprendizado de máquina para analisar comentários automaticamente, extraíndo as características do produto e determinando a polaridade de cada característica. Seus experimentos mostraram que o classificador teve uma precisão de 79.67% quando aplicados a cinco produtos populares.

[Farina et al. 2014] introduziram um *framework* onde permite aos desenvolvedores a executar um classificador de sentimentos e a validação dos resultados tanto em um ambiente local quanto em um ambiente distribuído (*web*), permitindo seleções de comentários em tempo real e representação gráfica de várias dimensões dos comentários (tempo, local, palavras-chave, fontes e o sentimento).

Redes Sociais também têm a atenção dos pesquisadores quando se trata de Análise de Sentimentos. [Sarlan et al. 2014] desenvolveram uma protótipo de uma aplicação *web* que através de uma vasta quantidade de *tweets*² é possível classificar as perspectivas dos consumidores através do sentimento extraído da mensagem textual.

A utilização de aprendizado de máquina é comum na análise do sentimento de um comentário. Trabalhos utilizando Máquinas de Vetores de Suporte (*Support Vector Machine*, SVM) [Mullen and Collier 2004, Tan et al. 2011], Sistemas Fuzzy [de Sousa et al. 2015, Santos et al. 2016, Indhuja and Reghu 2014, Dragoni et al. 2014] e Redes Neurais Artificiais (RNA) [Sharma and Dey 2012a, Sharma and Dey 2012b] são amplos na área. [Laryea et al. 2015] utilizaram um aprendizado de máquina supervisionado para classificar comentários da *Amazon* em cinco tipos: positivo, negativo, conselho, neutro e sem sentimento. Segundo eles, alguns comentários contém soluções técnicas, maneiras improvisadas de resolver os problemas, frustrações, felicidade, entre outros.

Para trabalhos onde existe um domínio específico, ontologias são desenvolvidas para facilitar a análise e extração de características. [De Freitas and Vieira 2013] apresentam métodos de avaliação para identificar polaridade em comentários escritos em Português de acordo com características descritas na ontologia definida. [Yannis and Evanthia 2012] propõem uma aplicação que constrói uma base de ontologia automaticamente através de textos de resultados de mecanismos de pesquisa com suas respectivas regras semânticas. No âmbito de ontologias referentes a hotéis, [Hasany and Selamat 2011] apresentam um sistema que coleta informações de hotéis através de ontologia e linguagem natural para tomada de decisão. Experimentos mostraram eficácia na ajuda da aplicação na tomada de decisão do usuário.

Neste trabalho o foco está além do sentimento do comentário. Métricas que são importantes para uma futura utilidade do comentário em outros contextos são coletadas e

²Mensagens de texto da rede social Twitter

disponibilizadas para o usuário através de uma aplicação *web*. Além disso, nenhum dos trabalhos relacionados citados nesta seção analisam as características individualmente, conhecendo a polaridade de cada uma, o que é uma das propostas desse trabalho.

4. Protótipo

O protótipo da aplicação *Web* recebe um comentário como entrada e retorna como saídas: i) a polaridade do comentário (positivo, negativo ou neutro); ii) as características e suas polaridades, analisando as palavras opinativas referentes; iii) a quantidade de tuplas <característica, palavra opinativa> encontradas; e iv) a corretude das palavras no comentário. As etapas de execução são mostradas na Figura 1.

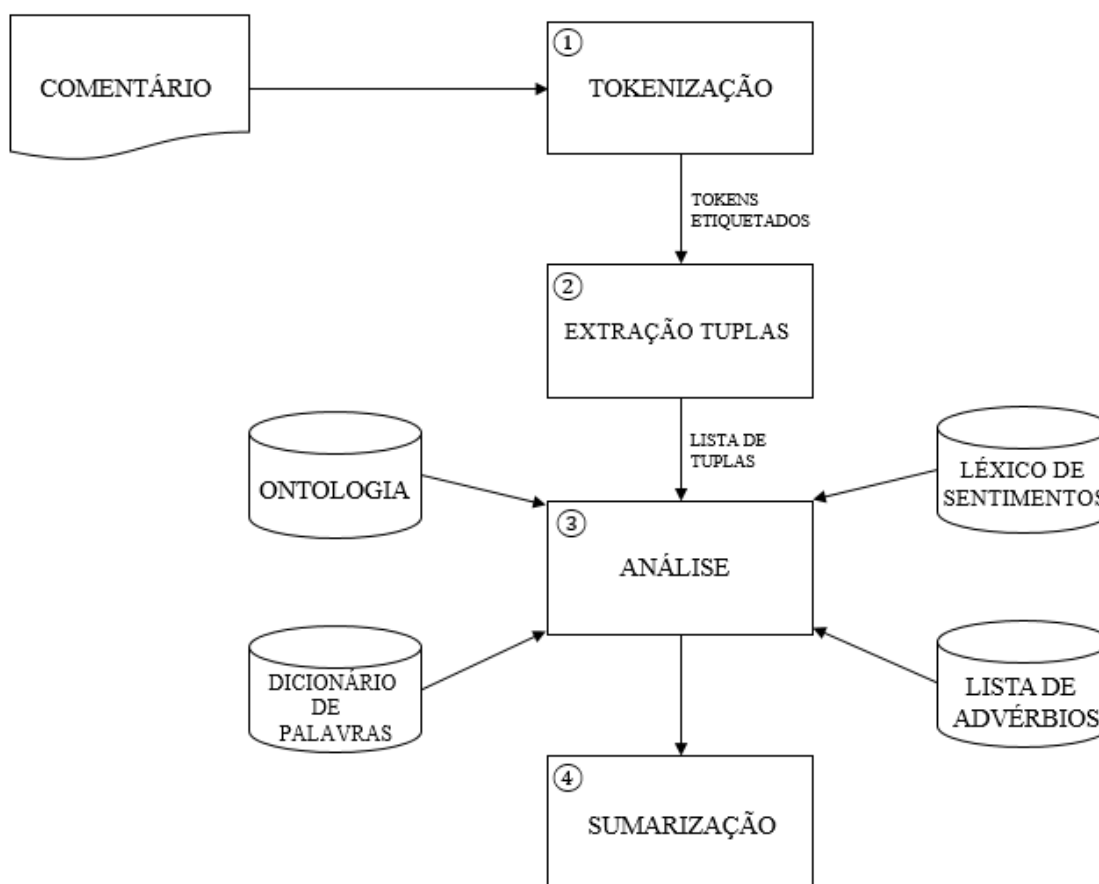


Figura 1. Protótipo: etapas de execução

Dado um comentário, a primeira etapa do protótipo foi realizar a separação das palavras do comentário em *tokens* (tokenização). Antes dos *tokens* serem submetidos ao etiquetador, é realizado um pré-processamento que substitui as preposições “da”, “do”, “das” e “dos” em uma estrutura “preposição + artigo”, por exemplo, da = de+a. Esse procedimento foi necessário pois o etiquetador não reconhece essas contrações como preposições e conseqüentemente não reconheceria palavras compostas como “café da manhã”.

A segunda etapa do protótipo consistiu em extrair as tuplas (palavra, classe gramatical) do comentário. Uma lista com tuplas foi gerada como resultado do etiquetador.

As classes gramaticais foram definidas pelo etiquetador *Mac-Morpho* presente no NLTK (*Natural Language Toolkit*) [Bird et al. 2012].

Logo após, na terceira etapa, ocorreu a análise do comentário. Para a realização dessa análise, foram construídas quatro bases de dados. Inicialmente, para a construção das bases, foram coletados 2.216 comentários do site *Booking.com*³ de um hotel escolhido aleatoriamente. A coleta foi feita utilizando um *Web crawler* construído com a ferramenta *Scrapy* que captura os comentários através do arquivo HTML da página. Foram coletados a data, o autor, a nota e o comentário em geral.

Esses comentários serviram de base para uma análise manual. Essa análise manual foi executada coletando 30 comentários aleatoriamente da base de comentários e identificando características, palavras opinativas, advérbios e verbos presentes nos comentários. Em alguns comentários foi possível identificar os sintagmas (nominal, verbal e preposicional). Foram identificados vários padrões linguísticos que foram compactados em cinco padrões principais, descritos na Figura 2. O padrão 1 coleta as características e palavras opinativas presentes no comentário, contendo ou não advérbios antes do adjetivo. Caso exista um verbo de ligação entre a característica e a palavra opinativa, o padrão 1 também reconhece a estrutura “sujeito + verbo de ligação + predicativo do sujeito”. O padrão 2 coleta as características e um ou mais adjetivos que se referenciam à ela. O padrão 3 por sua vez coleta as características e suas palavras opinativas quando aparecem no comentário na ordem inversa. O padrão 4 verifica se junto das características se encontram advérbios e verbos, uma vez que alguns verbos também transmitem sentimentos que podem ser explorados [Klavans and Kan 1998]. Por fim, o padrão 5 verifica o inverso do padrão 4, já que em alguns casos o verbo pode vir antes do advérbio.

```
padrão 1: (<SUBST>|<SUBST><PREP><SUBST>)<ADV>?<VERBO>?<ADV>?<ADJ>
padrão 2: (<SUBST>|<SUBST><PREP><SUBST>)<VERBO>?(<ADV>?<ADJ><, >)*<ADV>?<ADJ><CONJ><ADV>?<ADJ>
padrão 3: <ADJ>(<SUBST>|<SUBST><PREP><SUBST>)
padrão 4: <ADV><VERBO><PREP>(<SUBST>|<SUBST><PREP><SUBST>)
padrão 5: <VERBO><ADV><PREP>(<SUBST>|<SUBST><PREP><SUBST>)
```

Figura 2. Padrões linguísticos identificados

Com o uso das características extraídas na análise manual, foi criada uma ontologia com as principais características de hotéis encontradas. O *Booking.com* define 7 características em cada página de cada hotel. Da análise manual foram encontradas mais 51 características, totalizando 58. Com o intuito de aumentar o escopo da ontologia, foi utilizado o ConceptNet [Liu and Singh 2004] coletando os sinônimos das características previamente encontradas. Foram encontradas mais 28 características, totalizando 86 características na base. Uma filtragem manual foi realizada a fim de eliminar as características com o mesmo *stemming*⁴, finalizando com 81 características. Essa ontologia força a aplicação a ser direcionada ao domínio de hotéis e está disponível online⁵.

³<http://www.booking.com>

⁴Termo usado na linguística para descrever o processo de redução (flexionado ou derivado) da palavra.

⁵<https://goo.gl/HuKuGY>

A fim de fazer uma análise mais profunda do sentimento, foi construída uma lista de advérbios, sugerida inicialmente por [de Sousa et al. 2015], contendo 76 advérbios, a qual engloba boa parte dos advérbios da língua portuguesa e suas polaridades. Proveniente da análise manual realizada anteriormente, 6 advérbios coletados foram adicionados na base de advérbios, totalizando 82 advérbios.

Para analisar o sentimento de cada comentário foi utilizado o Sentilex-PT [Silva et al. 2012]. Uma base foi construída extraíndo as palavras do arquivo disponibilizado pelo Sentilex-PT. Dentre os atributos de cada palavra, apenas a polaridade da palavra foi considerada. Para a realização da tarefa de correção das palavras do comentário, foi construído um dicionário de palavras. A base de palavras foi retirada da *WordNet* [Dias-Da-Silva 2010] e contém 250.196 palavras.

Com todas as bases construídas, a aplicação foi capaz de analisar a polaridade de cada característica, a quantidade de tuplas e a correção de cada comentário. A partir do momento que um comentário é submetido para análise, a aplicação utiliza todas as bases construídas, uma vez que a aplicação segue o domínio de hotéis.

Tomando como base os padrões encontrados (ver Figura 2), a lista de tuplas foi analisada para cada padrão a fim de salvar quais tuplas contém características, adjetivos, advérbios e verbos. A aplicação retorna o número de tuplas encontradas no comentário.

Ao mesmo tempo a polaridade do comentário foi analisada. Os adjetivos e verbos foram submetidos ao Sentilex-PT que retornou um número inteiro referente à sua polaridade: 1 para positivo, -1 para negativo ou 0 para neutro. Caso junto ao adjetivo contenha um advérbio, a polaridade do adjetivo é alterada após o advérbio ser submetido à base de advérbios. Por exemplo, se o adjetivo “bom” estiver acompanhado do advérbio “muito”, sua polaridade é aumentada para 1,5. Por outro lado, se o adjetivo “ruim” estiver acompanhado do mesmo advérbio, sua polaridade é diminuída para -1,5. Ao final, as polaridades de cada adjetivo ou do conjunto advérbio+adjetivo do comentário são somadas e caso a soma resulte em um valor maior que 0 o comentário é considerado positivo; caso resulte em valor menor que 0 o comentário é considerado negativo e por fim, caso resulte em 0 o comentário é considerado neutro.

Por fim foram extraídos todos os adjetivos referentes a cada característica para analisar a impressão geral do comentário sobre cada uma delas. Essa informação é importante devido ao fato de que o usuário pode pesquisar apenas por algumas características do hotel, como quartos e limpeza, em vez do hotel em geral. Essa análise se deu pela ontologia construída. Para cada característica do comentário presente na base de características, os adjetivos (e advérbios quando o acompanhando) eram salvos para a análise de polaridade. Essa análise seguia as mesmas regras da polaridade do comentário, verificando a polaridade do adjetivo pelo Sentilex-PT, dos advérbios pela base de advérbios, somando os valores e resultando na polaridade daquela característica.

Ao final de todo o processo, na quarta etapa, a aplicação mostra na tela de resultados a polaridade do comentário, as características encontradas no comentário, um gráfico informando quais são as características positivas, negativas e neutras, a quantidade de tuplas encontradas e a correção do comentário. A Figura 3 mostra a tela de resultados da aplicação.

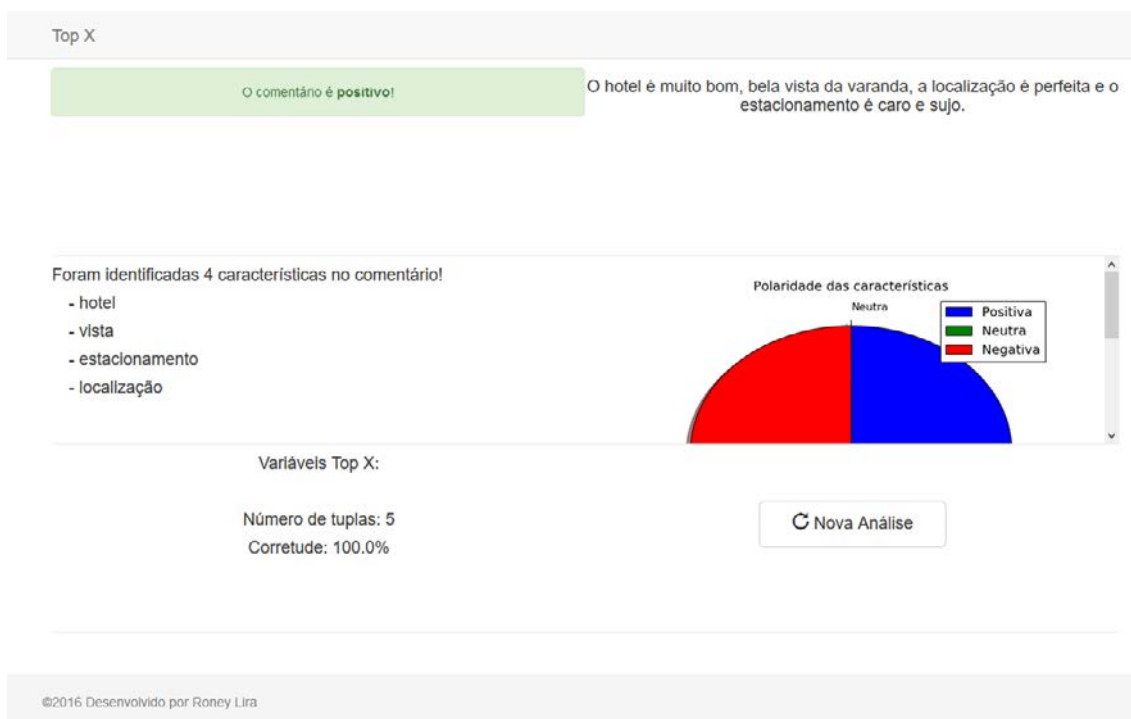


Figura 3. Protótipo: tela de resultados

5. Experimentos e Resultados

Para a validação do protótipo da aplicação *Web* foi realizado um experimento que verifica a precisão e cobertura (*recall*) da polaridade dos comentários analisados pelo protótipo, sendo mostrados os resultados obtidos.

A realização do experimento contou com uma base de 328 comentários aleatoriamente coletados da base existente, construída na etapa de análise. O número de comentários foi escolhido através de uma análise estatística que leva em consideração o nível de confiança e a margem de erro da amostra. Para esse trabalho o nível de confiança escolhido foi de 95% e a margem de erro escolhida foi de 5%.

Os 328 comentários foram analisados manualmente por três especialistas da área de linguística, graduandos em Letras Português no mesmo nível de curso, definindo cada comentário como positivo, negativo ou neutro. Primeiramente foi verificada a análise dos dois primeiros especialistas, onde se houvesse conflito de escolha da polaridade, a análise do terceiro especialista definiria a polaridade para o comentário.

Após a conclusão da análise manual, os comentários foram submetidos ao protótipo da aplicação *Web* que retornou, entre outras saídas, a polaridade do comentário. Dos 328 comentários submetidos, o protótipo retornou 212 comentários positivos, 73 comentários negativos e 43 comentários neutros. Em comparação com os resultados da análise manual, houve aumento de 25% na detecção dos comentários positivos e queda de 34% e 26% na detecção dos comentários negativos e neutros, respectivamente.

O aumento na detecção dos comentários positivos e a queda na detecção dos comentários negativos estão relacionados devido ao modo com que o protótipo define a polaridade do comentário, somando as polaridades das palavras opinativas referentes a

cada característica encontrada. Em alguns comentários de contexto negativo continham tuplas com palavras opinativas de polaridade positiva, tendo como consequência o incremento positivo na polaridade final do comentário e transformando um comentário de contexto negativo em positivo. Por exemplo, a Figura 4 mostra um comentário de contexto negativo que o protótipo identificou como positivo. Tuplas com polaridade positivas (suíte ótima, vista maravilhosa) são maioria em comparação com tuplas com polaridade negativas (equipe, meio lenta, serviço, meio devagar).

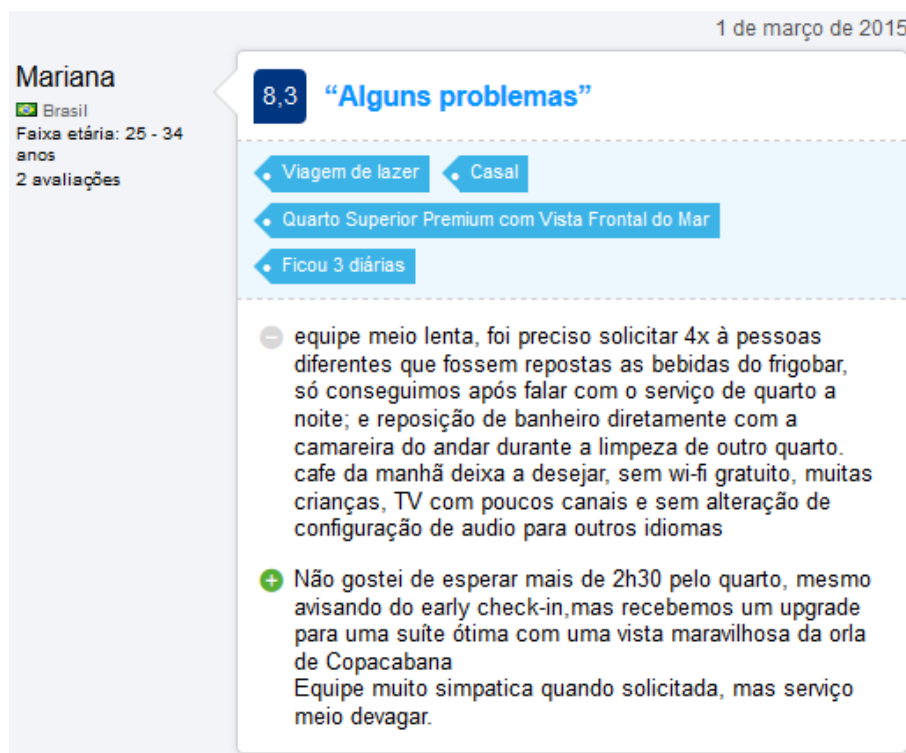


Figura 4. Comentário do *Booking.com*

Utilizando os resultados da execução de todos os comentários, foi feita a análise da polaridade dos comentários verificando a precisão, cobertura (*recall*) e medida-f de cada. Tomando como base a análise manual, os comentários foram separados por polaridade e comparados com os resultados do protótipo, que também foram separados da mesma maneira. A Tabela 1 mostra os resultados.

Tabela 1. Precisão, *Recall* e Medida-F de cada polaridade

	Precisão	<i>Recall</i>	Medida-F
Positivos	70.75%	94.33%	80.86%
Negativos	84.93%	55.85%	67.39%
Neutros	55.81%	41.37%	47.52%

Exemplificando para melhor entendimento: o experimento de precisão verificou o quão preciso foi o protótipo, conseguindo acertar os comentários que são realmente positivos pela análise dos especialistas. O protótipo encontrou 212 comentários positivos mas apenas 150 foram considerados positivos pela análise manual, alcançando assim a pre-

cisão de 70.75% ($150/212 = 0.7075$) dos comentários positivos. O experimento de cobertura (*recall*) verificou quantos foram os comentários positivos encontrados pelo protótipo que são realmente positivos de acordo com a análise manual. A análise manual descreveu 159 comentários como positivos enquanto o protótipo encontrou 150 comentários positivos, cobrindo então 94.33% dos comentários ($150/159 = 0.9433$). A medida-F é uma média harmônica da precisão e cobertura *recall*. O cálculo da medida-F é realizado segundo a equação

$$\frac{2 * p * r}{p + r} \quad (1)$$

onde p representa o valor da precisão e r representa o valor da cobertura (*recall*).

A partir dos resultados do experimentos verificou-se uma grande cobertura de comentários positivos e uma boa precisão de comentários negativos. Os comentários considerados neutros pelo protótipo podem ter a polaridade alterada para positivo ou negativo analisando o contexto do comentário.

6. Conclusão

Este trabalho apresentou o protótipo de uma aplicação *Web* baseada no domínio de hotéis que analisa os comentários e retorna como resultado informações como a polaridade do comentário, quantidade de tuplas <característica, palavra opinativa> a corretude das palavras escritas no comentário e a polaridade das características encontradas no comentário.

Observando a polaridade dos comentários, os resultados foram satisfatórios. Com bons números na precisão e *recall* tanto de comentários com polaridade positiva quanto negativa, pode-se considerar que os padrões linguísticos utilizados na detecção das tuplas <característica, palavra opinativa> são corretos. Porém, em alguns comentários de contexto negativo as tuplas encontradas eram de polaridade positiva, o que de acordo com o método utilizado para definir a polaridade do comentário, a polaridade era invertida. A utilização do contexto pelo ConceptNet é um dos trabalhos futuros planejados.

Na análise manual foi observado que alguns comentários poderiam ter uma melhor análise se houvesse um dicionário de gírias, neologismos, erros mecânicos e contrações. Em relação às contrações, um método de pré-processamento foi implementado para o protótipo se adaptar ao etiquetador das palavras. Um outro trabalho futuro seria a construção do dicionário que englobaria os outros problemas citados anteriormente, já que eles não dependem do domínio inserido do comentário.

A corretude do comentário também pode ser melhorada. Em [Torruella and Capsada 2013] são citadas algumas medidas de análise da riqueza léxica do texto, verificando entre outras características, o tamanho do texto em diferentes tipos de escrita.

A fim de incrementar o protótipo da aplicação *Web*, algumas funcionalidades futuramente serão implementadas. Uma delas é a condição do usuário ter uma opção de submeter um conjunto de comentários e a aplicação *Web* mostrar os resultados em geral, separando as polaridades dos comentários e das características encontradas no conjunto de comentários para melhor análise.

Uma outra funcionalidade a ser implementada no futuro é a importância do comentário. A ideia é baseada em [de Sousa et al. 2015] que utiliza Sistema Fuzzy com três entradas: reputação do autor, quantidade de tuplas e correteude do comentário. A diferença para o trabalho futuro é que uma outra métrica será utilizada: o tempo que o comentário foi postado, uma vez que em relação ao domínio de hotéis, o tempo é importante para o usuário tomar sua decisão. Além disso, as entradas serão ponderadas com o que o usuário mais se identifica e os comentários serão sumarizados em ordem de importância de acordo com as métricas utilizadas.

Referências

- Aravindan, S. and Ekbal, A. (2014). Feature extraction and opinion mining in online product reviews. pages 94–99.
- Bird, S., Klein, E., and Loper, E. (2012). *Natural Language Processing with Python*. O’Reilly.
- De Freitas, L. and Vieira, R. (2013). Ontology-based feature level opinion mining for portuguese reviews. pages 367–370.
- de Sousa, R. F., Rabelo, R. A. L., and Moura, R. S. (2015). A fuzzy system-based approach to estimate the importance of online customer reviews. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8.
- Dias-Da-Silva, B. C. (2010). Brazilian portuguese wordnet: A computational linguistic exercise of encoding bilingual relational lexicons. *International Journal of Computational Linguistics and Applications*, 1(1-2):137–150.
- Dragoni, M., Tettamanzi, A. G. B., and Costa Pereira, C. (2014). *Semantic Web Evaluation Challenge: SemWebEval 2014 at ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers*, chapter A Fuzzy System for Concept-Level Sentiment Analysis, pages 21–27. Springer International Publishing.
- Farina, J., Mazuran, M., and Quintarelli, E. (2014). *New Trends in Databases and Information Systems: 17th East European Conference on Advances in Databases and Information Systems*, chapter Extraction, Sentiment Analysis and Visualization of Massive Public Messages. Springer International Publishing.
- Hasany, N. and Selamat, M. (2011). Answering user queries from hotel ontology for decision making. pages 123–127.
- Htay, S. S. and Lynn, K. T. (2013). Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal*, 2013(394758).
- Hu, M. and Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pages 755–760.
- Indhuja, K. and Reghu, R. P. C. (2014). Fuzzy logic based sentiment analysis of product review documents. In *Computational Systems and Communications (ICCSC), 2014 First International Conference on*, pages 18–22.
- Khan, K., Baharudin, B., Khan, A., and e Malik, F. (2009). Mining opinion from text documents: A survey. In *3rd IEEE International Conference on Digital Ecosystems and Technologies*, pages 217–222.

- Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 423–430.
- Klavans, J. and Kan, M.-Y. (1998). Role of verbs in document analysis. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1, ACL '98*, pages 680–686.
- Laryea, B., Choi, C.-H., Jung, I.-S., Lee, K.-H., and Cho, W.-S. (2015). Web application for sentiment analysis using supervised machine learning. *International Journal of Software Engineering and its Applications*, 9(1):191–200.
- Li, M., Huang, L., Tan, C.-h., and Wei, K. K. (2013). Helpfulness of online product reviews as seen by consumers : source and content features. *International journal of electronic commerce : IJEC*, 17(4):101–136.
- Li, Z., Zhang, M., Ma, S., Zhou, B., and Sun, Y. (2009). *Information Retrieval Technology: 5th Asia Information Retrieval Symposium, AIRS 2009, Sapporo, Japan, October 21-23, 2009. Proceedings*, chapter Automatic Extraction for Product Feature Words from Comments on the Web, pages 112–123. Springer Berlin Heidelberg.
- Liu, B. (2010). *Handbook of natural language processing*, chapter Sentiment analysis and subjectivity. Chapman and Hall/CRC.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan and Claypool.
- Liu, H. and Singh, P. (2004). Conceptnet — a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4):211–226.
- Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of EMNLP 2004*, pages 412–418.
- Santos, R. L. S., Rabelo, R. A. L., and Moura, R. S. (2016). An experimental study based on fuzzy systems and artificial neural networks to estimate the importance of reviews about product and services. In *2016 International Joint Conference on Neural Networks (IJCNN 2016)*, page ”to appear”.
- Sarlan, A., Nadam, C., and Basri, S. (2014). Twitter sentiment analysis. pages 212–216.
- Seerat, B. and Azam, F. (2012). Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications*, 49(9):42–51.
- Sharma, A. and Dey, S. (2012a). An artificial neural network based approach for sentiment analysis of opinionated text. In *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, pages 37–42.
- Sharma, A. and Dey, S. (2012b). A document-level sentiment analysis approach using artificial neural network and sentiment lexicons. *SIGAPP Appl. Comput. Rev.*, 12(4):67–75.
- Silva, M. J., Carvalho, P., and Sarmiento, L. (2012). Building a sentiment lexicon for social judgement mining. In *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*, pages 218–228.

- Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., and Li, P. (2011). User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1397–1405.
- Torruella, J. and Capsada, R. (2013). Lexical statistics and tipological structures: A measure of lexical richness. *Procedia - Social and Behavioral Sciences*, 95:447 – 454.
- Vinodhini, G. and Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(6).
- Yannis, P. and Evanthia, K. (2012). An ontology-based question answering system exploiting search engines' results. pages 426–429.

Formal Information Flows Among Top Authorities of the Brazilian Federal Government based on Co-word Analysis of Data Published in the Official Gazette

Rafael Henrique Santos Soares ¹, Jorge H. C. Fernandes¹ Ricardo Sampaio¹

¹Grupo de Pesquisas sobre Inteligência Cooperativa em Redes Sociais Complexas – Programa de Pós-Graduação em Ciência da Informação – Faculdade de Ciência da Informação - Universidade de Brasília (UnB)
Campus Universitário Darcy Ribeiro, CEP: 70.910-900, Brasília- DF, Brazil
rafaelhss@gmail.com, {jhcf|rbsam}@unb.br

Abstract. *This paper describes a methodology for modeling social networks of formal communications among top authorities of the Brazilian Federal Government grounded on data available in the government official gazette (diário oficial da união). The text of a large number of official publications such as presidential decrees, ministerial orders and authority nominations was analyzed for identification of citations to organizations and persons. The co-occurrence of names of persons in such publications created a network of relations among such persons. An ego-network was built around the president Dilma Rousseff. Metrics of social network analysis were collected and analyzed in an exploratory fashion.*

1. Introduction

Information flows every day in public and private agencies, either between persons either between organizations. Communication is a natural and necessary process that supports collaboration and coordination of work. Public organizations must behave in accordance with the public interest and during the last decade Brazil's federal government has strongly evolved in publishing data about high level government decisions using portals over the web and thus enhancing its transparency.

Under a political point of view, information communication defines and is defined by the interaction among personal agents in fluid networks that exhibit heterogeneous and emergent properties. People talk through a lot means, trying to get necessary information to accomplish something or to convince others that something must be accomplished.

Federal agencies at the top of the formal hierarchy, such as ministries, are public institutions whose agents naturally develop heavy political bias, but they also must act in order to implement and evaluate public policies. This way, understanding the information network formed in a political scenario of a public organization is important to understand the dynamics of power and influence of public agents in the development of such policies.

This paper postulates that by focusing on the networks established by the persons that are top government authorities makes it possible to infer on the

collaborations and opportunities that are explored by these persons, based on their structural position on the network.

However, it is hard to map this type of network given that data about communication between top authorities is often not readily available and interviewing such persons on political questions is hard and expensive to do, even with a moderated number of respondents.

This paper explores the use of data available on web sites to map networks of communications among public agents in the Brazilian Federal Administration. We collect text from D.O.U. (“Diário Oficial da União” – Federal Official Gazette) to build social networks of public agents and. Through the analysis of this network of formal communications we aim to understand how these agents interact. With the data collected we have several options of analysis, but in this case we focused on networks established around the president of republic, and in how this interaction influences other opportunities to mediate information. We also analyze the dynamics of information flow and the effect of a structural position in nominations on the government.

2. Information communication between public agents and organizations and its effects in official media

Acts published at official media such as the D.O.U. (“Diário Oficial da União” – Federal Official Gazette) are manifestation of communication between public agents and organizations. In that sense, we explore Le Coadic’s (2004, p.35) concepts of formal and informal communications, Costa’s (2013) idea of a gradual transition from informal to formal and the Spiral of Communication Model proposed by Dance (1967, p.374). Based on these concepts, a model of information communication between public organizations and agents and its manifestation in official media like D.O.U. is presented.

When discussing information communication, Le Coadic (2004, p.35) argues that the communication process can be divided into two aspects: a written/formal and an oral/informal. The author asserts that the formal process produces stable information while the informal one delivers information more subject to changes and repetition. The author also states that the informal processes precede the formal ones.

An analogy to Le Coadic’s (2004, p.35) theory can be made to understand the information communication process between public agents and organizations which also has a formal and an informal component. Decisions are made internally after conjectural analysis, discussions, studies among other supplementary processes that, depending on the matter, include public agents outside the organization. Some published decisions about collaboration deals, working groups creation or project implementations are made after months or years of negotiation and informational exchange between the public organizations involved.

Table 1 shows an adapted version of Le Coadic’s (204, p.37) scheme comparing elements of formal and informal information communication. In this table, we included two columns presenting examples of each element stated by the author in the context of the information communication between public agents that can be found in official publications like D.O.U.

Table 1 – Examples of elements found in D.O.U.

	Formal	Informal
Cited by Le Coadic	Public (big potential audience)	Private (restricted audience)
Example in D.O.U.	Published online	Phone calls, e-mails, meetings.
Cited by Le Coadic	Information stored permanently	Information not stored and not recoverable
Example in D.O.U.	Files and online repository with open access.	Sometimes some poor registries can be found, like minutes and reports, but are often well structured.
Cited by Le Coadic	Information relatively old	Fresh new information.
Example in D.O.U.	A publication at D.O.U. is a communication to society of something already done	The decision ins been built during the communication process
Cited by Le Coadic	Information proved truth. Can be trusted.	Not proved information.
Example in D.O.U.	Publicity is a requirement for an act to take effect. Therefore, it is truth.	When communicating informally, people often do not worry about accuracy.
Cited by Le Coadic	Uniform spread.	The flow of the information is chosen by the
Example in D.O.U.	Access to D.O.U is open to anyone.	Only those invited to interact have access to the
Cited by Le Coadic	Low redundancy.	Redundancy is often an important element.
Example in D.O.U.	One single publication is enough for an act to take effect.	Complex matter are discussed by various agents inside organizations with gradual involvement of upper hierarchy, requiring redundancy.
Cited by Le Coadic	No direct interaction.	Direct interaction by default.
Example in D.O.U.	Publications in D.O.U are one-way	Public agents interact while discussing.

Source: Adaptation of Le Coadic’s (2004, p.36).

Costa (2013) made a graphic view of the organizational communication channels’ classification. Figure 01 synthetizes this scenario and allows us to identify a hierarchical relation between types of channels.

Analyzing figure 01 we can see that there is no polarity between formal and informal, but, a gradient that begins at one side completely formal and ends in the other, completely informal. In the middle lays the communication channels distributed gradually in this ‘rank’ according to its ‘formality level’.

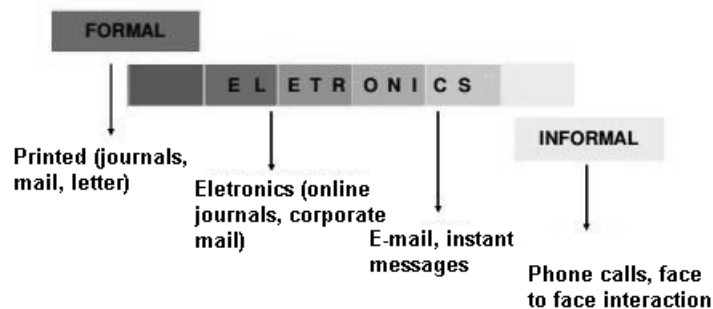


Figure 01 – Classification of communication channels.
Source: Costa (2013), translated

Dance (1967, p.378) argues that communication must be understood as a continuous and incremental process. He criticizes the linear models (for not represent feedback) and the circular ones (for not represent the effect of information exchange in individuals), and suggests a spiral model. The author believes that we should not use lines or circles but a helix (or a spiral) instead as the basic shape for represent human communication. The spiral, he says, is better for it shows the direction of the

communication flow, like the line, the feedback, like the circle, and the incremental impact that communication has on itself, making it bigger and aggregating more information every iteration.

McQuail and Windahl (1993) emphasize the dynamic aspect of Dance's model comparing to the "frozen picture" of the process provided by traditional communication models. The authors point out, however, that Dance (1967) did not provide a detailed tool or model for deep analysis but, instead, identified a core characteristic of the communication process that could be easily neglected because was absent in existing models that time.

With this background, the communication and mediation of information between public agents and organizations can be studied through an analysis of the official acts published in official communications like D.O.U. The model presented in figure 02 aims to show how the communication of information between public agents happen in an informal dimension, how this communication evolves until it breaks through the edge between the informal and formal and how this interaction finally is manifested in an publication at official media.

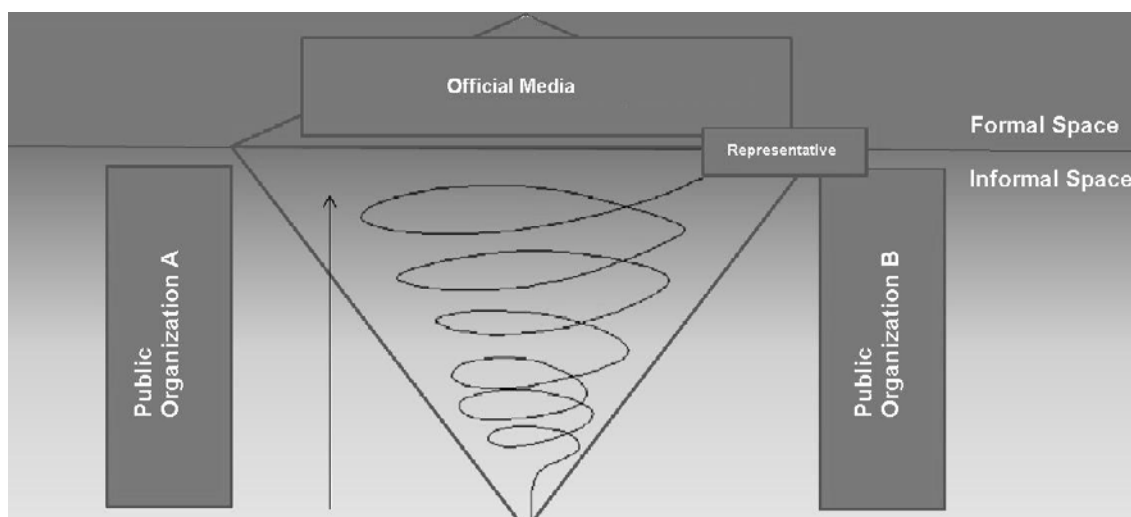


Figure 02 – public agents and organization information communication model

There is a mediation role played by the organization that prepares and publishes the decision. The elaboration process of the text to be sent to official media makes this player the representative of the group, who has the power of decide what is communicated and how.

The diamond shape in the back of the image represents this mediation phenomenon. The analogy to an iceberg intends to illustrate the fact that there are more information generated and a more intense communication flow going on before the decision in published in official media. And this publication only shows one piece of what really happened before. This understanding allows us to expand the analysis and infer that the evidence identified in official media represents a much more intense and significant interaction that precedes this evidence.

A pseudo-algorithm for building such network would be:

- For each page of the gazette:
 - Identify publication boundaries;
 - Identify person names inside the text.
 - For each pair of person found, establish a relationship.

Data was collected from the D.O.U (“Diário Oficial da União” – Federal Official Gazette), available online at <http://portal.imprensa.nacional.gov.br/>, from 2012-january-01 to 2013-April-01, using an automated software solution developed for this task. This software downloads data, finds names using heuristics and produces files representing the networks. The software was developed using the methodology described by Soares (2014) and authors can promptly provide collected data upon request.

The following table shows data about the resulting network.

Table 02 – Numbers about the network built according to the methodology

Period analyzed	2012-january-01 to 2013-April-01
Documents analyzed	About 2100 files. (2.5 GB of text data)
D.O.U Sections	All 3: Acts, Human Resources and Contracts
Full Network size	Vertices: 221392, Edges: 11699909

3.2 Ego Networks

Ego networks are especially useful to address the position of an individual among other actors and his opportunities to exercise power raised from his structural location in the network and to measure his social capital provided by this position and the social support one can receive and offer in each network setup.

The hypothesis investigated in this paper is strongly related to power and hierarchy. Ego networks are a very useful tool to study hierarchical phenomena and the opportunities actors have to “escape” from formal restrictions in influence and information mediation.

In that sense, Cross and Cummings (2003) offer a very interesting point of view of the use of ego networks in studies in individual performance:

Research assessing individual (or egocentric) networks has established links between performance and structural holes, defined in terms of lack of connectivity among people in one’s network. People with networks rich in structural holes are more likely to be promoted early, enjoy greater career mobility and adapt to changing environments more successfully.

4. Analysis of the information flow and mediation in President Dilma Rousseff’s ego network

Mediation in Social Network Analysis can be seen as a phenomenon that makes “a person with a lot of friends and contacts have much better chances to get help or information. So, social ties are a measure of social capital, an asset to be used for

advantage” (NOOY, MVAR e BATAGELJ, 2005, p.138). Social place in the structure of a network can allow one to mediate communication between others, which is an opportunity to play relevant roles in the network leading, eventually, to benefits just because of the structural position.

In this section, we show how information mediation in a network of public agents including President Rousseff can be studied to identify prominent agents and patterns on communication.

4.1. Network definition and manipulation

We studied the information flow in ego networks using a graph mapped through the identification of all person names in the D.O.U. Each full name, identified using the methodology described above, was considered a node in the network. Relationships between these nodes were discarded if their number of occurrence were less than eight, which is the mean of the values of lines with more than 1 occurrence. This decision was enough to discard sporadic connections. Examples of such connections are listings of granted retirements or projects receiving government funds. In situations like these the co-citation of names is circumstantial, do not fit the “public agents and organization information communication model” presented in this paper and is unlikely to be found again, being, therefore, disposable.

Table 02 below shows the distribution of the connections by their number of occurrence in the network of agents whose name was found by the text processing method applied. It is clear that the line disposal policy suggested is not only feasible, but necessary to make the dataset useful, since the initial amount of connections identified, mostly with only one occurrence, is too high.

The agent chosen as central for the analysis of the network of agents participating in public Administration of Brazil was President Dilma Rousseff. The hypothesis is that around such an important actor would be raised a network where the communication flow would be intense enough to create evidence in the D.O.U. in the way described by the “public agents and organization information communication model” presented before.

Table 02 - Distribution of the connections by their number of occurrence

Line Values		Frequency	Freq%	CumFreq	CumFreq%
Lowest value of line: 1.00000000					
Highest value of line: 124.00000000					
(... 1.0000]	9831124	93.7047	9831124	93.7047
(1.0000 ... 42.0000]	660173	6.2924	10491297	99.9971
(42.0000 ... 83.0000]	279	0.0027	10491576	99.9998
(83.0000 ... 124.0000]	23	0.0002	10491599	100.0000
Total		10491599	100.0000		

So, the network, after removal of the low value lines, was reduced to a network containing only the nodes connected to Rousseff (the ego) directly or through one other actor. In other words, we created a network with the *2-neighbors* of the ego and the connections between this subset of actors: The ego (gray node in the middle), her first

level neighbors (white nodes) and her second level neighbors (black), as seen in figure 04 below.

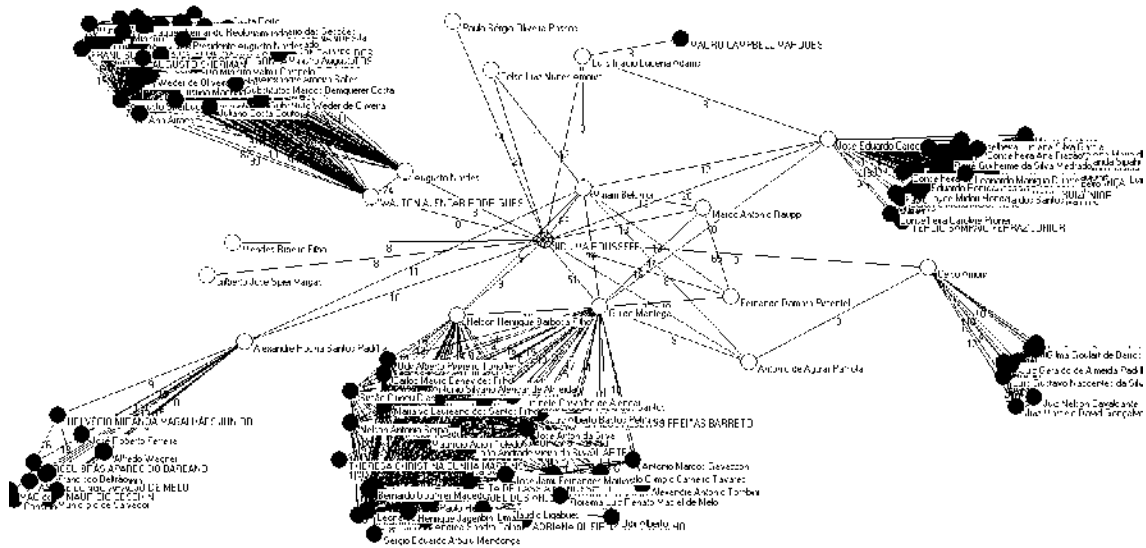


Figure 04 - President Dilma Rousseff's ego 2-neighbor network

4.2. Structural metrics and visual inspection

Visual inspection allows us to see a clear segregation between agents on the second level (blacks), who, structurally, are connected, at most, to two level one (white) individuals. There are five groups of agents on the second level, three of them connect to only one agent of level one and two of them connect to two agents of level one. There is a sixth “group” but it has only one member.

We can assume that these actors between the ego and the second level act as mediators of the interaction of those with the President. It comes from the way this view was built. A second level actor that eventually could have a closer access to the president would be shown as a first level one. We suppose this is what happened to “Nelson Henrique Barbosa Filho”, who is hierarchically below “Guido Mantega” but structurally at the same level.

Two agents in two groups played the mediator role. Nelson and Guido were Secretary and Minister of Economy respectively. Augusto Nardes and Walton Rodrigues are ministers of the Federal Court of Accounts (TCU – Tribunal de Contas da União). We can see in the network that they are ahead of the biggest groups of agents, where the information flow is probably bigger and demands more than one mediator. Yet, in the case of TCU it is interesting to see that Nardes and Rodrigues are still prominent for there are other seven ministers who are at the same top hierarchical level and do not seem to be as connected to the President as they are.

In this scenario, one interesting hypothesis to be cleared is that communication with the President is extremely restrict, since in the two level network only 8,45% (16 agents) have direct access to Rousseff. The fact that a second level is made by only 172 agents, a very small fraction of the 221387 agents in the original network reinforces this theory.

Finally, a structural aspect of this network should be highlighted: The clique formed by Mirian Belchior, Marco Antônio Raupp, Fernando Damata Pimentel, Guido Mantega e Dilma Rousseff. Cliques in social networks mean strong cohesion between the members what can be an indicator of a high level of cooperation, coordination and information exchange between these agents and the organizations they are ahead of. Identifying what causes this higher interaction requires further investigation. It can be either related to the mission of the organization, the personal affinity of these agents, party related issues or some other reason.

4.3. Information flow in Rousseff's neighborhood

Figure 05 shows a variation of the network shown in Figure 04 discussed in last section. To study dyadic restrictions, aggregated restrictions and structural holes in the ego network of President Rousseff, we removed all agents not connected directly to Dilma.

Generally speaking, dyadic restriction of a connection over an actor measures the opportunity this actor has to play a mediator role for others and the risk of having his communication to others mediated in case he loses this connection. As a rule of thumb, the bigger the restriction is the less mediation opportunities an actor has and more worried he should be about losing that connection.

This interpretation of the dyadic restriction metric, however, do not feels right to an ego network of a President. After all, since we used an open data resource to map these relationships, we assume the information flow between these agents is not secret, but the opposite: These are important information for cooperation and coordination of individuals involved, as discussed when the “public agents and organization information communication model” was shown in this paper.

What is expected, then, is more communication between agents in the Administration so the work would be done more efficiently, and from the President's point of view, there should be no interest in mediating information flow. In fact, big centralization of communication in the President's cabinet would represent an unnecessary overload, elevating risks of noise, delaying decisions and inflating bureaucracy.

So, the interpretation of this metric should be the inverse of what the literature often do, that is, from the President's point of view, a low restriction points out to a situation where the central power would be playing a mediator role which should be avoided. On the other hand, a high dyadic restriction mean low dependency of the other agents to the President, indicating a straight flow of information between them, situation that would lead to a better communication and should be pursued.

Table 03 shows values for dyadic restriction relative to the President and the lines between each agent and other neighbors. The positive correlation between these two metrics is expected and reinforces the notion that the more communication an agent has with others without the mediation of the President, bigger is the relative President's dyadic restriction, indicating more freedom for the agent in the network and less intervention of the central power in the information flow. We can see, for example, that Miriam Belchior has more connections with peers than any other, therefore, shows the biggest restriction to the ego.

Table 03 – Mediation metrics in the ego network

Agent	Dyadic restriction to President	Lines between other neighbors
Paulo Sergio Oliveira Passos	0,002222	0
Augusto Nardes	0,002626	2
Celso Luiz Nunes Amorim	0,006887	2
Marco Antonio Raupp	0,0147581	12
Miriam Belchior	0,1191908	32
Fernando Damata Pimentel	0,0177688	12
Guido Mantega	0,0873331	24
Alexandre Rocha Santos Padilha	0,0021115	2
Antonio de Aguiar Patriota	0,0138375	8
Celso Amorim	0,0017714	2
Jose Eduardo Cardoso	0,0096853	10
Gilberto Jose Spier Vargas	0,0007255	0
Luis Inacio Lucena Adams	0,0036674	6
Mendes Ribeiro Filho	0,0007255	0
Nelson Henrique Barbosa Filho	0,0029322	6
Walton Alencar Rodrigues	0,002626	2

These numbers can be aggregated in two indicators, to make the analysis process simpler. The aggregated restriction calculated for the President, calculated summing up all restrictions associated with her, was 0,2889 and the egocentric density, calculated as the density of the network when the ego is excluded, was 0.1484.

Aggregated restriction and egocentric density are hard to interpret alone. However, they are very useful to compare networks with similar size. So, in order to assess the impact of a more intense communication between the neighbors of the President in the network, these same metrics were calculated for a sub-network where the identified clique (Mantega, Belchior, Raupp, Pimentel and Rousseff) was removed, leading to a less interconnected network and, therefore, more dependent on the President. This new network showed 0.1501 for aggregated restriction and 0.0416 for egocentric density.

The removal of only four agents of the network dropped the aggregated restriction in 50% and the egocentric density in almost 70%. Such sensitivity in these metrics show that the values are small, because a very small change in the number of actors made a big change in the metrics. If these values were big, small perturbations in the network wouldn't lead to big impact on these metrics.

These calculations and comparisons of aggregated restriction and egocentric density allow us to believe that the communication of information between agents of public Administration in Brazil are strongly centered on President Rousseff, what may point out to difficulties implement direct articulation between these agents. A hypothesis raises then: An increment on the communication between the agents who deal directly with the President would lead to a situation where the information would flow more smoothly, fast, with less noise, and the President would have more resources to deal with other issues.

4.4. Structural proximity to President and nomination for high budget positions.

We analyzed another sub-network built from the original network after discarding every actor who did not connected directly to the President at least three times according to the methodology described. It is interesting to see that all the agents in Rousseff's 1-

neighbor ego network are public agents in first or second class positions, namely ministers and secretaries

There are several positions with different levels and responsibilities that the President can, freely, nominate or drop as she wishes. Over this 1-neighbor ego network we studied the structural proximity to the President and its relation to the budget of the ministry assigned to each element. This is a very unique characteristic of President's ego network and allow us to investigate if the closer they are to Rousseff, the more powerful they could be.

We assume that the bigger is the budget greater is the power and influence the one who controls it have. Thus, being nominated to head a ministry with big budget is something desirable from the individual point of view.

The hypothesis is that being structurally close to the President would lead to better opportunities to reach and keep these positions.

Table 04 shows the budget for each actor in the 1-neighbor network. Some of them have R\$0,00 budget because they aren't the top authority of the ministry and, thus, do not have full control of the budget. These information were manually extracted from government sources, specially the website *planalto.gov.br*.

Figure 05 shows this network drawn using the budget of the ministry as the size of the vertices. Using kamada-kawai energy algorithm, fixing Rouseff in the middle and using line values as similarities, we could produce a visualization that would place the most connected elements to the President closer the her correspondent node. In other words, the closer an element is of Rouseff, the greater is the number of connections it has with her, since more weigh means more communication.

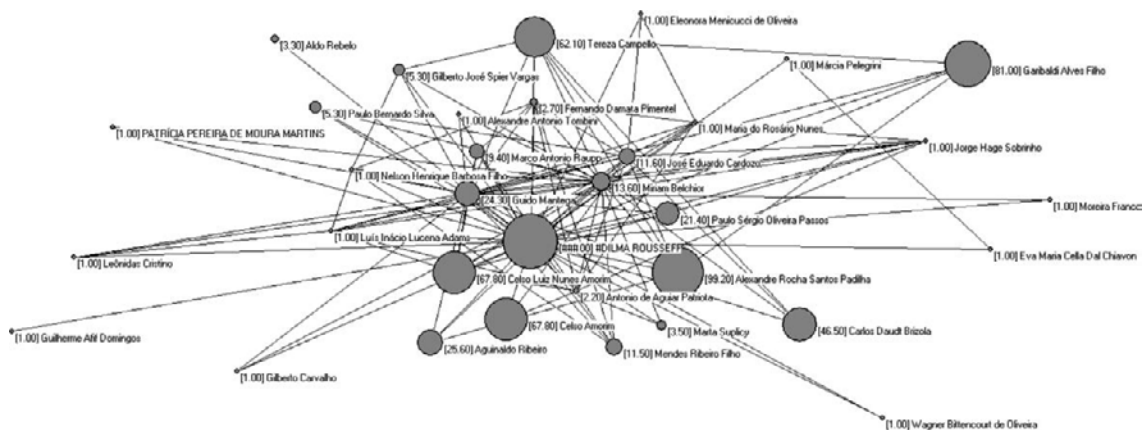


Figure 05 – President's 1-Neighbor network with proximity and time in position information

A visual inspection in this network leads us to believe that there really is a positive correlation between budget and number of connections with Dilma, because it seems that the bigger nodes are clustered together in the middle of the network, close to Rousseff, while the small ones are in the periphery of the net. Of course there are some exceptions in both groups.

Garibaldi Alves Filho, for instance, is represented by a big node far from Dilma. He was in charge of a ministry responsible for payments of retired workers in Brazil. So,

despite that his ministry has a huge budget, most of it is already designated. There is much less space for decisions that it seems.

To understand better the relation between the structural proximity to the President and the budget, we calculated the Pearson's correlation between these two numbers. It turned out that it was 0,6012, a statistically significant correlation that let us believe that, in fact, the elements structurally closer to the President are more likely to occupy more interesting positions in the Administration.

It is important to remember that such choices are political decisions, subject to lots of factors and influences and further additional research over this subject is essential to come to a stronger conclusion about the dynamics of power in politics.

5. Final comments

This paper makes an exploratory analysis of communications networks established around President Dilma Rousseff, based on data available at the federal official gazette. The structural metrics used were the agent's degree of relationships, dyadic restriction, aggregate restriction and egocentric density. Positive correlation coefficients indicate that the stronger the level of communication with the president, stronger is the tendency to being nominated for a position with large budget.

References

- COSTA, S. M. S. Comunicação para negócios: notas de aula da disciplina Fundamentos em Comunicação e Mediação da Informação. Programa de Pós-Graduação da Universidade de Brasília. 2013.
- COURTIAL, J.; LAW, J. A co-word study of artificial intelligence. *Social Studies of Science*, v. 19, n. 2, p. 301-311, maio 1989. Disponível em: <<http://www.jstor.org/stable/pdfplus/285145.pdf?acceptTC=true>>. Acesso em: 15 mar. 2016
- CROSS, R. CUMMINGS, J. N. Tie and Network Correlates of Individual Performance in Knowledge-Intensive Work. *Academy of Management Journal*. V. 47, n. 6, p. 928-937. December 2004
- DANCE, F. *Human communication theory: original essays*. Michigan: Holt, Rinehart and Winston, 1967.
- LE COADIC, Y. *A ciência da informação*. 2. ed. Brasília: Brinquet de Lemos, 2004.
- MCQUAIL, D.; WINDHAL, S. *Communication models for the study of mass communications*. Londres: Longman. 1993
- NOOY, W.; MRVAR, A.; BATAGELJ, V. *Exploratory network analysis with Pajek*. Cambridge: Cambridge University Press, 2005.
- SOARES, Rafael Henrique Santos. *Métodos para análise da comunicação e mediação da informação em organizações públicas por meio de redes sociais mapeadas a partir de publicações oficiais*. 2014. 155 f., il. Dissertação (Mestrado em Ciência da Informação) - Universidade de Brasília, Brasília, 2014.

How Congressmen Connect: Analyzing Voting and Donation Networks in the Brazilian Congress

Victor S. Bursztyn^{1,2}, Marcelo Granja Nunes¹, Daniel R. Figueiredo¹

¹Programa de Engenharia de Sistemas e Computação/COPPE
Universidade Federal do Rio de Janeiro (UFRJ) – Brazil

²EMC Brazil Research & Development Center
Parque Tecnológico, Ilha do Fundão – Rio de Janeiro, RJ – Brazil

{vbursztyn,mgranja}@cos.ufrj.br, daniel@land.ufrj.br

Abstract. *The goal of this work is to analyze two of the most central activities in the life of a congressman: raising funds and voting bills. We investigate the Brazilian Congress to shed light on the relationships between the donations received by congressmen elected in 2014 and their voting behaviors during the year of 2015. We merged publicly available data obtained from the Brazilian House of Representatives and the Superior Electoral Court (TSE) in order to create a tripartite network containing campaign donors, elected congressmen, and legal bills. Using this data, we create two projected networks having congressmen as nodes and links given as follows: 1) congressmen who received donations from the same donors (donation network); and 2) congressmen who voted in accordance to each other on legal bills (voting network). After characterizing these networks, we propose three fundamental questions on the behavior of congressmen that could benefit from the methods and concepts provided by Network Science. Finally, we analyze the results and compare them to general domain knowledge.*

1. Introduction

In November 2014, the influential Brazilian newspaper Folha de São Paulo [Folha de São Paulo 2014] reported that the top 10 corporate donors accounted for 70% of the funds raised by deputies elected earlier that year. Political scientists have long debated about possible negative aspects of electoral campaigns funded by large private companies. Notably, a study conducted by Datafolha suggests that the new House of Representatives' strong position in favor of those donations opposes to the position of the Senate and, more worth-noticing, to public opinion – 79% of the population was contrary in July 2015, according to Datafolha's polls [O Globo 2015].

Almost one year after the election, in September 2015, Brazil's Federal Supreme Court (STF) ruled that the Constitution does not allow corporations to donate to electoral campaigns, meaning that corporate actors are no longer allowed in upcoming elections [BBC 2015]. Nevertheless, the availability of donations data and the fact that they were legal for the most part of 2015 provide an opportunity – maybe a historical one – to shed light on the dynamics between congressmen funding and their behavior in Brazil's legislature.

According to David Lazer [Lazer 2011], the use of Network Science by political scientists is fairly recent, despite the long-lasting notion that power is an intrinsically relational phenomenon: it rises from the capacity that one actor has to affect other actors. Over the past two decades, a series of works focused on studying power and networks of legislative institutions in countries with pioneering transparency initiatives. Previous works have studied parliaments using public data such as bill co-sponsorship [Zhang et al. 2008][Fowler 2006], committee participation [Porter 2006], roll call voting [Dal Maso et al. 2014], social media [Peng et al. 2014], or voting advice applications [Etter et al. 2014]. Some of the goals of these works were to characterize congressmen communities, predict issue voting results, and verify ideological consistency over time. We observed an opportunity to apply similar methods to the national context, specifically on the Brazilian House of Representatives.

This work expects to deliver the following contributions: first, we created a unified dataset by integrating data regarding the Brazilian House of Representatives previously scattered across different sources. This resulted in a tripartite network with the following types of nodes: campaign donors, deputies elected in 2014, and legal bills voted during the year of 2015. Moreover, for each bill we mine the voting position of each congressman and for each donor we mine the set of congressmen to whom donations were given. Based on this data, we can measure the similarity between congressmen through two points of view: 1) congressmen who received donations from the same donors (donation network); and 2) congressmen who voted the same positions on legal bills (voting network). To the best of our knowledge, we could not find any work that has analyzed these two aspects together, nor a dataset connecting all three domains (donors, congressmen and votings). Therefore, we stress that the dataset itself might be a valuable contribution as it could enable other researchers to explore such relationships or to unveil other phenomena.

Secondly, we explore these relationships posing and addressing three fundamental questions, briefly put as follows:

1) Homophily: How co-partisanship (congressmen from a same political party) and co-regionality (congressmen from the same region) reflect on campaign donations and on voting behaviors?

2) Partisan cohesion: How cohesive are political parties in terms of voting?

3) Influence prediction: If we use a list of top influencing congressmen as ground truth (i.e., from a broadly acknowledged report), could we retrieve that same list by mining voting network's topological features?

Beyond data collection and classic network characterization, another key contribution of this work is the proposal of a specific network-based methodology (and metrics) to address the above questions. In particular, our approach to study homophily and cohesion can be applied to any network where nodes are subject to a similarity metric and belong to a category (in our case, a political party or a geographical location).

The remainder of this paper is organized as follows. In Section 2 we describe our dataset; in Section 3 we explain how donation and voting networks are created and their associated characteristics; in Section 4 we analyze higher level network features; and finally, in Section 5, we discuss our findings as well as future directions.

2. Data Sources

We considered two main sources of publicly available data: 1) the Superior Electoral Court (TSE), where all official donations are available for bulk download; and 2) the House of Representatives, where all voting records for bills are open for access through an API. A process was developed in order to fetch those sources and to merge them into our tripartite network. Our resulting network was persisted in a Neo4J database, as illustrated in Figure 1. The final database records 52245 campaign donors, 456 elected congressmen and 234 bills (legal bills voted between January and November 2015, right after the new Congress was established). Our consolidated dataset is made available at <http://github.com/vbursztyn/HowCongressmenConnect>, in various formats.

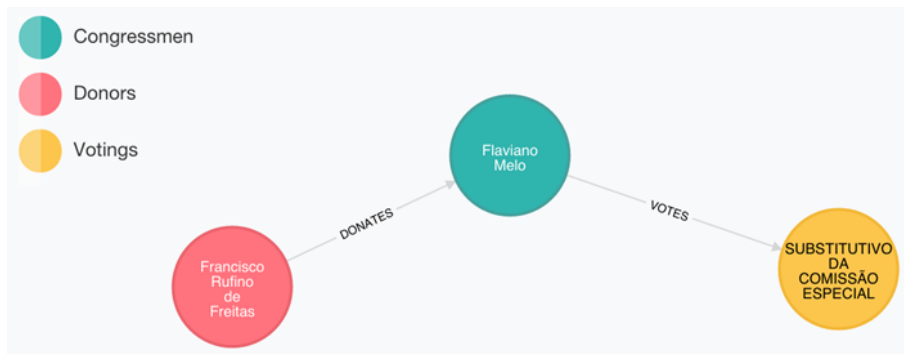


Figure 1. A single example illustrating all types of nodes and links

3. Donation and Voting Networks

From the tripartite network, we created two projected networks having congressmen as nodes and links as follows: 1) congressmen who received donations from the same donors (donation network); and 2) congressmen who voted the same positions on legal bills (voting network). In both networks, nodes are deputies and links are weighted similarities with values between zero (dissimilar) and one (identical). The metric used for similarity is the Jaccard Index, which is among the most frequent metrics in network applications. It is defined as the ratio between the cardinality of the intersection set and the cardinality of the union set. In our context, each congressman has a set of donors or a set of bills on which he has voted. Thus, for every pair of congressmen with a nonempty intersection of voting positions we have a similarity score in the voting network, and for every pair of congressmen with a nonempty intersection of donors we have a similarity score in the donation network.

Still, there are additional details on how this definition applies to each network. For the donation network, we only considered the presence of donors, disregarding the value of the donation. Although this application does flatten some information, it also prevents unwanted dissimilarity between candidates with a similar portfolio of donors but vastly different donations amounts. In essence, donors manage limited resources – regardless of their scales –, and the fact that a campaign attracted a person or a lobbyist to the point of making a donation is fairly significant.

In the voting network, we defined similarity considering the roll call votes made by elected congressmen on legal projects in 2015's agenda. Considering a pair of congressmen, the union set would be every bill they voted, and the intersection set would be

the bills in which their votes converged (options are: yes, no, abstention or obstruction). A special case to be mentioned is the speaker of the House of Representatives, who does not vote and therefore was excluded from this network.

The main features of these two networks are summarized in Table 1 below, where link weights have not been considered (and thus all links with non-zero weight are present in the network). Also, the voting attendance rate for deputies is shown on Figure 2.

Table 1. Characterization of the two projected networks

	Donation Network	Voting Network
Nodes count	456	456
Link count	60492	182256
Maximum degree	323	454
Minimum degree	0	0
Average degree	132.66	399.68
Density	0.29	0.88
Number of components	20	1
Relative size of the largest component	96%	100%
Global clustering index	0.69	0.95
Average local clustering index	0.68	0.96
Average distance	1.77	1.12
Pseudodiameter	5.0	3.0

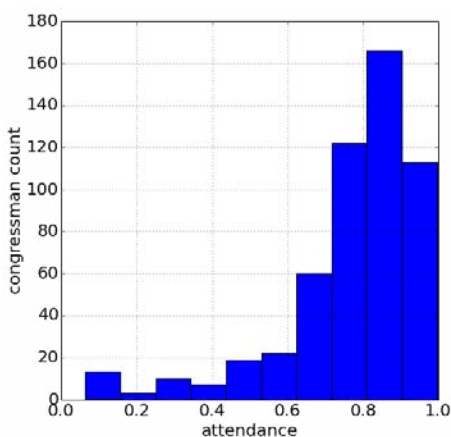


Figure 2. Voting attendance rate for deputies

Within those features, it is worth highlighting the average degree observed in both networks. In particular, the high connectivity in the voting network suggests that it is very unlikely for a congressman not to find a minimum degree of consensus with his peers, in a year’s agenda. This happens because congressmen are intrinsically constrained: they must declare a position if participating on a voting section, which in turn is limited to four options (yes, no, abstention or obstruction). All alone, the fact that there are more donors than bills could make more likely that a random pair of congressmen shares a donor. However, voting in a limited range of options imposes a very low probability for not sharing at least one single position throughout the year (for a random pair of congressmen:

$0.75^{234} \approx 6 * 10^{-30}$). Values observed for features such as the clustering coefficient and the average distance between nodes can also be explained from these observations.

Regarding voting attendance, Figure 2 shows that most deputies have a relatively high attendance, while a few have high levels of absence. The majority of these absent deputies are replacements, i.e. deputies who occupied the seat of another deputy. Since these deputies joined the Congress at a later date, they tend to have attended fewer votings.

4. Domain Specific Network Features

After a summarized view on the two projected networks, we propose and address the three fundamental questions in the following subsections.

4.1. Homophily – How co-partisanship and co-regionality reflect on campaign donation and on voting behaviors?

According to Anmol Madan et al. [Madan et al. 2011], homophily is a fundamental phenomenon in social networks, which refers to the tendency of individuals to form relationships with others that have similar attributes. Let's define homophily, in this particular case, by means of two attributes: congressmen who belong to the same party (i.e., co-partisanship), and congressmen elected from the same region (i.e., co-regionality). Once homophily is contextualized in our political setting, then we may propose our central questions: can we detect homophily in the donation network and in the voting network? More than uncovering such tendencies, it would be interesting to actually measure how strongly each attribute (co-partisanship and co-regionality) acts in both networks. In order to address this issue, we developed the following approach:

Consider that for each network each congressman holds an average similarity score defined by the average weight of its links (i.e., w_{avg}). Alternatively, for that same congressman, we may also calculate an average similarity selecting only the links that exhibit homophily: w_{avg}^p is the average score restricted to co-partisans, and w_{avg}^r restricted to co-regionality. Thus, for each node u , there are three averages based on the Jaccard Indexes: $w_{avg}(u)$, $w_{avg}^p(u)$ and $w_{avg}^r(u)$. Suppose our focus is on node u , then it is possible to calculate two similarity gains associated with our two homophily hypothesis:

$$\text{w.r.t. co-partisanship: } \text{gain}^{party}(u) = g^p(u) = (w_{avg}^p(u) - w_{avg}(u)) / (w_{avg}(u))$$

$$\text{w.r.t. co-regionality: } \text{gain}^{region}(u) = g^r(u) = (w_{avg}^r(u) - w_{avg}(u)) / (w_{avg}(u))$$

We can apply this similarity gain metric for every node in the network, for each network separately. Figure 3 shows the rank order of similarity gains in the donation network: green for co-partisanship, red for co-regionality. For instance, regarding co-regionality (red) congressman Mendonça Filho (DEM-PE) has a similarity gain of 1.43, Marco Antônio Cabral (PMDB-RJ) a gain of 1.13 and Tiririca (PR-SP) a negative gain of -0.04. It means that, on average, Mendonça Filho and Marco Antônio Cabral share more donors with congressmen from their respective states (AL and RJ) if compared to their respective average similarities to all peers. Conversely, Tiririca shares more donors with all of his peers, on average, than he shares with his peers from SP. This observation conforms to domain knowledge: Tiririca is not originally from SP and may have donors from other regions, that do not donate to other SP deputies.

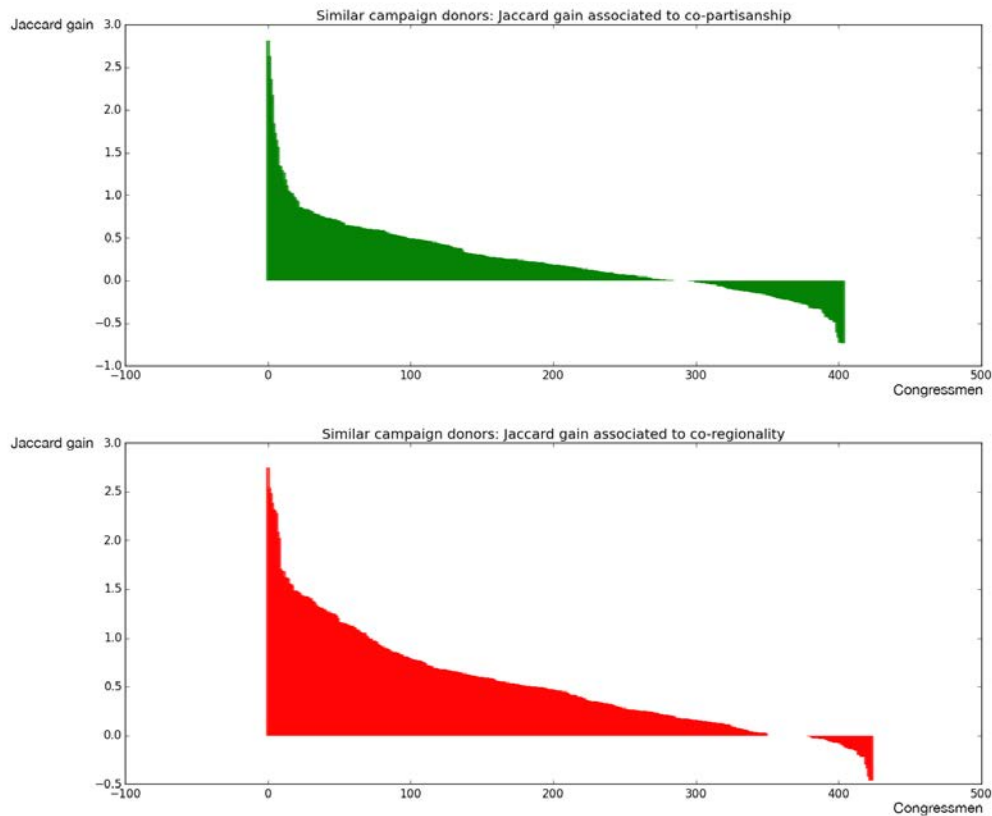


Figure 3. Similarity gains w.r.t. donations for deputies within the same political party (above) and based on the same federation unit (below)

The y-axis is the Jaccard Index gain (loss if negative) and the x-axis are the deputies. On average (across all nodes), the Jaccard Index gain in the donation network is 51% (0.51) if deputies represent the same state (co-regionality). Also on average, the Jaccard Index gain is 25% (0.25) if deputies belong to the same political party (co-partisanship). Therefore, in an election that allowed two sorts of donors (corporate donors and regular citizens), co-regionality was a stronger driver of donors similarity. If compared to co-partisanship, which also revealed to be a fairly relevant driver of donors similarity, co-regionality was twice stronger.

Similarly, Figure 4 shows the rank order of similarity gain in the voting network: green for co-partisanship, red for co-regionality. For instance, regarding co-partisanship (green) congressman Chico Alencar (PSOL-RJ) has a similarity gain of 1.94, Mara Gabrielli (PSDB-SP) a gain of 1.01 and Alessandro Molon (PT-RJ) a gain of 0.50. It means that, on average, their voting behaviors are more similar to the behaviors of their respective co-partisans, if compared to their average similarities to all peers. This also means that such identity is stronger for Chico Alencar regarding PSOL, than to Mara Gabrielli regarding PSDB. However, her identity with respect to her party's average behavior is stronger if compared to Alessandro Molon's regarding his party, PT.

Oppositely to what was noted in the donation network, Jaccard Index gain in the voting network is much stronger for co-partisanship if compared to co-regionality. On average, the Jaccard Index gain in the voting network is 46% (0.46) if deputies belong to

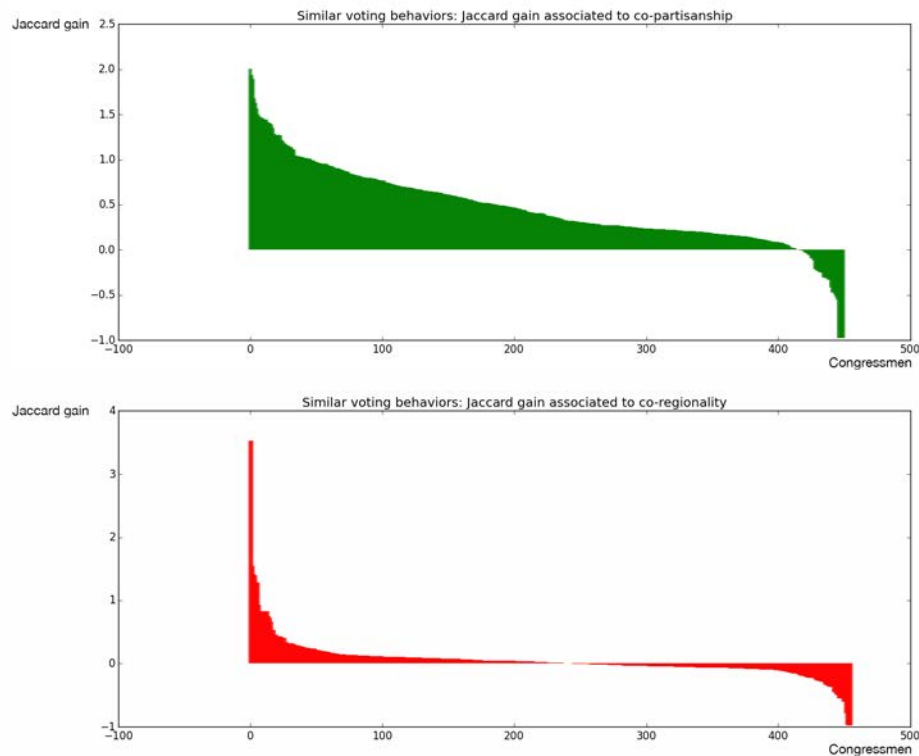


Figure 4. Similarity gains w.r.t. votings for deputies within the same political party (above) and based on the same federation unit (below)

a same political party, versus 5% (0.05) if they represent the same state (co-regionality). On these grounds, one may generally conclude that political parties are almost 10 times more effective in channeling voting behavior, if compared to pure locality. This gives a good measure of the role of political parties in today’s legislative – in a general and comparative view, parties do own expressive voting identities, in particular, much stronger than regionality.

In short, we may conclude that co-regionality drives donations more than co-partisanship, while co-partisanship drives voting behavior much more than co-regionality.

4.2. Partisan cohesion – How cohesive are political parties?

We can use the Jaccard Index gain value to identify which congressmen are more dissimilar with respect to his co-partisans. Such analysis indicates that the following deputies not only have greater tendency to vote unlike their co-partisans, but also are more similar to congressmen outside their respective parties: Guilherme Campos (PSD-SP), Giovanni Queiroz (PDT-PA), João Ananias (PC do B-CE), Betinho Rosado (PP-RN), Roberto de Lucena (PV-SP), Silvio Costa (PSC-PE), Danilo Cabral (PSB-PE), Zoinho (PR-RJ), Davi Alves Silva Júnior (PR-MA), Sandra Rosado (PSB-RN). From the voting pattern point of view, these congressmen are not aligned with their respective parties.

A second method that we propose to address partisan cohesion is as follows. For each political party, hereby represented by C , we calculate its intraparty average similarity $d_{int}(C)$: the average similarity of the links among party members; the interparty average similarity $d_{ext}(C)$: the average similarity of links between members of party C

and deputies who are not a member of C ; and the overall average similarity across all network links. These definitions have the following equations:

$$d_{ext}(C) = \frac{\sum_{i \in C, j \notin C} w_{ij}}{n_c(n - n_c)} \quad d_{int}(C) = \frac{\sum_{ij \in C} w_{ij}}{n_c(n_c - 1)/2} \quad d(C) = \frac{\sum_{ij} w_{ij}}{n(n - 1)/2}$$

Where n is the number of deputies and n_c is the number of deputies belonging to political party C . In theory, cohesive parties should have $d_{int} > d > d_{ext}$, and the wider the gap between d_{int} and d_{ext} , the more cohesive is a political party. Finally, we limited this analysis to parties with more than three congressmen.

The values measured for the averages are shown in Figure 5. In order to support the evaluation of our results, Figure 6 shows the number of deputies in each party. The ratio between intraparty similarity and overall average similarity suggests that some political parties are indeed much more cohesive than others. On the other hand, all parties have the interparty similarity lower than the overall average similarity, indicating all parties have some cohesion in comparison to average behavior. The most cohesive party is PSOL, which has the highest d_{int}/d ratio and the second lowest d/d_{ext} . Furthermore, contrary to anecdotal evidence, PMDB does have a mild internal cohesion. Also contrary to common knowledge, PV's intraparty similarity is about the same as the average similarity, indicating low cohesion in the voting behavior despite having only 8 deputies. Finally, we stress the fact that the most and least cohesive parties are small ones, but we also find parties with larger sizes with both high and low levels of cohesion, such as PSDB and PT (high cohesion) and PP (low cohesion).

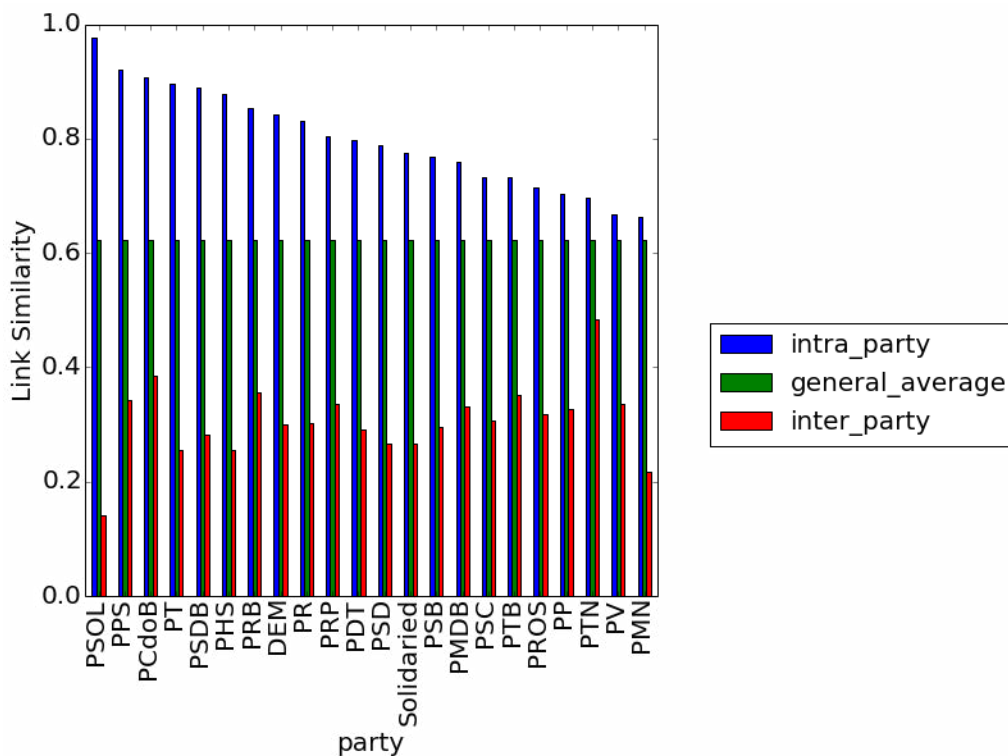


Figure 5. Average link similarities for each political party with more than three congressmen

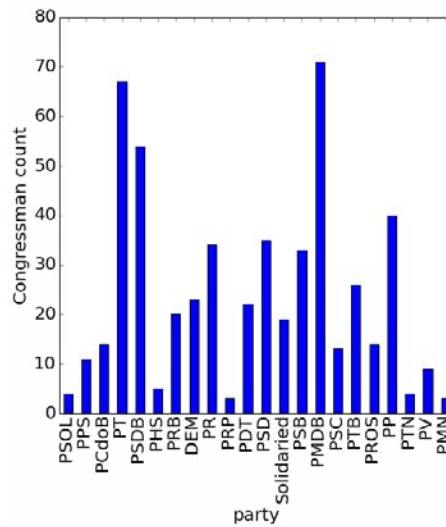


Figure 6. Number of deputies in each party with more than three congressmen

4.3. Influence prediction – If we use a list of top influencing congressmen as ground truth, could we retrieve that same list by mining voting network’s topological features?

The third question is based on the hypothesis that topological characteristics of a network can often be used to identify the most influential nodes, in our case members in the House of Representatives. To address this question we use the yearly report “Heads of Congress”, published by the Inter-Union Department for Parliamentary Advisory (DIAP) [Departamento Intersindical de Assessoria Parlamentar 2014], [Departamento Intersindical de Assessoria Parlamentar 2015], and considered an official source for the list of top influencing deputies. DIAP’s reports are often referred by major Brazilian newspapers when conveying political analysis, and thus it is a reasonable source for ground truth.

We consider node centrality as the topological feature used to identify top influential nodes. Nodes are sorted based on the following centrality metrics: Strength, Betweenness, Closeness, Katz, and PageRank. These metrics were chosen not only because they are the most common ones, but mainly because they capture different types of centrality. The Strength is defined as the sum of weights on links connected to a node. Both the Betweenness and the Closeness were normalized. The PageRank is used with a 0.85 damping coefficient and the Katz has $\alpha = 0.1$ and $\beta = 1$. Finally, all the metrics were calculated considering the Jaccard Index for the voting network as weights.

An important modification was made on the network. The link weight definition previously used in this paper was a similarity index, i.e. the higher its value, the more similar are the nodes connected by that link. However, the use of those centrality methods require link weights to be defined as a dissimilarity index, i.e. the lower its value, the more similar are the nodes. Therefore, we calculated this new link weight as being the complement of the old link weight, i.e. $weight_{dissimilarity} = 1 - weight_{similarity}$.

Afterward, we consider the top 55 members (about 10% of the House’s population) in the centrality ranks, resulting in five separate lists eligible to being tested against

DIAP’s Heads of Congress. The quality of each centrality metric is evaluated by its precision in identifying 55 names from the DIAP’s report. It is worth noticing that DIAP’s original report contains 100 congressmen, but only 55 are deputies. Also, we abstained from using deeper information contained in the report, such as the type of influence exerted by each congressman.

The distributions for four centrality metrics in the voting network may be found in Figure 7. As can be seen, the Strength distribution is similar to PageRank’s. This result was not unexpected as it is widely reported in the literature that, for undirected networks, the PageRank for nodes are proportional to their degrees [Grolmusz 2012]. More interestingly, the distribution for Betweenness was highly concentrated on a few nodes whose Betweenness value was zero. These nodes are deputies with the highest levels of abstention. A possible direction for further improvements could be to recreate the voting network with a minimum threshold for attendance, therefore decreasing this disproportional concentration of shortest paths on a few nodes.

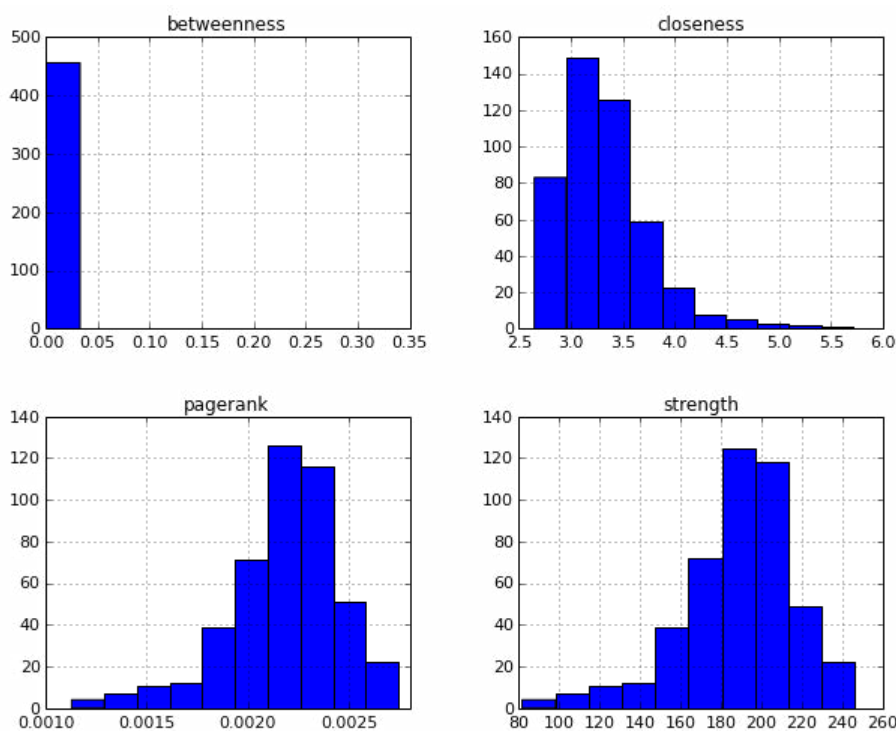


Figure 7. Histogram of centrality metrics for the nodes on the voting network

Finally, Table 2 indicates the precision of the five ranks, highlighting the best approach (underlined). The best performance is assigned to the top PageRank (and Strength) centrality, which manages to identify 27% of DIAP’s list. Such result surpasses a naive benchmark by 2.5 times, as random selection would only identify 11% on average. Nevertheless, this performance still is relatively low. A possible explanation is the high density of the network causing nodes to become similar while it should not be the case. Further work must be done in order to verify this hypothesis.

Table 2. Performance of centrality ranks w.r.t. precision@55 on DIAP's list

Metric	Precision@55
Random	0.11
Strength	<u>0.27</u>
Betweenness	0.18
Closeness	0.13
Katz	0.07
PageRank	<u>0.27</u>

5. Conclusions and Future Work

In this paper, we consolidate data from different public sources concerning donations to and votes by Brazilian congressmen. Donations refer to the election of 2014 and votes refer to legal bills in an 11 month agenda: from January 2015 to November 2015. Using this data we construct and analyze two networks where nodes are congressmen: the donation and voting networks. These relationships are fundamental to the political process, making the study of both networks a relevant subject for research in Network Science.

Furthermore, this paper finds evidence of homophily regarding co-partisanship and co-regionality in both networks. However, homophily patterns differ greatly between these networks, revealing the richness of the phenomena comprised in the original tripartite network. It is worth highlighting that we find quantitative evidence on the importance of parties in voting behavior, being almost 10 times stronger than co-regionality. Moreover, using the voting network we measured the degree of party cohesion and polarization with respect to the rest of the House of Representatives, with results that partially match our domain intuition, but also reveal new insights (e.g., the least cohesive parties). Finally, recognizing the most influential deputies using these networks and an external ground truth proved to be a difficult task that requires further study. A few promising directions for future work would be to test other similarity metrics besides the Jaccard Index, or to extend the dataset with other information such as congressmen participation in special commissions.

Also, due to the Federal Supreme Court's ruling, the landscape of donors is expected to change dramatically in the 2018 elections. That said, it is appealing to compare the networks uncovered by this work, to the ones that are going to arise after the prohibition of corporate donors. For example, homophily for co-regionality could increase significantly since corporate donors are more likely to be transversal to states than regular citizens.

References

- Dal Maso, C., Pompa, G., Puliga, M., Riotta, G., and Chessa, A. (2014). Voting behavior, coalitions and government strength through a complex network analysis. *PLoS one*, 9(12):e116046.
- Departamento Intersindical de Assessoria Parlamentar (2014). Radiografia do Novo Congresso: Legislatura 2015-2019. Technical report, Departamento Intersindical de Assessoria Parlamentar, Brasilia.

- Departamento Intersindical de Assessoria Parlamentar (2015). Os Cabeças do Congresso Nacional : uma pesquisa sobre os 100 parlamentares mais influente. Technical report, Departamento Intersindical de Assessoria Parlamentar, Brasília.
- Etter, V., Herzen, J., Grossglauser, M., and Thiran, P. (2014). Mining democracy. In *Proceedings of the second ACM conference on Online social networks*, pages 1–12. ACM.
- Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(4):456–487.
- Grolmusz, V. (2012). A note on the pagerank of undirected graphs. *CoRR*, abs/1205.1960.
- Lazer, D. (2011). Networks in political science: Back to the future. *PS: Political Science & Politics*, 44(01):61–68.
- Madan, A., Farrahi, K., Gatica-Perez, D., and Pentland, A. S. (2011). Pervasive sensing to model political opinions in face-to-face networks. In *Pervasive Computing*, pages 214–231. Springer.
- Peng, T.-Q., Liu, M., Wu, Y., and Liu, S. (2014). Follower-follower network, communication networks, and vote agreement of the us members of congress. *Communication Research*, page 0093650214559601.
- Porter, M. (2006). A network analysis of committees in the united states house of representatives. In *APS Meeting Abstracts*, volume 1, page 33006.
- BBC (2015). STF decide que doação de empresas a campanhas é inconstitucional; veja em que países isso já é proibido.
- Folha de São Paulo (2014). As 10 empresas que mais doaram em 2014 ajudam a eleger 70% da Câmara.
- O Globo (2015). Nova ferramenta digital revela doadores das ultimas campanhas eleitorais.
- Zhang, Y., Friend, A., Traud, A. L., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2008). Community structure in congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications*, 387(7):1705–1712.

Predição de Relacionamentos em Redes Sociais, uma Revisão Sistemática

William Takahiro Maruyama¹, Luciano Antonio Digiampietri¹

¹Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP)
Av. Arlindo Béttio, Ermelino Matarazzo – 03828-000 – São Paulo – SP – Brasil

Abstract. *The social network analysis area is on the rise. An important task in this area is link prediction, in which the goal is to predict connections between users. For this task it is necessary the use of attributes, methods, algorithms and techniques that measure, somehow, the possibility of a relationship be created. However, there are many approaches and combinations of attributes for predicting links. This paper aims to conduct a comprehensive survey of the attributes/characteristics that can be used to predict links in various contexts of social networks, based on the Systematic Review methodology.*

Resumo. *A área de análise de redes sociais está em ascensão. Uma importante tarefa desta área é a predição de relacionamentos, na qual o objetivo é prever conexões entre usuários. Para a realização desta tarefa são utilizados atributos, métodos, algoritmos e técnicas que medem, de alguma forma, a possibilidade de um relacionamento ser criado. No entanto, existem muitas abordagens e combinações de atributos para prever relacionamentos. Este trabalho tem como objetivo realizar um levantamento abrangente dos atributos ou características que podem ser utilizados na predição de relacionamentos nos diversos contextos das redes sociais, a partir da metodologia de Revisão Sistemática.*

1. Introdução

Cada vez mais presente no dia-a-dia das pessoas, as redes sociais online (tais como Facebook, LinkedIn, Google+, Twitter, Flickr, etc) são compostas por inúmeros indivíduos, os quais podem estabelecer entre si diversos tipos de interação ou relação (genericamente conhecidas como *links*). Em geral, as redes online crescem continuamente e expressivamente, além disso, muitas delas são enormes e esparsas. Adicionalmente, elas são naturalmente dinâmicas, pois a todo instante novas relações são estabelecidas ou desfeitas [Shin et al. 2012, Song et al. 2009, Song et al. 2012, Vasuki et al. 2011]. Neste contexto, este trabalho objetiva o levantamento das referências que estudam a dinâmica dos *links* em uma rede social, focando nos trabalhos que visam a prever novos relacionamentos que são mais prováveis de ser formados. Essa tarefa, da área Análise de Redes Sociais, é conhecida como Predição de Relacionamentos ou de Ligações (*Link Prediction*) [Liben-Nowell and Kleinberg 2003, Hasan et al. 2006, Murata and Moriyasu 2008, Hasan and Zaki 2011, Lu and Zhou 2011, Dhote et al. 2013].

A predição de novos relacionamentos dentro de uma rede social é uma tarefa que ganhou bastante destaque nos últimos anos e serve para, desde encontrar amigos que ainda não estavam ligados em numa rede social *online* [Vasuki et al. 2010, Tian et al. 2010, Perez et al. 2012, Fire et al. 2011,

Zhong et al. 2013, Quercia and Capra 2009], até para potencializar a realização de trabalhos em empresas ou na academia [Maruyama and Digiampietri 2016, Digiampietri et al. 2015, Digiampietri and Maruyama 2014, Digiampietri et al. 2013, Hsieh et al. 2013, Dong et al. 2012, de Sa and Prudencio 2011].

Para realizar a predição utiliza-se de métodos que de alguma maneira medem a proximidade ou similaridade entre as entidades (nós) da rede. Esses métodos fornecem medidas que podem ser utilizadas por si só para prever, mas podem ser adotadas como atributo (ou característica) a serem utilizados por uma estratégia de predição supervisionada [de Sa and Prudencio 2011, da Silva Soares and Bastos Cavalcante Prudencio 2012]. A identificação dos melhores conjuntos de atributos relevantes, dentre as combinações possíveis, é de muita importância para a melhoria de precisão dos modelos preditivos.

2. Metodologia e Condução

Com base em uma pesquisa exploratória realizada preliminarmente, foram identificadas *Link*, *co-authorship*, *prediction*, *social networking* e *scientific collaboration networking* como as principais palavras-chaves relacionadas ao assunto. Posteriormente, através da metodologia de revisão sistemática foi criado o protocolo que define e formaliza os procedimentos seguidos no presente estudo [Biolchini et al. 2005].

Este artigo tem como objetivo responder a seguinte pergunta: quais atributos existentes estão sendo utilizados na predição de coautorias em Redes Sociais? Para responder a esta questão foram feitas pesquisas nas principais bibliotecas digitais científicas da área, as quais disponibilizam os trabalhos via *web*. As bibliotecas digitais utilizadas foram: IEEEExplore Digital Library e ACM Digital Library.

Embora seja possível encontrar outros artigos sobre o assunto dispersados em diferentes revistas ou anais de eventos, optou-se pela consulta apenas a estas duas bibliotecas digitais por serem consideradas as que congregam um maior número de artigos na área.

Com base nas palavras-chaves selecionadas foram criadas e submetidas as expressões e opções de busca em cada uma das bibliotecas. Objetivou-se tanto encontrar artigos sobre predição de relacionamentos em qualquer tipo de rede social, assim como predição de relacionamentos em redes de colaboração científica. Para não restringir o resultado da busca, não foi considerado um período de publicação. A chave de busca submetida à biblioteca digital da ACM foi “((Abstract: “*Link*” OR Abstract: “*co-authorship*”) AND (Abstract: “*Prediction*”) AND (Abstract: “*social network*” OR Abstract: “*scientific collaboration network*”))”, além disso foi ativada a busca avançada, com utilização apenas do campo *abstract*. A chave de busca submetida à biblioteca digital IEEEExplore foi “ ((Abstract: “*Link*” OR Abstract: “*co-authorship*”) AND (Abstract: “*Prediction*”) AND (Abstract: “*social network*” OR Abstract: “*scientific collaboration network*”))”, além disso foi ativada a busca avançada, com filtro “Metadata only” ativo.

Todos os artigos encontrados na busca foram avaliados e selecionados segundo os critérios de inclusão e de exclusão que se seguem. Para aceitação do artigo, ele deve se enquadrar em todos os critérios de inclusão e nenhum de exclusão.

Critérios de inclusão: (i) Serão incluídos trabalhos completos publicados e disponíveis integralmente nas bases de dados científicas utilizadas. (ii) Serão incluídos trabalhos que analisem Redes Sociais. **Critérios de exclusão:** (i) Serão excluídos trabalhos de estudos secundários. (ii)

Serão excluídos trabalhos que não discutam os atributos que foram usados ou como foram usados para a predição de *links*. (iii) Serão excluídos trabalhos publicados que não estejam disponíveis integralmente nas bases de dados científicas especificadas.

Os resultados obtidos das duas etapas do protocolo da revisão sistemática estão organizados a seguir. O próximo parágrafo apresenta o resultado da triagem dos artigos (condução). Enquanto que a Seção 3 contém os resultados da sumarização das informações dos artigos incluídos na condução, análise das informações extraídas e a descrição de alguns artigos.

Com a submissão das expressões em cada uma das *engines* de busca das bibliotecas, foram encontrados inicialmente: 39 artigos na ACM e 37 artigos na IEEE. Desses artigos, ocorreu apenas um caso de repetição. Portanto, 75 artigos foram analisados. Uma seleção inicial foi realizada a partir dos critérios de inclusão e exclusão, aplicados sobre a leitura dos resumos (*abstracts*) de cada artigo. Nesta etapa, 12 artigos encontrados na ACM foram rejeitados e 14 na IEEE.

3. Extração

Os 49 artigos incluídos nesta revisão, foram lidos na íntegra e suas principais informações foram extraídas. Além dos dados bibliográficos, a Tabela 1 sumariza as informações extraídas de cada artigo, levando-se em consideração o foco do presente trabalho.

Com o levantamento realizado, foi observado que as publicações sobre este assunto são recentes. Foram incluídos: 1 artigo de 2003, 1 artigo de 2007, 3 artigos de 2009, 6 artigos de 2010, 15 artigos de 2011, 17 artigos de 2012 e 6 artigos de 2013. Nota-se que de 2010 a 2011 ocorreu um aumento de aproximadamente 71% nas publicações sobre assunto. Além disso, é nos últimos três anos que se concentra cerca de 77% das publicações. Portanto, é possível observar que este assunto tem, recentemente, atraído a atenção da comunidade científica da área.

Ao analisar a distribuição geográfica das publicações e tomando como base os dados de localização do primeiro autor, os resultados mostram que as pesquisas nessa área se concentram nos EUA, com 21 publicações, seguido pela China com 10. O Brasil localiza-se em quarto lugar com 3 publicações. Sobre os conjuntos de dados utilizados nos 49 artigos incluídos neste trabalho, foram registradas 57 fontes de dados diferentes (das 107 encontradas). Para tal análise, deve-se levar em consideração que um artigo pode ter utilizado mais de uma base de dados. A quantidade de fontes demonstra que há grande variedade de domínios.

Dentre todos os 49 artigos, o primeiro artigo publicado encontrado sobre a predição de *links* foi o de [Liben-Nowell and Kleinberg 2003], no qual os autores propuseram a predição de arestas (*links*) futuras com base nas arestas atuais, utilizando diversas medidas de proximidade (atributos) de nós em uma rede. Para tal, os autores utilizaram conjuntos de dados do arXiv, nos quais realizaram a predição de *links* de coautoria. Como resultado, eles concluíram que o atributo Katz e as variantes utilizadas apresentaram bom desempenho na maioria dos conjuntos de dados, sendo os melhores resultados obtidos em três dos cinco conjuntos. Além disso, segundo os autores, os atributos simples como CN e AA apresentaram resultados satisfatórios.

Em [Gao et al. 2011], os autores propuseram um modelo unificado de múltiplas informações da rede para prever *links* de coautoria em função do tempo. Essas informações são de três tipos: da estrutura global da rede, o conteúdo dos nós e as informações de proximidade nos grafos para capturar os padrões de evolução ao longo do tempo das ligações nas redes. Utilizando quatro conjuntos de dados do arXiv entre 1992 e 2002, os resultados apresentados demonstram, conforme os autores, que o método proposto é eficiente em vários conjuntos de dados, podendo, de acordo com os valores de *Area Under Curve* (AUC), superar os métodos tradicionais para predição de *links* temporais. Os autores comentam também a possibilidade do uso do modelo proposto em redes de larga escala.

Tabela 1. Tabela de extração dos dados

Referência	Base de dados	Atributos utilizados	Domínio de aplicação
[Aiello et al. 2012]	Last.fm e aNobii	Informações do perfil do usuário	Predição de <i>links</i> de amizade
[Almansoori et al. 2011]	Matriz com 24 encaminhamentos médicos	<i>Ethnicity</i> (E), <i>Professional Activity Match</i> (PAM), <i>Sum of Patients</i> (SofP), <i>Sum of Neighbors</i> (SofN) e <i>Jaccard Similarity</i> (JS ou JC de <i>Jaccard Coefficient</i>)	Predição de <i>links</i> positivos entre médicos
[Chang and Yao 2011]	Enron Email	<i>Singular value decomposition</i> (SVD), <i>Affinity measure</i> (AF)	Predição de <i>links</i> de trocas de e-mail
[Chelmis and Prasanna 2012]	Serviço de microblog corporativo (semelhante ao Twitter)	<i>Random</i> , <i>Shortest Distance</i> , <i>Common Neighbors</i> (CN), <i>Shared Vocabulary</i> , <i>Uniform Weighting Scheme</i> (SS_Uniform)	Predição <i>links</i> de intenção de comunicação
[Corlette and Shipman 2010]	Live Journal	Adamic-Adar (AA) e <i>Local Clustering Coefficient</i>	Predição de <i>links</i> de amizade (com efeito da abertura da rede)
[Costa and Ortale 2012]	Small World network e Enron Email	Bayesian Hierarchical Community-and-Role Model (BH-CRM), Latent Dirichlet Allocation for Graphs (LDA-G.)	Predição de <i>links</i> de interações de e-mails e citações
[Cukierski et al. 2011]	Flickr	Katz, CN, Adamic-Adar (AA), Cosseno, <i>Preferential Attachment</i> (PA), <i>Bayesian Sets</i> , SVD Features, SimRank, EdgeRank, <i>Commute Time</i> , <i>Bounded WalkPageRank</i> , <i>Maximum Flow</i> , <i>Betweenness Centrality</i> , <i>Core Number</i> , <i>Shortest Paths Histogram</i> , <i>Power Law Exponent</i>	Predição de <i>links</i> para separar relacionamentos reais de falsos
[Dong et al. 2012]	Epinions, Slashdot, Wikivote e Twitter	CN, AA, JC, PA, <i>ranking factor graph</i> , <i>out-degree</i> , <i>in-degree</i> e <i>all-degree</i>	Predição de <i>links</i> de interações em rede homogêneas e heterogênea
[Dong et al. 2011a]	CDRs de duas operadoras anônimas em uma cidade	CN, AA, JC, PA, <i>Hub Promoted Index</i> (HPI), <i>Hub Depressed Index</i> (HDI), <i>Salton Index</i> (SA), <i>Unweighted Random Walk</i> (URW), <i>Weighted call times random walk</i> (TRW), <i>Weighted call duration random walk</i> (DRW), <i>High-Performance Link Prediction</i> (HPLP), <i>Resource allocation based on weighted random walk</i> (RAURW), <i>Resource allocation based on weighted call times random walk</i> (RATRW), <i>Resource allocation based on weighted call duration random walk</i> (RADRW)	Predição de <i>links</i> de chamadas
[Dong et al. 2011b]	Power grid Network (PG), Political blogs network (PB), High-energy theory collaborations Network (Hep-th), Alex Arenas's Jazz, Alex Arenas's Email Network, Neural network of Elegans Network e US Air Network	CN, SA, Leicht-Holme-Newman Index (LHN), Sorensen Index (SOR), JC, HPI, HDI, PA, AA	Predição de <i>links</i> de interação em diversos tipos de redes
[Fire et al. 2011]	Academia, Facebook, Flickr, TheMarker e YouTube	<i>Vertex degree features</i> , <i>Vertex subgraphs features</i> , CN, <i>Total-Friends</i> , JC, <i>Transitive Friends</i> , PA, Katz, <i>Friends-measure</i> , <i>Opposite direction friends</i> , <i>Edge subgraphs edges number</i> , <i>Edge subgraphs components number</i> , <i>Shortest Path</i> (SP)	Predição de <i>links</i> faltantes de relacionamento em redes direcionadas e não direcionadas
[Gao et al. 2011]	Condensed Matter (Cond-mat), General relativity and quantum cosmology (Gr-qc), High energy physics phenomenology (Hep-ph) e High energy physics theory (Hep-th)	<i>Dependent Prediction method</i> , <i>Weighted Dependent Prediction method</i> , CN, PA, Katz, <i>Nonnegative Matrix Factorization</i> (NMF), <i>Graph Nonnegative Matrix Factorization</i> (GNMF) e <i>Graph Regularized Joint Matrix Factorization</i> (GRJMF)	Predição de <i>links</i> temporal de coautoria
[Gao et al. 2012]	Live Journal e arXiv	NMF, <i>Mixed Membership Stochastic Blockmodels</i> (MMSB), <i>Multiplicative Latent Factor Model</i> (MLFM), <i>Generalized Latent Factor Model</i> (GLFM) e <i>Latent Factor BlockModel</i> (LFBM)	Predição de <i>links</i> de relacionamento social e de coautoria
[Gomez Rodriguez and Rogati 2012]	LinkedIn	AA, CN normalizado, <i>Common attendees</i> (CAe) e Adamic-Adar baseado em evento (AAe).	Predição <i>links</i> de conexão entre usuários após participarem do mesmo evento
[Guo and Guo 2010]	DBLP e TakingItGlobal.org (TIG)	<i>Merge Weighted Features Algorithm</i> (MWF)	Predição temporal de <i>links</i> de amizades e coautoria baseado em matriz para combinação de características
[Hsieh et al. 2013]	LinkedIn, Enron Email e WikiTalk	CN, AA, <i>Time overlap</i> , <i>Company size</i> , <i>Company average age</i> , <i>Company cluster coefficient</i> , <i>Node propensity</i> e <i>Join time difference</i>	Predição de <i>links</i> interação de usuário aderido há um tempo à rede (com <i>links</i>) e de usuário recém-aderido (sem <i>links</i>)

Continua na próxima página.

Referência	Base de dados	Atributos utilizados	Domínio de aplicação
[Huang et al. 2012]	Epinions	<i>Average Filling (AF)</i> , <i>JC</i> , <i>SimRank</i> , <i>SVD</i> , <i>Matrix Completion (MC)</i> e <i>Joint Manifold Factorization (JMF)</i>	Predição de <i>links</i> de confiança e desconfiança na rede social através da agregação de redes sociais heterogêneas
[Jamali et al. 2011]	Flixster e Epinions	<i>Generalized Stochastic Blockmodel (GSBM)</i> e <i>Mixed Membership Stochastic Blockmodel (MMB)</i>	Predição de <i>links</i> entre usuários em uma Rede Social de Avaliação
[Kamei et al. 2012]	@cosme	<i>JC</i> , <i>Cosine Similarity (CS)</i> e Modelo probabilístico proposto com características latentes	Predição de <i>fan-links</i> faltantes com base nos dados observados de atividades do usuário
[Kunegis et al. 2013]	Epinions e Slashdot	<i>JC</i> , <i>AA</i> , <i>Exponential kernel</i> , <i>PageRank product</i> , <i>CN</i> , <i>Paths of length three</i> , similaridade por cosseno, <i>PA</i> e <i>PageRank condicional</i>	Predição de <i>links</i> negativos em rede sociais
[Kuo et al. 2013]	Foursquare, Twitter, e DBLP	<i>User friendship (UF)</i> , <i>Item ownership (IO)</i> , <i>Category popularity (CP)</i> , <i>BC</i> , <i>JC</i> , <i>PA</i> , <i>Attractiveness (AT)</i> , <i>PageRank with Priors (PRP)</i> , <i>AT-PRP</i> , <i>Infer e Learn</i>	Predição de <i>links</i> unseen-type em uma rede heterogenea
[Lerman et al. 2012]	Digg e Twitter	<i>CN</i> , <i>JC</i> , <i>AA</i> , <i>CS</i> , <i>Attention-limited Conservative Metric (CS-AL)</i> , <i>Non-Conservative Proximity (NC)</i> e <i>Attention-Limited Non-Conservating Proximity (NC-AL)</i>	Predição de <i>links</i> de atividade
[Leroy et al. 2010]	Flickr	<i>CN</i> , <i>Katz</i> e <i>rooted PageRank (PR)</i>	Predição de <i>links</i> entre os usuários em cold start
[Liben-Nowell and Kleinberg 2003]	astro-ph, cond-mat, gr-qc, hep-ph e hep-th	<i>CN</i> , <i>JC</i> , <i>AA</i> , <i>PR</i> , <i>Katz</i> , <i>Hitting time</i> , <i>SimRank</i> e <i>Meta-abordagens: Low-rank approximation, unseen bigrams e clustering</i>	Predição temporal de <i>links</i> de coautoria
[Lin et al. 2012]	Interactome, USAir, C. elegance e CGScience	<i>CN</i> , <i>AA</i> , <i>Resource Allocation (RA)</i> , <i>Weighted CN (WCN)</i> , <i>Weighted AA (WAA)</i> , <i>Weighted Resource Allocation (WRA)</i> , <i>BenefitRanked CN (BrCN)</i> , <i>BenefitRanked AA (BrAA)</i> e <i>BenefitRanked RA (BrRA)</i>	Predição de diversos tipos de <i>links</i> faltantes em redes ponderadas
[Lu et al. 2010]	Hep-th, CiteSeer e Society of Industrial and Applied Mathematics publications (SIAM)	<i>Katz single source (Katz-S)</i> , <i>Katz all source (Katz-C)</i> , <i>Truncated Katz single source (tKatz-S)</i> , <i>Truncated Katz all source (tKatz-C)</i> , <i>Supervised Learning single source (SL-S)</i> , <i>Supervised Learning pure color path (SL-P)</i> , <i>SL-P com L1</i> , <i>Supervised Learning hybrid color paths (SL-H)</i> , <i>SL-H com regularização L1</i> e <i>SL-H com regularização hierarchical structured (HS)</i>	Predição de <i>links</i> de coautoria
[Makrehchi 2011]	Informação bibliográfica de publicações em 20 domínios científicos coletados da web	<i>Latent Dirichlet Allocation (LDA)</i> com <i>Katz</i> , <i>LDA</i> com <i>SP</i> , <i>Bag-Of-Words (BOW)</i> e <i>Latent Semantic Indexing (LSI)</i>	Predição de <i>links</i> de coautoria, a partir da semelhança entre resumos em coautoria entre os autores
[Nie et al. 2012]	Wikipedia e Slashdot	<i>CN</i> , <i>SVD</i> , <i>Fixed Point Continuation (FPC)</i> , <i>Accelerated Proximal Gradient (APG)</i> e Método proposto pelos autores	Predição de <i>links</i> faltantes de interação entre usuários
[Perez et al. 2012]	Um conjunto de redes sociais (Address Book, Twitter, Google+ e Facebook) extraído de iPhones e um conjunto de contatos (amigos e não amigos) extraídos do Facebook	<i>CN</i> , <i>SA</i> , <i>JC</i> , <i>HPI</i> , <i>HDI</i> , <i>LHN</i> , <i>PA</i> , <i>AA</i> , <i>RA</i> e <i>WRA</i>	Predição de <i>links</i> para detecção de contatos ilegítimos
[Quercia and Capra 2009]	Parte dos dados do projeto Reality Mining do MIT	<i>SP</i> , <i>PageRank</i> , <i>HITS</i> e <i>KmarkovChain</i>	Predição de <i>links</i> para recomendar amigos com base na proximidade do celular
[de Sa and Prudencio 2011]	DBLP	<i>CN</i> , <i>JC</i> , <i>PA</i> , <i>Path Distance (PD)</i> , <i>RA</i> , <i>Local Path (LP)</i> e <i>Local Clustering Coefficient</i>	Predição de <i>links</i> temporal de coautoria em uma rede ponderada
[Shin et al. 2012]	Flickr, LiveJournal, MySpace e Epinions	<i>PA</i> , <i>AA</i> , <i>RWR</i> e <i>CN</i> . <i>Eigen-decomposition (EIG)</i> : <i>EIG-CN</i> e <i>EIG-Katz</i> . <i>Clustered Low Rank Approximation (CLRA)</i> : <i>CLRA-CN</i> e <i>CLRA-Katz</i> . <i>Multi-Scale Link Prediction (MSLP)</i> : <i>MSLP-CN</i> e <i>MSLP-Katz</i>	Predição de <i>links</i> explorando diferentes escalas de aproximação para redes sociais de grade escala
[da Silva Soares and Prudencio 2012]	Hep-th e Hep-lat	<i>AA</i> , <i>PA</i> , <i>CN</i> , <i>JC</i> , <i>Moving Average (MA)</i> , <i>Average (Av)</i> , <i>Random Walk (RW)</i> , <i>Linear Regression (LR)</i> , <i>Simple Exponential Smoothing (SES)</i> e <i>Linear Exponential Smoothing (LES)</i>	Predição de <i>links</i> de coautoria considerando séries temporais
[Song et al. 2009]	Digg, Flickr, LiveJournal, MySpace, YouTube e Wikipedia	<i>PA</i> , <i>PageRank product (PRP)</i> , <i>CN</i> , <i>AA</i> , <i>Katz</i> , <i>Graph distance (GD)</i>	Predição de <i>links</i> de relacionamentos em redes sociais de alta escala
[Song et al. 2012]	Flickr, LiveJournal e MySpace	Aprendizagem espectral com <i>Clustered Spectral Graph Embedding (CSGE)</i> , <i>Katz</i> com <i>Spectral Graph Embedding (SGE)</i> , <i>CN</i>	Predição de <i>links</i> e <i>links</i> faltantes de relacionamento

Continua na próxima página.

Referência	Base de dados	Atributos utilizados	Domínio de aplicação
[Steurer and Trattner 2013]	Second Life (posição dos usuários) e My Second Life	CN, JC, AA, PA, <i>Physical Distance</i> (MD), <i>Common Regions</i> (RC), <i>Regions Seen Concurrently</i> (RS) e <i>Observations Together</i> (RO)	Predição de link de interação entre usuários, através de análise dos dados de posição e da rede social
[Tian et al. 2010]	Facebook e CALL	Link <i>trend</i> , Número de interações totais, Número de recentes interações, Tempo da última interação, número de intervalos de tempo ativo, CN, JC, CN ativos, Número total de amigos, Número total de interações	Predição de <i>links</i> para reconexão de <i>links</i> . em redes de interação social
[Tylenda et al. 2009]	DBLP e astro-ph	Versões de PR e AA padrões e ponderadas por Ano da mais recente colaboração (<i>last</i>), Número de colaborações (<i>count</i>) e Número mínimo de coautores (<i>min. coauth</i>). <i>Maximum Entropy</i> (ME), <i>Time-Aware Maximum Entropy</i> (TME) <i>avg.</i> , <i>exp.</i> , TME <i>avg. lin.</i> , TME <i>avg. sqrt.</i> , TME <i>sum lin.</i> , Distance (<i>dist</i>), JC, CN, <i>last count</i> , <i>count last</i> , <i>min. coauthors</i> , <i>dist. last count</i> , <i>dist. count last</i> , <i>dist. min. coauth</i> , ordenação por <i>count last</i> , ordenação por <i>last count</i>	Predição de <i>links</i> de coautoria, novos e repetidos
[Valverde-Rebaza and Lopes 2012]	Twitter	<i>Within And Inter Cluster</i> (WIC), CN, AA, JC, RA e PA	Predição de <i>links</i> de seguidores no Twitter
[Vasuki et al. 2010]	Orkut e Youtube	tkatz e SVD	Predição de <i>links</i> para recomendação de comunidades
[Vasuki et al. 2011]	Rede combinada com componentes altamente conectados do Orkut (Orkut-lcc) e Youtube (Youtube-lcc)	tKatz-C, tKatz-A, tKatz com <i>latent factor model</i> (tKatz-LFM), tKatz com <i>common subspace model</i> (tKatz-CS) e tKatz <i>clustered latent factor model Equation</i> (tKatz-LFM-c)	Predição de <i>links</i> para recomendação de grupos ou comunidades em redes de grande escala
[Wang et al. 2011A]	CDRs	Katz, AA, CN, JC, <i>Spatial Cosine Similarity</i> , <i>Weighted Spatial Cosine Similarity</i> , <i>Extra-role Co-Location Rate</i> <i>Weighted</i> , <i>Weighted Co-Location Rate</i> <i>Common</i> e <i>Co-Location Rate</i>	Predição de <i>links</i> de chamadas com medidas de mobilidade
[Wang et al. 2011b]	CORA	<i>Dynamic Relational Topic Model</i> (dRTM) e RTM	Predição de <i>links</i> de citação com um modelo capaz de lidar com <i>links</i> ruidosos
[Wang et al. 2007]	DBLP, Genetics e Biochemistry	<i>Approximate Katz measure</i> (aKatz), <i>Co-occurrence probability</i> , AA e PA	Predição de <i>links</i> de coautoria utilizando um novo modelo probabilístico em rede de coautoria
[Xia et al. 2012]	Internet Movie Database	CN, JC, AA, <i>Collaborative Filtering</i> (CF), PA, Katz, <i>Minimum Description Length</i> (MDL), <i>Absent Links</i> (AL) e <i>Random Walk with Restart</i> (RWR)	Predição de <i>links</i> entre diretor e ator de filmes com métodos adaptados de métodos tradicionais baseados em vizinhança para redes sociais bipartidas
[Yin et al. 2011]	Twitter	Método proposto, PropFlow, Katz, JC, AA, CN, PA e <i>Matrix factorization</i>	Predição de <i>links</i> de seguidores em uma rede híbrida
[Yu et al. 2011]	Geraram quatro conjuntos de dados sintéticos e o conjunto de dados MIT Reality Mining Project	<i>Random</i> , <i>Same Edge</i> , <i>GPS Similarity</i> , RWR e <i>Geo-Friends Recommendation Framework</i> (GEFR)	Predição de <i>links</i> para recomendação de amigos em uma rede social cyber-physical
[Zhang et al. 2013]	Sina Microblog	<i>Exponential random graph model</i> (ERGM), JC e Katz	Predição de <i>links</i> de relacionamento nas comunidades de um microblog
[Zhong et al. 2013]	Tencent, SinaWeibo, Xiaonei, Facebook, Twitter, Github, Stackoverflow e Epinions	<i>Time-evolving Composite Network Models</i> (ITCom), <i>Mixed Membership Stochastic Blockmodels</i> (MMSB), <i>dynamic Mixed Membership Stochastic Blockmodels</i> (dMMSB), <i>Nonparametric Metadata Dependent Relational Model</i> (NMDR), <i>dynamic Infinite Relational Model</i> (dIRM) e <i>Tensor Factorization</i> (TF)	Predição temporal de <i>links</i> de interação e amizade entre usuários

Com o intuito de prever novos *links* considerando o comportamento dos *links* conforme a série temporal, [da Silva Soares and Bastos Cavalcante Prudencio 2012] utilizaram duas sessões do arXiv: *Theoretical high energy physics area* (1991 a 2010) e *High energy physics - lattice area* (1993 a 2010). A ideia básica é a construção de séries temporais para cada par de nós não conectados, usando um *score* de similaridade calculado por uma métrica topológica. Um modelo de previsão é então utilizado, a fim de prever o valor seguinte da série. Esse valor é o resultado final do par de nós para ser usado pelos métodos de predição de *link*, testado conforme uma abordagem supervisionada e não supervisionada. Conforme os autores, a abordagem supervisionada foi

melhor em todos os modelos de previsão em relação à abordagem não supervisionada, mas este trabalho ainda apresenta limitações quanto ao número de redes utilizadas nos experimentos e seus domínios.

Em [de Sa and Prudencio 2011] os autores trabalharam com redes acadêmicas e investigaram a relevância do uso de pesos nas ligações (arestas). Os autores utilizaram o conjunto de dados do DBLP, dividido em três conjuntos: não ponderada caso os dois autores já foram coautores de um mesmo artigo, ponderada de acordo com o número total de trabalhos em que o par de autores foi coautores e ponderada pela contribuição dos autores em seus trabalhos de coautoria. Conforme mostram os autores, em quase todas as comparações entre as redes, a rede não ponderada obteve um desempenho inferior em relação a, pelo menos, uma das redes ponderadas. Os autores concluem que, embora estes resultados não sejam conclusivos, é possível realizar melhorias no desempenho da predição de *links* ao se considerar os pesos de cada ligação.

Algumas pesquisas como a de [Gomez Rodriguez and Rogati 2012] são mais complexas ao considerar não só a interação *online*, mas também a interação *offline*, através de encontros sociais ou profissionais, entre os usuários. Com o intuito de mostrar como os eventos profissionais e encontros sociais no mundo real se relacionam com a dinâmica temporal e evolução de uma rede profissional, os autores concluíram que novos *links* são realizados em curto período após a data do evento e que sua predição é mais eficiente nesse período. Além disso, a conexão entre nós distintos possui influência dos nós em comum que ambos compartilham.

Em [Wang et al. 2011a], o foco principal do trabalho é explorar o poder preditivo de mobilidade individual comparado e combinado com atributos topológicos. Para tal, utilizaram as trajetórias e os padrões de comunicação de uma base anônima de um país, cujos dados são obtidos de CDRs (*Call Detail Record*). Segundo os autores, os resultados demonstram que a mobilidade tem forte influência na predição de *links*, conforme a correlação entre a semelhança entre os movimentos dos indivíduos, suas conexões sociais e a força das interações entre eles. Combinando as medidas de mobilidade e de rede, os autores mostraram que a precisão na predição pode ser significativamente melhorada com aprendizado supervisionado.

4. Considerações Finais

Com a revisão realizada é visível que o tema predição de *links* para redes sociais é recente e ao analisar todos os trabalhos revisados como um conjunto, pode-se inferir que há alguns atributos considerados tradicionais (ou de referência) como CN, Katz, JC, AA e PA, já que podem ser aplicados em diferentes domínios. Além disso, são utilizados como base para comparar o desempenho de métodos propostos ou são utilizados como base para criar variações.

O sistema de estabelecimento de relacionamentos é muito complexo, além da dinamicidade, às vezes engloba diversos fatores (como acontecimentos externos) e a alta escala na maioria das redes. Além disso, foi possível entender que cada rede social possui suas características, logo não há atributos ideais que satisfaçam todas as redes sociais. Com isso, nota-se uma grande variedade de atributos, a qual pode ser verificado na Tabela 1.

Deste modo, pode-se concluir, com base no levantamento realizado no presente trabalho, que o tema tem ganho destaque nos últimos anos e, sendo um tema recente, há muito a se pesquisar sobre novos atributos e domínios de aplicação para entendimento da complexidade de uma rede social. Mesmo assim, o presente trabalho pode ser utilizado como leitura básica para os pesquisadores ou desenvolvedores que objetivem criar novos sistemas de predição em uma rede social específica ou um algoritmo de predição multi-domínio.

Acknowledgments

O trabalho apresentado neste artigo foi parcialmente financiado pela CAPES e pelo CNPq (processos 306046/2013-0 e 477246/2013-3).

Referências

- Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., and Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Trans. Web*, 6(2):9:1–9:33.
- Almansoori, W., Gao, S., Jarada, T., Alhajj, R., and Rokne, J. (2011). Link prediction and classification in social networks and its application in healthcare. In *Information Reuse and Integration (IRI), 2011 IEEE International Conference on*, pages 422–428.
- Biolchini, J., Mian, P. G., Candida, A., and Natali, C. (2005). Systematic Review in Software Engineering. Technical Report May, Systems Engineering and Computer Science Department, COPPE/UFRJ, Rio de Janeiro.
- Chang, C. and Yao, X. (2011). Social network link predict based on af model. In *Computer Science and Network Technology (ICCSNT), 2011 International Conference on*, volume 1, pages 415–418.
- Chelmis, C. and Prasanna, V. (2012). Predicting communication intention in social networks. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 184–194.
- Corlette, D. and Shipman, III, F. M. (2010). Link prediction applied to an open large-scale online social network. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, HT '10*, pages 135–140, New York, NY, USA. ACM.
- Costa, G. and Ortale, R. (2012). A bayesian hierarchical approach for exploratory analysis of communities and roles in social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 194–201.
- Cukierski, W., Hamner, B., and Yang, B. (2011). Graph-based features for supervised link prediction. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 1237–1244.
- da Silva Soares, P. and Bastos Cavalcante Prudencio, R. (2012). Time series based link prediction. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–7.
- de Sa, H. and Prudencio, R. (2011). Supervised link prediction in weighted networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2281–2288.
- Dhote, Y., Mishra, N., and Sharma, S. (2013). Survey and analysis of temporal link prediction in online social networks. In *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*, pages 1178–1183.
- Digiampietri, L. A., do Nascimento Santiago, C. R., and Alves, C. M. (2013). Predição de co-autorias em redes sociais acadêmicas: um estudo exploratório em Ciência da Computação. In *II Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2013)*, page 12, Maceió, Alagoas, Brasil.

- Digiampietri, L. A. and Maruyama, W. T. (2014). Predição de novas coautorias na rede social acadêmica dos programas brasileiros de pós-graduação em ciência da computação. In *III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2014)*, pages 243–248.
- Digiampietri, L. A., Maruyama, W. T., Santiago, C. R. N., and da Silva Lima, J. J. (2015). Um sistema de predição de relacionamentos em redes sociais. In *XI Simpósio Brasileiro de Sistemas de Informação (SBSI 2015)*, pages 139–146.
- Dong, Y., Ke, Q., Rao, J., Wang, B., and Wu, B. (2011a). Random walk based resource allocation: Predicting and recommending links in cross-operator mobile communication networks. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 358–365.
- Dong, Y., Ke, Q., Rao, J., and Wu, B. (2011b). Predicting missing links via local feature of common neighbors. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, volume 2, pages 1038–1042.
- Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N., Rao, J., and Cao, H. (2012). Link prediction and recommendation across heterogeneous social networks. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 181–190.
- Fire, M., Tenenboim, L., Lesser, O., Puzis, R., Rokach, L., and Elovici, Y. (2011). Link prediction in social networks using computationally efficient topological features. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 73–80.
- Gao, S., Denoyer, L., and Gallinari, P. (2011). Temporal link prediction by integrating content and structure information. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*, pages 1169–1174, New York, NY, USA. ACM.
- Gao, S., Denoyer, L., and Gallinari, P. (2012). Link prediction via latent factor blockmodel. In *Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion*, pages 507–508, New York, NY, USA. ACM.
- Gomez Rodriguez, M. and Rogati, M. (2012). Bridging offline and online social graph dynamics. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12*, pages 2447–2450, New York, NY, USA. ACM.
- Guo, J. and Guo, H. (2010). Multi-features link prediction based on matrix. In *Computer Design and Applications (ICCD), 2010 International Conference on*, volume 1, pages V1–357–V1–361.
- Hasan, M. and Zaki, M. (2011). A survey of link prediction in social networks. In Aggarwal, C. C., editor, *Social Network Data Analytics*, pages 243–275. Springer US.
- Hasan, M. A., Chaoji, V., Salem, S., and Zaki, M. (2006). Link prediction using supervised learning. In *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*.
- Hsieh, C.-J., Tiwari, M., Agarwal, D., Huang, X. L., and Shah, S. (2013). Organizational overlap on social networks and its applications. In *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13*, pages 571–582, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

- Huang, J., Nie, F., Huang, H., and Tu, Y.-C. (2012). Trust prediction via aggregating heterogeneous social networks. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12*, pages 1774–1778, New York, NY, USA. ACM.
- Jamali, M., Huang, T., and Ester, M. (2011). A generalized stochastic block model for recommendation in social rating networks. In *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11*, pages 53–60, New York, NY, USA. ACM.
- Kamei, T., Ono, K., Kumano, M., and Kimura, M. (2012). Predicting missing links in social networks with hierarchical dirichlet processes. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–8.
- Kunegis, J., Preusse, J., and Schwagereit, F. (2013). What is the added value of negative links in online social networks? In *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, pages 727–736, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- Kuo, T.-T., Yan, R., Huang, Y.-Y., Kung, P.-H., and Lin, S.-D. (2013). Unsupervised link prediction using aggregative statistics on heterogeneous social networks. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13*, pages 775–783, New York, NY, USA. ACM.
- Lerman, K., Intagorn, S., Kang, J.-H., and Ghosh, R. (2012). Using proximity to predict activity in social networks. In *Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion*, pages 555–556, New York, NY, USA. ACM.
- Leroy, V., Cambazoglu, B. B., and Bonchi, F. (2010). Cold start link prediction. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '10*, pages 393–402, New York, NY, USA. ACM.
- Liben-Nowell, D. and Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management, CIKM '03*, pages 556–559, New York, NY, USA. ACM.
- Lin, Z., Yun, X., and Zhu, Y. (2012). Link prediction using benefitranks in weighted networks. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT '12*, pages 423–430, Washington, DC, USA. IEEE Computer Society.
- Lu, L. and Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6):1150 – 1170.
- Lu, Z., Savas, B., Tang, W., and Dhillon, I. (2010). Supervised link prediction using multiple sources. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 923–928.
- Makrehchi, M. (2011). Social link recommendation by learning hidden topics. In *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11*, pages 189–196, New York, NY, USA. ACM.
- Maruyama, W. and Digiampietri, L. (2016). Co-authorship prediction in academic social network. In *BRASNAM 2016*.

- Murata, T. and Moriyasu, S. (2008). Link prediction based on structural properties of online social networks. *New Generation Computing*, 26(3):245–257.
- Nie, F., Wang, H., Cai, X., Huang, H., and Ding, C. (2012). Robust matrix completion via joint Schatten p-norm and lp-norm minimization. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, pages 566–574.
- Perez, C., Birregah, B., and Lemercier, M. (2012). The multi-layer imbrication for data leakage prevention from mobile devices. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, pages 813–819.
- Quercia, D. and Capra, L. (2009). Friendsensing: Recommending friends using mobile phones. In *Proceedings of the Third ACM Conference on Recommender Systems, RecSys '09*, pages 273–276, New York, NY, USA. ACM.
- Shin, D., Si, S., and Dhillon, I. S. (2012). Multi-scale link prediction. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12*, pages 215–224, New York, NY, USA. ACM.
- Song, H. H., Cho, T. W., Dave, V., Zhang, Y., and Qiu, L. (2009). Scalable proximity estimation and link prediction in online social networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, IMC '09*, pages 322–335, New York, NY, USA. ACM.
- Song, H. H., Savas, B., Cho, T. W., Dave, V., Lu, Z., Dhillon, I. S., Zhang, Y., and Qiu, L. (2012). Clustered embedding of massive social networks. *SIGMETRICS Perform. Eval. Rev.*, 40(1):331–342.
- Steurer, M. and Trattner, C. (2013). Predicting interactions in online social networks: An experiment in second life. In *Proceedings of the 4th International Workshop on Modeling Social Media, MSM '13*, pages 5:1–5:8, New York, NY, USA. ACM.
- Tian, Y., He, Q., Zhao, Q., Liu, X., and Lee, W.-c. (2010). Boosting social network connectivity with link revival. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10*, pages 589–598, New York, NY, USA. ACM.
- Tylenda, T., Angelova, R., and Bedathur, S. (2009). Towards time-aware link prediction in evolving social networks. In *Proceedings of the 3rd Workshop on Social Network Mining and Analysis, SNA-KDD '09*, pages 9:1–9:10, New York, NY, USA. ACM.
- Valverde-Rebaza, J. and de Andrade Lopes, A. (2012). Structural link prediction using community information on twitter. In *Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on*, pages 132–137.
- Vasuki, V., Natarajan, N., Lu, Z., and Dhillon, I. S. (2010). Affiliation recommendation using auxiliary networks. In *Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10*, pages 103–110, New York, NY, USA. ACM.
- Vasuki, V., Natarajan, N., Lu, Z., Savas, B., and Dhillon, I. (2011). Scalable affiliation recommendation using auxiliary networks. *ACM Trans. Intell. Syst. Technol.*, 3(1):3:1–3:20.
- Wang, C., Satuluri, V., and Parthasarathy, S. (2007). Local probabilistic models for link prediction. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 322–331.

- Wang, D., Pedreschi, D., Song, C., Giannotti, F., and Barabasi, A.-L. (2011a). Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 1100–1108, New York, NY, USA. ACM.
- Wang, E., Silva, J., Willett, R., and Carin, L. (2011b). Dynamic relational topic model for social network analysis with noisy links. In *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pages 497–500.
- Xia, S., Dai, B., Lim, E.-P., Zhang, Y., and Xing, C. (2012). Link prediction for bipartite social networks: The role of structural holes. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 153–157.
- Yin, D., Hong, L., and Davison, B. D. (2011). Structural link analysis and prediction in microblogs. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*, pages 1163–1168, New York, NY, USA. ACM.
- Yu, X., Pan, A., Tang, L.-A., Li, Z., and Han, J. (2011). Geo-friends recommendation in gps-based cyber-physical social network. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 361–368.
- Zhang, C., Zhai, B. Y., and Wu, M. (2013). Link prediction of community in microblog based on exponential random graph model. In *Wireless Personal Multimedia Communications (WPMC), 2013 16th International Symposium on*, pages 1–6.
- Zhong, E., Fan, W., Zhu, Y., and Yang, Q. (2013). Modeling the dynamics of composite social networks. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 937–945, New York, NY, USA. ACM.

Revealing User Influence in an Online Newspaper

Gilberto Flores Pochet, Carlos Kamienski
 Universidade Federal do ABC (UFABC), Brazil,
 {gilberto.pochet, cak}@ufabc.edu.br

***Abstract** – Web-based online newspapers have become a very popular way to share information as well as to allow users to comment news and each other's comments. In such environments some users play a prominent role and eventually influence other users thus leading discussions as they please. So far, there is no way of identifying and quantifying such influence. This paper proposes and analyzes a methodology for identifying influential users that builds implicit social networks upon users who post comments in the same news, suggests three different ways of identifying influential users and measures influence based on similarity of comments. We applied it to data collected from a Brazilian online newspaper and results confirm its effectiveness by revealing a significant similarity between comments of identified influential users and the remaining ones.*

1. Introduction

User generated content and communication among users who produce content take place in different ways, such as in online news portals, blogs, messaging and social networks. An Online Social Network (OSN) generates a huge amount of data that can be used for different purposes, such as for understanding the reasons that make users like an idea, a product or a service. Users can share they thoughts freely and therefore they influence each other's opinions and decisions. Web-based newspapers have become an interesting way to share information in the past 15 years [13]. They provide Webpages and mobile applications where users can read a large variety of news, as well as post their comments and impressions about the news themselves and about other users' comments.

The effect of users who actively participate by posting comments in a social interaction driven by news is as significant as the effect of the content of the news itself [1]. The Internet may promote user active engagement because it allows users to access information on demand and to receive news in easier ways. Also, Web 2.0 features allow them to customize information to suit their interests and to provide deeper information about any topic as they see fit [3]. This flexibility can influence other users who are not only interested on news but also on other users' opinions.

In this paper we propose and analyze a methodology for identifying influential users, called Star Users, with six steps but with three key components. Firstly, we build implicit networks based on user participation on an online environment that is not officially a social network. Secondly, we propose three methods for identifying influential users: a) who posted comments in news together with the highest number of other users (Star-D); b) who posted the highest number of comments (Star-M) and; c) who posted comments in the highest number of news (Star-N). And thirdly, a method for measuring the level of similarity in the comments posted by the prospective influential users and the other remaining users, using text mining, data mining and similarity indexes. The key purpose of our work is to pinpoint more precisely the influential users. For the scope of this paper we define influence as the ability of certain users to generate discussion on a piece of news and particularly to lead other users to follow their line of

reasoning. We measured the former aspect by creating implicit social networks and identifying influential users and the latter by comparing comments using text mining.

We crawled public data from five different sections - Politics, Daily, World, Education and Technology - of the Brazilian Folha de São Paulo online newspaper. We collected news, comments and users for a period of three months, from November 2013 to January 2014 and only used public data, since privacy is a major concern of many users. Based on our results, we can see that the methodology works and confirms the influence inside the network, since the similarity metrics (Jacard's index and Dice's coefficient) yield high scores. For example, the Daily and Politics sections, the most popular ones, present scores as high as 0.76 for Dice and 0.41 for Jaccard. In other words, the ideas concealed by the Star user through their comments influence other users. Also, we observed that commenting the highest number of news (Star-N) and posting the highest number of comments (Star-M) generates more influence than having the highest degree (Star-D). The key contribution of this paper is the innovative methodology for identifying and measuring influence of users who comment news in an online newspaper.

The remainder of this paper is organized as follow. Section 2 discusses related work. Section 3 presents our methodology followed by the main results in section 4. Section 5 discusses lessons learned and section 6 draws some conclusions.

2. Related Work

There has been some research on identifying and measuring influence on online social networks, mainly analyzing Twitter due to the practicality of collecting data [18][2]. As stated by Castells [5], influence refers to the occurrence where the action of individuals can induce their friends to act in a related way. Centola [6] focused on the spread of behaviors among users and concluded that the adoption rate is higher when individuals receive multiple reinforcement signals from their neighbors. In our case, it may be mapped to a case where users start posting comments after a certain number of neighbors did it. Katona et. al [12] analyzed how highly connected users influence their neighbors, in the context of a traditional online social network, whereas our work analyzes an implicit network formed by users of an online newspaper. Chen et. al [7] studied influence maximization using greedy algorithms, but in a way not related to our problem.

Other line of existing research uses qualitative approaches for analyzing motivation and analysis of discourse of users who comment about online news. For example, Weber [17] discusses factors influencing participation and interactivity in online news and Diakopoulos & Naaman analyze the quality of discourse in the comments posted by readers of online news. We infer user influence quantitatively analyzing a large amount of news and comments and modelling the problem as a graph.

To the best of our knowledge, this paper is the first source of information in the literature that particularly deals with online newspapers using a network-oriented approach. There are some other areas of related work as online social networks, data mining, but we did not find anything similar to our methodology and our evaluation based on real data taken from an online newspaper.

3. Methodology

Our methodology has been developed specifically for this research and therefore it is considered as a key contribution of this paper. Its main idea is the creation of implicit

social networks based on the behavior of users who post comments about news in an online newspaper. This methodology is based on a sequence of six steps: 1) data collecting; 2) building social networks; 3) exploratory data analysis; 4) selecting star users and groups; 5) text mining and discovery of word clusters; 6) finding similarities.

Together, all steps of our methodology contribute to better understanding the two related key challenges: a) to identify the most influential users among those who post comments on news in terms of driving discussions or calling the attention of other users; b) to learn how to measure and where influence can be found more frequently inside a newspaper section.

3.1. Data Collecting

The newspaper chosen for gathering data is Folha de São Paulo (well-known as Folha), considered the number one in Brazil for both the traditional paper-based as well as the online web-based versions. Folha provides public and free access to most of its online news and comments posted by individual users. In order to obtain this data, we developed a crawler, using the Selenium tool and Java, that collected all news and comments from five Folha sections (Politics, World, Daily, Technology and Education) from November 1st 2013 to January 31th 2014. We collected 5,478 news and 37,672 comments coming from 6,950 different users. Please notice that the amount of data is low compared to what one can gather from online social networks. However, our goal was to analyze comments and influence in an online newspaper and not in Facebook or Twitter. We collected the following fields: user ID, date, news ID, total of comments by news and comments.

3.2. Building Social Networks

Here we propose a new method for building an implicit social network between users who post comments about news in a newspaper. The idea is that users who frequently interact by commenting the same news are in some way related to each other and form a social network. The more users comment the same piece of news, the more intense is their relationship. Therefore, we create a link between two users whenever they post comments in the same news. Please notice that we do not differentiate between commenting the same news and commenting each other comments, because users use both options interchangeably.

The process of creating implicit social networks from news, users and comments uses a bipartite graph $G=(U,V,E)$ where U denotes the set of users, V the set of news and E the set of edges between disjoint sets V and U . As emphasized by Newman [15], some social networks actually are formed by affiliation when individuals join common group memberships. Therefore they are naturally modeled by bipartite graphs, as individuals and groups are represented by two different vertex types and edges between them represent group membership. In our case, whenever users post comments in a piece of news, they become members of that news group. This method creates a bipartite network that is used for building the implicit social network afterwards (also called one-mode projection [15]).

Figure 1 illustrates the process of generating implicit networks out of news, users and comments. Let us assume users U_1 and U_2 wrote comments about news N_1 , users U_1 and U_3 wrote comments about news N_2 and users U_4 and U_5 wrote comments about news N_3 (Figure 2a). In our method, first we generate a bipartite graph for the entire set of news and users (Figure 2b). Next, we generate links between users, i.e., user U_1 is

linked to user U2 via N1, user U1 is linked to user U3 via N2 and user U4 is linked to user U5 via N3 (Figure 2c). After analyzing all comments and creating all links, our method creates two separate networks, one connecting users U1, U2 and U3 and the other connecting users U4 and U5 (Figure 2d). This method aims at creating a single connected network, but depending on the user behavior some smaller disconnected networks may appear.

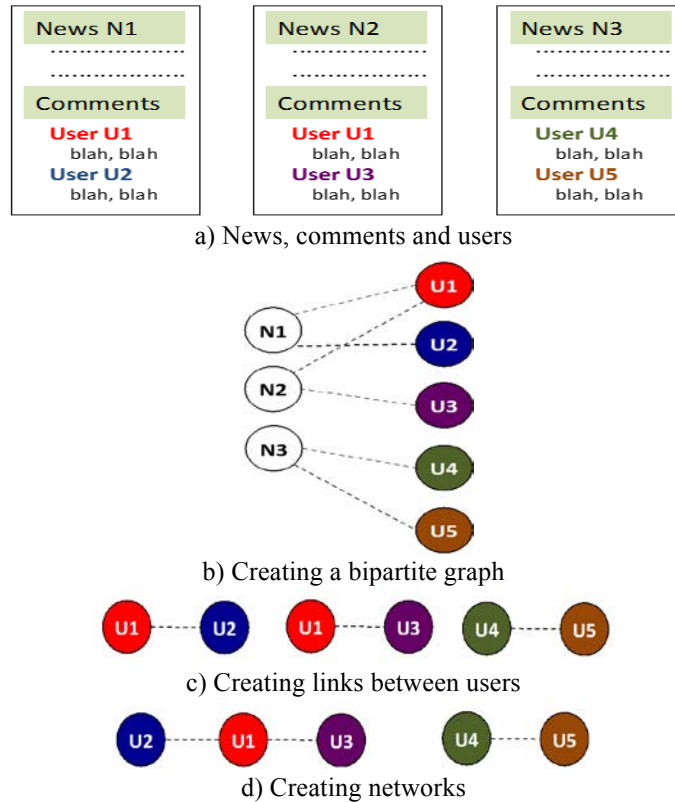


Figure 1 – Building Social Networks

3.3. Exploratory Data Analysis

After the implicit social networks are created, we analyze our new expanded dataset by observing some key metrics:

- Number of comments posted by each user
- Number of news each user posted comments
- Number of implicit networks
- User degree in implicit networks (connectivity level)
- Number of comments posted for each news
- Number of users who posted comments in each news

The EDA step plays an important role here, helping us to better understand our dataset and eventually to remove outliers.

3.4. Selecting Star Users and Groups

Key to our methodology is to identify users with high potential to exert influence on other users, called Star Users. Although a variety of characteristics may be used to find Star Users, for the scope of our research we considered three classes:

- Star-M: users with the highest number of comments, measuring their absolute performance when it comes to commenting news or participation level;
- Star-N: users who posted comments to the highest number of news, measuring their interest spread;
- Start-D: users with the highest degree, i.e. who posted comments to news that attracted a high number of other users, measuring their socialization level.

Eventually, classes Star-M, Star-N and Start-D may select the same user or three different users, depending on the newspaper section. Whenever a set of keywords a star user posts overlaps the set of keywords posted by other groups of users in the same implicit social network, we assume the former has influence over the latter, i.e., we say that this user group is under the influence zone of the star user. The comments we collected are divided into three groups:

- Group-Star: comments posted by the star users;
- Group-First: comments posted by all users directly connected to star users, i.e. their first level neighbors;
- Group-Other: comments posted by all other users who do not belong to Group-Star and Group-First.

Our method of identifying influential users is based on comparing the words of Group-Star to the words of Group-First and Group-Other. The idea is that if there is a high overlap level between Group-Star and Group-First, it may mean that star users influence their first level neighbors. On the other hand, if there is a significant overlap between words of Group-Star and Group-Other it may mean that the spread of influence of star users goes beyond their first level. As this is our basic assumption, it may not be always true. However, this method yield results, as we advocate, confirming that star users have influence over other users.

Figure 2 illustrates a distribution of information into three star users and consequently into nine different groups for the World newspaper section.

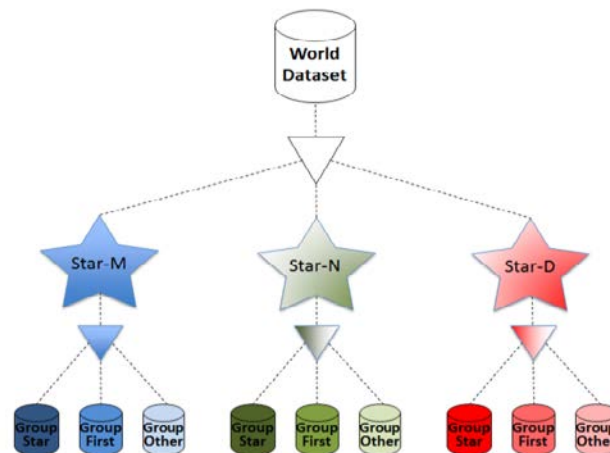


Figure 2 – Distribution of data into groups by section

Please notice that this arrangement means that we will end up with a range of information sets that may vary from 15 (5 sections x 1 star user x 3 groups) to 45 (5 sections x 3 start users x 3 groups). As mentioned before star users Star-M, Star-N and Star-D may be three different users at one extreme or they can be the same user in the other extreme.

A. *Text Mining and Discovery of Clusters*

Our methodology relies on well-known technique for text mining: tokenizing, stop list, stemming, creation of a bag of words and matrix of terms and TF-IDF. This method starts by reading the datasets and sequentially applying each of the text processing steps. We also remove terms with less than two characters, for reducing the size of the matrix of terms. The outcome of this step is a set of matrices of terms (between 15 to 45) ready for feeding the machine learning algorithms.

We use K-means [16] for machine learning, applying it over each matrix of terms coming from the previous text mining step. This process is comprised of three stages, where the first one is identifying the best number of clusters (K) [10]. We used the silhouette coefficient for finding the number of clusters. The second stage applies K-Means with the correct K [8] and finally the last stage obtains the results of the clustering method. Each cluster needs a depuration, which means that we need to extract their most frequent words for generating the Main Words or the ideas of the cluster (also keywords). At this point, one should recall that the number of cluster used here is the number of groups, which vary from 3 to 9 for each section of the newspaper, depending on the star users. We also considered other techniques in preliminary experiments, such as Latent Dirichlet Allocation [4], but since they yielded similar results, we stuck with K-Means, due to its simplicity.

3.5. Finding Similarities

A final computational step is to take the main words and apply techniques for finding similarities between them, using Jaccard’s index and the Dice’s coefficient. Figure 3 depicts the process made in the last three steps, starting with the groups of information depicted by Figure 2, passing through the text mining and data mining steps and finally being processed by this step for finding similarities between different clusters of words. An additional non-computational step is the analysis and interpretation of the final results of the similarity metrics.

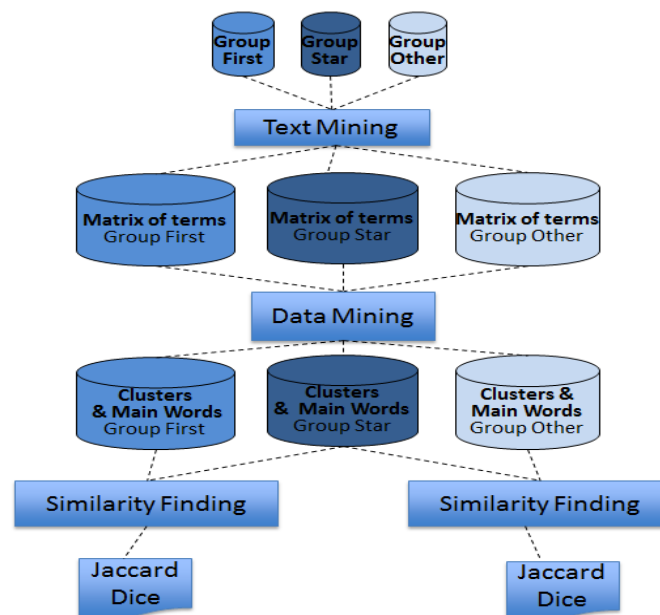


Figure 3 – Distribution of data into groups by section

Both Jaccard and Dice compare sets or strings and yield result values between 0 and 1 [11][14]. Jaccard compares words using exact matches, whereas Dice compares bigrams of each string of words. We use the results of Jaccard and Dice methods because they are two different techniques in order to obtain a clear decision for similarity. Jaccard is influenced by the number of words in a set, whereas Dice is flexible with the length of the string but is slightly less precise. Sometimes the results of the two techniques are similar, but other times they are completely different, which help us to understand better our results.

4. Results

4.1. Exploratory Data Analysis

The key results from the Exploratory Data Analysis come from Politics and Daily sections, the most popular ones. Figure 4 shows the histogram and ECDF for the Daily section. The Daily section was the most popular among users, maybe due to the bad news that always catch people's attention such as robbery, murder and poverty.

A total of 3,030 users posted 12,232 comments during the period measured, where 28 users commented more than 40 different pieces of news and 13 of them had more than one hundred comments.

As illustrated by Figure 4, more than 300 users have a degree higher than 181, which means that they commented news where this number of users also posted comments. This is the effect of 34 news that received comments from 51 different users. The whole Daily network has a clustering coefficient of 0.3, which means that it rendered a quite high triangulation of connections between users in the network.

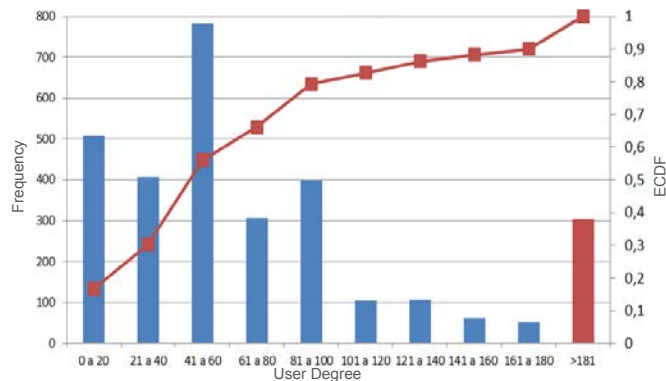


Figure 4 – Histogram/ECDF of user degree in the Daily section

On the other hand, the connectivity of the Politics section is smaller but much populated with a total of 2,560 participating users and 16,973 comments. Most news received between 0 and 10 comments, as shown by Figure 5.

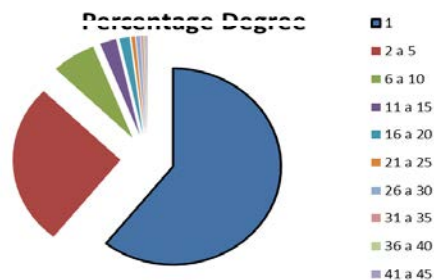


Figure 5 - Degree distribution of the Politics section

4.2. Clustering

Figure 6 presents the K-Means plot of a single star user (the same user was identified as Star-M, Start-N and Star-D) for the World section with clusters of the most common words. In this plot, we can observe a clear separation between two groups of words, where the first one in the left is larger with more ideas and documents, and the second group in the right contains less ideas and documents.

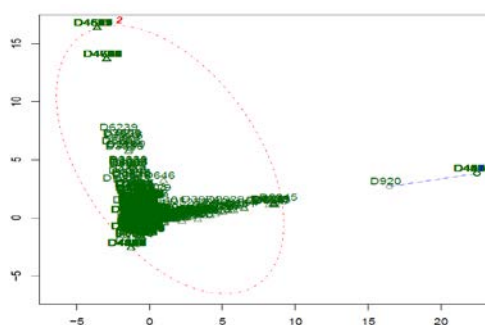


Figure 6 - K-Means plot for Group-Star – World Section

In addition, Table 1 presents a list of main words generated in the analysis of each cluster, for groups Star, First and Other. This is just an example to help understanding which words were mostly used, since the most significant results are generated by the similarity indexes. In this case, it is clear that most comments are related to war, which is quite reasonable.

For all five sections of the newspaper the results were similar to Figure 6 and Table 1. The purpose of Figure 7 is to highlight how results were obtained for Daily and Politics sections. These sections generated three Star Users (M, N and D) each, whereas for Technology, Education and World sections the three methods coincided in the same star user.

Table 1 - World section Main Words

Group	Cluster 1 Main words	Cluster 2 Main words
Other	Espionage,white,cryptology,security, service,software,stealth,companies, President	Brazil,government,Americans,military, president,Snowden,invade,trust,conflicts, nation,war,policy
First	Terror,American,military,technology, espionage,ronald,merkel,angel	Americans,fear,government,spying,speech, motives,Brazil,president,fidel,venezuela, canada,israel,Australia
Star	Religion,Planet,Terror,Americans,funding, tech,spy,software,flag	Church,Fidel,government,Islam,religion,abortion,Catholic,Vatican,Venezuela,Brazil,Cuba n,war,Israel

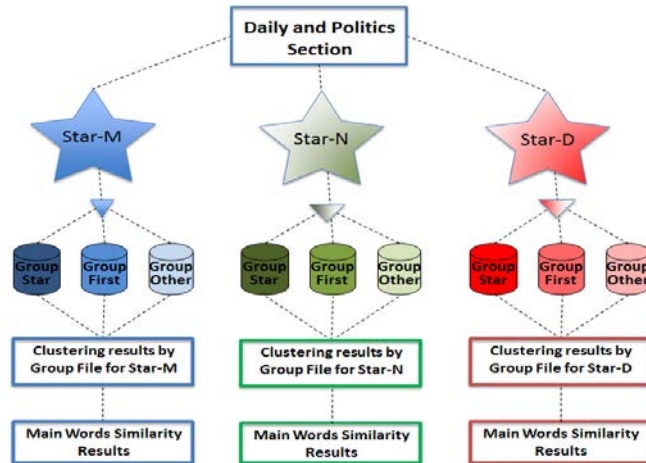


Figure 7 - Structure Results for Politics and Daily sections

4.3. Similarity results: Jaccard and Dice

Using the main words for the five sections we proceed to the next step and obtain the similarity indexes using Jaccard and Dice. In order to do this, we compare the main words (as a string) of all clusters of Group-Star with the main words of Group-First and Group-Other. For example, assuming Group-Star and Group-First have 2 clusters each, we make four cross comparisons. Next, we select the highest value of the similarity metrics among all clusters as our final result. Since clusters of words come from the same news, this procedure reveals whether star users effectively exerted influence over the remaining group of users.

In order to clarify the results, it is important to understand that Jaccard usually renders lower values than Dice, which does not mean that the former is worst than the latter. Jaccard is deeply affected by the length of the strings, so that we consider that a value higher than 0 means some similarity. On the other hand, Dice is not affected by the length of the string and therefore renders higher scores. In addition, sections Technology and Education did not generate convincing results due to the lower participation level, so that results are quite inconclusive, unfortunately.

Figure 8 presents encouraging results for the World section, where Jaccard and Dice produced scores of 0.31 and 0.65 respectively when compared the words from Group-Star to Group-First. It means that there is relevant similarity in the comments posted by the prospective influential users and their neighbors. Also, the results show that for Dice influence was spread from the Star user also beyond their first level neighbors, i.e., to Group Other.

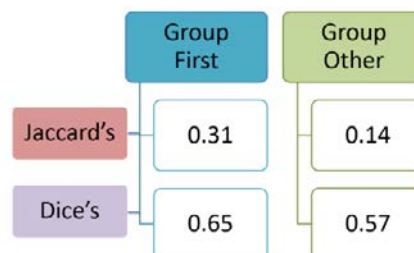


Figure 8 – Similarity results for the Star user – World section

Figure 9 presents the Jaccard and Dice results for the Daily section. Starting from Figure 10a, for Star-D, one may observe that its results are the lowest from this section, which is significant because it means that the highest degree user is not as influential as the users who posted the highest number of messages (Star-M) or commented the highest number of news (Star-N). For Star-M we have excellent results of 0.29 for Jaccard and 0.65 for Dice. For Star-N we have the best results from the Daily section, with 0.20 for Jaccard 0.71 for Dice. We observed something peculiar in this network, because the overlap of words coming from the Star users is more similar to Group-Other than to Group-First.

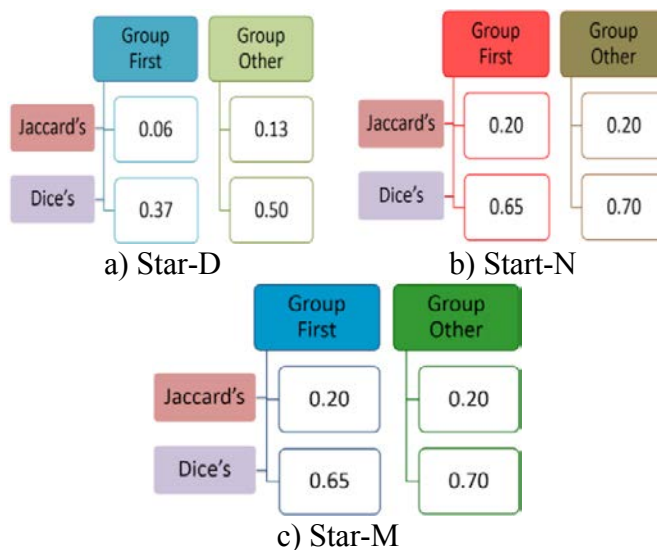


Figure 9 - Similarity results for Star users – Daily section

Figure 10 presents results for the Politics section, which are similar to those of the Daily section presented in Figure 9, where Star-D users have the lowest similarity scores and Star-M and Star-N the highest ones. It should be noticed that Star-D has a higher Jaccard score for Group-Other than for Group-First, which means that their influence is spreading into the network. The highest scores for the Politics section are for the Star-M user, although close to the similarity for Star-N so that it is not possible to state which one is more significant.

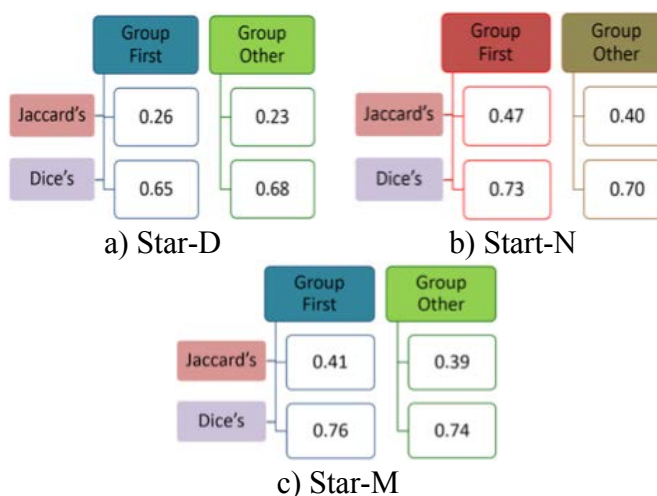


Figure 10 - Similarity results for Star users – Politics section

5. Discussion

The key contribution of this paper is an innovative methodology for identifying and measuring influence of users who comment news in an online newspaper. Based on our results, we can observe that the methodology works because the similarity metrics present high scores. For example, the Daily and Politics sections, the most popular ones, present high scores for the similarity metrics, such as 0.76 for Dice and 0.41 for Jaccard. In other words, the ideas concealed by the Star user through their comments influence other users such as Group-First and Group-Other. Also, we observed that commenting the highest number of news (Star-N) and posting the highest number of comments (Star-M) generates more influence than having the highest degree (Star-D).

As an initial and innovative methodology, there is plenty of room for improving and adapting it for being more precise in identifying influential users. For example, in this paper we assume that whenever two users commented the same piece of news, there is a connection between them. An interesting and simple variation is to consider that an edge in the implicit network will only be created when two users commented together a number of n pieces of news, where $n \geq 1$. For $n > 1$ the resulting network will be less connected, i.e., the clustering coefficient will be lower, but the relationship between those users will be stronger.

The key idea is to be able to pinpoint more precisely the influential users. Our methodology has six steps but we can identify three main components that make it appropriate for what it is aimed at. Firstly, the building of implicit networks based on participation on an environment that is not officially an online social network. Secondly, the three methods for identifying influential users, which are those who posted comments in news together with the highest number of other users (Star-D), those who posted the highest number of comments (Star-M) and those who posted comments in the highest number of news (Star-N). Of course different methods may be applied, such as the well-known centrality metrics. And thirdly, a method for measuring whether there is an overlap in the comments posted by the prospective influential users and the other remaining users, using text mining, data mining and similarity indexes.

Finally, we consider that influence is evidenced by the number of similar words used by users in their comments. However, we did not explore the cases when all words forming a cluster coincide with those of the news itself. In that case, the evidence might be less strong, since users might be just repeating the same words. In other words, we do not distinguish correlation between words from causality, i.e., one user influencing others, which we considered a derived open research problem.

6. Conclusion

This paper proposed and evaluated the effectiveness of a methodology for identifying influential users, called Star Users, who comment news in an online newspaper. This methodology builds implicit social networks upon users who post comments in the same news, propose three different ways of identifying influential users and measure influence based on an overlap between comments of these users.

Our methodology was validated by collecting news, users and comments for three months of five sections of the Brazilian Folha de São Paulo online newspaper. Results show that there is effectively a high overlap between comments of influential users and

the remaining users, up to scores of 0.76 for Dice's coefficient and 0.41 for Jaccard's index.

As future work we intend to confirm the influence of Star Users by observing their behavior over time and to confirm the effectiveness of our methodology applying it to other datasets.

References

- [1] Adamic, L. A., Adar, E., "How to Search a Social Network", *Social Networks*, 27(3), July 2005, p 187–203, 2003.
- [2] Bakshy, E., Hofman, J., Mason, W., Watts, D., "Everyone's an Influencer: Quantifying Influence on Twitter", *ACM WSDM'11*, pp. 65-74, February 2011.
- [3] Bampo, M, Ewing. M. T., Mather, D. R., Stewart, D. e Wallace, M., "The Effects of the Social Structure of Digital Networks on Viral Marketing Performance", *Information Systems Research*, Vol. 19, No. 3, p. 273-290, 2008.
- [4] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.
- [5] Castells, M. "The Information Age: Economy, Society and Culture", Oxford, 1998.
- [6] Centola, D., "The Spread of Behavior in an Online Social Network Experiment", *Science* 329, 1194, 2010.
- [7] Chen, W., Wang, Y., Yang, S., "Efficient influence maximization in social networks", *KDD '09*, pp. 199-208
- [8] Chiang, M. M. T., & Mirkin, B., "Intelligent choice of the number of clusters in K-Means clustering: an experimental study with different cluster spreads", *Journal of classification*, 27(1), 3-40, 2010.
- [9] Diakopoulos, N., Naaman, M., "Towards Quality Discourse in Online News Comments", *ACM 2011 Conference on Computer Supported Cooperative Work (CSCW 2011)*, pp. 133-142, March 2011.
- [10] Frahling, G., & Sohler, C., "A fast k-Means implementation using coresets", *International Journal of Computational Geometry & Applications*, 18(06), 605-625, 2008.
- [11] Huang, A., "Similarity measures for text document clustering", 6th New Zealand computer science conference, pp. 49-56, 2008.
- [12] Katona, Z., Zubcsek, P. P., & Sarvary, M. (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research*, 48(3), 425-443.
- [13] Katz J. "Here Come the weblogs", *Slashdot.org*, 1999.
- [14] Majumder, P., Mitra, M., & Chaudhuri, B. B., "N-gram: a language independent approach to IR and NLP", *ICUKL 2002*, November 2002.
- [15] Newman, M. E., "The Structure and Function of Complex Networks", *SIAM Rev.*, 45(2), 167–256, 2003.
- [16] Teknomo, K., "K-Means clustering tutorial", *Medicine*, 100(4), 3, 2006.
- [17] Weber, P, "Discussions in the comments section: Factors influencing participation and interactivity in online newspapers' reader comments", *New Media & Society* 16(6), pp. 941-957, August 2013.
- [18] Ye, Shaozhi, Wu, S. F., "Measuring Message Propagation and Social Influence on Twitter.com", *2nd Intl Conference on Social Informatics (SocInfo'10)*, pp. 216-231, October 2010.

Uma Análise da Polaridade Expressa nas Manchetes de Notícias Brasileiras

Pedro Ramos, Julio Reis, Fabrício Benevenuto

¹Universidade Federal de Minas Gerais (UFMG) – Brasil

{pedroramos, julio.reis, fabricio}@dcc.ufmg.br

Abstract. *In the last years, with the growth of the internet as a means of communication, a large number of people changed the way they consume news, replacing the traditional newspapers and magazines with the virtual and online versions. In order to lure people's attention and attract their clicks, news websites have to create effective strategies. Considering this, we investigate the possible strategies used by Brazilian news websites in the design of their headlines. Specifically, we analyze the number of 59.510 news produced by 8 different and important Brazilian newspapers over a 3-month period in 2015. To find the strategies that could be used to attract clicks, we extracted news text features with the polarity within the headlines. We also present a methodology to evaluate the headline polarity's strength over time. Our results reveal a number of interesting observations on polarity within the news and, in addition, we identified strong evidence that the polarity of the headlines is impacted by the occurrence of specific events.*

Resumo. *Nos últimos anos, com o crescimento da internet como meio de comunicação, um grande número de pessoas mudou a maneira de consumir notícias, substituindo os tradicionais jornais e revistas impressos por suas versões virtuais online e/ou web blogs. Sites de notícias têm que criar estratégias eficazes para chamar a atenção das pessoas e atrair seus cliques. Considerando isso, neste trabalho investigamos as possíveis estratégias utilizadas por jornais de notícias online na concepção de suas manchetes. Especificamente, analisamos o conteúdo de 59.510 notícias produzidas por 8 importantes e diferentes jornais brasileiros em um período de aproximadamente 3 meses do ano de 2015. A fim de descobrir as estratégias que poderiam ser usadas para atrair cliques, foram extraídas características do texto das notícias relacionadas com a polaridade expressa nas manchetes. Apresentamos ainda uma metodologia para avaliar a força e o teor da polaridade contida nas manchetes ao longo do tempo. Nossos resultados revelam várias observações interessantes sobre a polaridade expressa em notícias e, além disso, identificamos fortes evidências de que a polaridade das manchetes é impactada pela ocorrência de eventos específicos.*

1. Introdução

Nos últimos anos, com o crescimento da internet como meio de comunicação, um grande número de pessoas mudou a maneira de consumir notícias, substituindo os tradicionais jornais e revistas impressos por suas versões virtuais online e/ou *web blogs* [Mitchelstein and Boczkowski 2009, Tewksbury 2005].

Atualmente, as notícias online têm sido responsáveis pela captação ou atração de uma quantidade significativa de usuários na Internet [Tatar et al. 2014]. A facilidade e o baixo custo para criação e compartilhamento deste tipo de conteúdo online mudou, inclusive, as regras tradicionais de competição pela atenção do público. É comum que as grandes fontes de notícias concentrem parte de sua atenção em mídias online, onde podem divulgar suas notícias de forma eficaz e de forma mais acessível para uma grande parte da população [Bandari et al. 2012].

Neste contexto, o ambiente de jornais online tem se tornado bastante competitivo. Considerando que as pessoas estão dispostas a disponibilizarem um tempo limitado para o consumo de notícias [Varian 2010], é fundamental que esses jornais criem esquemas eficazes capazes de atrair a atenção das pessoas e conquistarem seus cliques. Entretanto, um aspecto ainda não totalmente compreendido está relacionado às estratégias adotadas por estas empresas para tornarem as notícias, e consequentemente os jornais online, mais atrativos para os usuários.

Uma das características fundamentais relacionadas com esta divulgação de notícias e sua atratividade é a manchete que a descreve. O objetivo da manchete é de forma breve e rápida chamar a atenção do leitor para a história que será apresentada. Este recurso tem sido utilizado há vários anos como mecanismo para atração dos leitores [Evans 1974].

Potencialmente as manchetes podem determinar a forma com que muitas pessoas leem as notícias [Konnikova 2016], e com o grande aumento desta ampla concorrência no mundo online, não é surpreendente nos depararmos com manchetes cada vez mais agressivas, exageradas, e em alguns casos até um pouco enganadoras. Entretanto, as manchetes não são apenas a primeira impressão que adquirimos em relação à determinada notícia. Esforços recentes sugerem que elas podem até mesmo determinar a forma de percepção dos leitores em relação ao restante do conteúdo associado, afetando inclusive a maneira como as pessoas vão se lembrar destas notícias [Ecker et al. 2014]. Apesar da importância absoluta das manchetes, pouco se sabe sobre estratégias envolvendo o seu uso durante a produção e consumo de notícias online. Acreditamos que uma melhor compreensão deste ambiente nos permita realizar observações com inúmeras aplicações para a concepção e adequação de sistemas de apoio ao usuário neste contexto. E é sobre esta lacuna que firma-se o objetivo deste trabalho. Nesta pesquisa damos um primeiro passo nesta direção ao realizarmos uma caracterização de manchetes em notícias online.

Para isso, propomos uma metodologia experimental baseada em análise de sentimentos como forma de capturar a força das polaridades expressas nas manchetes de notícias (ou seja, se elas são exageradamente negativas ou positivas, ou se elas são neutras). Mais especificamente, propomos uma análise da polaridade das notícias produzidas pelos jornais online brasileiros para respondermos as seguintes questões: Será que existe uma tendência em relação à quantidade de notícias de polaridade positiva, neutra ou negativa produzidas? Será que a polaridade das notícias publicadas é afetada pela ocorrência de algum evento específico?

Para respondermos a estas questões, coletamos e analisamos o conteúdo de 59.510 notícias produzidas por 8 importantes e diferentes jornais brasileiros durante um período de aproximadamente 3 meses do ano de 2015. Nossos resultados apresentam descobertas interessantes. Mostramos que existe um padrão em relação à polaridade de notícias produzidas pelos jornais brasileiros. Apresentamos ainda que a polaridade das manchetes é

impactada pela ocorrência de alguns eventos específicos. Além disso, apresentamos uma estratégia para avaliar a força e o teor da polaridade contida nas manchetes ao longo do tempo. Conforme discutiremos mais tarde, estas observações têm inúmeras implicações para a concepção de sistemas que apoiem o usuário neste contexto. Em última análise, nossos resultados contribuem para a compreensão da comunicação no ambiente contemporâneo, bem como da sociedade em que vivemos.

O restante desse artigo está organizado da seguinte forma: a seguir, descrevemos os trabalhos relacionados. Depois, descrevemos a nossa metodologia experimental proposta e a base de dados utilizada para o desenvolvimento do trabalho. Em seguida, apresentamos os resultados obtidos. Por fim, concluímos este trabalho e apresentamos direções para trabalhos futuros.

2. Trabalhos Relacionados

Jornais têm estabelecido portais de notícias digitais para fornecerem conteúdo público. Esses portais têm sido bastante utilizados e conseqüentemente, atraído cada vez mais e mais visitantes [Esiyok et al. 2014]. Logo, este ambiente tem sido alvo de estudos em diversos domínios, incluindo ciência da computação.

Recentemente, pesquisadores realizaram um experimento cujo objetivo era examinar o efeito de manchetes enganosas [Ecker et al. 2014]. Para realização do referido estudo, as notícias (factuais ou de opinião) foram apresentadas aos participantes, mas com diferentes manchetes. Os autores concluíram que as manchetes enganosas podem afetar a memória, o raciocínio e intenções comportamentais dos leitores. Eles argumentam que esses efeitos surgem não somente pelo fato das manchetes restringirem mais o processamento das informações e a polarização de leitores no sentido de uma interpretação específica, mas também porque os leitores lutam para atualizar sua memória, com o propósito de que sejam corrigidos equívocos iniciais. Estes esforços destacam a importância das manchetes, que vão além de atrair usuários para a leitura, podendo inclusive até alterar a percepção ou atitudes do indivíduo em relação ao conteúdo, o que motiva a realização de um estudo que se dedique exclusivamente à caracterização de manchetes.

Outro trabalho investigou, por meio de questionários, se a ordem de leitura das notícias, considerando conteúdo negativo ou positivo, apresentava impactos comportamentais no usuário [Nguyen et al. 2011]. Os resultados indicaram que usuários que primeiramente leram más notícias relataram maior intenção para a mudança de comportamento, o que não aconteceu para usuários que primeiramente leram boas notícias. Esse estudo mostra que a polaridade do conteúdo lido apresenta impacto sobre o comportamento das pessoas. Um estudo mais recente mostra que, inclusive, o usuário percebe que exista uma relação entre a polaridade do conteúdo lido e o seu humor [Reis et al. 2014a], o que intensifica a relevância na realização deste estudo a partir da polaridade expressa em manchetes.

Uma análise da polaridade expressa em manchetes de notícias no idioma inglês é apresentada em [Neves et al. 2013]. O trabalho, cujo objetivo era visualizar e detectar eventos positivos e negativos ao longo do tempo a partir da observação de notícias, analisou manchetes de diferentes fontes de jornais publicadas durante os anos de 2008 até 2013, com a utilização da plataforma AlchemyAPI¹. Os autores propuseram uma análise da tendência geral da polaridade das manchetes e desenvolveram uma ferramenta para

¹<http://www.alchemyapi.com/>

visualização do resultado do experimento (disponível em https://github.com/kelmerp/headline_sentiment_rating). O projeto, que foi desenvolvido em Chicago, visou estudar a correlação entre época do ano e sentimento das notícias, em especial, a correlação entre o atual momento político e social e os sentimentos das notícias nos diferentes jornais online.

Ainda neste contexto, em [Reis et al. 2015a] os autores investigaram estratégias utilizadas por organizações de notícias online para projetarem suas manchetes. Eles coletaram e analisaram o conteúdo de 69.907 manchetes produzidas por quatro grandes empresas de mídia conhecidas mundialmente durante um período mínimo de oito meses consecutivos do ano de 2014. Os resultados obtidos mostram que a maioria das manchetes de notícias produzidas possui polaridade negativa, e que esta situação é constante ao longo do tempo. Apresentaram ainda que manchetes extremamente negativas ou positivas geram uma maior popularidade, enquanto manchetes neutras tendem a ser menos atraentes. Uma análise sobre os comentários dos usuários mostrou que os mesmos tendem a ser mais negativos, independentemente da polaridade (negativa, neutra ou positiva) associada à manchete da notícia.

Nossa pesquisa em relação às notícias é complementar aos estudos existentes, entretanto, colocamos em perspectiva a análise do contexto brasileiro, para o qual não identificamos trabalhos que o explorem. Entretanto, este objetivo traz consigo um grande desafio que está relacionado à extração de sentimentos de um texto no idioma português. Este aspecto tem sido alvo de várias pesquisas [dos Santos et al. 2014, Reis et al. 2015b], uma vez que grande parte dos métodos de análise de sentimentos estão disponíveis apenas para o idioma inglês.

Na próxima seção apresentamos a metodologia experimental proposta para este trabalho, que inclui o processo de coleta da base de dados e a nossa estratégia adotada para inferência da polaridade expressa em uma manchete.

3. Metodologia experimental

3.1. Base de Dados

Para determinar as fontes de coleta das notícias no contexto brasileiro, tomamos como base a lista dos maiores jornais em circulação impressa e digital disponibilizada pela *Associação Nacional dos Jornais (ANJ)*². Consideramos os jornais que disponibilizam o conteúdo de notícias via *Feed RSS*. Esse mecanismo funciona como um facilitador no oferecimento do conteúdo web (ou resumos deste conteúdo) juntamente com links para as versões completas e outros metadados, permitindo ao mecanismo utilizador se manter informado sobre as atualizações efetuadas em diversos sites sem precisar visitá-los um a um. A partir disto, coletamos e utilizamos dados oriundos de 8 importantes e distintos jornais online brasileiros: G1 (O Globo) (<http://g1.globo.com/>), R7 (<http://www.r7.com/>), Yahoo Notícias (<https://br.noticias.yahoo.com/>), UOL (<http://www.uol.com.br/>), A Tarde (<http://atarde.uol.com.br/>), O Tempo (<http://www.otempo.com.br/>), Estado de Minas (<http://www.em.com.br/>) e Folha de São Paulo (<http://www.folha.uol.com.br/>).

A coleta foi realizada por meio de um *web crawler* executado durante o período de 06 de Setembro de 2015 até 30 de Novembro de 2015. Para cada uma das páginas web

²<http://www.anj.org.br/>

recuperadas, foi realizado um *parsing* do conteúdo retornado e recuperadas informações relativas à cada uma das notícias. A Tabela 1 sumariza informações relacionadas ao conjunto de dados coletado.

Tabela 1. Sumário de informações do conjunto de dados

Jornal	Total	Média Diária
A Tarde	658	8
Estado de Minas	8.687	101
Folha	7.228	84
G1	13.664	159
O Tempo	3.984	46
R7	8.325	97
UOL	7.707	90
Yahoo	9.257	108
Total	59.510	692

3.2. Inferência da polaridade

Um dos principais desafios deste trabalho consiste em definir o método de análise de sentimentos a ser utilizado para realização da pesquisa. Os métodos para análise de sentimentos existentes estão focados principalmente sobre a análise de texto em inglês, e o aspecto multilíngue tem sido bastante analisado em pesquisas recentes [Araújo et al. 2016]. Logo, ao invés de utilizarmos, por exemplo, um método que seja mais popular, identificamos a necessidade de realização de uma análise específica sobre o método mais adequado para utilização no contexto em questão.

Inicialmente criamos um subconjunto de 100 notícias aleatórias com número proporcional de notícias de cada uma das fontes coletadas. Cada uma das manchetes foi rotulada manualmente por 3 voluntários. Cada registro foi analisado por todos os voluntários com objetivo de que fossem evitados empates e alcançada uma classificação mais consistente e robusta baseada na maioria de votos, com análise do grau de concordância. Para cada manchete, foi solicitado aos voluntários que rotulassem o seu conteúdo como negativo, neutro ou positivo. A percentagem de concordância entre os voluntários foi de 76%. O resultado da classificação foi: 36% de manchetes rotuladas como positivas, 27% neutras e 37% negativas. Com esta base rotulada foi possível analisar a capacidade de predição de cada um dos métodos para o contexto proposto.

O processo de escolha do método de análise de sentimentos a ser utilizado foi baseado em [Gonçalves et al. 2013], que aponta para a não existência de um método de análise de sentimentos que seja melhor em todos os cenários. A partir daí utilizamos este conjunto de dados rotulados e comparamos o desempenho dos métodos de análise de sentimentos disponibilizados considerando 2 situações: (i) execução de um método de análise de sentimentos desenvolvido especificamente para o idioma português; (ii) tradução do texto de entrada para o idioma inglês e execução dos métodos nativos para o respectivo idioma [Araújo et al. 2016].

Nós utilizamos o conjunto de dados rotulado como *baseline* para compararmos o desempenho das estratégias de análise de sentimento. No caso (i) utilizamos a versão do método SentiStrength [Thelwall et al. 2010] desenvolvida e disponibilizada pelo autor em <http://sentistrength.wlv.ac.uk/>. Já no caso (ii) utilizamos a *Microsoft Translator API*³ para a tradução das manchetes para o idioma inglês. A escolha da desta

³<https://www.microsoft.com/en-us/translator/translatorapi.aspx>

API deve-se ao fato dela ter sido avaliada, num estudo comparativo, como a *API* mais eficiente de tradução de máquina disponível⁴, além de sua disponibilidade para utilização de forma gratuita. Em seguida, submetemos as manchetes ao iFeel [Araújo et al. 2014, Araújo et al. 2016], uma ferramenta online que disponibiliza a implementação e execução automática de vários métodos de análise de sentimento.

O método SentiStrength [Thelwall et al. 2010] específico para o idioma português (caso (i)) foi o método que apresentou melhor desempenho neste contexto, com uma acurácia de 64%, considerando as 3 classes (negativa, neutra e positiva). Outros métodos apresentaram resultados competitivos com o SentiStrength, entretanto uma propriedade atraente deste método é o formato de saída escalar, que proporciona a medição da força da polaridade em uma escala que varia de -4 (muito negativa) a 4 (muito positiva). Consideramos que isto fornece uma perspectiva interessante da intensidade da polaridade expressa na manchete. A Tabela 2 apresenta exemplos de títulos ou manchetes classificados de acordo com o SentiStrength.

Tabela 2. Exemplo de manchete por pontuação de sentimento

Força da Polaridade	Manchete
-4	Júri absolve policiais “Highlanders” acusados de decapitar deficiente em SP
-3	Acidente com ônibus deixa ao menos 15 mortos em estrada de Paraty (RJ)
-2	Lobista preso na Lava Jato nega propina e diz não conhecer Cunha
-1	ONU denuncia tchecos por prisão sistemática de migrantes e refugiados
0	Comissão bilateral Cuba-EUA se reúne em Havana
1	Ex-camelô vira empresário de sucesso e fatura R\$ 12 mil com palestras
2	Escola realiza “aulão” de redação do Enem de graça, em Porto Velho
3	Aprenda a fazer um delicioso pirulito de brigadeiro para a criançada
4	SP recebe sírios “com muito orgulho e muito prazer”, diz Haddad

Nas próximas seções apresentaremos os resultados das análises efetuadas com base nas estratégias definidas nesta etapa do trabalho, incluindo uma discussão das possíveis implicações dos resultados.

4. Resultados

4.1. Polaridade das Manchetes

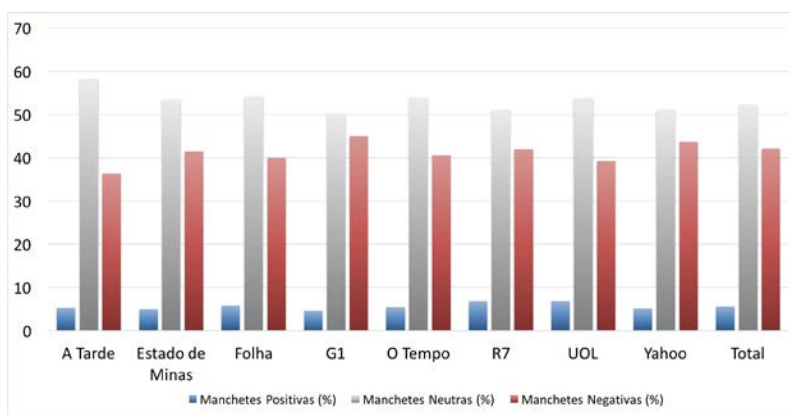
Inicialmente, medimos a força da polaridade de todas as manchetes coletadas. O resultado desta análise inicial é apresentado na Tabela 3. Observamos que o número de notícias no contexto brasileiro classificadas com força de polaridade extremas, ou seja, -4 e +4, é bastante inferior em comparação ao número de notícias classificadas com outras forças de polaridade. Notamos ainda que a maior concentração é de notícias classificadas com polaridade 0 (neutra).

Em seguida analisamos o agrupamento das notícias em 3 classes: negativa, neutra e positiva. O resultado desta análise é apresentado na Figura 1. Notamos uma tendência em todas as oito fontes de notícias analisadas: o número de manchetes neutras é o maior, seguido do número de notícias negativas e positivas. Particularmente, O jornal *GI* apresentou maior quantidade relativa de notícias negativas (45,06%), enquanto o jornal *A Tarde* conteve 36,32% de notícias negativas, que corresponde ao menor valor observado. O mesmo jornal *A Tarde* foi o que deteve a maior proporção de notícias neutras (58,36%). Percebemos ainda que a quantidade de notícias positivas publicadas está abaixo de 7%

⁴<https://www.neno-translate.com/en/blog/48-best-machine-translation-api>

Tabela 3. Percentagem de manchetes por força de polaridade

Jornal Online	Força de Polaridade								
	Negativa				Neutra	Positiva			
	-4	-3	-2	-1	0	1	2	3	4
A Tarde	36.32%				58.36%	5.32%			
	0.00%	1.67%	7.29%	27.36%		3.80%	1.52%	0.00%	0.00%
Estado de Minas	41.44%				53.57%	4.98%			
	0.02%	1.50%	8.02%	31.90%		4.03%	0.90%	0.03%	0.02%
Folha	39.98%				54.21%	5.81%			
	0.03%	1.44%	8.41%	30.11%		4.36%	1.43%	0.03%	0.00%
G1	45.06%				50.31%	4.63%			
	0.00%	1.76%	6.14%	37.16%		3.05%	1.55%	0.03%	0.00%
O Tempo	40.59%				53.99%	5.42%			
	0.00%	1.36%	7.23%	32.00%		4.49%	0.90%	0.03%	0.00%
R7	41.98%				51.20%	6.82%			
	0.01%	1.53%	7.89%	32.55%		5.32%	1.49%	0.01%	0.00%
UOL	39.29%				53.86%	6.85%			
	0.01%	2.13%	7.85%	29.30%		5.59%	1.23%	0.03%	0.00%
Yahoo	43.73%				51.15%	5.12%			
	0.01%	2.87%	9.96%	30.88%		4.25%	0.85%	0.00%	0.02%
Total	42.13%				52.31%	5.56%			
	0.01%	1.84%	7.84%	32.44%		4.29%	1.24%	0.02%	0.01%

**Figura 1. Polaridade das manchetes**

para todos os jornais, indicando que notícias com polaridade positiva são mais raras em comparação a notícias classificadas com outras polaridades.

Para uma melhor compreensão dos resultados obtidos, particionamos as análises numa perspectiva temporal, com o objetivo de verificar se a polaridade das notícias publicadas é afetada por uma determinada época ou devido a ocorrência de algum evento específico. Os resultados são apresentados na seção subsequente.

4.2. Polaridade ao longo do tempo

Nós investigamos a polaridade das manchetes em função do tempo considerando diferentes aspectos. Em primeiro lugar, nós verificamos se a polaridade das manchetes mudou ao longo do período para os dados coletados. Em seguida, analisamos se houve mudanças na polaridade das notícias que poderiam indicar reflexo de temas quentes ou eventos específicos. Depois, verificamos se existem flutuações de polaridade em determinados dias da semana, ou seja, investigamos se os dias da semana apresentam influência sobre a polaridade de notícias publicadas. Por fim propomos uma estratégia para analisar o teor da polaridade durante o período.

Considerando a distribuição de notícias publicadas ao longo do período, observamos na Figura 2 que o número de manchetes neutras e negativas é constantemente

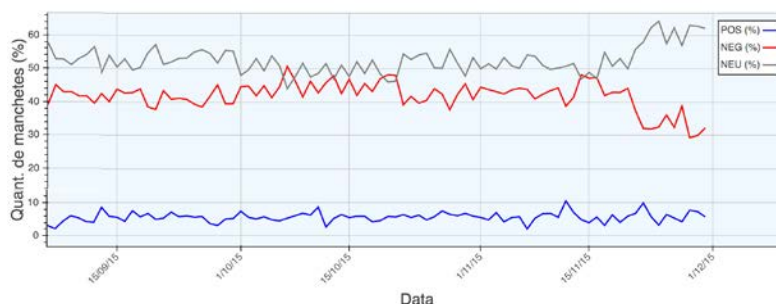


Figura 2. Percentual de notícias negativas, neutras e positivas no decorrer do tempo



(a) 13 de outubro

(b) 14 de novembro

Figura 3. Nuvem de palavras de manchetes - por dia

superior ao número de notícias positivas, com ligeiras flutuações entre as duas primeiras onde, em algumas datas, o percentual de notícias negativas inclusive superou o de notícias neutras. Com base neste aspecto averiguamos e percebemos que, particularmente, para os dias 7, 13, 20 e 21 de Outubro e 14 de Novembro de 2015, é possível notarmos algumas variações significativas (picos ou curvas) que indicam que o número de manchetes negativas publicadas é maior. Assim, para investigarmos a hipótese de que os acontecimentos deste período influenciaram a polaridade das manchetes, selecionamos as notícias veiculadas nestas datas e utilizamos um extrator de entidades⁵ para obtermos uma nuvem de palavras contendo as entidades mais citadas nas respectivas manchetes. O resultado para as datas de 13 de outubro e 14 de novembro é exibido na Figura 3⁶.

Para as datas analisadas, percebemos que o dia 13 de outubro (ver Figura 3(a)) foi marcado por notícias do contexto político nacional que afetaram de forma negativa a polaridade associada às manchetes publicadas durante este período. Além disso, podemos observar que o dia 14 de novembro foi marcado por um atentado terrorista à cidade de Paris (ver Figura 3(b)). Esse acontecimento social impactou fortemente a polaridade das manchetes veiculadas nesta data. Desta forma, acreditamos que existe forte influência de determinados acontecimentos na polaridade das notícias publicadas durante o período analisado.

Por fim, em um contexto temporal mais específico, analisamos a evolução da polaridade das manchetes no decorrer dos dias da semana. O resultado apresentado na Figura 4 nos permite notar um aumento na quantidade de notícias negativas publicadas nas segundas feiras, e uma queda na quantidade de notícias neutras publicadas, também no

⁵Tag Crowd: <http://www.tagcrowd.com/>

⁶A nuvem de palavras para as demais datas não é apresentada por limitação de espaço.

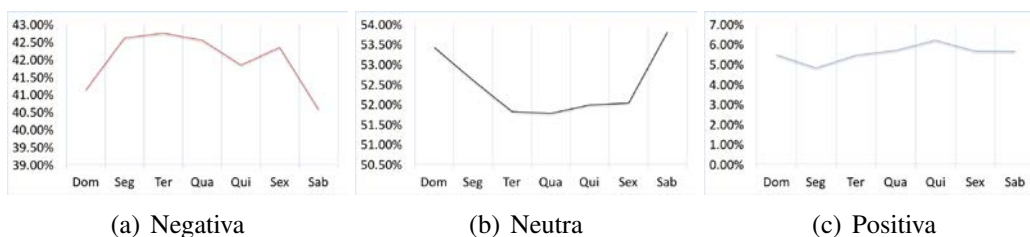


Figura 4. Percentual de manchetes publicadas - dias da semana

começo da semana. Observamos ainda que, em finais de semana, as notícias negativas diminuem. Em relação às notícias positivas, enxergamos apenas leves flutuações, mas em geral, esse percentual se mantém.

Finalmente, utilizamos nossa análise em relação à força da polaridade (ver Seção 4.1 e Tabela 3) para propormos uma forma de mensurar e analisar o teor de negatividade ou positividade expressa nas manchetes durante o período analisado. Uma vez que estamos interessados apenas no teor de negatividade ou positividade diários, nesta análise as manchetes classificadas com polaridade neutra não foram consideradas para o cálculo. Inicialmente, calculamos a média de polaridade diária, ou seja, se tivermos uma notícia negativa classificada como -4 e uma positiva classificada como 2, o teor de polaridade do dia será -1 (ou seja, $((-4) + (+2)) / 2 = -1$, teor negativo, neste caso). O resultado é apresentado na Figura 5 (linha cinza).

Observamos que o teor de polaridade é negativo durante todo o período analisado. Entretanto a média considera o número absoluto de notícias classificadas com determinada polaridade, dificultando, de certa forma a percepção da força da polaridade de cada uma das classes. Para analisarmos este aspecto, propomos uma análise que considera a diferença do módulo das médias positiva e negativa separadamente, em outras palavras, qual média é maior em módulo. Suponhamos por exemplo uma situação em que tenhamos 10 notícias negativas classificadas com força de polaridade -1, a média de polaridade negativa para este caso seria -1. Consideremos também que para esta mesma data tenham sido publicadas somente 2 notícias positivas classificadas com força de polaridade +4. Neste caso a média de polaridade positiva seria +4. A média de polaridade geral para este exemplo seria -0,2, entretanto a diferença do módulo das médias positiva e negativa separadamente seria +3 ($+4-1$). Acreditamos que isso representa de forma mais adequada a força e a expressividade das notícias nesta data. Ou seja, apesar de haver mais notícias negativas as notícias positivas são mais expressivas em termos de força de polaridade. O resultado desta análise também é apresentado na Figura 5 (linha laranja).

Percebemos que para algumas datas a força da polaridade das notícias positivas foi maior, apesar de estarem em menor número. Entretanto, não foi possível identificarmos um padrão claro, uma vez que os valores variam muito. Consideramos que ainda assim esta análise é útil uma vez que nos permite identificar em quais datas as notícias de cada uma das classes foram mais impactantes.

5. Considerações Finais e Trabalhos Futuros

O número de leitores de notícias na web tem crescido drasticamente e com isso os jornais online têm investido cada vez mais em estratégias para atrair os cliques dos usuários. Entretanto, pouco se sabe sobre estes aspectos. Propomos neste trabalho um primeiro

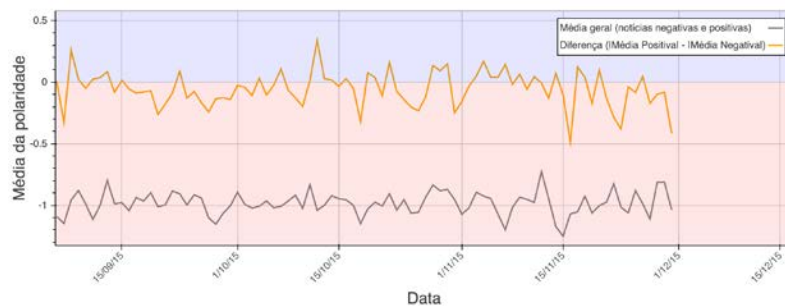


Figura 5. Teor da polaridade expressa nas manchetes

passo no entendimento deste universo com foco no contexto brasileiro. Exibimos uma caracterização de notícias neste ambiente online, que tem se tornado cada vez mais competitivo.

Para isso apresentamos uma metodologia experimental projetada para realização de nossas análises onde colocamos as manchetes como artefato central, a partir da hipótese de que elas são um importante elemento para atração da atenção dos usuários de notícias online. Nossa metodologia propõe a utilização de métodos de análise de sentimentos como meio para o entendimento de aspectos intrínsecos às manchetes.

Inicialmente, caracterizamos os dados de um conjunto de 59.510 notícias produzidas por oito grandes empresas de mídia brasileiras, reunidos durante um período de aproximadamente 3 meses do ano de 2015. Nossos resultados revelam várias observações interessantes sobre a polaridade expressa em notícias e o seu uso neste contexto, e fornecem características que podem ser utilizadas para a concepção de novos sistemas, e finalmente, para compreendermos a sociedade em que vivemos.

Revelamos uma tendência em todos os jornais analisados: o número de notícias neutras é maior, seguido do número de notícias negativas (que é competitivo em relação ao número de notícias neutras) e por fim, as notícias positivas (cuja quantidade é bastante baixa). Diferente de [Reis et al. 2015a], mostramos ainda que, para o cenário brasileiro existe grande concentração de notícias classificadas como neutras, e poucas classificadas no extremo negativo ou positivo. Sabemos que as notícias lidas podem impactar o comportamento dos usuários [Nguyen et al. 2011], e indicadores que apontam nessa direção também surgiram no estudo qualitativo realizado em [Reis et al. 2014a]. Outro aspecto está relacionado ao fato de que notícias neutras não são muito atrativas para os usuários [Reis et al. 2015a]. Logo, esses resultados contribuem para que outros pesquisadores investiguem características ou estratégias que possam apoiar o usuário na seleção de conteúdo neste cenário. Por exemplo, existem esforços recentes para a proposição de ferramentas que proporcionem ao usuário a possibilidade de selecionar um conteúdo com base na polaridade expressa pelo mesmo [Reis et al. 2014b]. Como trabalho futuro, pretendemos explorar este aspecto para o cenário brasileiro.

Já nossa análise temporal revelou que o número de manchetes neutras e negativas é constantemente superior ao número de notícias positivas. Além disso, mostramos evidências de que existe forte influência de determinados acontecimentos na polaridade das notícias publicadas durante o período analisado. Acreditamos que este tipo de descoberta possa contribuir para trabalhos que explorem esquemas de classificação com base

na previsão da popularidade de notícias. Como trabalho futuro pretendemos analisar a correlação entre a popularidade e a polaridade de uma notícia, sob a perspectiva, por exemplo, de um período ou determinado acontecimento.

Mostramos que é perceptível um aumento na quantidade de notícias negativas publicadas em segundas feiras além de uma queda na quantidade de notícias neutras publicadas no começo da semana. Apresentamos ainda que, em finais de semana, as notícias negativas diminuem. Acreditamos que uma possível explicação para esta variação esteja relacionada ao *Efeito Dia Da Semana*. Esta é uma teoria amplamente estudada em áreas do conhecimento como psicologia e economia que aponta para uma possível correlação entre dias úteis com momentos de estresse e mau-humor, e fins-de-semana com momentos de felicidade e prazer [M. Ryan et al. 2010]. Como trabalho futuro, pretendemos explorar a validação desta hipótese por meio de uma análise mais aprofundada focada especificamente neste aspecto.

Propomos também uma estratégia para análise da força da polaridade das notícias veiculadas numa mesma data. Mostramos que para algumas datas a força de polaridade das notícias positivas é maior, apesar de estarem em menor número, entretanto esse número é bastante variável. Diferente da média, esse enfoque sobre a força da polaridade explicita a expressividade da polaridade de cada uma das classes, e acreditamos que este aspecto possa ser utilizado por outros trabalhos que façam uso da análise de sentimentos como mecanismo para captura do sentimento de um texto qualquer.

Ainda como trabalho futuro, pretendemos investigar de forma mais aprofundada os resultados da execução do método de análise de sentimentos para o idioma português com objetivo de propormos estratégias que melhorem o desempenho destas abordagens no contexto de notícias online.

Por fim, acreditamos que nossa metodologia proposta, de forma geral, possa facilitar ou apoiar diferentes esforços de investigação neste domínio, impulsionando assim a produção de conhecimento.

Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto FAPEMIG - PRONEX - MASWeb, Modelos, algoritmos e sistemas para a Web (número do processo APQ - 01400-14), e por projetos individuais do CNPq, CAPES e Fapemig.

Referências

- Araújo, M., Diniz, J. P., Bastos, L., Soares, E., Júnior, M., Ferreira, M., Ribeiro, F., and Benevenuto, F. (2016). ifeel 2.0: A multilingual benchmarking system for sentence-level sentiment analysis. In *Proceedings of the International AAAI Conference on Web-Blogs and Social Media*, Cologne, Germany.
- Araújo, M., Gonçalves, P., Benevenuto, F., and Cha, M. (2014). ifeel: A system that compares and combines sentiment analysis methods. In *WWW (Companion Volume)*. International World Wide Web Conference (WWW'14).
- Araújo, M., Reis, J., Pereira, A., and Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the ACM Symposium on Applied Computing (SAC'16)*. ACM.
- Bandari, R., Asur, S., and Huberman, B. A. (2012). The pulse of news in social media: Forecasting popularity. In *ICWSM*, pages 26–33.

- dos Santos, A. G. L., Becker, K., and Moreira, V. (2014). Um estudo de caso de mineração de emoções em textos multilíngues. In *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM'14)*.
- Ecker, U. K., Lewandowsky, S., Chang, E. P., and Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, 20(4):323.
- Esiyok, C., Kille, B., Jain, B.-J., Hopfgartner, F., and Albayrak, S. (2014). Users' reading habits in online news portals. In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 263–266. ACM.
- Evans, H. (1974). *News headlines*. Holt, Rinehart and Winston.
- Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. (2013). Comparing and combining sentiment analysis methods. In *1st ACM Conference on Online Social Networks (COSN'13)*, pages 27–38.
- Konnikova, M. (2016). How headlines change the way we think. http://www.newyorker.com/science/maria-konnikova/headlines-change-way-think?mbid=social_twitter. Acessado em Março, 2016.
- M. Ryan, R., H. Bernstein, J., and Warren Brown, K. (2010). Weekends, work, and well-being: Psychological need satisfactions and day of the week effects on mood, vitality, and physical symptoms. In *Journal of Social and Clinical Psychology*, volume 29, pages 95–122.
- Mitchelstein, E. and Boczkowski, P. J. (2009). Between tradition and change: A review of recent research on online news production. *Journalism*, 10(5):562–586.
- Neves, L., Perez, K., Speisman, C., and Westerfield, C. (2013). Headlines sentiment project. <https://headlines-and-data.herokuapp.com/>. Acessado em Março, 2016.
- Nguyen, A., Legg, A., and Sweeny, K. (2011). Do you want the good news or bad news first? news order influences recipients' mood, perceptions, and behaviors. *UCR Undergraduate Research Journal*, 5:31–36.
- Reis, J., Benevenuto, F., Vaz de Melo, P., Prates, R., Kwak, H., and An, J. (2015a). Breaking the news: First impressions matter on online news. *Proceedings of the 9th International AAAI Conference on Web-Blogs and Social Media*.
- Reis, J., Gonçalves, P., Miranda, L., Prates, R., and Benevenuto, F. (2014a). A percepção dos usuários sobre a utilização de análise de sentimentos como apoio a seleção de notícias. In *Proceedings of the XIII Brazilian Symposium on Human Factors in Computing Systems*. Sociedade Brasileira de Computação.
- Reis, J., Gonçalves, P., Vaz de Melo, P. O., Prates, R., and Benevenuto, F. (2014b). Magnet news: You choose the polarity of what you read. *Proceedings of the International Conference on Web and Social Media (ICWSM)*.
- Reis, J., Oliveira, P., Araújo, M., Pereira, A., and Benevenuto, F. (2015b). Uma abordagem multilíngue para análise de sentimentos. In *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM'15)*.
- Tatar, A., Antoniadis, P., De Amorim, M. D., and Fdida, S. (2014). From popularity prediction to ranking online news. *Social Network Analysis and Mining*, 4(1):1–12.
- Tewksbury, D. (2005). The Seeds of Audience Fragmentation: Specialization in the Use of Online News Sites. *Journal of Broadcasting & Electronic Media*, 49(3):332–348.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.
- Varian, H. (2010). Newspaper economics: Online and offline. In *presentation at FTC Workshop on How Will Journalism Survive the Internet Age Washington, March*, volume 9.

Using Topological Properties to Measure the Strength of Co-authorship Ties

Michele A. Brandão, Matheus A. Diniz, Mirella M. Moro

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais – Belo Horizonte – MG – Brazil

{micheleabrandao, matheusad, mirella}@dcc.ufmg.br

Abstract. *Studying the strength of ties in social networks allows to identify impact at micro-macro levels in the network, to analyze how distinct relationships play different roles, and so on. Indeed, the strength of ties has been investigated in many contexts with different goals. Here, we aim to address the problem of measuring ties strength in co-authorship social networks. Specifically, we present four case studies detailing problems with current metrics and propose a new one. Then, we build a co-authorship social network by using a real digital library and identify how the strength of ties relates to the quality of publication venues when measured by different topological properties. Our results show the best ranked venues have similar patterns of strength of co-authorship ties.*

1. Introduction

The strength of ties in social networks has been explored in different contexts, such as information diffusion, analysis of patterns in communication logs, and so on. Specially, analyzing how strong a tie is allows to investigate the different roles of relationships, identify impact at micro-macro levels in the network [Burt 2010] and its influence in the patterns of communications [Wiese et al. 2015]. Despite the importance of analyzing the strength of ties, there are not many studies on evaluating how to measure the strength of ties in co-authorship social networks, where nodes are researchers and edges (ties) represent co-authorship between them. Studying the strength of co-authorship ties may reveal how its behaviors relate to research. Thus, any application based on co-authorship patterns may benefit. For instance, this study may help to improve Ductor [2015]’s work, which uses co-authorship to investigate if relationships lead to higher productivity.

Tie strength may be measured by a combination of the amount of time, the cooperation intensity and the reciprocal services that characterize the tie [Granovetter 1973]. Such strength may also be measured by using the *neighborhood overlap* metric [Brandão and Moro 2015, Easley and Kleinberg 2010], a topological property that captures the total number of collaboration between the two ends of each edge and identifies edges forming bridges in a community (set of nodes that are densely connected). The advantages of using such metric are: its simplicity to compute, the possibility to identify if ties are bridges or not, and the consideration of the neighbors to calculate the ties strength (then allowing to analyze how a tie is in the social network, for example, isolated or not).

Another metric that has been largely used to measure the intensity of co-authorship between ties is *weight* (the number of publications between pairs of researchers) [Silva et al. 2014, Yan et al. 2012]. Besides the simplicity to calculate such topological property, another advantage is the representation of the exact frequency of

collaboration between ties. However, we find a few problems in both metrics that complicate their sole use to measure the strength of co-authorship ties. The existence of such problems suggests the metrics should be considered together and with other social network (SN) properties to better measure ties strength.

Considering the problems of neighborhood overlap and weight to capture alone the strength of co-authorship ties, this work aims to propose a new metric, called *tieness*, that helps to define a tie as weak or strong. Note the goal of *tieness* is not to replace neighborhood overlap and weight, but being an additional feature. Another goal of this work is to study the relationship of the three metrics with the quality of publication venues. We measure such quality considering Qualis rankings¹, because most publications venues are classified according to such ranking [Laender et al 2008]. Such study may reveal another feature that may be used to measure the strength of co-authorship ties.

Indeed, after discussing related work (Section 2), we present the contributions of this paper, summarized as follows:

- We discuss four case studies where neighborhood overlap and weight alone have problems to measure the strength of ties. Also, we show the relationship between both metrics in a real dataset – DBLP² (Section 3).
- We propose a new metric called *tieness* that is a combination between a modification in neighborhood overlap and weight (Section 4).
- We describe a nominal scale to *tieness* based on the values of a modified neighborhood overlap and weight. Such nominal scale allows to identify when a tie is weak or strong and if it links researchers from different communities or not (Section 5).
- We verify the existence of a relation between the quality of a publication venue and the strength of ties when measured by different metrics (Section 6).

2. Related Work

There are many studies that address the ties strength [Brandão and Moro 2015, Goulas et al. 2014, Granovetter 1973, Wiese et al. 2015]. They contextualize the importance of our study to use a different SN to corroborate previous insights, such as distinct relationships play different roles, ties have large impact micro-macro level in the network depending on their strength, the influences in the patterns of communications, and so on.

Specifically, the strength of ties has been analyzed in different social networks. For example, Pappalardo et al. [2012] propose a definition of tie strength by measuring the interaction between two individuals over three different social channels: Facebook, Twitter and Foursquare. The existence of interactions on different channels indicate a greater chance of a stronger tie. Moreover, Gilbert and Karahalios [2009] classify friendships in Facebook as strong or weak based on variables obtained through an interaction history, such as inbox messages exchanged, days since first or last communication, and so on. Also using Facebook, Kahanda and Neville [2009] reach 87% accuracy in distinguishing strong and weak ties by mapping four different categories of features: transactional (such as picture postings and groups), network-transactional (considers the interaction between a pair of users and the overall interaction of these two users with the remaining users), topological (e.g., node degree and number of shared neighbors), and attribute

¹Qualis: <https://qualis.capes.gov.br/>

²DBLP: <http://dblp.uni-trier.de/>

based features (such as gender and interests). Furthermore, McGee et al. [2011] study if the geographic distance influence the strength of ties among users in Twitter. In order to do so, the authors consider users' friends, followers and recent tweets.

All of these methods require an interaction through the history (messages on timeline, tweets, etc) to build the predictive model (or to measure the strength of ties). Furthermore, Wiese et al. [2015] shows that the accuracy of methods based only on interaction history may be misleading. They classify strong ties through phone calls and SMS data, and managed to achieved 91.6% of accuracy even though only half of the strong ties were correctly classified. This happens because of the low proportion of strong ties regarding weak ties. Also, Zignani et al. [2016] classify ties in Facebook as interactive (strong) or non-interactive (weak) at their creation time, i.e, links without interaction history. In order to do so, the authors consider topological features, interaction-graph features and temporal features in supervised learning classifiers. Thus, it is important to explore and develop features different from interaction history.

Additionally, some datasets from social networks do not offer data on interaction among people (nodes). For example, in academic social networks, which have been largely studied [Cheng et al. 2014, Digiampietri and Maruyama 2014, Machado et al. 2013], the only data available is about the collaboration between different authors or features from authors or/and publications. This indicates the need of better topological features. In this context, we propose a new topological feature that helps to measure ties strength in co-authorship social networks.

3. An Analysis of Neighborhood Overlap and Weight

We have identified four cases where the existing metrics neighborhood overlap and weight cannot be solely used to measure strength of ties. They are defined based on empirical analysis of co-authorship social networks, as detailed next.

Case 1: A pair of collaborators without any common neighbor. One of the problems of using only neighborhood overlap to measure the strength of ties is when an author has a high frequency of collaboration with another author, but they do not have any *common neighbor*. In this case, the neighborhood overlap is zero, which does not represent the reality. Figure 1 exemplifies this case. Another problem here is that neighborhood overlap and weight present contradictory results. Analyzing neighborhood overlap, the pair AC is a bridge and the strength of co-authorship is very weak. At the same time, weight may indicate that such tie is not very weak. Therefore, considering both metrics may help to better analyze how strong a tie is.

Case 2: Determining if two collaborators are from the same community or not. One of the problems of measuring the strength of ties using only the weight is that such metric provides a simple vision of the relationship. It is not possible to know if the relationship is intra a community or not. This case is exemplified by Figure 2. Since ties with low weight may be intra a community and ties with high weight may be inter communities, the problem is that using only weight, we cannot verify Granovetter's theory [Granovetter 1973] (ties are weak when they serve as bridges in the network by connecting users from different groups, and strong when they link individuals in the same group).

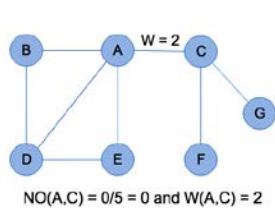


Figure 1: Case 1 – No common co-author.

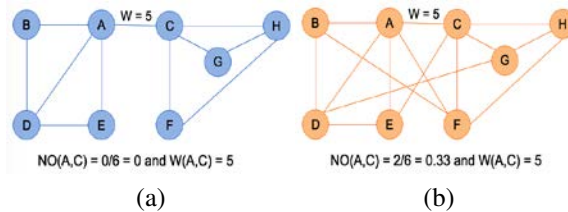


Figure 2: Case 2 – No representation if the tie is inside a community.

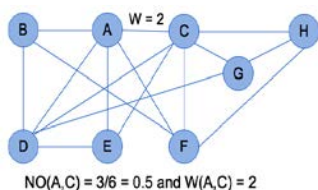


Figure 3: Case 3 – Many common co-authors.

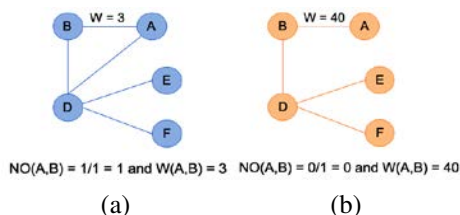


Figure 4: Case 4 – Results very small or too high.

Case 3: Little collaboration between a pair of collaborators and plenty of common neighbors. In this case, the neighborhood overlap and weight give values with opposite meaning, i.e. high neighborhood overlap and low weight. Thus, it is hard to identify if the tie would be weak or strong. Certainly, it depends on the analysis of the context. However, following Granovetter’s theory, such tie should be strong. Figure 3 gives an example of this case.

Case 4: Results with extreme values. Here, the problem is the neighborhood overlap or weight has extreme values that may not represent the reality. Figure 4a shows a maximum value to neighborhood overlap, because the edge is part of a triad. Nevertheless, the weight of the same edge is very small, which means that the tie is not necessarily very strong. Figure 4b presents a similar situation, but when the weight is very high and neighborhood overlap has the minimum value (zero). In this case, defining a tie as weak or strong based on only one of the metrics may provide a misleading interpretation.

Given all such problems, developing a new metric is necessary. In order to do so, we initially analyze the relationship between neighborhood overlap and weight in DBLP, downloaded in September 2015. We consider publications from articles and inproceedings with Qualis rankings at levels A1, A2, B1, B2, B3, B4 and B5, and from the year 1964 to 2015. Then, we build a co-authorship SN using such dataset whose features are: 145,604 authors and 208,808 links between them from 1,096,797 publications.

We now analyze the relationship between both on DBLP, as shown in Figure 5. The horizontal and vertical lines for axis x and y represent the frequency of neighborhood overlap and weight in each value, respectively. Most edges have neighborhood overlap lower than 0.50 and weight smaller than 15. Also, the correlation coefficient between the two variables is small for the three coefficients (Kendall, Spearman and Pearson).

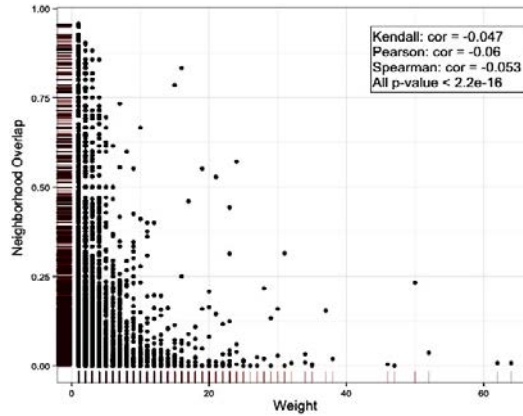


Figure 5: DBLP: Neighborhood overlap versus weight.

Therefore, neighborhood overlap and weight are monotonically and linearly independent.

Both metrics are important to measure the strength of ties and they capture different characteristics of the SN. In the next steps of this work, we intend to consider both as features to a computational model to automatically classify the strength of ties.

4. Tieness: A New Metric to Measure the Strength of Ties

Given the problems generated by using neighborhood overlap and weight alone to measure tie strength, in this section we introduce a new metric called *tieness*. Specifically, tieness results from a combination between a modification in neighborhood overlap (entitled modified neighborhood overlap) and weight, as shown by Equation 1.

$$tieness = \frac{|(A_{c_1} \cap A_{c_2}) + 1|}{|A_{c_1} \cup A_{c_2}| - (A_1 \text{ or } A_2 \text{ themselves})} \text{ weight} \quad (1)$$

where A_{c_1} represents the co-authors of researcher A_1 , and A_{c_2} the co-authors of A_2 . Note that we sum one at the numerator of neighborhood overlap to indicate that there is a link between A_1 and A_2 . This solves the problem when a pair of authors does not have any co-author in common. Also, we emphasize for unweighted social networks the tieness metric is equal to the modified neighborhood overlap.

Regarding computation time cost of tieness, the operations with the highest time cost are intersection ($O(A_{c_1} + A_{c_2})$) and union ($O(\min(A_{c_1}, A_{c_2}))$) using hash tables. Thus, the time complexity of the tieness is $O(\max(A_{c_1}, A_{c_2}))$ (Big O notation property: $O(\min(A_{c_1}, A_{c_2})) + O(A_{c_1} + A_{c_2}) = O(\min(A_{c_1}, A_{c_2}) + A_{c_1} + A_{c_2}) = O(\max(\min(A_{c_1}, A_{c_2}), A_{c_1}, A_{c_2}))$) [Cormen 2009].

Also, the overall level of tieness in a network is measured by the average of the tieness of all the edges:

$$\overline{tieness} = \frac{1}{|E|} \sum_{i=1}^{|E|} tieness_i \quad (2)$$

Table 1: Tieness for each case study and an extra case study representing the situation when NO and weight are in accordance.

Case	Image	NO	Weight	Tieness
Case 1: A pair of researchers without any common neighbor	Figure 1	0	2	0.4
Case 2: Determining if two researchers are from the same community or not	Figure 2a	0	5	0.83
Case 2: Determining if two researchers are from the same community or not	Figure 2b	0.33	5	2.5
Case 3: Little collaboration between a pair of researchers and a plenty of common neighbors	Figure 3	0.5	2	1.33
Case 4: Results with extreme values	Figure 4a	1	3	6
Case 4: Results with extreme values	Figure 4b	0	40	40
Regular Case: NO and weight in agreement	Figure 3 with $w = 12$	0.5	12	8.0

Note: NO is neighborhood overlap.

where $|E|$ is the number of edges in the social network. Also, the time complexity of the algorithm to measure the overall tieness is $O(|E| \max(A_{c_1}, A_{c_2}))$.

In order to understand how tieness represents ties in SN, Table 1 shows tieness' values for each case study. In Case 1, tieness gives a small value that indicates the presence of interactions (opposite of neighborhood overlap). However, analyzing only the final result of tieness for Cases 1, 2 and 3 is not enough to identify if a pair of researchers is intra a community or not. Also, regarding Case 4, the tieness is the same as weight when neighborhood overlap is zero and $2 \times \text{weight}$ when neighborhood overlap is one. In the Regular Case, when neighborhood overlap and weight are in accordance indicating that a tie is strong, tieness also provides a high value that may represent a strong tie. Such results cannot be used to identify if the tie belongs to a community and if it is a bridge or not. Indeed, an advantage of using our new metric is the values of ties strength are more different one from another, then allowing to better differentiate the strength of a tie from another. Also, we can consider the value of the modified neighborhood overlap and weight separately to evaluate the final result of tieness. Thus, the definition of a nominal scale is necessary to identify when a tie is weak or strong, and intra a community or not.

5. A Nominal Scale to Tieness

Here, we define a nominal scale to tieness by comparing the modified neighborhood overlap and weight. In doing so, we follow concepts discussed by Easley and Kleinberg [2010]: a weak tie has a small neighborhood overlap and a strong tie has a large one. Also, we define a nominal scale to neighborhood overlap comparing with weight.

In order to do so, Figure 6 shows the ECDFs (Empirical Cumulative Distribution Function [Lewis and McKenzie 1988]) and quartiles for neighborhood overlap, weight, modified neighborhood overlap and tieness. The analysis of ECDFs shows that neighborhood overlap and weight provide many repeated results to the strength of ties. For neighborhood overlap, 50% of data are equal to 0. In a similar manner, 50% of data have weight equal to 1. On the other hand, the modified neighborhood overlap and tieness provide different results for each quartile. Such difference among the quartiles may indicate

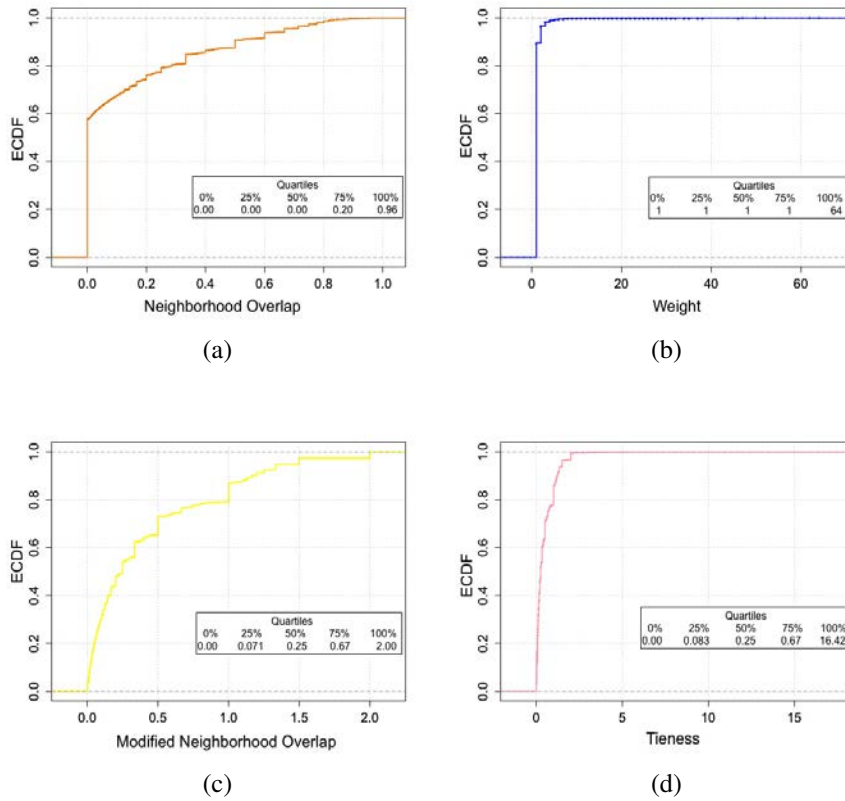


Figure 6: ECDF of each metric. In this scenario, modified neighborhood overlap and tieness metrics have more distinct values through the quartiles.

that tieness better represents the strength of ties than neighborhood overlap and weight, because it better distinguishes the relationship between nodes.

Here, we consider the values of quartiles to define a nominal scale, i.e., to identify when a tie is weak or strong, and if it connects different communities or not. Equation 3 shows the nominal scale to tieness based on the quartiles. Note for an unweighted social network, such scale is also valid because modified neighborhood overlap has the same value as tieness to the second and third quartile.

$$\begin{cases} \mathbf{weak}, & tieness \leq 0.25 \\ \mathbf{moderate}, & 0.25 < tieness < 0.67 \\ \mathbf{strong}, & tieness \geq 0.67 \end{cases} \quad (3)$$

To validate the proposed nominal scale, we verify if Granovetter's theory governs the social network and the strength of ties with such values. In summary, his theory claims the network tends to be more disconnected when weak ties are removed (i.e., the number of connected components tends to increase). Hence, we analyze the number of connected components in the network after removing weak and strong ties.

Table 2 presents the number of edges and connected components after removing weak and strong ties. Also, we show results when the strength of ties is mea-

Table 2: Number of connected components when weak and strong ties are removed from the social network.

State of the SN	# edges	% edges	# c. components	# times changed
Original	208,808	–	14,445	–
Tieness - weak ties removed	100,264	48.01	87,453	6.05
Modified NO - weak ties removed	95,706	45.83	90,864	6.29
Tieness - strong ties removed	152,529	73.07	34,605	2.4
Modified NO - strong ties removed	155,911	74.67	33,305	2.3

Note: c. components is connected components.

Table 3: DBLP sub-networks according to Qualis rankings.

SN	#nodes	% DBLP nodes	#edges	% DBLP edges	AvgWeight	AvgNO	AvgTieness
A1	18,429	12.66	21,622	10.35	1.15	0.094	0.25
A2	31,248	21.46	47,658	22.82	1.12	0.206	0.594
B1	47,101	32.35	56,889	27.24	1.18	0.101	0.47
B2	16,045	11.02	18,405	8.8	1.08	0.146	0.586
B3	14,576	10.01	17,240	8.26	1.04	0.213	0.7
B4	12,328	8.47	14,971	7.2	1.13	0.17	0.66
B5	3,939	2.7	4,345	2.08	1.07	0.136	0.6

Note: SN is social network and NO is neighborhood overlap. Weight, NO and Tieness are the average value on the SN.

sured by tieness (weighted SN) and modified neighborhood overlap (considering the SN as unweighted). According this table, when weak ties are removed, the number of connected components is higher than when removing strong ties. Also, the number of removed edges is larger when weak ties are removed. Indeed, the larger number of connected components may be explained by the larger removal of edges. However, we compare the proportion of the number of connected components by the number of edges for tieness and modified neighborhood overlap: for tieness, the proportion is $87,453/100,264 = 34,605/152,529 \Rightarrow 0.87 = 0.22$, and for modified neighborhood overlap, the proportion is $90,864/95,706 = 33,305/155,911 \Rightarrow 0.94 = 0.21$. The analysis of such proportions shows that the number of connected components per edge is greater when weak ties are removed. Thus, the nominal scale is valid. Also, as the removal of weak ties (defined according to the nominal scale) breaks the connected components of the network. Hence, tieness is indeed able to identify when a tie connects different communities or not.

6. The Strength of Ties and Publications' Quality

Here, our goal is to answer the following research question: is there any relation between the quality of a publication venue and the strength of ties when measured by different metrics? How are the values of the three metrics different from each other when considering the Qualis rankings at levels A1, A2, B1, B2, B3, B4 and B5? Such analysis provides another feature that may influence the strength of ties. Thus, we may be able to define which relationship is more important to a researcher. For instance, if a researcher has two co-authors a and b , and the publications with a are solely in high quality venues and with b in low quality venues – such situation could happen when a is a senior researcher and b is still a undergrad or grad student, who is still publishing in small venues.

Hence, using DBLP dataset, we have built a sub-network for each Qualis ranking. For example, a sub-network of researchers considering only publications in A1, then in

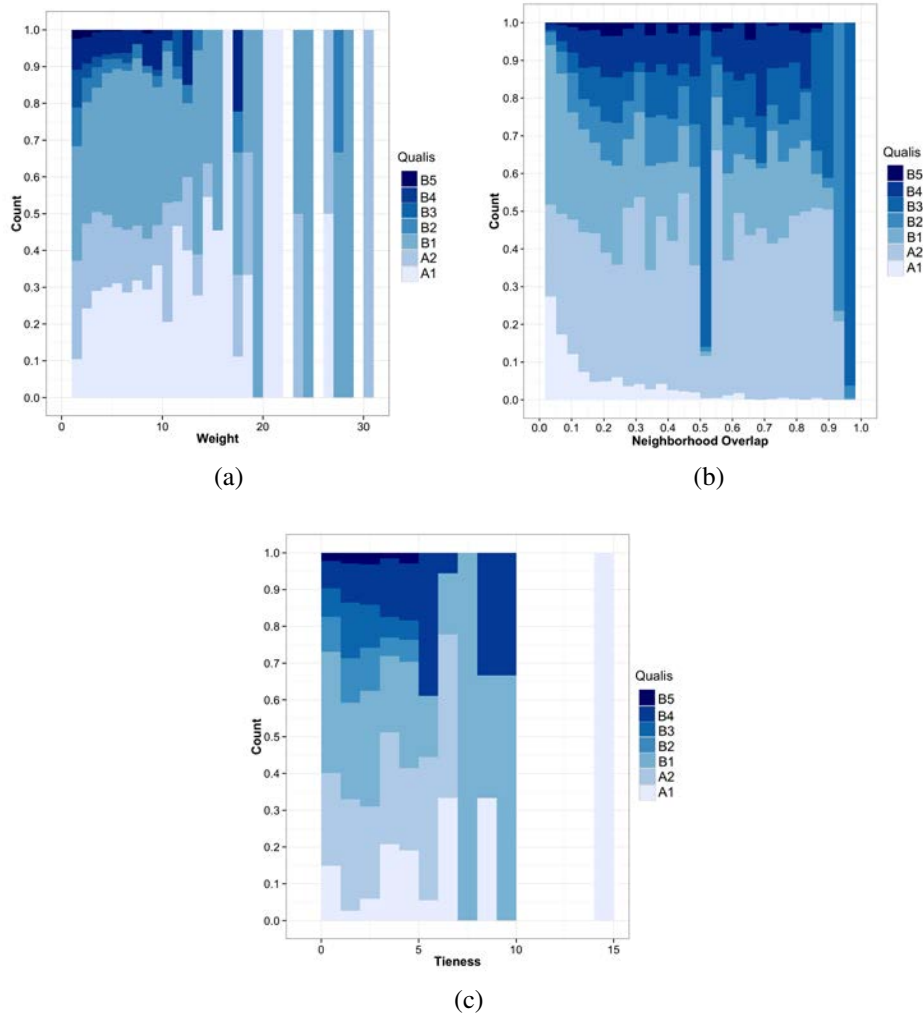


Figure 7: Proportion of co-authorships in venues grouped by Qualis rankings.

A2, and so on. Table 3 presents the main characteristics of each network: number of nodes, percentage of nodes from DBLP social network, number of edges, percentage of edges from DBLP SN, average weight of the edges in the sub-network, average neighborhood overlap and average tieness. Also, we applied a t-test (a statistical examination to compare two distributions [Jain 1991]) for the means of weight, neighborhood overlap and tieness. Most means are significantly different at 95% confidence level. The exceptions are the difference between weight mean of A2 and B4 sub-networks, and B2 and B5. In both cases, the confidence interval includes zero. Indeed, Table 3 reveals that co-authorships from publications in B1 have the highest weight average, and the co-authorships have the highest values to neighborhood overlap and tieness in B3. Additionally, B3 has the smallest weight, and A1 the lowest neighborhood overlap and tieness. Also, we observe that the values of tieness through the sub-networks are more different than the values of neighborhood overlap and weight. Thus, tieness might better represent the overall strength of the ties in the social network.

In order to compare the frequency of the strength of ties, we plot a histogram for each metric. Figure 7 contrasts the social network of each Qualis rankings. Specially,

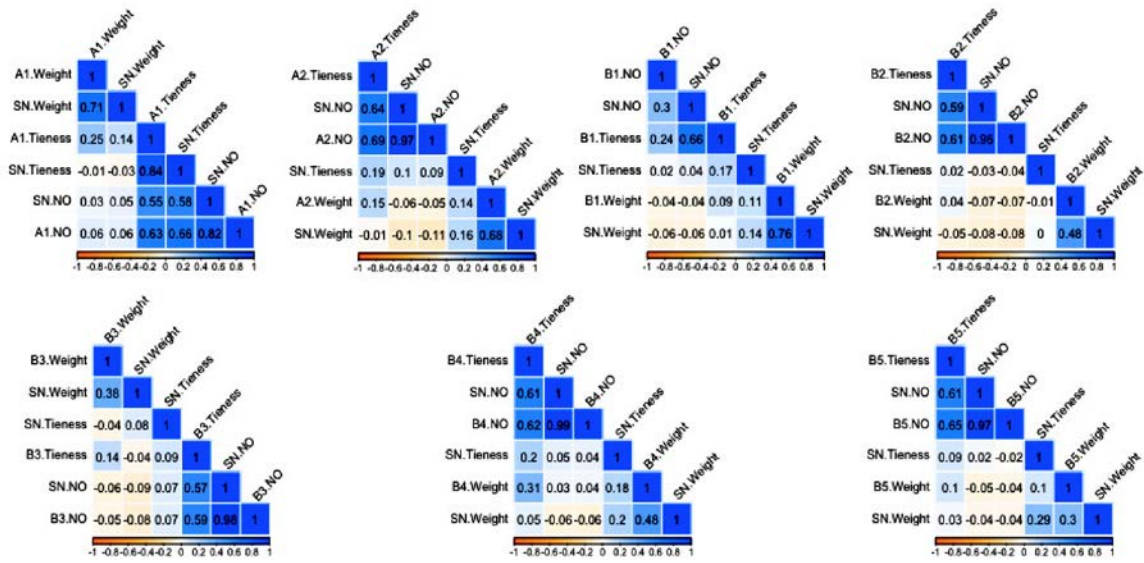


Figure 8: Correlation matrix based on Spearman's rank correlation coefficient for each metric in DBLP full social network and sub-networks.

Figure 7a shows pairs of researchers with weight higher than 20 are only in publications from A1 to B2. In addition, the proportion of co-authorships with large weight is greater to publications in classes A1 to B1 than classes B2 to B5. Figures 7b and 7c also present the difference between the strength of ties through the Qualis rankings. Also, the maximum value for tieness is only present in A1 social network (there are only 14 pairs of researchers with tieness larger than 10). Although the size of the sub-networks are different, such results indicate different behavior of co-authorship among Qualis rankings.

In this context, we also analyze the monotone correlation between weight, neighborhood overlap and tieness following Cohen [1988]'s conventions (*very large* when is greater than 0.7, *large* when is within [0.5;0.7), *moderate* when is within [0.3;0.5), *small* when is within [0.1;0.3), and *insubstantial* for any value smaller than 0.1). Figure 8 presents the correlation matrix for each Qualis rankings sub-network. Note that Figure 8 shows the value of each metric for pairs of researchers in the sub-networks (for example, A1.Tieness, B2.NO) and the equivalent value for the same pairs in the full social network of DBLP (for instance, SN.Weight, SN.Tieness and SN.NO). Hence, the correlation between SN.Weight and SN.NO in the A1 sub-network is different from A2, because each network considers the total value of the pairs of researchers in the full SN only if it exists in its sub-network.

Additionally, Figure 8 shows that changing the Qualis ranking reveals dramatic shifts in correlation. For instance, more metrics in A1 sub-network are correlated than in the others. Also, the weight in A1, A2 and B1 sub-networks are very largely and largely correlated with the weight of the full social network, whereas the weight in the remaining sub-networks are moderately correlated. Such result indicates that co-authorships from A1, A2 and B1 contribute more for the weights of the full social network. Regarding neighborhood overlap, we observe that most sub-networks have such metric largely or very largely correlated with neighborhood overlap of the full SN. The exception is B1 sub-network in which such correlation is moderate. This is expected since publications

in any Qualis ranking contribute to turn the full social network more connected. Considering tiness, only in the A1 sub-network such metric is very largely correlated with tiness of the full SN. In the other sub-networks, such correlation is small or insubstantial. This result may indicate that the weight and modified neighborhood overlap of pairs of researchers with publications in class A1 contribute more to the whole tiness of pairs of researchers. In addition, we note that there is monotone correlation between the tiness of the sub-network and the tiness of the full DBLP SN when there is also correlation between tiness and neighborhood overlap in the full SN. Such result may indicate whether the total value of neighborhood overlap of pairs of researchers in the sub-networks is not correlated to the total value of tiness, then the tiness also depends from other pairs of researchers from other sub-networks.

In summary, we observe that there are differences in the strength of ties when it is measured by neighborhood overlap, weight and tiness in each Qualis rankings. Our results show the best ranked venues (A1, A2 and B1) have similar patterns of strength of co-authorship ties – differently from the other Qualis rankings (B2, B3, B4 and B5).

7. Concluding Remarks

We identified some problems with using solely neighborhood overlap and weight to measure the strength of co-authorship ties. Then, we presented a new metric to measure such ties strength, called tiness, which has relatively low computational time cost and can be applied to other social networks types (since tiness is a topological feature). Also, we described a nominal scale to such metric that is based on the values of a modified neighborhood overlap and weight (since weight is part of the equation to compute the tiness). Such nominal scale allows to identify when a tie is weak or strong and if it links researchers from different communities or not. In addition, we checked the existence of a relation between the quality of a publication venue and the strength of ties when measured by neighborhood overlap, weight and tiness, and how the values of the three metrics are different from each other considering the publications' quality. Finally, our study showed that tiness provides more different values through the ties than neighborhood overlap and weight. This allows to better compare how strong a tie is regarding another one. We also found that high quality publications tend to have stronger ties than others.

As future work, we intend to consider neighborhood overlap, weight, tiness, publications' quality, temporal aspects and other topological properties as features to a computational model to automatically define the strength of co-authorship ties.

Acknowledgments. The authors thank the research funding agencies CAPES and CNPq.

References

- Brandão, M. A. and Moro, M. M. (2015). Analyzing the strength of co-authorship ties with neighborhood overlap. In *Procs. of DEXA*, pages 527–542, Linz, Austria.
- Burt, R. S. (2010). *Neighbor networks: Competitive advantage local and personal*. Oxford University Press.
- Cheng et al., C.-B. (2014). Study of scientific collaborations in the intelligence and security informatics research community by social network analysis. In *Procs. of CSBC – BraSNAM*, Rio de Janeiro, Brazil.

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second edition.
- Cormen, T. H. (2009). *Introduction to algorithms*. MIT press.
- Digiampietri, L. and Maruyama, W. (2014). Predição de novas coautorias na rede social acadêmica dos programas brasileiros de pós-graduação em ciência da computação. In *Procs. of CSBC – BraSNAM*, pages 243–248, Rio de Janeiro, Brazil.
- Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, 77(3):385–407.
- Easley, D. and Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press.
- Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Procs. of SIGCHI*, pages 211–220, New York, USA.
- Goulas et al., A. (2014). The strength of weak connections in the macaque cortico-cortical network. *Brain Structure and Function*, pages 1–13.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380.
- Jain, R. (1991). *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling*. John Wiley & Sons.
- Kahanda, I. and Neville, J. (2009). Using transactional information to predict link strength in online social networks. *ICWSM*, 9:74–81.
- Laender et al, A. H. F. (2008). Assessing the research and education quality of the top brazilian computer science graduate programs. *ACM SIGCSE Bulletin*, 40(2):135–145.
- Lewis, P. and McKenzie, E. (1988). *Simulation methodology for statisticians, operations analysts, and engineers*, volume 1. CRC press.
- Machado, M., Andrade, R., and Serpa, R. (2013). Teambuilder: Uso de mídias sociais para a colaboração de grupos na organização de tarefas. In *Procs. of CSBC – BraSNAM*, Rio de Janeiro, Brazil.
- McGee, J., Caverlee, J. A., and Cheng, Z. (2011). A geographic study of tie strength in social media. In *Procs. of CIKM*, pages 2333–2336, New York, USA.
- Pappalardo, L., Rossetti, G., and Pedreschi, D. (2012). How well do we know each other? detecting tie strength in multidimensional social networks. In *Procs. of ASONAM*, pages 1040–1045, Washington, USA.
- Silva et al., T. H. P. (2014). Community-based endogamy as an influence indicator. In *Procs. of JCDL*, pages 67–76, Piscataway, USA.
- Wiese et al., J. (2015). You never call, you never write: Call and sms logs do not always indicate tie strength. In *Procs. of CSCW*, pages 765–774, New York, USA.
- Yan et al., R. (2012). To better stand on the shoulder of giants. In *Procs. of JCDL*, pages 51–60, New York, USA.
- Zignani, M., Gaito, S., and Rossi, G. P. (2016). Predicting the link strength of “newborn” links. In *Procs. of WWW*, pages 147–148, Switzerland.

Detecção de Casos de Violência Patrimonial a partir do Twitter

João Paulo Clarindo¹, Fábio Coutinho¹, André Lage Freitas¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL) – Maceió, AL – Brasil

{jpcs, fabio, andre.lage}@ic.ufal.br

Abstract. *In recent years, cases of patrimonial violence have become part of the daily lives of Brazilians. This situation imposes on public security managers a big challenge: look for alternatives to the official statistics in order to reflect a more consistent overview of the reality. The popularity of social networks has resulted in continuous data generation about stories, feelings and opinions of their users. This work presents DETECT, a system which analyzes Twitter data in order to detect messages (tweets) that indicate the occurrence of patrimonial violence crimes, filling the gap presented in official data. We have experienced DETECT and initial results showed to be equivalent to the official data.*

Resumo. *Nos últimos anos, casos de violência patrimonial passaram a fazer parte do cotidiano dos brasileiros. Tal situação impõe aos gestores da segurança pública um grande desafio: buscar alternativas às estatísticas oficiais a fim de refletir uma visão mais condizente com a realidade. A popularização das redes sociais tem resultado na geração contínua de dados sobre relatos, sentimentos e opiniões de seus usuários. Este trabalho apresenta o sistema DETECT, que analisa dados da rede social Twitter com o objetivo de detectar mensagens (tweets) que indiquem a ocorrência de crimes de violência patrimonial, suprimindo a lacuna existente em dados oficiais. Um experimento foi realizado e os resultados encontrados mostraram-se equivalentes a dados oficiais.*

1. Introdução

Inicialmente, as redes sociais na Web eram vistas apenas como ambientes que permitiam a seus membros se comunicarem com amigos e familiares a partir das conexões mantidas entre os usuários. Com o decorrer do tempo, tais plataformas transformaram-se em um profícuo mecanismo de comunicação por meio do qual usuários são capazes de expressar sentimentos e manifestar suas opiniões.

O Twitter é uma das redes sociais mais utilizadas no mundo atualmente. O Brasil ocupa a segunda posição entre os países com mais usuários no Twitter [Nisha 2016]. Em janeiro de 2016, o Twitter possuía cerca de 320 milhões de usuários ativos que geravam aproximadamente 1 bilhão de mensagens diariamente [Twitter 2016]. Naturalmente, essa rede social representa uma importante fonte para a troca de opiniões.

Nos últimos anos, a segurança pública tem se tornado um dos maiores desafios da sociedade brasileira. De fato, os dados são alarmantes e evidenciam uma escalada da violência que atinge principalmente os grandes centros urbanos. De acordo com uma pesquisa realizada em São Paulo [CPP/Insper 2013], no período entre 2003 e 2013 registrou-se um aumento de 37% nos casos de furto, quando considerados os dados oficiais do governo estadual. Todavia, o mesmo estudo conclui que o aumento real foi de 81,5%. Tal diferença pode ser explicada pelo fato de que muitas vítimas não registram

a ocorrência desses crimes. Em se tratando de casos de roubo, apenas 35,3% das pessoas entrevistadas fizeram o registro, sendo o “medo de represália” (30%) e a “descrença que faria alguma diferença” (18,9%) os principais motivos alegados por quem não registrou o boletim de ocorrência.

Sendo assim, faz-se necessário buscar alternativas que permitam melhorar as estatísticas utilizadas pelos gestores e, preferencialmente, que reflitam uma visão mais recente dos acontecimentos e mais condizente com a realidade. Uma fonte viável para a obtenção desses dados são as redes sociais. De fato, devido à ocorrência cotidiana de assaltos, roubos e furtos nas principais cidades brasileiras, é natural que os seus cidadãos se utilizem das redes sociais para manifestar indignação e alertar amigos mediante o relato dos crimes aos quais foram submetidos.

Este trabalho apresenta o sistema DETECT (**DE**Tecção de **E**ventos **C**riminais a partir do **T**witter), que analisa dados da rede social Twitter com o objetivo de detectar mensagens (*tweets*) que relatem a ocorrência de crimes de violência patrimonial. Além de auxiliar aos gestores, tais dados podem servir também para ajudar na orientação dos cidadãos em áreas que não estão familiarizados, possivelmente, alimentando sistemas de navegação a fim de alertar os usuários sobre trechos com maior incidência de crimes.

Este artigo encontra-se organizado da seguinte maneira. Inicialmente, a Seção 2 discute os principais trabalhos relacionados encontrados na literatura. A Seção 3 detalha o funcionamento do sistema enquanto a Seção 4 descreve uma prova de conceito do sistema DETECT. Finalmente, a Seção 5 apresenta as considerações finais.

2. Trabalhos Relacionados

O problema de extrair e analisar *tweets* relacionados à violência pode ser visto como um problema de categorização ou classificação de tópicos, ou seja, decidir se um *tweet* será classificado como violento ou não-violento. Nesta seção, são discutidos os principais trabalhos encontrados na literatura que analisam dados do Twitter no contexto da violência.

[Cano Basave et al. 2013] apresentam um trabalho de detecção de *tweets* sobre violência através de um modelo de detecção de violência (VDM – *Violence Detection Model*). Este modelo propõe uma classificação da violência em nível de documento para domínios gerais em conjunção com detecção de tópicos e análise de tópicos relacionados à violência. O trabalho apresenta um modelo interessante para identificação de *tweets* violentos, contudo, não possui abordagem para a definição e tratamento da localização onde o *tweet* foi submetido.

[Li et al. 2012] introduzem um sistema de análise e detecção de eventos relativos à criminalidade e a desastres naturais baseados no Twitter, denominado TEDAS. O sistema permite detectar novos eventos, analisar padrões temporais e espaciais e identificar a importância dos eventos através do uso de mapas de calor gerados a partir de dados de geolocalização de *tweets*. TEDAS utiliza uma amostra menor de *tweets* para a análise com enfoque principal voltado para a detecção de desastres naturais.

ReDites [Osborne et al. 2014] constitui um sistema de detecção, rastreamento e monitoramento em tempo real de eventos em redes sociais. No experimento realizado neste trabalho, foi avaliada a possível detecção do ataque terrorista ao shopping Westgate, que ocorreu em setembro de 2013 no Quênia, utilizando apenas *tweets* obtidos através da

Streaming API. ReDites propõe a análise de *tweets* geolocalizados em tempo real, identificando as localidades das ocorrências de imediato.

Similarmente aos trabalhos acima relacionados, DETECT extrai e analisa os *tweets* em busca de padrões. Um dos diferenciais desta proposta é a possibilidade de integração dos dados encontrados com dados governamentais relacionados à segurança pública, possibilitando a validação dos mesmos.

3. O Sistema DETECT

Esta seção descreve as etapas seguidas neste trabalho para o processamento dos dados da plataforma Twitter: extração dos *tweets*, análise dos dados e visualização.

A Figura 1 apresenta uma visão geral do sistema, cuja execução é dividida em quatro passos: *extração*, *análise*, *identificação de localização* e *resultados*. O Passo 1 representa a fase de extração, que dispõe do módulo de extração, responsável por obter os *tweets* através da *Streaming API* e gerar arquivos de saída para serem utilizados pelo módulo de seleção executado no Passo 2. O módulo de seleção irá recuperar os *tweets* que contêm termos sobre violência patrimonial. Durante a fase de análise, a cada *tweet* é atribuído um índice de relevância a fim de detectar aqueles que de fato relatam um crime de violência patrimonial. Em seguida, no Passo 3, DETECT tentará identificar o local da ocorrência do crime através de seu próprio conteúdo ou de tags de geolocalização a partir do módulo de localização do *tweet*. Finalmente, no Passo 4, os resultados são visualizados e armazenados pelo módulo de visualização e armazenamento.

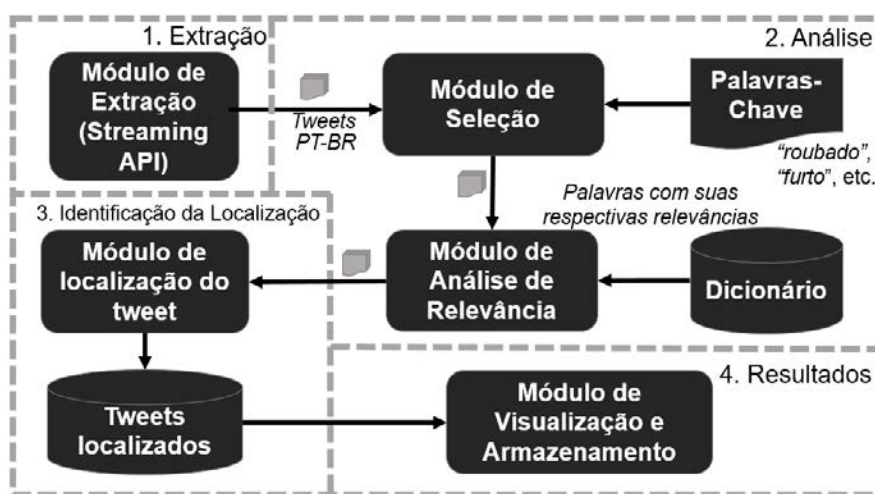


Figura 1. Visão Geral do Sistema DETECT

As seções a seguir, descrevem em detalhes os processos de extração, análise e visualização destes dados.

3.1 Extração de *tweets*

Para o módulo de extração de *tweets*, foi escolhido o mecanismo de extração *Streaming API*, que permite a extração em modo *stream*, ou seja, no momento da geração dos *tweets*.

O módulo de extração implementado em DETECT foi desenvolvido em Python. A abrangência da extração são os *tweets* gerados em tempo real e escritos em língua portuguesa, sem nenhuma filtragem de termos específica.

Além da funcionalidade básica de extração, o módulo exclui os *retweets* (replacação de *tweets*) no momento da extração para evitar repetições.

3.2 Análise

O processo de escolha para as palavras-chave utilizado na seleção de *tweets* teve origem a partir de uma pesquisa acerca de termos comuns presentes em relatos de crimes de violência patrimonial. As derivações destas palavras também foram utilizadas como palavras-chave. A base de termos que serviu como referência para a seleção neste trabalho foi construída levando em conta dados de boletins de ocorrência e comentários postados na Web a partir do site “Onde fui roubado”¹. Dentre as palavras utilizadas na seleção de *tweets* incluem-se *roubo*, *assalto*, *furto*, *arrombamento*, *mão armada*, e *arrastão*.

Para implementar o módulo de seleção de *tweets*, foi utilizado o Pentaho Data Integration (Kettle), que disponibiliza um ambiente ETL². A transformação de seleção feita pelo Kettle é representada pela Figura 2, onde os *tweets* são filtrados de acordo com as palavras-chave e enviados para o módulo de análise de relevância.

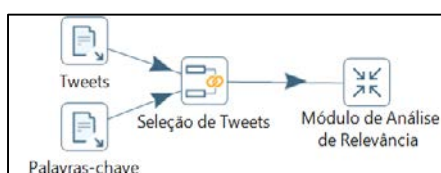


Figura 2: Transformação para seleção de tweets

Após os *tweets* terem sido selecionados de acordo com as palavras-chave, um dicionário com índices de relevância foi criado a partir de uma solução desenvolvida em Python, utilizando a biblioteca de análise léxica NLTK³. A criação do dicionário teve como base alguns *tweets* selecionados para estudo. Para cada *tweet*, as *stopwords* (artigos, pronomes, etc.) são removidas e às palavras restantes são atribuídas relevâncias de acordo com a frequência das mesmas. Para aquelas que possuem relação com o tema, atribui-se uma relevância positiva. Às palavras mais frequentes que não possuem relação com o tema e que remetem um novo sentido ao *tweet*, designa-se uma relevância negativa. Por fim, para palavras que não dizem respeito ao contexto da violência patrimonial e não interferem no sentido da mensagem define-se uma relevância neutra.

Considerando, por exemplo, o *tweet* “*Ontem à noite uma mulher foi assaltada no ponto de ônibus do meu lado*”, a análise o destacaria como resultado devido à ocorrência de palavras com alta relevância (“ponto de ônibus”). Por outro lado, o *tweet* “*Está na hora de assaltar a geladeira*” não seria retornado pelo DETECT devido ao termo “geladeira” estar muitas vezes associado à expressão popular *assaltar a geladeira* e, conseqüentemente, possuir baixa relevância.

Sendo assim, através de uma transformação no Kettle, é atribuída a relevância de cada *tweet* a partir do dicionário. A relevância total de um *tweet* é calculada a partir da soma das relevâncias de cada palavra. Caso o resultado seja positivo, o *tweet* é considerado relevante.

3.3 Localização de Tweets

Os *tweets* sem informações de geolocalização correspondem a maior parte de *tweets* gerados [Gonzalez et al. 2012]. Para identificar a origem dessas mensagens, foi implementada uma abordagem que se baseia em: (i) verificar se algum logradouro foi

¹ www.ondefuirobado.com.br

² <http://community.pentaho.com/projects/data-integration/>

³ <http://www.nltk.org/>

incluído no conteúdo do *tweet*; e (ii) identificar a localização do usuário que gerou o *tweet* através da obtenção de dados do perfil de usuário. Em (i), os *tweets* passam por uma análise léxica a partir de um conjunto de palavras-chave (ex. rua, avenida, estrada, etc). Mediante a obtenção do logradouro e da cidade do usuário, tal informação é analisada pelo Google Maps API para atribuir a geocodificação correspondente. Finalmente, os dados de latitude e longitude retornados pela Google Maps API são enviados para o módulo de visualização, que gera um mapa de calor relacionado com as áreas de maior incidência de *tweets* retornados pelo DETECT. Os mapas de calor foram implementados a partir de um script Python que utiliza a Google Maps API. Os resultados são gerados em um arquivo .html que inclui a exibição do mapa.

4. Prova de Conceito

A fim de avaliar as funcionalidades do DETECT, foi realizada uma extração de *tweets* entre os meses de setembro de 2015 e janeiro de 2016. Destes *tweets*, 210.881.326 não possuíam informações de geolocalização e 1.679.008 possuíam informações de geolocalização. O total de *tweets* extraídos no período mencionado foi de 212.560.334.

O sistema DETECT foi executado a partir de um ambiente construído sobre a plataforma de nuvem Microsoft Azure. Tal fato garantiu a estabilidade necessária durante a fase de coleta dos *tweets*. Para o protótipo utilizado neste experimento, foram implementados os seguintes módulos: módulo de extração, módulo de seleção, criação de dicionário e módulo de análise de relevância para *tweets* geolocalizados.

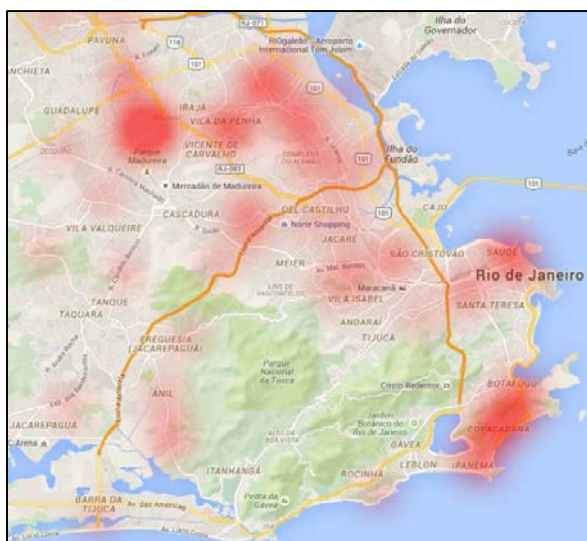


Figura 3: Mapa de calor refletindo tweets submetidos em áreas da cidade do Rio de Janeiro

O sistema detectou 569 tweets geolocalizados com relevância ao tema de violência patrimonial de acordo com o dicionário gerado. Dentre aqueles mais relevantes, a maioria foi gerada na cidade do Rio de Janeiro. Um mapa de calor com as principais áreas de geração destas mensagens no Rio de Janeiro é apresentado na Figura 3.

5. Considerações finais

Conforme descrito na seção anterior, DETECT retornou apenas *tweets* geolocalizados, o que explica o número diminuto do resultado (apenas 0,78% dos *tweets* extraídos possuíam informações de geolocalização). Tal fato ocorreu porque o módulo de identificação da

localização ainda não foi finalizado, logo, os trabalhos futuros incluirão também os *tweets* com localização a partir do conteúdo.

Além do desafio de prover a localização do tweet, outra questão importante é a redução dos “falsos positivos”, que são os *tweets* que não são relevantes ao tema, embora o sistema tenha detectado como relevante. *Tweets* que envolvem ironias, piadas ou erros de digitação podem induzir o sistema à geração de falsos positivos. Para resolver este problema, em trabalhos futuros, está prevista a implementação de um módulo de análise dos *tweets* utilizando aprendizado de máquina para identificar os padrões de *tweets* relacionados ao tema. Mesmo com uma faixa limitada de *tweets* para análise, foi possível perceber um índice maior de *tweets* em áreas nobres da cidade do Rio de Janeiro (como Ipanema e Copacabana) e também em comunidades na periferia da cidade, apresentando correspondência com os dados oficiais [ISP-RJ 2016].

Nos trabalhos futuros, os resultados serão contrapostos frente aos dados populacionais do IBGE. Além disso, será finalizada a implementação dos módulos restantes e desenvolvida uma versão para processamento distribuído do sistema DETECT para garantir um melhor desempenho diante do grande volume de dados.

Referências

- Cano Basave, A. E., He, Y., Liu, K. and Zhao, J. (out 2013). A weakly supervised Bayesian model for violence detection in social media. In 6th International Joint Conference on Natural Language Processing (IJCNLP)
- CPP/Insper (2013). Relatório da Pesquisa de Vitimização em São Paulo - 2003-2013. São Paulo.
- Gonzalez, R., Figueroa, G. and Chen, Y.-S. (2012). TweoLocator: A Non-intrusive Geographical Locator System for Twitter. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-Based Social Networks. , LBSN '12. ACM.
- ISP-RJ (2016). Incidências Criminais e Administrativas do Estado do Rio de Janeiro do ano de 2015. <http://www.isp.rj.gov.br/Conteudo.asp?ident=108>, [acessado em Abr 22].
- Li, R., Lei, K. H., Khadiwala, R. and Chang, K. C.-C. (apr 2012). TEDAS: A Twitter-based Event Detection and Analysis System. In 2012 IEEE 28th International Conference on Data Engineering (ICDE).
- Nisha (13 jul 2015). Top Ten Countries With Most Twitter Users. <http://www.perfectinsider.com/top-ten-countries-with-most-twitter-users-in-the-world/>, [acessado em Fev 10].
- Osborne, M., Moran, S., McCreadie, R., et al. (2014). Real-time detection, tracking, and monitoring of automatically discovered events in social media. Association for Computational Linguistics
- Twitter (2016). About Twitter. <https://about.twitter.com/company>, [acessado em Jan 26].

Inferindo o sexo de usuários de redes sociais utilizando o LIWC em português do Brasil

Luiz Antonio da Ponte Junior¹, Gustavo Paiva Guedes¹, Eduardo Bezerra¹

¹CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
Av. Maracanã, 229 - Rio de Janeiro - RJ - Brasil.

lapjunior@yahoo.com.br, {gustavo.guedes, eduardo.silva}@cefet-rj.br

Abstract. *This work presents preliminary results of an experimental evaluation about the inference of users gender in a Brazilian social network. This is done using a feature extraction process associated with those users. To achieve this goal, we use a Portuguese version of the linguistic feature called LIWC. Initial experimental results allow us to conclude that the classification task using datasets produced with LIWC is able to present satisfactory results. This occurs with no direct influence of words that manifest gender traits.*

Resumo. *Esse trabalho apresenta resultados preliminares de uma avaliação experimental sobre a possibilidade de se inferir o sexo dos usuários de uma rede social brasileira. Isso é feito a partir de um processo de extração de características associadas a esses usuários. Para alcançar esse objetivo, utilizamos uma versão em português do recurso linguístico denominado LIWC. Os resultados experimentais iniciais permitem concluir que a tarefa de classificação sobre os conjuntos de dados produzidos com o LIWC é capaz de apresentar resultados satisfatórios. Isso é feito sem influência direta das palavras que possuem traços de gênero.*

1. Introdução

O estudo das características de usuários de redes sociais online vem se expandindo nos últimos anos. Essa expansão se deve, em grande parte, ao entendimento de que os resultados desses estudos podem ser aplicáveis em áreas diversas como Marketing, Sociologia, Computação, e Linguística. Diversas redes sociais populares (e.g. Twitter, Facebook) têm tido seus dados coletados para serem estudados por pesquisadores de diversas áreas. Esses estudos adotam diferentes representações dos usuários, seja a partir de seus dados pessoais (e.g., idade, religião), regionais (e.g. local de nascimento, local de trabalho) ou mesmo a partir dos termos presentes em suas postagens. Alguns estudos indicam que é possível determinar os estados emocionais de um indivíduo a partir das palavras por ele utilizadas [Pennebaker 2013].

O objetivo deste trabalho é realizar uma avaliação experimental acerca da possibilidade de se inferir o sexo dos usuários de rede sociais a partir dos aspectos emocionais presentes em seus textos. Para isso, utilizamos o recurso linguístico denominado LIWC, que permite calcular o grau de uso de diferentes categorias de palavras. Esse recurso foi proposto por [Pennebaker et al. 2001]. Em particular, utilizamos neste trabalho a versão em português do LIWC [Filho et al. 2013]. Na avaliação experimental, utilizamos dados de usuários de uma rede social online brasileira denominada *Meu Querido Diário* (MQD).

O restante desse trabalho está organizado como segue. Na Seção 2, são apresentados os trabalhos relacionados. Na Seção 3, após uma descrição introdutória acerca do LIWC e do MQD, apresentamos o processo de extração de características dos usuários. Na Seção 4 são discutidos os resultados experimentais. Por fim, a Seção 5 expõe as conclusões e cenários futuros de expansão deste trabalho.

2. Trabalhos Relacionados

Existem trabalhos similares ao apresentado neste artigo que atendem à língua inglesa (e.g. [Peersman et al. 2011, Kokkos and Tzouramanis 2014]). Também podemos observar a existência de estudos capazes de inferir a idade e o sexo em textos em português do Brasil [Araújo et al. 2014]. Entretanto, nesse último, não foram considerados aspectos emocionais, cognitivos ou estruturais dos textos, o que caracteriza uma diferença e uma contribuição do presente trabalho.

O trabalho realizado em [Schler et al. 2006] identifica que o conteúdo dos textos escritos por homens e mulheres possui diferenças. Esse trabalho indica que essas diferenças podem ser identificadas não apenas por temas como política, dinheiro e vida pessoal, mas também pelos estilos de escrita, como o uso de pronomes, artigos e preposições. Essas diferenças presentes no conteúdo dos textos não delimitam apenas o sexo do usuário, mas também são observadas de acordo com faixas etárias.

O estudo apresentado em [Goswami et al. 2009] indica que também existe relação entre o tamanho médio dos textos, o tamanho médio de cada palavra de um texto e o uso de gírias com a faixa etária de grupos de usuários. Por meio de um conjunto montado para treinamento e do uso de um método de classificação, foi possível inferir a faixa etária com uma precisão de 90%. Foi identificado que, em grupos de usuários jovens, é predominante o uso de gírias e mensagens mais curtas.

O trabalho de [Filho et al. 2014] visa inferir o sexo e a idade por meio de textos de usuários coletados do Twitter. O procedimento consistiu na coleta dos últimos 200 textos de cada usuário, no pré-processamento dos textos e na classificação da relevância dos termos utilizados pelos usuários. Foram empregados 4 diferentes algoritmos de classificação (Naive Bayes, Naive Bayes Multinomial, SVM e Random Forest). Os autores utilizaram a medida de ganho de informação para selecionar os atributos mais relevantes. Os métodos propostos alcançaram bons resultados.

O estudo produzido em [Nguyen et al. 2013] destina-se a determinar a relação entre o uso da linguagem e a idade de usuários do Twitter. O trabalho analisou contas holandesas e constatou que entre o comportamento de pessoas jovens havia o maior uso da primeira pessoa e o alongamento de palavras. Por outro lado, o comportamento observado em pessoas mais velhas foi o maior uso de preposições e mensagens mais longas.

3. Inferência do sexo em usuários de redes sociais utilizando o LIWC

O objetivo dessa seção é descrever os procedimentos adotados para avaliar a adequação do dicionário da língua portuguesa do LIWC para inferir o sexo em usuários de redes sociais. Na Seção 3.1 apresentamos uma breve descrição do LIWC, em seguida, na Seção 3.2 detalhamos as características do conjunto de dados MQD1016. Na Seção 3.3 descrevemos o processo utilizado na extração de características dos usuários.

3.1. LIWC

O LIWC (*Linguistic Inquiry and Word Count*) é uma ferramenta proposta em [Pennebaker et al. 2001] e tem o objetivo de analisar os componentes emocionais, cognitivos e estruturais de textos. Isso é feito com base em um dicionário de palavras, que pode ser encontrado em diversas línguas. Nesse trabalho utilizamos o dicionário do LIWC para o português do Brasil, encontrado em [Filho et al. 2013].

O dicionário do LIWC em português possui 127.149 palavras, em que cada uma pode ser assinalada a uma ou mais categorias. Essas categorias representam perspectivas linguísticas, psicológicas, dentre outras. O dicionário é composto por um total de 64 categorias (e.g. *posemo*, *pronoun*, *time*). Com isso, cada palavra presente no texto dos usuários pode ser expressa por diferentes perspectivas (categorias).

Diversos estudos existentes utilizam a versão inglesa do LIWC, dentre eles pode-se destacar [Golbeck et al. 2011] e [Schwartz et al. 2013]. O primeiro tem o objetivo de auxiliar na inferência da personalidade de usuários do twitter. O último tem estuda a correlação entre comportamento e personalidade. Embora haja diversos trabalhos com a versão inglesa do LIWC, pouco se encontra sobre o LIWC em português.

3.2. A Rede Social MQD

A rede social MQD (<http://www.meuqueridodiario.com.br>) foi desenvolvida em 2009 e funciona como um diário online. Seus usuários escrevem entradas registrando o que fizeram no seu dia-a-dia, descrevendo seus sentimentos e emoções em forma de texto. Desse modo, as palavras utilizadas podem refletir os estados psicológicos e emocionais dos indivíduos que as escrevem [Pennebaker 2013].

O conjunto de dados MQD1016 foi criado em 2016 com textos de usuários do MQD. Ele é composto por textos de 510 usuários do sexo feminino e 506 usuários do sexo masculino. As palavras presentes nesse conjunto de dados foram preprocessadas de forma que apenas as palavras existentes na língua portuguesa fossem utilizadas. Assim, foram removidas palavras escritas de forma incorreta, como por exemplo as palavras com repetições de letras (e.g. *felizzzz*, *Oiiii*). Esse conjunto de dados contém 7480 atributos (palavras diferentes).

3.3. Extração das características dos usuários

Neste trabalho utilizamos o dicionário de palavras do LIWC em português [Filho et al. 2013] para filtrar as palavras não existentes na língua portuguesa. Com isso, o conjunto de dados MQD1016 possui, para cada usuário, o conjunto de palavras utilizadas em seus textos que existiam no dicionário do LIWC. Vale ressaltar que todas as palavras que continham letras maiúsculas foram substituídas pela mesma palavra em caixa-baixa.

Nesse conjunto de dados, cada usuário é representado como um vetor de p posições, em que p é o tamanho do conjunto de palavras formado pelos textos de todos os usuários. Cada posição do vetor representante de um usuário contém o número de vezes que a palavra no índice x_i ocorreu nos textos desse usuário. Assim, todos os usuários possuem um vetor de tamanho p .

Primeiramente, utilizamos o MQD1016 para a criação de um novo conjunto de dados denominado MQD1016-LIWC-PT, no qual cada usuário é representado por um

vetor de 64 posições. As posições do vetor de cada usuário representam a contagem de palavras (em seus textos) relativas a cada categoria do dicionário de dados do LIWC. Para produzir o vetor de um usuário u , as palavras utilizadas nos textos do respectivo usuário é pesquisada no dicionário do LIWC em português. Isso retorna, para cada palavra, um conjunto de n categorias $1 < n < m$ em que n é o número de categorias de cada palavra e m é o número total de categorias, no caso do LIWC, 64. Em seguida, as posições (no vetor) de cada categoria retornada (x_j) são incrementadas.

A Figura 1 ilustra um vetor de palavras representando um usuário do conjunto de dados MQD1016-LIWC-PT. Cada posição do vetor reflete o número de palavras em cada uma das 64 categorias do LIWC (x_i). Por exemplo, pode-se notar que 10 termos se enquadraram na categoria x_1 . Todos os usuários são representados segundo esse modelo.

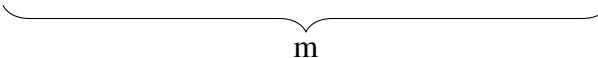
x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	...	x_{63}
1	10	13	7	5	21	3	9	...	11
									

Figura 1. Vetor representando um usuário utilizando as categorias do LIWC.

Em seguida, foi gerado um terceiro conjunto de dados, denominado MQD1016-LIWC-PT-20p. Para gerar esse conjunto de dados, foram selecionadas 20% das 64 categorias seguindo o teorema de Pareto [Koch 1999]. Nesse cenário, foram selecionadas as 13 categorias que apresentavam maior ganho de informação (*information gain*). As categorias selecionadas foram: pronoun, ppron, i, preps, humans, negemo, ingest, relativ, space, time, leisure, money, relig, class.

Os três conjuntos de dados produzidos (i.e. MQD2016, MQD2016-LIWC-PT e MQD2016-LIWC-PT-20p) foram empregados para produzir modelos de classificação. Em seguida, esses modelos foram utilizados para inferência do sexo. Os experimentos foram executados utilizando a técnica de validação cruzada denominada *k-fold validation* com dez partições. A média *F1-score* foi utilizada para a avaliação dos resultados obtidos. Esses resultados são apresentados na Seção 4.

4. Resultados experimentais

Essa seção apresenta os resultados experimentais encontrados na inferência do sexo de usuários de redes sociais utilizando o LIWC em português do Brasil. Para geração desses resultados, utilizamos algoritmos de classificação conhecidos na literatura: ZeroR, Random Forest (RF), Naive Bayes (NB), NB Multinomial e SMO.

A Tabela 1 descreve os resultados obtidos. Os valores em negrito indicam o conjunto de dados que obteve o melhor F1 para os algoritmos apresentados. O algoritmo ZeroR foi utilizado como *baseline*. Pode-se notar que os algoritmos Random Forest RF, NB e NB Multinomial apresentaram melhor F1 para o conjunto de dados MQD1016. O algoritmo SMO apresenta melhores resultados para o conjunto de dados MQD1016-LIWC-PT-20p.

É importante ressaltar que, embora os algoritmos RF, NB e NB Multinomial apresentem F1 inferiores nos conjuntos de dados MQD1016-LIWC-PT e

Tabela 1. Classificação de sexo - Média F1

	ZeroR	RF	NB	NB Multinomial	SMO
MQD1016	0.336	0.645	0.611	0.656	0.602
MQD1016-LIWC-PT	0.336	0.574	0.565	0.566	0.598
MQD1016-LIWC-PT-20p	0.336	0.599	0.582	0.565	0.607

MQD1016-LIWC-PT-20p, muitas palavras do conjunto de dados MQD1016 são flexionadas em gênero, o que pode influenciar o resultado da classificação, Essa característica pode ser observada ao selecionarmos as palavras que apresentam maior ganho de informação. As 10 palavras que produziram maior ganho de informação no conjunto de dados MQD1016 são: ele, cansado, ela, para, sozinha, dele, eu, meu, mais, menino. Pode-se notar três adjetivos flexionados em gênero (i.e. cansado, sozinha, menino), dois pronomes possessivos que possuem traços de marcação de gênero (i.e. dele, meu) e dois pronomes pessoais que também apresentam traços de marcação de gênero (i.e. ele, ela). Apenas três palavras não possuem informações de gênero (i.e. para, eu, mais).

Nesse cenário, fica evidente que a tarefa de classificação nos conjuntos de dados produzidos com o LIWC é capaz de apresentar resultados satisfatórios, sem que haja influência direta das palavras que possuem traços de gênero. Assim, o procedimento de extração de características aqui proposto é relevante, visto que, um dos algoritmos utilizados (SMO) apresentou melhores resultados para o conjunto de dados MQD1016-LIWC-PT-20p. Essa evidência fica mais clara quando removemos as 7 palavras com traços de gênero (i.e. cansado, sozinha, menino, dele, meu, ele, ela) do conjunto de dados MQD1016, o que totaliza uma remoção de menos de 0.01% dos 7480 atributos (palavras). Ao executar novamente o algoritmo que apresentou o melhor resultado para esse conjunto de dados (i.e. NB Multinomial), houve uma perda de 3% na média F1.

5. Conclusões e trabalhos futuros

Nesse trabalho, apresentamos uma avaliação da inferência do sexo de usuários de redes sociais utilizando o dicionário de dados do LIWC para o português do Brasil. Essa avaliação tem o objetivo de contribuir na lacuna de trabalhos que envolvem o dicionário mencionado, visto que sua versão inicial foi proposta recentemente em [Filho et al. 2013].

Para isso, produzimos três conjuntos de dados utilizando textos provenientes de uma rede social brasileira. Avaliamos esses três conjuntos de dados com quatro algoritmos: RF, NB, NB Multinomial e SMO. De forma geral, os resultados obtidos foram satisfatórios. Vale ressaltar que para o caso do algoritmo de classificação SMO, os melhores resultados foram alcançados com o conjunto de dados MQD1016-LIWC-PT-20p. Esses resultados podem servir como base para trabalhos que realizem classificação de textos em português do Brasil. Os resultados preliminares alcançados indicam bons resultados.

É interessante ressaltar que alguns trabalhos que realizam a inferência de sexo em textos são sensíveis a palavras flexionadas em gênero. Assim, palavras como *cansada*, *obrigado*, *obrigada*, poderiam ter forte influência na inferência do sexo. Essas palavras são evidenciadas em [Araújo et al. 2014] como parte de um grupo de 20 palavras (atributos) que possuem maior ganho de informação na separação entre as duas classes do sexo dos usuários. Por outro lado, não existe no LIWC uma classe sensível ao gênero. As-

sim, o que define os resultados apresentados no decorrer desse trabalho são características de preocupações pessoais, processos linguísticos, psicológicos, etc.

Durante o desenvolvimento desse trabalho, emergiram algumas ideias para trabalhos futuros, dentre elas, a inferência da idade de usuários utilizando o dicionário do LIWC em português do Brasil. Além disso, seria interessante que alguns esforços fossem investidos para o estudo da inferência da personalidade de usuários utilizando esse mesmo dicionário. Conforme mencionado, já existem trabalhos nesse âmbito utilizando o dicionário do LIWC em inglês. Vale ressaltar que, como trabalho futuro, iremos realizar a comparação com outras abordagens da literatura.

Referências

- [Araújo et al. 2014] Araújo, M., Gonçalves, P., and Benevuto, F. (2014). Métodos para análise de sentimentos no twitter. In *CSBC 2014 - BraSNAM* ().
- [Filho et al. 2013] Filho, P. P. B., Pardo, T. A. S., and Aluísio, R. M. (2013). An evaluation of the brazilian portuguese liwc dictionary for sentiment analysis.
- [Filho et al. 2014] Filho, R. M., Carvalho, A. I. R., and Pappa, G. L. (2014). Inferência de sexo e idade de usuários no twitter. In *CSBC 2014 - BraSNAM* ().
- [Golbeck et al. 2011] Golbeck, J., Robles, C., Edmondson, M., and Turner, K. (2011). Predicting personality from twitter. In *SocialCom/PASSAT*, pages 149–156. IEEE.
- [Goswami et al. 2009] Goswami, S., Sarkar, S., and Rustagi, M. (2009). Stylometric analysis of bloggers’ age and gender. In *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, California, USA, May 17-20, 2009*.
- [Koch 1999] Koch, R. (1999). *The 80/20 Principle: The Secret of Achieving More with Less*. A Currency book. Doubleday.
- [Kokkos and Tzouramanis 2014] Kokkos, A. and Tzouramanis, T. (2014). A robust gender inference model for online social networks and its application to linkedin and twitter. *First Monday*, 19(9).
- [Nguyen et al. 2013] Nguyen, D., Gravel, R., Trieschnigg, D., and Meder, T. (2013). “how old do you think i am?”: A study of language and age in twitter. In *Proceedings of the Seventh International AAI Conference on Weblogs and Social Media, ICWSM 2013*.
- [Peersman et al. 2011] Peersman, C., Daelemans, W., and Van Vaerenbergh, L. (2011). Predicting age and gender in online social networks. In *Proceedings of the 3rd International Workshop on Search and Mining User-generated Contents, SMUC ’11*, pages 37–44, New York, NY, USA. ACM.
- [Pennebaker 2013] Pennebaker, J. (2013). *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury USA.
- [Pennebaker et al. 2001] Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). *Linguistic Inquiry and Word Count*. Lawrence Erlbaum Associates, Mahwah, NJ.
- [Schler et al. 2006] Schler, J., Koppel, M., Argamon, S., and Pennebaker, J. (2006). Effects of Age and Gender on Blogging. In *Proc. of AAI Spring Symposium on Computational Approaches for Analyzing Weblogs*.
- [Schwartz et al. 2013] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., and Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, 8(9):e73791.

BreSci - 10° Brazilian e-Science Workshop

Apresentação

Nas últimas décadas, tem havido uma revolução no modo como a ciência e a engenharia têm sido conduzidas, ao se utilizar de forma intensiva as tecnologias de informação e comunicação (TIC). Essa nova forma de realizar a ciência, denominada de e-Science ou e-Ciência, desempenha hoje um papel fundamental na metodologia de trabalho adotada por muitos grupos de pesquisa em todo o mundo.

O X BreSci tem como objetivo colaborar com os esforços de e-Science propondo um fórum de discussão sobre temas relevantes dessa área de estudo. Além da trilha principal, que tem um escopo mais amplo e mais relacionado com as TIC, contamos também com uma trilha de aplicações específica para discutir temas relacionados às áreas particulares de aplicação da e-Science, incluindo (mas não sendo restrito a) bioinformática, astronomia, química, e biodiversidade.

A aproximação com pesquisadores dessas áreas da ciência visa a estreitar o relacionamento entre os participantes das diversas áreas. Ademais, essa aproximação propicia a identificação de demandas relativas à infraestrutura computacional sob o ponto de vista das áreas fim. De outro lado, essa colaboração também propicia às áreas fim uma melhor difusão das soluções elaboradas pela comunidade de computação.

O BreSci 2016, em sua décima edição, sendo a sétima edição colocada no CSBC, manteve o nome de BreSci e manteve o formato do evento definido em 2012, incluindo a trilha principal e trilhas específicas para as áreas de bioinformática, astronomia, química, e informática na biodiversidade. Foram submetidos 21 trabalhos para trilha principal (6 no total para as trilhas de aplicações), sendo 9 aceitos para publicação como trabalhos completos. Ademais, nas trilhas de bioinformática e astronomia foram aceitos, respectivamente, 1 e 2 resumos. As trilhas de química e informática na biodiversidade não tiveram artigos aceitos.

Kary A. D. C. S. Ocaña (Laboratório Nacional de Computação Científica - LNCC)

Regina Braga (Universidade Federal de Juiz de Fora - UFJF)

Comitê de Organização

Coordenação do Workshop

Kary A. D. C. S. Ocaña (LNCC)

Regina Braga (UFJF)

Coordenador Local

Avelino Francisco Zorzo (PUCRS)

Comitê Diretivo

Kary Ocaña – LNCC (co-chair 2016)

Regina Maciel Braga – UFJF (co-chair 2016)

Luiz M. R. Gadelha Jr. – LNCC (co-chair 2015)

Sergio Manuel Serra da Cruz – UFRRJ (co-chair 2015)

Antônio Tadeu A. Gomes – LNCC (co-chair 2014)

Francisco Brasileiro – UFCG (co-chair 2013)

Comitê de Programa

Alberto Krone-Martins (Universidade de Lisboa)

Alcione Oliveira (UFV)

Alexandre Lima (COPPE/UFRJ)

Ana Maria de Carvalho Moura (LNCC)

Andrey Brito (UFCG)

Antônio Basilio (FIOCRUZ)

Antonio Saraiva (USP)

Antônio Tadeu Azevedo Gomes (LNCC)

Bruno Schulze (LNCC)

Carla Osthoff (LNCC)

Carolina Guimarães (FIOCRUZ)

Cristina Boeres (UFF)

Daniel de Oliveira (UFF)

Debora Drucker (EMBRAPA)

Diogo Tschoeke (UFRJ)

Duncan Ruiz (PUC-RS)

Eduardo Bezerra (CEFET/RJ)

Eduardo Dalcin (Inst. Pesq. Jardim Botânico)

Eduardo Ogasawara (CEFET/RJ)

Esther Pacitti (INRIA & LIRMM)

Fabiano Thompson (UFRJ)

Fábio Custódio (LNCC)

Fabio Andre Porto (LNCC)

Fernanda Baião (UNIRIO)

Fernanda Campos (UFJF)

Francisco Brasileiro (UFCG)

Gilberto Z. Pastorello (Universidade de Alberta)
 Glauber Wagner (UFSC)
 Guilherme Loss (LNCC)
 Hubert Stassen (UFRGS)
 J. Miguel Ortega (ICB-UFMG)
 Joao Setubal (USP)
 João Eduardo Ferreira (IME/USP)
 Jonas Dias (EMC Corporation)
 José Antonio Macêdo (UFC)
 José Maria David (UFJF)
 Kaline Coutinho (USP)
 Kary Ocaña (LNCC), co-chair
 Kele Belloze (CEFET/RJ)
 Laurent Dardenne (LNCC)
 Leonardo Azevedo (IBM/UNIRIO)
 Leonardo Murta (UFF)
 Lucia Drummond (UFF)
 Luciano Digiampietri (USP)
 Luiz M. R. Gadelha Jr. (LNCC)
 Marcelo de Macedo Brígido (UnB)
 Marcia Chame (FIOCRUZ)
 Marco Antônio Pereira Araújo (UFJF)
 Marcos Catanho (FIOCRUZ)
 Maria Claudia Cavalcanti (IME/RJ)
 Maria Emilia Walter (UnB)
 Marinez F. de Siqueira (JBRJ)
 Marta Mattoso (COPPE/UFRJ)
 Pedro Correa (USP)
 Rafael Loyola (UFG)
 Rafaelli Coutinho (CEFET/RJ)
 Regina Braga (UFJF), co-chair
 Reinaldo de Carvalho (INPE)
 Ricardo Ogando (ON-MCTI & LIneA)
 Ricardo Torres (Unicamp)
 Rita de Almeida (UFRGS)
 Ronaldo Goldschmidt (IME)
 Sandra Fabbri (UFSCar)
 Sergio Lifschitz (PUC-Rio)
 Sergio Manuel Serra da Cruz (UFRRJ)
 Silvia Olabariaga (University of Amsterdam)
 Tainá Raiol Alencar (FIOCRUZ)
 Vanessa Braganholo (UFF)
 Victor Stroelle (UFJF)
 Wagner Arbex (EMBRAPA)
 Vinod Rebello (UFF)

Palestras

Dados de Proveniência: não comece sua análise sem eles

Profa. Marta Lima de Queirós Mattoso (COPPE/UFRJ)

O BreSci faz 10 dez anos!!! Seu fórum de apresentação de artigos de apoio computacional ao desenvolvimento da ciência é a própria tradução do tema do CSBC 2016: #Computação e Interdisciplinaridade.

Nesta palestra apresentaremos uma evolução do papel da proveniência em experimentos científicos. Dados de proveniência registram como o de fluxo de dados resultante das simulações computacionais foi produzido, o que é essencial para se fazer um experimento reprodutível e confiável. Daremos ênfase ao encadeamento de transformações de dados científicos abordando soluções para: geração, rastreamento, adaptação e análise de dados. Mais especificamente, serão discutidos os desafios em prover dados que apoiem o cientista na concepção e reconfiguração do experimento durante a sua condução em ambientes de alto desempenho computacional. Serão apresentados casos reais de aplicações em bioinformática com execução paralela em nuvens e aplicações em geofísica em supercomputadores.

Trabalhos aceitos

Trabalhos Completos

Técnica de Clusterização não-hierárquica Aplicada para a Caracterização in silico de Promotores Associados a Genes de Choque Térmico de Escherichia coli

Gabriel Dall'Alba (University of Caxias do Sul - Brazil),
Scheila de Avila e Silva (Universidade de Caxias do Sul - Brazil)

Comparing Provenance Data Models for Scientific Workflows: an Analysis of PROV-Wf and ProvOne

Wellington Oliveira (Universidade Federal Fluminense - Brazil)
Paolo Missier (Newcastle University - Great Britain)
Daniel de Oliveira (Universidade Federal Fluminense - Brazil)
Vanessa Braganholo (UFF - Brazil)

Verificação da Reprodução de Workflows Científicos por meio de Algoritmos de Detecção de Plágio

Filipe Santiago (UFF - Brazil),
Daniel de Oliveira (Universidade Federal Fluminense - Brazil)

E-SECO ProVersion: Manutenção e Evolução de Experimentos Científicos

Tassio Ferenzini Martins Sirqueira (IF Sudeste MG - Brazil)
Humberto Dalpra (UFJF - Brazil)
Regina Braga (Universidade Federal de Juiz de Fora - Brazil)
Marco Antonio Araujo (UFJF / IF Sudeste MG - Brazil)
José Maria David (Universidade Federal de Juiz de Fora - Brazil)
Fernanda Campos (Universidade Federal de Juiz de Fora - Brazil)

Ciência Cidadã Baseada em Big Data Aplicada ao Planejamento Urbano

Eliza Gomes (Universidade Federal de Santa Catarina - Brazil)
Mario Dantas (UFSC - Brazil)
Douglas Macedo (UFSC - Brazil)
Carlos Roberto De Rolt (Universidade do Estado de Santa Catarina - Brazil)
Marcelo Brocardo (Universidade do Estado de Santa Catarina - Brazil)
Luca Foschini (University of Bologna - Italy)

Workflows Científicos com Apoio de Bases de Conhecimento em Tempo Real

Victor Bursztyrn (Federal University of Rio de Janeiro - Brazil)
Jonas Dias (EMC Corporation - Brazil)
Marta Mattoso (COPPE/UFRJ - Brazil)

Integrating Scientific Workflows with Scientific Gateways: A Bioinformatics Experiment in the Brazilian National High-Performance Computing Network

Maria Luiza Mondelli (LNCC - Brazil)
Marcelo Galheigo (LNCC - Brazil)
Vívian Medeiros (LNCC - Brazil)
Bruno Fernandes Bastos (LNCC - Brazil)
Antonio Tadeu Gomes (LNCC - Brazil)
Marta Mattoso (COPPE/UFRJ - Brazil)
Ana Tereza Vasconcelos (LNCC - Brazil)
Luiz Manoel Rocha Gadelha Júnior (LNCC - Brazil)

BioDSL: A Domain-Specific Language for Mapping and Dissemination of Biodiversity Data in the LOD

Kleberson Serique (ICMC/USP - Brazil)
José Campos dos Santos (Instituto Nacional de Pesquisas da Amazônia - Brazil)
Dilvan Moreira (ICMC-USP - Brazil)

Modelagem e Análise de Conformidade do Processo de Atracamento Molecular

Miller Biazus (Universidade Federal do Rio Grande do Sul - Brazil),
Lucineia Heloisa Thom (Federal University of Rio Grande do Sul - Brazil),
Marcio Dorn (Universidade Federal do Rio Grande do Sul - Brazil)

Resumos

Laboratório Interinstitucional de e-Astronomia: Passado, Presente, e Futuro

Luiz da Costa (LIneA & ON - Brazil)
Angelo Fausti Neto (LIneA, LSST - USA)
Marcio Maia (Observatorio Nacional - Brazil)
Ricardo Ogando (ON, LIneA - Brazil)
Riccardo Campisano (LIneA, CEFET/RJ - Brazil)

Avaliação do Desempenho Relativo de Bancos de Dados NoSQL para a Arquivos de Genótipos

Wagner Arbex (Empresa Brasileira de Pesquisa Agropecuária - Embrapa - Brazil)
Vinícius Schettino (Universidade Federal de Juiz de Fora - UFJF - Brazil)
Arthur Almeida (Universidade Federal de Juiz de Fora - UFJF - Brazil)
Leojayme Silva (Universidade Federal de Juiz de Fora - UFJF - Brazil)

Data Release Interface para o Dark Energy Survey

Luiz da Costa (LIneA & ON - Brazil)
Angelo Fausti Neto (Observatório Nacional - Brazil)
Ricardo Ogando (ON, LIneA - Brazil)
Marcio Maia (Observatório Nacional - Brazil)
Riccardo Campisano (LIneA, CEFET/RJ - Brazil)

Técnica de Clusterização não-hierárquica aplicada para a caracterização *in silico* de promotores associados a genes de choque térmico de *Escherichia coli*

Gabriel Dall'Alba¹, Scheila de Avila e Silva¹

¹ Instituto de Biotecnologia – Universidade de Caxias do Sul (UCS)
Rua Francisco Getúlio Vargas, 1130– 95070-560 – Caxias do Sul – RS – Brazil

{gdalba@ucs.br sasilva6@ucs.br}

Abstract. *Computational techniques play an important role in the post-genomic era, due to the amount of biological data generated and released. In this context, this paper carry out an in silico analysis of σ_{24} and σ_{32} -dependent promoter sequences. It was made by using the clustering technique with stability values as input data. The content of the clusters was analyzed by average purity obtained for the clusters. In general, all the clusters presented 63% of purity. However, the clusters 7, 10 and 11 obtained, respectively, 85%, 82% and 83% of purity. As conclusions, was possible to identify degrees of degeneration in the sequences and grouping features. After all, this paper contributes to the comprehension of the different biological bacterial promoter's profiles. Furthermore, these results can be applied in the reducing of false positives of in silico promoter predicting tools.*

Resumo. *As técnicas computacionais têm importante papel na era pós-genômica, devido ao crescente número de dados biológicos gerados. Um campo bem desenvolvido envolve as questões relacionadas a regulação gênica. Com base no exposto, este trabalho dedica-se a uma análise in silico dos promotores relacionados aos fatores σ_{24} e σ_{32} da bactéria *E. coli*. Para isso, foi utilizada a técnica de Clusterização e a codificação dos dados nos valores de estabilidade. Os agrupamentos foram analisados com base na pureza obtida. A pureza média obtida foi de 63%, com destaque para os agrupamentos 7, 10 e 11, que obtiveram uma pureza média de 85%, 82% e 83% respectivamente. Foi possível observar diferentes graus de degeneração das sequências e as características que os grupos estudados apresentam. Deste modo, contribui-se para a compreensão dos diferentes perfis biológicos encontrados nos promotores bacterianos. Além disso, os resultados podem auxiliar na redução de falsos positivos em ferramentas de predição de promotores.*

1. Introdução

A era pós-genômica apresenta uma série de desafios para as diferentes áreas de pesquisa, uma vez que a quantidade de dados biológicos chega, atualmente, aos Petabytes [Marx 2013]. Portanto, há a necessidade de aplicar diferentes técnicas que auxiliem no armazenamento e acesso destas quantidades de dados, procurando reduzir o espaço entre a geração de dados e sua análise [Kanehisa et al. 2014]. Neste contexto, a tecnologia computacional permite o manejo de dados e geração de inferências nos diversos segmentos biológicos, pois permite a análise de grandes quantidades de dados, possibilitando o cruzamento de informações e extração de informações que não seriam

possíveis sem o auxílio destas técnicas [Attwood et al. 2011]. Dentro deste contexto, técnicas de aprendizado de máquina, como SVM (*support vector machine*), Redes Neurais Artificiais e Clusterização são aplicadas em diferentes áreas das ciências da vida, como a genômica, proteômica, metabolômica e regulação gênica [de Avila e Silva e Echeverrigaray 2012] [Callebaut 2012].

A regulação da transcrição gênica em seres procariotos desempenha um papel importante na resposta adequada destes seres às mudanças em seu ambiente, possibilitando sua sobrevivência em determinadas condições. A presença de mecanismos de proteção do genoma e formas de seletividade de expressão gênica (a qual ocorre, principalmente, no momento da transcrição) são reflexos da ausência de núcleo [Krebs, Goldstein e Kilpatrick 2014].

O processo de expressão de genes consiste em várias etapas, iniciadas com a interação da enzima chamada RNA polimerase (RNAP) com segmentos de DNA chamados de promotores, os quais antecedem a região codificante. Esta interação possui um valor essencial, uma vez que o reconhecimento dos promotores é fundamental para a transcrição do gene associado a ele. Deste modo, o promotor pode ser considerado um elemento regulatório da expressão gênica [Krebs, Goldstein e Kilpatrick 2014]. No processo de identificação de um promotor, a RNAP possui uma subunidade chamada sigma (σ), a qual auxiliará na identificação de um promotor específico. Por exemplo, em *Escherichia coli*, os fatores σ_{24} e σ_{32} são responsáveis pela expressão de genes relacionados ao estresse por choque térmico, já o σ_{58} possui função relacionada à assimilação de nitrogênio.

O reconhecimento de um determinado promotor se dá em regiões específicas da sequência, denominadas motivos consensuais. Estes motivos estão localizados em duas regiões distintas do promotor, chamadas de região -10 e região -35, em referência ao primeiro nucleotídeo transcrito, que recebe a numeração +1. Ambas as regiões possuem função distinta, a região -35 atua como sinal para o seu reconhecimento pela RNAP, enquanto a região -10 situa-se na região que ocorre a abertura da fita de DNA [Krebs, Goldstein e Kilpatrick 2014] [De Avila e Silva e Echeverrigaray 2012]. Apesar de levarem o nome de “sequências consenso”, o grau de conservação varia tanto entre promotores de um fator σ como entre sequências reconhecidas por diferentes fatores σ . Deste modo, evidencia-se a dificuldade em executar uma análise global dos dados a partir deste critério. As divergências observáveis entre os motivos consensuais justificam seu reconhecimento por diferentes fatores σ , considerando as peculiaridades de cada grupo de sequências. Na tabela 1, é possível visualizar a composição de nucleotídeos dos motivos consensuais para cada fator σ em *Escherichia coli*.

Tabela 1. Composição de nucleotídeos dos motivos consensuais para os diferentes fatores σ em *Escherichia coli*. Onde “pb” representa a distância em pares de base.

Fator σ	Função	Consenso -35 / -10
24	Estresse por choque térmico	GGAAGT 15 pb GTCTAA
28	Mobilidade celular e patogenicidade	CTAAA 15 pb GCCGATAA
32	Estresse por choque térmico	CCCTTGAA 13-15 pb CCCGATNT
38	Resposta a estresse	TTGACA 16-18 pb TATACT

54	Assimilação de Nitrogênio	CTGGNA 16-18 pb TTGCA
70	Sigma constitutivo	TTGACA 16-18 pb TATAAT

Fonte: DE AVILA E SILVA E ECHEVERRIGARAY, 2012.

Além do conteúdo de nucleotídeos, os promotores também possuem características estruturais próprias, como a estabilidade e a curvatura, oferecendo a possibilidade de diferentes abordagens para o estudo dos promotores [Kanhere e Bansal 2005]. Trabalhos previamente publicados já demonstraram que as regiões promotoras são menos estáveis que as regiões gênicas [Kanhere e Bansal 2005] [Ramprakash e Schwarz 2007] [Jáuregui *et al.* 2003]. Deste modo, justifica-se o potencial do uso desta característica como critério de classificação das sequências promotoras. Uma metodologia desenvolvida em Kanhere e Bansal (2005) e aprimorada por Rangannan e Bansal (2007) mostra, em seus resultados, que a estabilidade apresenta-se como uma medida mais eficaz que os motivos conservados para diferenciar regiões promotoras e não-promotoras [Rangannan e Bansal 2007]. A metodologia é baseada nas diferenças de estabilidade (ΔG^0) entre as regiões promotoras e codificantes.

Com base no que foi apresentado, este trabalho dedica-se a análise *in silico* (computacional) dos promotores reconhecidos pelos fatores σ_{24} e σ_{32} em *Escherichia coli*, aplicando a técnica de aprendizado de máquina denominada clusterização e utilizando o critério de estabilidade para a codificação dos dados. A escolha destes dois fatores σ deve-se ao fato de ambos estarem descritos com a mesma função (estresse por choque térmico). O estresse por choque térmico pode alterar alguns processos biológicos da bactéria como, por exemplo, o envelopamento, fixação de proteínas e alterações na forma e estrutura do DNA [Lim *et al.* 2013].

2. Metodologia

Os dois conjuntos de sequências promotoras da bactéria *E. coli* foram obtidos no banco de dados RegulonDB [Salgado *et al.* 2013]: (i) 521 sequências reconhecidas pelo σ_{24} ; (ii) 324 sequências reconhecidas pelo σ_{32} . A estabilidade do DNA pode ser expressada através de sua energia livre (ΔG), a qual depende da composição de mononucleotídeos e dinucleotídeos. Portanto, a estabilidade do DNA pode ser predita a partir de sua sequência e das interações com cada vizinho mais próximo. A contribuição de cada dinucleotídeo está descrito por SantaLucia e Hicks (2004). Para aplicação da técnica utilizando a energia livre (ΔG), ΔG foi calculada utilizando a seguinte equação, conforme SantaLucia e Hicks (2004) e Kanhere e Bansal (2005).

$$\Delta G^0 = \Delta G_{ij} \quad (1)$$

Onde ΔG^0_{ij} é a variação padrão de energia livre para os dinucleotídeos de tipo *ij*. A equação original, descrita por Kanhere e Bansal (2005), foi adequada para os objetivos propostos neste trabalho. Os valores foram obtidos por meio da técnica de janela deslizante (move window), passando por uma janela de um nucleotídeo por vez. Tanto a preparação dos dados, quanto a ferramenta de execução do algoritmo K-means foram implementadas pelos autores em linguagem de programação python e C#.

Após esta etapa, as sequências estão adequadas para a aplicação da técnica computacional de clusterização, uma técnica de classificação não supervisionada que pode ser dividida em dois métodos: o hierárquico e o não-hierárquico. O algoritmo *K-Means* encontra-se dentro da categoria não-hierárquica, sendo capaz de separar um conjunto de dados em agrupamentos de acordo com um critério de distância predefinido, sendo o critério de cálculo de distância Euclidiano o mais usado. O algoritmo requer um valor numérico K (atribuído pelo usuário da ferramenta) equivalente ao número de agrupamentos desejados.

A técnica de clusterização apresenta-se eficaz para o reconhecimento de padrões escondidos dentro de um conjunto de dados. Apesar de ser um método simples e efetivo, cada cluster possui uma sensibilidade apurada quanto ao seu centroide inicial, ou seja, arranjos totalmente diferentes podem surgir a partir de uma nova escolha randômica do centroide inicial [Witten e Frank 2005].

Devido à metodologia para obtenção do número ótimo de K ser empírica, as simulações foram realizadas utilizando valores de $K = 8$ até 20, conforme trabalhos prévios realizados pelos autores, os quais estão em processo de publicação. Os agrupamentos foram analisados em relação à sua pureza, ou seja, quantas sequências de um mesmo fator σ foram agrupados em um mesmo cluster. Posterior a esta etapa, foi utilizada a ferramenta WebLogo [Crooks 2004] para visualização dos motivos consensuais encontrados em cada agrupamento, a qual está disponível on-line.

3. Resultados e Discussão

3.1 Agrupamentos resultantes

Com a análise dos resultados obtidos, foi possível verificar que a simulação com $K = 12$ obteve uma pureza média de 63%. A pureza de cada agrupamento gerado, pode ser observada na figura 1.

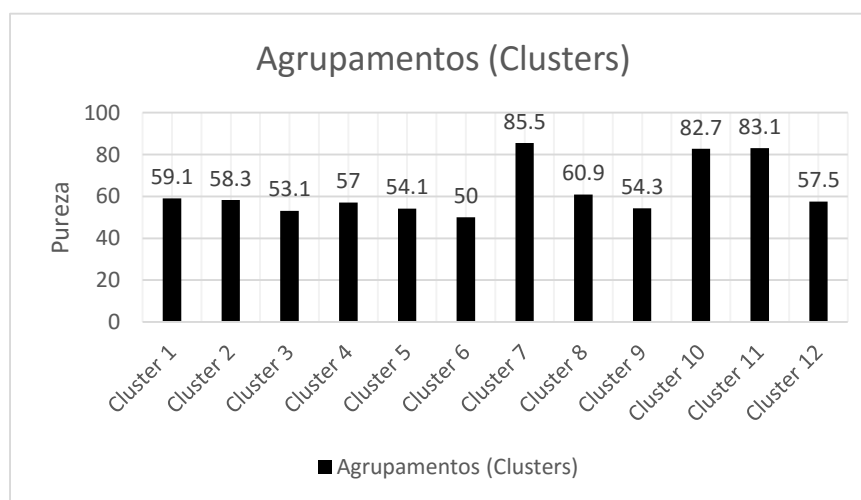


Figura 1. Pureza dos agrupamentos gerados na simulação com valor de $K = 12$.

Com base na pureza obtida, uma investigação mais aprofundada foi realizada somente para os clusters que obtiveram uma pureza média acima de 80%, ou seja, os clusters 7, 10 e 11, os quais tiveram predominância do σ_{24} . Com o auxílio da ferramenta WebLogo [Crooks 2004], foi possível visualizar os seguintes motivos

consensuais para cada um dos clusters, conforme a figura 2.

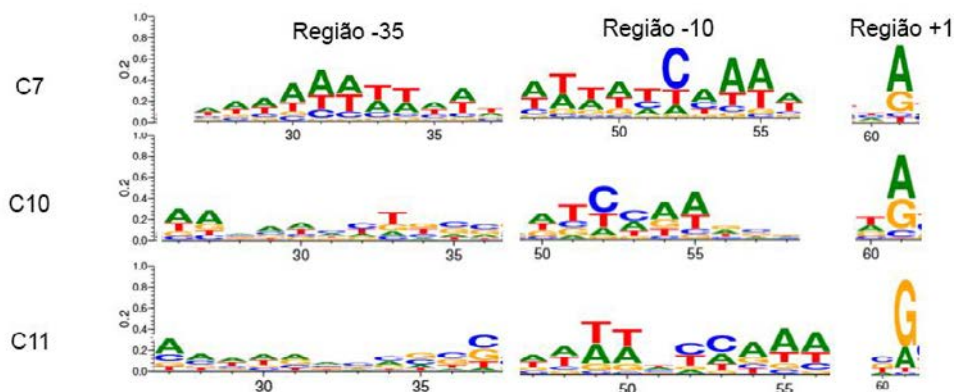


Figura 2. Composição de nucleotídeos para as 2 regiões consensuais, -35 e -10, e para a região +1 encontradas nos clusters com pureza acima de 80%.

É possível observar que a região canônica denominada -10 apresentou similaridades ao consenso biológico previamente descrito na literatura conforme tabela 1 (De Avila e Silva e Echeverrigaray, 2012). Por outro lado, não foi possível verificar relação entre o consenso da região -35 estabelecido na literatura e o consenso obtido. Essas observações são fundamentadas pela presença da região -10 expandida, principalmente para os promotores do σ_{32} , ou seja, a informação da região -35 foi realocada para a posição -10, a fim de proporcionar um melhor funcionamento da RNAP [Lim et al. 2013] [Koo et al. 2009].

Nos 3 clusters analisados, há a prevalência de nucleotídeos A e G localizados na região +1 (início da transcrição). Conforme relatos prévios na literatura, a presença do nucleotídeo G nesta posição pode estar relacionada com um mecanismo de pausa e varredura do gene a ser transcrito, com o objetivo de reparar erros e, conseqüentemente, diminuir a incidência de mutações. Os mecanismos de resposta ao estresse por choque térmico são importantes para a sobrevivência de um organismo, apresentando-se de forma rápida e eficaz para reduzir as chances de possíveis danos ao genoma [Herbert et al. 2006] [Houten e Kisker 2014] [Vvedenskaya et al. 2014].

Dentro do contexto biológico, percebe-se que um certo grau de degeneração das regiões consenso é tolerado no processo de transcrição gênica, ou seja, a transcrição ocorre de forma apropriada apesar das alterações na composição de nucleotídeos (Figura 1). Porém, computacionalmente, os desafios para a predição de promotores bacterianos são evidenciados. A pureza média obtida para a simulação com $k = 12$ foi de 63%, o que reforça a dificuldade em separar os dois fatores σ estudados.

Uma das possíveis explicações para a dificuldade apresentada encontra-se nos motivos consensuais do fator σ_{32} . A maior incidência da região -10 extendida é relatada para este fator. Utilizando a ferramenta Weblogo [Crooks 2004] para os agrupamentos que resultaram em uma pureza abaixo de 80%, foi possível observar que, em sua grande maioria, as seqüências relacionadas ao fator σ_{32} possuem o motivo consensual da região -10 degenerado ou similar ao consenso -10 relacionado ao fator σ_{24} . Portanto, a falta de especificidade nesta região para o σ_{32} pode explicar a dificuldade em separar computacionalmente os dois fatores

3.2 Trabalhos Relacionados

Os trabalhos relacionados à análise *in silico* de promotores bacterianos possuem perfis distintos, com o uso de abordagens diferentes. Kaushik et al. (2016) explica as distintas características estruturais do DNA, como a formatação cruciforme do DNA, a curvatura, cadeias paralelas de DNA e quádruplos de Guanina. Todas as características descritas possuem um amplo potencial para a pesquisa, por exemplo, os quádruplos de Guanina estão ligados com a transcrição / regulação gênica e a formatação cruciforme com seu impacto direto no enrolamento do DNA, podendo sobrepor sítios de ligação de proteínas e afetar as interações que promovem a regulação gênica. A importância das características estruturais do DNA na regulação gênica bem como na necessidade de compreender os mecanismos de controle estrutural em múltiplas posições do genoma é reforçada. O conhecimento aprofundado de tais características pode auxiliar na criação de alvos terapêuticos e no desenvolvimento de estratégias eficientes para o combate de doenças, tendo como alvo direto a expressão gênica.

As diferentes técnicas de inteligência artificial, como a *Support Vector Machine* (SVM) e redes neurais artificiais, são utilizadas para a predição de promotores bacterianos. Não foram encontrados trabalhos utilizando o algoritmo K-means em sequências promotoras. Os resultados não são diretamente comparáveis, visto as diferenças técnicas, mas é possível contextualizar os resultados obtidos neste trabalho. Além disso, há uma baixa quantidade de trabalhos que analisam outros fatores σ além do $\sigma 70$. Gordon et al. (2003), utilizando um *kernel* de alinhamento de sequências (através da abordagem de SVM), obtiveram valores para a exatidão de 84%, especificidade de 84% e sensibilidade de 82% para sequências reconhecidas pelo fator $\sigma 70$. Utilizando redes neurais artificiais, Rani et al. (2007) obteve valores acima de 90% para exatidão, especificidade e sensibilidade, utilizando as sequências de promotores reconhecidos pelo fator $\sigma 70$.

Os trabalhos de Gordon et al. (2003) Rani et al. (2007) utilizaram a composição de nucleotídeos como parâmetros para o treinamento das redes neurais. Em contraste, autores como Askary et al. (2009) e de Avila e Silva et al. (2014) utilizaram os valores de estabilidade da sequência como parâmetro. No caso de Askary et al. (2009), foram utilizadas sequências com tamanho de 413 nucleotídeos, em um total de 467 sequências, todas com seus Sítios de Início de Transcrição estabelecidos experimentalmente. O valor obtido para a exatidão foi de 94%. Em uma abordagem mais ampla, utilizando 6 fatores σ de *E. coli*, de Avila e Silva et al. (2014) apresenta a estabilidade como característica que distingue os promotores reconhecidos pelos diferentes fatores σ . No trabalho, ficou evidenciada a distinção entre os fatores $\sigma 28$ e $\sigma 54$, cada um com valores de exatidão de 80,2% e 78,8%, respectivamente. Para os fatores $\sigma 24$ e $\sigma 32$, foi encontrada uma exatidão de, respectivamente, 58% e 64%. Uma das diferenças do trabalho aqui proposto em relação aos demais apresentados é o uso da técnica de clusterização em conjunto com um parâmetro de estrutura física do DNA. Além de trabalhar com dois fatores σ que compartilham a mesma função, evidenciando as dificuldades em trabalhar *in silico* com tais conjuntos de dados.

4. Considerações Finais

O presente trabalho dedicou-se à análise *in silico* dos promotores relacionados ao estresse por choque térmico, regulados pelos fatores $\sigma 24$ e $\sigma 32$. A análise foi realizada utilizando a técnica da clusterização, com os dados codificados em seus valores de estabilidade. Foram realizadas diversas simulações com um valor de K entre 8 até 20 agrupamentos. A

pureza média obtida para a simulação de $K = 12$ foi de 63%, com o destaque de alguns agrupamentos com pureza acima de 80% (*clusters* 7, 10 e 11). A análise posterior à utilização da clusterização foi realizada com o auxílio da ferramenta WebLogo [Crooks 2004]. Dentro dos grupos com maior pureza, apenas a região -35 apresentou-se degenerada, enquanto a região -10 não divergiu do consenso descrito na literatura. Foi possível observar que o fator σ_{32} apresenta uma maior incidência da região -10 estendida, descrita por Lim et al. (2013) e Koo et al. (2009). Adicionalmente, não é possível encontrar similaridades com o consenso descrito na literatura para as duas regiões consensuais desde fator, a falta de especificidade neste conjunto de dados pode explicar uma baixa pureza na geração dos agrupamentos.

Biologicamente, as estratégias para funcionamento da RNAP e a degeneração encontrada nos motivos consensuais não acarreta em dificuldades para a transcrição gênica. Porém, evidenciam as dificuldades em analisar computacionalmente os dados biológicos. Deste modo, os diferentes agrupamentos apresentados neste trabalho contribuí-se para a compreensão dos diferentes perfis biológicos encontrados nos promotores bacterianos. Além de evidenciar o uso da estabilidade como parâmetro para análises *in silico* e auxilia na redução de falsos positivos em ferramentas de predição de promotores, como a ferramenta BacPP [de Avila e Silva, et al., 2011]. Com base no que foi exposto, pretende-se expandir a pesquisa para os demais fatores σ alternativos, ampliando os parâmetros para codificação dos dados, a fim de utilizar a curvatura como critério de classificação.

4. Referências Bibliográficas

- Askary, A., Masoudi-Nejad, A., Sharafi, R., Mizbani, A., Parizi, S. N. e Purmasjedi, M. (2009). N4: A precise and highly sensitive promoter predictor using neural network fed by nearest neighbors. In *Genes & Genetic Systems* (84) (6), páginas 425-430.
- Attwood, T. K. et al. (2011). Concepts, Historical Milestones and the Central Place of Bioinformatics in Modern Biology: A European Perspective, In *Trends and Methodologies*, Editado por Mahmood A. Mahdavi, InTech, Croácia.
- Callebaut, W. (2012). Scientific perspectivism: A philosopher of science's response to the challenge of big data biology. In *Studies in History and Philosophy of Biological and Biomedical Sciences*, páginas 69-80.
- Crooks, G. E. et al. (2004). WebLogo: A Sequence Logo Generator. In *Genome Research* (14) (6), páginas 1188-1190.
- de Avila e Silva, S. e Echeverrigaray, S. (2012). "Bacterial Promoter Features Description and Their Application on *E. coli* in silico Prediction and Recognition Approaches, In Bioinformatics, Editado por Horácio Pérez-Sánchez, InTech, Croácia.
- de Avila e Silva, S., et al. (2014). DNA duplex stability as discriminative characteristic for *Escherichia coli* σ_{54} - and σ_{28} - dependent promoter sequences. In *Biologicals* (42) (1), páginas 22-28.
- de Avila e Silva, S., Echeverrigaray, S. e Gerhardt, G. J. L. (2011). BacPP: Bacterial promoter prediction - A tool for accurate sigma-factor specific assignment in enterobacteria. In *Journal of Theoretical Biology* (287), páginas 92-99.
- Gordon, L., et al. (2003). Sequence alignment for recognition of promoter regions. In *Bioinformatics* (19) (15), páginas 1964-1971.

- Herbert., M. K. et al. (2006). Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. In *Cell* (125) (6), páginas 1083-1094.
- Houten, B. V. e Kisker, C. (2014). Transcriptional pausing to scout ahead for DNA Damage. In *Proceedings of the National Academy of Sciences* (111) (11), páginas 3905-3906.
- Jáuregui, R. et al. (2003). Conservation of DNA curvature signals in regulatory regions of prokaryotic genes. In *Nucleic Acids Research*, páginas 6770-6777.
- Kanehisa, S. et al. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. In *Nucleic Acids Research* (42), páginas D199-D205.
- Kanhere, A. e Bansal, M (2005). A Novel method for prokaryotic promoter prediction based on DNA stability. In *Bioinformatics* (6) (1), páginas 1-10.
- Kaushik, M. et al. (2016). A bouquet of DNA structures: Emerging diversity. In *Biochemistry and Biophysics Reports* (5), páginas 388-395.
- Koo, B. M. et al. (2009). Dissection of recognition determinants of *Escherichia coli* σ_{32} suggests a composite -10 region with a 'extended -10' motif and a core -10 element. In *Molecular Microbiology* (72) (4), páginas 815-829.
- Krebs, J., Goldstein, S. e Kilpatrick, S. T. (2014) Genes XI, ed. Sudbury, Massachusetts: Jones and Bartlett, 930 p.
- Lim, B. et al. (2013). Heat Shock Transcription Factor σ_{32} Co-opts the Signal Recognition Particle to Regulate Protein Homeostasis in *E. coli*. In *PLOS Biology* (11) (12), páginas 1-15.
- Marx, V. (2013). The Big Challenges of Big Data. In *Nature* (498), páginas 255-260.
- Ramprakash, J. e Schwarz, F. P. (2007) Identification and annotation of promoters regions in microbial genome sequences on the basis of DNA stability. In *Journal of Biosciences* (32), páginas 851-862.
- Rangannan, V. e Bansal, M. (2007). Identification and annotation of promoter regions in microbial genome sequences on the basis of DNA stability. In *Journal of Biosciences*, páginas 851-862.
- Rani, T. S., Bhavani, S. D. e Bapi, R. S. (2007). Analysis of *E. coli* promoter recognition problem in dinucleotide feature space. In *Bioinformatics* (23), páginas 582-588.
- Salgado, H. et al. (2013). RegulonDB v. 8: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. In *Nucleic Acids Research* (41), páginas D203-D213.
- SantaLucia J and Hicks D (2004). The thermodynamics of DNA structural motifs. *Annual review of biophysics and biomolecular structure* (33), páginas 415-440.
- Vvedenskaya, I. O. et al. (2014). Interactions between RNA polymerase and the "core recognition element" counteract pausing. In *Science* (344), páginas 1285-1289.
- Witten, I. H. e Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. Ed. San Francisco: Morgan Kaufman, 560p.

Comparing Provenance Data Models for Scientific Workflows: an Analysis of PROV-Wf and ProvOne

Wellington Oliveira^{1,2}, Paolo Missier³, Daniel de Oliveira¹, Vanessa Braganholo¹

¹Instituto de Computação, Universidade Federal Fluminense (UFF), Brazil

²DACC, Instituto Federal do Sudeste de Minas Gerais – Rio Pomba Campus, Brazil

³School of Computing Science, Newcastle University, UK

{wellmor, danielcmo, vanessa}@ic.uff.br, paolo.missier@ncl.ac.uk

Abstract. *Scientific workflows rely on provenance to be understandable, reproducible and trustworthy. Nowadays, there is a growing demand for interoperability between provenance data generated from heterogeneous workflow management systems. To address this issue, some provenance models have been proposed by extending PROV to support specific requirements of scientific workflows. In this paper, we present two prominent provenance models for scientific workflows, PROV-Wf and ProvOne, which are specializations of PROV, and compare their elements and relationships. Our goal is to provide an overview of each one and to support the choice for the most suitable for a specific context.*

1. Introduction

With the popularization of the Workflow Management Systems (WfMS), many *ad-hoc* provenance models emerged. Their goal was to represent provenance that was captured during the workflow execution, allowing scientists to compare different workflow executions, their parameters, and input data. Since different WfMS capture provenance using different perspectives, the provenance models of two different WfMS are never the same. To make matters worse, different WfMS use different storage models such as relational databases, RDF, XML, and specific models such as Virtual Data Model (VDL) [Foster *et al.* 2002].

Well aware of this scenario, the scientific community organized the *Second Provenance Challenge*¹ to discuss a standard provenance model. The result was the *Open Provenance Model* (OPM) [Moreau *et al.* 2008] that was implemented for many WfMS aiming at reaching provenance interoperability. Later, covering some gaps of OPM and adding new relationships, the W3C Provenance Working Group² proposed the PROV³ model.

PROV provides a generic data model (*i.e.*, PROV-DM) to outline provenance. However, it is not able to represent information about the workflow structure itself in the scientific experiment domain (prospective provenance [Freire *et al.* 2008]). Its focus is solely on retrospective provenance [Freire *et al.* 2008]). Thus, new specializations

¹ <http://twiki.ipaw.info/bin/view/Challenge/SecondProvenanceChallenge>

² https://www.w3.org/2011/prov/wiki/Main_Page

³ <https://www.w3.org/TR/prov-dm/>

were created to bridge this gap. Initiatives like PROV-Wf [Costa *et al.* 2013] and ProvONE [Missier *et al.* 2013] introduced new elements and relations to PROV to represent the prospective provenance and their links to retrospective provenance, much needed in this context. However, each model has approached prospective provenance in different ways. This may pose some difficulties to integrate provenance from WfMS that use different models to represent provenance. For example, one could ask “*Can I map PROV-Wf Program entity directly to ProvONE Program entity?*” If no, “*is there another entity or relationship that represents the same element in both models?*”

Since both prospective and retrospective provenance are fundamental for analyzing workflow evolution and results, scientists should be able to identify a model that better represents their provenance dataset or map their elements for exchanging purposes. In this paper, we describe these provenance models, map and compare their elements. We chose ProvONE and PROV-Wf models because they are intended for representing and exchanging provenance. Our goal is to show the similarities and differences among them and raise some issues about their representations that can help scientists to make a decision about which one is more suitable to represent their provenance dataset or guide a provenance integration process.

The paper is structured as follows. Section 2 provides a background about provenance data models. The PROV-Wf and ProvONE models are described in Section 3 and Section 4, respectively. Section 5 presents a mapping between those models. Finally, Section 6 provides final remarks about the mapping and discusses future work.

2. Provenance Types and Models

The term *data provenance* can be defined as the source or lineage of the data and it can be used to interpret and reproduce the results of scientific experiments [Freire *et al.* 2008]. Especially for experiments modeled as scientific workflows, provenance can be classified as *prospective* and *retrospective* [Freire *et al.* 2008]. *Prospective* provenance (henceforth called just *p-prov*) represents the specification of computational tasks. It corresponds to the steps to be followed to achieve a result. *Retrospective* provenance (henceforth called just *r-prov*) consists in a structured and detailed history of the execution of computational tasks (metadata associated to the execution of activities and environment characteristics).

In 2006, a discussion about standard provenance representations at the *International Provenance and Annotation Workshop (IPAW)* culminated with the creation of the *Provenance Challenge*, aiming at verifying and comparing existing provenance representations. Since the *Second Provenance Challenge*, the community began to investigate interoperability issues. That culminated with the OPM 1.0, later extended to OPM 1.1 after the *Third Provenance Challenge*⁴. Most participants of the OPM initiative joined an effort at W3C to develop another general provenance model to serve as a W3C standard. It was called PROV, and it became a W3C recommendation in April 2013. A comparison of these models was conducted a couple of years ago [Bivar *et al.* 2013].

⁴ <http://twiki.ipaw.info/bin/view/Challenge/ThirdProvenanceChallenge>

PROV-DM is the data model of PROV. It is an agnostic conceptual model and it can be applied to model different domains. It incorporates a core structure, composed by *entity*, *agent*, and *activity* elements and their relationships *WasDerivedFrom*, *WasInformedBy*, *Used*, *WasGeneratedBy*, *WasAssociatedWith*, *WasAttributedTo*, *ActedOnBehalfOf*. These elements and relationships are shown in Figure 1. PROV has also extended elements such as (i) *subtyping*, including software agent (from agent element) and *revision* (from entity element); (ii) *expanded relations*, including activity association and other *n-ary* relationships; (iii) *optional identification* that can identify an instance of an association of two or more elements; and (iv) *new relations*, consisting of subtypes or expanded versions of existing ones. The complete PROV-DM structure is composed of six components: (i) entities or activities; (ii) derivations of entities; (iii) agents; (iv) bundles; (v) properties; and (vi) collections.

As stated before, in the scientific workflows context provenance can be classified as p-prov and r-prov. PROV model is able to represent r-prov elements and their relationships, but there is just a *Plan* entity to represent p-prov aspects. In this way, new representation models such as PROV-Wf and ProvONE emerged aiming to extend PROV, adding elements and relationships to represent p-prov and r-prov in the context of scientific workflows. In the next sections, we describe these two conceptual models, PROV-Wf and ProvONE that have been proposed to extend PROV adding new domain elements and cover the p-prov representation gap.

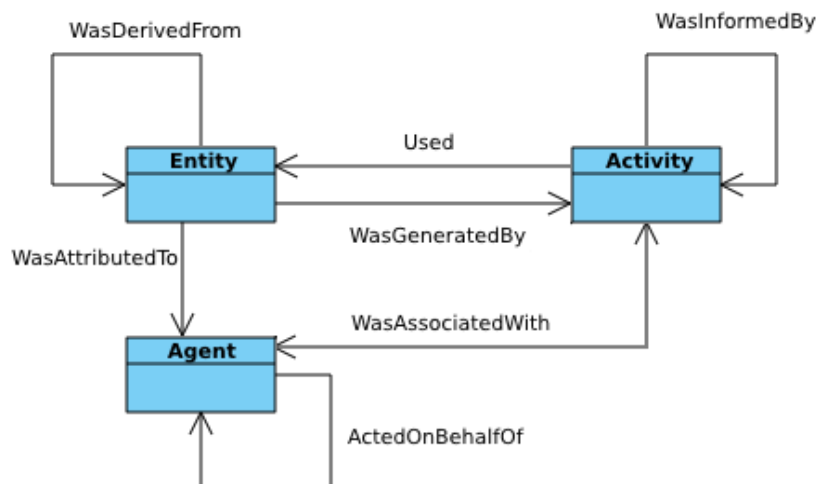


Figure 1. PROV core structures⁵

3. PROV-Wf

PROV-Wf is a conceptual model for the representation of p-prov and r-prov generated from scientific workflows [Costa *et al.* 2013]. PROV-Wf is a specialization of the W3C PROV model, which is designed to be a generic model for representing provenance in a diversity of domains different from scientific experiments (arts, industry, *etc.*). Because PROV is a generic model, it is far from trivial for several users to map its elements to the scenario of scientific experiments. PROV-Wf was thus proposed aiming to specializing PROV to the scientific experimental scenario by providing specific

⁵ <https://www.w3.org/TR/prov-dm/#core-structures>

elements for this context. PROV-Wf can explicitly define what information is captured and how it is stored.

PROV-Wf is agnostic about the environment and the WfMS and works with a set of elements that, according to Costa *et al.* [2013], can be classified into three main types: (i) Structure of the Experiment; (ii) Execution of the Experiment; and (iii) Environment Configuration. The Structure of the Experiment element is formed by the set of planning objects (*Workflow* and *Activity*) and entities (*Program*, *Field*, *Relation*, *Value*, and *Value Type*). Figure 2 shows this structure as blue rectangles. The Execution of the Experiment element consists of entities (*File*, and *Domain Data*) and activities (*Execute Workflow*, *Execute Activity*, *Execute Extractor*, and *Program Invocation*) shown as dark yellow rectangles in Figure 2. Finally, the Environment Configuration includes the *Machine* and *Scientist* agents, shown in Figure 2 as light grey rectangles. The elements *Domain Data* and *Execute Extractor* were added in an extension of PROV-Wf by de Oliveira *et al.* [2015] to represent domain-specific data.

In the PROV-Wf model, a *Workflow* class is composed by *Activities* that have one or more *Programs* and *Fields* defined by a *Value Type*. The *Execute Workflow* class represents an execution of a *Workflow* and it is performed in a *Machine* that can have many *Execute Activities*. The *Scientist* class is responsible for controlling a *Program Invocation* and *Execute Activity* that can have many *Execute Extractor* classes to capture and store *Domain Data* from *Files* (specializations of *Value Type*).

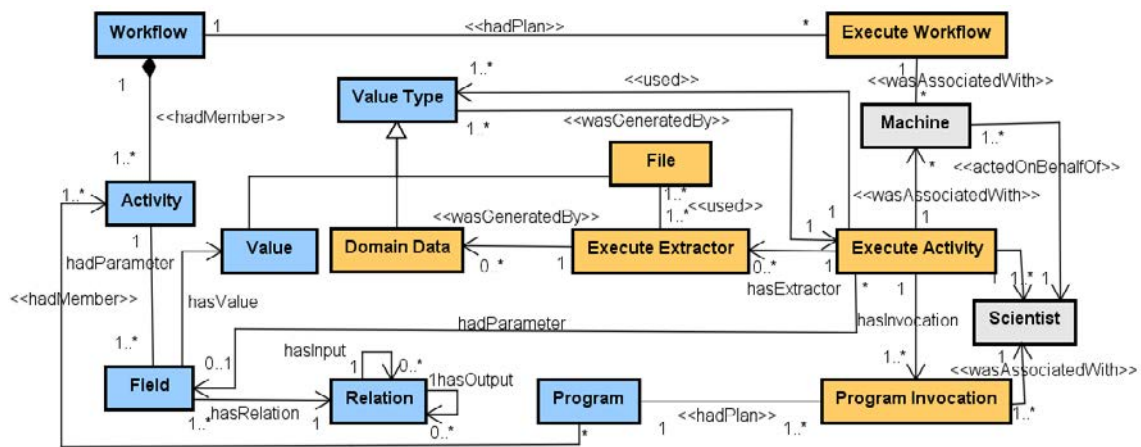


Figure 2. PROV-Wf conceptual model. Adapted from [De Oliveira *et al.* 2015]

4. ProvONE

ProvONE also extends the PROV model with an explicit representation of p-prov, thus capturing the most relevant information on scientific workflow processes, and is designed to accommodate extensions for specific scientific workflow systems [Missier *et al.* 2013]. It is fairly comprehensive including both p-prov and r-prov and allows for easy integration of terms from external vocabularies, including Dublin Core or WfMS. ProvONE is also fairly stable, and supported by a large data conservation project,

Table 1 lists and maps the entities from PROV-Wf and ProvONE models and classifies them as p-prov, r-prov, and domain data. As can be seen, some entities do not match. This is because some of them were represented as entities in one model, and as relationships in the other. For example, the entity *Program* in PROV-Wf can be represented by the *Program's* auto-relationship *hasSubProgram* in ProvONE. Similarly, the entity *ExecuteWorkflow* in PROV-Wf can be represented by the *Execution's* auto-relationship *wasPartOf* in ProvONE.

Table 1. Mapping PROV-Wf and ProvONE entities

PROV-Wf	ProvONE	Provenance Type
Workflow	Workflow	p-prov
Activity	Program	p-prov
Program	-	p-prov
Field	-	p-prov
ValueType	Entity	p-prov
ExecuteWorkflow	-	r-prov
ExecuteActivity	Execution	r-prov
ProgramInvocation	-	r-prov
ExecuteExtractor	-	domain data
DomainData		domain data
File	Document	r-prov
Value	Data	r-prov/p-prov
Relation	Port	r-prov/p-prov
Machine	-	r-prov
Scientist	User	r-prov
-	Controller	p-prov
-	Visualization	r-prov/p-prov
-	Collection	r-prov/p-prov
-	Channel	p-prov
-	Association	r-prov/p-prov
-	Usage	r-prov/p-prov
-	Generation	r-prov/p-prov

PROV-Wf provides specific elements to represent domain data such as *Execute Extractor* and *Domain Data*. On the other hand, ProvONE does not have an element to represent that kind of object, but it has a different specialization (*Visualization*) and grouping element (*Collection*) for the *Entity* object. ProvONE can also represent provenance from WfMS that use *channel* concept (*i.e.*, Kepler [Ludäscher *et al.* 2006]). A *Channel* element can connect output to input ports. PROV-Wf, in turn, identifies a computer or a virtual *Machine* as an element where some *trial* (workflow execution) may run. ProvONE cannot represent that element, but it represents a *Controller* to a specific *Program* that does not exist in PROV-Wf.

ProvONE has also three entities that connect p-prov to r-prov using a ternary relationship: *Association*, *Usage*, and *Generation*. These kinds of entities are not present in Prov-Wf, but some similar features are encountered in the *hadPlan*, *wasAssociatedWith*, *used*, and *wasGeneratedBy* relationships. ProvONE allows for easy integration of terms from external vocabularies, including Dublin Core or WfMS and it is supported by a large data conservation project, DataONE⁸.

⁸ <https://www.dataone.org/>

The mapping outlined by Table 2 relates PROV-Wf and ProvONE relationships. Most relationships of both models come from the PROV model. On the other hand, PROV-Wf and ProvONE models added new elements to represent the relationship among p-prov entities. Here we also have some gaps and relationships that were represented by entities. For example, the relationship *wasInformedBy* in ProvONE model is represented by the entity *Relation* in the PROV-Wf model. This later also represents the same information of the relationship *wasDerivedFrom* (for the entity *Data*) in the ProvONE model.

Table 2. Mapping PROV-Wf and ProvONE relationships

PROV-Wf	ProvONE
hadMember	hasSubProgram
hasInput	hasInPort
hasOutput	hasOutPort
hadParameter	hasDefaultParam
used	used
wasGeneratedBy	wasGeneratedBy
wasAssociatedWith	wasAssociatedWith
hadPlan	hadPlan
hasValue	hadEntity
hasInvocation	wasPartOf
hasExtractor	-
actedOnBehalfOf	-
hasRelation	-
-	wasInformedBy
-	wasDerivedFrom (<i>Data</i>)
-	wasDerivedFrom (<i>Program</i>)
-	controlledBy/control
-	hadInPort
-	hadOutPort
-	connectsTo
-	hadMember
-	agent
-	qualifiedAssociation
-	qualifiedUsage
-	activity

ProvONE has a special relationship *wasDerivedFrom* (for the entity *Program*) to represent different versions of *programs*. There is no such element in PROV-Wf, but different from ProvONE, it may define relationships to inform links between agents such as *Scientist* and *Machine*. The relationships *hadInPort* and *hadOutPort* link p-prov to r-prov entities in the ProvONE model, but there are no representation in PROV-Wf. Finally, *controlledBy/control* relationships relates *Controller* to *Programs* and the relationships *qualifiedAssociation*, *qualifiedUsage*, and *activity* relationships connect the aforementioned entities *Association*, *Usage*, and *Generation* in a ternary relationship among p-prov and r-prov elements. Those later are only represented in ProvONE.

6. Final Remarks

In this paper, we expose PROV-Wf and ProvONE features, their similarities and differences. Both models can represent p-prov and r-prov by extending the PROV data model, but do it using different entities and relationships. We do not compare them to

indicate the best one. Rather we describe their features intending to support the best choice for a specific context of provenance design and exchange.

As future work, we plan to develop a mechanism to convert PROV-Wf to ProvONE model and *vice-versa*. This kind of transformation may help scientists to exchange provenance among WfMS that use heterogeneous models to represent p-prov and r-prov at same time.

7. Acknowledgments

The authors would like to thank CNPq and FAPERJ for partially supporting this work.

References

- Bivar, B., Santos, L., Kohwalter, T., *et al.* (jul 2013). Uma Comparação entre os Modelos de Proveniência OPM e PROV.
- Costa, F., Silva, V., De Oliveira, D., *et al.* (2013). Capturing and querying workflow runtime provenance with PROV: a practical approach. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops*. . ACM.
- De Oliveira, D., Silva, V. and Mattoso, M. (2015). How Much Domain Data Should Be in Provenance Databases? In *Proceeding of the 7th USENIX Workshop on the Theory and Practice of Provenance (TaPP 15)*. . USENIX Association.
- Foster, I., Vöckler, J., Wilde, M. and Zhao, Y. (2002). Chimera: a virtual data system for representing, querying, and automating data derivation. In *14th International Conference on Scientific and Statistical Database Management, 2002. Proceedings*.
- Freire, J., Koop, D., Santos, E. and Silva, C. T. (2008). Provenance for Computational Tasks: A Survey. *Computing in Science Engineering*, v. 10, n. 3, p. 11–21.
- Ludäscher, B., Altintas, I., Berkley, C., *et al.* (2006). Scientific workflow management and the Kepler system. *Concurrency and Computation: Practice and Experience*, v. 18, n. 10, p. 1039–1065.
- Missier, P., Dey, S., Belhajjame, K., Cuevas-Vicentín, V. and Ludäscher, B. (2013). D-PROV: Extending the PROV Provenance Model with Workflow Structure. In *TaPP 13*. , TaPP '13. USENIX Association. <http://dl.acm.org/citation.cfm?id=2482949.2482961>, [accessed on Apr 30].
- Moreau, L., Freire, J., Futrelle, J., *et al.* (2008). The Open Provenance Model: An Overview. In *IPAW, Lecture Notes in Computer Science*. Springer. http://link.springer.com/chapter/10.1007/978-3-540-89965-5_31.

Verificação da Reprodução de *Workflows* Científicos por meio de Algoritmos de Detecção de Plágio*

Filipe Tadeu Santiago, Daniel de Oliveira

Instituto de Computação – Universidade Federal Fluminense (IC/UFF)
Niterói – RJ – Brazil

filipe_santiago@id.uff.br, danielcmo@ic.uff.br

Resumo. *Atualmente, diversos experimentos científicos são realizados por intermédio de simulações computacionais. Além da especificação, execução e monitoramento desses experimentos, um dos principais desafios da comunidade científica é como podemos verificar a reprodução de tais experimentos. Para que um experimento possa ser classificado como científico, seus resultados devem ser passíveis de reprodução por terceiros em condições similares. Em experimentos modelados como workflows científicos e executados em sistemas de gerência de workflows científicos, os dados de proveniência capturados são capazes de descrever tanto a especificação do workflow quanto o histórico da execução do mesmo. Dessa forma, esses dados oferecerem a base para verificar se uma determinada execução de um experimento é a reprodução de uma execução anterior. Entretanto, como comparar os dados de proveniência de execuções distintas de workflows não é uma tarefa trivial. Para verificar a reprodução desses workflows, nesse artigo fazemos uso de algoritmos de detecção de plágio. Tais algoritmos são capazes de verificar a semelhança entre execuções bastando ter acesso à base de proveniência prospectiva e retrospectiva do experimento.*

1. Introdução

Os avanços científicos em ciência da computação nas últimas décadas possibilitaram que os cientistas de diversas áreas, como a biologia e a química, se tornassem capazes de utilizar simulações computacionais em suas pesquisas, e com isso obter resultados com mais qualidade e em menor tempo [1][2]. Muitas dessas simulações são compostas por uma série de programas que formam um fluxo coerente onde a saída produzida por um programa é a entrada do próximo no fluxo. Tal encadeamento de programas pode ser modelado como um *workflow* científico. Um *workflow* científico é uma abstração capaz de representar experimentos, onde o cientista escolhe os programas que serão executados, a ordem em que tais programas serão executados, as dependências de dados entre os programas, e após isto, submete o *workflow* a um Sistema de Gerência de *Workflows* Científicos (SGWfC). Os SGWfCs são responsáveis por gerenciar a complexa execução do *workflow*, distribuindo a execução do mesmo em ambientes de processamento de alto desempenho, além de coletar dados de proveniência [3], que descrevem tanto a especificação do *workflow* quanto o seu histórico de execução.

Em um mesmo experimento científico, um *workflow* pode ser executado n vezes. Cada execução do *workflow* pode consumir um subconjunto de dados e valores

* Os autores gostariam de agradecer ao CNPq, CAPES e FAPERJ por financiarem parcialmente esse trabalho

de parâmetros de forma a confirmar ou refutar uma hipótese científica. Entretanto, mesmo que uma determinada hipótese científica seja confirmada ou refutada por meio de diversas execuções de um *workflow*, o seu resultados só será considerado de fato *científico* caso o mesmo seja passível de reprodução por terceiros em condições semelhantes. Isto é, o experimento só é considerado científico caso existam garantias de reprodutibilidade [4].

Dessa forma, para que o experimento e os resultados obtidos sejam válidos perante a comunidade científica, devemos ser capazes de comparar diferentes execuções de *workflows*. Uma das formas de verificar se uma execução de um *workflow* é uma reprodução de uma execução anterior é comparar os dados de proveniência prospectiva (a partir daqui chamada de *p-prov*) e proveniência retrospectiva (a partir daqui chamada de *r-prov*) das execuções. Entretanto, essa comparação não é trivial. Além de informações que normalmente são armazenadas em bancos de dados de proveniência (*p-prov* e *r-prov*), se torna necessário comparar os conteúdos de um grande volume de arquivos gerados pelas execuções dos *workflows* (que contém efetivamente o resultado do experimento). Os SGWfCs apesar de oferecerem um arcabouço completo para especificação, execução e monitoramento de *workflows*, ainda carecem de ferramentas de análise, principalmente no que tange a verificação da reprodução de experimentos. Dessa forma, se torna crucial que novas soluções para verificação da reprodução de *workflows* sejam propostas [1].

Um problema similar é considerado na área de mineração de dados, mais especificamente na área de detecção de plágio [5]. O plágio, ou simplesmente a cópia de um conteúdo desenvolvido por terceiros sem a devida autorização, é um problema crescente na academia e na indústria [5]. Comparar conteúdos para se calcular a similaridade entre os mesmos é uma tarefa crucial na detecção de plágio. Assim, diversas técnicas e algoritmos diferentes de detecção de plágio foram propostos, alguns focados na detecção de plágio em código fonte e outros em documento de texto. Tais algoritmos e técnicas oferecem um arcabouço interessante que pode ser usado para comparar execuções de *workflows*.

Nesse artigo, é proposta uma abordagem para verificação da reprodução de *workflows* científicos chamada Sličnost. A Sličnost se beneficia de algoritmos e técnicas de detecção de plágio já existentes para verificar a reprodução de *workflows*. A ideia principal por trás da abordagem é incorporar a um SGWfC um componente que importa do banco de dados de proveniência a *p-prov*, a *r-prov* e o conteúdo de arquivos de diferentes execuções de um *workflow*, os formata e submete a algoritmos de detecção de plágio. A saída do algoritmo é então um percentual de similaridade entre as execuções do *workflow*. Todo o arcabouço foi implementado no SGWfC SciCumulus [6].

Esse artigo se encontra organizado em 4 seções além dessa introdução. A seção 2 apresenta o referencial teórico sobre detecção de plágio. A Seção 3 apresenta a abordagem proposta e a validação realizada. A Seção 4 discute trabalhos relacionados e, finalmente, a Seção 5 conclui esse artigo.

2. Referencial Teórico: Detecção de Plágio em Documentos

Com a ascensão da Internet, a quantidade de ocorrências de plágio aumentou consideravelmente, visto que apenas acessando a *Web* e buscando por um determinado assunto tem-se acesso a diversos materiais sobre os mesmos [5]. O plágio é considerado um crime, semelhante à pirataria e à falsificação [5]. Dessa forma, detectar documentos plagiados se tornou uma prioridade em todo setor acadêmico.

Devido ao grande número de informação disponível na internet, e ao grande número de trabalhos publicados na academia, é impossível que uma abordagem manual seja capaz de detectar todos os possíveis casos de plágio que existam em um documento apresentado. Com isto surge a necessidade de um procedimento automatizado que execute tal tarefa. Para cumprir esta tarefa, os algoritmos de detecção de plágio foram propostos, e até hoje, infelizmente, são necessários. Existem dois métodos para analisar se um documento foi plagiado ou não [7]: extrínseco e intrínseco.

O primeiro método se baseia basicamente em comparar um documento específico com uma base de dados de documentos. São realizados vários testes que verificam a similaridade entre o documento entregue e cada um dos que estão disponíveis na base de dados, e, uma vez que seja encontrado um trecho semelhante a um documento da base que não fora citado, é detectado um nível de plágio [7]. Tal detecção é influenciada por diversos fatores, como por exemplo, o idioma em que o documento entregue foi redigido, o idioma que os documentos presentes na base foram redigidos, presença de sinônimos, e até mesmo a ordem em que uma frase tenha sido escrita.

O segundo método, também chamado de detecção de plágio interna, propõe verificar se no próprio documento que está sendo analisado existem trechos com padrões de escrita diferentes. Um mesmo texto com mais de um padrão de escrita pode indicar que partes foram plagiadas. Entretanto o documento deverá ser melhor analisado para atestar se é realmente um plágio ou apenas uma modificação brusca do estilo de escrita do autor. Os dois métodos são complementares, uma vez que o método intrínseco não é capaz de dizer a partir de qual documento o plágio foi realizado.

Para cada um dos métodos, diversos algoritmos para detecção de plágio já foram propostos [7]. O *N-Gram* [8] é um dos mais famosos entre os algoritmos de detecção de plágio. Ele tem como ideia principal criar uma espécie de impressão digital de um documento, e esse documento poderia ser identificado por essa impressão. Uma vez que outro documento contiver um fragmento de tal impressão, este fragmento é considerado como plágio do primeiro. O primeiro desafio desse algoritmo é como deve-se montar os *N-Grams*. O primeiro passo é remover os caracteres especiais do documento, visto que eles sempre existem em textos, e não é desejado que eles influenciem a semelhança entre os resultados. Após isto devemos separar as palavras em *substrings* de tamanho n .

Cada um desses *N-Grams* é submetido a um algoritmo de *hash*, formando um conjunto de *hashes*, e com isto obtêm-se um identificador para o documento, ou sua impressão digital. Se submetermos outro documento ao mesmo processo e compararmos suas impressões digitais, poderemos observar a similaridade entre ambos. Um ponto crítico nesse algoritmo é o parâmetro n , pois este definirá a granularidade de sua impressão digital, e se este for muito pequeno, palavras comuns em nosso vocabulário como “de”, “as”, *etc.* Seriam consideradas no texto que é usado para comparação. Por outro lado, se n for demasiadamente grande, utilizaremos *substrings* muito grandes, o que ignoraria muitas semelhanças entre os arquivos a serem comparados.

Para diminuir o impacto do parâmetro n , foi proposta a técnica dos múltiplos *N-Grams*. Nessa técnica, é utilizada uma faixa de valores para o parâmetro n , de modo que a taxa de semelhança entre dois documentos é dada pela média das taxas de semelhança encontradas por cada *N-Gram*. Dessa forma, o algoritmo de múltiplos *N-Grams* não fixa um único valor de *N-Gram*, podendo assim se livrar de singularidades para um

determinado valor do parâmetro n . Porém, esse algoritmo é computacionalmente custoso, uma vez que realiza o cálculo de similaridade entre documentos diversas vezes.

Uma alternativa menos custosa computacionalmente é o algoritmo *Winnowing* [9]. Esse algoritmo, também é baseado no uso de *N-Grams*, mas tem como objetivo principal reduzir o tamanho da impressão digital, sem fazê-la perder a qualidade, visto que existem textos muito extensos, e apenas a comparação de suas impressões digitais se tornaria deveras custosa. Basicamente, o *Winnowing* remove os caracteres não representativos do documento, forma todas as *N-Grams* possíveis e essas são submetidas a um algoritmo de *hash*. Porém, de posse dos *hashes*, o *Winnowing* procura reduzir a quantidade de informação na impressão digital do texto. Para alcançar tal objetivo ele introduz uma ideia de janela e esta janela irá conter w *hashes*, e dentre estes o de menor valor é selecionado. Uma vez escolhido esse valor de *hash*, a janela se desloca uma unidade para a direita, e seleciona o *hash* de menor valor da mesma, se não for o mesmo escolhido anteriormente. Deste modo, a comparação entre dois documentos se torna mais eficaz, pois o tamanho de sua impressão digital é menor. Porém, isto insere uma preocupação nova que é definir o valor do parâmetro w , visto que quanto maior a janela, mais informação estamos dispostos a perder. Esses três algoritmos foram usados na abordagem proposta nesse artigo para calcular a similaridade entre execuções de *workflows* e, assim, verificar a reprodução.

3. Abordagem Proposta: Sličnost

Nessa seção apresentamos a abordagem proposta, denominada Sličnost (similaridade de origem em Bósnio). O Sličnost é um componente que pode ser acoplado a um SGWfC e que se baseia nos dados de proveniência coletados pelo SGWfC e presentes na base de proveniência para comparar execuções diferentes de um *workflow*. Uma visão abstrata da arquitetura do Sličnost é apresentada na Figura 1. Em sua versão atual o Sličnost foi configurado para ser acoplado ao SGWfC SciCumulus e sua base de proveniência que segue o modelo PROV-Wf [10].

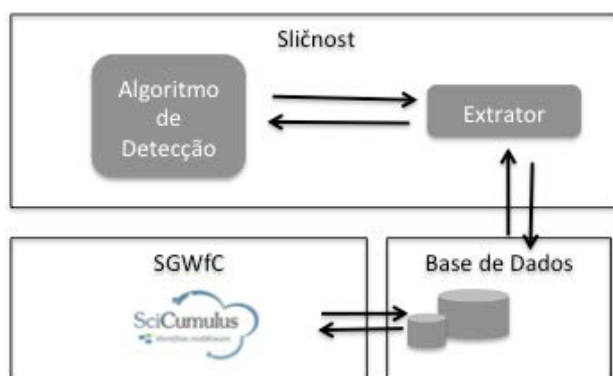


Figura 1 Arquitetura simplificada do Sličnost

O Sličnost é composto por dois componentes principais: o componente extrator e o algoritmo de detecção de plágio. O extrator é o responsável por consultar o banco de dados de proveniência do SGWfC escolhido (nesse caso o SciCumulus) e obter tanto a *r-prov* quanto a *p-prov*. Esse componente deve ser estendido para cada um dos SGWfCs que se desejar trabalhar, uma vez que cada um dos SGWfCs existentes possui um modo diferente de armazenar dados de proveniência. No caso do SciCumulus, toda a base de proveniência se encontra representada em um banco de dados relacional. Para se obter

tanto a *r-prov* quanto a *p-prov* foram realizadas consultas simples em SQL como a apresentada a seguir onde as informações das execuções das atividades são obtidas:

```
SELECT * FROM eworkflow ew, cworkflow cw, eactivity ea
WHERE ew.tagexec like '%'||cw.tag||%'
      AND ea.wkfid = ew.ewkfid
      AND Cw.tag = %Nome do workflow%
      AND ew.tagexec = %Nome do Experimento%
      AND ca.tag = %Nome da Atividade%
```

Uma vez que esses dados são obtidos, o componente extrator identifica quais arquivos foram produzidos pela execução do *workflow* (por meio de consultas a *r-prov*) e extrai o conteúdo de cada um desses arquivos como texto plano (*i.e.*, *plain text*). Para conseguirmos realizar as comparações entre as execuções dos *workflows* utilizando algoritmos de detecção de plágio é necessário que todas as informações estejam contidas em apenas um documento. Dessa forma, todos os dados extraídos (*r-prov*, *p-prov* e conteúdo de arquivos) são representados em um documento com um *template* bem definido. Sendo assim, nessa primeira versão do Sličnost as seguintes informações foram consideradas e inseridas no documento: nome do *workflow*, quantidade de atividades, tempo de execução do *workflow*. Para cada atividade do *workflow* lista-se o nome da atividade, a quantidade de execuções da mesma e quais os valores de parâmetros associados a cada execução. Além disso, cada arquivo produzido é listado e seu conteúdo inserido no documento. Dessa forma, é gerado um documento textual que descreve por completo uma execução de um *workflow*, de forma que ela possa ser analisada por um algoritmo de detecção de plágio conforme apresentado na Figura 2.

```
sciphysciPHY1.mafftalignwithmafftMAP./experiment.cmdworkflow-alfatime2015-12-2720:15:4
1.158+002015-12-2720:22:17.86+0044522mafft2015-12-2720:15:41.158+002015-12-2720:2
2:17.86+001ip-10-154-226-89.ec2.internal10.154.226.89ORTHOMCL2643513FASTA_FILE/
root/exp_SciPhy2/exp/mafft/13/ORTHOMCL264.mafft2441MAFFT_FILE/root/exp_SciPhy2/e
xp/mafft/13/44557mafft2015-12-2720:15:41.158+002015-12-2720:22:17.86+001ip-10-232-9
1-163.ec2.internal10.232.91.163ORTHOMCL297.mafft2441MAFFT_FILE/root/exp_SciPhy2/
exp/mafft/48/ORTHOMCL2973513FASTA_FILE/root/exp_SciPhy2/exp/mafft/48/44577mafft2
015-12-2720:15:41.158+002015-12-2720:22:17.86+004ip-10-154-226-89.ec2.internal10.154
.226.89ORTHOMCL5262621FASTA_FILE/root/exp_SciPhy2/exp/mafft/68/ORTHOMCL526.
mafft2301MAFFT_FILE/root/exp_SciPhy2/exp/mafft/68/44568mafft2015-12-2720:15:41.158
+002015-12-2720:22:17.86+001ip-10-159-151-224.ec2.internal10.159.151.224ORTHOMCL
4243200FASTA_FILE/root/exp_SciPhy2/exp/mafft/59/ORTHOMCL424.mafft1741MAFFT_FI
LE/root/exp_SciPhy2/exp/mafft/59/44540mafft2015-12-2720:15:41.158+002015-12-2720:22:
17.86+001ip-10-142-204-143.ec2.internal10.142.204.143ORTHOMCL2823513FASTA_FILE
/root/exp_SciPhy2/exp/mafft/31/ORTHOMCL282.mafft2441MAFFT_FILE/root/exp_SciPhy2/
exp/mafft/31/44550mafft2015-12-2720:15:41.158+002015-12-2720:22:17.86+002ip-10-159-
```

Figura 2 Exemplo de fragmento de documento gerado contendo *r-prov*, *p-prov* e dados de domínio extraídos de arquivos produzidos

Uma vez que o componente de extração gera os documentos, o algoritmo de detecção de plágio pode então ser executado. Nessa primeira versão do Sličnost foram implementados os algoritmos *N-Grams*, múltiplos *N-Grams* e *Winnowing*. Todos os algoritmos descritos na Seção 2 foram implementados em C++ e o algoritmo de *hash* escolhido foi o djb2 que é conhecido como um dos melhores algoritmos de *hash*, apesar de sua eficiência não ter sido formalmente provada. Todo o código desenvolvido pode ser acessado em https://bitbucket.org/danielcmo/deteccao_plagio.

Para avaliar o Sličnost foi realizada uma série de experimentos, com cada algoritmo de detecção de plágio implementado, e para determinados cenários possíveis em uma situação de experimentação real. Escolhemos como caso de estudo o workflow SciPhy [11] que gera árvores filogenéticas a partir de sequências de DNA, RNA e

aminoácidos. Para realizar as comparações, escolhemos uma execução do SciPhy como execução-base. A execução-base escolhida consumiu um arquivo multi-fasta contendo informações do *Plasmodium falciparum*, que é um protozoário e causa a malária em humanos. A partir da execução-base (exp01) extraímos informações de 3 outras execuções com diferentes características: (i) uma segunda execução (exp02) que consome o mesmo arquivo multi-fasta de entrada e há variação no parâmetro *bootstrap replicate* da atividade RAxML; (ii) uma terceira execução (exp03) que consome o mesmo arquivo multi-fasta e há variação no modelo evolutivo e no parâmetro *bootstrap replicate* e (iii) uma quarta execução (exp04) que consome um arquivo multi-fasta diferente (que não corresponde ao *Plasmodium falciparum*). Os parâmetros utilizados nos algoritmos foram $n=4$ para o *N-Gram* (gerando o chamado *4-Gram*), o parâmetro n variando de 4 a 6 para o algoritmo de múltiplos *N-Grams* e $w=3$ para o tamanho de janela do algoritmo *Winnowing*. Os resultados das comparações são apresentados na Tabela 1, na Tabela 2 e na Tabela 3, onde são apresentadas a similaridade entre as execuções e o tempo em milissegundos decorrido para realizar a comparação.

Tabela 1 Resultados 4-Gram

Comparação 4-Gram			
Experimento Base	Experimento Comparado	Similaridade (%)	Tempo Decorrido (ms)
exp01	exp02	96.65	981
exp01	exp03	66.97	883
exp01	exp04	25.46	865
exp02	exp01	96.68	888
exp03	exp01	66.99	886
exp04	exp01	27.37	867

Tabela 2 Resultado Múltiplos N-Gram

Comparação Múltiplos N-Gram [N = 4,5,6]			
Experimento Base	Experimento Comparado	Similaridade (%)	Tempo Decorrido (ms)
exp01	exp02	89.99	2861
exp01	exp03	58.80	2847
exp01	exp04	19.56	2780
exp02	exp01	90.02	2792
exp03	exp01	58.81	2863
exp04	exp01	21.03	3089

Tabela 3 Resultados do algoritmo Winnowing

Comparação Winnowing N = 4 W = 3			
Experimento Base	Experimento Comparado	Similaridade (%)	Tempo Decorrido (ms)
exp01	exp02	95.79	668
exp01	exp03	69.24	639
exp01	exp04	26.07	611
exp02	exp01	95.92	667
exp03	exp01	68.61	613
exp04	exp01	28.62	551

Conforme podemos observar na Tabela 1, na Tabela 2 e na Tabela 3, a suposição realizada de que podemos verificar a reprodução de experimentos foi comprovada após a execução dos diversos algoritmos. Nesse experimento consideramos que a similaridade acima de 90% indica uma reprodução (tal *threshold* irá variar de experimento para experimento). A execução com pouca variação nos dados (exp02) em relação ao experimento-base (exp01) tem alta taxa de similaridade (aproximadamente 95%), e quanto maior a variação nos dados do experimento menos similares os resultados se tornam. Porém, esse resultado depende também dos valores de parâmetros escolhidos para os algoritmos de detecção de plágio. Dependendo da escolha desses valores o resultado pode não condizer com a realidade.

Podemos notar também que como o esperado o *Winnowing* foi a técnica mais eficaz em termos de tempo de execução, uma vez que otimiza as escolhas dos *N-Grams* que compõem a impressão digital do documento que descreve a execução do *workflow*.

A técnica menos eficaz, como esperado, foi a de múltiplos *N-Grams* por fazer a verificação do documento para mais de um valor do parâmetro n . É importante ressaltar que os resultados foram bastante similares utilizando as 3 técnicas, ocorrendo apenas uma queda maior no valor da similaridade quando utilizamos a técnica de múltiplos *N-Grams*, o que era relativamente esperado, pois utilizamos apenas dois valores diferentes para o parâmetro n do que fora usado nas execuções do *4-Gram* e do *Winnowing*. Ainda sim, os valores se mantiveram relativamente estáveis, o que comprova que a similaridade não é apenas uma peculiaridade para o parâmetro $n=4$.

Outro ponto que merece esclarecimento é que o valor da similaridade entre um documento A em comparação com um documento B pode diferir do valor entre a comparação do B em relação ao documento A, visto que os dados de um experimento podem estar “contidos” no outro. Dessa forma, da maneira que os algoritmos foram implementados o documento que se encontra contido será semelhante ao que o contém, mas o contrário não será necessariamente verdade.

4. Trabalhos Relacionados

Existem algumas abordagens na literatura que visam garantir a reprodutibilidade de experimento e comparar execuções de *workflows* científicos para verificar a reprodução de experimentos. Nenhuma delas utiliza algoritmos de detecção de plágio para executar essa verificação. Dessa forma, apresentamos as abordagens mais semelhantes nessa seção. O *ReproZip* [12] é uma ferramenta que consegue copiar o *workflow*, e suas dependências, assim como os dados que o mesmo gerou utilizando uma máquina virtual. O *ReproZip* possibilita que o *workflow* possa ser executado *a posteriori* com configurações idênticas, de forma a garantir a reprodutibilidade. Entretanto, o *ReproZip* não compara as execuções do *workflow*. O *ReproZip* apenas mantém o mesmo ambiente para garantir que uma será reprodução fiel da outra.

Missier et al. [13] propõem um algoritmo para verificar a semelhança entre os dados de proveniência entre um experimento base e suas possíveis reproduções. A abordagem de *Missier et al.* se baseia na hipótese que a reprodução do experimento tenha uma especificação similar ao original, de modo que a similaridade é medida comparando-se cada atividade de um experimento reproduzido com a sua correspondente no original. Se uma execução possuir uma atividade diferente, é identificada uma nova ramificação e indica-se uma diferença entre as execuções. Apesar de realizar uma comparação estrutural eficaz, a abordagem de *Missier et al.* não utiliza o conteúdo dos arquivos produzidos na comparação, o que faz com que o resultado se torne menos eficaz.

5. Conclusão

Há muito tempo a ciência se preocupa não só em realizar grandes descobertas, mas também validar a capacidade dos resultados de experimentos serem passíveis de reprodução por terceiros em condições semelhantes. Com o advento da computação no processo de experimentação científica e a adoção de *workflows* como abstrações para modelar experimentos baseados em simulação, novas limites para as pesquisas surgiram, com isso aumentando a produtividade, e ao mesmo tempo dificultando a forma como se valida a reprodução desses experimentos baseados em simulação.

Como tal problema é, em termos científicos, uma prioridade, várias instituições incentivam a pesquisa sobre a verificação da reprodução de experimentos baseados em simulações. Nesse artigo apresentamos uma abordagem para verificação da reprodução

de experimentos modelados como *workflows* chamada Sličnost. A Sličnost utiliza dados de proveniência extraídos de bases de dados de proveniência (*r-prov* e *p-prov*) e algoritmos de detecção de plágio para verificar se uma execução de *workflow* é uma reprodução (ou não) de uma execução anterior. A avaliação da abordagem proposta foi realizada com o *workflow* SciPhy de análise filogenética e os resultados obtidos mostraram que a Sličnost é capaz de verificar a reprodução e a similaridade entre execuções de *workflows*. Trabalhos futuros incluem a análise de escalabilidade da Sličnost, uma vez que dependendo do tamanho do *workflow* analisado o documento gerado pode se tornar demasiadamente grande e a comparação se tornar pouco eficaz.

6. Referências Bibliográficas

- [1] E. Deelman, D. Gannon, M. Shields, and I. Taylor, “Workflows and e-Science: An overview of workflow system features and capabilities,” *Future Generation Computer Systems*, vol. 25, no. 5, pp. 528 – 540, 2009.
- [2] Y. Gil, E. Deelman, M. Ellisman, T. Fahringer, G. Fox, D. Gannon, C. Goble, M. Livny, L. Moreau, and J. Myers, “Examining the Challenges of Scientific Workflows,” *Computer*, vol. 40, no. 12, pp. 24–32, 2007.
- [3] J. Freire, D. Koop, E. Santos, and C. T. Silva, “Provenance for Computational Tasks: A Survey,” *Computing in Science and Engineering*, vol. 10, pp. 11–21, May 2008.
- [4] C. Goble, “The Reality of Reproducibility in Computational Science,” *Fapesp*, May-2013. [Online]. Available: <http://www.fapesp.br/eventos/2013/05/escience/Goble.pdf>. [Accessed: 27-Jan-2014].
- [5] A. M. E. T. Ali, H. M. D. Abdulla, and V. Snasel, “Survey of Plagiarism Detection Methods,” in *2011 Fifth Asia Modelling Symposium*, 2011, pp. 39–42.
- [6] D. Oliveira, E. Ogasawara, F. Baião, and M. Mattoso, “SciCumulus: A Lightweight Cloud Middleware to Explore Many Task Computing Paradigm in Scientific Workflows,” in *International Conference on Cloud Computing*, Washington, DC, USA, 2010, pp. 378–385.
- [7] S. M. zu Eissen and B. Stein, “Advances in Information Retrieval: 28th European Conference on IR Research, ECIR 2006, London, UK, April 10-12, 2006. Proceedings,” M. Lalmas, A. MacFarlane, S. Rüger, A. Tombros, T. Tsikrika, and A. Yavlinsky, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 565–569.
- [8] W. B. Cavnar, “Using An N-Gram-Based Document Representation With A Vector Processing Retrieval Model,” in *TREC*, 1994.
- [9] S. Schleimer, D. S. Wilkerson, and A. Aiken, “Winnowing: Local Algorithms for Document Fingerprinting,” in *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 2003, pp. 76–85.
- [10] W. Oliveira, D. Oliveira, and V. Braganholo, “Experiencing PROV-Wf for Provenance Interoperability in SWfMSs,” in *IPAW*, Cologne, German, 2014, pp. 294–296.
- [11] K. A. C. S. Ocaña, D. de Oliveira, E. Ogasawara, A. M. R. Dávila, A. A. B. Lima, and M. Mattoso, “SciPhy: A Cloud-Based Workflow for Phylogenetic Analysis of Drug Targets in Protozoan Genomes,” in *Advances in Bioinformatics and Computational Biology*, 2011, pp. 66–70.
- [12] F. Chirigati, D. Shasha, and J. Freire, “ReproZip: Using Provenance to Support Computational Reproducibility,” in *Proceedings of the 5th USENIX Conference on Theory and Practice of Provenance*, Berkeley, CA, USA, 2013, pp. 1–4.
- [13] Z. Bao, S. Cohen-Boulakia, S. B. Davidson, A. Eyal, and S. Khanna, “Differencing Provenance in Scientific Workflows,” in *IEEE International Conference on Data Engineering*, 2009, pp. 808–819.

E-SECO ProVersion: Manutenção e Evolução de Experimentos Científicos

**Tassio F. M. Sirqueira^{1,2}, Humberto L. O. Dalpra¹, Regina Braga¹,
Marco A. P. Araújo^{1,2}, José Maria N. David¹, Fernanda Campos¹**

¹Programa de Pós-graduação em Ciência da Computação –
Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora
(UFJF) – Juiz de Fora – MG – Brasil

²Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais
(IF Sudeste MG) – Campus Juiz de Fora – Juiz de Fora – MG – Brasil

{tassio.sirqueira, marco.araujo}@ice.ufjf.br, humbertodalpra@gmail.com
{regina.braga, jose.david, fernanda.campos}@ufjf.edu.br

***Abstract.** This paper discusses some characteristics of a scientific software ecosystem and the life cycle of an experiment, with emphasis on maintenance and evolution aspects. The E-SECO ProVersion is presented, that aim to support maintenance and evolution of experiments using data provenance. An extension of PROV model is presented together with an ontology, named PROV-OEXT. The article also presents existing workflows repositories and discuss the support that they provide to evolution and maintenance.*

Resumo. Este trabalho discute algumas características de um ecossistema de software científico e do ciclo de vida de um experimento, com ênfase em manutenção e evolução dos experimentos, usando dados de proveniência. Uma extensão do modelo PROV é apresentada conjuntamente com uma ontologia, PROV-OEXT. O artigo também apresenta repositórios de *workflows* existentes e discute o suporte provido por esses no que tange a manutenção e evolução.

1. Introdução

Um experimento científico pode ser definido como um conjunto de atividades (análises) interligadas entre si [GOBLE *et al.*, 2010]. O ciclo de vida de um experimento é composto por diversas etapas, desde a concepção do problema até a obtenção dos resultados conforme proposto por Belloum *et al.* (2011). Todas essas etapas do experimento devem ser registradas de forma a manter o registro da modelagem e execução do mesmo.

Um *workflow* científico é um modelo ou *template* que representa a sequência de atividades implementadas por ferramentas, programas ou serviços [DEELMAN *et al.*, 2009]. Um ou mais *workflows* científicos podem ser utilizados para a execução de um experimento. *Workflows* científicos são interpretados e executados por Sistemas Gerenciadores de *Workflows* Científicos (SGWfC). No geral, os SGWfC limitam-se a gerenciar a execução de *workflows* científicos de forma isolada ao experimento do qual fazem parte. Assim, de acordo com Hasan *et al.* (2007), faz-se necessário o uso de ferramentas independentes do SGWfC para apoiar o desenvolvimento do experimento

científico, sendo necessário o registro das variações dos *workflows* associados ao experimento, devido as modificações no decorrer da pesquisa [MATTOSO *et al.*, 2009].

Neste contexto, podemos considerar que todas as informações a respeito do experimento científico fazem parte de sua gerência de configuração [MATTOSO *et al.*, 2009]. Assim, avanços e mudanças no experimento devem ser registrados, criando uma base de conhecimento sobre a pesquisa. O registro dessas informações pode se tornar ainda mais útil se aplicado ao conceito de laboratórios colaborativos [VAZ *et al.*, 2012], onde pesquisadores, geograficamente dispersos, estão trabalhando em um mesmo experimento e as atividades e dados de análises devem ser registrados e compartilhados com os demais pesquisadores do grupo, evitando a perda de conhecimento e do controle sobre os dados do experimento. Para isso, é necessário o armazenamento tanto dos dados quanto dos processos que os geraram. Uma das abordagens para se realizar esse registro é o uso de modelos de proveniência [MATTOSO *et al.*, 2009].

Alguns SGWfC como o Taverna [OINN *et al.*, 2007], Kepler [ALTINTAS *et al.*, 2004] e Pegasus [GIL *et al.*, 2007], permitem capturar os passos do *workflow* (processo e dados) durante sua execução. Esses sistemas, em geral, adotam modelos proprietários para capturar os traços de proveniência gerados nas execuções. Com a ausência da padronização entre os SGWfC é difícil a interoperabilidade dos dados, bem como a consulta e a análise pelo pesquisador. Faz-se necessário um modelo de proveniência de dados que permita a captura de proveniência retrospectiva e prospectiva, mantendo o padrão para qualquer SGWfC. Proveniência de dados é o registro da história da derivação dos dados, que possibilita a reprodutibilidade, interpretação dos resultados e diagnóstico de problemas [LIM *et al.*, 2010]. Estas informações podem ser usadas para identificar métodos, regras, auxiliar os usuários na criação de *workflows* semelhantes, na compreensão de correlações de dados e na experiência para futuros experimentos [MOREAU *et al.*, 2011]. Neste trabalho é utilizado o modelo de proveniência de dados PROV [MOREAU e MISSIER, 2013], padronizado pela W3C.

Todos os dados referentes ao experimento fazem parte de sua gerência de configuração, a qual deve acompanhar todos os passos da pesquisa, utilizando-se da proveniência de dados. Além desta, informações como derivações, manutenções e evolução do experimento e *workflow* vinculados, contribuem para a compreensão do mesmo. Para apoio ao respectivo controle, pode-se utilizar, como arcabouço, as técnicas de manutenção e evolução de software.

Assim, propõe-se o desenvolvimento de uma arquitetura para suporte a manutenção e evolução de experimentos científicos, fornecendo, de maneira automatizada, informações estratégicas relacionadas a evolução e manutenção do experimento, de forma que os cientistas possam tomar decisões ou obter maior conhecimento em relação ao mesmo. Esta arquitetura denomina-se E-SECO ProVersion, e é parte da abordagem E-SECO [FREITAS *et al.*, 2015]. O E-SECO ProVersion busca adicionar funcionalidades específicas para a gerência de configuração de experimentos científicos no contexto de ecossistemas de software científicos e de laboratórios colaborativos

O artigo é composto por 3 seções, além da introdução. A seção 2 apresenta os trabalhos relacionados. Na seção 3 é apresentada a proposta do E-SECO ProVersion,

sua arquitetura, seu desenvolvimento e uma breve análise junto a base do myExperiment. Por fim, a seção 4 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

O SimiFlow é uma arquitetura para comparação e agrupamento de *workflows* pré-existentes, por similaridade, visando a construção de vários experimentos por meio de abordagem ascendente [SILVA *et al.*, 2010]. O CollabCumulus é um portal com conteúdo de repositórios de proveniência que realiza a análise dos dados gerados ou consumidos [MIRANDA *et al.*, 2014]. O PBASE [CUEVAS-VICENTTÍN *et al.*, 2014] é uma extensão do modelo de proveniência PROV destinada a *workflows* científicos, permitindo análise e replicação de experimentos.

O myExperiment é um ambiente colaborativo de compartilhamento e publicação de *workflows* [GOBLE *et al.*, 2010]. O CrowDLabs é um repositório similar que foi desenvolvido pelo grupo do SGWfC VisTrails [CALLAHAN *et al.*, 2006]. Permite a execução de *workflows*, importação de dados e reutilização por terceiros, porém não permite a análise detalhada de seus dados.

O diferencial da proposta do E-SECO ProVersion em relação a estes trabalhos é o suporte explícito a manutenção e evolução de experimentos científicos. Para isso utiliza uma base de dados de proveniência, modelada segundo o padrão PROV, que garante a interoperabilidade dos dados entre experimentos. Por meio de uma ontologia, apoio de regras de inferência para a descoberta de informações implícitas, e o acesso à repositórios de *workflows*, objetiva-se a formação de uma base histórica de dados de experimentos, com vistas a um melhor suporte a manutenção e evolução dos mesmos.

3. E-SECO ProVersion

O uso de *workflows* científicos é uma abordagem bastante utilizada no contexto de e-Science e existem muitas pesquisas voltadas para o gerenciamento e execução de experimentos baseados em *workflows*. No entanto, experimentos complexos envolvem interações entre pesquisadores geograficamente distribuídos, podendo caracterizar-se como laboratórios colaborativos, e que demandam a utilização de grandes volumes de dados, serviços e recursos computacionais distribuídos. Este cenário categoriza um ecossistema de experimentação científica [FREITAS *et al.*, 2015].

Neste contexto, foi proposta uma abordagem baseada em ecossistemas de software [BOSCH, 2009], denominada E-SECO (E-Science Software ECOsystem) [FREITAS *et al.*, 2015]. O E-SECO possui um ciclo de vida baseado na proposta de Belloum *et al.* (2011). Este ciclo de experimentação foi expandido para englobar a abordagem E-SECO ProVersion, a fim de viabilizar o suporte a gerência de configuração de experimentos, conforme Figura 1.

A gerência de configuração deve ser entendida como uma etapa de suma importância para o ciclo de vida de um experimento e dos *workflows* vinculados. Através desta pode-se determinar o estado do experimento em um determinado momento, o que, considerando o contexto de um laboratório colaborativo, pode derivar informações importantes para acertos futuros, tais como resultados parciais, erros encontrados ou gerados, entre outras informações, durante a execução do experimento.

Além disso, permite-se prever o comportamento do experimento no futuro e, no caso dos *workflows*, como os mesmos devem ser mantidos e evoluídos.

Conforme detalhado por Mattoso *et al.* (2009), o apoio a reutilização e a gerência de configuração deve ser tratado com um item importante a ser explorado, de forma a diminuir o retrabalho e apoiar o aumento de produtividade e de qualidade dos experimentos dos pesquisadores. São necessárias soluções mais abrangentes, que permitam tratar as manutenções e evoluções dos *workflows* ao longo do seu ciclo de experimentação e permitir a reutilização do próprio experimento.

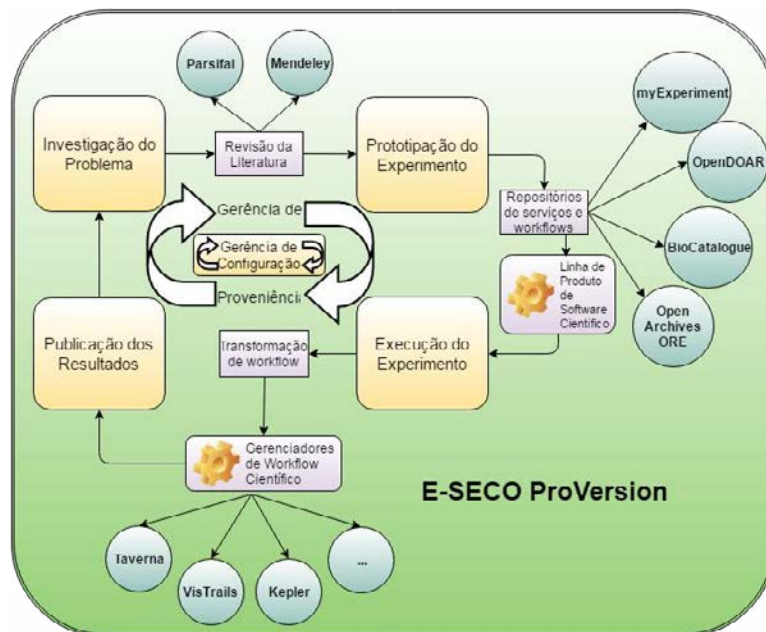


Figura 1. Ciclo de vida de um experimento científico no E-SECO ProVersion.

Na proposta E-SECO ProVersion, a gerência de configuração de um experimento está dividida em duas etapas, Gerência de Proveniência e Gerência de Manutenção e Evolução. A Figura 2 apresenta a arquitetura do E-SECO ProVersion, integrando os módulos relacionados à gerência de configuração. O gerente de manutenção e evolução é responsável pelo controle de dados de proveniência prospectiva e retrospectiva de um experimento, com o objetivo de controlar a evolução e as manutenções existentes no experimento e *workflows* a ele vinculado, além disso, considera que o experimento pode ser composto por múltiplos *workflows*, e para cada um são geradas diversas versões que devem ser controladas. Além das informações do experimento a qual faz parte, informações acerca do comportamento do *workflow*, resultados de execução e controle dos dados consumidos e gerados no experimento também são controlados pelo módulo. O módulo de proveniência, também pertencente ao gerente de configuração, é responsável pela coleta e armazenamento dos dados capturados durante a execução dos *workflows*, que são utilizados pelo módulo de manutenção e evolução. Como o E-SECO ProVersion está inserido no contexto de um ecossistema e a interoperabilidade dos dados é um fator importante, o uso de um modelo de proveniência padrão, amplamente aceito pela comunidade, é importante. Neste sentido, o modelo PROV (MOREAU & MISSIER, 2013) foi o escolhido para ser utilizado na proposta deste trabalho, compondo a gerência de proveniência. A modelagem do banco, a qual segue as regras do modelo PROV, pode ser acessada no

seguinte endereço <http://goo.gl/LvP8Aq>. Este modelo de proveniência permite acompanhar as etapas do experimento, segundo o modelo de ciclo de vida definido, criando uma base de conhecimento sobre a pesquisa. Além da coleta e o armazenamento em uma base de dados, uma ontologia integrada ao módulo é utilizada, juntamente com os dados do experimento, para auxiliar na extração do conhecimento por meio de inferências, possibilitando a identificação de *workflows* com tarefas similares, fluxos de trabalho próximos e características de manutenção e evolução.

O modelo PROV já conta com o suporte de uma ontologia denominada PROV-O [LEBO *et al.*, 2013], que é pública e disponibilizada pela W3C. No entanto, a PROV-O não expressa todo o conhecimento necessário para o suporte a evolução e manutenção de *workflows* e experimentos. O trabalho de expansão da ontologia PROV-O, através da criação da ontologia PROV-OEXT, permite a descoberta de novas informações, tanto para a captura da proveniência prospectiva quanto retrospectiva. Algumas das relações causais da ontologia PROV-OEXT são exibidas na Figura 3. Com essas informações, o E-SECO ProVersion pode sugerir mudanças estratégicas tanto na modelagem quanto na execução do experimento ou *workflows* associados.

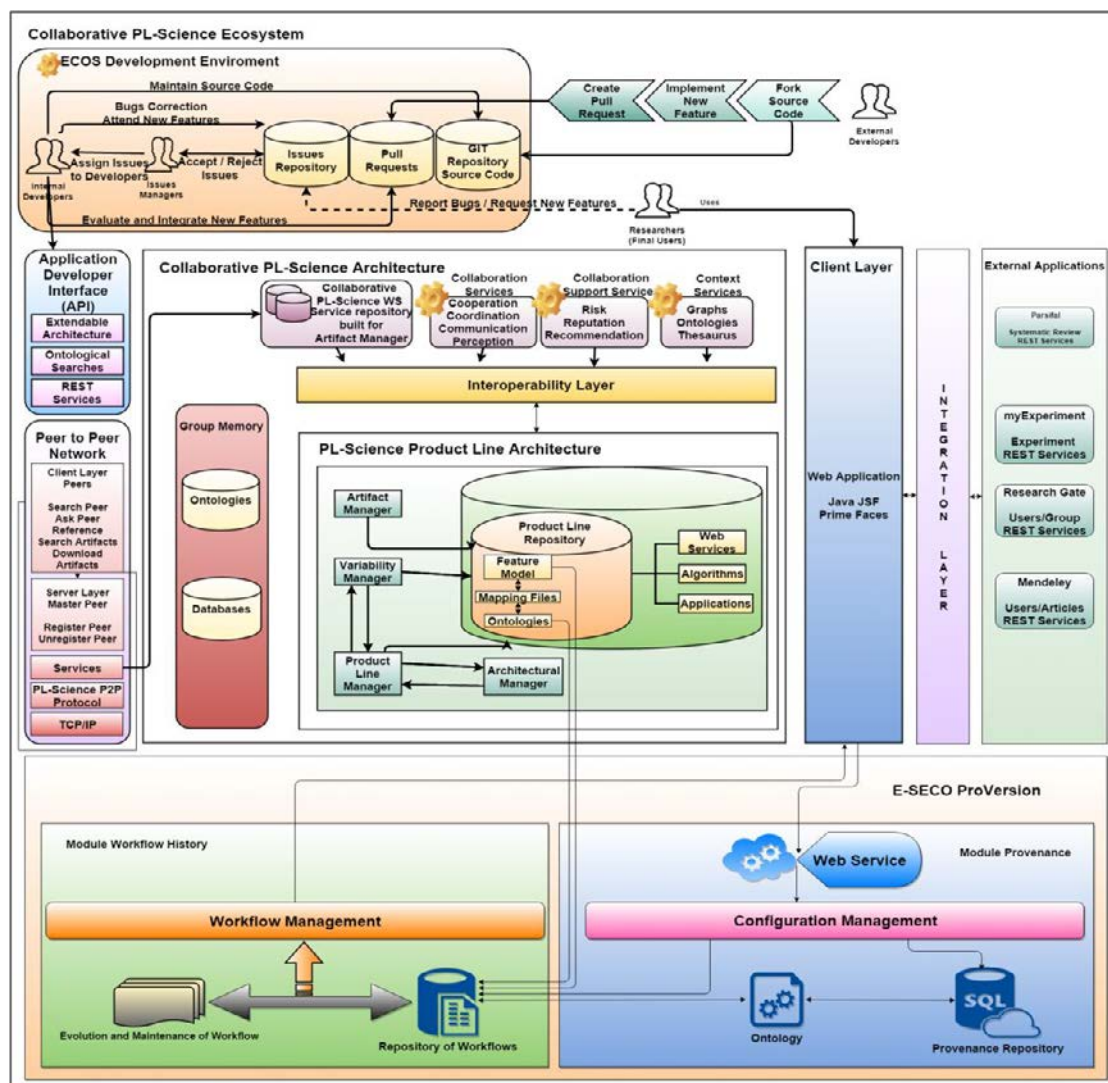


Figura 2. Arquitetura do E-SECO ProVersion.

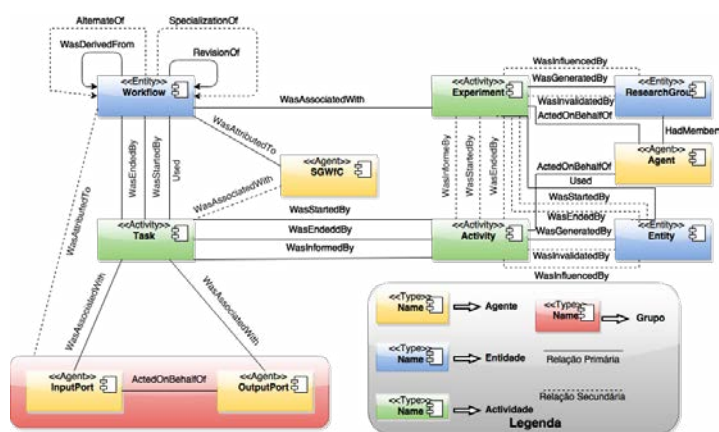


Figura 3. Relações causais da ontologia PROV-OEXT.

Além disso, no contexto de manutenção e evolução de software, uma fonte importante de informação são os dados históricos. A falta deste tipo de informação dificulta a análise dos resultados do experimento, o que pode impedir ou atrapalhar a reutilização dos *workflows* e dos experimentos, visto que não se conhece claramente a origem dos mesmos. Com o objetivo de mapear dados históricos, extraídos de repositórios existentes, foi realizado um estudo nos repositórios de *workflows* disponíveis, a fim de analisar a manutenção e evolução dos *workflows*. O referido estudo foi realizado junto aos repositórios myExperiment e CrowDLabs, ao longo do mês de outubro de 2015. No repositório CrowDLabs não foi possível extrair nenhum dado, devido a necessidade de permissão de acesso a base. No myExperiment constatou-se que entre os 3692 *workflows* disponíveis na base, 1571 utilizam os SGWfCs Taverna, Kepler ou VisTrails, os quais são os principais SGWfC, representando mais de 40% do total. Destes, 1520 são *workflows* desenvolvidos no SGWfC Taverna, 47 no SGWfC Kepler e 4 no VisTrails, destacando o Taverna como o principal em utilização entre os membros do repositório. Apesar de ser um número considerável, a quantidade é inferior a metade do total disponível no repositório, o que pode enviesar o estudo. Com relação ao histórico de versão dos *workflows*, observou-se que entre os 1571 *workflows* analisados, apenas 29% possuem dados de versionamento, o que reitera a falta de informações sobre o ciclo de vida do *workflow*, dificultando a sua reutilização.

Foi também realizada uma análise para a verificação das tarefas mais utilizadas entre os *workflows* estudados. Essa análise é útil por apresentar a quantidade de *workflows* que são afetados, caso uma tarefa seja modificada, o que irá impactar diretamente nos experimentos a que o *workflow* está vinculado. A lista das 10 tarefas mais utilizadas pode ser vista na Figura 4. Considerando a mineração de dados históricos, no repositório do myExperiment não existem informações sobre a proveniência dos *workflows* disponibilizados e os experimentos em que foram aplicados. Do ponto de vista da manutenção e evolução, tais informações se tornam essenciais para entender como os mesmos foram mantidos e evoluídos ao longo de um ciclo de experimentação.

A falta de dados históricos adequados dificulta a análise dos *workflows* e consequentemente dos experimentos pelo E-SECO ProVersion. Entretanto foi desenvolvido no contexto do E-SECO ProVersion um meta-repositório já preparado para extração e mineração destes dados de forma a contribuir com as pesquisas em e-

Science. O objetivo é acessar repositórios de *workflows* existentes e complementar com informações necessárias para o controle de manutenção e evolução.

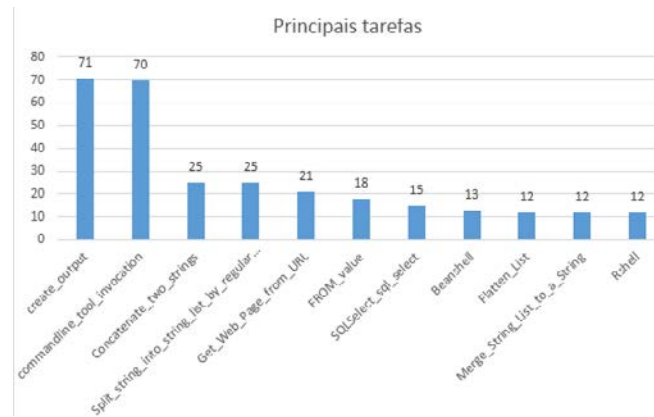


Figura 4. Lista das tarefas mais utilizadas nos *workflows*.

4. Conclusão

Atualmente, o E-SECO ProVersion permite ao pesquisador capturar os dados do *workflow* por meio de um serviço Web disponível na ferramenta. Os dados coletados junto aos *workflows* são referentes às atividades executadas, suas portas de comunicação, valores de entrada e saída, tempo de execução e informações sobre falhas na execução, além das relações causais entre as atividades. Estes dados alimentam a ontologia PROV-OEXT, que por meio de regras específicas do domínio detectam informações sobre a evolução e manutenção em *workflows* e experimentos, sendo essas informações disponibilizadas ao pesquisador por meio de uma interface Web.

A disposição destas informações ao pesquisador, de forma automatizada, facilita a identificação da origem das falhas e conseqüentemente a busca pela solução. Também contribuem para formação de uma base de conhecimento sobre o experimento, fortalecendo o uso de laboratórios colaborativos e trazendo ganhos ainda maiores no que tange a gerência de configuração dos experimentos.

Agradecimentos

Os autores agradecem o apoio da Fapemig, CAPES e CNPq.

Referências

- Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B. and Mock, S. (2004). Kepler: an extensible system for design and execution of scientific workflows. In Scientific and Statistical Database Management. 16th International Conference on (p. 423-424). IEEE.
- Belloum, A., Inda, M. A., Vasunin, D., Korkhov, V., Zhao, Z., Rauwerda, H. and Hertzberger, L. O. (2011). Collaborative e-science experiments and scientific workflows. Internet Computing, IEEE, 15(4), p. 39-47.
- Bosch, J. (2009, August). From software product lines to software ecosystems. In Proceedings of the 13th international software product line conference (pp. 111-119). Carnegie Mellon University.
- Callahan, S. P., Freire, J., Santos, E., Scheidegger, C. E., Silva, C. T. and Vo, H. T. (2006). Managing the evolution of dataflows with VisTrails. In Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on (pp. 71-71). IEEE.

- Cuevas-Vicenttín, V., Kianmajd, P., Ludäscher, B., Missier, P., Chirigati, F., Wei, Y. and Dey, S. (2014). The PBase scientific workflow provenance repository. *International Journal of Digital Curation*, 9(2), 28-38.
- Deelman, E., Gannon, D., Shields, M. and Taylor, I. (2009). Workflows and e-Science: An overview of workflow system features and capabilities. *Future Generation Computer Systems*, 25(5), p. 528-540.
- Freitas, V.; David, J. M. N.; Braga, R. M. and Campos, Fernanda. (2015). An architecture for a Scientific Ecosystem. In: 9th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (WDES 2015), Belo Horizonte, Brazil, SBC, p. 41-48.
- Gil, Y., Ratnakar, V., Deelman, E., Mehta, G. and Kim, J. (2007, July). Wings for pegasus: Creating large-scale scientific applications using semantic representations of computational workflows. In *Proceedings of the National Conference on Artificial Intelligence (Vol. 22, No. 2, p. 1767)*.
- Goble, C. A., Bhagat, J., Aleksejevs, S., Cruickshank, D., Michaelides, D., Newman, D. and De Roure, D. (2010). myExperiment: a repository and social network for the sharing of bioinformatics workflows. *Nucleic acids research*, 38(suppl 2), p. 677-682.
- Hasan, R., Sion, R. and Winslett, M. (2007, October). Introducing secure provenance: problems and challenges. In *Proceedings of the 2007 ACM workshop on Storage security and survivability (p. 13-18)*. ACM.
- Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D. and Zhao, J. (2013). PROV-O: The PROV Ontology [www Document]. URL <http://www.w3.org/TR/prov-o/>
- Lim, C., Lu, S., Chebotko, A., & Fotouhi, F. (2010, July). Prospective and retrospective provenance collection in scientific workflow environments. In *Services Computing (SCC), 2010 IEEE International Conference on (pp. 449-456)*. IEEE.
- Mattoso, M., Werner, C., Travassos, G., Braganholo, V., Murta, L., Ogasawara, E. and Martinho, W. (2009). Desafios no apoio à composição de experimentos científicos em larga escala. *Seminário Integrado de Software e Hardware, SEMISH*, 9, 36.
- Miranda, G; De Souza, J. A.; Braganholo, V.; Oliveira, D. de. (2014). CollabCumulus: Uma Ferramenta de Apoio à Análise Colaborativa de Proveniência em Workflows Científicos. SBSC.
- Moreau, L. and Missier, P. (2013). PROV-DM: The PROV Data Model [www Document]. URL <http://www.w3.org/TR/prov-dm/>
- Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P. and Plale, B. (2011). The open provenance model core specification (v1.1). *Future generation computer systems*, 27(6), 743-756.
- Oinn, T., Li, P., Kell, D. B., Goble, C., Goderis, A., Greenwood, M. and Zhao, J. (2007). Taverna/myGrid: aligning a workflow system with the life sciences community. In *Workflows for e-Science (p. 300-319)*. Springer London.
- Silva, V., Chirigati, F., Maia, K., Ogasawara, E., Oliveira, D., Braganholo, V. and Mattoso, M. (2010). SimiFlow: Uma Arquitetura para Agrupamento de Workflows por Similaridade. *IV e-Science*, 1-8.
- Vaz, G. J., Giachetto, P. F., Torres, T. Z. and Massruhá, S. M. (2012). Um Modelo de Estrutura Organizacional em Plataformas de E-Science. In *Anais do Congresso da Sociedade Brasileira de Computação (Vol. 32)*.

Ciência Cidadã Baseada em Big Data Aplicada ao Planejamento Urbano

Eliza Gomes¹, M.A.R. Dantas¹, Douglas D. J. de Macedo²
Júlio Dias³, Carlos De Rolt³, Marcelo Brocardo³, Luca Foschini⁴

¹Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC)
Caixa Postal 476 – 88.040-900 – Florianópolis – SC – Brasil

²Departamento de Ciência da Informação
Universidade Federal de Santa Catarina (UFSC)
88040-900 – Florianópolis – SC – Brasil

eliza.gomes@posgrad.ufsc.br, {mario.dantas, douglas.macedo}@ufsc.br

³Centro de Ciências da Administração e Socioeconômicas
Universidade do Estado de Santa Catarina (UDESC)
Av. Madre Benvenuta, 2037 – 88.035-001 – Itacorubí – Florianópolis – SC

{julio.dias, rolt}@udesc.br, marcelo.luiz.brocardo@gmail.com

⁴Universidade de Bologna (DISI), Bologna, Itália

luca.foschini@unibo.it

Abstract. *In this paper we presented a model and an architecture for building and structuring environments that need to store, search and analyze, efficiently, of large amount of data. For the implementation of the model we used ParticipACT Brazil project. This project aims to deploy in Florianópolis city the citizen science process to obtain relevant information to help managers to solve urban problems. We can be observed that the implementation of the structure has facilitated the environment and helped in choice of specialized tools in manipulating big data.*

Resumo. *Neste artigo são apresentados um modelo e uma arquitetura para construção e estruturação de ambientes que necessitam realizar, eficientemente, armazenamento, busca e análise de grandes volumes de dados. Para a implementação do modelo foi utilizado o projeto ParticipACT Brasil. Este projeto visa implantar na cidade de Florianópolis o processo de ciência cidadã para obter informações relevantes que auxiliem os gestores a resolverem problemas urbanos. Pode-se observar, então que a implementação do modelo facilitou a estruturação do ambiente, e ajudou na escolha de ferramentas especializadas na manipulação de big data.*

1. Introdução

O constante uso da internet em diferentes equipamentos e lugares (*Internet of Things*) faz com que a implementação de cidades inteligentes tenha um aumento significativo

[Jara et al. 2014]. Uma cidade com sensores pode obter uma variedade de dados e, conseqüentemente, gerar uma grande quantidade de informações que podem ser utilizadas tanto pelos cidadãos quanto pelos administradores para auxiliar na tomada de decisões [Moreno-Cano et al. 2015]. Além disso, a interação entre dispositivos de tecnologia da informação e comunicação (TIC) e os cidadãos pode contribuir com o desenvolvimento e a manutenção da cidade inteligente. Por exemplo, informações cedidas pelos habitantes sobre a situação do tráfego em uma determinada rua e horário pode ajudar a tornar o sistema de transporte mais inteligente [Bicocchi et al. 2013]. Esta grande quantidade e variedade de dados gerados pelas cidades inteligentes cria um *big data*.

Adicionalmente, dados de entidades prestadoras de serviços públicos podem ser disponibilizados para integrar o *big data*, pois sua utilização pode ajudar na resolução de problemas urbanos. No entanto, por serem fornecidos por diferentes fontes, esses dados podem apresentar uma variedade de formatos. Para resolver isso, são necessárias ferramentas de integração de dados. Em outras palavras, ferramentas que transformem dados diferentes em um único formato a fim de possibilitar o armazenamento em uma única base de dados.

Contudo, o aumento de técnicas e ferramentas para *big data* [Inacio and Dantas 2014] torna o processo de implementação de infraestrutura computacional um desafio, uma vez que é necessário conhecer e escolher as ferramentas mais adequadas para o ambiente, de acordo com a quantidade e o tipo de dado. Uma solução seria desenvolver um modelo de infraestrutura de *big data* que apresente como os dados devem se comportar no ambiente e quais os tipos de ferramentas que devem ser usadas e escolhidas. Entretanto, existem poucos artigos que apresentam um modelo de plataforma para *big data* [Cheng et al. 2015], [Ma and Liang 2015].

Para resolução deste problema, é proposto nesse artigo um modelo de infraestrutura para *big data*. O principal objetivo do modelo é estruturar um ambiente para receber e trabalhar eficientemente com grandes quantidades de dados. O modelo também visa fornecer as tecnologias que atendam aos requisitos para o desenvolvimento do ambiente dando autonomia para o desenvolvedor na escolha de ferramentas.

Para implementar o modelo foi realizado um estudo de caso utilizando como base o projeto ParticipACT Brasil [ParticipACT 2016b]. Este projeto tem como base os conceitos de cidades inteligentes (*smart cities*), bem como o objetivo de projetar, promover e desenvolver um sistema de gerenciamento sócio-técnica para formar gradual e progressivamente um *big data* para analisar problemas de uma área urbana.

Este artigo tem a seguinte estrutura organizacional: na Seção 2 é feita uma descrição do projeto ParticipACT Brasil, detalhando seus objetivos e estrutura; na Seção 3 é detalhado o modelo apresentado nesse artigo; um estudo de caso realizado com o projeto ParticipACT Brasil é apresentado na Seção 4; na Seção 5 são apresentados alguns trabalhos relacionados à proposta deste artigo; por fim, na Seção 6 são apresentadas as conclusões e indicações para trabalhos futuros.

2. Projeto ParticipACT Brasil

ParticipACT Brasil [ParticipACT 2016b] é uma extensão da plataforma *crowdsensing* ParticipACT [ParticipACT 2016a] da Universidade de Bologna. Este projeto está sendo

desenvolvido pela Universidade do Estado de Santa Catarina em parceria com a Universidade Federal de Santa Catarina e recebe apoio financeiro e de disponibilização de dados de entidades públicas e privadas. O objetivo do ParticipACT Brasil é utilizar os dados disponibilizados por entidades que prestam serviços básicos para os habitantes e promover “campanhas” de *crowdsensing*, para que os habitantes da cidade, voluntariamente, enviem e disponibilizem informações que possam auxiliar os gestores na resolução de problemas urbanos. Em outras palavras, com o projeto ParticipACT será possível desenvolver o processo de ciência cidadã, não somente com a ajuda dos habitantes, mas também com a ajuda de instituições públicas e privadas com a disponibilização de dados úteis para o desenvolvimento e aprimoramento dos serviços oferecidos aos cidadãos [Gomes et al. 2016].

O projeto está sendo desenvolvido, inicialmente, na cidade de Florianópolis/SC. O município de Florianópolis foi escolhida por ser a cidade sede das universidades, turística e com problemas de mobilidade e infraestrutura, o que aumenta a necessidade de um estudo concreto para a resolução desses problemas.

ParticipACT Brasil tem como objetivo inicial apresentar evidências científicas que determinem, com maior exatidão possível, a quantidade de turistas que visitam a cidade, ou seja, a população flutuante, principalmente, nos meses de Dezembro a Janeiro. Para isso, pretende-se em um primeiro momento obter dados de instituições fornecedoras de serviços básicos como: companhia de coleta de resíduos sólidos, companhia de distribuição de energia elétrica, companhia de água e saneamento, secretarias da prefeitura do município e censo (para determinar o número de habitantes).

A infraestrutura do ParticipACT Brasil, conforme mostra a Fig. 1, é dividida em três partes principais: *Big Data*, *Crowdsensing* e *Website*.

- **Big Data:** é formado pelo *Database Server*, *Server* e *Big Data Server*. *Database Server* é responsável por armazenar as informações capturadas pelos cidadãos através de “campanhas” de *crowdsensing*. *Server* é um servidor fornecido pelo ParticipACT da Bologna/Itália responsável por gerenciar as “campanhas” de *crowdsensing*. Finalmente, o *Big Data Server* integra as bases de dados, *Database Server* e dados provenientes de diferentes instituições e correlaciona esses dados para gerar informações importantes para tomada de decisão.
- **Crowdsensing:** neste está incluso o componente *App - Phone* e o *Server*. O *App - Phone* coleta dados resultantes das campanhas de participação e cooperação voluntária dos cidadãos. O *crowdsensing* é realizado por meio de smartphones (inicialmente para sistemas Android e IOS), com acesso ao aplicativo do ParticipACT Brasil, para coleta de dados. O objetivo de se gerenciar uma campanha de *crowdsensing* é coordenar um grupo de pessoas para coletar um certo tipo, e talvez complexo, de dados. Os participantes do *crowdsensing* podem acessar as informações coletadas nas campanhas através portal.
- **Website:** é o portal disponibilizado para a interação entre o usuário e o sistema ParticipACT. Esse *website* permite que o usuário visualize os resultados das análises dos dados, faça download do aplicativo, entre outros.

Diante disso, o ParticipACT Brasil está diretamente relacionado ao processo de ciência cidadã, uma vez que o projeto incentiva cidadãos, agências governamentais, em-

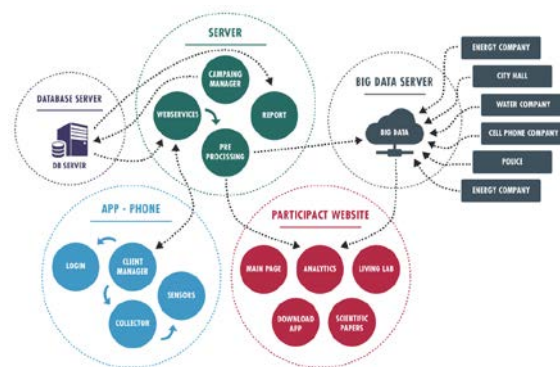


Figura 1. Infraestrutura do projeto ParticipACT Brasil [Gomes et al. 2016]

presas privadas e academia, a participarem de campanhas e assim fornecerem dados úteis para pesquisas e estudos [Conrad and Hilchey 2011].

3. Proposta

Nessa seção são apresentados o modelo e a arquitetura para um ambiente que necessite manipular grandes quantidades de dados.

Big Data é um conjunto de dados caracterizado pelo armazenamento de grandes volumes e variedades de dados, bem como pela velocidade e flexibilidade que devem ser oferecidas para manipulação dos dados como armazenamento, consulta e análise [Kitchin 2013]. Diante disso, um modelo e arquitetura de plataforma para *big data* se torna importante uma vez que o uso correto e estruturado de ferramentas determina positivamente o uso eficiente dos dados.

O principal objetivo do modelo é estruturar um ambiente que atenda aos requisitos exigidos por uma plataforma para *big data*, ou seja, volume e variedade de dados, velocidade de busca e armazenamento, flexibilidade para manipulação dos dados e possibilidade de análise sistemática dos dados.

A Figura 2 apresenta o modelo composto por cinco camadas, bem como a arquitetura desenvolvida com base no modelo proposto.

1. A primeira camada consiste dos dados fornecidos pelas instituições que são armazenados na base de dados. Diferentes instituições fornecem dados de diferentes tipos e formatos. Os tipos mais comuns são: txt, csv, planilha e base de dados relacional.
2. A segunda camada representa a interação dos dados para que sejam armazenados em uma única base de dados. Para converter os dados em um único formato é necessário utilizar uma ferramenta para extrair, transformar e carregar (*Extract, Transformation and Load* - ETL) dados. ETL, por sua vez, é um mecanismo utilizado para migrar dados heterogêneos de uma ou mais fontes de dados para um repositório de dados, *data marts* ou *data warehouse* [Albrecht and Naumann 2008]. A ferramenta utilizada, inicialmente, para o desenvolvimento da arquitetura foi o Pentaho Data Integration (também conhecido como Kettle) [Pentaho 2016].
3. A terceira camada apresenta a base de dados. De acordo com o tipo e quantidade de dados, foi utilizado a base de dados NoSQL Apache Cassandra

[Cassandra 2016]. Cassandra é um sistema de gerenciamento de banco de dados distribuído, orientado a coluna, para gerenciamento de grandes quantidades de dados estruturados distribuídos por muitos servidores.

4. A quarta camada representa a etapa de processamento de dados. Nesta fase os dados são relacionados, manipulados e analisados. A ferramenta utilizada foi o Programa R, um ambiente para análises e gráficos estatísticos. Este programa possui um pacote, RCassandra, que possibilita a conexão entre a base de dados e o programa R [R 2016].
5. A última camada consiste de uma interface amigável que apresenta ao usuário as informações geradas. Essas informações podem ser visualizadas através de gráficos, mapas ou desenhos.

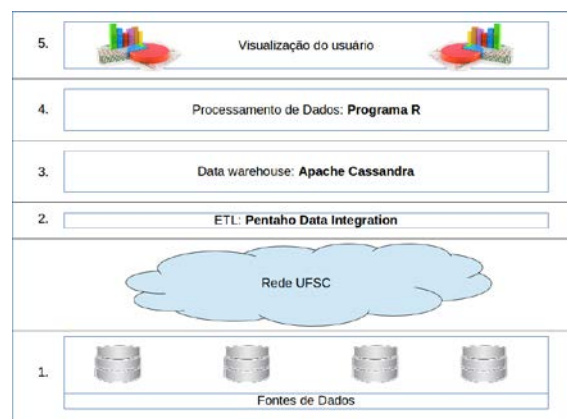


Figura 2. Modelo e Arquitetura de infraestrutura para ambientes de big data

As ferramentas Kettle e Cassandra foram escolhidas para os experimentos iniciais porque o Kettle apresenta uma interface amigável e todas as funcionalidades necessárias para a transformação e integração dos dados. Além disso, possui uma versão para a comunidade *open source*. Por outro lado, o Apache Cassandra foi escolhido, pois além de ser uma base de dados NoSQL colunar, se destaca pela boa escalabilidade, alta disponibilidade, baixa latência e por permitir a replicação dos dados em múltiplos *datacenters* [Deka 2014].

4. Estudo de Caso

Para implementar o modelo e arquitetura apresentados, foi utilizado o projeto ParticipACT Brasil, descrito na seção 2. Um dos objetivos desse projeto é receber dados de companhias prestadoras de serviços públicos e analisá-los para fornecer resultados científicos relevantes para os administradores e para a população de um modo geral. O objetivo específico desse estudo de caso é realizar uma comparação entre dois bairros, com a quantidade de habitantes relativamente parecida, da cidade de Florianópolis (Bairro C e Bairro T). O Bairro C é um bairro que atrai muitos turistas, pois oferece amplo serviço de hotéis e restaurantes, além de estar localizado em uma região cuja praia possui água com temperatura mais agradável e ampla faixa de areia. Já o Bairro T é mais residencial, ou seja, não possui grandes quantidades de hotéis ou restaurantes, por isso não é frequentado por turistas. O objetivo da comparação entre esses bairros é de verificar a quantidade

de resíduos sólidos produzidos durante o ano de 2015, considerando as características de cada bairro.

O funcionamento da arquitetura pode ser visto na Fig. 2. Os dados recebidos do provedor estão, normalmente, em formatos csv ou txt. Por isso, os dados são carregados no Kettle, que os transforma, em um formato compatível com o Cassandra, e os envia para a base de dados para armazenamento. Uma vez armazenados, os dados são capturados pelo Programa R e, a partir disso, podem ser manipulados. Após a manipulação, a informação gerada a partir dos dados é apresentada para o usuário final através de gráficos ou na forma escrita por meio de um *website*.

A Figura 3 apresenta o gráfico comparativo entre os Bairro C e Bairro T, no qual é possível observar a variação da produção de resíduos sólidos no decorrer do ano. Entretanto, está em fase de desenvolvimento um conjunto de algoritmos para avaliação dos resultados (conforme aqueles apresentados na Fig. 3).

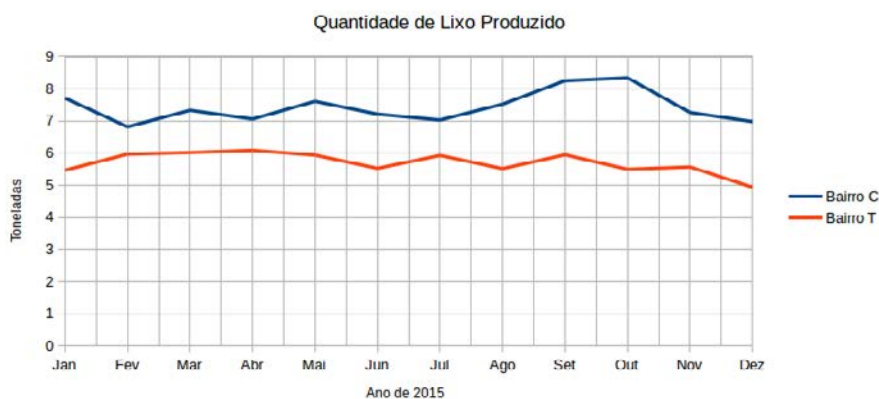


Figura 3. Quantidade de resíduos sólidos produzidos em dois bairros de Florianópolis

Outras análises poderiam ser realizadas para que seja possível explicar o motivo dessa variação, como por exemplo, se o aumento da produção de resíduos sólidos no Bairro T no mês de Julho se deve ao período de férias escolares ou se existe alguma relação entre a primavera, a safra de frutas e verduras e o aumento da produção de resíduos sólidos nos meses de Setembro e Outubro. Vale ressaltar que o objetivo desse estudo de caso é implementar o modelo e a arquitetura propostos e verificar o seu funcionamento em uma situação real. Em outras palavras, o foco não é proporcionar uma análise estatística ampla e completa.

4.1. Experiências

Para o desenvolvimento do ambiente de *big data* foram encontradas algumas dificuldades no que diz respeito às versões e instalações dos softwares utilizados. Em relação às ferramentas Kettle e Apache Cassandra, a versão atual (até a escrita desse artigo) do Kettle 6.0, não é compatível com a versão atual do Cassandra (3.4). Para que pudesse ser realizada a integração dos dois softwares foi necessário instalar o Cassandra 2.0.17.

Esses softwares exigem uma grande quantidade de memória disponível e dedicada à sua execução. Em outras palavras, computadores simples de laboratório não são

suficientes para que a arquitetura proposta seja testada.

5. Trabalhos Relacionados

O artigo de [Cheng et al. 2015] apresenta o *City Data and Analytics Platform (CiDAP)*, uma arquitetura para trabalhar com plataformas para *big data*. CiDAP processa tanto dados históricos quanto de tempo-real, enquanto expõe dados para várias aplicações. A arquitetura proposta projeta um ambiente para *big data* e descreve as ferramentas que foram utilizadas para a construção da plataforma. Os testes para validar a plataforma CiDAP foram realizados na cidade de Santander na Espanha, uma das maiores cidades inteligentes para testes chamada SmartSantander.

Por outro lado, o artigo de [Ma and Liang 2015] apresenta uma arquitetura distribuída de plataforma para *big data* para cidades inteligentes. Nessa arquitetura os autores utilizaram ferramentas como Kafka, Storm e Spark para a implementação de transmissão em tempo real de dados estruturados e não estruturados e bases de dados. Adicionalmente, desenvolveram um sistema de análise estatística para dados em tempo real e para dados *offline*.

No artigo de [Khan and Kiani 2012] é proposta uma arquitetura baseada em nuvem computacional para armazenar e processar informações fornecidas pelos habitantes de uma cidade inteligente (através do processo de ciência cidadã). A proposta inclui a possibilidade do cidadão tanto fornecer dados, através do uso de smartphones, quanto acessar informações contextuais, pré-definidas pelo usuário, a partir de um sistema integrado.

Os trabalhos mencionados apresentam uma arquitetura desenvolvida para ambientes e com ferramentas específicas. A proposta desse artigo se diferencia, pois apresenta um modelo não condicionado à ferramentas ou ambientes, dando assim maior autonomia ao desenvolver na escolha da arquitetura.

6. Considerações Finais e Trabalhos Futuros

Este artigo apresentou um modelo e uma arquitetura de infraestrutura para armazenamento, busca e análise eficientes para grandes volumes de dados. Para a implementação inicial da proposta foi utilizado o projeto ParticipACT Brasil como estudo de caso. Este projeto visa unir informações fornecidas tanto pelos cidadãos, através de campanhas de *crowdsensing*, quanto por empresas públicas e privadas, através da disponibilização voluntária de seus dados. O modelo e a arquitetura propostos estruturou de maneira eficiente o ambiente de armazenamento e análise de dados. Pôde-se observar também, que o modelo ofereceu certa autonomia para a escolha de outras ferramentas.

Como trabalhos futuros, pretende-se utilizar outras fontes de dados, como a companhia de água e saneamento, de distribuição de energia elétrica, dados do censo, bem como outras ferramentas no modelo proposto para ratificar sua autonomia e eficiência. Quanto ao aspecto de rastreabilidade dos resultados das análises pretende-se desenvolver um portal que apresente para o usuário final os dados e a forma de análise utilizada para sua manipulação. Pretende-se também realizar um estudo comparativo das ferramentas, de modo que seja possível determinar qual a ferramenta mais apropriada para o ambiente e os requisitos específicos, realizar análises mais completas dos dados, além de inserir ao modelo restrição temporal para que as respostas sejam fornecidas dentro de um tempo pré-determinado.

Referências

- Albrecht, A. and Naumann, F. (2008). Managing etl processes. *NTII*, 8:12–15.
- Bicocchi, N., Cecaj, A., Fontana, D., Mamei, M., Sassi, A., and Zambonelli, F. (2013). Collective awareness for human-ict collaboration in smart cities. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2013 IEEE 22nd International Workshop on*, pages 3–8. IEEE.
- Cassandra (2016). Apache cassandra. <http://www.planetcassandra.org/>. Acessado: Março.
- Cheng, B., Longo, S., Cirillo, F., Bauer, M., and Kovacs, E. (2015). Building a big data platform for smart cities: Experience and lessons from santander. In *Big Data, 2015 IEEE International Congress on*, pages 592–599. IEEE.
- Conrad, C. C. and Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental monitoring and assessment*, 176(1-4):273–291.
- Deka, G. C. (2014). A survey of cloud database systems. *IT Professional*, (2):50–57.
- Gomes, E., Dantas, M., Macedo, D. D. J., Rolt, C. D., Brocardo, M., and Foschini, L. (2016). Towards an infrastructure to support big data for a smart city project. In *25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*. Artigo aceito pelo evento.
- Inacio, E. C. and Dantas, M. A. (2014). A survey into performance and energy efficiency in hpc, cloud and big data environments. *International Journal of Networking and Virtual Organisations*, 14(4):299–318.
- Jara, A. J., Genoud, D., and Bocchi, Y. (2014). Big data in smart cities: from poisson to human dynamics. In *Advanced Information Networking and Applications Workshops, 2014 28th International Conference on*, pages 785–790. IEEE.
- Khan, Z. and Kiani, S. L. (2012). A cloud-based architecture for citizen services in smart cities. In *Proceedings of the 2012 IEEE/ACM 5th International Conference on Utility and Cloud Computing*, pages 315–320. IEEE Computer Society.
- Kitchin, R. (2013). Big data and human geography opportunities, challenges and risks. *Dialogues in human geography*, 3(3):262–267.
- Ma, S. and Liang, Z. (2015). Design and implementation of smart city big data processing platform based on distributed architecture. In *Intelligent Systems and Knowledge Engineering, 2015 10th International Conference on*, pages 428–433. IEEE.
- Moreno-Cano, V., Terroso-Saenz, F., and Skarmeta-Gomez, A. F. (2015). Big data for iot services in smart cities. In *Internet of Things, 2015 IEEE 2nd World Forum on*, pages 418–423. IEEE.
- ParticipACT (2016a). Participact. <http://participact.unibo.it/>. Acessado: Abril.
- ParticipACT (2016b). Participact brasil. <http://labges.esag.udesc.br/participact/>. Acessado: Abril.
- Pentaho (2016). Pentaho data integration. <http://www.pentaho.com/>. Acessado: Março.
- R (2016). R project. <https://www.r-project.org/>. Acessado: Março.

Workflows Científicos com Apoio de Bases de Conhecimento em Tempo Real

Victor S. Bursztyn^{1,2}, Jonas Dias², Marta Mattoso¹

¹Programa de Engenharia de Sistemas e Computação/COPPE – Universidade Federal do Rio de Janeiro (UFRJ), Brasil

²EMC Brazil Research & Development Center – Parque Tecnológico, Ilha do Fundão, Rio de Janeiro, RJ, Brasil

{vbursztyn,marta}@cos.ufrj.br

jonas.dias@emc.com

Abstract. *One major challenge in large-scale experiments is the analytical capacity to contrast ongoing results with domain knowledge. We approach this challenge by constructing a domain-specific knowledge base, which is queried during workflow execution. We introduce K-Chiron, an integrated solution that combines a state-of-the-art automatic knowledge base construction (KBC) system to Chiron, a well-established workflow engine. In this work we experiment in the context of Political Sciences to show how KBC may be used to improve human-in-the-loop (HIL) support in scientific experiments. While HIL in traditional domain expert supervision is done offline, in K-Chiron it is done online, i.e. at runtime. We achieve results in less laborious ways, to the point of enabling a breed of experiments that could be unfeasible with traditional HIL. Finally, we show how provenance data could be leveraged with KBC to enable further experimentation in more dynamic settings.*

1. Introdução

Devido a um maior acesso a recursos do Processamento de Alto Desempenho (PAD), tanto o mundo científico quanto o corporativo têm visto uma ascensão das aplicações com intensa manipulação de dados. Nesse contexto, a modelagem por meio de *workflows* surge como uma maneira apropriada para capturar os passos do fluxo de transformações de dados (ou *dataflow*) de tais aplicações. Na gerência da execução de um experimento científico com PAD, os conceitos relacionados a workflows servem como uma linguagem comum entre especialistas de domínio e cientistas da computação, habilitando pesquisas e esforços multidisciplinares. Entretanto, um conjunto de desafios ainda em aberto envolvem, por exemplo, a inclusão de especialistas de domínio nas iterações do ciclo de vida do experimento. Isso torna o *human-in-the-loop* (HIL) [W.F.J. 2007] um tópico bastante relevante, sendo atualmente abordado na comunidade acadêmica de workflows com dados em larga escala [Jagadish et al. 2014].

Trabalhos anteriores [Mattoso et al. 2015] abordam questões ligadas a este tópico sob o ponto de vista da proveniência de dados [Davidson and Freire 2008]. Neles, experimentos reais foram conduzidos com especialistas de domínio utilizando o Chiron, um motor de execução de workflows [Ogasawara et al. 2013]. Seus objetivos incluem: o uso efetivo de recursos de PAD; o suporte ao direcionamento dinâmico de workflows, para que workflows de longa duração possam ser ativamente gerenciados por especialistas de domínio; e consultas a uma base de proveniência de dados em tempo real.

Durante o ciclo de vida de um experimento científico, cientistas realizam a submissão de diferentes configurações de workflows até chegar a um resultado conclusivo. Para isso, analisam os resultados e frequentemente checam se fatos bem estabelecidos em seus domínios estão sendo respeitados. Caso o workflow em execução esteja refutando fatos conhecidos, ao invés de validá-los, é mais provável que ele seja interrompido, reconfigurado e só então recommençado; a ciência normalmente incrementa o conhecimento previamente estabelecido ao invés de romper com os seus fundamentos. Durante esse processo, especialistas de domínio costumam recorrer à literatura, a anotações textuais e a resultados de experimentos anteriores. Assim, a checagem de fatos em workflows centrados em dados pode extrapolar a habilidade humana para administrar informações (i.e., relembrar fatos), sendo necessário interagir com fontes de dados externas e heterogêneas, de tabelas estruturadas a arquivos PDF.

De forma complementar, a construção de bases de conhecimento (CBC) aborda o conhecimento armazenado em fontes de dados altamente heterogêneas. Segundo Ré et al. [2014], "CBC é o processo de popular uma base de conhecimento (BC) com fatos (ou asserções) extraídos dos dados (e.g., texto, áudio, vídeo, tabelas, diagramas etc)". A ferramenta de código aberto DeepDive [Ré et al. 2014], para CBC, vem se destacando em uma diversidade de domínios tais como: paleontologia e geologia, genética e farmacogenética, no apoio a órgãos investigativos, e no enriquecimento da Wikipedia.

A motivação deste trabalho é aumentar o poder de análise de dados em tempo real de um Sistema de Gerência de Workflows Científicos (SGWfC), propondo a sua integração a um sistema automático de CBC. Esperamos, como resultado ainda não encontrado na literatura, prover uma solução declarativa para mineração, em tempo real, de fatos de domínio sobre dados de proveniência. Desta forma, durante a execução de um workflow em larga escala, podemos avaliar se os resultados obtidos até um dado momento respeitam os fatos conhecidos do domínio. Se necessário, o workflow pode ser reconduzido pelo cientista ou adaptado automaticamente. O objetivo é propor uma solução de integração de SGWfC e CBC que proporcione os seguintes benefícios: (i) poupar o tempo de cientistas com a inspeção manual *offline* dos dados de proveniência; (ii) economizar recursos de PAD, já que cientistas podem reconduzir um workflow à direção correta mais rapidamente; e (iii) ampliar o potencial de reprodutibilidade de um workflow, pois a publicação da BC usada naquele domínio complementaria a definição do próprio workflow e as consultas de proveniência, compondo um pacote abrangente para a reprodução do experimento.

Fora da literatura de workflows, destacamos que alguns métodos [Hu, Zhang et al. 2009-1, Hu, Sun et al. 2009-2] para clusterização de textos já se apóiam em bases de conhecimento externas para alcançar resultados superiores. Entretanto, essas utilizações são internas à clusterização. Em nossa proposta, a base de conhecimento é utilizada para validação externa dos resultados, como um especialista de domínio.

Este artigo está organizado da seguinte maneira: na seção 2 apresentamos os conceitos de CBC e o K-Chiron; na seção 3, apresentamos o caso de uso realizado e resultados experimentais; por fim, na seção 4, apresentamos as conclusões.

2. K-Chiron: integração de CBC a SGWfC

A solução de integração de um SGWfC e CBC é apresentada por meio do K-Chiron,

uma extensão do poder analítico de dados do Chiron para contemplar CBC. Três conceitos fundamentais definem a nossa abordagem para o K-Chiron, que são: (1) a estrutura básica de CBC; (2) uma breve organização das fontes de dados observadas ao longo do atual processo de depuração conduzido por especialistas de domínio; e (3) uma visão geral sobre como essas partes (CBC e a base de fatos) podem integrar um SGWfC.

2.1. Três camadas para CBC

De acordo com Ré et al. [2014], a perspectiva do usuário de um sistema de CBC pode ser minimamente definida por um sequência de três passos, como ilustrado na Figura 1. A figura ilustra um sistema de CBC destinado a popular uma BC sobre deputados federais que pertencem a bancadas suprapartidárias, com fatos extraídos de várias fontes de dados. Na figura, três deputados claramente associados à bancada empresarial, todos da unidade federativa ES, estão mencionados em uma entrada de dados não-estruturados (i.e., um texto), legível por humanos. O primeiro passo (passo 0) consiste no pré-processamento do texto: extração dos termos nas frases, classificação das Partes Do Discurso (POS) e reconhecimento de estruturas que tenham algum valor semântico útil ao domínio. O passo seguinte (passo 1) baseia-se no resultado anterior e corresponde à extração de características (*features*), incluindo: i. Extração de menções a deputados e bancadas (entidades para o sistema de CBC); ii. Candidatos a fatos verdadeiros (combinações de deputados e bancadas que possam estar associados); e iii. Características que ajudem a descrever tal relação, como as palavras entre as menções a essas entidades. Finalmente, o último passo (passo 2) demonstra uma regra de inferência unindo os passos anteriores. A regra é modelada segundo uma função lógica, e é o elemento de mais alto nível que um usuário de um sistema de CBC deve declarar antes que o sistema prossiga à fase de treinamento, processe os dados de entrada e gere fatos classificados como verdadeiros ou falsos, com seus respectivos níveis de certeza. Ao final desse exemplo, verificaríamos as bancadas com suas respectivas composições.

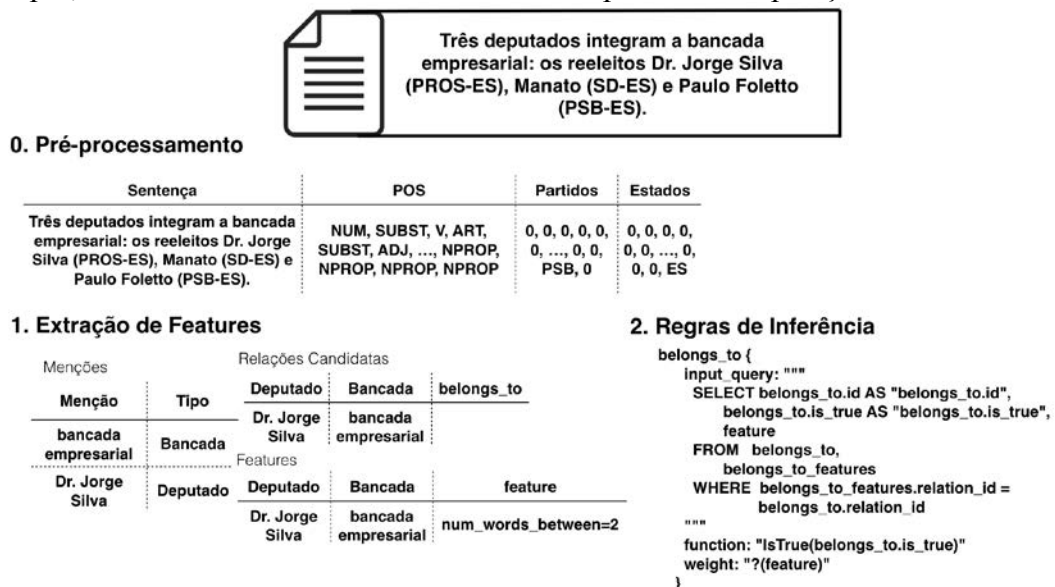


Figura 1. Adaptado de (Ré et al.): as três camadas para CBC.

2.2. A base de fatos: as fontes de dados do especialista de domínio

Como apresentado, especialistas de domínio mantêm seu conhecimento armazenado em

fontes de dados heterogêneas. Verificamos a existência de fontes de dados em três níveis de exploração do domínio: i. O mais interno diz respeito às anotações pessoais tomadas pelo próprio especialista; ii. O segundo nível representa fatos descobertos em conjunto com um grupo de pesquisa, frequentemente relacionados a explorações mais ambiciosas que exigem coordenação e comunicação entre membros do grupo, abrangendo fatos não necessariamente publicados; iii. E o nível mais externo representa fatos conhecidos bem estabelecidos na comunidade científica daquele domínio. Assim, a base de fatos pode ser considerada como uma agregação de textos, de diversas tabelas de dados e de coleções de arquivos PDF, todos carregando fatos com diferentes níveis de confiança, tratando-se de um cenário recorrente à comunidade de CBC.

2.3. CBC no contexto do Chiron e os dados de proveniência

A arquitetura original do Chiron [Ogasawara et al. 2013] é apresentada na Figura 2 em conjunto com a proposta da base de fatos. O Chiron requer dados de entrada em um sistema de arquivos compartilhado (Figura 2, item 1), a definição de um workflow (item 2) e os usuários podem executar consultas sobre os dados de proveniência (item 3) armazenados em um banco de dados relacional durante a execução do workflow. O uso de um sistema de CBC adiciona à arquitetura do Chiron: uma base de fatos (item 5) e definições de características e de regras de inferência (item 4). As caixas azuis (itens 2 a 4) representam o que há de mais específico sobre o experimento e os itens numerados (1 a 5) representam o pacote completo do experimento, que poderia ser compartilhado para facilitar a reprodução do mesmo.

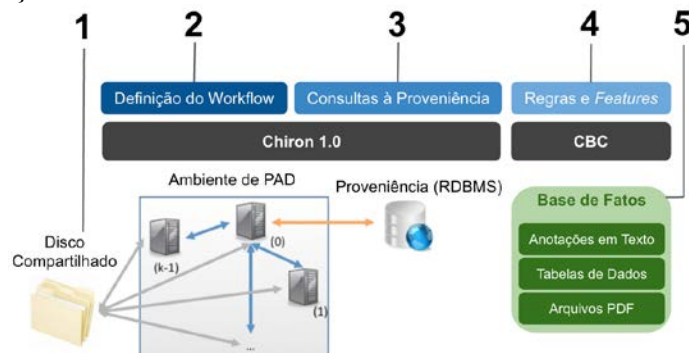


Figura 2. K-Chiron: CBC no contexto do Chiron e os elementos de dados.

A checagem de fatos por meio de dados de proveniência é uma atividade laboriosa e envolve interações com várias fontes de dados, uma vez que há geralmente muitos fatos importantes para um experimento. Em trabalhos anteriores [Dias et al. 2011, 2015], mostrou-se que a recondução de um workflow no Chiron pode, em tempo de execução, economizar um tempo substancial no ciclo de um experimento. A mineração de fatos deve ser pensada como um mecanismo para amplificar essa capacidade pois avalia automaticamente se os resultados obtidos respeitam os fatos conhecidos do domínio. Podem existir diferentes níveis de confiança para diferentes conjuntos de fatos conhecidos, cada nível causando decisões específicas na recondução do workflow.

3. Resultados Experimentais

O caso de uso realizado é uma aplicação no campo das Ciências Políticas, baseada em dados públicos disponíveis na Câmara dos Deputados brasileira. O seu propósito é processar dados brutos sobre como os 513 deputados federais votam em projetos de lei e

então permitir que um especialista de domínio identifique as bancadas mais proeminentes (i.e., que votam de maneira mais coesa) com o apoio de um workflow científico no Chiron. Como existem atualmente no Brasil 35 partidos políticos, o estudo das bancadas suprapartidárias torna-se especialmente interessante, embora desafiador por conta de sua natureza transversal à organização partidária oficial. Entende-se que o especialista de domínio queira interpretar agrupamentos de deputados federais que votam de maneira semelhante de acordo com as bancadas suprapartidárias mais tradicionais (agronegócio, empresarial, sindical, feminina e religiosa), ou a segmentos delas. Este experimento agrupa as votações na Câmara por meio de um algoritmo de clusterização [Hartigan & Wong 1979]. Como a quantidade de agrupamentos é desconhecida, o objetivo do workflow é encontrar a parametrização que resulta nos agrupamentos mais homogêneos em relação à identificação das bancadas.

O workflow é modelado para utilizar o algoritmo de clusterização K-Means [Hartigan & Wong 1979] da biblioteca Scikit Learn versão 0.17 [“Scikit Learn” 2016]. Variou-se o parâmetro k (i.e., a quantidade de agrupamentos) para varrer um espaço de busca pré-definido. Neste caso, com cinco bancadas de tamanhos variados, também convém testar se as bancadas não formariam subgrupos ou até mesmo um supergrupo diante das questões votadas. Assim, a varredura do parâmetro k foi definida no intervalo entre 4 e 18. Para fins de ilustração, a Figura 3 mostra uma visualização do K-Means executado com $k = 6$ e projetado para 2D (com uma análise dos componentes principais). A maneira escolhida para avaliar a qualidade dos agrupamentos nesta aplicação é pela medida de precisão [Manning et al. 2008]. Portanto, conduziremos a varredura em busca da maior precisão média, usando como *ground truth* a BC populada a partir do processo explicado no item 2.1.

O conjunto de dados usado neste experimento é extraído da API da Câmara [“Dados Abertos - Legislativo” 2016]. Ele cobre as votações de 384 deputados acerca de 14 projetos de lei considerados polêmicos tais como: reforma política, a diminuição da maioria penal, o relaxamento de regulações do agronegócio, dentre outros. Os votos são computados como valores inteiros, conforme: sim (+2), obstrução (+1), não (0), abstenção (-1) e nulo (-2). A obstrução fica entre os valores "sim" e "não" porque é uma posição que, dependendo do contexto da votação, pode corresponder a um ou a outro. Também vale destacar que a lista com 384 nomes é um subconjunto dos 513 deputados eleitos, já que a parte que falta não é retornada pela API nas votações observadas.

3.1. O uso do DeepDive para CBC

Ao final de cada eleição, o DIAP (Departamento Intersindical de Assessoria Parlamentar) faz o lançamento de uma série de relatórios à imprensa analisando a composição das novas bancadas suprapartidárias. O departamento busca interpretar a nova Câmara dos Deputados independentemente das fronteiras partidárias, e tradicionalmente busca preencher os novos nomes das bancadas: a do agronegócio, a empresarial, a sindicalista, a feminina e a religiosa. Portanto, a BC deste exemplo consiste no entendimento dos seis relatórios lançados pelo DIAP e popular fatos do tipo "*deputado d_i pertence à bancada suprapartidária b_j* ".

O uso do DeepDive começa com um passo de pré-processamento, que neste caso é feito sobre todos os relatórios do DIAP. Em conformidade com o exemplo oficial do DeepDive, usamos a biblioteca NLTK versão 3.1 [“Natural Language Toolkit” 2016] para

realizar os passos básicos para o processamento de linguagem natural: a extração de todas as frases dos relatórios com seus respectivos termos; a classificação de cada POS usando o corpus MacMorpho do NLTK, que consiste em um conjunto anotado de textos de notícias (mais de um milhão de termos) extraídas de diversos jornais brasileiros. Além disso, reconhecemos dentro das frases dois conjuntos de termos: as abreviações de unidades federativas brasileiras e os acrônimos dos partidos políticos, pois ambas informações são úteis durante a detecção de entidades. Portanto, para cada frase pré-processada deve existir: a frase bruta; os termos dela extraídos; os termos seguidos de suas classificações como POS; os termos identificados como unidades federativas brasileiras; e os termos identificados como acrônimos de partidos políticos.

Uma vez que o passo de pré-processamento está concluído, construímos dois extratores, criando código especificamente para esta BC. O primeiro é o extrator de deputados, que aplica algumas heurísticas para extrair candidatos a menções – assume que, em relatórios, deputados são mencionados próximos de seus respectivos partidos políticos e de suas unidades federativas originárias. Para ampliar os candidatos a menções, o extrator considera todos os unigramas, bigramas e trigramas anteriores à heurística apresentada (i.e., um acrônimo de partido político seguido de um estado brasileiro, ou vice-versa). Já o segundo extrator é o que obtém as bancadas suprapartidárias. Similarmente, ele aplica heurísticas para extrair candidatos a menções – assume que essas entidades são caracterizadas pelo termo "bancada" seguido de um termo classificado como adjetivo. Os extratores foram implementados em Python e declarados na configuração do DeepDive.

Com todos os candidatos a menções a entidades extraídos, seguimos para extrair candidatos à relação *belongs_to*. Para o propósito do workflow a ser executado após a CBC, todas as co-ocorrências de deputados e bancadas em uma mesma seção são consideradas candidatas à relação-alvo (*belongs_to*). Entretanto, este extrator aplica um filtro adicional: ele remove candidatos à relação-alvo se o deputado em questão não constar entre os 384 nomes obtidos pela API da Câmara, admitindo uma distância de Levenshtein de até 2 [Navarro 2001]. Assim, todos os fatos que serão obtidos a partir dos candidatos à relação estarão ancorados aos dados de entrada do workflow; só conheceremos as bancadas de quem interessar ao experimento.

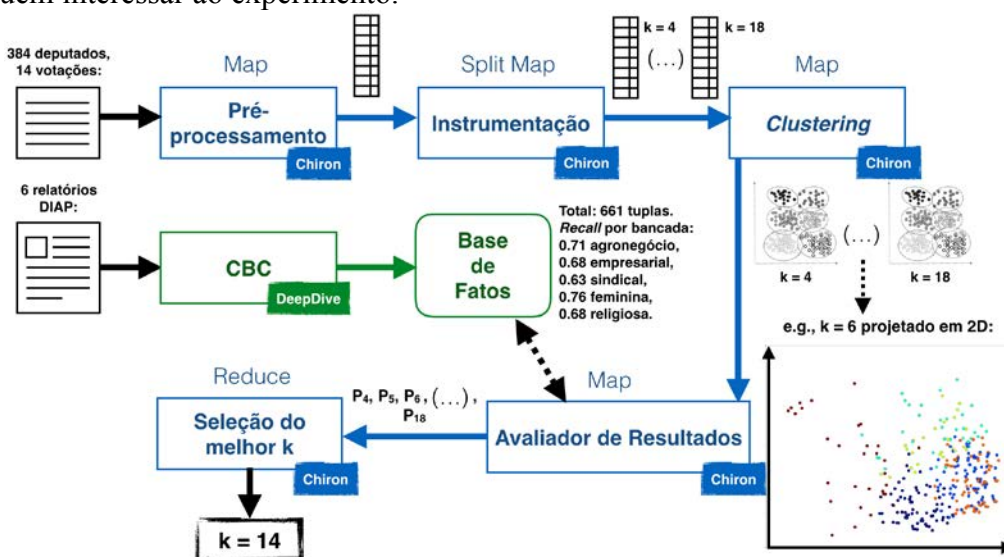


Figura 3. O workflow analítico da Câmara com o K-Chiron.

Para treinar o modelo estatístico do DeepDive, compila-se uma lista com 60 fatos verdadeiros (i.e., deputados específicos que de fato pertencem a bancadas específicas). A geração de uma lista de fatos verdadeiros constitui a chamada "supervisão distante", necessária ao DeepDive. Além disso, cria-se um extrator de características genéricas para cada sentença extraída como uma candidata à relação *belongs_to*. Tal extrator emprega a biblioteca Python fornecida pelo DeepDive, chamada DDLib, para emitir características genéricas a partir do texto das frases. O uso da DDLib para extrair features é a abordagem recomendada para tratar de atributos de texto no DeepDive. Por fim, declara-se uma regra de inferência no arquivo de configuração do DeepDive conforme apresentada na Figura 1. Ela ao mesmo tempo captura os exemplos de fatos verdadeiros e, nos casos preditos pelo DeepDive, retorna como saída o valor da predição. A regra de inferência popula a BC tanto com os fatos conhecidos usados no treinamento, quanto com os fatos preditos como verdadeiros ou falsos a partir dos candidatos extraídos à relação *belongs_to* com uma cobertura de 63% a 76% das bancadas desejadas.

3.2. O workflow integrado à base de conhecimento

O workflow do experimento com o K-Chiron são descritos na Figura 3. Nesta prova de conceito, a integração com o DeepDive foi feita a partir de uma atividade do workflow programada para realizar consultas à base de fatos no DeepDive. Em um cenário mais geral, as atividades do workflow no K-Chiron podem ser configuradas de maneira declarativa para minerar os dados de proveniência, checando os fatos na BC. Neste caso, o SGWfC é responsável pela varredura do espaço de busca enquanto a base de fatos possibilita a verificação dos resultados em tempo real. Resumindo a sequência de passos para o workflow, temos: i. Uma atividade do tipo *Map* consolida as votações; ii. Uma atividade do tipo *Split Map* instrumenta o conjunto de dados para varrer todo o espaço de busca de k ; iii. Uma atividade *Map* lê as votações e aplica o K-Means para cada valor distinto de k ; iv. Cada resultado da clusterização, em outra atividade *Map*, é testado com respeito à BC populada pelo DeepDive, culminando em uma medida de precisão para cada agrupamento e uma precisão média para cada valor de k . Por fim, o workflow termina indicando para qual valor de k a precisão média é ótima, significando a produção dos agrupamentos mais homogêneos (em que proporção, na média, os agrupamentos foram preenchidos por membros de uma única bancada suprapartidária).

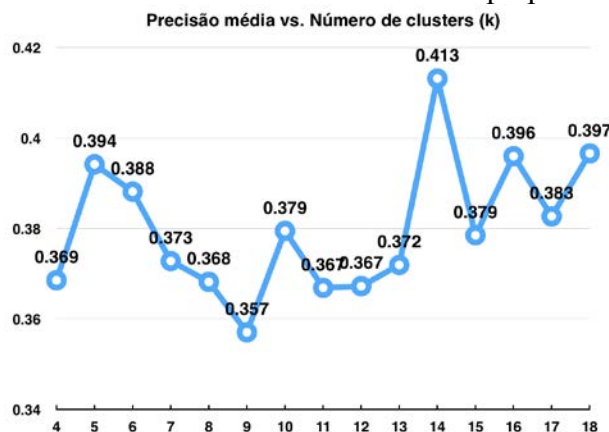


Figura 4. A precisão média dos agrupamentos variando k entre [4,18].

A Figura 4 apresenta a qualidade média dos agrupamentos e pode-se observar que a precisão ótima é alcançada com $k = 14$, lembrando que tal verificação diz respeito à

base de fatos obtida com o DeepDive. Se o especialista de domínio fosse navegar pelos agrupamentos resultantes para todos os valores de k testados, sua avaliação passaria pelos 384 nomes de deputados 15 vezes (dado que $|\{4,18\}| = 15$). Isso significaria uma avaliação de quase seis mil nomes, possivelmente consultando os seis relatórios do DI-AP nos quais são armazenados um total de 582 fatos relevantes (i.e., deputados específicos pertencem a bancadas suprapartidárias específicas).

4. Conclusões

Encapsular o conhecimento do especialista de domínio usando CBC, e com isso levá-lo para dentro do ciclo do workflow, mostrou ser uma ideia bastante produtiva. É difícil mensurar com precisão a quantidade de tempo efetivamente poupada, mas é importante destacar que, de seis mil avaliações originalmente demandadas, apenas uma única avaliação precisou ser feita tirando proveito das consultas em tempo real. Aplicar o DeepDive ao HIL de workflows pode ser uma ideia ainda mais produtiva se explorada em cenários de execução dinâmica de workflows, em problemas que também dependam da administração de uma grande quantidade de fatos por parte do especialista de domínio. Nesses casos, a mineração de fatos em larga escala pode: i. poupar o tempo de cientistas com a inspeção manual dos dados de proveniência; e ii. economizar recursos de PAD, já que cientistas poderiam reconduzir um workflow à direção correta mais rapidamente.

5. Referências

- Dados Abertos - Legislativo (2016). <http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo/dados-abertos-legislativo>, [accessed on Apr 7].
- Davidson, S. B. and Freire, J. (2008). Provenance and Scientific Workflows: Challenges and Opportunities. In *Proceedings of the 2008 ACM SIGMOD*
- Dias, J., Guerra, G., Rochinha, F., et al. (may 2015). Data-centric iteration in dynamic workflows. *Future Generation Computer Systems*, v. 46, p. 114–126.
- Dias, J., Ogasawara, E., Oliveira, D., et al. (2011). Supporting Dynamic Parameter Sweep in Adaptive and User-Steered Workflow. In *WORKS '11*. ACM.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), p. 100–108.
- Hu, X., Sun, N., Zhang, C., & Chua, T. S. (nov 2009). Exploiting internal and external semantics for the clustering of short texts using world knowledge. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 919-928).
- Hu, X., Zhang, X., Lu, C., Park, E. K., & Zhou, X. (june 2009). Exploiting Wikipedia as external knowledge for document clustering. In *Proceedings of the 15th ACM SIGKDD* (pp. 389-396).
- Jagadish, H. V., Gehrke, J., Labrinidis, A., et al. (1 jul 2014). Big data and its technical challenges. *Communications of the ACM*, v. 57, n. 7, p. 86–94.
- Manning, C. D., Raghavan, P. and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Mattoso, M., Dias, J., Ocaña, K. A. C. S., et al. (may 2015). Dynamic steering of HPC scientific workflows: A survey. *Future Generation Computer Systems*, v. 46, p. 100–113.
- Natural Language Toolkit (2016). <http://www.nltk.org/>, [accessed on Apr 7].
- Navarro, G. (mar 2001). A Guided Tour to Approximate String Matching. *ACM Computing Surveys*, v. 33, n. 1, p. p 31–88.
- Ogasawara, E., Dias, J., Silva, V., et al. (2013). Chiron: A Parallel Engine for Algebraic Scientific Workflows. *Concurrency and Computation*, v. 25, n. 16, p. 2327–2341.
- Ré, C., Sadeghian, A. A., Shan, Z., et al. (23 jul 2014). Feature Engineering for Knowledge Base Construction. *arXiv:1407.6439 [cs]*,
- Scikit Learn (2016). <http://scikit-learn.org/stable/>, [accessed on Apr 7].
- W.F.J., B. (2007). Human-in-the-loop simulation: the right tool for port design. In *Port Technology International*.

Integrating Scientific Workflows with Scientific Gateways: A Bioinformatics Experiment in the Brazilian National High-Performance Computing Network

Maria Luiza Mondelli¹, Marcelo Monteiro Galheigo¹, Vívian Medeiros¹,
Bruno F. Bastos¹, Antônio Tadeu Azevedo Gomes¹,
Marta Mattoso², Ana Tereza R. Vasconcelos¹, Luiz M. R. Gadelha Jr.¹

¹Laboratório Nacional de Computação Científica (LNCC)
Petrópolis – RJ – Brazil

²Programa de Engenharia de Sistemas e Computação (PESC/COPPE)
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brazil

{mluiza, galheigo, vivian, bfbastos, atagomes, atrv, lgadelha}@lncc.br

marta@cos.ufrj.br

Abstract. *Bioinformatics experiments are rapidly and constantly evolving due to improvements in sequencing technologies. These experiments usually demand high performance computation and produce huge quantities of data. They also require different programs to be executed in a certain order, allowing the experiments to be modeled as workflows. However, users do not always have the infrastructure needed to perform these experiments. Our contribution is the integration of scientific workflow management systems and grid-enabled scientific gateways, providing the user with a transparent way to run these workflows in geographically distributed computing resources. The availability of the workflow through the gateway allows for a better usability of these experiments.*

1. Introduction

Biology is generating data sets that are increasingly large due, for instance, to high-throughput sequencing technologies. Various computational tools can be composed to perform complex analyses of these sequences in areas such as metagenomics, transcriptomics and comparative genomics. To better manage these analyses, scientists can use scientific workflow management systems [Liu et al. 2015] (SWfMS), which allow them for easily composing and executing many-task computations through the automation of the steps commonly involved: scheduling activity execution based on data dependencies; parallel execution of independent activities; synchronization of workflow execution; managing data transfer and activity execution in distributed computing environments. In addition to these steps, it is also interesting to gather provenance information [Freire et al. 2008] describing the history of composition and execution of computational processes, such as scientific workflows. This allows, for instance, determining how data sets were derived from other data sets and computational activities. They allow for the precise description of how a computational process was planned, which is called *prospective provenance*, and what occurred during execution, which is called *retrospective provenance*. Applications of provenance information include reproducibility and validation of computational

scientific experiments, scientific workflow reuse, data quality evaluation and attribution of scientific results.

Parallel and distributed computing is frequently used for scaling up the execution of computationally and data-intensive scientific workflows. These computations often involve the collaboration of a large number of geographically distributed individuals and heterogeneous resources. Grid computing [Foster 2001] was proposed to allow for managing users and computational resources in these distributed computing environments that often include multiple institutions. Finally, scientific gateways [Wilkins-Diehr et al. 2008] consolidate this functionality to provide grid-enabled scientific applications through web interfaces. In this work, we describe an experiment of implementing a scientific gateway to provide pre-existing Bioinformatics workflows in the Brazilian National High-Performance Computing Network (SINAPAD). It allows users to execute these scientific workflows using computational resources geographically distributed through an easy-to-use web interface.

This work is organized as follows: in section 2, we describe how SciCumulus and Swift were integrated to the grid middleware [Lima et al. 2005] and scientific gateway toolkit [Gomes et al. 2015] of SINAPAD. In section 3, we evaluate this integration by experimenting with two pre-existing scientific workflows, SwiftGecko [Mondelli et al. 2015] for comparative genomics, and SciPhy [Ocana et al. 2011] for phylogenetics. In the same section we present a performance evaluation of SwiftGecko. In section 4, we compare the integration to related efforts in this area. Finally, in section 5, we describe future work and make some concluding remarks.

2. Integrating Scientific Workflows with Grid Middleware

SINAPAD uses the CSGrid [Lima et al. 2005] middleware to support the usage and management of distributed computational resources. It offers features to integrate resources and applications as well as to manage data and users with a very rich and flexible authentication and authorization scheme. Through the system, users have access to a workspace in which they can find and execute all available applications as well as add new applications. On top of this infrastructure, CSGrid offers two key entry points for users: a Java desktop client and a service bus based on CORBA (called OpenBus) through which job (OpenDreams) and file (HDataService) management services are provided as secure APIs. To facilitate the access to these CSGrid services through the Web, the mc2 toolset [Gomes et al. 2015] has been developed for supporting the implementation of web-based science gateways and Webservice/REST-based APIs.

For the integration of scientific workflows and CSGrid, two parallel SWfMSs were used: Swift [Wilde et al. 2011] and SciCumulus [de Oliveira et al. 2010]. Swift supports the specification, execution and analysis of scientific workflows through a high-level scripting language with many characteristics commonly found in functional programming. Swift has the capability to evaluate and execute in parallel the expressions whose data dependencies are met. It also has support for gathering and querying for core provenance information [Gadelha et al. 2012], related to datasets and computational activities managed during the workflow execution, and additional information that scientists are usually interested in querying, such as domain-specific annotations. The provenance is stored and can be accessed through a relational database (SQLite3 or PostgreSQL).

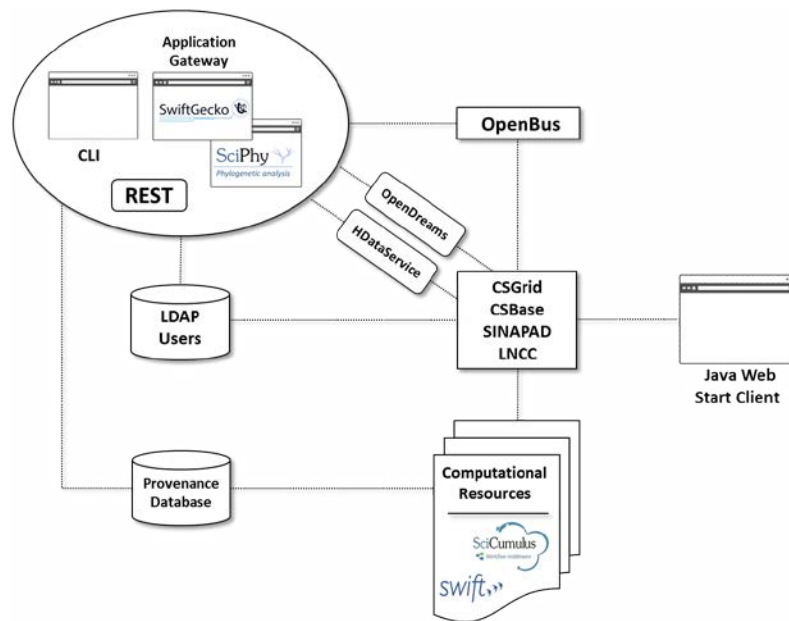


Figure 1. Conceptual schema of the integration between scientific workflows and CSGrid middleware

SciCumulus also allows scientific workflow modeling, execution and analysis. It was first developed as a cloud middleware to explore parameter sweep and data fragmentation parallelism in scientific workflow activities. Currently, SciCumulus supports the parallel execution in both clusters and cloud environments. The activities and the data flow between them are defined by the user through a file in XML format. SciCumulus uses an algebraic approach, focused on data, to the execution of workflows. This approach makes the provenance, which is stored using a relational database (PostgreSQL), a key point both for the execution and for the analysis of the workflow.

In order to deploy workflows in SINAPAD, as depicted in Figure 1, we first installed and configured SciCumulus, Swift and the applications used by the workflows on the computational resources managed by CSGrid. Through the CSGrid platform, we created two *algorithms* for executing these two workflows. “Algorithm” is a term used in CSGrid for describing a non-interactive application that can be dispatched directly by CSGrid clients with a clearly-defined set of input and output parameters. A CSGrid algorithm comprises an XML-based configuration file that defines these parameters and one or more submission scripts that deal with the specifics of the involved applications. For instance, in the case of the Swift we created scripts to configure the environment to allow the SWfMS to be executed from a submitting node in the HPC clusters managed by CSGrid. For SciCumulus, because of execution peculiarities and common allocation policy settings in HPC clusters available in SINAPAD, it was necessary to use a network proxy to allow SciCumulus access to an external provenance database.

3. Evaluation

For the evaluation, we will present two workflows, SciPhy [Ocana et al. 2011] and Swift-Gecko [Mondelli et al. 2015], that were deployed in SINAPAD.

In bioinformatics, phylogeny is an area that process biological sequences with the

objective of obtaining phylogenetic trees and other statistical calculations that support inferences about the evolutionary life and relationships of organisms of interest. Designing, executing and analyzing phylogenetic (or phylogenomic) experiments is not an easy task since they are complex and execute time/CPU-intensive applications. Then, they require technologies such as SWfMSs and HPC environments. To support these scenarios, the scientific workflow SciPhy for phylogenetic analyses was modeled using SciCumulus. SciPhy consume amino acid or nucleotide datasets and generate phylogenetic trees. It presents the following four main activities (Figure 2):

1. Multiple Sequence Alignment (MSA): To construct individual MSA. This activity receives files in the multi-FASTA format as input and it produces an MSA as output.
2. MSA Conversion: converts the output from the activity 1 to PHYLIP format.
3. Search for the best evolutionary model: It tests the PHYLIP files to find the best evolutionary model.
4. Construction of phylogenetic trees: it receives the outputs from activities 2 and 3 and constructs phylogenetic trees.

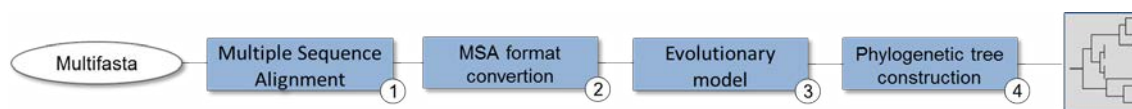


Figure 2. Conceptual view of the workflow SciPhy

Comparative genomics is an example of another area of bioinformatics in which experiments also aim to infer relationships between organisms. They are considered large-scale bioinformatics workflows as they require HPC to process the huge amount of data, known as biological big data. Addressing these data-intensive problems provides measures to perform, for instance, several analysis with the ultimate objective to better understand the biology of the organisms. The GECKO [Torreno and Trelles 2015] application is an example of workflow for genome comparisons designed to identify collections of high-scoring segment pairs (HSPs). To take benefit from distribution and paralelism, this application was implemented in Swift as the SwiftGecko¹ workflow, which is composed by three main computational modules, as presented in Figure 3 and described as follows:

1. Dictionary Calculation: Correspond to the creation of dictionaries for each sequence, composed by the activities 1-4 in blue boxes.
2. HSP Detection: Detection of the seed points or hits used to identify the HSP locations. It is composed by the activities 5-9 in green boxes.
3. Post-processing Module: to attach complementary tasks such as for performing statistical calculations or functional annotations, composed by the activity 10 in orange box.

In Figure 4 we show the user web interface of the SwiftGecko algorithm, obtained as a result of the integration. In this interface, the user must enter the required parameters to execute the SwiftGecko workflow without the need to install any program related to the workflow or the SWfMS that will manage its execution. The CSGrid middleware along

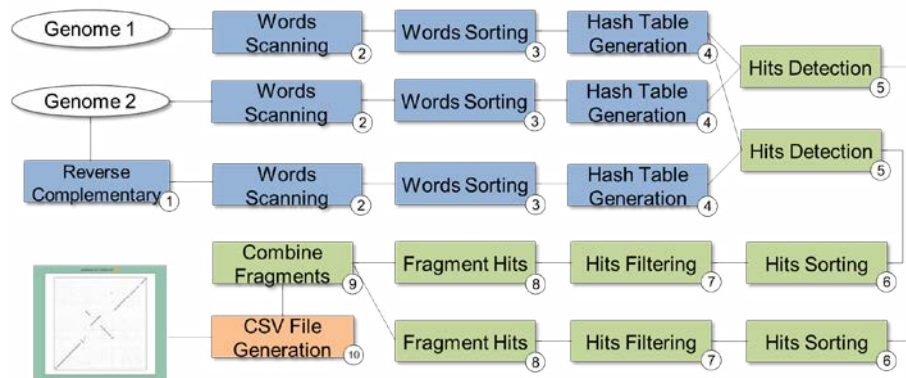


Figure 3. Conceptual view of the workflow SwiftGecko

Figure 4. Example of interface for workflow execution

with the execution environment of each SWfMS are responsible for the distribution of jobs through the SINAPAD network.

SwiftGecko was used as a case example in order to evaluate the performance of the integration. The workflow was executed with five bacterial complete genomes as input, resulting in 135 task executions. The executions were performed in a 72-node cluster with 16GB of RAM and 8 computing cores per node. In this evaluation, presented in Figure 5, we measure the performance (total execution time) of SwiftGecko for serial and parallel executions directly in the cluster and parallel through CSGrid, also using the same cluster. Results present performance improvements of up to 47% in the parallel execution time when compared to the sequential execution, which drops from around 8 minutes (using a single core) to around 4 minutes using 8 cores. The execution through CSGrid presents improvements of approximately 23% when compared to the sequential execution, dropping to 6 minutes. It is worth mentioning that the SwiftGecko applications use an out-of-core strategy and are I/O intensive. Therefore, scalability could be improved by using higher throughput storage systems or more efficient data management

¹Available at <https://github.com/mmondelli/swift-gecko>

strategies. Regarding the execution of the workflow through the portal, in addition to the submission and execution times, the file transfer time between CSGrid and the cluster is also taken into account. For this reason, we can understand the difference between the parallel executions in the cluster and through the portal.

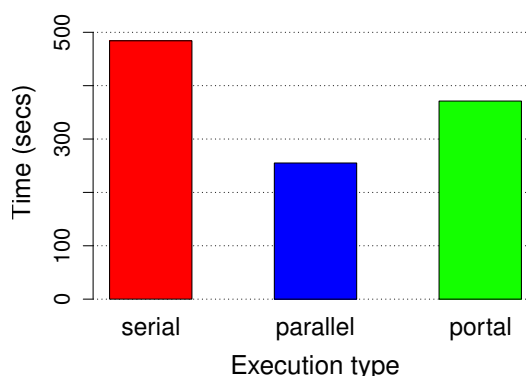


Figure 5. Evaluation of the execution times of SwiftGecko workflow

4. Related Work

There are some approaches in the literature that present mechanisms to execute large-scale scientific workflows. Globus Genomics [Madduri et al. 2014] presents a system for rapid analysis of large quantities of next-generation sequencing (NGS) genomic data. It integrates Galaxy WfMS for the specification of a data analysis workflow, and Globus tools for data movement and access-control management. However, the work is focused only in the cloud infrastructure, using Amazon services to storage data and to execute the data analysis pipeline.

Askalon [Nadeem et al. 2007] proposes an application development and computing environment to provide an invisible Grid to the application developers. It uses the Unified Modelling Language (UML) diagrams to represent the workflow graphically. It also provides an XML-based representation, to translate the UML diagram and pass it to the runtime system to be scheduled and executed. Although the application involves the concepts of workflows and Grids, instead of integrating existent systems, they built a single system from scratch.

5. Conclusion

Providing scientific workflows through an application management system that operates on a computational Grid environment as CSGrid, allows users to have access to more resources to execute their experiments. This is an important aspect for users who do not have the computing infrastructure needed to perform their executions. In this work, we presented an approach to integrate SWfMSs with grid-enabled scientific gateways, providing a way to run through an interface workflows that previously needed to be run via command line in the cluster. Thus, since the workflow is available through CSGrid, the execution process becomes a transparent task for the user. We evaluate a bioinformatics

workflow for genome comparisons, which have demonstrated satisfactory performance when executed through this infrastructure. As future work, we intend to support the provenance that is already captured by SWfMSs [de Oliveira et al. 2010] [Gadelha et al. 2011] to allow the analysis of information that is generated through the workflow execution.

In addition, we aim at integrating the SWfMS functionality with grid middleware in a native way, through a prototype of a SWfMS with support to CSGrid developed in SINAPAD called OSC Workflow Manager (GWO). This prototype uses a Open Scientific Connectors (OSC) language to describe workflows [Medeiros and Gomes 2012] in which the main feature consists in the interaction of tasks mediated by connectors. Tasks and connectors interfaces are called, respectively, from ports and roles. A workflow described in OSC is composed by linking tasks ports and connectors roles. Tasks, connectors, ports and roles can be associated with properties that allow its parameterization.

OSC has basic types, used to associate abstract workflows concepts with its concrete implementation. A task can be, for instance, an executable program or an encapsulated workflow and a connector may represent a file transfer or a character pipe. The language also allows types of non-functional attributes to be combined with basic types, such as task parallelism, fault tolerance and provenance tracking [Medeiros and Gomes 2013]. This combination is defined by the workflow developer, that may decide with great flexibility in which parts the treatment is performed.

The OSC language was developed based on the architectural description language known as Acme [Garlan et al. 2010]. Therefore, workflows described in OSC may be graphically modeled by using AcmeStudio tool. The executable tasks of the workflow defined during its modeling correspond to the CSGrid algorithms. After modeling, the workflow can be stored in the GWO workflow repository and then executed, so that the workflow model is transparent to the end user. Thus, to execute the workflow through the GWO, the user must define the input and output parameters required for its execution. The GWO is in charge of managing the execution through CSGrid, including parallel execution of independent tasks. It also takes responsibility for the processing of non-functional attributes included in the model by the workflow developer. In this way, the various features that are available externally to SINAPAD infrastructure will be available natively. This will allow for a better integration and control of the workflows within the platform.

References

- de Oliveira, D., Ogasawara, E., Baião, F., and Mattoso, M. (2010). SciCumulus: A Lightweight Cloud Middleware to Explore Many Task Computing Paradigm in Scientific Workflows. In *2010 IEEE 3rd International Conference on Cloud Computing*, pages 378–385. IEEE.
- Foster, I. (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of High Performance Computing Applications*, 15(3):200–222.
- Freire, J., Koop, D., Santos, E., and Silva, C. (2008). Provenance for Computational Tasks: A Survey. *Computing in Science & Engineering*, 10(3):11–21.
- Gadelha, L. M. R., Wilde, M., Mattoso, M., and Foster, I. (2011). Exploring provenance in high performance scientific computing. In *Proc. of the 1st Annual Workshop on High*

- Performance Computing meets Databases - HPCDB '11*, pages 17–20. ACM Press.
- Gadelha, L. M. R., Wilde, M., Mattoso, M., and Foster, I. (2012). MTCProv: a practical provenance query framework for many-task scientific computing. *Distributed and Parallel Databases*, 30(5-6):351–370.
- Garlan, D., Monroe, R., and Wile, D. (2010). Acme. In *CASCON First Decade High Impact Papers on - CASCON '10*, pages 159–173. ACM Press.
- Gomes, A. T. A., Bastos, B. F., Medeiros, V., and Moreira, V. M. (2015). Experiences of the Brazilian national high-performance computing network on the rapid prototyping of science gateways. *Concurrency and Computation: Practice and Experience*, 27(2):271–289.
- Lima, M. J. d., Melcop, T., Cerqueira, R., Cassino, C., Silvestre, B., Nery, M., and Uruahy, C. (2005). CSGrid: um sistema para integração de aplicações em grades computacionais. In *Salão de Ferramentas do XXIII SBRC. Anais do XXIII SBRC*, pages 1207–1214.
- Liu, J., Pacitti, E., Valduriez, P., and Mattoso, M. (2015). A Survey of Data-Intensive Scientific Workflow Management. *Journal of Grid Computing*, 13(4):457–493.
- Madduri, R. K., Sulakhe, D., Lacinski, L., Liu, B., Rodriguez, A., Chard, K., Dave, U. J., and Foster, I. T. (2014). Experiences building Globus Genomics: a next-generation sequencing analysis service using Galaxy, Globus, and Amazon Web Services. *Concurrency and Computation: Practice and Experience*, 26(13):2266–2279.
- Medeiros, V. and Gomes, A. T. A. (2012). Towards Fully Configurable Support to Non-Functional Attributes in Scientific Workflows. In *IEEE eScience Early Results and Works-in-Progress Poster Papers*, pages 2–3.
- Medeiros, V. and Gomes, A. T. A. (2013). Expressando Atributos Não-Funcionais em Workflows Científicos. In *Proc. of VII Brazilian e-Science Workshop*.
- Mondelli, M. L., Torreño, O., Ocaña, K. A. C. S., Mattoso, M., Wilde, M., Vasconcellos, A. T., Trelles, O., and Gadelha, L. M. R. (2015). SwiftGECKO: a provenance-enabled parallel comparative genomics workflow. In *Proceedings X-Meeting 2015*, page 268.
- Nadeem, F., Nerieri, F., Podlipnig, S., Qin, J., Siddiqui, M., Truong, H.-I., and Villazon, A. (2007). ASKALON : A Development and Grid Workflows. In *Workflows for e-Science*, pages 450–471. Springer.
- Ocana, K. A., Oliveira, D. d., Ogasawara, E., Davila, A. M., Lima, A. A., and Mattoso, M. (2011). SciPhy: A Cloud-Based Workflow for Phylogenetic Analysis of Drug Targets in Protozoan Genomes. In *Advances in Bioinformatics and Computational Biology - 6th Brazilian Symposium on Bioinformatics, BSB 2011. Proceedings*, pages 66–70.
- Torreno, O. and Trelles, O. (2015). Breaking the computational barriers of pairwise genome comparison. *BMC bioinformatics*, 16(1):250.
- Wilde, M., Hategan, M., Wozniak, J. M., Clifford, B., Katz, D. S., and Foster, I. (2011). Swift: A language for distributed parallel scripting. *Parallel Computing*, 37(9):633–652.
- Wilkins-Diehr, N., Gannon, D., Klimeck, G., Oster, S., and Pamidighantam, S. (2008). TeraGrid Science Gateways and Their Impact on Science. *Computer*, 41(11):32–41.

BioDSL: A Domain-Specific Language for mapping and dissemination of Biodiversity Data in the LOD

Kleberson J. do A. Serique¹, José L. Campos dos Santos², Dilvan A. Moreira¹

¹University of São Paulo (USP)
CEP: 13566-590 – São Carlos – SP – Brazil

²National Institute for Amazonian Research (INPA)
CEP.: 69060-001 – Manaus – AM – Brazil

junio.serique@gmail.com, laurindo.campos@inpa.gov.br, dilvan@gmail.com

Abstract. *Currently, Linked Open Data (LOD) have enabled integrated data sharing across disciplines over the Web. However, for LOD users, in areas such as biodiversity (which massively use the Web to disseminate data), the task of transforming data file contents in CSV (Comma Separated Value) to RDF (Resource Description Framework) is not trivial. We have developed a new approach to map data files in CSV to RDF format based on a domain-specific language (DSL) called BioDSL. Using it, biodiversity data users can write compact programs to map their data to RDF and link them to the LOD. Biodiversity vocabularies and ontologies, such as Darwin Core and OntoBio, can be used with BioDSL to enrich user data. Existing tools are exclusively focused on mapping (CSV to RDF), offering little or no support for linking data to the LOD (interconnecting user entities to LOD entities). They also are more complex to use than BioDSL.*

1. Introduction

In biology, data are understood as a collection of facts that require interpretation of their meaning in order to become knowledge. This interpretation is performed only by humans, but the interpretation of extreme large data sets is only possible with the use of computers, ideally, with high performance capabilities.

In recent years, biodiversity open data have become openly available online in sites such as GBIF¹ and SpeciesLink². However, these data are available, in most cases, in CSV (Comma Separated Values) and other formats that do not have explicit semantic information about data [Van Der Waal et al. 2014]. While online data aggregators (such as GBIF) have helped in increasing the amount of biodiversity data available in digital format, the meaning of these aggregated data is often ambiguous [Moura et al. 2012].

To tackle this problem, a recent and growing movement, called Linked Open Data (LOD), uses Semantic Web technologies for sharing and linking data over the Web. LOD refers to a set of best practices for publishing data with semantic information so that it can be interlinked, referenced across datasets, and explored by automatic processing [Berners-Lee 2006]. LOD sites use the RDF model to represent knowledge in a simple structure. In RDF, data is represented as statements. Each statement is represented

¹<http://www.gbif.org/>

²<http://splink.cria.org.br/>

by a triple (S, P, O), where S is a URI (Uniform Resource Identifier) for the subject, P is a URI for the predicate, and O is either an URI or literal for the object.

Nowadays, LOD sites contain a large number of linked openly available datasets that cover a wide variety of disciplines, including biology. They form the LOD data cloud. Even though there is a large amount of biological datasets in the LOD cloud, no significant amount of biodiversity datasets are available for reuse and sharing. Also, biodiversity is still absent from works related to the development of ontologies and Semantic Web technologies [Walls et al. 2014]. Ontologies are used to provide a common vocabulary for a domain of interest and define the logical relationships that hold between these vocabulary items [Matsubara et al. 2009]. Furthermore, formal ontologies facilitate the discovery of implicit knowledge, through the reasoning process.

Once available on the LOD cloud, biodiversity datasets become more easily accessible by analytical tools, such as scientific workflows. Such tools can be used not only to test and integrate biodiversity data, from different sources, but also to integrate them to sources from other scientific fields (equally important to understand biodiversity phenomena).

To increase the availability of biodiversity data on the LOD cloud, one approach is to convert data, in CSV and other tabular formats used by biodiversity researchers, to RDF. Current tools to map data from CSV format to RDF offer no or little support to link these data to data already available on the LOD cloud.

This paper presents BioDSL, a new approach for the mapping of biodiversity data, in tabular format (CSV), to RDF. BioDSL is part of a bigger project to build a semantic infrastructure for biodiversity data.

The remainder of this paper is structured as follows. Section 2 presents the related work. Section 3 presents the BioDSL language. Section 4 concludes with some suggestions for future work.

2. Related Work

The Semantic Web provides technologies conducive to scientific data integration and dissemination. The first step, to take advantage of such technologies, is to transform tabular data (CSV) into a format readable by machines, such as RDF. The next step is to follow Linked Open Data principles when generating these data[Berners-Lee 2006].

There are two general approaches to map CSV data to RDF, the row and cell-based translations. The row-based translation is the most common. It assumes that each row describes a subject and that each column represents a property[Dimou et al. 2014]. This approach allows us to quickly get RDF from a CSV file. The cell-based translation assumes that each row can represent more than one subject, or, in many cases, that properties may come from a vocabulary or ontology.

Currently, there are several tools to convert structured data, as CSV files, to RDF³. In order to find works related to BioDSL, a literature review was performed. Among the works found, we highlight here the ones closely related to our proposal, such as RDF

³<http://www.w3.org/wiki/ConverterToRdf>

Refine⁴, the SemantEco Annotator, the csv2rdf4lod⁵, the Apache Any23⁶, the recommendations of RDB2RDF⁷, the RML mapping language and the Sparqlification mapping language (SML)⁸. A more general overview of mapping tools for structured sources, such as CSV, is given in [Unbehauen et al. 2012].

The RDB2RDF working group of the W3C produced two recommendation for the conversion of Relational Database (RDB) and CSV data to RDF [Scharffe et al. 2012]. The first recommendation is the Direct Mapping of Relational Data to RDF⁹ that is similar to row-based translation. This approach allows to quickly get RDF from a CSV file, provided you follow the RFC4180¹⁰ standard: the file should provides a header for columns. The second uses the R2RML language¹¹ to perform the mapping and enable the use of terms from vocabularies and ontologies. R2RML mappings are themselves RDF graphs and are written down in Turtle syntax[Stadler et al. 2015].

The Apache Any23 (Anything To Triples) is a library, a web service and a command line tool that extracts structured data in RDF format from a variety of Web documents (CSV, HTML, Microformats, etc). It is used in a large number of applications, such as Sindice. Apache Any23 converts files in CSV to RDF following an extraction algorithm¹². This algorithm requires that: i) the CSV files should be compatible with the RFC4180 standard, headers will be used as RDF properties; ii) a base URI should identify the CSV file. However, to perform the interconnection of the CSV file records with other LOD entities, the user must put their entities URIs in a new column in the original CSV file.

The RDF Refine [Maali et al. 2011] is an extension to OpenRefine¹³, where the data can be converted to RDF. It requires that users define a skeleton for the RDF as a tree structure, then it creates the mappings between the columns in the table and the nodes of the tree. One of the highlights feature of the RDF Refine is the OpenRefine data reconciliation tool. It allows the user to perform searches for URI entities on the LOD, such as DBpedia¹⁴ to associate them with entities in the CSV file columns [Maali et al. 2011]. However, this procedure requires some adjustments, transposition and additions of columns, thereby altering the structure of the original CSV file.

The SemantEco Annotator is similar to RDF Refine. It provides a simple user interface, keeping all interactions for mapping the RDF separate from the content of the original data table, which remains unchanged [Seyed et al. 2013]. This annotator serves as a front-end to the csv2rdf4lod conversion tool.

The csv2rdf4lod tool [Lebo and Williams 2010] converts tabular data files, such as CSV, to RDF according to the interpretation of coded parameters using a specific vo-

⁴<http://refine.deri.ie>

⁵<https://github.com/timrdf/csv2rdf4lod-automation/wiki>

⁶<http://any23.apache.org/>

⁷<http://www.w3.org/2001/sw/rdb2rdf>

⁸<http://sparqlify.org/wiki/SML>

⁹<http://www.w3.org/TR/rdb-direct-mapping/>

¹⁰<http://www.ietf.org/rfc/rfc4180.txt>

¹¹<http://www.w3.org/TR/r2rml/>

¹²<http://any23.apache.org/dev-csv-extractor.html>

¹³<http://openrefine.org>

¹⁴<http://dbpedia.org>

cabulary for conversion¹⁵. This tool has two types of conversion: raw conversion and enhanced conversion. The first uses the row-based translation, while the latter needs a RDF file encoding a vocabulary to make the conversion. This file also provides the provenance history of the operations that led from the original CSV file.

SML is an RDB2RDF human readable mapping language with the same expressiveness as R2RML. SML provides virtual RDF graphs over relational databases or CSV files. This language is part of the Sparqlify platform that integrate several components of a Web application in a SPARQL endpoint. SML is based on R2RML. It has equal expressiveness, but it is less verbose than R2RML[Stadler et al. 2015].

The RML mapping language is an extended version of the R2RML language. RML keeps the mapping definitions, as in R2RML, but excludes database-specific references from the core model[Dimou et al. 2014]. The main difference between them are the input sources. In R2RML, they can be a database table or CSV file, while in RML they can be a broad set of input sources that, together, describe a certain domain.

Table 1. Comparison of the most common features of tools mapping CSV to RDF

Feature	RDF Refine	SemantEco	csv2rdf4lod	Apache Any23	RML	SML	BioDSL
Add Ontologies	-	O	X	-	X	X	X
Not Change CSV columns	-	X	X	X	X	X	X
LOD Instance matching	X	-	-	-	-	-	X
Make RDF model	X	X	X	X	X	X	X
Make OWL model	-	-	-	-	-	-	X
Infers OWL model	-	-	-	-	-	-	X
Export to SPARQL endpoint	-	-	-	-	-	X	X
Show mapping incoisistence	-	O	-	-	-	-	X
UI editor interface	X	X	-	-	-	O	O
Row-based translation	X	X	X	X	X	X	X
Cell-based translation	X	X	X	-	X	X	X
Visual-based mapping	X	X	-	-	-	-	O
Code-base mapping	-	-	X	-	X	X	X
CSV Data type verification	X	O	-	-	-	-	X
Reusability of the mappings	-	-	X	-	X	X	X
	X	Developed					
	O	Under Development					
	-	No Information					

Table 1 presents a feature list with comparisons between these tools features and BioDSL's. The main drawback of most of these data mapping solutions is the assumption that each row describes a single entity, such as row-based translation. Moreover, existing tools are exclusively focused on mapping the CSV data to the RDF model. They do not attempt to interconnect their source entities with existing entities in the LOD cloud.

Just RDF Refine makes an instance matching to automatically find LOD cloud entities URIs for reuse. This allows linking the data sets with other facts from other

¹⁵<http://purl.org/twc/vocab/conversion>

datasets (which allows data enrichment). Also, these tools do not allow that users create an integrated model with all used ontologies and mapping data as an OWL file. That could be used in other ontologies tools, such as Protégé. A missing, but useful (specially for novices), feature in these tools is a SPARQL endpoint. It allows other users to make queries against datasets using a Semantic Web standard. It also allows instance matching to improve the linking between datasets. A much desired feature is to show inconsistent mappings: it can be an ontological inconsistency or a syntax error, in case of script-based mapping. Data type errors must also be shown, thus the users can improve their data with accurate data type values.

To address the remaining gaps in the CSV to RDF conversion process of biodiversity data, BioDSL was developed as a DSL to make the complex process of conversion transparent to end users, covering from data integrity checks to ontological reasoning in OWL. In addition, BioDSL allows mapping of entities, present in CSV data sets, to existing entities in the LOD cloud, thus facilitating the integration of new data with other datasets already in the LOD cloud. It has the potential of enabling the discovery of new knowledge implicit in the data source.

3. The BioDSL Language

The BioDSL implementation uses the Groovy programming language. Groovy has native support for the development of DSLs and is compatible with the Java programming language and many Semantic Web's APIs.

In a typical biodiversity CSV file, each column may represent entities of different types (classes) and each line encode several entities related to each other, such as collected species, collection sites, institutions, collectors. These entities can also be repeated in different lines.

BioDSL has a declarative syntax based on objects and functions. For instance, the CSV file is represented by an object, called *csv*, whose properties refer to the column names (such as *csv.institutionID*). The main function, *Map*, does the actual mapping between CSV file and RDF. Other auxiliary functions define how URIs will be created, for instance, they can associate entities, from the CSV file, to LOD entities through their URIs. The following subsections will describes BioDSL syntax.

3.1. CSV file

Figure 1 shows the way to load a CSV file into a BioDSL script. The *addCsv* method, in Figure 1 line 1, receives a path or a URI to load a CSV file. The function *ignoreRow* declares numbers of lines that will be ignored (shown in Figure1 line 2). To skip more than one line, users must use a list of values corresponding to the line numbers (shown in Figure 1 line 3).

3.2. Loading Ontologies and defining Prefixes

Ontologies can be loaded into BioDSL and their vocabularies used to name and classify entities. The syntax to load ontologies is shown in Figure 1, the variables *ontobio* and *dwcterms*, on lines 6 and 7 respectively, store instances of ontology classes from a file and a web resource. BioDSL will load these ontologies into its RDF model.

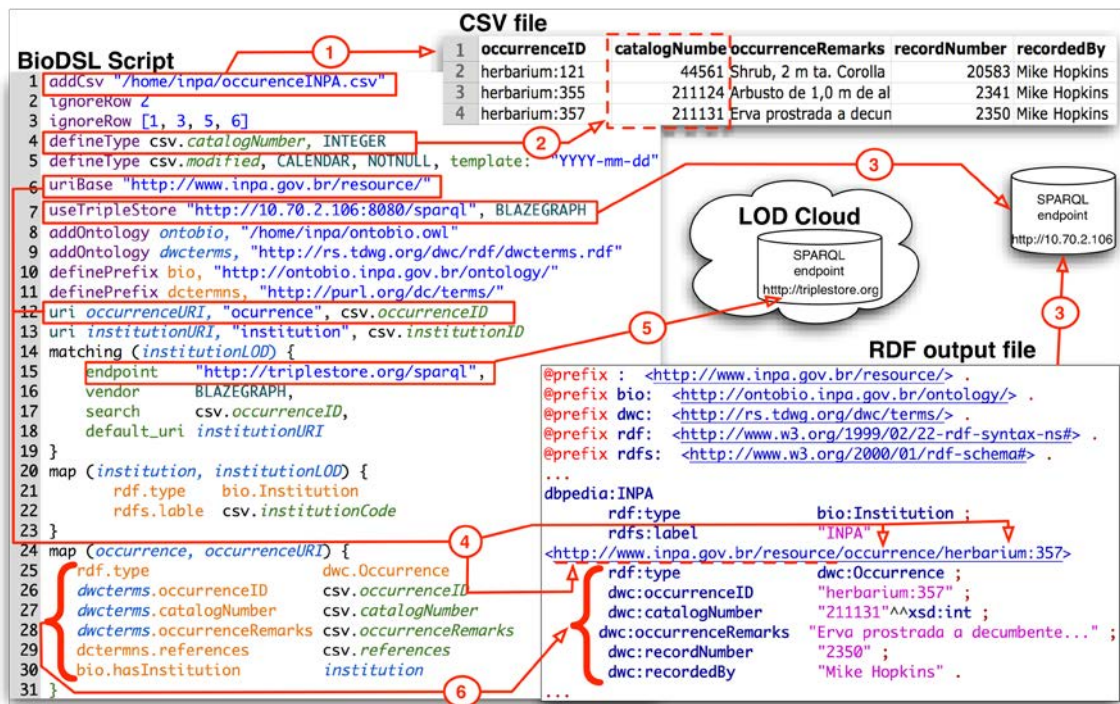


Figure 1. Configuring the CSV file and mapping to RDF in BioDSL

The *definePrefix* function, shown in Figure 1 lines 8 and 9, is used to simplify mapping ontology vocabularies to CSV entities. The objects in *bio* and *dwc* variables hold the ontology entities (classes, properties and individuals). These entities share the same prefix. In Figure 1, the *bio* and *dwc* variables are used later in the mapping, matching and gazetteer functions.

3.3. Defining a Base URI

The *uriBase* function, Figure 1 line 10, allows users to set a base URI for the generated RDF. By default, individuals (resources) will be created, using this base URI. Unless an individual reuses a URI from the LOD supplied by the Matching and Gazetteer functions.

3.4. The *uri* function

The *uri* function (Figure 1, line 12) is used to build URIs of individuals to be used in the final RDF. To build a URI, the *uri* function gets the URI Base value and concatenates it with its parameters. The user can pass as parameters a string, column name, using the *csv* object, or a list of column names. In Figure 1, line 12, the function uses the string *occurrence* and the values from the *occurrenceID* column. Arrow 4.2 (Figure 1) indicates the final value in the RDF file.

3.5. Matching function

The *matching* function is the most relevant BioDSL function. It allows users to link their entities (from the CSV file) to entities from SPARQL endpoints. Then, using a semantic search engine, BioDSL can retrieve URIs from known entities in the endpoints to be reused to enrich data, during the RDF generation. Many entities from CSV files are already present in the LOD cloud, in sites such as Geonames, DBpedia, the SWI

Gazzetter[Cardoso et al. 2015]. Entity reuse can give the final RDF a high degree of interconnectivity to the LOD cloud. This may represent a way to the discovery of new knowledge.

The first parameter of *matching* function is its identifier. It is used by the *map* function to call a particular mapping when it needs it. The other matching functions parameter are all named. The *endpoint* parameter (Fig1. line 15) contains the SPARQL endpoint URL. The *vendor* parameter contains information about the SPARQL endpoint vendor. There are different strategies for different vendors when BioDSL makes a semantic search. Currently, the BioDSL implementation supports two endpoint vendors, Blazegraph and Virtuoso. The *search* parameter contains a column name or set of column names (from the CSV file). The search is performed using the contents of the named column. If a set is used, the contents of each column are concatenated to form the search string. BioDSL uses the endpoints to searches string fields, such as labels, to find matching entities. If more than one entity is returned, the one with the best search score is used (different endpoints may use different search algorithms). If no entities are returned, the URI supplied by the *default* parameter is used. The optional parameter *type* is used to restrict the class (type) the entities, returned by the search string, should belong to. For instance, if the search string “Harpia harpyja” and type “http://dbpedia.org/ontology/species” are used, BioDSL will only return entities that belongs to this OWL class. This combination of best score and type restriction can be very effective in finding correct matchings, specially when using scientific terms, like species names. However, mistakes can happen and users should make tests to get acceptable error rates.

When the *matching* function retrieves an URI and reuses it in the final RDF as an entity name, BioDSL is saying that the entity, from the CSV file, and the entity, the URI represents, are the same. A link is created connecting the entity from the data set (being converted) to the LOD. Latter, that entity can be easily enriched using the facts already known about it in the LOD cloud.

3.6. Map function

The *map* function builds the final entities in RDF. It uses URIs, supplied by the user or generated by the *matching* and *uri*, and build RDF triples using the entities from the CSV file. Each triple add some information about the entity. For each data entity type, an user wants to map, a *map* function should be created.

The first parameter of the *map* function is its identifier. Using it, entities mapped by it can be referenced in other maps. In Figure 2, line 31, *institution* refers to the entity mapped by the *map*, in line 20. More specifically, to the institution referenced on the same line, in the CSV file, as the occurrence. By default, the *map* function identifier is also used to build the entity URI as the concatenation of the base URI (section 3.3), identifier and the line number in the CSV file. However, users can provide a *uri* or *matching* function identifier, as shown in Figure 2, labels “a” and “b”. In this case, the URI returned by the functions will be used.

Once a URI is defined for the RDF entity, BioDSL maps its properties. In the function body, lines between the characters, users can define RDF property/values pairs. For property URIs and values (RDF object), they use prefix objects (created by a prefix

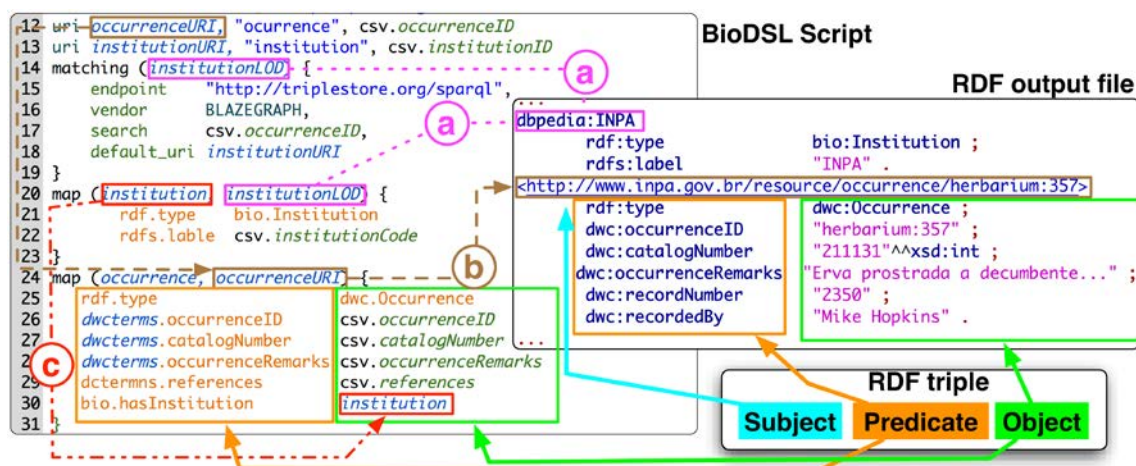


Figure 2. Details of how BioDSL generates RDF triples from maps, matches and uris: a) Using the matching function to find subject URLs from the LOD; b) Using the uri function to generate subject URL; c) Links between entities

function), shown in Figure 2 line 21, or a string representing a valid URI. Additionally, values can also use CSV column names (in this case, the column value is used, Figure 2 lines 22, 26 to 29) or an identifier of another map (Figure 2 label “c”). For each property/value pair, a triple is generated using the entity URI as subject, the property URI as predicate and the value as object.

4. Conclusion

In this paper, we presented a new approach for mapping open biodiversity data to RDF triples. It enables biodiversity users to easily take advantage of Semantic Web technologies, such as data enrichment with LOD information, data interlinking with known LOD entities, SPARQL endpoint, etc. This is all done in a less complicated way to users than other current mapping tools.

A small group of test users, involving experts in Semantic Web and biodiversity (from the National Institute for Amazonian Research - INPA), has been involved in the BioDSL development. For testing, it is being used open biodiversity data from Brazilian institutions, users of the GBIF, SpeciesLink platform and SiBBR¹⁶. These data are in spreadsheet format and can be converted to CSV format, without loss of information. The current BioDSL implementation only supports data in CSV format.

A Web based IDE for BioDSL is under development. Users will be able to write BioDSL scripts with suggestions and feedback in a rich interactive environment. Users will be able to create, share, and publish these scripts and their data, as RDF triples (in a file or in a SPARQL endpoint) contributing to the LOD cloud growth.

5. Acknowledgment

We would like to thank ICMC-USP, LIS-INPA and CAPES (The National Council for the Improvement of Higher Education - Brazil), for partially supporting this work. Thanks are also due to our colleagues for ideas and for revising this text.

¹⁶<http://www.sibbr.gov.br>

References

- Berners-Lee, T. (2006). Design issues: Linked data.
- Cardoso, S. D., Amanqui, F. K., Serique, K. J. A., dos Santos, J. L. C., and Moreira, D. a. (2015). SWI: A Semantic Web Interactive Gazetteer to support Linked Open Data. *Future Generation Computer Systems*, pages –.
- Dimou, A., Sande, M. V., Colpaert, P., Verborgh, R., Mannens, E., and Van De Walle, R. (2014). RML: A generic language for integrated RDF mappings of heterogeneous data. *CEUR Workshop Proceedings*, 1184.
- Lebo, T. and Williams, G. T. (2010). Converting governmental datasets into linked data. In *Proceedings of the 6th International Conference on Semantic Systems - I-SEMANTICS '10*, page 1, New York, New York, USA. ACM Press.
- Maali, F., Cyganiak, R., and Peristeras, V. (2011). Re-using Cool URIs: Entity Reconciliation Against LOD Hubs. *LDOW*.
- Matsubara, W., Kusano, K., Bannai, H., and Shinohara, A. (2009). Language and Automata Theory and Applications. *Lata*, 5457:578–587.
- Moura, A. M. D. C., Porto, F., Poltosi, M., Palazzi, D. C., Magalhães, P., and Vidal, V. (2012). Integrating Ecological Data Using Linked Data Principles. In *Joint V Seminar on Ontology Research in Brazil*, pages 156–167.
- Scharffe, F., Ateazing, G., and Troncy, R. (2012). Enabling linkeddata publication with the datalift platform. *Proc. AAAI workshop on . . .*
- Seyed, P., Chastain, K., Ashby, B., Liu, Y., Lebo, T., Patton, E., and McGuinness, D. (2013). Semanteco annotator. *CEUR Workshop Proceedings*, 1035:161–164.
- Stadler, C., Unbehauen, J., Westphal, P., Sherif, M. A., and Lehmann, J. (2015). Simplified RDB2RDF Mapping. *Proceedings of the 8th Workshop on Linked Data on the Web (LDOW2015), Florence, Italy*.
- Unbehauen, J., Hellmann, S., Auer, S., and Stadler, C. (2012). Knowledge extraction from structured sources. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7538:34–52.
- Van Der Waal, S., Wecl Cel, K., Ermilov, I., Janev, V., Milosevic, U., and Wainwright, M. (2014). Lifting open data portals to the data web. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8661:175–195.
- Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P. L., Davies, N., Endresen, D., Gandolfo, M. A., Hanner, R., Janning, A., Krishtalka, L., Matsunaga, A., Midford, P., Morrison, N., Tuama, É. Ó., Schildhauer, M., Smith, B., Stucky, B. J., Thomer, A., Wieczorek, J., Whitacre, J., and Wooley, J. (2014). Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. *PLoS ONE*, 9(3).

Modelagem e Análise de Conformidade do Processo de Atracamento Molecular

Miller Biazus, Lucinéia Heloisa Thom, Márcio Dorn

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

mbiazus@inf.ufrgs.br, lucineia@inf.ufrgs.br, mdorn@inf.furb.br

Abstract. *The molecular docking process is complex, specially because it includes terms of a specialized vocabulary from the bioinformatics domain, which can lead to misunderstandings between the stakeholders. This article proposes a validated and verified molecular docking process model using the Business Process Model and Notation (BPMN) as graphical notation. The process model exposes a common understanding - based on the bioinformatics literature and on the opinion of domain experts - on how the process should be defined and executed and is used as a starting point to perform a conformance analysis of two different molecular docking softwares.*

Resumo. *O processo de atracamento molecular é complexo, especialmente devido ao vocabulário especialista da área de bioinformática, podendo levar a problemas de interpretação entre os stakeholders do processo. O presente artigo propõe um modelo validado e verificado para o processo de atracamento molecular utilizando a Notação e Modelo de Processos de Negócio (BPMN), reconhecida como padrão pela OMG. O modelo de processo expõe um entendimento consensual - com base na literatura de Bioinformática e na opinião de especialistas de domínio - sobre como o processo deve ser definido e executado, e é utilizado como ponto de partida para uma análise de conformidade entre dois softwares existentes para realizar atracamento molecular.*

1. Introdução

Bioinformática estrutural visa o entendimento de processos biológicos e a aplicação de técnicas computacionais para aprender e organizar informação estrutural associada a macromoléculas [Altman and Dugan 2005]. Atracamento molecular pode ser definido como uma metodologia que lida com o problema de prever a ligação não-covalente de um receptor (macromolécula, geralmente uma enzima ou proteína) e um ligante (molécula menor) em um nível atômico. A afinidade dessa ligação depende de uma função de avaliação e dos métodos computacionais aplicados [Guedes et al. 2013].

A modelagem de processos com vocabulário especialista e características dinâmicas, tais como os presentes em estratégias computacionais para atracamento molecular é muito complexa, principalmente porque requer o conhecimento de termos de um domínio específico que pode levar a problemas de interpretação, ambiguidades e falta de concordância entre os *stakeholders*. O processo de atracamento molecular inclui macroatividades como preparação do ligante e receptor, as quais podem ser simuladas por uma variedade de softwares (como DockThor [de Magalhães et al. 2014], Autodock Vina

[Trott and Olson 2010]). Contudo, cada software possui suas próprias técnicas para executar tais atividades, o que implica em diferentes processos de atracamento molecular, particularmente quando comparado com o processo proposto pela literatura de bioinformática estrutural .

Considerando o Gerenciamento de Processos de Negócio (BPM), a Notação e Modelo de Processo de Negócio (BPMN) é reconhecida como padrão pela *Object Management Group* (OMG) [Recker 2010]. BPMN vem sendo utilizada para uma grande variedade de propósitos de modelagem de processos, incluindo áreas como análise de negócios, análise de requerimentos, descrição, documentação, simulação e execução de processos [Mateja et al. 2015].

Este artigo propõe a modelagem do processo de atracamento molecular (validada e verificada), baseada na extração de informações da literatura do processo de atracamento molecular, representada através da BPMN. As principais contribuições do presente trabalho são a) documentação padronizada e sistemática do processo de atracamento molecular em uma notação de modelagem de processos; b) motivar o uso da BPMN para representação de processos complexos incorporados na literatura; c) apresentar um estudo de análise de conformidade entre os processos de dois softwares existentes de atracamento molecular.

O restante deste artigo está estruturado da seguinte forma: na seção 2 são introduzidos conceitos de atracamento molecular e de BPM necessários para o entendimento da abordagem proposta; na seção 3 é proposto um modelo em BPMN para o processo de atracamento molecular; nas seções 4 e 5 são descritas, respectivamente, a validação e verificação do modelo de processo; na seção 6 é apresentada uma análise de conformidade entre dois softwares de atracamento molecular; na seção 7 são apresentados trabalhos relacionados; por último, na seção 8 é apresentada a conclusão do trabalho.

2. Atracamento Molecular e Gerenciamento de Processos de Negócio

Atracamento molecular é uma abordagem *in silico* que consiste na predição de modos e afinidade de ligação entre um receptor (macromolécula), o qual possui um sítio de ligação de interesse, e um ligante (molécula), o qual atuará no sítio de ligação do receptor, inibindo ou ativando determinada função do mesmo. As moléculas receptor e ligante podem possuir um certo grau de flexibilidade, o que possibilita diferenciar o atracamento molecular em três abordagens: corpos rígidos (*Rigid-body docking*), ligante flexível (*Flexible-ligand docking*) e receptor flexível (*Flexible-receptor docking*). *Flexible-ligand docking* é a abordagem de atracamento molecular utilizada neste trabalho por ser o tipo de atracamento molecular mais comumente utilizado [Dhanik and Kavraki 2001].

O Gerenciamento de Processos de Negócio (BPM) contribui para a redução de custos, de tempo e taxas de erro em execuções de processos [Dumas et al. 2013]. Através da modelagem de processos e automação, organizações podem documentar, otimizar e controlar melhor seus processos de negócio. BPM tem sido explorado e utilizado em inúmeros domínios de aplicação, incluindo o domínio de negócios onde processos são, em geral, bem conhecidos e estáticos, e o domínio da bioinformática onde processos científicos apresentam características dinâmicas e requerem aspectos de flexibilidade e adaptabilidade quando implementados. O ciclo de vida de BPM tipicamente inclui as seguintes fases [Dumas et al. 2013]: identificação do processo, descoberta do processo,

análise do processo, redesenho do processo, implementação do processo e controle e monitoramento do processo. Neste trabalho são aplicadas as duas primeiras fases do ciclo de vida de BPM a fim de obter como resultado o processo *as-is* (ou seja, a modelagem do processo em seu estado atual) de atracamento molecular.

Considerando a fase de descoberta de processo, a notação gráfica BPMN possibilita a representação dos mais diferentes aspectos de um processo. Seu foco é ser um ponto intermediário de conexão entre a modelagem do processo de negócio e a implementação. Em vista disso, é utilizada para uma grande variedade de propósitos de modelagem de processos [Mateja et al. 2015], incluindo, principalmente, análise de negócios e documentação de processos. Diversos softwares podem ser utilizadas para modelagem de processos (por exemplo *Bizagi*, *Signavio* e *Bonita*).

3. Modelagem do processo de atracamento molecular

Em métodos tradicionais para realizar a modelagem o analista depende de informações obtidas e organizadas a partir de especialistas no domínio [Dumas et al. 2013]. Foram analisadas ao todo 37 referências¹ da literatura de bioinformática. As principais atividades de atracamento molecular foram identificadas e extraídas manualmente a partir da convergência dos dados entre as referências, ou seja, atividades que eram definidas ou pressupostas como existentes em todas as referências. Atividades específicas de técnicas particulares de atracamento molecular não foram incluídas.

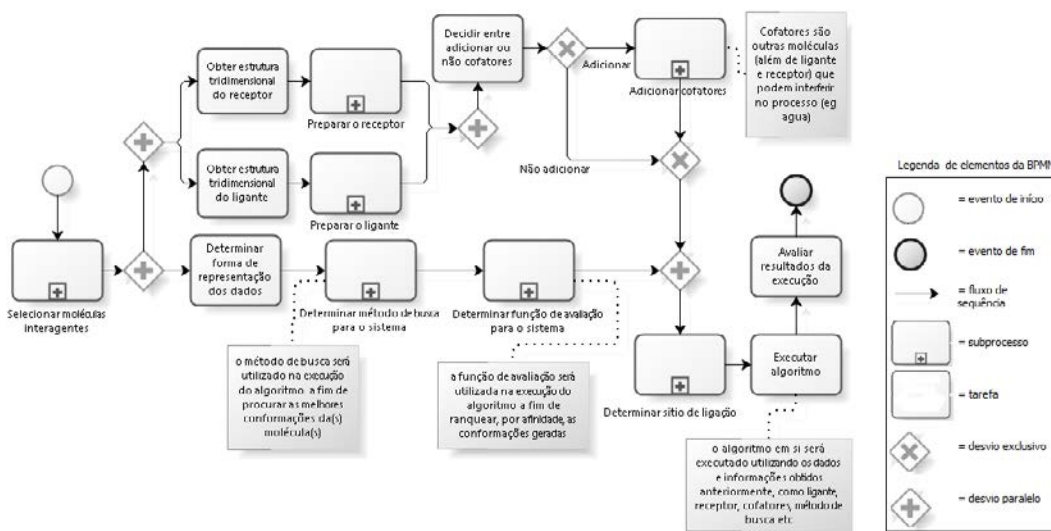


Figura 1. Modelo de processo - Atracamento molecular

Com o primeiro esboço do modelo do processo de atracamento molecular em mãos, foram conduzidos debates com três especialistas do domínio de bioinformática voltado a atracamento molecular. O objetivo foi refinar e complementar o modelo de processo de atracamento molecular obtido na literatura. O processo de atracamento molecular apresentado neste artigo foi modelado usando elementos da notação BPMN 2.0 [Allweyer 2010] e o software de modelagem Bizagi versão 2.9. A figura 1 exhibe

¹Disponibilizadas em <https://goo.gl/kIMwnt>

o processo principal de atracamento molecular. A modelagem do processo foi realizada com base nas sete diretrizes de modelagem de processo, incluindo utilização do padrão verbo-objeto para descrever tarefas, eventos de início e fim únicos, entre outras [Mendling et al. 2010]. Tais diretrizes são relevantes para minimizar a probabilidade de erros estruturais no modelo, assim como *deadlocks* e *livelocks*.

4. Validação da Modelagem do Processo de Atracamento Molecular

Para validar o modelo de processo de atracamento molecular apresentado na figura 1 foi utilizado um questionário². O questionário foi respondido por 9 pesquisadores da área de biologia molecular (especificamente com envolvimento em atracamento molecular), sendo três pesquisadores com conhecimento avançado em atracamento molecular, três pesquisadores com conhecimento intermediário e três pesquisadores com conhecimento básico.

Foi disponibilizado acesso ao material de um tutorial sobre BPMN para os pesquisadores que não conheciam a notação. A principal pergunta do questionário buscava investigar se o pesquisador, de fato, determinava o modelo como correto. Os 9 pesquisadores relataram que a modelagem do processo estava correta e, portanto, sem necessidade de alterações adicionais, indicando que o processo modelado está em conformidade com o processo existente na vida real.

5. Verificação da Modelagem do Processo de Atracamento Molecular

A notação BPMN, apesar de suas características favoráveis, como fácil entendimento e capacidade de representação de processos complexos, é ambígua em suas construções e deficiente em semântica formal, o que pode levar os modelos a exibirem estados indesejáveis como *deadlocks* e *livelocks*, além de inconsistência, ambiguidade e incompletude [Van Nuffel et al. 2009]. Para suprimir esses potenciais problemas, foi feita uma verificação na modelagem seguindo uma abordagem relatada por Dijkman e colaboradores (2008), onde utiliza-se um mapeamento de BPMN para Redes de Petri. Com tal mapeamento foi gerado um modelo na forma de Rede de Petri, o qual foi verificado com um software Woflan [van der Aalst 1999] a fim de analisar determinadas propriedades, como: verificação sintática (tarefas sem condição de entrada ou saída), detecção de construções suspeitas (fragmentos do modelo que podem ocasionar algum estado indesejável, como *deadlock*), detecção de transições mortas (nunca disparada em nenhuma instância do processo) e transições não-vivas (existe marca alcançável para a qual a transição é morta) e *Soundness*. Aalst (1997) define *soundness* como: “Para qualquer caso, o processo irá eventualmente terminar, e no momento que o processo terminar existirá uma marca em um lugar x e todos os outros lugares estarão vazios”.

Segundo Dijkman(2008) um processo BPMN é *bem formado* se possui as seguintes características: um evento de início possui apenas um fluxo de sequência de saída e não possui nenhum fluxo de sequência de chegada; um evento de fim possui apenas um fluxo de sequência de chegada e não possui nenhum fluxo de sequência de saída; atividades e eventos intermediários possuem apenas um fluxo de sequência de chegada e apenas um fluxo de sequência de saída; desvios divergentes possuem apenas um fluxo de sequência de chegada e mais de um fluxo de saída; desvios convergentes possuem vários fluxos de

²Disponibilizado em <https://goo.gl/aOwKQk>

sequência de chegada, porém apenas um fluxo de sequência de saída. O processo de atracamento molecular modelado, assim como cada um de seus subprocessos possui todas as condições citadas.

A Rede de Petri resultante do mapeamento é exibida na figura 2. O diagnóstico do software Woflan não revelou transições mortas ou não vivas, erros sintáticos e construções suspeitas. Logo, todas as propriedades estão corretas, e isso indica que a Rede de Petri relacionada ao processo de atracamento molecular modelado não apresenta inconsistências semânticas e está verificada formalmente e, portanto, tanto o modelo de processo de atracamento molecular mapeado para Redes de Petri quanto o modelo em BPMN não apresentem problemas semânticos e estruturais em sua modelagem.

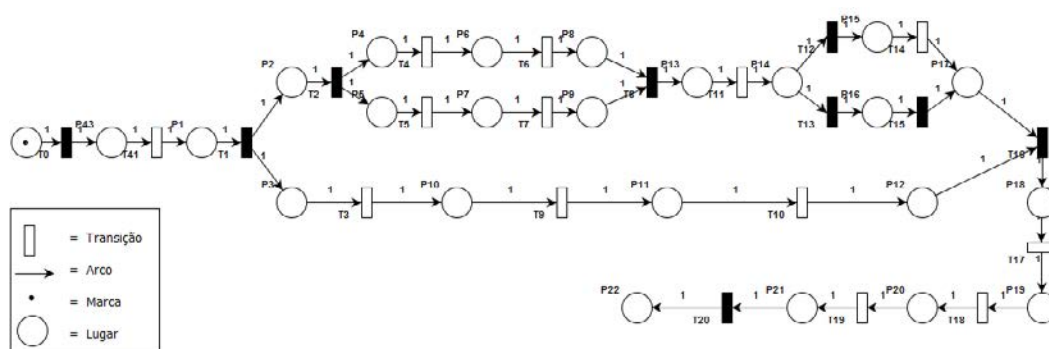


Figura 2. Rede de Petri do Processo de Atracamento Molecular

6. Análise de conformidade do modelo de processo de atracamento molecular com os processos de softwares de atracamento molecular

A mineração de processos representa um papel importante na área de BPM. O principal objetivo é a descoberta de modelos de processo baseada nos dados disponíveis de *logs* de eventos [Gehrke and Werner 2013]. Gehrke (2013) relata que a abordagem utilizada na mineração de processos para a obtenção de informações confiáveis sobre a execução real de um processo de negócio baseia-se na exploração de dados armazenados em sistemas, os quais são gerados durante o processamento de transações de negócios.

Foram analisados dois softwares de atracamento molecular: DockThor [de Magalhães et al. 2014] e Autodock Vina [Trott and Olson 2010]. Foram identificados 5 logs de eventos únicos em cada um dos dois softwares analisados. O *plugin* "BPMN analysis using heuristic miner" [Ayutaya et al. 2012] foi utilizado a fim de gerar um modelo em BPMN para DockThor (figura 3) e AutodockVina (figura 4) a partir dos *logs* de eventos gerados.

Fitness é uma medida de conformidade que indica o grau ao qual os traços de execução de um *log* podem ser associados com caminhos de execução válidos especificadas pelo modelo de processo [Rozinat and van der Aalst 2008]. No caso do Dockthor, existem diversas divergências (em relação ao modelo proposto) na ordem que as atividades são executadas, porém todas as atividades existentes nos *logs* de eventos do DockThor constam no modelo de processo de atracamento molecular, e não é necessária a criação artificial de uma marca na execução do processo: os *logs* podem ser executados do começo

ao fim sem nenhum problema, se adaptando perfeitamente ao modelo de processo, o que resulta em um *fitness* de 100%.

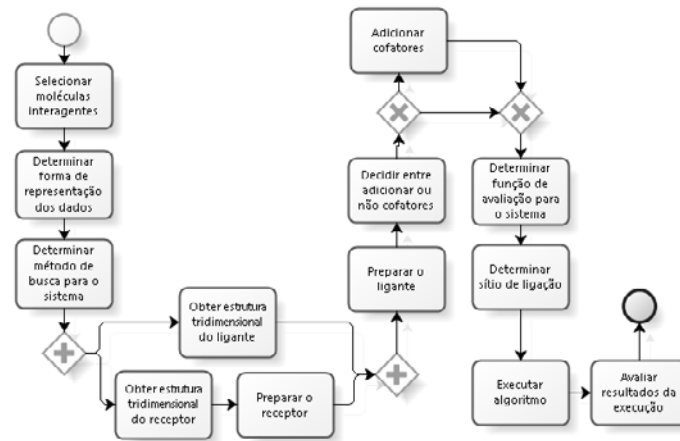


Figura 3. Processo minerado do software DokThor

Analogamente ao DockThor, todas as atividades dos *logs* de eventos do Autodock Vina constam no modelo de processo de atracamento molecular. São encontradas, no entanto, diversas divergências na ordem que essas atividades são executadas. O *replay* de todos os logs de eventos, ainda, requer a criação artificial de uma marca para a execução do processo, portanto não apresentando 100% de *fitness*.

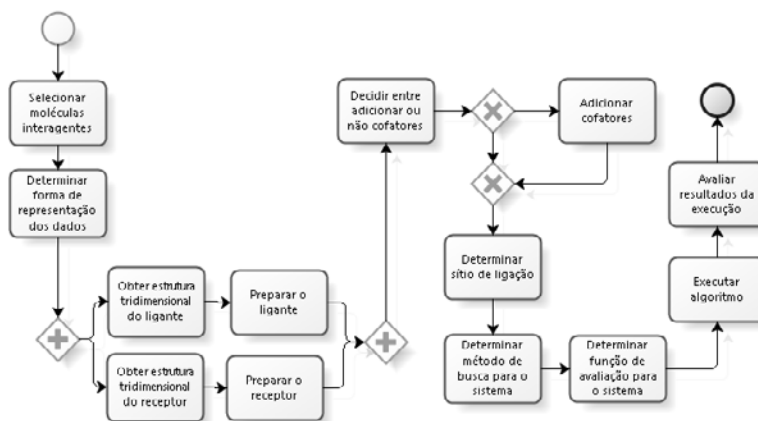


Figura 4. Processo minerado do software Autodock Vina

7. Trabalhos relacionados

Existem vários trabalhos que abordam técnicas para descoberta e mineração de processos. Em Gonçalves, Santoro e Baiao (2009) é explorada a técnica narrativa associada a mineração de texto e interpretação de linguagem natural para geração automática de modelos de processo. No trabalho consta que os participantes do processo descrevem a maneira de agir através de histórias de usuário.

Em Santoro e colaboradores (2010) os autores exploram um método para capturar histórias de um determinado grupo como alternativa para as entrevistas individuais

tradicionalmente utilizadas para elicitar processos. Segundo os autores, quando em coletivo, os participantes são capazes de verificar de forma cooperativa os processos descritos pelos outros integrantes por meio de suas perspectivas e adicionar novas histórias com descrições adicionais ou alternativas.

No contexto de extração de modelos a partir da literatura, existem algumas contribuições como o trabalho apresentado por Friedrich e colaboradores (2011), onde os autores apresentam um método automático para gerar modelos em BPMN a partir de textos em linguagem natural. Os métodos mais utilizados em trabalhos de elicitação de processos de negócio são a entrevista e a análise da literatura existente. Segundo Silva e colaboradores (2012), as informações obtidas a partir de diferentes técnicas de elicitação (por exemplo literatura, *brainstorms* e entrevistas) devem ser agrupadas e organizadas, e informações conflitantes devem ser resolvidas. Essas ações irão, então, refletir no modelo de processo desejado, que representa o conhecimento obtido na elicitação. A abordagem utilizada neste artigo para modelar processos a partir da literatura não é complementar às abordagens existentes, sendo que os modelos obtidos devem ser complementados (se necessário) e validados por especialistas de domínio.

Poucos trabalhos têm sido feitos em relação à documentação sistematizada de processos de bioinformática. A documentação é limitada à tutoriais e guias de usuário para softwares específicos de atracamento molecular.

8. Conclusão

No presente trabalho foi apresentado um modelo, na notação BPMN, para o processo de atracamento molecular, validado a partir de questionários com pesquisadores e validado a partir de um mapeamento para Redes de Petri, um formalismo que possibilita diagnósticos e análises formais complexas. Após, foi feita uma análise de conformidade com o modelo extraído a partir de mineração dos processos de dois softwares utilizados para atracamento molecular (DockThor e Autodock Vina). Tal análise mostra que apenas o *fitness* do processo do software Dockthor é perfeito.

O conceito de BPM, além de sua utilidade para o trabalho, foi particularmente útil para modelar um processo complexo com vocabulário especialista, como atracamento molecular, usando-se a literatura como referência e com interferência mínima de entrevistados. O modelo de processo gerado na notação BPMN apresentou alta aceitação dos usuários, os quais deram *feedback* positivo em relação ao nível de compreensão do modelo. Ainda, alguns usuários elogiaram a iniciativa e reiteraram que a modelagem utilizando BPMN é bastante clara e facilita o entendimento do processo.

Trabalhos futuros são sugeridos a fim de processar as demais fases do ciclo de vida de processo não abordadas no trabalho (por exemplo implementação e execução).

9. Agradecimentos

Esta pesquisa é suportada pelo Programa SticAmSud da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e pela Fundação de Ampara à Pesquisa do Estado do Rio Grande do Sul - Fapergs.

Referências

- Allweyer, T. (2010). *BPMN 2.0*. BoD.
- Altman, R. B. and Dugan, J. M. (2005). *Defining Bioinformatics and Structural Bioinformatics*, pages 1–14. John Wiley Sons, Inc.
- Ayutaya, N., Palungsantikul, P., and Premchaiswadi, W. (2012). Heuristic mining: Adaptive process simplification in education. In *ICT and Knowledge Engineering 2012 10th International Conference on*, pages 221–227.
- de Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C., and Dardenne, L. E. (2014). A dynamic niching genetic algorithm strategy for docking of highly flexible ligands. *Information Sciences*, pages 206–224.
- Dhanik, A. and Kavraki, L. E. (2001). Protein–ligand interactions: Computational docking.
- Dumas, M., Rosa, M. L., Mendling, J., and Reijers, H. A. (2013). *Fundamentals of Business Process Management*. Springer.
- Gehrke, N. and Werner, M. (2013). Process mining.
- Goncalves, J., Santoro, F., and Baiao, F. (2009). Business process mining from group stories. In *Computer Supported Cooperative Work in Design, 2009. CSCWD 2009. 13th International Conference on*, pages 161–166.
- Guedes, I. A., Magalhães, C. S., and Dardenne, L. E. (2013). Receptor–ligand molecular docking. *Biophysical Reviews*, 6(1):75–87.
- Mateja, K., Gregor, J., Marjan, H., and Gregor, P. (2015). Business process model and notation: The current state of affairs. *Computer Science and Information Systems*, 12:509, 539.
- Mendling, J., Reijers, H. A., and van der Aalst, W. M. P. (2010). Seven process modeling guidelines (7pmsg). *Inf. Softw. Technol.*, 52(2):127–136.
- Recker, J. C. (2010). Opportunities and constraints : the current struggle with bpmn. *Business Process Management Journal*, 16(1):181–201.
- Rozinat, A. and van der Aalst, W. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64 – 95.
- Santoro, F. M., Borges, M. R. S., and Pino, J. A. (2010). Acquiring knowledge on business processes from stakeholders’ stories. *Adv. Eng. Inform.*, 24(2):138–148.
- Trott, O. and Olson, A. J. (2010). Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry* 31, pages 455–461.
- van der Aalst, W. M. P. (1999). Woflan: A petri-net-based workflow analyzer. *Syst. Anal. Model. Simul.*, 35(3):345–357.
- Van Nuffel, D., Mulder, H., and Van Kervel, S. (2009). Enhancing the formal foundations of bpmn by enterprise ontology. In *Advances in Enterprise Engineering III*, volume 34 of *Lecture Notes in Business Information Processing*, pages 115–129. Springer Berlin Heidelberg.

Laboratório Interinstitucional de e-Astronomia: passado, presente, e futuro

Luiz A. Nicolaci da Costa^{1,2}, Angelo Fausti Neto^{2,3}, Marcio A. G. Maia^{1,2}, Ricardo L. C. Ogando^{1,2}, Riccardo Campisano^{2,4}

¹Observatório Nacional - MCTI

²Laboratório Interinstitucional de e-Astronomia – LIneA - MCTI

³LSST Corporation - USA

⁴Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – CEFET/RJ
{ldacosta,ogando,maia,angelofausti,riccardo.campisano}@linea.gov.br

Abstract. *We describe how LIneA was created, what it has produced in terms of software and hardware infrastructure in the last 10 years of work on Dark Energy Survey and in preparation to many other projects, such as LSST, when an e-science center in Brazil will be needed to handle huge data volume, velocity, and variability. Parallelism, provenance, and visualization are some of the challenges faced at LIneA in order to generate products and achieve scientific results.*

Resumo. *Descrevemos como o LIneA foi criado, e o que produziu em termos de software e hardware nos últimos 10 anos de trabalho no Dark Energy Survey e em preparação para vários outros levantamentos, tal como o LSST, quando um centro de e-ciência no Brasil será necessário para atender seu grande volume, velocidade, e variabilidade de dados. Paralelismo, proveniência, e visualização são alguns dos desafios encarados pelo LIneA a fim de gerar produtos e alcançar os resultados científicos.*

1. Introdução

O desafio de poder atuar no levantamento *Dark Energy Survey* (DES, Flaugher et al. 2015) motivou, há quase 10 anos, a formação de uma rede de pesquisa composta por pesquisadores de diversas instituições e a formação de um time de tecnólogos para dar suporte técnico ao desenvolvimento de ferramentas que facilitassem o manuseio de grandes volumes de dados. Estas necessidades levaram a proposta da criação de um projeto estruturante no Observatório Nacional (ON) conhecido como Astrosoft e finalmente a criação do Laboratório Interinstitucional de e-Astronomia (LIneA), um centro de e-Ciência voltado para a Astronomia. Este centro permite que pesquisadores brasileiros se engajem na ciência de grandes levantamentos astronômicos. A missão do LIneA é criar soluções de software e montar uma infra-estrutura física para lidar com o problema de *Big Data* gerado por projetos de *Big Science*. Por *Big Data* entenda-se projetos que envolvam a aquisição de dados que satisfaçam um dos seguintes quesitos: velocidade, volume ou variedade. Por *Big Science* entenda-se problemas de ciência fundamental na fronteira do conhecimento, os quais devido à complexidade inerente requerem projetos de *Big Data* e tipicamente grandes colaborações científicas, como por exemplo, o estudo da Energia Escura. A evolução do LIneA nos últimos 5 envolveu o desenvolvimento de um sofisticado portal científico que está em constante aprimoramento possuindo instâncias de operação em vários centros internacionais como Cerro Tololo Inter-american Observatory (CTIO), Fermilab, e National Center for Supercomputing Applications (NCSA). O laboratório estabeleceu colaborações técnicas

com centros como Lawrence Berkeley National Laboratory (LBL), Fermilab, NCSA, Large Synoptic Survey Telescope (LSST, Abell et al. 2009) e Stanford Linear Accelerator Center (SLAC). Expandiu suas metas promovendo a entrada de pesquisadores brasileiros em novos projetos internacionais como Sloan Digital Sky Survey IV (SDSS-IV), Dark Energy Spectroscopic Instrument (DESI, Eisenstein et al. 2015), e LSST em colaboração com os principais centros de pesquisa no mundo, apoiando pesquisa de ponta em astronomia. O LIneA também promove a formação de um novo tipo de profissional capaz de atuar como protagonista em grandes colaborações internacionais e capacitado para a nova ciência de dados.

2. Infraestrutura física

Para atender diferentes necessidades computacionais dos projetos apoiados pelo LIneA, um centro de armazenamento, processamento, análise e distribuição de dados foi construído ao longo dos anos. Atualmente o centro tem mais de 100 equipamentos que fazem parte de uma arquitetura desenhada para atender as principais demandas das aplicações científicas, tais como: altas taxas de transferência de dados; armazenamento de grandes quantidades de dados; consulta eficiente de banco de dados; processamento paralelo; hospedagem de serviços críticos para a operação e colaboração. Destacam-se os seguintes elementos: *Cluster* de processamento para produção (38 nós, 912 núcleos), *Cluster* de processamento para testes de integração (4 nós, 72 núcleos), *Cluster* Lustre (6 nós, sistema de arquivos compartilhado de alto desempenho), Mass Storage (armazenamento de dados), Banco de dados do SDSS (MS SQL), Banco de dados do DES (PostgreSQL), Cluster de serviços (VMs) onde se encontram as ferramentas colaborativas git, twiki, repositório de documentos, máquina dedicada ao desenvolvimento, e sistema de Transferência de Dados (DTS).

3. Desenvolvimento de software

O desenvolvimento do portal teve início em 2007 e seu desenho básico evoluiu muito desde então. Cerca de 30 profissionais de TI contribuíram para o seu desenvolvimento, sem contar alunos, pós-doutorandos e pesquisadores. Vários produtos já foram entregues e estão disponíveis para diferentes comunidades. Por exemplo o *Quick Reduce* (QR), o *Data Server*, Criação de Catálogos Científicos, e *Workflows* Científicos. O QR é um sistema para diagnóstico em “tempo real” das exposições da câmera do DES (DECAM) instalada no telescópio Blanco de 4 metros no CTIO, Chile (Fausti Neto et al. 2013). *Data Server* é um sistema disponível no LIneA e no Fermilab que permite: acompanhar o progresso do levantamento DES, verificar visualmente a qualidade das imagens, permitindo marcar defeitos, rejeitá-las e fazer comentários, sobrepor objetos de catálogos (internos e externos ao DES) às imagens do levantamento, realizar buscas nos catálogos de objetos produzidos a partir destas imagens, criar ou carregar listas de objetos de interesse (ex. lentes fortes), visualizar estes objetos e criar “recortes de imagens” para referência futura ou publicação. A criação de catálogos para análises científicas específicas é talvez um dos maiores desafios de levantamentos fotométricos de grandes áreas do céu tendo em vista o volume de dados, o grande número de atributos associados a uma fonte (~900), o grande número de decisões que devem ser tomadas ao longo do caminho (ex. cortes em magnitudes, *flags*, e regiões rejeitadas) e a necessidade de preservar sua memória de produção. O último estágio do sistema de análise *end-to-end* (da Costa et al. 2013) é feito pelos *workflows* científicos que usam os catálogos preparados no estágio anterior para realizar diferentes análises científicas. Para controlar a execução de todos esses passos foi criada uma tela de *Dashboard* (Figura 1). O *Dashboard* consulta o banco administrativo que registra processos executados, seu status, e proveniência para um dado *release* do DES. Nessa tela central também se pode acessar e compartilhar os resultados de cada processo.

The screenshot shows the DES Science Portal Dashboard with the following data:

Pipeline	Start	End	Duration	Runs	Status
Data Installation					
Install Catalogs	2016-01-29 08:27:26	2016-01-29 12:18:50	03:51:24	1	●
Install Merge Mask	2016-03-19 18:37:32	2016-03-19 20:01:01	01:23:29	2	●
Install Depth Mask	2016-03-11 17:25:05	2016-03-11 17:27:57	00:02:52	2	●
Install Images					●
Install Depth Maps	2016-03-11 17:30:33	2016-03-11 17:57:39	00:27:06	1	●
Systematic Maps	2016-03-15 11:41:15	2016-03-15 13:04:09	01:22:54	2	●
Zeropoint Correction	2016-03-23 13:49:02	2016-03-23 14:19:21	00:30:19	2	●
Spectroscopic Sample					●
QA Credit					●
Total: 7:38:4					
Data Preparation					
SO Separation	2016-05-04 03:48:31	2016-05-04 04:06:30	00:17:59	5	●
Training Set Maker	2016-05-06 21:44:29	2016-05-07 05:57:50	04:13:21	5	●
Photo-z Training	2016-05-12 14:56:38	2016-05-12 18:58:12	02:01:34	5	●
Photo-z Computer	2016-05-10 16:12:01	2016-05-10 17:15:22	01:03:21	5	●
Galaxy Properties	2016-05-03 23:36:56	2016-05-04 02:53:17	03:17:22	1	●
Total: 10:53:57					
Value-added Catalogs					
Cluster	2016-05-06 21:02:26	2016-05-06 21:20:00	00:17:34	10	●
GE					●
GA					●
Total: 0:17:34					

Figura 1. Dashboard usado para monitorar o número de execuções, início, fim, duração e status do último processo de cada pipeline executado no portal.

4. Resultados

Entre os resultados obtidos ao longo dos últimos cinco anos, destacamos a criação e a operação de um laboratório multi-usuário de médio porte para o processamento e distribuição de dados, dedicados aos grandes levantamentos astronômicos e integrados a vários centros internacionais. O laboratório tem mais de 100 usuários cadastrados, atendendo pesquisadores de universidades no Rio de Janeiro, São Paulo, Paraná e Rio Grande do Sul e algumas instituições no exterior.

O desenvolvimento de um complexo e abrangente sistema de software para atender de forma eficiente as necessidades de validação, análise e mineração do grande volume de dados envolvido. Estes sistemas estão em operação no observatório do CTIO no Chile atendendo a todos os usuários da DECam, no Fermilab com mais de 180 usuários registrados, no LIneA atendendo aos pesquisadores brasileiros e, em breve, no NCSA e em Berkeley. Como consequência dos termos acordados pelo LIneA com os vários projetos, o laboratório assumiu compromissos que se estendem pelos próximos cinco anos. Além disso, a experiência adquirida mostra o papel vital que o laboratório pode e deve ter na preparação da comunidade brasileira para a era do LSST. Além disso, temos a formação de pesquisadores produzindo até agora 12 teses de mestrado e seis de doutorado. Quatro pesquisadores jovens foram reconhecidos como “construtores” no projeto DES pela contribuição equivalente a 24 meses de trabalho dada para a infraestrutura do projeto, no caso o desenvolvimento de ferramentas de análise integradas ao portal científico desenvolvido pelo LIneA.

A produção de 95 publicações em revistas arbitradas. Com quase 20 artigos por ano e da ordem de 5.000 citações acumuladas, atestam a produtividade e impacto científico dos projetos.

5. Perspectivas

Do ponto de vista técnico, o LIneA desenvolveu, implantou e manteve dois complexos sub-sistemas do Portal que o validaram, consolidando a posição do laboratório como um parceiro confiável na entrega de produtos úteis e de qualidade à colaboração internacional do DES. Estes subsistemas já estão em uso pela comunidade internacional no CTIO e no Fermilab e, em breve, no NCSA. A qualidade do produto motivou, por exemplo, a colaboração do DESI em convidar o laboratório para participar do

desenvolvimento do *Quick Look Framework* para validar 15.000 espectros a serem obtidos durante cada exposição do instrumento. A assinatura do acordo com o LSST garante que o investimento feito pelo LIneA para o DES poderá ser estendido e melhorado, integrando novas tecnologias nos próximos 5 anos, dando a oportunidade única da equipe brasileira estar altamente preparada para capitalizar nos dados do LSST, já no período de verificação científica, previsto para começar em 2020-2021. Até lá, será possível usar o próprio DES para desenvolver os vários *workflows* científicos.

No momento, junto com o Laboratório Nacional de Astrofísica (LNA), uma grande prioridade é estabelecer o processo de montagem do *Brazilian Participation Group* do LSST, construir a infraestrutura de software, e montar a base do Centro Regional de dados do LSST – e se for o desejo da comunidade, trabalhar para estender o acordo para incluir toda a comunidade astronômica em um projeto que mudará o jogo da astronomia na próxima década.

References

- Abell et al. (2009) “LSST Science Book, Version 2.0” ArXiv 0912.0201L
- da Costa et al. (2013) “End-to-end scientific processing in the LIneA Science Portal” CSBC 2013 - BreSci - VII Brazilian e-Science workshop
- Eisenstein et al. (2015) “The Dark Energy Spectroscopic Instrument (DESI): Science from the DESI Survey” AAS 225 336.05
- Fausti Neto et al. (2013) “Quick Reduce: Dark Energy Survey Camera mountain-top Quality Assessment tool and its master calibration pipeline” CSBC 2013 - BreSci - VII Brazilian e-Science workshop
- Flaugher, B. et al. (2015) “The Dark Energy Camera” AJ 150 150

Avaliação do desempenho relativo de bancos de dados NoSQL para arquivos de genótipos

Vinícius Junqueira Schettino¹, Arthur Lorenzi Almeida¹,
Leojayme Rodrigues Manso Silva¹, Wagner Arbex^{1,2,*}

¹Universidade Federal de Juiz de Fora – UFJF
Dep. de Ciência da Computação – DCC
Campus Universitário, 36.036-900, Juiz de Fora, MG, Brasil

²Empresa Brasileira da Pesquisa Agropecuária – Embrapa
R. Eugênio do Nascimento, 610, 36.038-330, Juiz de Fora, MG, Brasil

vinicius.schettino@ice.ufjf.br, lorenzi@ice.ufjf.br,
leojayme.manso@ice.ufjf.br, wagner.arbex@{ufjf.edu.br, embrapa.br}

Abstract. *The bioinformatics and the genomics work with nonstandard database. The classic databases are characterized by tabular design of data set and handling thereof by RDBMS. The genotype files are instances of non-classical databases and are characterized by being generated as text files, with unbalanced data, high dimensionality and the large volume of data, among other things. The RDBMS have not shown a great solution for the processing of such bases and, therefore, this study is evaluating the performance of NoSQL databases representing of two different data model families from testing scenarios for handling genotype files.*

Resumo. *A bioinformática e a genômica trabalham com bases de dados fora do padrão tradicional ou clássico que, por sua vez, caracterizam-se pela organização tabular e pelo tratamento destas em SGBDRs. Arquivos de genótipos são exemplos de bases de dados não clássicas e são caracterizados por serem gerados como arquivos textos, com dados desbalanceados, com alta dimensionalidade e por ocuparem muito espaço, entre outros aspectos. Os SGBDRs não têm se mostrado uma boa solução para o tratamento de tais bases e, portanto, o presente trabalho busca avaliar o desempenho relativo entre bancos de dados NoSQL que representam duas famílias de diferentes modelo de dados, a partir de cenários de teste para a manipulação de arquivos de genótipo.*

1. Introdução

Uma das atividades mais complexas das pesquisas em bioinformática tem sido a manipulação de banco de dados, que, em geral, tratam com 3 diferentes tipos de conjuntos de dados genômicos, quais sejam: os dados de sequenciamento ou sequências de nucleotídeos, também conhecidos como *reads*; os dados de genotipagem por marcadores moleculares do tipo SNP ou, simplesmente, “genótipo” ou “marcadores SNP”; e, ainda, os muitos arquivos de metadados dos *reads* e genótipos.

* Autor correspondente.

Os genótipos com marcadores moleculares do tipo SNP, segundo [Caetano 2009], surgiram na década passada, com as plataformas de genotipagem de marcadores moleculares de polimorfismo de base única ou *single polymorphism nucleotide* (SNP). No genoma humano, cerca de 0,1% dos nucleotídeos podem ser SNPs, segundo [Brookes 1999], o que pode determinar, aproximadamente, 3 milhões de SNPs por indivíduo, se for considerada a genotipagem de todos os SNPs do genoma do indivíduo.

As tecnologias que permitem a “leitura” de *chips* de DNA para a genotipagem de marcadores SNP, são divididas em baixa, média e alta densidade, possibilitando a genotipagem de alguns milhares de nucleotídeos – p. ex., 3 mil –, até centenas de milhares, que ultrapassam a 700 mil o número de marcadores SNP genotipados em um único ensaio.

A redução do custo dos serviços de genotipagem tem permitido aos projetos fazerem experimentos com *chips* de média ou alta densidade e, pela relação “custo × benefício”, os *chips* de média densidade – que geram genótipos de algumas dezenas de milhares até mais de 100 mil marcadores SNP –, talvez sejam os mais utilizados.

Neste contexto, o presente trabalho tem como objetivo avaliar o desempenho relativo dos SGBDs MongoDB e Tarantool¹, que representam duas diferentes “famílias” de bancos de dados NoSQL, a partir da manipulação de arquivos com dados de genotipagem. Para tanto, foi desenvolvido e executado um *benchmark*, a partir da geração de um arquivo de genótipos simulados de 5 mil indivíduos, com 56 mil marcadores SNP para cada indivíduo.

2. Considerações sobre bancos de dados NoSQL e trabalhos relacionados

Os sistemas de bancos de dados NoSQL, ou “not-only SQL”, surgiram para atender às necessidades da computação científica e de novos paradigmas, tais como, *big data* e *data science*, assim como, uma alternativa para os problemas de escalabilidade em armazenamento, paralelismo e gestão de grandes volumes de dados não estruturados, comuns na manipulação de bases de dados de genótipos.

[Edlich 2016] lista cerca de 12 modelos ou “famílias” de bancos de dados NoSQL, sendo os sistemas baseados “em colunas” (*column family database* ou *wide column store*), “em documentos” (*document store*), “em chave-valor” (*key-value* ou *tuple store*) e “em grafos” (*graph databases*) os mais utilizados ou referenciados, como pode ser visto em [Hecht and Jablonski 2011], [Li and Manoharan 2013], [Veronika Abramova; Jorge Bernardino and Pedro Furtado 2014] e [Aniceto et al. 2015].

Para este trabalho foram utilizados os bancos de dados MongoDB e Tarantool que representam, respectivamente, os SGBDs baseados em documentos e chave-valor. Além de serem utilizados em diversos trabalhos relacionados, como os citados anteriormente.

3. O experimento e seus materiais e métodos

O experimento baseou-se no desenvolvimento e na execução de um *benchmark* com o uso do Yahoo! Cloud Serving Benchmark (YCSB) 0.7, descrito em [Cooper et al. 2010], o qual é amplamente utilizado para a comparação e avaliação de desempenho de SGBDs. Seu ambiente de execução utilizou os SGBDs MongoDB 3.2.3 e Tarantool 1.6.8, em suas configurações padrões, com Debian 7.0 em um hardware de 16 GB de memória RAM.

¹Informações sobre MongoDB e Tarantool em [MongoDB, Inc. 2016] e [Mail.Ru Group 2016].

O YCSB é composto de um gerador de dados e um conjunto de testes de desempenho para avaliar operações de leitura, inserção, atualização etc.. Cada um dos cenários de teste é associado a uma *workload*, ou “carga de trabalho”, e é definido por um conjunto de características, como, p. ex., a porcentagem de operações de leitura/atualização, o número total de transações ou o número de registros da *workload*. O conjunto padrão de *workloads* do YCSB não atendeu às exigências do experimento a ser realizado, por não reproduzir as características de um arquivo de genótipo. Assim, o primeiro passo do experimento foi a criação de uma *workload* personalizada, simulando uma população de 5 mil indivíduos genotipados e, cada indivíduo, com um genótipo hipotético de 56 mil marcadores. O que gerou uma carga de 5 mil registros, com 56 mil campos por registro e 1 byte por cada campo.

O passo seguinte consistia na criação do banco de dados, no ambiente do YCSB, para cada um dos SGBDs utilizados e, o terceiro passo, na execução das operações dos cenários de teste em cada SGBD e na recuperação dos resultados para análise.

Houve, ainda, a necessidade de se definir as operações a serem executadas nos testes, pois, pela natureza dos trabalhos com arquivos de genótipo, operações de edição e inserção são pouco relevantes. Para tais arquivos, os desempenhos de leitura e/ou atualização aliados a escalabilidade são mais importantes e mais próximos de situações relevantes ao contexto dos mesmos. Portanto, no terceiro passo, o experimento foi dividido em duas etapas, os cenários C1, com as operações de carga e inserção, e C2, com as operações de leitura e atualização.

4. Resultados e análise do experimento

A Tabela 1 apresenta o resultado do *benchmark* entre os bancos de dados MongoDB e Tarantool, nas duas etapas do experimento.

Tabela 1. Resultado do *benchmark* entre os sistemas MongoDB e Tarantool.

	C1			C2		
	tempo de exec. ms	<i>throughput</i> op/s	des. relativo %	tempo de exec. ms	<i>throughput</i> op/s	des. relativo %
MongoDB	272.229,0	18,37	111,13	152.252,0	65,60	1,84
Tarantool	302.449,0	16,53	89,98	2.802,1	3.568,75	5440,17

Para C1 foram executadas 5 mil operações de inserção, para os 5 mil registros da *workload*, quando o MongoDB obteve um desempenho ligeiramente superior ao Tarantool, explicado pelo *throughput*. Como consequência, o tempo de execução do MongoDB foi pouco superior a 4,5 minutos e o do Tarantool pouco superior a 5 minutos.

A execução de C2 consistia em 10 mil operações divididas uniformemente entre operações de leitura e atualização. O resultado para C2 mostrou-se diferente, pois enquanto o tempo de execução do MongoDB foi pouco superior a 2,5 minutos, o do Tarantool, não chegou a 3 segundos, o que, também, pode ser explicado pelo *throughput*.

O desempenho relativo foi tomado com base no *throughput* dos cenários do experimento e, pela análise de C1, MongoDB mostrou-se mais eficiente e cerca de 11% mais rápido do que Tarantool, que, por sua vez, não chegou a 90% do resultado obtido pelo MongoDB. Todavia, mediante à C2, o resultado se inverte e o MongoDB tem um

desempenho inferior a 2%, se comparado ao desempenho do Tarantool que, por sua vez, apresentou um desempenho marcante, sendo mais de 5400 vezes mais rápido do que o MongoDB.

5. Conclusão

O tratamento adequado de arquivos de genótipos por uma solução de banco de dados é de grande importância para projetos em bionformática e genômica e, notadamente, os recursos clássicos supridos por SGBDRs não cumprem esta tarefa.

A análise dos resultados permite concluir que, para armazenamento, leitura e atualização de arquivos de genótipos, o uso de um banco de dados NoSQL, como o Tarantool, deve ser considerado, se comparado a sistemas com modelo de dados baseados em documentos, como é o caso do MongoDB.

A partir desta conclusão, devem ser apreciadas em investigações futuras (i) a hipótese de que bancos de dados NoSQL, com modelos de dados baseados em chave-valor, apresentam desempenho superior em relação a bancos de dados NoSQL com outros modelos de dados, para arquivos de genótipo; e (ii) a complementação da abordagem utilizada nos experimentos deste trabalho com cenários que avaliem escalabilidade em bancos de dados NoSQL.

Referências

- Aniceto, R., Xavier, R., Guimarães, V., Hondo, F., Holanda, M., Walter, M. E., and Lifschitz, S. (2015). Evaluating the cassandra NoSQL database approach for genomic data persistency. *International Journal of Genomics*, 2015.
- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 2(234):177–186.
- Caetano, A. R. (2009). Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. *Rev. Bras. de Zootecnia*, 38:64–71.
- Cooper, B. F., Silberstein, A., Tam, E., Ramakrishnan, R., and Sears, R. (2010). Benchmarking cloud serving systems with ycsb. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, pages 143–154, New York, NY, USA. ACM.
- Edlich, S. (2016). NoSQL. <http://www.nosql-database.org/>.
- Hecht, R. and Jablonski, S. (2011). NoSQL evaluation: A use case oriented survey. In *2011 International Conference on Cloud and Service Computing*, pages 336–341. IEEE.
- Li, Y. and Manoharan, S. (2013). A performance comparison of sql and nosql databases. In *Communications, Computers and Signal Processing (PACRIM), 2013 IEEE Pacific Rim Conference on*, pages 15–19.
- Mail.Ru Group (2016). Tarantool. <http://tarantool.org/>.
- MongoDB, Inc. (2016). MongoDB for giant ideas. <https://www.mongodb.org/>.
- Veronika Abramova; Jorge Bernardino and Pedro Furtado (2014). Experimental Evaluation of Nosql Databases. *International Journal of Database Management Systems (IJDMs)*, 6(3):1–16.

Data Release Interface para o Dark Energy Survey

Luiz A. Nicolaci da Costa^{1,2}, Angelo Fausti Neto^{2,3}, Ricardo L. C. Ogando^{1,2}, Marcio A. G. Maia^{1,2}, Riccardo Campisano^{2,4}

¹Observatório Nacional - MCTI

²Laboratório Interinstitucional de e-Astronomia – LIneA - MCTI

³LSST Corporation - USA

⁴Centro Federal de Educação Tecnológica Celso Suckow da Fonseca – CEFET/RJ
{ldacosta,angelofausti,ogando,maia,riccardo.campisano}@linea.gov.br

Abstract. *Dark Energy Survey, after a proprietary period, releases to community processed data, such as images, catalogs, and maps. LIneA scientific portal allows that these different products, which are input to scientific analysis, are integrated and distributed to the collaboration and public. We describe a query builder to efficiently join these products.*

Resumo. *O Dark Energy Survey, após um período proprietário, disponibiliza para a comunidade seus dados processados na forma de imagens, catálogos, e mapas. O portal científico do LIneA permite que esses diferentes produtos necessários para a análise científica sejam integrados e distribuídos para a colaboração e o público. Descrevemos um query builder para juntar eficientemente esses produtos.*

1. Introdução

O *Sloan Digital Sky Survey* (SDSS) há mais de 15 anos mapeia o Universo estudando de exoplanetas a quasares distantes. Seu sucesso se traduz na publicação de quase 7 mil artigos com a marca SDSS que já foram citados quase 300 mil vezes. Uma das razões do sucesso desse projeto é a disponibilização de novos dados em *Data Releases* (DR) públicos. Imagens, espectros, e catálogos de estrelas e galáxias são disponibilizados através de um site (*SkyServer*) onde astrônomos podem fazer consultas e baixar dados. Desde o DR8, em 2010, o LIneA mantém um espelho desse site¹ que atende a comunidade astronômica brasileira, economizando tempo ao transferir dados e evitando filas de submissão no sítio americano. Além do mais, sob demanda, os pesquisadores brasileiros podem fazer requisições especiais para o uso do site, como por exemplo aumento de quota de espaço em disco, permitindo grandes consultas ao banco de dados do SDSS no Brasil.

O *Dark Energy Survey* (DES) teve sua primeira luz em Setembro de 2012, quando começou o período de Verificação Científica (*Science Verification*, SV), e em Setembro de 2013 começou oficialmente a observação de 1/8 do hemisfério Sul, onde o DES tem condições de repetir o sucesso que o SDSS obteve no Norte. Por exemplo, o DES já revelou a existência de várias vizinhas da Via-Láctea (Drlica-Wagner et al. 2015) e procurou pela contrapartida óptica da primeira fonte de ondas gravitacionais encontrada pelo LIGO (Abbott et al. 2016). Portanto, um *Data Release Interface* (DRI) é uma ferramenta fundamental para o DES e para a comunidade. A seguir discutimos o *Data Server*, protótipo do DRI, já implementado na instância do Portal Científico do DES no Fermilab.

1 <http://www.linea.gov.br/020-data-center/aceso-a-dados-2/>

2. Data Server

O Data Server é um complexo sistema entregue para a colaboração DES em Abril de 2014. Este sistema é composto por um conjunto de ferramentas disponíveis no LIneA e no Fermilab que permitem através de PostgreSQL, TurboGears, e ExtJS:

- Acompanhar o progresso do levantamento DES.
- Verificar visualmente a qualidade das imagens e realizar marcações.
- Sobrepor catálogos às imagens do levantamento.
- Realizar buscas nos catálogos de objetos produzidos a partir destas imagens.
- Criar ou carregar listas de objetos de interesse.
- Visualizar estes objetos e criar “recortes de imagens”

A ferramenta *Tile Viewer* usada na verificação dos dados do DES, é uma das muitas disponíveis no sistema. Nela é possível não só ver imagens, mas associá-las a objetos detectados e, as máscaras aplicadas, como aquelas que indicam estrelas brilhantes. Além disso, associada ao *Upload* de alvos, permite que Grupos de Trabalho verifiquem visualmente a seleção de alvos para observações de seguimento ou aferimento de algoritmos de detecção, sendo intensamente usada, por exemplo, pelo Grupo de Lentes Fortes, cuja detecção de lentes requer confirmação visual (Figura 1).

3. Data Release Interface

Um compromisso assumido com o DES é o desenvolvimento da *Data Release Interface* (DRI) para seu DR público em 2017. Esta interface é uma evolução do *Data Server* discutido acima. Esse processo também é fundamental como preparação para o *Large Synoptic Survey Telescope* (LSST, Abell et al. 2009) cujo volume de dados será ordens de magnitude maior que o do DES. O projeto da DRI, além da interface propriamente dita, envolve a criação de catálogos científicos através da composição de vários produtos, como os de catálogos co-adicionados, de algoritmos de classificação e medição de propriedades, e mapas e máscaras (ex.: estrelas saturadas), mantendo a proveniência dos mesmos (Figura 2). Atualmente, por conta do volume de dados, os catálogos co-adicionados originais, com cerca de 900 parâmetros e centenas de milhões de linhas, são mantidos no Hadoop, pois tentativas com Pgpool encontravam um gargalo ao processar um grande número de regiões do céu em paralelo. Em um *cluster Rocks* com 1000 núcleos, apenas o Hadoop permitiu o processamento de mil regiões do céu em paralelo. Produtos derivados como classificação estrela-galáxia (morfológicos) e *redshifts* fotométricos (ajuste de templates e *machine learning*), menos volumosos, por conterem menos colunas, são ingeridos no banco de dados de catálogos ao longo do processamento usando a ferramenta GAVO DaCHS (Demleitner et al. 2014). Essa ferramenta obedece aos protocolos do Observatório Virtual, melhorando a acessibilidade dos dados no que tange a sua descrição. Finalmente, todas as tabelas são combinadas usando um *pipeline* (*Value Added Catalog*, VAC) cuja componente *query builder* dá ao usuário a capacidade de configurar o catálogo final de modo a se adequar ao seu objetivo científico, seja através de cortes em parâmetros, seleção de regiões do céu, escolha de algoritmos e parâmetros relevantes. Executar todos os passos de processamento, gerando as diferentes máscaras e parâmetros, em um *cluster* com 1.000 núcleos pode levar dezenas de horas, como esperado para esse volume. Mas uma vez prontos, produzir um catálogo para uma região de cerca de 3.000 graus quadrados e cerca de 100 milhões de objetos a partir desses produtos auxiliares leva menos de duas horas. Um dos segredos está em aplicar as máscaras aos catálogos, removendo elementos que não serão utilizados. Finalmente, o *join* apenas entre as tabelas de interesse do cientista provê resultados com eficiência. Esses produtos estão sendo exportados para o banco Oracle do NCSA para a colaboração.

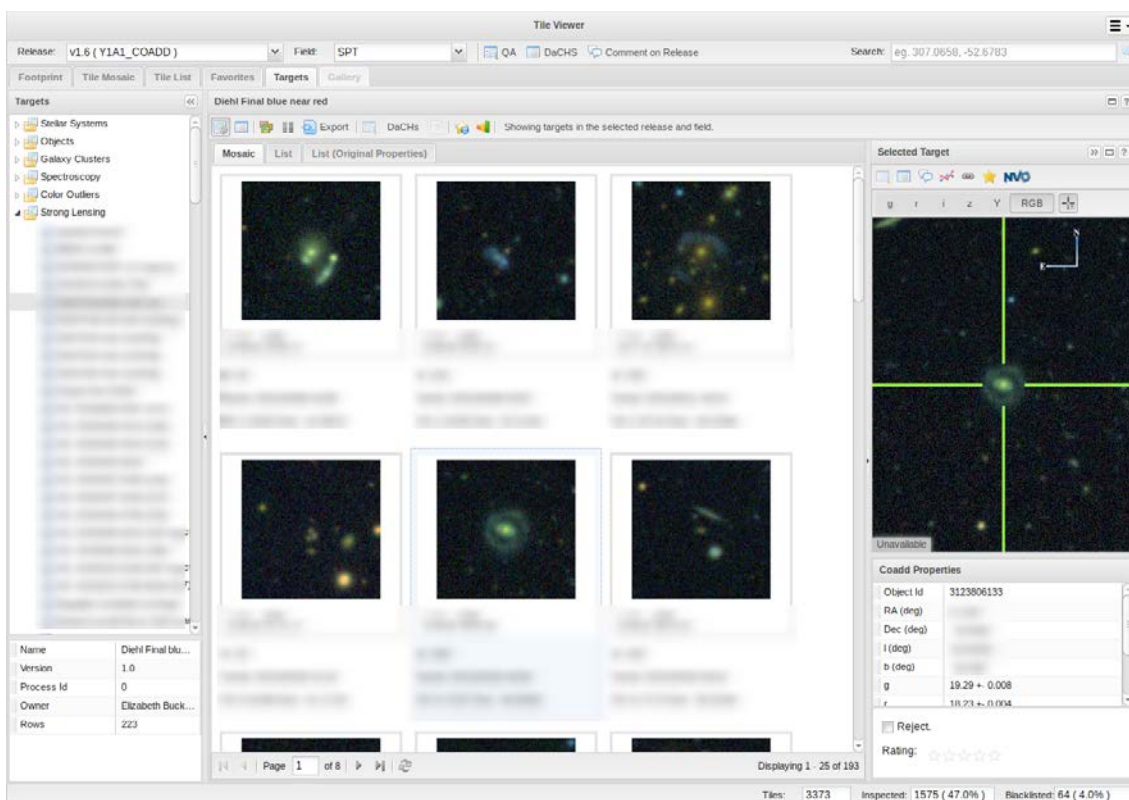


Figura 1. Captura de tela do Tile Viewer no Fermilab mostrando lentes gravitacionais a partir de um catálogo de usuários ingerido no sistema. Os dados das fontes foram encobertos para evitar vazamento.

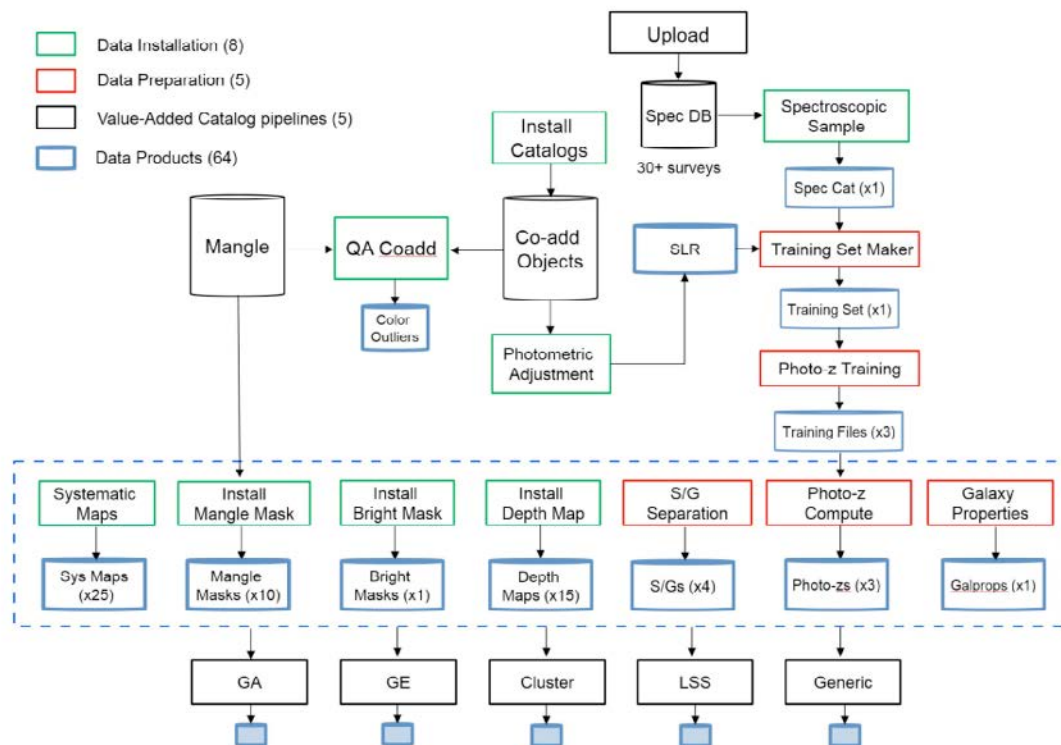


Figura 2. Elementos que compõem os catálogos científicos do DES. Iniciamos com *Install Catalogs* e *Upload* de catálogos espectroscópicos públicos. Os mapas (verde) e os produtos (vermelho) são ingeridos no banco e sua junção produz o VAC.

4. Perspectivas

O sucesso do *Data Server* com a colaboração foi grande e mostrou ao mesmo tempo que ao se disponibilizar um produto complexo como este, é fundamental se preparar para atender a demanda de novos casos de uso. Nos próximos meses, o portal deve continuar seu *refactory* agressivo, de modo a permitir atualizações mais ágeis no OpenStack do NCSA onde o DRI vai operar. Utilizando novas tecnologias como Python 3, Django, e Alladin. Essa interface vai servir inicialmente ao DES e posteriormente a comunidade em geral. Além da importância intrínseca de servir o DES, esse trabalho também serve como ponte para o centro de dados do LSST, o qual vai exigir um esforço ainda maior por parte do LIneA para implementar.

Referências

- Abell et al. (2009) “LSST Science Book, Version 2.0” ArXiv 0912.0201L
- Abbott et al. (2016) “Localization and broadband follow-up of the gravitational-wave transient GW150914” arXiv 160208492
- Alam et al. (2015) “The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III” ApJS 219 12
- da Costa et al. (2013) “End-to-end scientific processing in the LIneA Science Portal” CSBC 2013 - BreSci - VII Brazilian e-Science workshop
- Demleitner et al. (2014) “Virtual observatory publishing with DaCHS” A&C 7 27
- Drlica-Wagner et al. (2015) “Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey” ApJ 813 109
- Eisenstein et al. (2015) “The Dark Energy Spectroscopic Instrument (DESI): Science from the DESI Survey” AAS 225 336.05
- Fausti Neto et al. (2013) “Quick Reduce: Dark Energy Survey Camera mountain-top Quality Assessment tool and its master calibration pipeline” CSBC 2013 - BreSci - VII Brazilian e-Science workshop
- Flaugher, B. et al. (2015) “The Dark Energy Camera” AJ 150 150

Cloudscape Brazil

Apresentação

EUBrasilCloudFORUM Project (<http://eubrasilcloudforum.eu>) organizes the 3rd edition of Cloudscape Brazil with the ambition to shape cloud computing developments in both Brazil and Europe by bringing insights into current market trends, challenges slowing down mainstream adoption of cloud services, and best practices to address them. Further topics include: policy issues around taxation, bandwidth, energy pricing, security and data protection, which are slowing down cloud adoption in Brazil.

As a follow-up of the technical presentation of novel solution towards a more reliable and resilient cloud services and applications at the WCN (July 6th), Cloudscape Brazil 2016 will be a forum for discussing cloud-friendly policy measures to ensure the supply side meets the demand and needs of cloud services' consumers of more trustworthy and reliable cloud computing services. The aim of the event is to introduce a dialogue demonstrating how SMEs and end users can benefit and make the most out of adopting cloud computing.

Cloudscape Brazil takes a practical look at some of the major challenges & requirements to the adoption of cloud computing, exploring pathways for closer collaboration between Brazil and Europe on issues of common interest.

Priscila Solis Barreto (Universidade de Brasília)

Marco Vieira (Universidade de Coimbra)

Sara Pittonet (IT Trust Services)

Comitê de Organização

Coordenação Geral

Priscila Solis Barreto (Universidade de Brasília)

Marco Vieira (Universidade de Coimbra)

Sara Pittonet (IT Trust Services)

Coordenação Local

Tiago Ferreto (PUC/RS)

Painéis / Palestras

EU-Brazil Cloud Research excellence

Priscila Solis, University of Brasilia and EUBrasilCloudFORUM Brazilian coordinator

Wagner Meira, Federal University of Minas Gerais

Leandro Ciuffo, RNP

Philippe O.A. Navaux, Professor at the Universidade Federal of Rio Grande do Sul and Computer Science Area Director of CAPES, Brazil

Cristiano Both, UFRGS

Andrey Brito, UFCG

Presentation session. Funded by the EC and MCTI, EUBrasilCloudFORUM supports EU-BR collaborative projects in the collection and promotion of their results and activities. The results will be used to draft a research Roadmap on cloud computing, identifying collaboration needs and opportunities between Europe and Brazil for the European Commission and to MCTI, thus contributing to the definition of future cooperation priorities between the two regions. This session will provide an overview of the priorities and outputs of each jointly funded project with a live demo from the FIBRE project and insight into issues such as cloud security personal data protection, trans-Atlantic flows of data, portability of data, and appropriate remedies to vendor lock-in.

A common EU-BR technological & legal framework for managing security and privacy aspects on the cloud

Fabio Martinelli, CNR, Italy

Heber Fialho Maia Junior, General Coordination of Management and IT Governance, Secretariat of Logistic and IT, Dept. of Governance and Information Systems, Ministry of Planning, Brazil

Felipe Matos, Chief Start-up Farmer and Advisory Board Member of the Brazilian Association of Start-ups

Andrey Brito, UFCG and SecureCloud project

Europe and Brazil have different might have legal framework for privacy and data management, thus making solutions addressing security and privacy not always aligned. Current needs to process large amounts of data, potentially containing sensitive information, may require rethinking big data programming models for building scalable and secure applications. Application developers should carefully consider these issues while implementing services that handle personal data; policies should be defined and enforced at all stages during service provision. The ultimate goal is to address the concerns of the end-users and made them aware of the risks when data is not handled in a secure and privacy-friendly way. This session will address the challenges in privacy and data protection, providing a close look at legal implication for cloud services mainly in Brazil, and will contribute to identify areas for collaboration and alignment between the two regions, supported by relevant stakeholders including research, industry and policy makers.

Cloud federation & Open Science Cloud at cross-regional level

Ignacio Blanquer, Professor of the Polytechnic University of Valencia, representing the EUBRA-BIGSEA project

José Luiz Ribeiro Filho, Director of Services and Solutions of the Brazilian National Education and Research Network (RNP), Brazil

Marco Carvalho, SNIA Brasil

Free flow of data and portability of both data and applications are among the main challenges to address for cloud computing to help drive the innovation by enabling the free movement of data for reasons other than the protection of personal data and by avoiding vendor lock-in. Federation of cloud resources is the way forward to address these data flows and portability issues. The actual need is to have computing resources aside data stores to reduce the latency and optimise the usage of the network infrastructure, and to replicate and synchronize data to make services and applications more resilient. Cloud marketplaces are the business equivalent of a cloud federation by offering customers the access to a wide variety of commercial services within the same platform. Capitalizing and leveraging upon the Open Science Cloud within the EU's DSM strategy is a priority policy action to define a public, open cloud infrastructure that all scientific researchers can use in an integrated way, removing silos. This session will shed lights on the federation of cloud resources across the two regions and give a better understanding of how the European Open Science model could be adopted in Brazil, in order to establish a public cloud infrastructure for the Brazilian science community

Cutting edge cloud technologies: 5G, Cloud and IoT, Fog computing

Andrea Bondavalli, Professor of Computer Science at the University of Firenze, Italy

Karin Breitman, General Manager of the Brazil R&D Center, EMC, Brazil

Carsten Oliver Schirra, Senior Manager Strategy & Business Development at Philips Research Brazil

Afonso Lamounier Jr. Vice President Government Relations, SAP Latin America and Caribbean, SAP

Flavio Lenz Cesar, Special Minister Counselor, Ministry of Communications, Brazil

Sergio Takeo Kofuji, Assistant Professor at the University of Sao Paulo, Coordinator to FI WARE LAB in University of São Paulo, Brazil

The EC commission in a recent communication (April 19th) has identified 5G and Internet of Things amongst the ICT standardisation priorities for the Digital Single Market. This session will discuss the emergence of the mobile edge computing paradigm to reduce the latency for processing near the source large quantity of data and the need of the emerging 5G technology to satisfy the requirements of different verticals. Mobile Edge Clouds have the potential to provide an enormous amount of resources, but it raises several research challenges related to the resilience, security, , data portability and usage due to the presence of multiple trusted domains, as well as energy consumption of battery powered devices.

Large and centralized clouds have been deployed and have shown how this paradigm can greatly improve performance and flexibility while reducing costs. However, there are many issues requiring solutions that are user and context aware, dynamic, and with the capability to handle heterogeneous demands and systems. This is a challenge triggered by the Internet of Things (IoT) scenario, which strongly requires cloud-based solutions that can be dynamically located and managed, on demand and with self-organization capabilities to serve the purposes of different verticals.

Towards a EU-Brazil Open Science Cloud environment & future directions for 2017

Priscila Solís Mendez, professor of the Department of Computer Science at the Universidade de Brasília

Marco Vieira, University of Coimbra, Portugal

Philippe O.A. Navaux, Professor at the Universidade Federal of Rio Grande do Sul and Computer Science Area Director of CAPES, Brazil

Fabio Martinelli, CNR, Italy

Ignacio Blanquer, Professor of the Polytechnic University of Valencia, representing the BIGSEA project

Andrea Bondavalli, Professor of Computer Science at the University of Firenze, Italy

Panel discussion & wrap-up with concrete calls to action. All the chairs of previous sessions will join the panel and report the main topics raised to be collected for the EUBrasilCloudFORUM Roadmap.

18º COMPUTEC – Seminário de Computação e Mercado

Apresentação

O COMPUTEC aborda a área de Computação desde uma perspectiva da tecnologia e da inovação, defendendo o incremento da cooperação e da interação entre as universidades, as empresas, a sociedade e os diferentes níveis e esferas de governo. Esta reflexão proposta pelo COMPUTEC está inserida em um contexto nacional de forte prioridade das agências de fomento governamentais para a área de inovação e transferência de tecnologia, estimulando a ampliação da cooperação entre as universidades, as empresas, a sociedade e o governo. Neste sentido, o COMPUTEC está organizado de forma a identificar e analisar grandes e importantes desafios e oportunidades em inovação e desenvolvimento tecnológico, principalmente na área de Computação, no contexto da Hélice Tripla. Ao propiciar esta reflexão e discussão, com destacados representantes da academia, das empresas, da sociedade e do governo, pretende-se contribuir consideravelmente para o avanço da Computação no Brasil. No COMPUTEC as ideias são apresentadas e discutidas por meio de painéis e palestras que abordam os temas propostos em diversos aspectos tais como a interação universidade-empresa, inovação tecnológica, transferência autossustentada para a sociedade, pesquisa aplicada, oportunidades e desafios e programas governamentais na área de inovação. Logo, os pronunciamentos de representantes dos setores público e privado, abordando políticas públicas, estratégias de desenvolvimento, visões de mercado e perspectivas futuras, deverão trazer à discussão temas extremamente relevantes àqueles que estão prospectando novas oportunidades e buscando preparar-se para um melhor entendimento dos desafios para os diversos atores públicos e privados envolvidos, nas áreas acadêmicas (docentes, pesquisadores e estudantes), empresarias e governamentais.

Rafael Prikladnicki (PUCRS) – coordenador geral do Computec

Ricardo Melo Bastos (PUCRS) – coordenação local do Computec

Palestras

Bloco 1: Empreendedorismo, inovação, e criatividade e interdisciplinaridade

Criatividade e interdisciplinaridade na computação – caso 1

Palestrante: Cirano Silveira - diretor de P&D da HP Inc

Criatividade e interdisciplinaridade na computação – caso 2

Palestrante: Luis Saraiva - diretor de P&D da Hewlett Packard Enterprise

Inovação e interdisciplinaridade no setor público

Palestrante: Thiago Ribeiro - empreendedor com atuação no setor público

Empreendedorismo e interdisciplinaridade

Palestrante: Artur de Franseschi - fundador e empreendedor da Goga, empresa incubada na Raiar/Tecnopuc

Bloco 2: Casos de interdisciplinaridade

Caso 1 - Upy

Palestrante: Alexandre Schumacher - fundador da Upy, empresa instalada no Feevale Techpark

Caso 2 - DBServer

Palestrante: Eduardo Peres - fundador da DBServer, empresa instalada no Tecnopuc

Caso 3 - Aegro

Palestrante: Pedro Dusso - fundador da Aegro, empresa instalada no CEI/UFRGS

Bloco 3: Ambiente interdisciplinar

Ambiente de trabalho interdisciplinar na computação

Palestrante: Matheus Tait - Gerente Geral da ThoughtWorks Porto Alegre

Interdisciplinaridade e desenvolvimento de software

Palestrante: representante da SAP

Parques tecnológicos e computação

Palestrante: Márcio Machado - Presidente da Reginp - Rede Gaúcha de Parques Tecnológicos e Incubadoras de Empresas

CQ 2016 - XVII CURSO DE QUALIDADE

Apresentação

O Curso de Qualidade no Ensino da Computação (CQ 2016) é um evento realizado pela Comissão de Educação da SBC e tem como finalidade principal propiciar a discussão com a comunidade acadêmica, representada principalmente pelos professores e coordenadores de curso da área de Computação, temas relacionados a metodologias de ensino, projeto e avaliação de cursos, escopo de disciplinas e o processo de ensino e aprendizagem.

Nesta edição, o CQ 2016 tem como finalidade apresentar e debater com os envolvidos na educação – diretores, coordenadores, professores, pesquisadores e profissionais – aspectos de gestão acadêmica e estratégias para implementação de interdisciplinaridade nos currículos, além de cuidados para se implementar um curso EAD a partir de um curso presencial. Os assuntos serão abordados por meio de cursos, painel e oficina, além da apresentação de relatos de experiências submetidos e aprovados. Houve a inscrição de 14 trabalhos e submissão final de 12 relatos. Dos relatos submetidos 5 foram aprovados, o que gera uma taxa de aceitação de 41,66%.

Aos palestrantes, painelistas, autores e demais membros da comunidade que contribuíram para a realização deste evento, um agradecimento especial pelo reconhecimento e confiança no evento. Ao Comitê de Programa pelo excelente trabalho de avaliação de artigos, e pela disponibilidade dedicada, nossa gratidão.

A todos os participantes nossas boas vindas e desejos de que todos possam aproveitar ao máximo as atividades da programação desta edição do Curso de Qualidade.

Marcelo Hideki Yamaguti (PUCRS)

Ronaldo Celso Messias Correia (FCT/UNESP)

Comitê de Organização

Coordenação Geral

Ronaldo Celso Messias Correia (FCT/UNESP)

Coordenação Local

Marcelo Hideki Yamaguti (PUCRS)

Comitê de Programa

Ana Paula Bacelo (PUCRS)

Andreia Malucelli (PUCPR)

Celso Olivete Junior (FCT/UNESP)

Daltro Nunes (UFRGS)

Igor Steinmacher (UTFPR)

Jair Cavalcanti Leite (UFRN)

Marcelo Hideki Yamaguti (PUCRS)

Renata Mendes de Araujo (UNIRIO)

Rogério Eduardo Garcia (FCT/UNESP)

Ronaldo Celso Messias Correia (FCT/UNESP)

Simone Martins (UFF)

Cursos, Painel e Oficina

Curso I: Gerenciamento de Conflitos

Prof. Dr. André Hartmann Duhá (PUCRS)

Este curso tem como objetivo abordar aspectos práticos do processo de geração e administração de conflitos no contexto universitário. Por meio de estudos de caso e exercícios de simulação, busca promover a reflexão crítica sobre o tema e desenvolver as competências necessárias para a obtenção de resultados positivos na administração dos conflitos no ambiente acadêmico.

Curso II: Do Presencial para a EaD

Profa. Dra. Mára Lúcia Fernandes Carneiro (UFRGS)

Este curso tem como objetivo abordar os aspectos relacionados a como tornar um curso presencial em um curso EaD. Algumas questões serão discutidas no curso, tais como: Quais são os requisitos legais envolvidos? Quais são os principais desafios e dificuldades enfrentados na concepção e implementação? Um curso a distância de qualidade é realmente mais barato? É uma oportunidade de compreender os vários aspectos envolvidos na implementação de um curso a distância a partir das experiências que temos em cursos presenciais.

Painel: Interdisciplinaridade na Estrutura Curricular do Curso

Profa. Dra. Claudia Bauzer Medeiros (UNICAMP)

Prof. Dr. João Batista Siqueira Harres (PUCRS)

Prof. Dr. Marcelo Walter (UFRGS)

Este painel tem como objetivo promover um espaço para debater os aspectos relacionados à implementação da interdisciplinaridade nos currículos de cursos de ensino superior, contribuindo com soluções, estratégias e experiências que possam ser aplicadas pelos gestores acadêmicos e docentes em seus cursos.

Oficina: Interdisciplinaridade nos Cursos de Graduação

A oficina envolverá a discussão entre os participantes do CQ de formas de implementação de interdisciplinaridade nos cursos de graduação face às Diretrizes Curriculares Nacionais da área de Computação. Como resultado desta oficina será gerada uma proposta de formas de implementação de interdisciplinaridade nos cursos de ensino superior da área de Computação. Esta proposta será encaminhada à Diretoria de Educação da SBC como material adicional à discussão dos Currículos de Referência da SBC dos cursos da área de Computação.

Trabalhos aceitos

A Prática Interdisciplinar no Curso de Engenharia de Computação do Centro Universitário de Anápolis (UniEVANGÉLICA)

Viviane Carla Batista Pocivi (UniEvangélica)
Adrielle Beze Peixoto (UniEvangélica)
Rosana Machado de Souza (UniEvangélica)
Luciana Nishi (UniEvangélica)

AGES – Um Espaço Interdisciplinar, Baseado em Projetos, para Aprendizagem de Engenharia de Software

Marcelo Hideki Yamaguti (PUCRS)
Flávio Moreira de Oliveira (PUCRS)
Cássio André Wailer Trindade (PUCRS)

Aplicação e Condução do Projeto Aplicado Interdisciplinar no Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas

Roberto Marcos Kalili (USJT)
Ana Paula Gonçalves Serra (USJT)

Atividades Interdisciplinares em um Curso de Engenharia de Software: Desenvolvendo Softwares Tradicionais e Jogos

Andre Roberto Ortoncelli (UTFPR)
Marisângela Pacheco Brittes (UTFPR)
Rafael A. P. Oliveira (UTFPR)

Interdisciplinaridade no IF Farroupilha - Campus Santo Ângelo através da Prática Profissional Integrada

Karlise Soares do Nascimento (Instituto Federal Farroupilha - Campus Santo Ângelo)
Fábio Weber Albiero (Instituto Federal Farroupilha - Campus Santo Ângelo)
Andréa Pereira (Instituto Federal Farroupilha - Campus Santo Ângelo)
Joice Machado (Instituto Federal Farroupilha - Campus Santo Ângelo)

A prática interdisciplinar no Curso de Engenharia de Computação do Centro Universitário de Anápolis (UniEVANGÉLICA).

Viviane Carla Batista Pocivi¹, Adrielle Beze Peixoto¹, Rosana Machado de Souza¹,
Luciana Nishi¹

¹Curso de Bacharelado em Engenharia de Computação – Centro Universitário de Anápolis (UniEvangélica) – 75083-515 – Anápolis – GO – Brasil

[{viviane.batista, adrielle.peixoto,
rosana.souza, luciana.nishi}@unievangelica.edu.br](mailto:{viviane.batista, adrielle.peixoto, rosana.souza, luciana.nishi}@unievangelica.edu.br)

Abstract. *This report aims to present the interdisciplinary practice which has been developed at the Computing Engineering course (ECC) at the University Center in Anápolis – UniEVANGÉLICA. The interdisciplinary purpose is organized through the course curriculum and the students activities called Projeto Interdisciplinar (PI) – Interdisciplinary Project (IP), which have been developed from the second to the eight grades in the course. All of them involve the contribution and orientation of the teachers and the coordination of the course. The results of the PI are selected and presented at the Interdisciplinary Seminar of Technology and Society (SITES), which are in the XIV year. We present the model of activities which are develop through the interdisciplinary activities, the main results, the challenges, the learning experiences and the futures strategies to keep on the interdisciplinary at the ECC.*

Resumo. *Este relato de experiência objetiva apresentar a prática interdisciplinar vivenciada no curso de Engenharia de Computação (CEC) do Centro Universitário de Anápolis - UniEVANGÉLICA. A proposta interdisciplinar encontra-se organizada curricularmente por meio das atividades discentes denominadas Projeto Interdisciplinar (PI), desenvolvidas do segundo ao oitavo períodos do curso, sob orientação e coordenação de docentes. Os resultados dos PI são apresentados semestralmente no Seminário Interdisciplinar de Tecnologia e Sociedade (SITES), que se encontra na XIV edição. Serão apresentados o formato em que as atividades interdisciplinares são desenvolvidas, os principais resultados alcançados, os desafios vivenciados, as lições aprendidas e as estratégias futuras para condução da interdisciplinaridade no CEC.*

1. Introdução

Motivados pelo desafio da melhoria da formação dos acadêmicos dos cursos de Computação da UniEVANGÉLICA, desde o ano de 2006, se inicia a prática interdisciplinar nestes cursos. Em 2008, como um marco dessas ações, é realizado o Primeiro Seminário Interdisciplinar de Tecnologia e Sociedade (SITES). Com a

maturidade adquirida ao longo dos anos, a prática interdisciplinar foi formalizada nos Projetos Pedagógicos dos Cursos de Sistemas de Informação e Engenharia de Computação.

Atualmente, no Curso de Engenharia de Computação (CEC), os Projetos Interdisciplinares são atividades discentes formais do 2^a ao 8^o períodos. Existe uma coordenação de projetos interdisciplinares e em cada período há um professor orientador, que acompanha os grupos de trabalho. Em cada período, espera-se a formação de habilidades e competências pessoais, interpessoais e técnicas distintas. A temática dos projetos é selecionada pelos próprios alunos.

Para que haja o envolvimento do corpo docente, foi instituído no calendário acadêmico o Momento SITES, no qual em um dia da semana todos os professores do curso recebem os alunos de projetos interdisciplinares e orientam em conformidade com sua linha de formação e/ou vivência profissional. O Momento SITES ocorre de duas a três vezes no semestre letivo.

Um dos marcos de avaliação dos projetos é chamado de Qualificação. Neste momento, para cada período, uma banca é formada por três professores que avaliam o desenvolvimento do trabalho. A avaliação final é realizada no SITES, evento com duração de três a cinco dias, no qual os alunos apresentam para os demais colegas, docentes e convidados os resultados de seus projetos. A avaliação é realizada por professores e convidados e a nota obtida reflete em 30% da nota final de todas as disciplinas em que o aluno está matriculado.

Nos itens a seguir serão apresentados: a proposta pedagógica dos projetos interdisciplinares em cada período do CEC; os principais resultados alcançados por meio da prática interdisciplinar; os desafios encontrados; as lições aprendidas e as estratégias futuras.

2. Proposta pedagógica dos projetos interdisciplinares

Nesta seção serão apresentados os objetivos, metodologia, estratégias de avaliação e formato de cada projeto interdisciplinar.

1.1. Projeto Interdisciplinar I (PI I)

Desenvolvido no 2^o período do CEC, conhecimentos teóricos aprendidos no 1^o e 2^o períodos são colocados em prática por meio do desenvolvimento de projeto de software, por equipes de quatro a seis alunos. Estimula-se o desenvolvimento de um projeto real. Os artefatos são produzidos pelos alunos e avaliados na seguinte sequência: (a) proposta de sistema – 1^a avaliação; (b) documento textual com os artefatos de documento de visão, especificação de casos de uso, protótipos, fundamentação teórica e banner para a qualificação – 2^a avaliação; e (c) o banner corrigido e apresentado no SITES – 3^a avaliação. O grande desafio encontrado é o de se trabalhar em equipe, gerenciando e mantendo o foco do projeto.

1.2. Projeto Interdisciplinar II (PI II)

Desenvolvido no 3º período do CEC, o PI II é continuidade do PI I, acrescentando-se o projeto conceitual, lógico e físico do Banco de Dados. A metodologia de desenvolvimento segue o trabalho em equipe, orientação do docente de PI II e orientação no Momento SITES. A apresentação no SITES se dá no formato de exposição de Banners.

1.3. Projeto Interdisciplinar III (PI III)

Desenvolvido no 4º período do CEC, tem por objetivo produzir um *paper* que relata sobre os resultados e experiências dos projetos interdisciplinares dos períodos anteriores. Acrescenta-se, ainda, uma fundamentação teórica mais rigorosa. O formato de *paper* adotado é o da Sociedade Brasileira de Computação (SBC). Percebe-se que o grande desafio do PI III é de se aperfeiçoar a escrita em conformidade com o formato estabelecido.

1.4. Projeto Interdisciplinar IV (PI IV)

No 5º período do CEC, os discentes são expostos ao desafio de desenvolverem minicursos que contemplem a área de conhecimento de Engenharia de Computação. Os minicursos são ofertados no SITES e abertos à comunidade. O processo de avaliação segue os quesitos de 1ª, 2ª e 3ª avaliações, sendo necessariamente a 2ª avaliação a Qualificação e a 3ª avaliação no SITES. Aqui se encontram três relevantes desafios: o da escrita, o da produção de conteúdo e o da iniciação à atividade de docência. Desde 2015, os minicursos passaram a ser ofertados em formato EAD.

1.5. Projeto Interdisciplinar V (PI V)

No PI V, 6º período do CEC, as equipes de trabalho desenvolvem uma solução (produto de software ou tecnologia) para problemas da sociedade, preferencialmente da comunidade local. Estimula-se visão empreendedora. Devem ser bem definidos o escopo, planejamento e a forma de gerenciamento de cada projeto. O resultado final é exposto em um formato de feira, em um dos dias da programação do SITES. As equipes tem a liberdade para utilizar recursos e tecnologias inovadoras na apresentação das soluções. Representantes do mercado são convidados a participar das apresentações.

1.6. Projeto Interdisciplinar VI (PI VI)

No 7º período do CEC, o PI VI desenvolve o projeto Feira de Ideias que busca promover a visão interdisciplinar com base no empreendedorismo. Alunos submetem seus projetos que passam por várias etapas de validação. Inicia-se com os projetos inscritos, os quais são avaliados por uma banca de qualificação. Aqueles que possuem potencial inovador e empreendedor são selecionados, modelados e apresentados por alunos consultores. As melhores ideias selecionadas são avaliadas por docentes, representantes do mercado. A primeira ideia classificada é premiada no último dia do SITES. O SEBRAE Anápolis e a UniINCUBADORA são parceiras na realização deste projeto.

1.7. Projeto Interdisciplinar VII (PI VII)

Desenvolvido no 8º período do CEC, tem por objetivo amadurecer o processo em construção, contando com a experiência que os alunos adquiriram ao longo de seis Projetos Interdisciplinares. Os alunos tornam-se os responsáveis pela organização do SITES, inicialmente avaliando as edições anteriores, identificando potencialidades, fragilidades e, principalmente, criando propostas de melhoria. A segunda fase consiste na construção de estratégias viáveis para solução dos problemas identificados. Neste aspecto, a autonomia e pró-atividade do aluno torna-se imprescindível. As ações, projetos, metodologias, formalidades institucionais e quaisquer outras necessidades são de responsabilidade dos envolvidos.

3. Principais resultados

Os alunos relatam a superação de dificuldades iniciais: falar em público, autonomia, capacidade de resolução de problemas, trabalho em equipe, liderança, pro-atividade. Afirmam a contribuição do perfil adquirido para a atuação profissional e melhor percepção sobre o perfil do curso e das possibilidades de atuação. Por parte dos docentes, há maior integração e compreensão do perfil que se deseja formar no CEC.

4. Desafios

Os principais desafios vivenciados são: a) Manter o corpo docente, em grande parte horista, integrado aos objetivos e atuantes nos projetos interdisciplinares; b) Manter o evento exequível em três dias, considerando crescente número de trabalhos; c) Tornar mais eficiente o processo de replicação dos 30% da nota de avaliação final em todas as disciplinas em que o aluno está matriculado; d) Gerar publicações acerca dos projetos que estão sendo desenvolvidos.

5. Lições aprendidas

Como lições aprendidas, ressalta-se: a) Envolver os alunos, inclusive em algumas tomadas de decisão acerca dos projetos interdisciplinares, foi fundamental para a quebra de resistência e aceitação desta nova realidade de trabalho. b) Evitar o engessamento dos trabalhos, proporcionando um ambiente favorável ao desenvolvimento de projetos que nascem entre os alunos; c) Delegar as disciplinas de Projeto Interdisciplinar a professores que tenham perfil de trabalho em equipe e que, naturalmente, se envolvam nos desafios que surgem ao longo do processo; d) Avaliar continuamente e propor melhorias a cada semestre, integrando a visão de todos os envolvidos: gestores, docentes, alunos e parceiros.

6. Estratégias futuras

Como estratégias futuras, pretende-se: a) Melhorar a logística do SITES, tornando-o operacional diante do calendário acadêmico e mais estimulante aos participantes; b) Revisar os objetivos de cada período, orientando-os às novas realidades; c) Orientar resultados para o aumento de publicações; d) Valer-se das metodologias interdisciplinares para dirimir deficiências em conteúdos de formação básica; e) Ampliar as propostas existentes para projetos de cunho multidisciplinar.

AGES – um espaço interdisciplinar, baseado em projetos, para aprendizagem de Engenharia de Software

Marcelo H. Yamaguti¹, Flávio M. de Oliveira¹, Cássio A. W. Trindade¹

¹Faculdade de Informática – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

Av. Ipiranga, 6681 – Prédio 32 – Sala 505 – 90619-900 Porto Alegre – RS

{marcelo.yamaguti,flavio.oliveira,cassio.trindade}@pucrs.br

Abstract. The Software Engineering education faces the challenge of qualifying professionals with competence to work in an interdisciplinary way and in teams, with flexibility to perform different roles and capable of adapting to change. This article describes AGES (Agência Experimental de Engenharia de Software), a space that allows students to apply their knowledge in an integrated and diversified way. They participate in teams engaged in real projects. We present the assumptions, the implementation process of AGES, its status and lessons learned until now.

Resumo. O ensino da Engenharia de Software enfrenta o desafio de formar profissionais com competência para atuar de forma interdisciplinar e em equipe, sendo flexíveis para exercer diferentes papéis e adaptar-se a mudanças. Este artigo descreve a experiência de implantação da AGES (Agência Experimental de Engenharia de Software), um espaço que permite ao estudante aplicar seus conhecimentos em uma experiência integradora e diversificada, participando de equipes engajadas em projetos reais. São apresentados os pressupostos teóricos, o processo de implantação da AGES, o estado atual e as lições aprendidas até o momento.

1. Introdução

O ensino da Engenharia de Software enfrenta, no Brasil e no mundo, o desafio de atender a uma demanda crescente de profissionais capazes de produzir software robusto e com qualidade de maneira sistemática e eficiente. As características da área demandam, cada vez mais, desenvoltura para atuar de forma interdisciplinar e em equipe, sendo flexível para exercer diferentes papéis e adaptar-se a mudanças tecnológicas com rapidez.

Neste contexto, a AGES - Agência Experimental de Engenharia de Software foi criada para ser um ambiente com foco na concretização da busca da integração de conhecimentos. A AGES possui suas atividades fortemente centradas na realização de projetos de desenvolvimento de software, resolvendo problemas reais e em equipes também baseadas no cotidiano profissional, oportunizando o desenvolvimento integrado de competências e favorecendo uma transposição didática a partir do que é iniciado em na sala de aula.

Este artigo descreve a experiência de implantação de Agência Experimental de Engenharia de Software, como parte integrante do currículo do novo curso de Bacharelado em Engenharia de Software da PUCRS.

2. A AGES

A Agência Experimental de Engenharia de Software foi concebida para ser um ambiente prático de aprendizagem com as seguintes finalidades: propiciar aos estudantes o exercício de situações reais de projetos em um ambiente com foco na aprendizagem; viabilizar a integração interdisciplinar dos conteúdos do curso; integrar ensino, pesquisa e extensão; e permitir contato com organizações e empresas sempre que este contato agregar elementos para a formação do perfil do egresso.

Além disso, a AGES segue as seguintes diretrizes acadêmicas: desenvolvimento de competências do curso de Engenharia de Software, particularmente aquelas voltadas ao trabalho em equipe, respeitando as individualidades; forte vínculo com os conteúdos das disciplinas e professores do curso; autonomia em relação às turmas em andamento; realização de atividades com previsibilidade, mas sem urgências de prazo; foco na aprendizagem.

A Agência, além da realização de projetos de desenvolvimento de software, também prevê atividades de formação, tais como: participação dos alunos em palestras; capacitação em tecnologias (conhecimentos) necessárias para o desenvolvimento dos projetos; preparação para apresentação de palestras ou apoio a capacitação; prática em língua estrangeira, mais especificamente em língua inglesa.

Todos os alunos do curso de Bacharelado em Engenharia de Software participam das atividades da Agência, podendo desempenhar diferentes atividades ao longo do curso. Mais especificamente, o aluno deve realizar 4 módulos previstos no projeto do curso de ES. Cada módulo envolve 120 horas de atividades.

Na integração da AGES com o currículo do curso de Engenharia de Software, o estudante, à medida em que obtém novas habilidades pelas disciplinas da matriz curricular, desenvolve as competências correspondentes a cada Módulo da AGES. O Módulo I busca aprofundar as competências relacionadas a um Desenvolvedor de Software. O Módulo II engloba as competências de um Analista de Sistemas. No Módulo III o estudante executa as competências de um Arquiteto de Software; e finalmente no Módulo IV o estudante pode atuar como Gerente de Projeto ou aprofundar outra competência de um Engenheiro de Software.

Como se pode perceber, cada módulo está estruturado de forma que, sistematicamente, competências de diferente complexidade sejam desenvolvidas fomentando a formação do futuro profissional.

No início de cada módulo os alunos, sob a supervisão do professor orientador, escolhem os projetos de maior afinidade. Neste processo são desafiados na escolha do projeto e iniciam os trabalhos precisando compor com colegas de diferentes perfis (valores, estilos, modalidades de aprendizagem e experiências de vida) uma equipe.

Os alunos devem estar atentos aos conteúdos trabalhados nas disciplinas do curso e utilizá-los no contexto do projeto. Além de professores orientadores, a AGES conta com uma equipe técnica de profissionais de TI que auxilia nas atividades de especificação de software e na condução das tarefas técnicas dos alunos nos projetos. Isto se constitui em um aspecto diferencial para a simulação de situações reais em projetos de desenvolvimento de software.

Para relatar a sua atuação e o desenvolvimento de suas competências, cada aluno consolida o conjunto dos seus relatórios individuais de cada Módulo em um **portfólio individual**. O portfólio atesta as diferentes competências adquiridas pelo aluno ao longo de sua atuação na Agência. A composição do portfólio do aluno também pode agregar um conjunto de certificados (certificação) que o aluno adquira fora da Agência.

Após realização de todos os Módulos na Agência, o aluno deve escrever e apresentar a uma banca de avaliação, com supervisão de um professor orientador, um **memorial descritivo** do seu portfólio realizado na Agência, com a devida fundamentação teórica baseada em seus estudos nas disciplinas do Curso, e suas conclusões, contribuições e lições aprendidas a partir dessas vivências. Esta apresentação tem como objetivo a avaliação do conjunto de atividades e conhecimentos do aluno, considerando as competências e habilidades esperados de um egresso do curso de Engenharia de Software.

3. Implantação

A implantação física atual da AGES envolveu a criação de ambientes (figuras 1 e 2) que permitissem dinâmicas e formas de organização diversas, tanto no desenvolvimento de projetos, quanto de outras atividades de capacitação e qualificação dos estudantes.



Figura 1: ambiente superior.



Figura 2: ambiente térreo.

Na AGES há notebooks disponíveis e rede sem fio nos ambientes, o que permite maior mobilidade dos estudantes durante as atividades. Esta infraestrutura foi disponibilizada no final de julho de 2015. No 2º semestre daquele ano os alunos do Curso de Engenharia de Software participaram do primeiro módulo de atividades na AGES.

Inicialmente foram constituídas as equipes (times) para a execução dos projetos. Cada time ficou encarregado de um projeto que consistia em desenvolver uma aplicação Web. Antes de iniciar o desenvolvimento do software, ambos os times receberam um treinamento básico sobre tecnologias Web.

Durante o semestre, os times engajaram-se no desenvolvimento dos projetos, com apoio da equipe técnica da AGES e dos orientadores. A fim de permitir maior diversidade da experiência, foram realizadas sessões de retrospectivas. As retrospectivas utilizaram uma dinâmica na qual os alunos foram estimulados a relatar pontos positivos, pontos negativos, sugestões, e outros pontos. Assim, as sessões serviram para compartilhamento de vivências técnicas e também como *feedback* para a própria implantação da AGES.

Ao final do semestre houve a entrega da versão inicial de software por cada projeto.

4. Lições aprendidas e próximos passos

O cronograma de implantação da AGES, iniciado em 2015/2, acompanha a implantação do curso de Engenharia de Software em si, que se estende até 2018. Mesmo assim, já é possível fazer considerações iniciais da experiência, a partir do que foi observado no funcionamento dos times e dos relatos colhidos nas retrospectivas. Abaixo são sumarizados algumas das lições aprendidas neste semestre:

1. **Maturidade dos times:** ao escolher a metodologia a ser utilizada, ficou claro que metodologias ágeis são mais sensíveis à maturidade do time do que abordagens baseadas no modelo em cascata, com gestão do tipo comando-controle. Como os alunos eram muito iniciantes, o time que empregou um processo ágil, com ênfase na autogestão, teve mais dificuldade para gerar resultados.
2. **Importância das retrospectivas:** foi unânime o reconhecimento dos benefícios do compartilhamento das vivências para a aprendizagem de todos os participantes.
3. **Busca por aplicar competências já desenvolvidas:** os alunos relataram que havia grande expectativa em aplicar rapidamente competências de programação que eles já traziam do semestre anterior, o que demorou mais a acontecer, devido ao treinamento em tecnologias Web. Isso teve impacto no aproveitamento da vivência na AGES.
4. **Competências individuais X trabalho em equipe:** muitos alunos eram capazes de relacionar conhecimentos técnicos e encontrarem soluções por ensaio e erro. Contudo, a maioria não tinha o costume de trabalhar efetivamente em equipe (o que é diferente de dividir as tarefas e reunir tudo no final, como ocorre muitas vezes em trabalhos escolares), a valorizar o saber do outro e achar soluções coletivas. Este é um dos maiores ganhos obtidos com a vivência na AGES.
5. **Interdisciplinaridade:** a oportunidade de combinar conhecimentos vistos em diferentes disciplinas em uma situação de mundo real, já no 2º semestre do curso, realimentou a aprendizagem de forma extremamente positiva.
6. **Importância das lideranças:** a proatividade dos líderes de cada equipe foi decisiva no sucesso dos projetos.

Neste semestre (2016/1), a segunda turma da AGES iniciou suas atividades, na qual já foram incorporadas algumas das lições apresentadas. Atualmente há 4 projetos iniciados, sendo que alguns projetos já estão com a modelagem realizada e os alunos se engajaram em desenvolvimento, enquanto outros ainda estão na fase de especificação, com os alunos participando de reuniões com *stakeholders* e tarefas de mais alto nível. Já é possível constatar que o ambiente está mais rico, tanto em produtividade como em compartilhamento da aprendizagem, aumentando a interdisciplinaridade e a diversidade da experiência dos estudantes na AGES.

Na continuidade do projeto de implantação do curso de Engenharia de Software a AGES tem como previsão a sua expansão física para acomodar até 240 estudantes simultaneamente em projetos diversos. A implantação do Módulo II ocorrerá em 2016/2, do módulo III em 2017/2 e finalmente do módulo III em 2018/1, quando se prevê que em cada time de projeto haja estudantes do 2º, 4º, 6º e 7º níveis do curso trabalhando em equipe, interagindo e trocando experiências.

Aplicação e Condução do Projeto Aplicado Interdisciplinar no Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas

Roberto Marcos Kalili, Ana Paula Gonçalves Serra

Faculdade de Tecnologia e Ciências Exatas – Universidade São Judas Tadeu (USJT)
São Paulo – SP – Brazil

{prof.kalili, prof.anapaula}@usjt.br

***Resumo.** O presente trabalho visa expor as etapas de implantação de projetos aplicados interdisciplinares implantados nos dois semestres de 2015 para o curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas da Universidade São Judas, cujo objetivo fundamental é consolidar todas as competências desenvolvidas em todas as disciplinas de um mesmo módulo. A proposta de um projeto aplicado envolve uma solução prática, de cunho real, em que o aluno deve, para colocá-la em prática, elaborar uma pesquisa detalhada como forma de complementação aos assuntos estudados no módulo, implementar o projeto e realizar uma apresentação formal.*

1. Introdução

Em 2015 o curso Superior em Análise e Desenvolvimento de Sistemas foi implantado na Universidade São Judas Tadeu. A matriz curricular semestral e modular contempla o Projeto Aplicado que é um projeto interdisciplinar realizado na disciplina de Projeto Aplicado em cada semestre, considerada como uma disciplina integradora, o seu objetivo fundamental é consolidar todas as competências desenvolvidas em todas as disciplinas do mesmo módulo. Esse é o motivo pelo qual a avaliação de todas as demais disciplinas fica, parcialmente, vinculada ao desempenho acadêmico do aluno nessa disciplina integradora.

Do ponto de vista mais específico da formação, o Projeto Aplicado é concebido, no contexto formativo dos Projetos Pedagógicos dos Cursos, como conjunto de atividades estruturadas estrategicamente para promover a progressiva autonomia intelectual do aluno, habilidade preconizada pelas Diretrizes Curriculares Nacionais. Organiza-se, assim, a partir de um problema formulado sobre temática interdisciplinar, utilizando uma metodologia que leve os alunos a desenvolverem competências e habilidades para: Identificar, planejar e resolver problemas; Abstrair, analisar, sintetizar e produzir conhecimentos; Levantar e investigar hipóteses; Aprender com autonomia e atualizar-se permanentemente; Trabalhar em equipe; Analisar cenários e tomar decisões; Comunicar-se oralmente e por escrito; Criar e inovar; Adquirir conhecimentos sobre a área de estudo e a profissão; Buscar, processar e analisar informações.

Trata-se de uma prática pedagógica que prevê uma forma flexível de aprendizagem, pautada no trabalho, ora individual, ora em equipe, e capaz de despertar no aluno, pelo desafio da autoaprendizagem, a curiosidade e o estímulo à investigação na busca de soluções para assuntos de seu interesse, em articulação com os propósitos da formação específica.

O tema do Projeto Aplicado está diretamente relacionado à formação de habilidades (específicas e globais) e competências do curso e de competências descritas nas Diretrizes Curriculares Nacionais, de modo a corroborar com a formação profissional, humana e cidadã dos alunos da instituição, ajudando-os na sua inserção político-social. Adota ainda, como princípio, o papel ativo dos estudantes na construção do conhecimento, em que o processo de aquisição do saber é mais importante que o próprio saber.

A inclusão da disciplina na estrutura curricular dos cursos da Universidade São Judas Tadeu vem propiciar, por meio da elaboração coletiva e da troca de experiências necessárias à sua realização, uma constante avaliação e revitalização dos processos de ensino e aprendizagem, levando os professores a se organizar para planejar suas ações, avaliar suas consequências e replanejá-las. Os discentes trabalham de forma sistemática, organizada e solidária em grupos, tanto em sala de aula quanto em outros espaços de aprendizagem, visando construir sua autonomia acadêmica, intelectual, política e profissional.

O trabalho é desenvolvido em todos os módulos do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas. Para tal, os alunos são orientados pelo professor da disciplina sobre as regras de construção, apresentação e avaliação do projeto, como descrito nas próximas seções.

2. Aplicação e Condução do Projeto Aplicado

O Projeto Aplicado foi implementado de forma integral em duas turmas (1º. Semestre de 2015 e 2º. Semestre de 2015). Sendo que a primeira turma cursou o Módulo A - Ciclo 1, referente ao eixo de Conceitos de Sistemas de Informação e Infraestrutura, e a segunda turma cursou o Módulo B - Ciclo 1, referente ao eixo de Programação Básica.

2.1. Módulo A – Ciclo 1: Projeto Aplicado em Fundamentos de Sistemas de Informação e Infraestrutura

O Módulo A - Ciclo 1 – Eixo: Conceitos de Sistemas de Informação e Infraestrutura é composto pelas seguintes disciplinas: Ambientes de Programação; Fundamentos de Infraestrutura de TI; Fundamentos de Segurança da Informação; Fundamentos de Sistemas de Informação; Leitura e Produção de Texto; Projeto Aplicado em Fundamentos de Sistemas de Informação.

2.1.1. Planejamento

Para os alunos do Módulo A – Ciclo 1, o tema definido a ser implementado na disciplina de Projeto Aplicado em Fundamentos de Sistemas de Informação e Infraestrutura foi um sistema de controle de estoque em uma empresa de eventos. Os produtos de consumo (alimentos e bebidas) deveriam ter sua entrada e saída controlados, devidamente vinculados à cada evento, o que geraria um inventário no final de cada evento, com relatórios de consumo. Os produtos de patrimônio seriam cadastrados no estoque geral da empresa e vinculados à cada evento, devendo ser verificada a sua reintegração após a realização deste.

Os professores do módulo se reuniram diversas vezes, sendo apresentados ao tema e relatando como poderiam contribuir, ou seja, quais assuntos de suas disciplinas seriam

considerados no escopo do projeto. Houve uma plena colaboração dos professores, que foi um fator fundamental para o sucesso do projeto.

Uma das regras do Projeto Aplicado é que o escopo deveria abordar elementos reais da área de tecnologia, abrangendo elementos e conceitos das disciplinas envolvidas, estimulando pesquisas e a busca de conteúdos adicionais, além do conteúdo programático de cada disciplina. Ou seja, cada professor ensinava o conteúdo da sua disciplina, e de acordo, com o conteúdo ensinado os alunos deveriam pesquisar como aplica-los no Projeto Aplicado.

2.1.2. Escopo

O escopo do projeto foi dividido basicamente em duas partes:

1. Estudo do conteúdo de disciplinas teóricas, como: Segurança, Infraestrutura e Fundamentos de Sistemas de Informação e sua aplicação no tema central de controle de estoque em uma empresa de eventos. Para isso, o professor do projeto aplicado, apresentava algumas questões, como por exemplo: Como garantir a segurança nesse sistema? Quais equipamentos e infraestrutura seriam necessária? Que tipo de sistema deveria ser desenvolvido e quais poderiam ser integrados? Como artefato desse estudo teórico os alunos geravam um documento apoiado pela disciplina de Leitura e Produção de Texto.
2. Aplicação prática realizada na disciplina de Ambientes de Programação, na qual os alunos deveriam desenvolver um site com navegação, porém sem banco de dados, utilizando HTML, CSS e jQuery, considerando conceitos de usabilidade e navegabilidade, além da obrigatoriedade da navegação estar funcional, por botões e menus.

2.1.3. Grupo de Alunos

A turma desse módulo tinha aproximadamente 100 alunos, foram divididos grupos de 5 a 7 alunos. Além da aula de Projeto Aplicado os alunos tinham um laboratório de informática à disposição para os encontros dos grupos aos sábados, das 9:00 às 17:00hs com os softwares necessários.

2.1.4. Acompanhamento do Projeto

A implementação, bem como o acompanhamento do projeto seguiram os conceitos básicos de gestão de projetos. Os alunos deveriam entregar a cada 15 dias artefatos intermediários de pesquisas ou interfaces gráficas. Além das entregas programadas, em todas as aulas era realizado um sorteio para que um grupo apresentasse o que havia realizado nos últimos 15 dias de trabalho.

2.1.5. Entrega e Avaliação

A apresentação foi realizada por meio de banners expostos aos alunos de toda a Universidade, durante uma semana, e no dia agendado uma banca composta por 2 professores das disciplinas do módulo e 1 professor convidado avaliavam o projeto. Os alunos ainda apresentavam e explicavam o protótipo navegável.

Para a apresentação foram consideradas: desenvoltura e demonstração de conhecimento, grau de aplicação dos assuntos solicitados no projeto, conhecimento individual dos

integrantes, postura e comprometimento. Após a apresentação, os professores da banca realizaram arguição individual.

2.2. Módulo B – Ciclo 1: Projeto Aplicado em Programação Básica

Este módulo possuiu todas as etapas do Projeto Integrado realizado no Módulo A – Ciclo 1, mas com um diferencial, o eixo desse módulo era mais prático, visando a programação e desenvolvimento de software. O Módulo B – Ciclo 1 – Eixo: Programação Básica é composto pelas seguintes disciplinas: Aspectos Regulatórios e Legais em TI; Banco de Dados; Matemática Computacional; Programação Aplicada; Raciocínio Lógico; Projeto Aplicado em Programação.

2.2.1. Planejamento

Para os alunos do Módulo B – Ciclo 1, o tema definido a ser implementado na disciplina de Projeto Aplicado em Programação Básica foi um sistema de software de controle de estoque e vendas de uma loja de rua que deveria ser implementado em Java e MySQL. Os professores do módulo se reuniram diversas vezes, para sincronizar o conteúdo das disciplinas com os artefatos a serem entregues no projeto.

2.2.2. Grupos de Alunos

Foram seguidas as mesmas regras do Módulo A – Ciclo 1, descrito no item 2.1.3.

2.2.3. Acompanhamento do Projeto

Foram seguidas as mesmas regras do Módulo A – Ciclo 1, descrito no item 2.1.4.

2.2.4. Entrega e Avaliação

Foram seguidas as mesmas regras do Módulo A – Ciclo 1, descrito no item 2.1.5, com exceção do banner, que não foi solicitado, pelo perfil do eixo ser mais técnico, com isso, a apresentação foi realizada no laboratório de informática com a demonstração do software e análise de código e banco de dados.

3. Resultados Obtidos na Aprendizagem dos Alunos

Considerando que foram turmas dos primeiros semestres o resultado foi muito satisfatório, observou-se pelo grupo de professores que os alunos se sentiram motivados na busca de mais conhecimento e desafiados no desenvolvimento do projeto aplicado. As apresentações finais foram muito boas, sendo perceptível o orgulho e comprometimento na apresentação, apesar da timidez de alguns alunos.

No início do semestre observou-se uma falta de maturidade natural dos alunos, principalmente nos quesitos integração e relacionamento com os colegas do grupo. O professor do Projeto Aplicado orientou e auxiliou inclusive na distribuição de tarefas e prazos de forma detalhada por grupo. As entregas previamente agendadas, com desconto de notas por atrasos foram fundamentais para o comprometimento e a organização dos grupos.

Alguns professores inicialmente foram reticentes com o Projeto Aplicado, não acreditavam que a integração entre as disciplinas poderia acontecer, no final esses mesmo professores consideraram que o Projeto Aplicado é uma boa forma de ensino-aprendizado.

Atividades Interdisciplinares em um Curso de Engenharia de Software: Desenvolvendo Softwares Tradicionais e Jogos

André Roberto Ortoncelli¹, Marisângela Pacheco Brittes¹, Rafael A. P. Oliveira¹

¹Universidade Tecnológica Federal do Paraná - (UTFPR) Câmpus Dois Vizinhos
Estr. p/ Boa Esperança, S/n - Zona Rural, Dois Vizinhos - PR, 85660-000
Coordenação de Engenharia de Software – COENS
Dois Vizinhos-PR.

{ortoncelli, mbrittes, raoliveira}@utfpr.edu.br

Abstract. *This paper presents two approaches used to promote interdisciplinarity in a Software Engineering undergraduate course. Such approaches involve the design of games and traditional software, being carried out through Supervised Practical Activities (APS). The objective of the proposed activities is offering the opportunity to students develop a less superficial work, closer to the market reality, also aiming at increasing the synergy between the knowledge areas and the motivation of the students.*

Resumo. *Esse trabalho apresenta duas abordagens utilizadas para promover a interdisciplinaridade em um curso de graduação em Engenharia de Software. Tais abordagens envolvem projetos de jogos e softwares tradicionais, realizados por meio de Atividades Práticas Supervisionadas (APS). O objetivo das atividades propostas é oferecer aos alunos a oportunidade de desenvolverem um trabalho menos superficial e mais próximo da realidade do mercado, visando também aumentar a sinergia entre as disciplinas e a motivação dos alunos.*

1. Introdução

A interdisciplinaridade consiste da combinação de duas, ou mais, disciplinas acadêmicas para um intuito qualquer de uma atividade particular (por exemplo, um projeto ou trabalho). Atividades interdisciplinares são desafiadoras pelo fato de envolverem o rompimento de barreiras tradicionais e, por fim, por promoverem a interação e pensamento na relação entre as disciplinas envolvidas [Yamamoto et al. 2003]. Diversos recursos interdisciplinares têm sido usados no estado da prática de diferentes áreas acadêmicas e na indústria. No contexto de software, a interdisciplinaridade muitas vezes é um recurso indispensável em alguns projetos de pesquisa, bem como para a educação [Yamamoto et al. 2005].

Diversas limitações podem ser detectadas na qualidade do ensino de *Engenharia de Software* (ES) nas universidades contemporâneas. Tais limitações são variadas e, em geral, dependem da grade de disciplinas e da ementa dos cursos de computação. Visando identificar os principais problemas do ensino de ES em diversas universidades, [Paiva 2011] apresenta uma revisão sistemática de literatura na qual são destacados 11 desafios para o ensino de ES nos dias atuais: (i) distância entre o conteúdo ensinado e a realidade do mercado de trabalho; (ii) falta de interesse do discente em disciplinas massivamente teóricas; (iii) altas taxas de evasão e reprovação; (iv) índices de aprendizagem efetiva bastante questionáveis; (v) pouca ênfase no trabalho em grupo; (vi) pouco estímulo à interdisciplinaridade; (vii) o planejamento de conteúdos e práticas pedagógicas não tem o crivo institucional; (viii) de forma geral, o que se ensina e aprende em um semestre ou disciplina, não é apropriado pela comunidade como um patrimônio intelectual

para uso futuro; (ix) pouca ou nenhuma interação entre professores e alunos fora da sala de aula; (x) o conhecimento produzido em um curso/disciplina não é compartilhado com a comunidade externa; e (xi) cada instituição/docente segue sua própria linha de acordo com a sua visão/concepção.

Com o intuito de aliviar os desafios supracitados se apropriando de conceitos de interdisciplinaridade, o presente estudo descreve um relato de experiência conduzido na no curso Bacharelado em Engenharia de Software *Universidade Tecnológica Federal do Paraná* (UTFPR) câmpus Dois Vizinhos, no qual as *Atividades Práticas Supervisionadas* (APS) de duas disciplinas foram integradas de modo a serem complementares e, por fim, uma disciplina foi completamente programada de modo a atacar alguns dos desafios identificados em [Paiva 2011]. Entre as contribuições do estudo, destacam-se: (i) a descrição de duas iniciativas integradoras no quarto semestre do curso; e (ii) uma breve discussão acerca do uso de iniciativas integradoras em disciplinas de cursos de Computação.

Uma APS consiste em uma ou mais atividades extraclasse, que devem ser acompanhadas pelo professor da disciplina, pois integralizam a carga horária da disciplina, podendo compor também a nota do aluno na disciplina. Considera-se esse tipo de atividade de extrema importância para a formação do aluno pois ela permite autonomia para que o aluno aprenda a desenvolver suas competências resolvendo problemas em grupo ou individualmente.

Além desta seção introdutória, o restante desse trabalho está organizado da seguinte maneira: A Seção 2 descreve a proposta de uma APS interdisciplinar referente ao projeto de jogos. Na Seção 3 são descritas as atividades da disciplina de Oficina de Integração. Por fim, uma breve discussão e as conclusões do trabalho são apresentadas na Seção 4.

2. Atividade Interdisciplinar: Projeto de Jogos

No quarto período do curso de Engenharia de Software são ofertadas as disciplinas de Requisitos de Software e Análise Orientada a Objetos, as quais possuem conteúdos teórico-práticos complementares. Desse modo, optou-se por criar uma APS que pudesse contemplar o conteúdo das duas disciplinas, permitindo que os alunos matriculados nas duas disciplinas pudessem desenvolver uma atividade mais complexa, sendo avaliados simultaneamente nas duas disciplinas.

Essa APS consiste em duas atividades relacionadas a engenharia reversa de jogos digitais, na primeira atividade é apresentado aos alunos o projeto de um jogo simples desenvolvido em Java (para o qual, os alunos tem acesso ao código completo), e na segunda parte foi sorteado para cada grupo (de até 3 alunos) um jogo disponível na web¹.

A ideia de utilizar conceitos do desenvolvimento de jogos, surgiu pelo fato de que todos os projetos até então desenvolvidos durante as disciplinas do curso tiveram foco no desenvolvimento de aplicações comerciais e gerenciais. Entendeu-se a importância de se diversificar o tipo de aplicação a ser estudada, agregando novos conhecimentos para a formação dos alunos. O jogo utilizado na primeira parte foi desenvolvido utilizando o framework JPlay². na linguagem de programação Java, sendo um jogo muito simples,

¹acesso: <http://www.clickjogos.com.br/>

²acesso: <http://www2.ic.uff.br/jplay/>

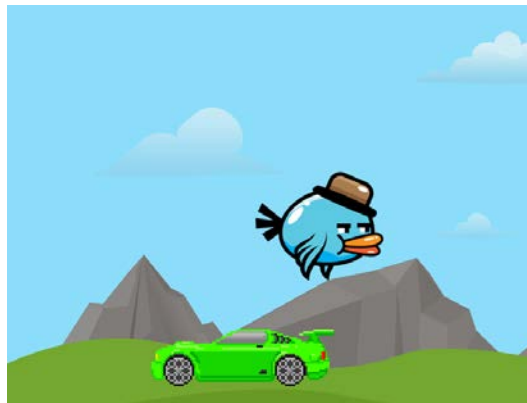


Figura 1. Jogo Utilizado na Primeira Atividade

consistindo em pássaro, controlado pelo usuário, que não pode colidir com um carro (caso essa colisão ocorra, o jogo é encerrado). A Figura 1 apresenta a interface desse jogo. Apesar do enredo simples desse jogo, o objetivo é que os alunos aprendam os conceitos de desenvolvimento de jogos que serão necessários para a APS.

Apesar do curso de Engenharia de Software em questão não ter uma disciplina regular relacionada ao desenvolvimento de jogos, nas disciplinas, foram dedicadas aulas específicas para capacitar os alunos nos conhecimentos básicos relacionados ao projeto de jogos digitais.

Na disciplina de Análise Orientada a Objetos foi cobrado que os alunos produzissem quatro diagramas da UML para cada jogo: (i) Diagrama de Casos de Uso; (ii) Diagrama de Classes; (iii) Diagramas de Sequência; e (iv) Diagramas de Estados.

Já na disciplina de Requisitos de Software, cada uma das equipes deveria identificar os requisitos funcionais e não funcionais de cada um dos jogos, e documentá-los, conforme a especificação de requisitos utilizado na disciplina.

O resultado final do trabalho foi entregue em uma única documentação e apresentado em forma de seminário, assistido e avaliado pelos professores das duas disciplinas.

3. Oficina de Integração

A disciplina de Oficina de Integração (também ofertada no quarto semestre do curso) visa à revisão dos conhecimentos das disciplinas de formação básica e profissionalizante ocorridas até o período corrente do curso, tendo como foco a aplicação dos conhecimentos adquiridos no desenvolvimento de um sistema computacional que contemple tal integração.

Diferentemente da APS descrita na Seção 2, nessa disciplina os alunos não desenvolvem um jogo, mas sim, um software "tradicional", que envolve a persistência de dados em um banco de dados relacional.

Nessa disciplina os alunos são divididos em grupos, que devem implementar um produto de software completo, utilizando um processo de desenvolvimento de software definido previamente (para cada grupo é definido por sorteio um tema diferente).

Os alunos desenvolvem a maior porcentagem desse trabalho nas aulas reservadas para disciplina de Oficina de Integração, porém algumas atividades são trabalhadas em

disciplinas que ocorrem no mesmo semestre, por exemplo, em uma ou duas aulas da disciplina de Análise Orientada, na qual os alunos realizam parte da modelagem do software.

A avaliação dos alunos nessa disciplina é processual e composta por entregas fracionadas até a entrega completa e final do produto implementado, por meio da confecção de relatórios e documentação e também por apresentações orais durante as aulas, com participação nas atividades propostas.

4. Discussões e Considerações Finais

Nesse trabalho é apresentado como estão sendo realizadas atividades interdisciplinares no quarto semestre do curso de Engenharia de Software da UTFPR (Campus Dois Vizinhos). Com as atividades descritas busca-se aumentar a sinergia entre as disciplinas, criando um ambiente de ensino-aprendizagem mais efetivo e motivador para alunos.

Outro fator que motiva a realização de atividades interdisciplinares é o fato de que enquanto os alunos trabalham em torno de uma situação problema interdisciplinar, além de promover a cooperação entre as disciplinas, os alunos desenvolvem um trabalho menos superficial e mais próximo (ou até iguais) dos trabalhos quando estiverem trabalhando com projetos reais. Observa-se claramente a importância da percepção dos alunos compreenderem a relação prática entre as áreas do conhecimento no desenvolvimento de um projeto de software. Enquanto teoria, há uma nítida dificuldade em abstrair conceitos que na prática, são trabalhados em conjunto. Essa estratégia de interdisciplinaridade apresentou-se de fundamental importância não somente para fixação e correlação de conteúdos das disciplinas, mas também para apoiar o direcionamento dos futuros profissionais em formação no curso, através de práticas que possibilitam vislumbrar o mercado de desenvolvimento de software em ambiente acadêmico.

Como objetivos futuros pretende-se discutir estratégias para implementação de iniciativas interdisciplinares nos demais semestres do curso, além de definir uma metodologia formal para validar a eficácia das atividades interdisciplinares apresentadas nesse trabalho.

Referências

- [Paiva 2011] Paiva, S. R. (2011). Uma revisão sistemática das pesquisas realizadas sobre a melhoria no ensino de engenharia de software. Technical report, Universidade Federal da Paraíba – UFPB/CCEN.
- [Yamamoto et al. 2003] Yamamoto, F. S., da Silva, A. F., and Zanutto, J. (2003). Projetos interdisciplinares no ensino de ciência da computação. In *I WEIBASE - Workshop de Educação em Computação Bahia-Sergipe*, Bahia-Sergipe.
- [Yamamoto et al. 2005] Yamamoto, F. S., da Silva, A. F., Zanutto, J., and Zampiroli, F. A. (2005). Interdisciplinaridade no ensino de ciência da computação. In *Anais do Workshop sobre Educação em Computação (WEI 2005) integrado com o XXV Congresso da Sociedade Brasileira de Computação*, São Leopoldo, RS.

Interdisciplinaridade no IF Farroupilha - Campus Santo Ângelo através da Prática Profissional Integrada

Karlise Soares do Nascimento¹, Fábio Weber Albiero¹, Andréa Pereira¹, Joice Machado¹

¹Instituto Federal Farroupilha – Campus Santo Ângelo
Caixa Postal 98.806-700 – Santo Ângelo – RS – Brasil

{karlise.nascimento, fabio.albiero}@iffarroupilha.edu.br

{andrea.pereira, joice.machado}@iffarroupilha.edu.br

Abstract. *This paper includes reports on Professional Practice Integrated (PPI) developed at the Federal Institute Farroupilha - Campus Santo Angelo in the year 2015, on Information and Communication shaft, which covers the following courses: Technical Course on Maintenance and Support in Computer - Integrated to High School, Technical Course on Internet Information - Subsequent (post high school) and Technology on Internet Systems course. The PPI is a series of activities aimed the integration of different disciplines. As an example, on the Technical Course on Maintenance and Support in Computer, there is the disciplines integration of basic area (Art, Biology, etc.) and disciplines of technical area (Computer Organization and Architecture, etc.).*

Resumo. *Este relato apresenta a Prática Profissional Integrada (PPI) desenvolvida no Instituto Federal Farroupilha - Campus Santo Ângelo, no ano de 2015, no eixo Informação e Comunicação, o qual abrange os seguintes cursos: Curso Técnico em Manutenção e Suporte em Informática - Integrado ao Ensino Médio, Curso Técnico em Informática para Internet - Subsequente (Pós Ensino Médio) e Curso Superior de Tecnologia em Sistemas para Internet. A PPI é uma série de atividades que visam a integração de diferentes disciplinas. Como exemplo, no Curso Técnico em Manutenção e Suporte em Informática, há a integração de disciplinas da área básica (Arte, Biologia, etc.) com disciplinas da área técnica (Organização e Arquitetura de Computadores, etc.).*

1. Introdução

O Instituto Federal Farroupilha, em 2014, fundou na cidade de Santo Ângelo, um de seus onze campi no estado do Rio Grande do Sul. O Campus Santo Ângelo atua nos eixos tecnológicos de Informação e Comunicação, Ambiente e Saúde e Recursos Naturais, nos diferentes níveis de educação: Técnico Integrado ao Ensino Médio, Subsequente (Pós Ensino Médio), Superior, entre outras modalidades. Desta forma, faz-se necessário relatar sobre a interdisciplinaridade nos diferentes níveis, uma vez que os mesmos professores atuantes no curso superior também podem atuar nas demais modalidades.

O Corpo Docente, o Setor de Apoio Pedagógico e a Direção de Ensino, planejam conjuntamente, a interdisciplinaridade prevista na legislação, através de reuniões que discutem como será abordada e quais as disciplinas serão articuladas no ano ou semestre letivo e documentam as decisões no projeto da Prática Profissional Integrada (PPI). É

por meio desta prática que o IF Farroupilha desenvolve a interdisciplinaridade entre os componentes curriculares de cada curso.

Este artigo apresenta, na seção 2, relatos das Práticas Profissionais Integradas desenvolvidas no ano de 2015, que ocorreram no Curso Técnico em Manutenção e Suporte em Informática - Integrado ao Ensino Médio, no Curso Técnico em Informática para Internet - Subsequente (Pós Ensino Médio) e no Curso Superior de Tecnologia em Sistemas para Internet. A seção 3 apresenta as considerações finais.

2. Interdisciplinaridade no IF Farroupilha - Campus Santo Ângelo

A interdisciplinaridade tem como objetivo principal desenvolver um trabalho de integração dos conteúdos de uma disciplina com outras áreas de conhecimento, o que contribui para o aprendizado do aluno. Esta interação entre disciplinas aparentemente distintas, é uma maneira complementar que possibilita a formação de um saber crítico-reflexivo. É através dessa perspectiva que a interdisciplinaridade surge como uma alternativa para superar a fragmentação entre as disciplinas, proporcionando um diálogo entre estas e relacionando-as entre si para a compreensão da realidade [Carlos 2007].

No Instituto Federal Farroupilha, a interdisciplinaridade se faz presente por meio da Prática Profissional Integrada (PPI), que constitui parte do Projeto Pedagógico de Curso (PPC) dos três cursos já citados, do eixo Informação e Comunicação.

3. Prática Profissional Integrada

De acordo com a instrução normativa nº 02, de 18 de fevereiro de 2013, do IF Farroupilha, entende-se por Prática Profissional Integrada a articulação da teoria no processo de ensino e aprendizagem [Farroupilha 2013], visando a interdisciplinaridade assegurada no currículo e na prática pedagógica; e, superando a fragmentação de conhecimentos, imposta na organização curricular. A PPI tem por finalidade ser uma estratégia educacional favorável para a contextualização, flexibilização e integração entre o conhecimento teórico e profissional em prática.

3.1. PPI no Curso Técnico em Manutenção e Suporte em Informática - Integrado ao Ensino Médio

No Curso Técnico em Manutenção e Suporte em Informática - Integrado ao Ensino Médio, a Prática Profissional Integrada, teve como objetivo principal receber doações de equipamentos eletrônicos, reciclá-los ou descartá-los, de forma correta. Os objetivos específicos constituíram em: promover o reuso de equipamentos eletrônicos, aumentando seu tempo de vida e reduzindo a quantidade de lixo eletrônico; oportunizar aos alunos uma reflexão sobre o lixo eletrônico; incentivar os alunos na questão da cidadania solidária; promover a inclusão digital através do reuso desses equipamento; e, por fim, evitar a contaminação ambiental e prejuízos a saúde pública devido à disposição incorreta do lixo eletrônico.

Esta PPI integrou disciplinas da área básica, tais como: Arte, Biologia, Filosofia, Física, Geografia, Língua Portuguesa, Matemática, Química e Sociologia; com as disciplinas da área técnica: Introdução à Informática e Organização e Arquitetura de Computadores, por meio de uma série de atividades propostas.

A metodologia empregada consistiu, primeiramente, no recebimento de doações de componentes eletrônicos através dos alunos e dos servidores do campus. Após a etapa de coleta desses componentes, realizaram-se as atividades propostas em cada uma das disciplinas que integraram a PPI. Por exemplo, na disciplina de Química, a atividade proposta consistiu no estudo sobre os elementos químicos que compõe cada componente físico do dispositivo eletrônico. Por sua vez, na disciplina de Organização e Arquitetura de Computadores, a atividade proposta consistiu na realização de testes dos componentes físicos, verificando quais são as funções de cada componente e quais poderiam ser reproveitados para a construção de novos dispositivos eletrônicos. Os componentes físicos que apresentaram defeitos ou falhas foram utilizados como matéria prima na disciplina de Artes para a elaboração de obras de arte.

3.2. PPI no Curso Técnico em Informática para Internet - Subsequente (Pós Ensino Médio)

No Curso Técnico em Informática para Internet - Subsequente (Pós Ensino Médio) foram realizadas duas Práticas Profissionais Integradas, uma em cada semestre letivo do ano de 2015. A primeira PPI teve como objetivo principal a elaboração de uma página web, tendo como conteúdo um glossário de palavras técnicas de origem inglesa, que são utilizadas na área de informática. Desta forma, esta PPI buscou integrar as disciplinas de Aplicativos Web, Inglês I, Fundamentos da Computação e Metodologia Científica.

A metodologia utilizada para a primeira PPI consistiu na divisão da turma em pequenos grupos, os quais foram orientados a selecionar conceitos técnicos abordados na disciplina de Fundamentos da Computação. Após esta etapa, cada grupo de alunos produziu um glossário dos termos em inglês com seus respectivos significados em português. Posteriormente, cada um dos grupos construiu uma página web (utilizando conceitos aprendidos na disciplina de Aplicativos Web), aonde apresentaram os seus glossários. A conclusão desta atividade, ocorreu em forma de seminário, com a apresentação da página web criada e do respectivo glossário para os colegas e docentes do curso, além de um relatório integrando assim a disciplina de Metodologia Científica. Esta PPI proporcionou aos alunos do curso um aprofundamento nos conceitos técnicos trabalhados nas disciplinas de Fundamentos da Computação e Aplicativos Web, assim como um aprimoramento da capacidade de compreensão e escrita do vocabulário da língua inglesa e portuguesa.

A segunda PPI teve como objetivo principal ampliar o conhecimento dos alunos sobre a construção de páginas web, levando em consideração a modelagem de um banco de dados e do sistema proposto. Esta PPI previu atividades interativas nos laboratórios de informática do campus, integrando as disciplinas de Banco de Dados, Análise e Modelagem de Sistemas, Programação Web Orientada a Objetos e Português Instrumental.

Quanto a metodologia, os discentes, de modo individual, construíram uma página web, implementada em PHP, e modelaram o banco de dados e as funcionalidades do sistema, apresentando os diagramas produzidos em forma de relatório. O tema para a construção da página web foi escolhido pelo próprio discente. A etapa final, consistiu na socialização do trabalho, em forma de seminário para os colegas e os docentes do curso.

3.3. PPI no Curso Superior de Tecnologia em Sistemas para Internet

No Curso Superior de Tecnologia em Sistemas para Internet foram realizadas duas Práticas Profissionais Integradas, uma em cada semestre letivo do ano de 2015. A pri-

meira PPI teve como objetivo principal a elaboração de uma página web, tendo como conteúdo um glossário de palavras técnicas de origem inglesa que são utilizadas na área de informática. Diante disso, esta PPI integrou as disciplinas de Introdução à Informática, Fundamentos da Computação, Inglês Técnico e Leitura e Produção Textual.

A metodologia utilizada para o desenvolvimento desta PPI ocorreu de modo semelhante à metodologia empregada no curso subsequente: 1) divisão da turma em pequenos grupos, sendo orientados a selecionarem conceitos abordados na disciplina de Fundamentos da Computação; 2) produzir um glossário dos termos em inglês com seus respectivos significados em português; 3) construir uma página web; e, por fim, 4) socializar, em forma de seminário, a página web criada para os colegas e docentes do curso, além de um relatório escrito.

Também de modo semelhante à PPI aplicada ao curso subsequente, a segunda PPI do curso superior teve como objetivo principal ampliar os conceitos dos alunos no que tange a construção de páginas web, levando em consideração a modelagem do banco de dados e a interação humano-computador (acessibilidade). Esta PPI também previu atividades interativas nos laboratórios de informática e a interdisciplinaridade entre os seguintes componentes curriculares: Construção de Páginas Web, Banco de Dados I, Interação Humano-Computador e Metodologia Científica.

4. Considerações Finais

Este artigo apresentou relatos da Prática Profissional Integrada (PPI) desenvolvida no eixo Informação e Comunicação, do Instituto Federal Farroupilha - Campus Santo Ângelo, no ano de 2015. A PPI é uma forma encontrada pelo IF Farroupilha de pôr em prática a interdisciplinaridade, contribuindo para o aprendizado do aluno.

A execução dessas práticas em nossa instituição de ensino gerou como resultado a publicação de alguns trabalhos, sendo um deles intitulado: “Números do lixo eletrônico, criatividade, criação e conscientização”, desenvolvido no Curso Técnico em Manutenção e Suporte em Informática - Integrado ao Ensino Médio. Diante dos resultados positivos, pretende-se dar continuidade ao desenvolvimento da Prática Profissional Integrada visando sempre melhora-lá, a fim de que, em um dado momento, a interdisciplinaridade esteja fortemente dissolvida na estrutura curricular de cada um dos cursos da instituição.

Além disso, percebe-se um grande envolvimento por parte dos alunos, com a pesquisa e realização das atividades propostas, permitindo que visualizem de forma prática, os conceitos trabalhados em sala de aula.

Referências

- Carlos, J. G. (2007). Interdisciplinaridade no ensino médio: desafios e potencialidades. Disponível em: http://repositorio.unb.br/bitstream/10482/2961/1/2007_JairoGoncalvesCarlos.pdf. Acesso em: 03/05/16.
- Farroupilha, I. F. (2013). Instrução normativa nº 02, de 18 de fevereiro de 2013. Disponível em: http://w2.iffarroupilha.edu.br/site/midias/arquivos/2013119154234593instrucao_normativa_n%C2%B0_02_2013_-_pratica_profissional_integrada.pdf. Acesso em: 29/04/16.

CTD – Concurso de Teses e Dissertações

Apresentação

Há quase 30 anos a Sociedade Brasileira de Computação vem realizando o Concurso de Teses e Dissertações da área de Computação do País. Na 29ª edição, realizada no ano de 2016 em Porto Alegre, o CTD teve um total de 100 trabalhos submetidos, sendo 37 teses de doutorado e 63 dissertações de mestrado. Todos os 100 trabalhos foram defendidos no ano de 2015. O processo de seleção dos trabalhos contou com a participação de revisores de diferentes estados brasileiros e de várias áreas da Computação. Cada trabalho foi avaliado por pelo menos 3 revisores. Posteriormente à revisão dos artigos, cada avaliador teve acesso às revisões dos co-revisores. Em seguida foi estimulada a discussão entre os co-revisores com o objetivo de diminuir possíveis discrepâncias, potencializando uma recomendação final mais consistente. Assim como em edições anteriores, foram selecionadas as 6 melhores teses de doutorado e as 10 melhores dissertações de mestrado para apresentação oral e publicação no CTD 2016.

André Soares (UFPI)
Rodrigo Barros (PUCRS)

Comitê de Organização

Coordenação do CTD

André Soares (UFPI)
Rodrigo Barros (PUC-RS)

Comitê de Avaliação

Alba Melo (UnB)
Alberto Raposo (PUC-Rio)
Alceu Britto Jr. (PUC-PR)
Alexandre Falcão (UNICAMP)
Altigran Soares da Silva (UFAM)
Ana Shirley F. da Silva (UFC)
André Santanchè (UNICAMP)
Andre Aquino (UFAL)
Angelo Brayner (UNIFOR)
Anselmo Paiva (UFMA)
Antonio Alfredo Ferreira Loureiro (UFMG)
Artur Ziviani (LNCC)
Aurora Pozo (UFPR)
Bruno Schulze (LNCC)
Carlos Kamienski (UFABC)
Carlos Eduardo Ferreira (USP)
Célio Vinicius Neves de Albuquerque (UFF)
Cecilia Rubira (UNICAMP)
Celina Figueiredo (UFRJ)
Celso Alberto Saibel Santos (UFES)
Christina Chávez (UFBA)
Clarisse de Souza (PUC-Rio)
Claudia Linhares Sales (UFC)
Claudia Ribeiro (UERN)
Claudio Jung (UFRGS)
Cláudia Nalon (UnB)
Clodoveu Davis (UFMG)
Daniela Barreiro Claro (UFBA)
Diego Aranha (UNICAMP)
Divanilson Campelo (UFPE)
Edson Cáceres (UFMS)
Eduardo Camponogara (UFSC)
Eduardo Cerqueira (UFPA)
Eduardo Nakamura (UFAM)
Esteban Clua (UFF)
Eulanda Santos (UFAM)

Fabio Andre Porto (LNCC)
 Fábio Cozman (USP)
 Francisco Carvalho Junior (UFC)
 Helio Pedrini (UNICAMP)
 Ismar Frango Silveira (Universidade Cruzeiro do Sul)
 Ivan Silva (UFPI)
 Jayme Szwarcfiter (UFRJ)
 Joberto Martins (UNIFACS)
 Joceli Mayer (UFSC)
 José de Souza (UFC)
 José Palazzo Moreira de Oliveira (UFRGS)
 José Antonio Macedo (UFC)
 Leila Silva (UFS)
 Leonardo Murta (UFF)
 Liang Zhao (USP)
 Luciana Buriol (UFRGS)
 Luciano Digiampietri (USP)
 Luis Lamb (UFRGS)
 Luis Carlos De Bona (UFPR)
 Marcelo Finger (USP)
 Marcelo Pimenta (UFRGS)
 Marco Spohn (UFFS)
 Marco Tulio Valente (UFMG)
 Marcos Gonçalves (UFMG)
 Marcus Ritt (UFRGS)
 Maria da Graça Campos Pimentel (USP)
 Mario Teixeira (UFMA)
 Martin Musicante (UFRN)
 Mário Benevides (UFRJ)
 Mirella Moro (UFMG)
 Mylene Farias (UnB)
 Nabor Mendonça (UNIFOR)
 Nelson Rosa (UFPE)
 Newton Vieira (UFMG)
 Patrick Brito (UFAL)
 Paulo Maciel (UFPE)
 Pedro de Alcântara dos Santos Neto (UFPI)
 Rafael Prikkladnicki (PUC-RS)
 Renata Galante (UFRGS)
 Renato Krohling (UFES)
 Ricardo Torres (UNICAMP)
 Rodolfo Azevedo (UNICAMP)
 Rohit Gheyi (UFCG)
 Ronaldo Ferreira (UFMS)
 Silvia Vergílio (UFPR)
 Simone Martins (UFF)
 Thiago Noronha (UFMG)

Thiago Pardo (USP)
Uirá Kulesza (UFRN)
Wagner Meira Jr. (UFMG)
William Giazza (UnB)

Trabalhos aceitos

Trabalhos de Mestrado

Centralidade de tempo em grafos variantes no tempo

Eduardo Chinelate Costa (UFJF)

Alex Borges Vieira (UFJF)

Ana Paula Couto da Silva (UFMG)

Synthesis of code anomalies: revealing design problems in the source code

Willian Nalepa Oizumi (PUC-Rio)

Alessandro Garcia (PUC-Rio)

Partial least squares for face hashing

Cássio Elias dos Santos Júnior (UFMG)

William Robson Schwartz (UFMG)

Energy-aware load balancing in distributed data centers

Rodrigo Augusto Cardoso da Silva (UNICAMP)

Nelson Fonseca (UNICAMP)

Verificação de locutores independente de texto: uma análise de robustez a ruído

Hector Natan Batista Pinheiro (UFPE)

Tsang Ing Ren (UFPE)

George Darmiton da Cunha Cavalcanti (UFPE)

A gamified peer assessment model for on-line learning environments: An experiment with MeuTutor

Thyago Tenorio Martins de Oliveira (UFAL)

Ig Ibert Bittencourt (UFAL)

Um algoritmo exato em clusters de GPUs para o hitting set aplicado à inferência de redes de regulação gênica

Danilo Carastan dos Santos (UFABC)

Luiz Carlos da Silva Rozante (UFABC)

Heurísticas para desambiguação incremental de nomes de autores em referências bibliográficas

Alan F. Santana (UFMG)

Marcos Gonçalves (UFMG)

Multiple parenting relationships in image phylogeny

Alberto Arruda Oliveira (UNICAMP)

Anderson de Rezende Rocha (UNICAMP)

Quality of service provisioning in LTE-EPON integrated networks

Carlos A. Astudillo (UNICAMP)

Nelson Fonseca (UNICAMP)

Juliana Freitag Borin (UNICAMP)

Trabalhos de Doutorado

Efficient high-dimensional filtering for image and video processing

Eduardo Simões Lopes Gastal (UFRGS)

Manuel M. Oliveira Neto (UFRGS)

Algoritmos paralelos exatos e otimizações para alinhamento de sequências biológicas longas em plataformas de alto desempenho

Edans Sandes (UnB)

Alba Cristina Magalhães A. Melo (UnB)

Content delivery in vehicular ad hoc networks

Fabício Aguiar Silva (UFMG)

Linnyer Beatrys Ruiz (UEM)

Antonio Alfredo Ferreira Loureiro (UFMG)

Técnicas de projeção para identificação de grupos e comparação de dados multidimensionais usando diferentes medidas de similaridade

Paulo Joia (USP)

Luis G. Nonato (USP)

Electromigration aware cell design

Gracieli Posser (UFRGS)

Ricardo Reis (UFRGS)

Sachin S. Sapatnekar (University of Minnesota)

Algorithms for sorting by reversals or transpositions, with application to genome rearrangement

Gustavo Rodrigues Galvão (UNICAMP)

Zanoni Dias (UNICAMP)

Centralidade de Tempo em Grafos Variantes no Tempo

Eduardo Chinelate Costa¹, Alex Borges Vieira¹, Ana Paula Couto da Silva²

¹ Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Juiz de Fora – Juiz de Fora – MG

²Departamento de Ciência da Computação
Universidade Federal de Minas Gerais – Belo Horizonte – MG

Abstract. *Centrality usually refers to metrics that assess the relative importance of vertices. However, in Time-Varying Graphs (TVGs) it is possible to assess the importance of time instants (or states) of a graph throughout its existence. Determining important time instants may be useful to defining best times to spread, generate models and predict the behavior of TVGs. In this paper, we define time centrality in TVGs. Time centrality evaluates the relative importance of time instants. We present and evaluate two time centrality metrics focused on information dissemination processes. Our results show that the best classified time instants, according to created metrics, can make the diffusion process up to 2.5 times faster and achieve twice the number of nodes in certain cases.*

Resumo. *O conceito de centralidade geralmente se refere a métricas que avaliam a importância relativa dos vértices. Entretanto, Grafos Variantes no Tempo (GVTs) possibilitam a avaliação da importância dos instantes de tempo (ou estados) de um grafo ao longo de sua existência. Determinar instantes de tempo importantes pode ser útil para definir melhores momentos para difusão, gerar modelos e prever o comportamento de GVTs. Neste trabalho, é definido o conceito de Centralidade de Tempo em GVTs. A centralidade de tempo avalia a importância relativa dos instantes de tempo. São apresentadas e avaliadas duas métricas de centralidade de tempo voltadas a processos de difusão de informação. Os resultados mostram que os instantes de tempo melhor classificados, de acordo com as métricas criadas, podem tornar o processo de difusão até 2,5 vezes mais rápido e atingir o dobro do número de nós em certos casos.*

1. Introdução

Grafos variantes no tempo (GVTs) ganharam interesse recente, em especial por contemplar a evolução temporal, tanto de nós, quanto de arestas. Nesse novo cenário, de maneira similar a grafos estáticos, o conceito de centralidade geralmente se refere a métricas que avaliam a importância relativa dos vértices. Não obstante, pesquisadores sempre tentaram mapear o conceito do cenário tradicional para o novo.

Entretanto, GVTs possibilitam uma nova abordagem para se determinar centralidade. De fato, como agora é possível acompanhar a evolução do grafo, não só vértices e arestas são importantes, mas o tempo se torna um novo personagem a ser avaliado. Assim, uma nova maneira de se determinar centralidade em grafos é proposta. Neste trabalho, a *Centralidade de Tempo em Grafos Variantes no Tempo* é definida e formalizada, avaliando a importância dos instantes de tempo (ou estados) de um grafo ao longo de sua existência.

Determinar instantes de tempo importantes nesse contexto pode ter fortes aplicações práticas. Em particular, pode ser útil para definir melhores momentos para difusão de uma informação, gerar modelos e prever o comportamento de GVTs.

Além da formalização de Centralidade de Tempo, são apresentadas duas métricas voltadas a processos de difusão de informação em uma rede. As métricas foram avaliadas em um conjunto de dados real e os resultados indicam que, se os instantes de tempo com maior centralidade fossem usados para difundir algo na rede, tal processo seria até 2,5 vezes mais rápido. Caso houvesse limitações de tempo, os resultados também indicam um melhor desempenho da difusão com a escolha do instante de tempo apropriado. Nesse caso, a cobertura da difusão é até 2 vezes maior, quando comparado a uma escolha aleatória do momento de início da difusão.

2. Objetivos e Contribuições

Até onde se sabe, este é o primeiro trabalho que aborda a centralidade de tempo em GVTs. De fato, os trabalhos anteriores estão mais interessados em qualificar a importância dos nós em uma rede do que em determinar os instantes de tempo mais importantes no contexto de um GVT.

Para algumas aplicações modeladas em GVTs, mais importante do que identificar os nós centrais sob uma determinada definição, é identificar os principais instantes de tempo para a tomada de certas decisões. Decidir *quando*, e não somente *onde* iniciar um processo de difusão pode ser de extrema importância para um resultado mais eficaz.

Por exemplo, é mais interessante determinar o melhor instante de tempo no qual uma vacina pode ser aplicada em uma população para atingir melhores respostas contra uma epidemia. Por sua vez, [Spasojevic et al. 2015] mostrou recentemente evidências de que recomendar melhores tempos para um usuário postar em redes sociais aumenta a probabilidade de obtenção de respostas do público.

Visto que muitos eventos modelados por GVTs são notoriamente cíclicos, como mobilidade em redes de transportes e surtos epidêmicos de doenças, a Centralidade de Tempo seria útil na construção de modelos de previsão. As estratégias para a possível identificação antecipada de instantes de tempo centrais podem basear-se na descoberta de evidências no passado recente ou no momento presente da evolução do GVT, indicando que um determinado instante de tempo é relativamente melhor (isto é, mais central) com relação a um conjunto de outros.

Os resultados experimentais mostram que processos de difusão de informação iniciados em instantes de tempo melhor avaliados pelas métricas de centralidade de tempo — mais centrais — são efetivamente mais rápidos e/ou mais abrangentes que processos iniciados nos demais instantes de tempo. Quando comparado a uma escolha aleatória, um processo de difusão pode atingir 20% da rede até 2,5 vezes mais rápido. Já a cobertura da rede em um dado número de passos do processo de difusão praticamente dobra com relação aos instantes de tempo escolhidos aleatoriamente.

3. Centralidade de Tempo

Assim como na Centralidade de Nó, diferentes definições de Centralidade de Tempo podem refletir diferentes noções de importância para instantes de tempo visando aplicações

distintas. Em outras palavras, da mesma forma que a Centralidade de Nó cria um ranqueamento dos nós sob um critério de importância arbitrária, a Centralidade de Tempo cria um ranking de instantes de tempo usando algum critério para medir a importância relativa que um instante de tempo pode ter quando comparado com os outros.

São apresentadas duas métricas de centralidade de tempo que podem avaliar a difusão em GVTs a partir de diferentes perspectivas. Para definir tais métricas, um GVT genérico $H = (V, E, T)$ [Wehmuth et al. 2014] é considerado. Seja $N = |T(H)|$ o tamanho do conjunto $T(H)$. Cada instante de tempo individual $t_i \in T(H)$, onde $0 \leq i < N$, de modo que $T(H) = \{t_0, t_1, \dots, t_{N-1}\}$.

3.1. Tempo de Cobertura

A métrica de *Tempo de Cobertura (Cover Time (CT))* do ponto de vista da centralidade de tempo avalia o número mínimo de instantes de tempo que uma difusão leva para atingir uma dada fração dos nós. Mais formalmente, o tempo de cobertura $CT(t_i, \tau)$ retorna a quantidade média de instantes de tempo entre todos os nós para que uma difusão iniciada no tempo t_i alcance uma certa fração τ da rede, ou seja,

$$CT(t_i, \tau) = \begin{cases} \frac{1}{|V(H)|} \sum_{u \in V(H)} d_t(t_i, u, \tau), & \text{Se } \tau \text{ é atingido,} \\ \infty, & \text{Caso contrário.} \end{cases} \quad (1)$$

onde $d_t(t_i, u, \tau)$ é o número de instantes de tempo para que uma difusão iniciada no nó u no instante de tempo t_i alcance a fração τ de nós. Se a fração τ de nós não é alcançada pelo processo de difusão iniciado em qualquer nó em t_i , então $CT(t_i, \tau)$ é considerado ∞ .

O processo de difusão segue uma busca em largura (BFS), ou seja, um nó genérico (u, t_i) inicia a difusão por meio de envio de informações para todos os seus nós adjacentes. Em seguida, os nós temporais adjacentes retransmitem a informação para os seus próprios nós adjacentes no próximo instante de tempo, e assim por diante. A informação é distribuída até que a fração τ de nós seja atingida e o tempo necessário para a difusão é armazenado. Esta avaliação é repetida para cada nó $(u, t_i) \in VT(H)$. O tempo médio de difusão é a centralidade de tempo de cobertura do instante de tempo t_i .

3.2. Centralidade de Cobertura

A métrica de *Centralidade de Cobertura (Time-Constrained Coverage (TCC))* do ponto de vista da Centralidade de Tempo avalia a cobertura do GVT atingida por um processo de difusão depois de um número limitado de passos. Aqui é assumido o mesmo processo de difusão definido para o tempo de cobertura (Seção 3.1). Mais precisamente, para um dado instante de tempo t_i , $TCC(t_i, \Phi)$ retorna a fração média de nós do TVG atingidos pelo processo de difusão em Φ passos. Mais formalmente,

$$TCC(t_i, \Phi) = \frac{1}{|V(H)|^2} \sum_{u \in V(H)} d_c(t_i, u, \Phi), \quad (2)$$

onde $d_c(t_i, u, \Phi)$ é o número de nós atingidos a partir do nó u depois de Φ passos quando a difusão começa no tempo t_i . Novamente, é importante lembrar que a métrica de centralidade de cobertura considera uma média dos processos de difusão iniciados a partir de cada nó.

4. Avaliação da Centralidade de Tempo

As métricas de centralidade de tempo propostas foram avaliadas usando um conjunto de dados coletados no âmbito do projeto MOSAR [Lucet et al. 2012]. O MOSAR (*Mastering hOSpital Antimicrobial Resistance and its spread*) é um projeto de colaboração científica que compreende várias instituições de pesquisa, que se concentram na dinâmica de transmissão de bactérias resistentes aos antimicrobianos (AMRB) em ambientes de alto risco (unidades de terapia intensiva e centros cirúrgicos). O conjunto de dados adotado consiste no registro de contatos presenciais (de médicos, enfermeiros, funcionários e pacientes) em uma determinada ala hospitalar por um período de duas semanas. Cada um dos 160 voluntários que participaram no estudo foi equipado com um dispositivo de RFID que detectou a presença de outros dispositivos dentro de uma distância pequena (cerca de um metro). Os dispositivos de identificação eram únicos e sempre associados à mesma pessoa. A cada 30 segundos durante o período de duas semanas, cada dispositivo registrou a lista de todos os dispositivos (nós) que estavam dentro de sua área de cobertura, a fim de estabelecer o arranjo dos contatos entre eles (arestas).

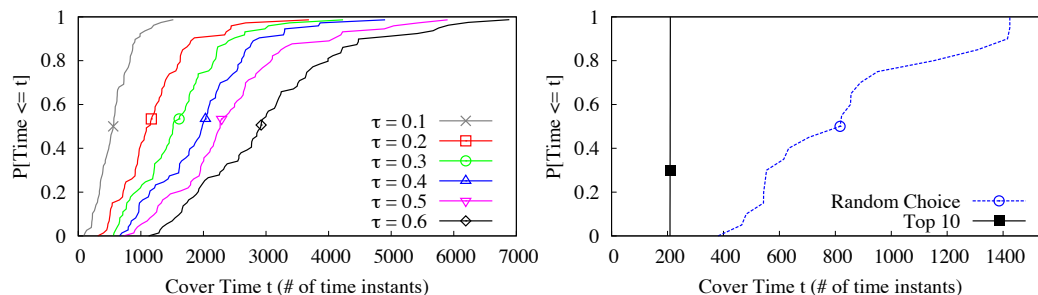
O conjunto de dados MOSAR é modelado como um GVT $M = (V, E, T)$, conforme descrito em [Wehmuth et al. 2014], onde $V(M)$ é o conjunto de todos os participantes, $E(M)$ é o conjunto de todos os contatos interpessoais, e $T(M)$ é o conjunto de instantes de tempo sequenciais com 30 segundos de granularidade. Dessa forma, $|V(M)| = 160$ e $|T(M)| = 40320$. Os contatos são representados por arestas não-orientadas, ou seja, a aresta entre duas pessoas é representada por uma aresta orientada e sua recíproca.

4.1. Tempo de Cobertura

Inicialmente são mostrados os resultados para a métrica de Tempo de Cobertura (Figura 1). Vale ressaltar que menores valores de tempo de cobertura indicam instantes de tempo mais centrais, que podem difundir uma informação em tempo.

A Figura 1(a) apresenta a função de distribuição acumulada (CDF) do tempo de cobertura para diferentes valores da fração τ de nós atingidos no GVT MOSAR. Seis diferentes cenários são considerados com $\tau = \{0.1, 0.2, \dots, 0.6\}$. Os resultados apresentam os primeiros 33000 instantes de tempo, para que haja espaço para a difusão se espalhar ao longo dos 40320 instantes de tempo do GVT MOSAR.

Cada valor em uma curva representa a fração dos instantes de tempo do GVT que alcançam a correspondente fração τ de nós em, no máximo, t instantes de tempo. A fim melhorar a visualização dos dados, os símbolos estão em posições arbitrárias nas curvas (o mesmo acontece com as demais figuras). Por exemplo, para atingir 60% do GVT MOSAR ($\tau = 0.6$), 20% dos instantes de tempo leva menos de 2000 instantes de tempo. Em outras palavras, a difusão da informação iniciada em até 20% dos instantes de tempo leva menos de 16 horas e 40 minutos para atingir 60% dos nós no GVT MOSAR, somente difundindo de maneira oportunista através do contato presencial entre os indivíduos.



(a) Distribuição do Tempo de Cobertura

(b) 10 Melhores x Escolha Aleatória

Figura 1. Função de Distribuição Acumulada do Tempo de Cobertura na rede MOSAR

Na Figura 1(b), para o GVT MOSAR, é comparado o desempenho dos 10 melhores instantes de tempo de acordo com a métrica de tempo de cobertura contra 10 instantes de tempo escolhidos aleatoriamente para iniciar uma difusão. Tomando o resultado médio como referência, para atingir 10% dos nós ($\tau = 0.1$) no GVT MOSAR, o tempo de cobertura foi de 266 instantes de tempo quando a difusão começa nos 10 melhores instantes de tempo; enquanto que o tempo de cobertura foi de 680 instantes de tempo, quando a difusão começa nos 10 instantes de tempo escolhidos aleatoriamente. Ou seja, considerando os instantes de tempo mais centrais, a difusão leva 2.5 vezes menos tempo quando comparado com a difusão iniciada em instantes de tempo aleatórios.

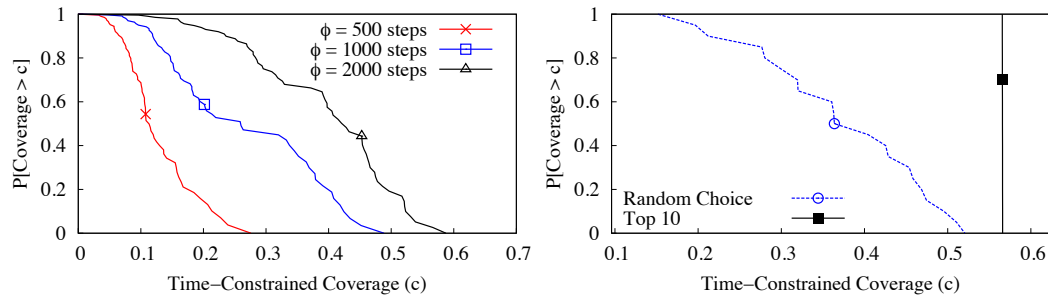
4.2. Centralidade de Cobertura

A centralidade de cobertura também é capaz de selecionar instantes de tempo que apresentem desempenhos de difusão distintos. A Figura 2(a) apresenta a função de distribuição acumulada complementar da centralidade de cobertura do GVT MOSAR para diferentes valores de Φ passos. Assim como no tempo de cobertura, os primeiros 33000 instantes de tempo são analisados.

Cada valor em uma curva representa a fração dos instantes de tempo do GVT que alcançam a correspondente centralidade de cobertura c em, no mínimo, Φ passos. Por exemplo, com um limite de $\Phi = 2000$ passos, a difusão iniciada em 80% dos instantes de tempo alcança no mínimo 28% dos nós no GVT MOSAR, observando que a máxima cobertura alcançada é de 58%.

Conforme esperado, quanto maior o número de passos Φ , maior a cobertura global da rede. Vale ressaltar, contudo, que em todos os casos, um número muito limitado de instantes de tempo — na verdade, os mais centrais — foram eficientemente superiores na cobertura do GVT MOSAR, com um número limitado de passos, quando a difusão foi iniciada neles. Na Figura 2(b), são comparados o desempenho dos 10 melhores instantes de tempo — os mais centrais — de acordo com a métrica de centralidade de cobertura, e o desempenho dos 10 instantes de tempo escolhidos aleatoriamente para iniciar uma difusão de informação. De fato, iniciar uma difusão em qualquer um dos 10 melhores instantes de tempo alcançaria, no mínimo, 56% da cobertura da rede, enquanto que nos

instantes de tempo escolhidos aleatoriamente, o processo de difusão alcançaria não mais que 33% dos nós do GVT MOSAR, considerando um limite de $\Phi = 2000$ passos.



(a) Distribuição da Centralidade de Cobertura

(b) 10 Melhores x Escolha Aleatória

Figura 2. Função de Distribuição Acumulada da Centralidade de Cobertura na rede MOSAR

5. Conclusões

Na dissertação, foi introduzido o conceito de Centralidade de Tempo em GVTs. A centralidade de tempo avalia a importância relativa de um instante de tempo dentro do contexto de um GVT. Além disso, foram apresentadas duas métricas de centralidade de tempo, baseadas em processos de difusão, e ambas foram avaliadas usando um conjunto de dados real, que representa uma rede dinâmica de contato interpessoal. Os resultados mostram que iniciar uma difusão em um instante de tempo mais central, de acordo com nossas métricas, pode resultar em um processo de difusão mais rápido e abrangente.

Como trabalho futuro, é cogitado o desenvolvimento de modelos de previsão baseados na centralidade de tempo para sistemas complexos que podem ser representados por GVTs. Para isso, é necessário elaborar estratégias que possibilitem a identificação de instantes de tempo mais centrais, baseadas em evidências no passado recente ou no presente da evolução do GVT. Dessa forma, é possível indicar prováveis instantes de tempo centrais à medida em que eles surgirem.

Referências

- Lucet, J. C., Laouenan, C., Chelius, G., Veziris, N., Lepelletier, D., Friggeri, A., Abiteboul, D., Bouvet, E., Mentre, F., and Fleury, E. (2012). Electronic Sensors for Assessing Interactions between Healthcare Workers and Patients under Airborne Precautions. *PLoS ONE*, 7(5):e37893.
- Spasojevic, N., Li, Z., Rao, A., and Bhattacharyya, P. (2015). When-to-post on social networks. In *Prof. of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM KDD)*.
- Wehmuth, K., Ziviani, A., and Fleury, E. (2014). A Unifying Model for Representing Time-Varying Graphs. *ArXiv e-prints*.

Synthesis of Code Anomalies: Revealing Design Problems in the Source Code

Willian N. Oizumi¹, Alessandro F. Garcia¹ (Advisor)

¹Informatics Department – Pontifical Catholic University of Rio de Janeiro (PUC-Rio)

{woizumi, afgarcia}@inf.puc-rio.br

Abstract. *Design problems affect most software projects and make their maintenance expensive and impeditive. Thus, the identification of potential design problems in the source code – which is very often the only available and up-to-date artifact in a project – becomes essential in long-living software systems. This identification task is challenging as the reification of design problems in the source code tend to be scattered through several code elements. However, state-of-the-art techniques do not provide enough information to effectively help developers in this task. In this work, we address this challenge by proposing a new technique to support developers in revealing design problems. This technique synthesizes information about potential design problems, which are materialized in the implementation under the form of syntactic and semantic anomaly agglomerations. Our evaluation shows that the proposed synthesis technique helps to reveal more than 1200 design problems across 7 industry-strength systems, with a median precision of 71% and a median recall of 78%. The relevance of our work has been widely recognized by the software engineering community through 2 awards and 7 publications in international and national venues.*

1. Introduction

Design problems are caused by the violation of key design principles or rules [Oizumi et al. 2016]. Software systems suffer from design problems, introduced either during original development or during evolution. *Fat Interface* and *Unwanted Dependency* [Garcia et al. 2009] are examples of design problems. Software systems – like Linux [Schach et al. 2002] and Mozilla Firefox [Godfrey and Lee 2000] – have had to be fundamentally reengineered or have been discontinued when design problems were allowed to persist in the code and to be compounded by other design problems introduced later [Oizumi et al. 2016]. Therefore, even presenting different degrees of severity, design problems should be identified and removed in early versions of a program.

Design problems are introduced and allowed to remain in a system because their localization in the source code is difficult. As design documentation is often informal or nonexistent, code anomalies – popularly known as code smells [Fowler 1999] – are the mechanisms used to locate possible design problems in the source code. However, each code anomaly represents only a partial indicator of a design problem [Oizumi et al. 2015]. Examples of typical code anomalies are *Long Method* and *God Class*. Even though each code anomaly can provide some hint to developers, it alone might not suffice to indicate the presence of a design problem. In fact, isolated anomalies are often irrelevant to software design [Oizumi et al. 2014a]. Each design problem is rarely localized in a single anomalous code element; instead, it is scattered into different anomalous code elements

of the implementation [Oizumi et al. 2016, Oizumi et al. 2015, Oizumi et al. 2014a]. Therefore, developers often have to analyze multiple code anomalies in order to locate and understand a single design problem.

Unfortunately, there is very limited knowledge about how design problems manifest in the source code. This occurs because the relation between design problems and their counterpart code anomalies is often complex. There is little to no understanding of which relationships between code anomalies are frequent indicators of design problems in complex software systems. There is a recent growing interest in conceptually characterizing interactions between code anomalies. However, the relation of code anomalies and design problems is rarely investigated. Empirical studies only address how individual occurrences of code anomalies emerge during software evolution or affect quality attributes. They do not analyze how individual anomalies and their relationships in the code might help developers to spot design problems. As a result, conventional techniques for anomaly detection are unable to effectively reveal design problems in the source code [Oizumi et al. 2015].

In this context, this work addresses the aforementioned gap in the literature, proposing and evaluating a technique for the **Synthesis of Code Anomalies (SCA, for short)** [Oizumi et al. 2014b]. In order to reveal design problems, SCA searches for coherent groups of inter-related code anomalies – the so called anomaly agglomerations. With the aim of assessing the effectiveness of SCA, we conducted two evaluations: (1) a multi-case study involving 7 systems with different sizes [Oizumi et al. 2016, Oizumi et al. 2015], and (2) a controlled experiment and interviews with several professional developers [Oizumi 2015]. Both evaluations confirmed that SCA is at least twice better than conventional techniques to reveal critical design problems in a software system. Moreover, we found that our algorithm for synthesizing semantic agglomerations helped to locate design problems with a precision higher than 80%.

2. Synthesis of Code Anomalies: An Overview of the Proposed Technique

Existing techniques (e.g. [Lanza and Marinescu 2006]) for detecting design problems in the source code are based on the assumption that single measures or smells help to locate design problems. Instead, SCA reveals design problems through the systematic search and summarization of information about code-anomaly agglomerations [Oizumi et al. 2014b]. We present below a summarized description of the main steps of SCA. Refer to Chapter 3 of [Oizumi 2015] for a detailed description.

Detection of Individual Code Anomalies. Using conventional algorithms [Lanza and Marinescu 2006], SCA analyzes the source code of the program aiming at detecting instances of code anomalies [Fowler 1999][Lanza and Marinescu 2006]. SCA covers the types of code anomalies documented in [Fowler 1999].

Search for Agglomerations. After the detection of code anomalies is completed, SCA explores information about code anomalies and relationships between code elements [Oizumi et al. 2015] to search for anomaly agglomerations. The search performed by SCA explores syntactic and semantic relationships to identify anomaly agglomerations in a program [Oizumi et al. 2014b]. Method calls and design concerns involving anomalous elements in the program are examples of syntactic and semantic relationships, respectively. A concrete example of agglomeration is provided in Section 3.

In this work, we proposed an evaluated 6 categories of agglomeration based on semantic and syntactic relationships: (1) concern-based, (2) cross-boundary, (3) intra-boundary, (4) hierarchical, (5) intra-method and (6) intra-class. The first category encompasses agglomerations formed through design concerns. Agglomerations composed by relationships that crosses the boundaries between design components fall in the second category. The third category contains agglomerations that occur in the internal structure of design components. Agglomerations based on hierarchical relationships (inheritance and interface implementation) fall in the fourth category. Finally, the fifth and sixth categories consider agglomerations occurring in common methods and common classes, respectively. Each category of agglomeration provides information that other categories may not provide. In other words, each of them presents a distinct perspective to analyze anomaly agglomerations [Oizumi et al. 2014a, Oizumi et al. 2016]. The identification of each category of relationship is realized through a search strategy. To better understand SCA search strategies, consider the concern-based category. The search strategy for this category consists of grouping into an agglomeration anomalous elements of different components, which are related to the same design concern (e.g. Concurrency).

Summarization of Relevant Information. To provide valuable information about the agglomerations found, SCA synthesizes relevant information about each agglomeration. This includes a textual description of the anomalies in each agglomeration, a list of code elements surrounding each agglomeration, and historical information about each agglomeration. The aforementioned information is intended to support developers in the process of analyzing, understanding and removing design problems revealed by agglomerations.

3. Evaluation of SCA

Our initial studies indicated that several isolated anomalies are often not related to design problems [Oizumi et al. 2014a, Oizumi et al. 2014b]. This finding motivated us to perform the evaluation of SCA based on two research questions:

- RQ1.** Is SCA an accurate technique to support the location of design problems?
- RQ2.** What is the most useful category of relationships to locate design problems?

The objective of RQ1 is to investigate whether SCA is better than a conventional technique to identify design problems. The conventional technique consist of state-of-the-art algorithms for code anomaly detection [Lanza and Marinescu 2006]. To answer this question, we performed two evaluations regarding the proportion of false positives and false negatives of each technique. RQ2 is aimed at investigating which of the proposed categories of relationships is the best to reveal design problems. These questions were answered based on the use of quantitative and qualitative methods. Procedures and results of our studies are described below.

Methodology. In order to answer research questions RQ1 and RQ2, we conducted two assessments: a multi-case study in the context of seven industry-strength systems, and a controlled experiment and interviews with eight professional developers. In the first assessment, we investigated which technique presented the highest precision and recall in the identification of design problems. As agglomerations involves more code elements than single anomalies, we also analyzed the proportion of elements in agglomerations that are related to design problems [Oizumi et al. 2015]. Finally, we compared the accuracy

of anomaly agglomeration categories to understand which is the best one to reveal design problems. To perform our analysis, original developers helped us to identify instances of 8 design problems in the 7 target systems. Details about each design problem are provided in Section 2.2 of [Oizumi 2015].

In the second assessment, we compared the use of SCA and of the conventional technique by professional developers through a controlled experiment. In this experiment, professionals were asked to identify design problems in two different systems using both techniques. This experiment helped us to analyze, based on the developers' perception, the benefits of SCA in actual tasks on the identification of design problems. A thorough description of both evaluations is presented in Chapters 4 and 5 of [Oizumi 2015].

SCA Outperformed the Conventional Technique. Regarding RQ1, SCA was at least twice better than the conventional technique to indicate the presence of design problems, presenting a median precision of 71% and a median recall of 78%. In four systems, SCA was five times more accurate than the conventional technique. Using SCA, developers could discard near by 4000 isolated code anomalies, which would be irrelevant to reveal design problems. Finally, we observed that developers make several mistakes as they are exposed to various false positives when using a conventional technique.

Example of Critical Design Problem Detected with SCA. Our findings in both studies also confirmed that SCA was accurate to reveal the most critical design problems in the analyzed systems. Consider the hierarchy of classes presented in Figure 1. This figure shows a snapshot of the *Versioner* class hierarchy in the OODT system [Mattmann et al. 2006]. The *Versioner* hierarchy is responsible for managing and storing versions of different *Product* types using alternative storage strategies. All classes in the *Versioner* hierarchy have to implement the *createDataStoreReferences* method. This method has two parameters: a *Product* instance and a *Metadata* instance. As there are no sub-classes for each type of *Product*, each *createDataStoreReferences* implementation has to decide if it is handling the correct *Product* type (e.g., the *MetadataBasedFileVersioner* must only process “flat” products). Consequently, the *Product* type handled by each *Versioner* implementation cannot be discovered from the *createDataStoreReferences* interface. Instead, this can only be discovered by analyzing details of each *createDataStoreReferences* implementation. Hence, the developer can conclude the *Versioner* implementations are affected by the Fat Interface design problem. This design problem occurs in interfaces that expose multiple functionalities through a general interface. This problem can only be identified through a careful analysis of the OODT source code.

Only through the use of SCA, developers were able to identify this critical design problem. All implementations of *Versioner* (*SingleFileBasicVersioner*, *BasicVersioner*, *DateTimeVersioner* and *MetadataBasedFileVersioner*) are affected by instances of the Feature Envy anomaly – a method that calls more methods of a single external class than of its own class. All these anomaly instances occur in classes that implement the same interface – which is the *Versioner* interface. This means the anomalous classes are inter-related through hierarchical relationships. SCA explores such relationships to detect this category of anomaly agglomeration. This agglomeration provides fundamental information to reveal the Fat Interface: anomalous counter-part elements of the design problem in the source code, and the relationships between the anomalous elements. We observed in the experiment that, analyzing this agglomeration, a developer did not waste

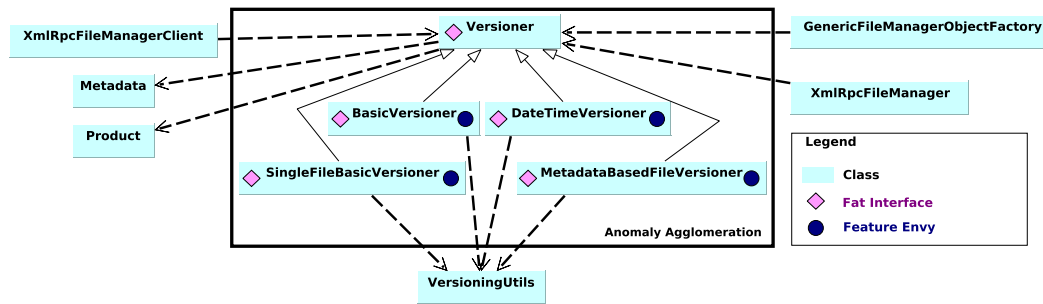


Figure 1. Example of agglomeration in the OODT system.

time in analyzing irrelevant non-agglomerated anomalies.

Other features of SCA were also responsible for achieving high accuracy. SCA provides information about the relationships between the agglomeration and its surrounding code elements. In the example of Figure 1, the code elements *Metadata*, *Product*, *VersioningUtils*, *XmlRpcFileManager*, *XmlRpcFileManagerClient* and *GenericFileManagerObjectFactory* are not anomalous. However, they are surrounding the agglomeration, as they are related to anomalous classes in the agglomeration. This information is important for a developer to recognize the extent and relevance of the design problem. Moreover, if a developer eventually removes the design problem, some surrounding elements would surely be affected.

Accuracy of Agglomeration Categories. Regarding RQ2, intra-component, cross-component and hierarchical categories cannot be considered very strong indicators of design problems. Their statistical significance was not high. Nevertheless, considering all the analyzed systems, approximately 50% instances of these categories of agglomerations were related to design problems, with a median recall of 55%. This accuracy is much higher than the accuracy of individual code anomalies. Design problems were often much more precisely indicated by semantic anomaly agglomerations. In general, the accuracy was approximately 80% when considering all the design problems and systems analyzed in our study. This result was also confirmed by our controlled experiment with professionals. Moreover, design problems identified with the help of such semantic anomaly agglomerations were considered the most complex and hard to spot. Considering our sample of systems, semantic agglomerations presented a median recall of 13%. Our analysis revealed that this low recall occurred due to inaccurate information provided by original architects, which can be improved by using automatically detected design concerns. When provided with more precise information about design concerns, semantic agglomerations revealed 39.58% more design problems [Oizumi et al. 2016].

4. Concluding Remarks

We designed and proposed SCA, a technique for the synthesis of code anomalies and support the location of design problems in the source code. The proposed technique was evaluated in the context of two empirical studies. Both studies provided evidence that SCA, in fact, significantly outperforms existing state-of-the-art techniques. We also created *Organic* [Oizumi and Garcia 2015], an open-source project for the practical use and evaluation of SCA. **Relevance.** The relevance of our idea and concrete solution

have been recognized by the software engineering community. We published our results in three international conferences [Oizumi et al. 2016] (**Qualis A1**), [Vidal et al. 2015], and [Vidal et al. 2016]; a symposium [Oizumi et al. 2014a]; an international journal [Oizumi et al. 2015]; and a workshop [Oizumi et al. 2014b, Albuquerque et al. 2014]. **Awards.** Our work was awarded twice at the Congresso Brasileiro de Software (CB-Soft 2014): **best paper** of WMod [Oizumi et al. 2014b] and **third best paper** of SBES [Oizumi et al. 2014a], which is the most traditional Brazilian symposium on Software Engineering. As a result, we were invited to submit an extended version of our work to the Journal of Software Engineering Research and Development (JSERD) [Oizumi et al. 2015]. **Acknowledgements.** This research was sponsored by PUC-Rio, FAPERJ and CAPES. We thank Nenad Medvidovic, Arndt von Staa and Leonardo da Silva Sousa for their technical contributions.

References

- Albuquerque, D., Garcia, A., Oliveira, R., and Oizumi, W. (2014). Deteccao interativa de anomalias de codigo: Um estudo experimental. In *WMod'14 at Cbsoft*.
- Fowler, M. (1999). *Refactoring: Improving the Design of Existing Code*. Addison-Wesley Professional, Boston.
- Garcia, J., Popescu, D., Edwards, G., and Medvidovic, N. (2009). Identifying architectural bad smells. In *13th European Conf. on Software Maintenance and Reengineering*.
- Godfrey, M. and Lee, E. (2000). Secrets from the monster: Extracting Mozilla's software architecture. In *CoSET-00*, pages 15–23.
- Lanza, M. and Marinescu, R. (2006). *Object-Oriented Metrics in Practice*. Springer.
- Mattmann, C. et al. (2006). A software architecture-based framework for highly distributed and data intensive scientific applications. In *Proceedings of the 28th ICSE*.
- Oizumi, W. (2015). Synthesis of code anomalies: Revealing design problems in the source code. MSc dissertation. In *Pontifical Catholic University, Informatics Department*.
- Oizumi, W. and Garcia, A. (2015). Organic. <http://wnoizumi.github.io/organic/>.
- Oizumi, W., Garcia, A., Colanzi, T., Staa, A., and Ferreira, M. (2015). On the relationship of code-anomaly agglomerations and architectural problems. *JSERD, Springer*.
- Oizumi, W., Garcia, A., et al. (2014a). When code-anomaly agglomerations represent architectural problems? An exploratory study. In *SBES'14; Maceio, Brazil*.
- Oizumi, W., Garcia, A., et al. (2016). Code anomalies flock together: Exploring code anomaly agglomerations for locating design problems. In *ICSE'16; Austin, USA*.
- Oizumi, W., Garcia, A., Sousa, L., Albuquerque, D., and Cedrim, D. (2014b). Towards the synthesis of architecturally-relevant code anomalies. In *WMod'14 at CBSOft*.
- Schach, S. et al. (2002). Maintainability of the linux kernel. *Software, IEE Proceedings*.
- Vidal, S., Guimaraes, E., Oizumi, W., et al. (2016). On the criteria for prioritizing code anomalies to identify architectural problems. In *SAC'16; Pisa, Italy*.
- Vidal, S., Vazquez, H., Diaz-Pace, J. A., Marcos, C., Garcia, A., and Oizumi, W. (2015). Jspirit: a flexible tool for the analysis of code smells. In *34th SCCC*.

Partial Least Squares for Face Hashing

Cássio E. dos Santos Jr., William Robson Schwartz

Department of Computer Science - Universidade Federal de Minas Gerais
Belo Horizonte - MG - Brasil

{cass,william}@dcc.ufmg.br

Introduction

In this work, we focus on the face identification task. Specifically, the main goal is to provide a face identification approach scalable to galleries consisting of numerous subjects and on which common face identification approaches would probably fail on responding in low computational time. There are several applications for a scalable face identification method: surveillance scenarios, human-computer interaction and social media. The few aforementioned applications show the importance of performing face identification fastly and, in fact, several works in the literature have been developed in the past years motivated by these same types of applications (surveillance, forensics, human-computer interaction, and social media). However, most of the works focus on developing fast methods to evaluate one test face and a single subject enrolled in the gallery. These methods usually develop low computational cost feature descriptors for face images that are discriminative and with low memory footprint enough to process several images per second. Note that these methods still depend on evaluating all subjects in the face gallery. Therefore, if the number of subjects in the gallery increases significantly, these methods will not be able to respond fastly and new methods shall be developed to scale the face identification to this larger gallery.

Face identification methods usually consists of a face representation or description in the feature vector where mathematical models can be applied to determine the face identity. In this case, it is used one model to determine each identity in the face gallery, therefore, being necessary a number of models equal to the gallery size. Note that the parameters in each model are learned using samples for each subject in the face gallery and every model must be evaluated to correctly identify a test sample. In this work, we propose a method to reduce the number of models evaluated in the face identification by eliminating identities that are somewhat clearly not the identity in the test sample. Figure 1 illustrates the common face identification pipeline employed in practice and the main component tackled in this work.

There is an extensive literature of works regarding large-scale image retrieval that could be employed in face identification. However, most of these works focus on returning a list containing images from the dataset that are similar to the test image. Although reasonable to recover images in large datasets, such approaches are not suitable to apply directly to the face identification task. The models from subjects in the face gallery should optimally be described regarding the discriminative features related to each subject identity, which might consume less memory, specially if several samples per subject are available, and less computational time since only discriminative features are evaluated to determine the face identity.

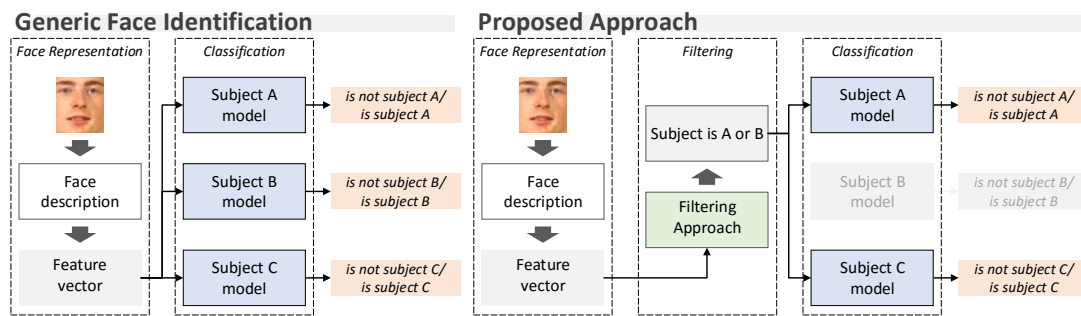


Figure 1: Common face identification pipeline and the proposed pipeline with the filtering approach which is used to reduce the number of evaluations in the classification step with low computational cost. The filtering approach is the main contribution in this work and it is tailored considering recent advances in large-scale image retrieval and face identification based on PLS.

The proposed approach is inspired by the family of methods regarded as *locality-sensitive hashing* (LSH), which are the most popular large-scale image retrieval method in the literature, and the *partial least squares* (PLS), which has been explored intensively in numerous past works regarding face recognition. We call the proposed approach *PLS for hashing*, abbreviated to PLSH and ePLSH in its extension. The main goal in LSH is to approximate the representation of samples in the high dimensional space using a small binary representation where the search can be implemented efficiently employing a hash structure to approximate near-identical binary representations. The idea in LSH is to generate random hash functions to map the feature descriptor in the high dimensional representation to bits in the binary representation.

In the PLSH approach, the random projection in the aforementioned example is replaced by a PLS regression, which provides discriminability among subjects in the face gallery and allow us to employ a combination of different feature descriptors to generate a robust description of the face image. PLSH is able to provide significant improvement over the brute-force approach (evaluating all subjects in the gallery) and compared to other approaches in the literature. Furthermore, since the evaluation of hash functions in PLSH requires a dot product between the feature and regression vectors, additional speedup can be achieved by employing feature selection methods, resulting on the extended version of PLSH (ePLSH).

The following contributions are presented in this work. (i) A fast approach for face identification that support a combination of several feature descriptors and high dimensional feature vectors. (ii) The proposed approach presents at least comparable performance with other methods in the literature and up to 58 times faster when enough samples per subject are available for train. (iii) Extensive discussion and experimentation regarding alternative implementations that may guide future development in scalable face identification methods. (iv) The proposed approach is easy to implement and to deploy in practice since only two trade-off parameters need to be estimated.

This work resulted in the following publications in which the former presents PLSH and the latter presents ePLSH. Jr, C. E. S., Kijak, E., Gravier, G., and Schwartz, W. R. (2015). Learning to hash faces using large feature vectors. In *Content-Based Multimedia Indexing (CBMI), 13th IEEE Workshop on*, pages 1–6, and C E Santos Jr, E Kijak, G. G. W. R. S. (2016). Partial least squares for face hashing. *Elsevier Neurocomputing – Special Issue on Binary Representation Learning in Computer Vision*, pages 1–44.

Experimental setup

We evaluate PLSH and ePLSH in two standard face identification datasets, FERET and FRGCv1. The facial recognition technology (FERET) dataset [Phillips et al. 2000] consists of 1,196 images, one per subject for training, and four test sets designed to evaluate the effects of lightning conditions, facial expression and aging on face identification methods. The test sets are: *fb*, consisting of 1,195 images taken with different facial expressions; *fc*, consisting of 194 images taken in different lightning conditions; *dup1*, consisting of 722 images taken between 1 minute and 1,031 days after the gallery image; *dup2*, is a subset of *dup1* and consists of 234 images taken 18 months after the gallery image. In our experiments, all images were cropped in the face region using annotated coordinates of the face, scaled to 128×128 pixels and normalized using the self-quotient image (SQI) method to remove lightning effects [Wang et al. 2004].

The face recognition grand challenge dataset (FRGC) [Phillips et al. 2005] consists of 275 subjects and samples that include 3D models of the face and 2D images taken with different illumination conditions and facial expressions. We follow the same protocol described by Yuan et al. [Yuan et al. 2005], which considers only 2D images and consists in randomly selecting different percentages of samples from each subject to compose the face gallery and using the remaining samples to test. The process is repeated five times and the mean and standard deviation of the rank-1 recognition rate and speedup (considering the brute-force approach) are reported. The samples were cropped in the facial region, resulting in size 138×160 pixels, and scaled to 128×128 pixels.

All experiments regarding parameter validation were performed on the FERET dataset, since it is the dataset with the largest number of subjects (1,196 in total). FERET consists of four test sets and we use *dup2* to validate parameters since it is considered the hardest of the dataset.

We consider four feature descriptors in this work, CLBP [Ahonen et al. 2006], Gabor filters [Randen and Husoy 1999], HOG [Dalal and Triggs 2005] and SIFT [Lowe 2004], which mainly captures information about texture and shape of the face image. This set of features was chosen because they present slightly better results in the face identification and indexing compared to the previous works [Schwartz et al. 2012, Jr et al. 2015].

The error rate of the pipeline as described in Figure 1 results from errors induced by the filter approach (fail to return identity of test sample in the candidate list) and by the face identification approach (fail to identify correctly the subject in the candidate list). Therefore, to assess the performance of the filter approach alone, we provide results considering the *maximum achievable recognition rate* (MARR) [Jr et al. 2015], which is calculated considering that a perfect face identification method is employed for different percentages of candidates visited in the list. Note that the MARR value is the upper bound for the recognition rate achieved by the filter and face identification pipeline.

Experimental results

Results regarding MARR and rank-1 recognition rate for PLSH in all test sets from the FERET dataset are presented in Figures 2(a) and 2(b). For the test sets *fb* and *fc*, about 1% of subjects in the candidates list is enough to achieve more than 95% of the rank-1 recognition rate of the brute-force approach (presented in the legend of Figure 2(b) for each test

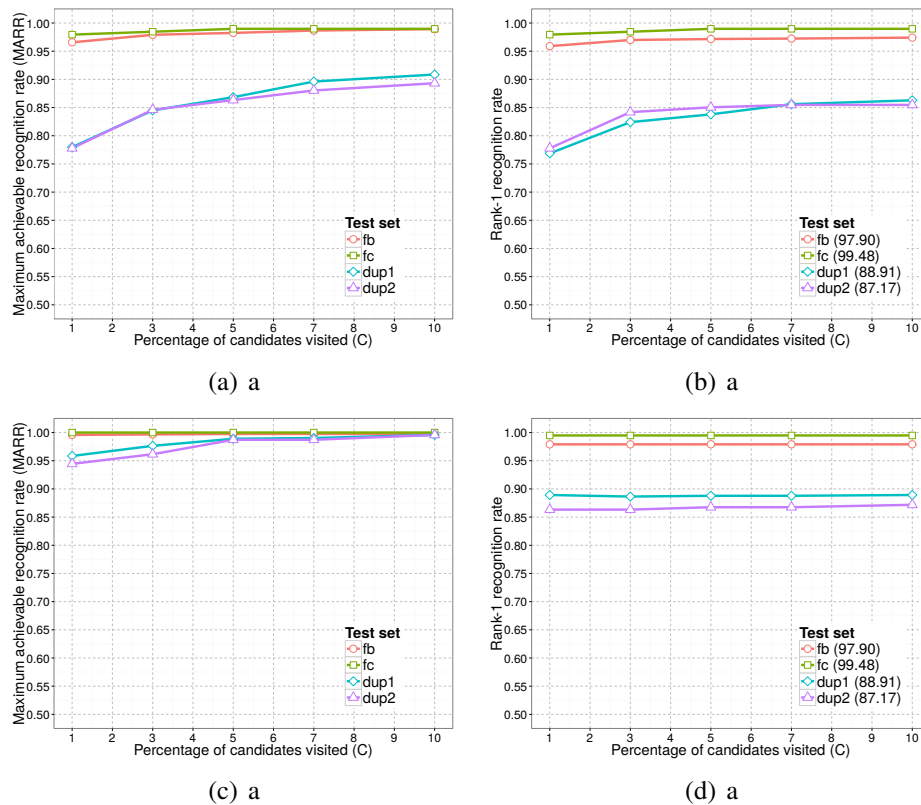


Figure 2: Results on the FERET dataset. (a) PLSH MARR curves, (b) PLSH rank-1 recognition rate, (c) ePLSH MARR curves and (d) ePLSH rank-1 recognition rate. Number in parenthesis indicate rank-1 recognition rate for the brute force approach.

set). However, for the test sets *dup1* and *dup2*, about 5% of subjects in the candidate list ensured at least 95% of the brute-force rank-1 recognition rate. The theoretical speedup in the worst case can be calculated considering the 150 PLSH hash function evaluations and the 5% of the gallery size, which consists of 60 PLS projections. In this case, the number of PLS projections would be 210 compared to the 1,196 projections necessary in the brute-force approach, which would still result in a 5.6 times speedup.

Results from ePLSH are presented in Figures 2(c) and 2(d). Using only 1% of subjects in the candidate list, it is possible to recover all subjects in the rank-1 recognition rate from brute-force approach for all four test sets. In this case, the rank-1 recognition rate from the ePLSH pipeline is the same as the brute-force approach, but with reduction to 1% of the subjects evaluated in the identification. Considering that the cost to evaluate all hash models in ePLSH is about the same as in PLSH, the theoretical speedup is 7.38 times compared to the brute-force approach in the worst case.

Results from the FRGC dataset for PLSH and ePLSH are presented in Table 1 along with results from three other methods as presented in the literature. The three methods are the cascade of rejection classifiers (CRC) from [Yuan et al. 2005], the PLS-based search tree [Schwartz et al. 2012], and our previous published work [Jr et al. 2015], which consists of PLSH with the combination of HOG, Gabor filter and LBP feature descriptors. For PLSH and ePLSH, we vary the number of hash models and the maximum percentage of subjects visited in the candidate list and we present the results with rank-1 recognition rate close to 0.95 and higher speedups. In this way, it is possible to compare

		% of samples for train				
		90%	79%	68%	57%	35%
CRC [Yuan et al. 2005]	Speedup	1.58×	1.58×	1.60×	2.38×	3.35×
	Rank-1 rec. rate	80.5%	77.7%	75.7%	71.3%	58.0%
Tree-based [Schwartz et al. 2012]	Speedup	3.68×	3.64×	3.73×	3.72×	3.80×
	Rank-1 rec. rate	94.3%	94.9%	94.3%	94.46%	94.46%
PLSH	Speedup	18.24×	8.61×	6.95×	3.96×	3.49×
	Rank-1 rec. rate	95.31%	95.31%	93.60%	94.67%	94.60%
ePLSH	Speedup	233.61×	98.93×	45.42×	22.29×	14.21×
	Rank-1 rec. rate	96.03%	95.02%	95.98%	94.67%	94.44%

Table 1: Comparison between the proposed approach and other approaches in the literature. The highest speedups are shown in bold. The full table with more information can be found in page 58 of the dissertation.

directly the maximum speedup achievable when using PLSH and ePLSH compared to the other approaches, which also provide rank-1 recognition rate close to 0.95.

According to Table 1, the speedup for PLSH and ePLSH decreases considerable as the number of samples per subject available for train reduce. The reason for that is the increase in the number of hash models and the maximum number of subjects visited in the candidate list to guarantee at least 0.95 rank-1 recognition rate. Even with reduced speedups considering 35% of samples available for train, ePLSH provides significant improvement over the speedup achieved by the tree-based approach (3.6 times faster), while PLSH provides competitive speedup.

The speedup provided by PLSH and ePLSH compared to the tree-based approach is noticed with 90% of the samples available for train, where PLSH is about 5 times faster than the tree-based approach while ePLSH is about 13 times faster than PLSH. Finally, in the worse case, ePLSH provides at least 14 times speedup considering the brute-force approach in the setup with 200 hash models and 10% of subjects in the candidate list.

Conclusions

In this work, we proposed and evaluated PLSH and its extension ePLSH for face indexing. PLSH is inspired by the well-known locality-sensitive hashing for large-scale image retrieval and PLS for face identification, which provides fast and robust results for face indexing. Additional gain in speedup was achieved with the ePLSH, a method that employs PLS-based feature selection to reduce the computational cost to evaluate hash functions, enabling a large amount of additional hash functions to be employed and raising the indexing precision. We evaluated several parameters and alternative implementations of PLSH in the hope that they will be useful for future face indexing development. The experiments were conducted on two face identification standard datasets, FERET and FRGCv1, with 1, 196 and 275 subjects, respectively. Although these datasets do not provide enough number of subjects for a proper evaluation regarding scalability to large galleries, PLSH and ePLSH still provide significant improvement in speedup compared to other scalable face identification approaches in the literature.

The conclusions and considerations regarding PLSH and ePLSH are the following: (i) they support for high dimensional feature vectors, allowing different complementary feature descriptors to be employed to increase the robustness of the face indexing; (ii) they are easy to implement and deploy in practice since the only parameters needed to be set are the number of hash models and subjects in the candidate list. (iii) they do not provide good performances when the number of samples per subject is reduced and (iv)

incremental enrollment of subjects in the framework requires re-training of the hash models, which may be prohibitive to perform in practice, specially for ePLSH which demands considerable more hash models.

Acknowledgments

The authors would like to thank the Brazilian National Research Council – CNPq (Grant #477457/2013-4), the Minas Gerais Research Foundation – FAPEMIG (Grant APQ-00567-14) and the Coordination for the Improvement of Higher Education Personnel – CAPES (DeepEyes Project).

References

- Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041.
- C E Santos Jr, E Kijak, G. G. W. R. S. (2016). Partial least squares for face hashing. *Elsevier Neurocomputing – Special Issue on Binary Representation Learning in Computer Vision*, pages 1–44.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society Conference on*, volume 1, pages 886–893.
- Jr, C. E. S., Kijak, E., Gravier, G., and Schwartz, W. R. (2015). Learning to hash faces using large feature vectors. In *Content-Based Multimedia Indexing (CBMI), 13th IEEE Workshop on*, pages 1–6.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.
- Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J., and Worek, W. (2005). Overview of the face recognition grand challenge. In *Computer vision and pattern recognition (CVPR). IEEE Conference on*, volume 1, pages 947–954.
- Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence*, 22(10):1090–1104.
- Randen, T. and Husoy, J. H. (1999). Filtering for texture classification: A comparative study. *Pattern Analysis and Machine Intelligence*, 21(4):291–310.
- Schwartz, W. R., Guo, H., Choi, J., and Davis, L. S. (2012). Face identification using large feature sets. *Image Processing, IEEE Transactions on*, 21(4):2245–2255.
- Wang, H., Li, S. Z., and Wang, Y. (2004). Face recognition under varying lighting conditions using self quotient image. In *Automatic Face and Gesture Recognition*, pages 819–824.
- Yuan, Q., Thangali, A., and Sclaroff, S. (2005). Face identification by a cascade of rejection classifiers. In *Computer Vision and Pattern Recognition (CVPR) - Workshops. IEEE Conference on*, pages 152–152.

Energy-aware load balancing in distributed data centers

Rodrigo A. C. da Silva¹, Nelson L. S. da Fonseca¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas - SP - Brazil

rodrigo@lrc.ic.unicamp.br, nelson@ic.unicamp.br

Abstract. *This paper summarizes the dissertation "Energy-aware load balancing in distributed data centers", which proposed two new algorithms for minimizing energy consumption in cloud data centers. Both algorithms consider hierarchical data center network topologies and requests for the allocation of groups of virtual machines (VMs). The Topology-aware Virtual Machine Placement (TAVMP) algorithm deals with the placement of virtual machines in a single data center. It reduces the blocking of requests and yet maintains acceptable levels of energy consumption. The Topology-aware Virtual Machine Selection (TAVMS) algorithm chooses sets of VM groups for migration between different data centers. Its employment leads to relevant overall energy savings.*

Introduction

Data processing and service provisioning have been intensively migrated to the cloud due to the flexibility of obtaining service on demand on a pay per use basis. In infrastructure as a Service (IaaS), users rent virtual machines (VMs) instantiated in data centers, which typically contain thousands of servers. Moreover, cloud providers may own several data centers for the processing of their demand.

A big concern in cloud computing is the energy consumption of data centers. In 2010, for instance, data centers consumed about 1.5% of all the energy consumed in the world. Moreover, energy consumption grows faster than the energy efficiency of servers, storage and network equipment [Koomey 2011]. Therefore, strategies for management of resources need to be developed to decrease the energy consumption. Large energy consumptions affect the environment due to the emissions of pollutant gases and strongly impact the financial costs of service providers. Therefore, decreasing the energy consumption in data centers impacts not only the cloud operation but also Society as a whole.

This paper summarizes the dissertation "Energy-aware load balancing in distributed data centers" [da Silva and da Fonseca 2015b] which proposed a load balancing scheme among the data centers of a provider with the aim of decreasing the energy consumption. Two novel algorithms were proposed. The Topology-aware Virtual Machine Placement (TAVMP) algorithm maps VM groups on a data center in a way to avoid unnecessary activation of servers and network switches. Results show that it is possible to reduce blocking of requests while maintaining acceptable levels of energy consumption. The second algorithm, the Topology-aware Virtual Machine Selection (TAVMS) algorithm chooses sets of groups of virtual machines for potential migration to other data centers when it will decrease the overall energy consumption. Results show that relevant energy savings can be achieved by employing TAVMS. The novelty of these algorithms

relies on the consideration of the data center network topology on resource allocation decisions, the characterization of users' demands as arrivals of groups of virtual machines, and the consideration of realistic inter-VM communications.

The remaining of this paper is organized as follows. Section 2 presents related work. Sections 3 and 4 summarize the first and second algorithms of the dissertation, respectively. Section 5 presents the main results obtained. Section 6 lists the publications of the dissertation. Finally, Section 7 draws conclusions from the dissertation.

Related work

This section briefly describes related work to virtual machine placement and VM migration between data centers. Different approaches have been used to the VM placement problem. The authors of [Beloglazov et al. 2012], for instance, proposed an approach to minimize the energy consumption considering a bin-packing model, using a best fit decreasing algorithm. The solution for the VM placement problem in [Meng et al. 2010] takes into consideration the data center network, proposing a traffic-aware placement algorithm based on the knowledge of the network topology and an estimation of the traffic matrix, matching clusters of virtual machines and servers to minimize VM communications. The TAVMP algorithm differs from proposals in the literature by considering a demand model based on the arrival of groups of virtual machines as well as their pattern of communication. Moreover, the selection of network paths is performed jointly with the decision of mapping VMs onto servers, so that the use of network switches is minimized.

In [Ilyas et al. 2012], the authors proposed an optimization formulation to decide on the migration of a data center workload based on electricity prices in different locations. The algorithm presented in [Mandal et al. 2013] decides to migrate long-lived VMs to sites where renewable energy is available. They show that green energy can replace brown energy with a low increase in the overall consumption. The TAVMS algorithm differs from the literature by the consideration of fine granularity in the decision of workload for migration rather than the whole workload of a data center, allowing fine tuning of energy savings. In addition, it considers the data center network topology in its decisions.

Topology-aware Virtual Machine Placement

The TAVMP algorithm places groups of virtual machines in a data center taking into account the network topology. It aims to place VM groups in small areas of hierarchical data centers so that few network switches are needed to serve the network flows. In order to achieve such goal, the data center is divided in hierarchical areas containing switches and servers. A parameter $j \geq 0$ is used to describe the level of an area in a hierarchical topology. For example, the Fat-tree topology has three levels: racks, PODs and the whole data center, as illustrated in Figure 1.

TAVMP and its notation are presented in Figure 2. The input is the data center network, the group of virtual machines (request) and the highest level of the hierarchy ($j = 2$ in Figure 1). The algorithm divides the topology into small areas, using an auxiliary algorithm to find subgraphs (SUB), Line 1. Recursion is then performed on each area (Line 4), and, when the lowest level is reached, a placement decision is made, specifying which servers will be used to host the virtual machines, as well as the network paths for the inter-VM flows. Another algorithm, named Placement in Current Area (PCA), evaluates

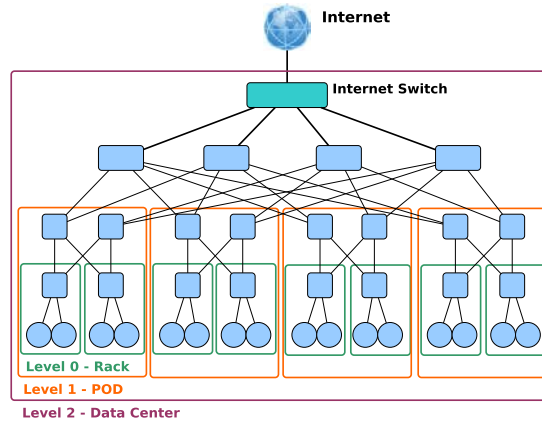


Figure 1. Fat-tree architecture divided into levels. Each circle is a physical server, and each rectangle a switch.

the adequacy of areas for the placement of a group of VMs (Line 7). When the recursion finishes, there can be various choices for the placement of the VMs. The area chosen is the one which involves the minimum energy consumption (Line 5). Furthermore, if none of the areas are suitable for accommodating the group of VMs due to lack of resources, the PCA algorithm is applied to the immediate higher level and, if necessary, to the data center level (Line 7). When TAVMP decides on the final placement of the group of virtual machines, the VMs are instantiated and the status of the network occupancy is updated.

Notation	Description
$\mathcal{D} = (\mathcal{E}, \mathcal{L})$	graph representing a data center area
$\mathcal{E} = \mathcal{H} \cup \mathcal{S}$	\mathcal{H} is the set of servers/hosts and \mathcal{S} is the set of switches
\mathcal{L}	links representing the physical connections
$\mathcal{G} = (\mathcal{V}, \mathcal{F})$	graph representing a VM group
\mathcal{V}	the set of virtual machines
\mathcal{F}	the set of network flows between the VMs

Input: $\mathcal{D} = (\mathcal{E}, \mathcal{L}), \mathcal{G} = (\mathcal{V}, \mathcal{F}), j$

Output: The placement

```

1 subgraphs  $\leftarrow$  SUB( $\mathcal{D}, j$ )
2 if subgraphs  $\neq$   $\emptyset$  then
3    $\forall$  subgraph  $\in$  subgraphs
4     placement  $\leftarrow$ 
       TAVMP(subgraph,  $\mathcal{G}, j - 1$ )
5   Choose placement with smallest power increase
6 if No placement was successful then
7   placement  $\leftarrow$  PCA( $\mathcal{D}, \mathcal{G}$ )
8 return placement

```

Figure 2. The TAVMP algorithm and its notation.

The input of the PCA algorithm is a graph representing an area in the data center topology and information of the demands of the group of VMs. The goal is to choose the servers with resources available to support the VMs computational demands, but minimizing the increase in total energy consumption. Moreover, these servers must be connected by links with sufficient bandwidth to support the communication demands. The PCA algorithm evaluates the placement of one VM at a time, and all the servers are analyzed to find the one on which the energy consumption for that VM workload is minimized. The decision of the paths to accommodate the inter-VM flows is also made in a way that the number of switches visited by a flow is minimized. The PCA algorithm succeeds if all the VMs in the group can be placed on servers, otherwise, the request is rejected.

The input to SUB is the area to be divided and its level, and its output is a set of subgraphs. Figure 1 shows the three levels ($j = 0, 1, 2$) and their subgraphs. Although the SUB algorithm finds subgraphs in the Fat-tree topology, the TAVMP algorithm can accept other topologies as long as SUB is changed to divide the subgraphs according

to the topology. Complete descriptions of the PCA and SUB algorithms can be found in [da Silva and da Fonseca 2015b].

Load balancing in distributed data centers

The TAVMS algorithm performs two tasks: the selection of potential sets of VMs in a data center to be migrated and the negotiation of migration of these workloads to other data centers. The former decision takes into account hierarchical network topologies, while the latter considers the energy effects on the destination data centers (by using the TAVMP algorithm) and the migration downtime. These two tasks are denoted as Selection (SEL) and Negotiation (NEG) algorithms, respectively.

The integration of the algorithms SEL and NEG in TAVMS is explained next. SEL analyzes the PODs in Fat-tree topology discriminating the sets of all virtual machine groups in sequential PODs. This strategy increases the chances of switching off sequential servers and switches, reducing therefore the energy consumption. Its output is used as input to the NEG algorithm, which decides on the migration of the chosen sets by iterating with all the other data centers based on the computation of the energy impact of the migration of each set of VM groups on the overall energy consumption. If the allocation of these VMs will spend less energy in another data center in relation to the source data center, migration is chosen to be performed. The transmission time over the optical infrastructure is calculated, and VMs terminating before this migration time are removed to avoid unnecessary migration. If no migration leads to energy savings, no workload is transferred. The SEL and NEG algorithms are not presented here due to space constraints, their complete description can be found in the dissertation [da Silva and da Fonseca 2015b].

Results

In this section, the main results obtained by the employment of the proposed algorithms are presented. Both algorithms were evaluated by using the Cloudsim Simulator. Energy consumption models were used to account the energy consumption of servers and network switches. The first algorithm was compared to the Power Aware Best Fit Decreasing (PABFD) algorithm from [Beloglazov et al. 2012], while the second was compared to algorithms based on migration of fixed loads, Topology-aware Threshold (TT) and Random Threshold (RT) algorithms, as well as a scenario without migrations (NM).

The parameter K of the Fat-tree topology, which defines the data center size, was varied. Heterogenous servers and VM instances were considered, and the arrival and departure processes of requests were modelled as self-similar series, as suggested in [Han et al. 2013]. Traffic scenarios considered medium (MG) and large (LG) VM groups as well as low (LT), medium (MT) and high (HT) traffic demands. In the second algorithm, four data centers were placed in the National Science Foundation's Network topology. Values for all parameters can be found in the dissertation [da Silva and da Fonseca 2015b].

Energy efficiency and blocking ratio of requests produced by the two algorithms were evaluated. For the first algorithm, energy efficiency was assessed as the total energy spent by a data center divided by the number of accepted VMs. In the second algorithm, it was evaluated as the total energy spent by all data centers. The blocking ratio is the percentage of VM requests which were not placed in relation to the total number of VMs.

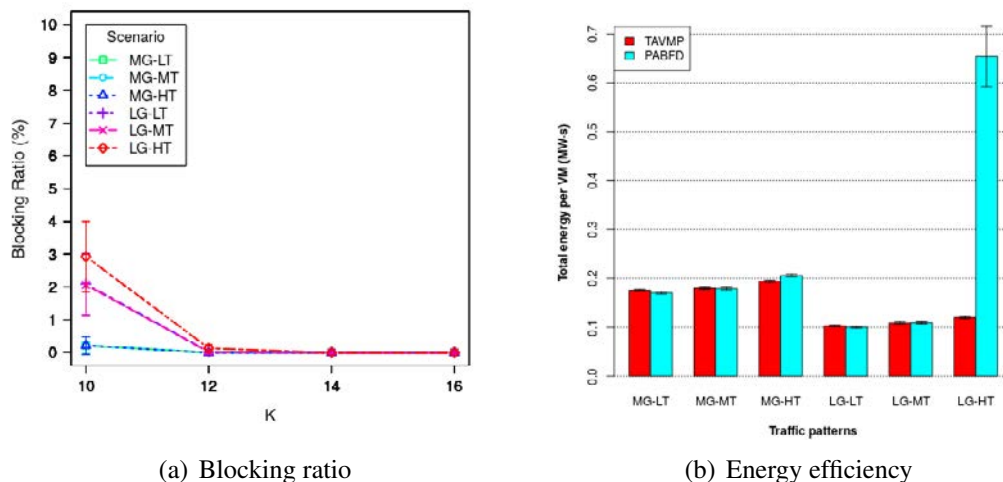


Figure 3. Results for the TAVMP algorithm.

Results for the first algorithm are presented in Figure 3. The blocking ratio of the algorithm is shown in Figure 3(a) for a Hurst parameter value of 0.7. The blocking ratio remains at low levels, even under high traffic scenarios. The blocking ratio produced by the PABFD algorithm for MG-HT remains between 12% and 14%, and for LG-HT it is close to 100%, while the TAVMP algorithm maintains all values under 4%. Figure 3(b) presents the energy efficiency of the TAVMP algorithm for larger data centers ($K = 16$). For low and medium traffic scenarios it remains similar to the PABFD algorithm, while for high traffic demands their VMs use much less energy. This happens due to the VM consolidation in small areas of the topology provided by the TAVMP algorithm, which prevents the formation of network bottlenecks.

Figure 4 presents results for the energy savings obtained by the TAVMS algorithm for data centers with different sizes. TAVMS outperforms the fixed load algorithms in all scenarios. The TT outperforms RT since it takes into account the topology, choosing VMs in close proximity. Moreover, the TAVMS algorithm considers more options of VM groups allocation, considering more precisely the data center network, and consequently saving more energy, ranging from 7.1% to 14% in relation to the NM scenario.

Publications

Results of the dissertation presented in this paper were published in the proceedings of the International Conference on Communications [da Silva and da Fonseca 2015a] (Qualis A2) and in the Journal of Grid Computing [da Silva and da Fonseca 2015c] (impact factor 1.5, Qualis B1). In addition, a third manuscript is under preparation to be submitted to an international journal containing the TAVMS algorithm and associated results.

Conclusion

The dissertation discussed in this paper introduced two algorithms for energy savings in distributed data centers, one for the virtual machine placement and the other to decide on the migration of workload between data centers. The consideration of the network topology and requests as VM groups with traffic demands, which is original consideration

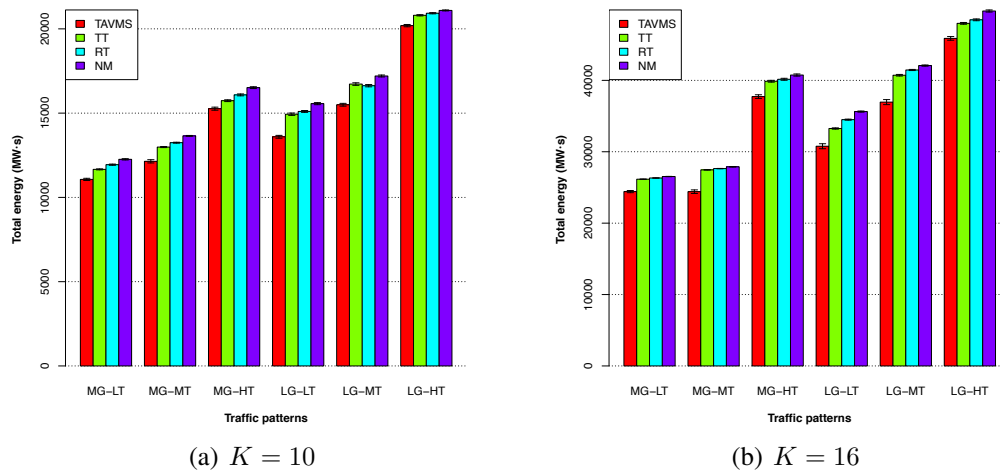


Figure 4. Total energy spent in all data center according to their size.

in this type of algorithms, showed to lead to energy savings while maintaining the quality of the offered services. Future directions of this work include a deeper analysis of traffic patterns of inter-VM communications and also the provisioning of differentiated services.

References

- Beloglazov, A., Abawajy, J., and Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Generation Computer Systems*, 28(5):755 – 768. Special Section: Energy efficiency in large-scale distributed systems.
- da Silva, R. A. C. and da Fonseca, N. L. S. (2015a). Algorithm for the placement of groups of virtual machines in data centers. In *Communications (ICC), 2015 IEEE International Conference on*.
- da Silva, R. A. C. and da Fonseca, N. L. S. (2015b). Energy-aware load balancing in distributed data centers. Master's thesis, Universidade Estadual de Campinas.
- da Silva, R. A. C. and da Fonseca, N. L. S. (2015c). Topology-aware virtual machine placement in data centers. *Journal of Grid Computing*, pages 1–16.
- Han, Y., Chan, J., and Leckie, C. (2013). Analysing virtual machine usage in cloud computing. In *Services (SERVICES), 2013 IEEE Ninth World Congress on*, pages 370–377.
- Ilyas, M., Raza, S., Chen, C.-C., Uzmi, Z., and Chuah, C.-N. (2012). Red-bl: Energy solution for loading data centers. In *INFOCOM, 2012 Proceedings IEEE*, pages 2866–2870.
- Koomey, J. (2011). Growth in data center electricity use 2005 to 2010. *Oakland, CA: Analytics Press. August*.
- Mandal, U., Habib, M., Zhang, S., Mukherjee, B., and Tornatore, M. (2013). Greening the cloud using renewable-energy-aware service migration. *Network, IEEE*, 27(6):36–43.
- Meng, X., Pappas, V., and Zhang, L. (2010). Improving the scalability of data center networks with traffic-aware virtual machine placement. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9.

Verificação de Locutores Independente de Texto: uma Análise de Robustez a Ruído

Hector N. B. Pinheiro, Tsang Ing Ren, George D. da C. Cavalcanti

Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brazil

{hnbp,tir,gdcc}@cin.ufpe.br

Abstract. *This work focuses on the development of text-independent speaker verification systems. The main challenge in the development of these systems comes from the mismatches which may occur in the acquisition of the speech signals. The techniques proposed to reduce the mismatch effects are referred as compensation methods. They may operate in three domains: in the feature extraction process, in the estimation of the speaker models and in the computation of the decision score. This work presents a wide description of the main techniques used in the development of text-independent speaker verification systems and the main feature-, model- and score-based compensation methods. In the experiments, we present a comprehensive comparison between the conventional techniques and the alternatively compensation methods. Furthermore, two compensation methods are proposed: one operates in the model domain and the other in the score domain, which outperformed the main compensation techniques.*

Resumo. *Este trabalho foca no desenvolvimento de sistemas de verificação de locutores independente de texto, cujo principal desafio provém das chamadas incompatibilidades que podem ocorrer na aquisição dos sinais de voz. As técnicas propostas para suavizá-las são chamadas de técnicas de compensação e três são os domínios onde elas podem operar: no processo de extração de características do sinal, na construção dos modelos dos locutores e no cálculo do score final do sistema, utilizado na autenticação. Esse trabalho apresenta uma vasta revisão da literatura do desenvolvimento de sistemas de verificação independentes de texto, das técnicas de compensação de características, modelos e scores. Na fase de experimentação, uma análise comparativa das principais técnicas propostas na literatura é apresentada. Além disso, duas técnicas de compensação são propostas, uma do domínio de modelagem e outra do domínio dos scores, que por suas vezes apresentaram desempenhos superiores às principais técnicas da literatura.*

1. Introdução

O processo de identificação de um determinado indivíduo é realizado milhões de vezes, todos os dias, por organizações dos mais diversos setores. Perguntas como "Quem é esse indivíduo?" ou "É essa pessoa quem ela diz ser?" são realizadas frequentemente por organizações financeiras, sistemas de comércio eletrônico, sistemas de telecomunicações e por instituições governamentais. Para esses tipos de sistemas, a autenticação errônea de

um indivíduo pode trazer consequências desastrosas. Nesse contexto, o desenvolvimento de um sistema de identificação pessoal automático com uma alta acurácia tem se tornado cada vez mais crítica. Identificação biométrica é o processo de identificar um indivíduo a partir de características físicas ou comportamentais. Devido ao fato de muitas dessas características serem únicas, elas são capazes de proporcionar uma maior acurácia ao processo de identificação pessoal do que os métodos tradicionais (utilização de cartões, documentos ou senhas). Alguns exemplos dessas características são: face, impressão digital, íris, assinatura e voz. Reconhecimento de locutores é uma modalidade biométrica que propõe realizar o processo de identificação pessoal a partir das informações presentes unicamente na voz do indivíduo. Este trabalho foca no desenvolvimento de sistemas de verificação de locutores independente de texto. Nesse tipo de modalidade, o sistema deve executar uma autenticação (verificar a identidade alegada pelo usuário) independentemente do conteúdo fonético presente na locução. Sistemas de verificação de locutores

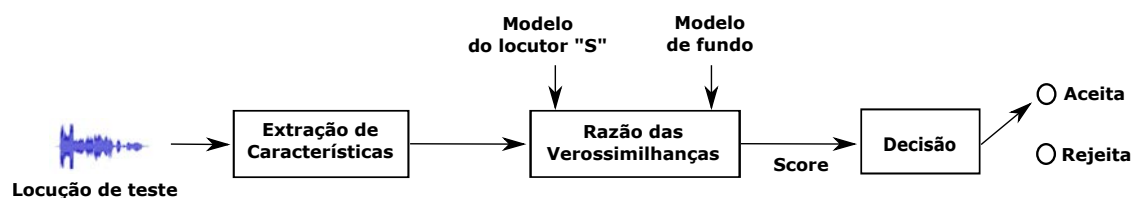


Figura 1. Arquitetura básica de um sistema de verificação de locutores independente de texto.

devem decidir se uma dada locução, X , foi gerada por um locutor específico, S , ou não. Dessa maneira, duas hipóteses são consideradas: $H_0 = \{X \text{ foi produzida por } S\}$ e $H_1 = \{X \text{ não foi produzida por } S\}$. O sistema deve, portanto, ser capaz de modelar tais hipóteses, de modo que seja possível o cálculo da verossimilhança delas ocorrerem. A arquitetura básica de um sistema desse tipo é mostrada na Figura 1. Na fase de cadastramento ocorre a estimação dos modelos associados às hipóteses. O modelo associado à hipótese nula é referenciado como modelo do locutor e é estimado utilizando locuções de treino produzidas por S . Já o chamado modelo de fundo, associado à hipótese alternativa, é estimado utilizando locuções produzidas por diversos locutores, geralmente diferentes de S . Na fase de verificação (ou autenticação), as verossimilhanças da locução de teste com respeito a esses modelos são computadas, de modo que a decisão final seja realizada. O módulo de extração de características possui o objetivo de realizar uma descrição mais compacta do sinal de voz, através de vetores de características. Até então, há um domínio absoluto da utilização de extratores espectrais de tempo curto, em especial, aqueles que extraem os Coeficientes Mel-cepstrais (*Mel-frequency cepstral coefficients*, MFCCs) acompanhados de suas componentes dinâmicas (coeficientes delta de primeira e segunda ordem). A abordagem tradicional para a modelagem dos locutores consiste na técnica GMM-UBM [Reynolds et al. 2000]. Nela, ambas as hipóteses são modeladas através de modelos de misturas Gaussianas (*Gaussian Mixture Models*, GMMs). Primeiramente, um modelo de fundo universal (*Universal Background Model*, UBM) e independente de locutor é estimado através do algoritmo EM (emphExpectation-Maximization) utilizando locuções de diversos locutores. O modelo do locutor S é então produzido pela adaptação das médias das distribuições do UBM utilizando as locuções de S . Essa adaptação é realizada maximizando a probabilidade a posteriori das misturas do modelo (adaptação *Maximum A Posteriori*, MAP). Outra abordagem mais recente é chamada de GMM-SVM

[Campbell et al. 2006], onde cada locução é descrita por um "supervetor GMM", que descreve o GMM resultante da adaptação MAP do UBM utilizando a locução. Um SVM (*Support Vector Machine*) é então treinado para realizar a classificação das locuções, no espaço dos supervetores GMM.

2. Técnicas de compensação de ruído

O grande desafio no desenvolvimento de sistemas de reconhecimento de locutores consiste das chamadas incompatibilidades apresentadas entre os sinais de voz utilizados na estimação dos modelos e aqueles utilizados na autenticação. Muitas são as possíveis fontes de incompatibilidade, que vai desde o tipo de microfone utilizado até o ruído acústico presente no ambiente ou a qualidade do canal de comunicação por onde o áudio é transmitido. Tais fatores impactam na geração do sinal de voz, produzindo distorções que dificultam o reconhecimento. As técnicas desenvolvidas para suavizar tais distorções são chamadas de técnicas de compensação e elas operam sobre três domínios: na extração de características, na modelagem dos locutores e no cálculo dos *scores*. As técnicas de compensação de características se concentram na remoção dos efeitos do ruído sobre o processo de extração das características do sinal e podem operar em diferentes fases da extração. As principais técnicas são: Subtração de Média Cepstral; Normalização de Média Cepstral; Filtragem RASTA; Deformação de Características e Coeficientes de Amplitudes Descorrelacionadas. Além disso, elas podem ser combinadas entre si, se operarem sobre etapas distintas da extração de características. Já as técnicas de compensação de *scores* tentam normalizar os *scores* produzidos por um determinado locutor, de modo a facilitar a escolha de um limiar de classificação. As principais normalizações aplicadas são: Normalização Zero, Normalização de Teste e Normalização *Handset*. Por fim, as técnicas de compensação de modelo propõem estimar modelos de locutores que sejam capazes de lidar com possíveis distorções existentes nas locuções utilizadas para autenticação. As primeiras técnicas surgiram especificamente para tratar as incompatibilidades de canal, mais precisamente aquelas que surgem pela utilização de diferentes tipos de microfones. Porém, os modelos atuais se propõem a lidar com distorções de quaisquer tipos, de uma maneira geral. As principais técnicas de compensação de modelo são: Síntese do Modelo do Locutor e Mapeamento de Características, ambas aplicadas à modelagem GMM-UBM; Projeção dos Atributos Indesejáveis, aplicada à modelagem GMM-SVM e a técnica PUM-GMM, que combina os chamados modelos de união *a posteriori* (*Posterior Union Models* - PUMs) e teoria dos dados ausentes.

Uma descrição detalhada de cada uma dessas técnicas é apresentada em [Pinheiro 2015]. Nesse trabalho, tais técnicas são avaliadas e comparadas entre si, apresentando assim um panorama da eficácia que as compensações proporcionam em conjunto às modelagens GMM-UBM e GMM-SVM. Além disso, duas técnicas de compensação são propostas nesse trabalho. A primeira delas é uma técnica de compensação de *scores* que realiza a normalização dos *scores* através da distribuição normal acumulada dos *scores* produzidos pelo locutor correspondente à alegação. A segunda delas, do domínio dos modelos, propõe uma modelagem baseada no PUM-GMM [Ming et al. 2007]. A formulação apresentada pelos autores da técnica não é completamente adequada para verificação de locutores independente de texto e considera uma abordagem *a posteriori* de uma quantidade comumente elevada de modelos de locutores (diferentes do locutor associada à alegação) para a formação do modelo de união necessário para a escolha do conjunto de

características adequado para a autenticação (seguindo a teoria dos dados ausentes). A técnica proposta apresenta uma formulação mais apropriada para a tarefa, combinando os dois conceitos utilizados pelos autores com um tipo de modelagem utilizando modelos de fundo para a formação do modelo de união. As duas técnicas propostas foram também comparadas às técnicas de compensação presentes na literatura e se mostraram bastante eficazes para seus propósitos. Detalhes desses resultados são mostrados mais adiante. Por conta de limitações ao escopo do trabalho [Pinheiro 2015], outra técnica de compensação de modelo proposta pelos autores não foi apresentada. Tal técnica é referenciada como Type-2 Fuzzy GMM-UBM, que combina os chamados GMMs difusos de tipo 2 [Tsang et al. 2012, Pinheiro et al. 2013] a métodos de treinamento multicondicional [Pinheiro et al. 2014] para estimar modelos difusos robustos a incompatibilidades. O modelo proposto foi avaliado utilizando o mesmo protocolo apresentado [Pinheiro 2015] e demonstrou desempenhos superiores à modelagem tradicional GMM-UBM [Pinheiro et al. 2016], apresentando um ganho de desempenho de 31,5% em ambientes com alto grau de ruído.

3. Experimentos

Em [Pinheiro 2015], experimentos foram realizados para analisar a robustez dos sistemas de verificação de locutores na presença de incompatibilidades provenientes de ambientes ruidosos. Para isso a base de dados pública *MIT Mobile Device Speaker Verification Corpus* foi utilizada. Ela apresenta locuções provenientes de 48 locutores gravadas em três sessões distintas (uma utilizada para treinamento e outras duas para teste). Um total de 54 locuções (para cada locutor em cada sessão) foram geradas em três ambientes distintos: um escritório silencioso, o *hall* de entrada de um prédio, com nível mediano de ruído, e no cruzamento de ruas movimentadas, com alto nível de ruído de ambiente. Os sistemas foram então treinados utilizando as locuções provenientes do local silencioso (sem ruído) e testado em todos os ambientes. Dessa maneira foi possível a análise da robustez dos sistemas quando há alta incompatibilidade de ruído entre as locuções. Os experimentos foram divididos em quatro etapas, com objetivos específicos. Em cada uma delas, as melhores configurações de sistemas foram identificadas. Essa identificação foi realizada utilizando um teste estatístico pareado, o *Wilcoxon Signed-Rank*, através dos desempenhos encontrados nos pontos de operação que apresentam taxas de falsa aceitação e rejeição iguais, chamado de *Equal Error Rate* (EER).

A primeira etapa consistiu na comparação dos métodos convencionais, sem a presença de nenhum tipo de compensação. Basicamente, foram utilizadas diferentes configurações de coeficientes MFCC como possibilidades de características, e as duas modelagens: GMM-UBM e GMM-SVM, variando o número de misturas dos GMMs. Um total de 21 sistemas foram testados para ambas as modelagens ao variar a utilização dos coeficientes dinâmicos e ao variar a quantidade de distribuições dos modelos (potências de 2, de 16 a 1024). Nesse experimento foi constatada a superioridade estatística da modelagem GMM-SVM em relação ao GMM-UBM, quando as melhores configurações encontradas para as modelagens foram comparadas.

Já na segunda etapa dos experimentos foram avaliadas as técnicas de compensação de características propostas na literatura e as possíveis combinações entre elas. Ao variar mais uma vez a quantidade de misturas dos modelos e a utilização dos coeficientes dinâmicos dos MFCCs, um total de 294 configurações foram testadas para ambas

as modelagens. Para a modelagem GMM-UBM, um total de 12 configurações foram identificadas como as melhores e cada uma das técnicas da literatura aparecem em alguma delas. Além disso, constatou-se que, para essa modelagem, combinar técnicas de compensação de características aumenta significativamente a robustez dos sistemas (todas as 12 configurações apresentam combinações de técnicas). Por outro lado, na modelagem GMM-SVM, apenas 4 configurações foram consideradas as melhores e absolutamente todas elas apresentam apenas técnicas de compensação baseadas na normalização das características (subtração e normalização de média cepstral). Tais técnicas de normalização aparentam ser significativamente mais adequadas à abordagem GMM-SVM do que as demais. Por fim, nessa etapa foi constatada uma equivalência estatística entre os desempenhos encontrados pelas melhores configurações das modelagens GMM-UBM e GMM-SVM, o que mostra que, ao utilizar técnicas de compensação de características, a modelagem GMM-UBM é capaz de equiparar-se em desempenho ao GMM-SVM.

Na terceira etapa dos experimentos foram avaliadas as técnicas de compensação de *scores* presentes na literatura e a técnica proposta em [Pinheiro 2015]. Para cada tipo de modelagem, as configurações utilizadas (de características e de modelagem) foram fixadas para aquelas que apresentaram as menores taxas de EER nas etapas anteriores dos experimentos. Em ambas as modelagens, foram analisados os casos em que há ou não a compensação de características. Para a modelagem GMM-UBM, foi constatada uma superioridade da técnica Normalização Zero, para ambos os casos em que compensação de características ocorre ou não. Nesse caso, a aplicação da técnica proposta diminuiu a taxa de EER, mas não apresentou ganho estatisticamente significativo. Já na modelagem GMM-SVM, a técnica proposta foi aquela que apresentou o maior ganho de desempenho e se mostrou estatisticamente superior às demais. Como isso ocorreu para ambos os casos em que a compensação de características ocorre ou não, a técnica proposta se mostrou uma excelente alternativa para ser utilizada em conjunto à modelagem GMM-SVM.

Na quarta e última etapa dos experimentos, foram avaliadas as técnicas de compensação de modelos propostas na literatura e a técnica proposta em [Pinheiro 2015]. Primeiramente, a técnica Projeção de Atributos Indesejáveis foi comparada à modelagem convencional GMM-SVM, uma vez que ela é aplicável somente a essa abordagem. Nesse caso as melhores configurações da modelagem GMM-SVM observadas na segunda etapa dos experimentos foram utilizadas. Porém, nenhum ganho estatisticamente significativo foi observado em nenhuma delas. Em seguida, a técnica de compensação PUM-GMM e a técnica proposta foram comparadas com a técnica GMM-UBM. Mais uma vez, as melhores configurações encontradas para essa modelagem na segunda etapa dos experimentos foram utilizadas. Nesse caso, o desempenho encontrado pela técnica PUM-GMM foi significativamente superior ao da abordagem GMM-UBM, com ganho de desempenho de 15,5%. O mesmo foi observado para a técnica proposta, que provocou um ganho de desempenho de 11,1%. Porém, como observado anteriormente, a técnica PUM-GMM não é completamente adequada para verificação de locutores e por essa razão requer uma quantidade elevada de modelos de outros locutores, o que aumenta o custo computacional na fase de autenticação. Nesse contexto, a formulação apresentada em [Pinheiro 2015] utiliza o modelo de fundo para esse propósito, adequando os conceitos do PUM-GMM para o contexto de verificação. Tais resultados mostram, portanto, que a formulação proposta é tão eficaz quanto a da técnica da literatura, requerendo, porém, um custo computacional bem inferior.

4. Contribuições

Em [Pinheiro 2015] podemos destacar as seguintes contribuições: (i) descrição dos principais métodos utilizados no desenvolvimento de sistemas de verificação de locutores independente de texto; (ii) descrição dos principais métodos de compensação de ruído presentes na literatura; (iii) avaliação e comparação dos principais métodos utilizados no desenvolvimento dos sistemas; (iv) avaliação e comparação dos principais métodos de compensação presentes na literatura; (v) proposta de um método de compensação de *scores* baseado na Distribuição Normal Acumulada; (vi) proposta de um método de compensação de modelos baseado no treinamento multicondicional, na Teoria dos Dados Ausentes e na modelagem GMM-UBM; (vii) desenvolvimento de um algoritmo quadrático que realiza o cálculo da probabilidade de um Modelo de União a Posteriori; (viii) avaliação e comparação entre os métodos propostos e os métodos presentes na literatura.

Referências

- Campbell, W. M., Sturim, D. E., and Reynolds, D. A. (2006). Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Processing Letters*, 13(5):308–311.
- Ming, J., Hazen, T. J., Glass, J. R., and Reynolds, D. A. (2007). Robust speaker recognition in noisy conditions. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5):1711–1723.
- Pinheiro, H. N., Ren, T. I., Cavalcanti, G. D., Jyh, T. I., and Sijbers, J. (2013). Type-2 fuzzy GMM-UBM for text-independent speaker verification. In *International Conference on Systems, Man, and Cybernetics (SMC)*, pages 4328–4331. IEEE.
- Pinheiro, H. N., Ren, T. I., Cavalcanti, G. D., Jyh, T. I., and Sijbers, J. (2014). Type-2 fuzzy GMMs for robust text-independent speaker verification in noisy environments. In *International Conference on Pattern Recognition (ICPR)*, pages 4531–4536. IEEE.
- Pinheiro, H. N. B. (2015). *Verificação de Locutores Independente de Texto: uma Análise de Robustez a Ruído*. Dissertação (Mestrado em Ciência da Computação) - Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil.
- Pinheiro, H. N. B., Vieira, S. R. F., Ren, T. I., Cavalcanti, G. D. C., and Mattos Neto, P. S. G. (2016). Type-2 fuzzy GMM for text-independent speaker verification under unseen noise conditions. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- Reynolds, D. A., Quatieri, T. F., and Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1):19–41.
- Tsang, R., Gabriel, D., Pinheiro, H. N., and Cavalcanti, G. D. (2012). Speaker verification using type-2 fuzzy gaussian mixture models. In *International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2336–2340. IEEE.

A gamified peer assessment model for on-line learning environments: An experiment with MeuTutor

Thyago Tenório¹, Ig Ibert Bittencourt²

¹Universidade Federal de Alagoas (UFAL) – Campus Arapiraca/Pólo Penedo
Av. Beira Rio – 57.200-000 – Penedo – AL – Brasil

²Instituto de Computação (IC) – Universidade Federal de Alagoas (UFAL)
Av. Lourival Melo Mota, s/n – 57.072-900 – Maceió – AL – Brasil

thyago.oliveira@penedo.ufal.br, ig.ibert@ic.ufal.br

Abstract. *Though the on-line learning environments provide scalable and automatic ways to present a content, forums and evaluate the progress of the students, they are still limited in their ability to evaluate complex activities such as correction of essays. Facing this issue, Peer Assessment (PA) offers a powerful solution. However, a major problem found when using PA is the presence of inadequate behaviours from students, which affect the learning and evaluation system. In this sense, this paper proposes a gamified peer assessment model, where the gamification elements improvement the motivational aspect of the students. In the experiments conducted, it was concluded that there were no statistical variations between the grades obtained compared to the experts, the correction time was reduced and the cost involved decreased 72.4%. The gamification encouraged students to use the platform: increased access (+64.28%), essays performed (+10.53%) and essays corrected (+20%).*

1. Introduction

The recent increase in the popularity of online courses became more accessible for anyone, with an internet connection, to enroll freely in university level courses [Piech et al. 2013]. However, while new web technologies allow scalable ways to deliver video lecture content, to implement social forums and to track student progress, we remain limited in our ability to evaluate and give feedback for complex and often open-ended student evaluations, such as mathematical proofs, design problems and essays [Piech et al. 2013]. This happens because this kind of activity requires manual and individually correction by teacher, which leads to an excessive workload. The further students in the environment, more activities will be done and the correction by teachers would quickly become infeasible.

Facing this issue, Peer Assessment (PA) offers a promising solution to scale the grading of complex activities in courses with tens or even hundreds of thousands of students [Piech et al. 2013]. PA stimulates students to reflect, to discuss and to collaborate in their learning process [Topping 1998]. Peer Assessment is a process by which students or their peers attach grades based on predefined benchmarks by professor [Sadler and Good 2006]. PA is recommended because reduces the teacher's workload [Rubin and Turner 2012], increases learning outcome [Murakami et al. 2012].

The effectiveness and quality of an assessment depend of their incorporation into of the learning process [Schuwirth 2004]. Peer Assessment is a collaborative evaluation

process, where the quality of the final results depends of good evaluations of their reviewers. However, a major problem found when using Peer Assessment is the presence of inadequate behaviours from students, which affect the learning and assessment system. Often, the students are not motivated to collaborate with the reviews of the activities. The emergence of this type of behaviour occurs by various cognitive and emotional factors, such as boredom, lack of motivation and the need to get results quickly [Kapp 2012]. To resolve this issue, it is necessary to use techniques that have the ability to “engage” positively the emotional/cognitive state of the student.

The paper proposes a gamified peer assessment model, where the gamification elements improvement the motivational aspect of the students. Gamification is the use of mechanical, ideas and aesthetics games (context, fast feedback, competition, stages, achievements, points, among others), to engage people, to motivate actions, to promote learning and to solve problems [Kapp 2012]. This term is commonly used to express the use of game elements (storyline, score, levels, quests, badges and rankings) in environments that are not games (educational environment, for example) to motivate or influence people to perform a certain activity. The combined use of peer assessment techniques with gamification become the process most powerful and complete, avoiding and/or decreasing the presence of inappropriate behaviour by students.

2. Peer Assessment Model

This section aims to describe the proposed model. The Figure 1 shows the model’s steps. The first step is the creation of the discursive activities and definition of the evaluation form containing the criteria on which the activities will be evaluated. The second step starts when the activity is released to student. They will answer it and will submit their responses to the system.

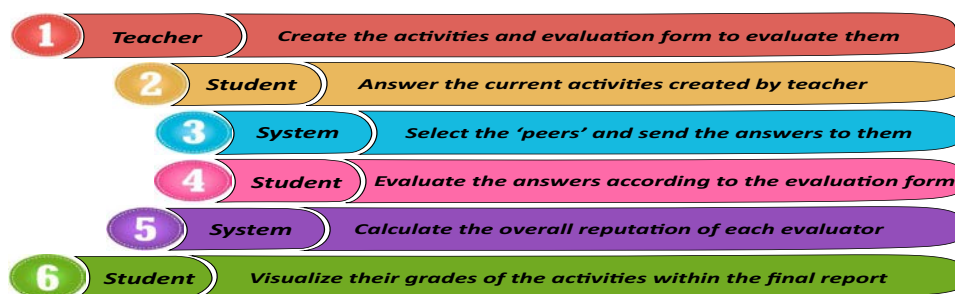


Figure 1. Model in details

In the third step, the system sends the submitted activities to others students correct them. Then, in the fourth step, the students, according to the distribution in the previous step, will evaluate the responses of their peers based on pre-defined evaluation form by the teacher (step 1), in other words, in this step the students will evaluate their peers and will send the grades to the system. The deadline for the correction of the activities in the system is called evaluation period. To assign his grade, the student will receive the responses of a particular student and the evaluation form with pre-defined criteria. He must analyse the activity from the point of view of the criteria and then simply fill out the form with their grades.

In order to have greater confidence in the final grade, the fifth step evaluates the overall reputation and the level of competence of each evaluator for each criterion. In addition to the general reputation in the system, it is necessary to calculate the reliability in the grades of the users of the system. So, the system calculates the reliability of each assessment, the competence's level of evaluator in each criteria and their reliability and can then calculate the final result as a weighted average. Finally, the sixth step consists in generate a report by system with the grades of each criteria as well as the final result of the activity, which to be presented to the student. It is possible include comments and there is possible to changes on the grades by teacher before their presentation to the student.

The details of each step of the model can be found in Chapter 4 (section 4.1) of the dissertation. This entire process is kept in functioning through the use of gamification elements, which aims to increase the motivation of users. The main objective is to ensure students feel motivated to participate on both performing activities and correcting them. In this sense, we defined the following target behaviours: To encourage students to participate more often in the activities; To challenge students to get badges; To provide medals to students who participated of the activities; To provide badges or points to students that answered/corrected activities; To promote a ranking parametrized for students; To encourage competition with rewards to the winners.

3. Controlled Experiments and Results

We applied our Peer Assessment model in a on-line learning environment called MeuTutor®. The educational environment MeuTutor is an intelligent tutoring system, which aims to monitor the learning of students in a personalized way, ensuring quality in teaching and improving the performance. The chosen version was the MeuTutor-ENEM, which aims to help high school students to prepare for the National High School Exam (ENEM). Thus, the environment offers courses related to high school subjects like Portuguese, mathematics, among others. Our goal was evaluate the effectiveness of to use our proposed model in the context of correction of essays.

We analysed two fundamental aspects: 1- Effectiveness of using Peer Assessment - It involves questions like **“How can we evaluate the quality of the corrections made by the students?”** and especially **“Can these assessments be compared with expert’s assessments?”**. We also inquired about **“Does the model applied in this context really reducing the cost and overwork of the teacher? ”**; 2- Gamification impact - It involves questions like **“What is the influence of gamification in the proposed model?”** and **“Does the gamification really influence and motivate students within the peer assessment process in this competitive environment?”**

Thus, two experiments were performed at separate. The first evaluates the effectiveness of the peer assessment model and the second evaluates the influence of gamification into model. For the first, we selected about 30 students who had some essays done in the MeuTutor environment. The correction process was done using the proposed model and the traditional way (where the experts corrected the essays). The variables analysed in experiment planning were: grade (N) (from 0 to 1000); Time for correction (T) (in minutes); Cost per student per correction (C) (em R\$).

The data of these metrics are summarized in Table 1. We considered three aspects: Application of the traditional model (correction by expert) within the environment Meu-

Tutor (T1); Application of the proposed peer assessment model within the environment MeuTutor (T2); traditional model without the use of educational environment (T3).

Table 1. Summarization of the data of experiment 1

Treatment	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Metric grade (N)						
T1	80.0	490.0	660.0	615.6	750.0	880.0
T2	380.0	520.0	573.0	577.7	624.5	960.0
T3	400.0	460.0	540.0	536.0	590.0	680.0
T1 - T2	-400.00	-79.75	40.00	37.89	110.00	480.00
Metric Time (T)						
T1	153.0	382.8	600.0	679.3	889.5	1833.0
T2	240.0	337.5	472.5	501.7	588.8	900.0
T3	420.0	555.0	660.0	669.1	780.0	917.0
Metric Coust (C)						
T1	1.36	3.40	5.33	6.03	7.90	16.29
T2	1.68	1.68	1.68	1.68	1.68	1.68
T3	3.03	4.01	4.77	4.83	5.63	6.62

Through statistical tests presented in Chapter 5 (more specifically in Section 5.3.3), we had the following results: 1- There were no statistically significant variations between the grades obtained in both the proposed model and the traditional model. With this, it is possible to use the proposed model to replace the traditional model in the environment and get their benefits without compromising the final results of the evaluations; 2- The correction time using the proposed model showed faster than expert correction time; 3- There were enough statistical evidence to indicate that the cost involved of the proposed model showed lesser than the cost of the traditional model. Through of the creation of a regression model, it shows that the cost reduction is about 72.4%.

The second experiment aims to evaluate the motivational impact of the inclusion of gamification in the proposed model in this work. So, our research question was **“Does the use of gamification elements in the peer assessment model bring significant differences on the motivational aspect of the students involved?”**. The metrics used to evaluate the model of this experiment were: Amount of access (A), number of essays performed (RF) and amount of corrected essays (RC). The experiment was conducted in a private school with high school classes (about 100 students). Randomly, it was distributed to each student a group (1 or 2) representing the control group (GC) and experimental group (GT). Students in the control group (GC) performed the treatment T1 (without gamification) while students of the experimental group (GT) performed the treatment T2 (with gamification).

The Table 2 briefly present some of the main results obtained in the experiment, in a direct comparison between the model without gamification (T1) and with gamification (T2).

The Table 2 showed that there were two more registers in the group of the treatment T2 (+11.76 %). The number of pendent essays was four, regardless of treatment. This may have been caused by the short deadline for the realization and correction of essays, implying that some students failed to correct enough essays. By have more es-

Table 2. Summarization of data on the variables analyzed

Data observed	T1	T2	Difference	% Dif.
Number of student records	17	19	+2	+11.76%
Total Pending Essays	4	4	0	0%
Average grades	669.33	563.92	-105.41	-15.75%
Number of logins (A)	42	69	+27	+64.28%
Total Essays Made (RF)	19	21	+2	+10.53%
Total Corrected Essays (RC)	45	54	+9	+20.0%

says made and due to this amount be equal, we concluded that the percentage difference between grades is greater in gamification (-15.75%). Moreover, the data showed that, in T2, the amount of access was 64.28% higher, the number of essays performed was 10.53% higher and 20% more essays corrected if compared to T1, which indicates the strong influence of gamification on the platform.

4. Discussion

The paper proposes a gamified peer assessment model that uses gamification elements as a motivational aspect for students inside Peer Assessment process. Two experiments in MeuTutor educational environment showed satisfactory results (the grades of both models are equivalent, lower time and a cost about 72% lesser on proposed model). Moreover, gamification influenced positively on the overall context of the peer assessment model (64.28% more access, 10.53% more essays performed and about 20% more essays corrected).

To analyse these results, we will simulate with a real scenario of application. Suppose that a single school with about 100 students resolves to promote the creation of essays in their on-line environment. Suppose that going to be performed one essay per week per student, which leads to four essays monthly per student. Suppose an annual period of 10 months of work in this project. So, at the end we will have a total of 4000 essays performed (100 students x 4 essays/month x 10 months).

Considering the number of essays and the average time obtained in our experiments, we have that an expert would take about 754 hours (4000 x 11.31min) to correct all essays in total. On the other hand, the simple use of the model without gamification would decrease this time to an average of 8.36min for essay, which would give us a total of 557 hours (4000 x 8.36min) spent with corrections (a gain of 197 hours, or 26.12%). The inclusion of gamification in the model would reduce the correction time for 6.68min. So, with the application of the complete model, the time total to correct the essays would be 445 hours (4000 x 6.68min), a reduction of 309 hours or 40.98% compared to the traditional model or reduction of 112 hours or 20.10% compared to the model without gamification.

Considering the same number of essays and analyzing the average of the costs involved in the experiments, we have that on average, an essay correction by a specialist would cost about R\$6.04. Thus, the total to correct all essays would be about R\$ 24.160 (4000 x 6.04). On the other hand, the simple use of the model without gamification decrease this cost to an average of R\$1.68 for essay, which would give us a total of R\$

6.720 (4000 x 1.68) (about R\$ 17.440 lesser or 72.18%). So, with the application of the complete model, the cost total to correct the essays would be R\$ 5.360 (4000 x 1.34), a reduction of R\$ 18.800 or 77,81% compared to the traditional model or reduction of R\$ 1.360 or 20.23% compared to the model without gamification. Now, imagine the gains if increase the number of schools, which means more students or increase the amount of essays performed by student.

5. Publications

The work presented here was reported in some publications in periodic and conferences directly and/or indirectly. Some publications were: 1- A gamified peer assessment model for on-line learning environments in a competitive context (accepted for publication in *Computers in Human Behavior - Qualis A2 - Impact Factor=2.694*); 2- "Avaliação por pares como ferramenta colaborativa na correção de redações: Um experimento com o ambiente educacional MeuTutor" (accepted for publication in *CSBC SBSC 2016 - Qualis B4*); 3- A gamified peer assessment model for on-line learning environments: An experiment with MeuTutor (*CSBC CTD 2016*); 4- Does Peer Assessment in On-line Learning Environments work? A systematic review of the literature (in analysis on *Computers in Human Behavior*); 5- Ontology-based Feature Modeling: An Empirical Study in Changing Scenarios (published on *Expert systems with application - Qualis A2 - Impact Factor=2.240*); 6- Java ontology integrated toolkit (same number 5); 7- A Semantic Tool to Assist Authors in the Instantiation of Software Product Lines for Intelligent Tutoring Systems Context (published on *IEEE Multidisciplinary Engineering Education Magazine*); Among other papers indirectly.

References

- Boud, D., Cohen, R., and Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4):413–426.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.
- Murakami, C., Valvona, C., and Broudy, D. (2012). Turning apathy into activeness in oral communication classes: Regular self-and peer-assessment in a tblt programme. *System*, 40(3):407–420.
- Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., and Koller, D. (2013). Tuned models of peer assessment in moocs. *arXiv preprint arXiv:1307.2579*.
- Rubin, R. F. and Turner, T. (2012). Student performance on and attitudes toward peer assessments on advanced pharmacy practice experience assignments. *Currents in Pharmacy Teaching and Learning*, 4(2):113–121.
- Sadler, P. M. and Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1):1–31.
- Schuwirth, L. (2004). Optimising new modes of assessment: in search of qualities and standards. *Tijdschrift voor Medisch Onderwijs*, 5(23):250–251.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, 68(3):249–276.

Um Algoritmo Exato em *clusters* de GPUs para o *Hitting Set* Aplicado à Inferência de Redes de Regulação Gênica

Danilo Carastan dos Santos - Autor^{1*}, Luiz Carlos da Silva Rozante - Orientador¹

¹Centro de Matemática, Computação e Cognição
Universidade Federal do ABC (UFABC)
Santo André – SP – Brasil

{danilo.santos, luiz.rozante}@ufabc.edu.br

Abstract. We propose a Hitting Set problem (HSP) algorithm suited for GRN inference applications by introducing innovations in the data structures and a sorting scheme to allow efficient discarding of Hitting Set non-solution candidates. We provided an implementation for multi-core CPUs and for (homogeneous and heterogeneous) GPU clusters. Experimental results show that the optimizations offered significant individual performance gains and a high scalability with increasing number of GPUs. The conjunction of these gains resulted in speedups above 60 for the parallel part of the algorithm. We publish two papers: one in conference (Qualis CC A1) and the other in journal (Qualis CC A2).

Resumo. Propomos um algoritmo para resolver o problema do Hitting Set (HSP), adequado para aplicações de inferência de GRNs, introduzindo inovações nas estruturas de dados e um mecanismo de ordenação que permite um descarte eficiente de candidatos que não são solução. Foram providas implementações em CPU multi-core e em clusters (homogêneos e heterogêneos) de GPU. Resultados experimentais mostraram que as otimizações ofereceram ganhos de desempenho individuais significativos e alta escalabilidade com o aumento do número de GPUs. A conjunção desses ganhos resultou em speedups acima de 60 para a parte paralela do algoritmo. Publicamos dois artigos: um em congresso (Qualis CC A1) e outro em periódico (Qualis CC A2).

1. Introdução e Motivação

Diversas áreas de estudo possuem problemas teóricos ou práticos os quais podem ser modelados, em parte ou como um todo, como uma instância do problema da Transversal Mínima (do Inglês, *Hitting Set Problem* ou HSP). A área de Biologia de Sistemas não é uma exceção, apresentando aplicações tais como distância reversa genômica [Kolman and Walen 2007], experimentos com reação em cadeia de polimerase [Pearson et al. 1996] e inferência de redes de regulação gênica (do Inglês *Gene Regulatory Networks* ou GRNs) [Ideker et al. 2000, Ruchkys and Song 2003]. Ideker et al. [Ideker et al. 2000], por exemplo, utiliza uma abordagem iterativa de perturbação de sinais, na qual se aplicam perturbações sucessivas nos níveis de expressão dos genes e as

*Financiado pelo CNPq

alterações consequentes são analisadas com o objetivo de inferir interações regulatórias entre os genes. Uma característica dessa abordagem é que um de seus passos pode ser modelado via HSP. Essas interações entre os genes são usualmente essenciais para diversos tratamentos de doenças [Madhamshettiwar et al. 2012] e estudos de um sistema biológico [Barrera et al. 2007]. Em geral, é bastante habitual que os subsequentes tamanhos de entrada para o HSP dessas aplicações sejam consideravelmente grandes, fazendo com que a obtenção das soluções exatas seja inviável ou até mesmo impossível para os algoritmos atuais.

Uma vez que o HSP é um problema NP-Difícil [Garey and Johnson 1999], existem alguns algoritmos que tentam contornar esse problema, tais como algoritmos exatos não polinomiais [Shi and Cai 2010], algoritmos de aproximação e heurísticas [Ruchkys and Song 2003, Cendic 2014]. Ruchkys e Song [Ruchkys and Song 2003] propuseram algoritmos de aproximação sequencial e paralelo para resolver o *Hitting Set* de modo a inferir quais genes afetam o nível expressão de um determinado gene-alvo. Mais recentemente, Steinbach e Posthoff [Steinbach and Posthoff 2012] propuseram um algoritmo exato que utiliza GPUs (do Inglês, *Graphics Processing Units*) para determinar a cobertura mínima de expressões Booleanas, o qual é um problema dual do *Hitting Set*. A exceção do algoritmo de aproximação de Ruchkys e Song [Ruchkys and Song 2003], nenhum dos algoritmos supracitados são capazes de lidar com grandes conjuntos de entrada (ordem de milhares de variáveis), o que é precisamente o caso no contexto de inferência de GRNs. Além disso, nenhum destes algoritmos permitem utilizar múltiplas GPUs.

Nesse trabalho foi proposto um algoritmo que faz uso de múltiplas GPUs para se obter as soluções exatas do HSP, incluindo um mecanismo de ordenação capaz de desclassificar eficientemente as não soluções do problema. O algoritmo foi utilizado para resolver o problema de inferência de GRNs e os resultados experimentais mostraram que o mecanismo de ordenação resultou em *speedups* de até 3,5, em comparação com o algoritmo sem esse mecanismo. Adicionalmente, a implementação em GPU foi capaz de fornecer um *speedup* adicional de 4,7, em comparação com uma implementação em CPU *multi-core*. Por fim, utilizando 8 GPUs Tesla K20c, dividida em duas *workstations*, foi obtido um *speedup* adicional de 6,6, em comparação com o algoritmo executando em apenas uma GPU. Todas essas otimizações combinadas fornecerem *speedups* acima de 60, levando em consideração as partes paralelizadas do algoritmo.

O restante desse texto está estruturado como segue. Na Seção 2 são apresentadas algumas das definições preliminares desse trabalho, incluindo a definição do HSP e o problema de inferência de GRNs. Na Seção 3 são apresentados os objetivos e as contribuições obtidas nesse trabalho. Os resultados alcançados, incluindo os artigos publicados, são apresentados na Seção 4 e as conclusões são apresentadas na Seção 5.

2. Algumas Definições

2.1. Inferência de Redes de Regulação Gênica

Dado um conjunto de dados de entrada e um determinado gene alvo, o problema de identificação dos preditores consiste em determinar os genes cujos níveis de expressão realizam uma melhor predição do nível de expressão desse gene alvo. Um gene *preditor* é um gene cujo perfil de expressão possui informação significativa (sozinho ou em conjunto com outros preditores) sobre o perfil de expressão do gene alvo [Barrera et al. 2007].

Sendo assim, o problema de inferência de GRNs consiste em determinar o melhor conjunto de preditores para todos os genes alvo em estudo, levando em conta os dados de entrada. Em geral, os dados de entrada são medidas de expressão dos genes de um determinado sistema biológico. Embora estas expressões sejam de natureza contínua, adotamos uma representação Booleana, que é um bom ponto de partida para a modelagem de GRNs [Kauffman 1969]. A literatura referente aos métodos de inferência de GRNs expande rapidamente. Mais informações relativas a esses diversos métodos de inferência podem ser encontradas na seguinte revisão: [Ristevski 2013].

Muitos dos métodos existentes de inferência de GRNs não fazem utilização de técnicas de paralelismo. Entretanto, publicações recentes indicam que há um crescimento em pesquisa de métodos paralelos de inferência de GRNs, em especial, métodos que utilizam o processamento paralelo em GPUs, como por exemplo os trabalhos de Shi *et al.* [Shi et al. 2011] e Borelli *et al.* [Borelli et al. 2012]. Nesse trabalho consideramos a abordagem adotada por [Ruchkys and Song 2003], na qual é utilizada uma estratégia baseada no HSP (ver Seção a seguir) para efetuar a inferência de GRNs.

2.2. Problema do *Hitting Set*

Em linhas gerais, o HSP é um problema de otimização que é capaz de modelar diversos problemas combinatórios e pertence à classe dos problemas NP-difíceis [Garey and Johnson 1999]. Formalmente, podemos definir o HSP da maneira a seguir. Dado um conjunto finito X , uma coleção \mathcal{S} de subconjuntos de X (alternativamente denominada como coleção de cláusulas) e um inteiro positivo k , o objetivo é encontrar um subconjunto $H \subseteq X$, de menor cardinalidade, tal que:

$$|H| \leq k \text{ e } H \cap S \neq \emptyset, \forall S \in \mathcal{S}. \quad (1)$$

Mais de um subconjunto de X pode satisfazer as condições acima. É importante notar que o inteiro positivo k , no contexto de inferência de GRNs, denota a quantidade de preditores por gene alvo e, portanto, o grau máximo dos vértices da GRN inferida. Neste trabalho foi definido o grau máximo $k \leq 5$, uma vez que existem várias evidências que sugerem que o grau médio dos genes de uma GRN está entre 2 e 3 [Kauffman 1993].

3. Objetivos e Contribuições

O objetivo desse trabalho consistiu no desenvolvimento de um algoritmo capaz de obter soluções exatas do HSP aplicado ao problema de inferência de GRNs. Essa proposta conseguiu lidar, de maneira eficiente, com conjuntos de entrada contendo milhares de variáveis, pela introdução de diversas inovações nas estruturas de dados utilizadas no algoritmo. Além disso, propusemos um mecanismo de ordenação o qual permite um descarte eficiente de candidatos que não são soluções do *Hitting Set*. Tais aprimoramentos permitiram que o algoritmo proposto seja utilizado no método de inferência de GRNs de Ruchkys e Song [Ruchkys and Song 2003]. Em seguida, desenvolvemos uma implementação do algoritmo proposto capaz de executar em aglomerados (*clusters*) de GPU de maneira eficiente e escalável. O uso de *clusters* de GPUs permitiu a execução de conjuntos de entrada maiores (ver mais detalhes na Seção 4).

As principais contribuições deste trabalho são:

1. Um mecanismo de ordenação que possibilita um descarte no conjunto de soluções candidatas do problema *Hitting Set* que proporciona, dependendo da configuração da entrada, um considerável ganho de desempenho (*speedups* entre 2,2 e 3,5).
2. Um algoritmo paralelo e eficiente baseado na arquitetura GPU/CUDA (*speedup* entre 1,6 e 4,7), para resolver o problema *Hitting Set* – aplicado à inferência de GRNs – com quantidade arbitrária de elementos (variáveis).
3. Uma representação computacional dos elementos do problema *Hitting Set* que permite o uso relativamente eficiente dos recursos de memória das GPUs.
4. Extensão do algoritmo mencionado no item 2, o qual é capaz de fazer uso de aglomerados (*clusters*) de GPUs homogêneas, com curva de *speedup* próxima do linear em função da quantidade de GPUs (*speedup* de 6,6 para até 8 GPUs).
5. Aprimoramento do algoritmo proposto no item 2, incluindo um novo mecanismo de geração e atribuição de candidatos a solução à *threads* da GPU, capaz de reduzir severamente as porções sequenciais do algoritmo e a transferência de dados para as GPUs, aumentando ainda mais a escalabilidade do algoritmo.
6. Aprimoramento do algoritmo do item 4, adicionando suporte para *clusters* heterogêneos, com um mecanismo automático de balanceamento de carga que determina os tamanhos de entrada mais apropriados para as GPUs disponíveis.

4. Resultados Obtidos

As implementações do algoritmo proposto foram avaliadas utilizando dois computadores, cada um com um processador Intel® Xeon E5-2690 - com 10 núcleos, 25MB de *cache*, 3.0 GHz e 256GB de memória RAM - e quatro placas de vídeo NVIDIA® Tesla K20c, com 2496 núcleos e 5GB de memória global. As implementações estão disponíveis através da URL <https://bitbucket.org/dancarastan/grnhspgpu>. Em todos os experimentos, os tamanhos de entrada adotados foram os tamanhos tipicamente encontrados no problema de inferência de GRNs, com instâncias do HSP com conjunto X de tamanho até 4096 (número de genes) ou instâncias cujas soluções possuam cardinalidade k até 5 (número de preditores por gene-alvo). Os tempos de execução apresentados são uma média dos tempos de execução medidos de dez execuções do algoritmo com instâncias distintas. O parâmetro de otimização `-O3` foi utilizado durante a compilação de todas as implementações os algoritmos. Foram elaborados três experimentos para mensurar a eficiência do algoritmo em relação ao tempo de execução e a escalabilidade em função do número de GPUs. Nesses experimentos foram utilizados dados artificiais gerados aleatoriamente utilizando a função `rand()` da biblioteca padrão da linguagem C, o que permitiu a criação de instâncias de tamanhos e cardinalidades definidos. Abaixo é apresentado um breve resumo dos resultados experimentais obtidos. Informações detalhadas sobre os experimentos podem ser obtidas nos artigos da Seção 4.1.

1. Foi mensurada a eficiência do mecanismo de ordenação para resolver o *Hitting Set*. Para tal, o algoritmo sequencial proposto foi executado com e sem o mecanismo de ordenação. Foi adotada a biblioteca OpenMP para paralelizar em CPU o algoritmo e foram utilizados 10 núcleos de CPU para esse experimento. Nos casos avaliados foram obtidos *speedups* entre 2,2 e 3,5, evidenciando a eficiência desse mecanismo de pré-processamento.
2. Executamos a implementação em GPU utilizando apenas uma GPU e seu tempo de execução foi comparado com o tempo de execução da implementação em

CPU utilizando dez núcleos. Em ambos os casos, foi efetuado o procedimento de ordenação antes de avaliar as soluções. Nesse experimento, levando em consideração as partes paralelizadas do algoritmo, foram obtidos *speedups* entre 1,6 e 4,7 e crescentes com o aumento do número total de genes e com o aumento do número de preditores por gene alvo. Esse resultado mostra que o algoritmo em GPU proposto é capaz de utilizar os recursos da GPU de maneira eficiente.

3. Avaliamos a escalabilidade da implementação utilizando um aglomerado de GPUs. Para isso utilizamos duas máquinas com 4 GPUs cada e executamos o algoritmo para 1, 2, 4 e 8 GPUs. Nas partes paralelizadas do algoritmo, foi obtida uma curva de *speedup* em função da quantidade de GPU crescente, quase linear e com valor de 6,6 para 8 GPUs, mostrando que o algoritmo proposto consegue lidar de maneira eficiente com a adição de mais GPUs para o processamento.

Ao combinarmos as técnicas avaliadas nos experimentos 1, 2 e 3, obtemos *speedups* acima de 60 nas partes paralelizadas do algoritmo, em comparação ao algoritmo sequencial paralelizado em dez núcleos e sem o mecanismo de ordenação.

4.1. Artigos Publicados

- **Artigo publicado nos anais de congresso internacional:** Danilo Carastan-Santos, Raphael Y. de Camargo, David C. Martins-Jr, Siang W. Song, Fabrizio F. Borelli e Luiz C. S. Rozante. A multi-GPU Hitting Set Algorithm for GRNs Inference. In *15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, Shenzhen, Guangdong, China, 2015. DOI: 10.1109/cc-grid.2015.29. Conferência classificada no comitê de CC da CAPES como Qualis A1. Publicamos as contribuições referentes aos itens 1 a 4 da Seção 3.
- **Artigo publicado em periódico:** Danilo Carastan-Santos, Raphael Y. de Camargo, David C. Martins-Jr, Siang W. Song, Luiz C. S. Rozante. Finding exact hitting set solutions for systems biology applications using heterogeneous GPU clusters. *Future Generation Computer Systems*. Elsevier. DOI: 10.1016/j.future.2016.02.009. Periódico com fator de impacto da Journal Citation Reports (JCR) de 2,786 e é classificado no comitê de CC da CAPES como Qualis A2. Publicamos as contribuições referentes aos itens 5 e 6 da Seção 3.

5. Conclusão

Nesse trabalho desenvolvemos um algoritmo exato para resolver o problema do *Hitting Set* aplicado à inferência de GRNs. Foram introduzidas inovações nas estruturas de dados e foi proposto um mecanismo de ordenação que permitiu efetuar um descarte eficiente dos candidatos que não são soluções do problema. Foram solucionadas instâncias do HSP na ordem de milhares de elementos em tempo razoável, alcançando bom desempenho e escalabilidade. Assim, mostramos que é possível resolver o HSP aplicado à inferência de GRNs fazendo o uso de GPUs, em tempo razoável e para entradas típicas no contexto de inferência de GRNs (milhares de genes). É importante observar também que o algoritmo proposto nesse trabalho pode ser adaptado para outras aplicações que possam ser modeladas via HSP e que possuam tamanho de entrada da ordem de milhares de elementos.

Referências

Barrera, J., Cesar-Jr, R. M., Martins-Jr, D. C., Vencio, R. Z. N., Merino, E. F., Yamamoto, M. M., Leonardi, F. G., Pereira, C. A. B., and del Portillo, H. A. (2007). Constructing

- probabilistic genetic networks of *Plasmodium falciparum* from dynamical expression signals of the intraerythrocytic development cycle. In *Methods of Microarray Data Analysis V*, chapter 2, pages 11–26. Springer.
- Borelli, F. F., Camargo, R. Y., Martins, D. C., Stransky, B., and Rozante, L. C. (2012). Accelerating gene regulatory networks inference through gpu/cuda programming. In *Computational Advances in Bio and Medical Sciences (ICCABS), 2012 IEEE 2nd International Conference on*, pages 1–6. IEEE.
- Cendic, B. L. (2014). A genetic algorithm for the minimum hitting set. *Scientific Publications of the State University of Novi Pazar*, 6(2):107.
- Garey, M. R. and Johnson, D. S. (1999). *Computers and Intractability - A guide to the Theory of NP-completeness*. W. H. Freeman and Company.
- Ideker, T. E., Thorsson, V., and Karp, R. M. (2000). Discovery of regulatory interactions through perturbation: inference and experimental design. *Pacific symposium on biocomputing*, 5:302–313.
- Kauffman, S. A. (1969). Homeostasis and differentiation in random genetic control networks. *Nature*, 224(215):177–178.
- Kauffman, S. A. (1993). *The Origins of Order*. Oxford University Press, New York.
- Kolman, P. and Walen, T. (2007). Reversal distance for strings with duplicates: Linear time approximation using hitting set. *The Electronic Journal of Combinatorics*, 14(1):11.
- Madhamshettiwar, P. B., Maetschke, S. R., Davis, M. J., Reverter, A., and Ragan, M. A. (2012). Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome medicine*, 4(5):1–16.
- Pearson, W. R., Robins, G., Wrege, D. E., and Zhang, T. (1996). On the primer selection problem in polymerase chain reaction experiments. *Discrete Applied Mathematics*, 71(1):231–246.
- Ristevski, B. (2013). A survey of models for inference of gene regulatory networks. *Nonlinear Analysis: Modelling and Control*, 18(4):444–465.
- Ruchkys, D. P. and Song, S. W. (2003). A parallel solution to infer genetic network architectures in gene expression analysis. *International Journal of High Performance Computing Applications*, 17(2):163–172.
- Shi, H., Schmidt, B., Liu, W., and Muller-Wittig, W. (2011). Parallel mutual information estimation for inferring gene regulatory networks on gpus. *BMC Research Notes*, 4:189.
- Shi, L. and Cai, X. (2010). An exact fast algorithm for minimum hitting set. *Int. Joint Conference on Computational Science and Optimization*, 1:64–67.
- Steinbach, B. and Posthoff, C. (2012). Sources and obstacles for parallelization—a comprehensive exploration of theunate covering problem using both CPU and GPU. *GPU Computing with Applications in Digital Logic*, page 63.

Heurísticas para desambiguação incremental de nomes de autores em referências bibliográficas

Alan Filipe Santana¹, Marcos André Gonçalves¹ (Orientador)

¹ Departamento de Ciência da Computação
Universidade Federal de Minas Gerais – Belo Horizonte, MG – Brasil

{alanfs, mgoncalv}@dcc.ufmg.br

1. Contextualização e Motivação

Nos últimos anos é possível observar um aumento exponencial da produção global de publicações científicas. Com o desenvolvimento da *World Wide Web* (WWW), grande parte destas publicações ficam acessíveis através da Internet. Nesse cenário, Bibliotecas Digitais (BD), como DBLP¹, MEDLINE² e BDBComp³, possuem grande importância ao prover serviços que facilitam o acesso a publicações relevantes pela comunidade acadêmica, além de possibilitar pesquisas e análises relacionadas com redes de colaboração, tendências, cobertura de tópicos e impacto de publicações.

Um dos principais problemas que prejudicam a qualidade dos serviços de recuperação de informações bibliográficas fornecidos pelas BDs é a ambiguidade de nomes de autores. Este problema ocorre quando dois ou mais autores compartilham um mesmo nome (nomes homônimos) ou quando um autor utiliza diferentes nomes em suas referências bibliográficas (nomes sinônimos). Os desafios ao lidar com este problema têm levado ao desenvolvimento de inúmeros métodos de desambiguação [Ferreira et al. 2012]. Dentre as estratégias possíveis, existem esforços para o desenvolvimento de identificadores únicos globais e métodos automáticos. Serviços de associação de identificadores a pesquisadores como a Open Researcher and Contributor ID (ORCID)⁴ dependem da colaboração voluntária e ativa de pesquisadores e autores, o que pode ser improvável de se conseguir dentro de um intervalo de tempo curto e em âmbito global. Portanto, nos últimos anos, uma grande atenção tem sido dada ao desenvolvimento de métodos automáticos de desambiguação para serem aplicados em BDs [Ferreira et al. 2012].

Métodos automáticos de desambiguação normalmente tentam resolver o problema agrupando referências de um mesmo autor baseado em medidas de similaridades entre seus atributos ou, diretamente, associando uma citação a um determinado autor. Historicamente, métodos supervisionados têm, empiricamente, produzidos os melhores resultados [Ferreira et al. 2012]. Entretanto, ao depender de dados de treinamento, esses métodos podem não ser adequados em situações reais nas quais novos nomes ambíguos, ausentes no conjunto de treinamento, aparecem todo o tempo, além de desconsiderar as mudanças que comumente ocorrem nos perfis de publicação dos autores. Mesmos os métodos de desambiguação baseados em técnicas tradicionais de clusterização não são práticos ao considerar as características de uma BD real [Carvalho et al. 2011].

¹<http://dblp.uni-trier.de/>

²<http://www.ncbi.nlm.nih.gov/pubmed/>

³<http://www.lbd.dcc.ufmg.br/bdbcomp/>

⁴<http://orcid.org>

A fim de lidar com os desafios relativos a evolução das BDs, alguns métodos utilizaram heurísticas, combinando-as com abordagens supervisionadas [Veloso et al. 2012, Ferreira et al. 2014] ou com o objetivo de se obter métodos incrementais de desambiguação [Carvalho et al. 2011, Esperidião et al. 2014]. Estes últimos visam desambiguar as referências no momento em que elas são inseridas na BD, considerando um conjunto de regras ou heurísticas para identificar a presença de um novo autor e a fragmentação de grupos de citações (ou *clusters*), portanto, são potencialmente mais eficientes. Apesar do avanço obtido com estas abordagens, alguns desafios ainda estão em aberto, especialmente os relacionados com as seguintes características: (i) presença de poucas informações nas citações, (ii) tolerância a erros, (iii) eficiência (baixa complexidade de tempo), (iv) tolerância a mudanças nos perfis de publicação dos autores e; (v) a constante inclusão de novos autores [Ferreira et al. 2012].

Essa dissertação teve como objetivo o desenvolvimento de um novo método de desambiguação de nomes autores em referências bibliográficas considerando, concomitantemente, todos os desafios encontrados em uma BD real listados acima. Particularmente, focamos no problema de *desambiguação incremental*, onde apenas as novas entradas incluídas em uma BD são desambiguadas, ao invés de aplicar ao processo de desambiguação à toda BD periodicamente, como é comumente proposto. Isso é um aspecto essencial para garantir a escalabilidade do processo, além de preservar eventuais correções manuais que porventura tenham sido efetuadas. Apesar da importância do problema, pouquíssimos trabalhos na literatura têm focado nele até bem pouco tempo atrás (apenas recentemente trabalhos têm começado a surgir, e.g., [Carvalho et al. 2011, Esperidião et al. 2014]), o que também demonstra o caráter inovador e o impacto prático de nossas soluções.

2. Principais Contribuições

As principais contribuições deste dissertação incluem:

- O desenvolvimento de um novo método de desambiguação incremental, baseado em heurísticas específicas do domínio do problema, altamente eficiente e efetivo, capaz de ser utilizado de forma supervisionada ou não.
- Análise de uma estratégia de incorporação de características baseadas em co-ocorrência de palavras na resolução da tarefa de desambiguação.
- Avaliação do método proposto utilizando coleções reais extraídas a partir da DBLP e BDBComp e comparação dos resultados com vários métodos supervisionados e não-supervisionados encontrados na literatura.
- Avaliação do método proposto em cenários que simulam a evolução de bibliotecas digitais durante um determinado período de tempo, utilizando coleções reais e sintéticas, e comparação dos resultados com os dois únicos métodos incrementais encontrados na literatura.

Uma parte dos trabalhos desenvolvidos foi publicada e apresentada na *ACM/IEEE Joint Conference on Digital Libraries (JCDL)* [Santana et al. 2014] (*Qualis A2*), a principal conferência mundial na área. Uma versão estendida desse trabalho foi convidada para uma edição especial dos melhores artigos da conferência e aceita para publicação, após nova revisão, no periódico *International Journal on Digital Libraries (IJDL, Qualis B3)* [Santana et al. 2015]. Além dessas publicações, um novo artigo abordando

a versão mais recente do método proposto foi aceita para publicação no periódico *Journal of the Association for Information Science and Technology* (JASIST, *Qualis A1*) [Santana et al. 2016], o principal periódico da área.

3. Desambiguação Incremental baseada no *Cluster* mais Similar

O método proposto na dissertação consiste basicamente em três fases: (i) seleção de *clusters* (conjuntos de citações bibliográficas) candidatos, (ii) cálculo das similaridades entre cada citação e *cluster* candidato e (iii) atualização do conjunto de treinamento. A última fase é composta pelas etapas: (a) identificação de novos autores, (b) atualização de associações duvidosas, e (c) identificação de *clusters* fragmentados. Cada fase explora heurísticas específicas do domínio com o objetivo de criar automaticamente um conjunto de treinamento utilizado para definir as associações entre referências e autores. As etapas de cada fase são detalhadas no Capítulo 3 da dissertação e resumidas a seguir.

Seleção de *clusters* candidatos: Dada uma citação de teste c_k e uma referência ambígua c_k^a , a primeira fase do método proposto consiste na seleção dos *clusters* que possuem pelo menos um nome compatível com c_k^a utilizando o algoritmo de comparação de nomes desenvolvido por [Oliveira 2005], chamado de Comparação de Fragmentos.

Cálculo das similaridades entre cada citação e *cluster* candidato: Cada termo encontrado na citação fornece uma evidência da associação entre uma referência da citação e um autor. A força desta evidência varia conforme o atributo ao qual o termo pertence e a sua capacidade discriminativa. Com base nestas observações foi escolhido utilizar uma função de similaridade baseada na soma ponderada das similaridades entre cada atributo da citação, que possui a referência ambígua, e das citações de cada *cluster* selecionado na fase anterior. Para o cálculo das similaridades entre os atributos foi utilizado uma função que pondera os termos da citação de acordo com algumas heurísticas: (i) cálculo da capacidade discriminativa do termo a partir do número de autores que o utilizam, (ii) estimativa da probabilidade condicional da referência pertencer a um dado autor dada a ocorrência do termo no conjunto de treinamento e (iii) estimativa da distribuição do termo dentro de um determinado *cluster*.

Identificação de novos autores: após o cálculo das similaridades, se o maior valor encontrado for maior ou igual a um limite γ , a citação é associada ao *cluster* correspondente, caso o contrário, um novo *cluster* é criado para representar o autor da referência.

Atualização de associações duvidosas: sempre que uma citação é associada a um *cluster* do conjunto de treinamento, uma métrica que estima a confiança da classificação é calculada. Se o valor dessa métrica for menor que γ , a citação é incluída em um conjunto \mathcal{E} para futuras reclassificações. Caso o contrário, todas as citações presentes no conjunto \mathcal{E} , que compartilham pelo menos um termo com a citação classificada, são reclassificadas no intuito de corrigir possíveis erros durante o processo de desambiguação. A métrica de confiança desenvolvida utiliza as similaridades obtidas na segunda fase do algoritmo e retorna um valor entre 0 e o maior valor de similaridade encontrado.

Identificação de *clusters* fragmentados: durante a desambiguação, os grupos de citações dos autores que trabalham em diferentes linhas de pesquisa podem ser fragmentados, mesmo se forem utilizados pequenos valores para o parâmetro γ . Para lidar com esse problema, sempre que uma nova citação é incluída no conjunto de treinamento e sua

Tabela 1. Valores médios da métrica K obtidos pelos métodos supervisionados.

Método	Coleção		
	DBLP	BDBComp	KISTI
DICS	0,919 ± 0,026	0,956 ± 0,024	0,952 ± 0,002
SLAND	0,878 ± 0,027	0,882 ± 0,031	0,927 ± 0,002
Cosine	0,884 ± 0,028	0,746 ± 0,041	0,883 ± 0,003
SVM	0,777 ± 0,038	0,579 ± 0,042	0,797 ± 0,004
NB	0,711 ± 0,045	0,537 ± 0,067	0,768 ± 0,005

classificação é considerada confiável, são calculadas as similaridades entre o grupo da citação e todos os outros grupos que compartilham pelo menos um termo em comum com a citação. Se essa similaridade for maior do que γ , os *clusters* são unidos. Para comparar dois *clusters* foi utilizada uma função similar a que compara uma citação com uma *cluster*, com a diferença na utilização de procedimento de normalização dos valores a partir do número de citações do menor *cluster*.

Para aumentar a praticidade e eficiência do método proposto, foram desenvolvidas heurísticas para a configuração automática do método com ou sem a presença de um conjunto de treinamento (veja as Seções 3.2 e 3.3 para mais detalhes). O algoritmo desenvolvido possui baixa complexidade de tempo, quando comparado com outras abordagens tradicionais, conforme mostrado na Seção 3.5 da dissertação.

4. Avaliação Experimental

Para avaliar o desempenho do método desenvolvido foram utilizados cinco métodos baseados em associação de autores - SAND [Ferreira et al. 2014], SLAND [Velo et al. 2012], Cosine [Lee et al. 2005], SVM [Han et al. 2004] e NB [Han et al. 2004] -, dois métodos baseados em agrupamentos de autores - LAVSM-DBSCAN [Huang et al. 2006] e HHC [Cota et al. 2010] - e dois métodos incrementais - INDi [Carvalho et al. 2011] e MINDi [Esperidião et al. 2014]. Os experimentos foram realizados utilizando duas coleções extraídas a partir da DBLP (DBLP e KISTI), uma a partir da BDBComp e quatro coleções sintéticas. Para comparar os resultados obtidos pelos métodos de desambiguação, foram utilizadas duas métricas: a métrica K e a *pairwise* F1. A seguir são destacados os principais resultados obtidos durante a avaliação do método proposto, chamado de DICS, em diferentes cenários de aplicação.

Comparação com métodos supervisionados: Para avaliação dos métodos supervisionados foram utilizadas as coleções DBLP, BDBComp e KISTI. Os grupos ambíguos de cada coleção foram divididos em conjuntos de treino e teste, cada um com 50% do número total de citações. Esta divisão foi realizada de maneira aleatória e repetida 10 vezes. A Tabela 1 mostra os resultados obtidos em termos da métrica K. O método DICS obteve os melhores valores em todas as coleções com apenas um empate estatístico na coleção DBLP com o método Cosine em relação a métrica pF1. Em comparação com os melhores *baselines*, considerando a métrica K, o ganhos foram de 3,9%, 8,4% e 2,7% nas coleções DBLP, BDBComp e KIST, respectivamente. Em relação a métrica pF1, o maior ganho médio foi de 28% na coleção BDBComp.

Comparação com métodos não supervisionados: Para avaliação dos métodos não supervisionados foram utilizadas as mesmas divisões de teste usadas nos experimentos com os métodos supervisionados. A Tabela 2 mostra os resultados obtidos em termos

Tabela 2. Valores médios da métrica K dos métodos não supervisionados.

Método	Coleção		
	DBLP	BDBComp	KISTI
DICS	0,791 ± 0,059	0,944 ± 0,025	0,942 ± 0,003
SAND	0,674 ± 0,091	0,942 ± 0,022	0,892 ± 0,003
HHC	0,692 ± 0,084	0,937 ± 0,021	0,862 ± 0,003
LASVM-DBSCAN	0,479 ± 0,097	0,883 ± 0,042	0,858 ± 0,004

da métrica K. O método DICS obteve os melhores resultados médios nas coleções DBLP e KISTI. Na coleção BDBComp houve empate estatístico entre DICS e os métodos SAND e HHC. Os ganhos obtidos em relação ao melhor *baseline* foram de 14% e 5,5% nas coleções DBLP e KISTI respectivamente. Comparando com os resultados obtidos com a utilização do conjunto de treinamento, a queda de desempenho foi de 13,9% na DBLP e menos de 2% na coleções BDBComp e KISTI.

Comparação com métodos incrementais: Para a avaliação dos métodos incrementais, foram utilizadas as coleções BDBComp e KISTI ordenadas em ordem cronológica e quatro coleções sintéticas que simulam cenários onde novos autores são inseridos no repositório da BD e onde os autores alteram seus perfis de publicação. Em relação à métrica K, DICS obteve os maiores valores com ganhos de até 8,7% e 50% comparando com INDi nas coleções BDBComp e KISTI, respectivamente. Analisando os resultados em termos de coesão (PMA) e pureza (PMC), DICS foi por volta de 7,3% menos efetivo em relação à métrica PMC quando comparado com o método INDi, mas cerca de 26% e 143% mais efetivo considerando a métrica PMA nas coleções BDBComp e KISTI, respectivamente. Nas coleções sintéticas DICS, superou todos os *baselines* em termos de coesão e equilíbrio entre fragmentação e pureza. Nos cenários simulando a introdução de novos autores, o ganho, comparando com o melhor *baseline*, alcançou 4,5% em relação à métrica K. Nos cenários simulando mudanças nos perfil de publicação dos autores, o método proposto obteve ganhos de até 19,8% comparado com o método MINDi.

Além destas avaliações, foram realizadas diversos outros experimentos que avaliaram: (i) a utilização da estratégia de incorporação de características baseadas na coocorrência de palavras proposto por [Figueiredo et al. 2011], (ii) o impacto de cada componente do algoritmo, (iii) a sensibilidade do método DICS aos valores dos seus parâmetros e (iv) a análise do tempo de execução.

5. Conclusão e Trabalhos Futuros

Neste trabalho, foi proposto um novo método incremental de desambiguação baseado em heurísticas capaz de criar e atualizar automaticamente um conjunto de treinamento utilizado para determinar os autores de cada citação. Foram propostos procedimentos para realizar a configuração automática dos parâmetros com e sem a presença de dados de treinamento. Foi realizada uma extensa avaliação experimental, utilizando coleções reais e sintéticas, a fim de avaliar o desempenho do método quando este é aplicado de forma supervisionada, não supervisionada e incremental. Em todos os cenários, o método proposto superou todos os *baselines*, com ganhos em quase todos os grupos ambíguos. Também foram apresentadas: a avaliação de uma estratégia de incorporação de características baseadas em coocorrência de palavras, uma análise das capacidades do método, uma análise de sensibilidade aos valores dos parâmetros e uma análise da complexidade do algoritmo.

No futuro, pretende-se: (i) avaliar estratégias para alteração automática dos valores dos parâmetros à medida que o método atualiza o conjunto de treinamento; (ii) avaliar a inclusão de novos atributos, quando disponíveis, como ano da publicação e endereço dos autores; e (iii) explorar estratégias de *relevance feedback* para permitir melhorias no modelo de desambiguação a partir da interação com um administrador da BD.

6. Agradecimentos

Esse trabalho foi parcialmente financiado pelo projeto INWeb (MCT/CNPq 573871/2008- 6) e por financiamentos individuais de CNPq, CAPES e FAPEMIG.

Referências

- Carvalho, A. P., Ferreira, A. A., Laender, A. H. F., and Gonçalves, M. A. (2011). Incremental Unsupervised Name Disambiguation in Cleaned Digital Libraries. *J. of Inf. and Data Manage.*, 2(3):289–304.
- Cota, R. G., Ferreira, A. A., Nascimento, C., Gonçalves, M. A., and Laender, A. H. F. (2010). An unsupervised heuristic-based hierarchical method for name disambiguation in bibliographic citations. *J. of the American Soc. for Info. Sci. and Tech.*, 61(9):1853–1870.
- Esperidião, L. V. B., Ferreira, A. A., Laender, A. H. F., Gomes, D. M., Tavares, A. I., and Assis, G. T. (2014). Reducing Fragmentation in Incremental Author Name Disambiguation. *JIDM*, 5(3).
- Ferreira, A. A., Gonçalves, M. A., and Laender, A. H. (2012). A Brief Survey of Automatic Methods for Author Name Disambiguation. *SIGMOD Record*, 41(2):15–26.
- Ferreira, A. A., Veloso, A., Gonçalves, M. A., and Laender, A. H. F. (2014). Self-training author name disambiguation for information scarce scenarios. *JASIST*, 65(6):1257–1278.
- Figueiredo, F., Rocha, L., Couto, T., Salles, T., Gonçalves, M. A., and Meira Jr., W. (2011). Word Co-occurrence Features for Text Classification. *Inf. Syst.*, 36(5):843–858.
- Han, H., Giles, L., Zha, H., Li, C., and Tsioutsoulouklis, K. (2004). Two Supervised Learning Approaches for Name Disambiguation in Author Citations. In *Proc. of the 4th ACM/IEEE-CS JCDL, JCDL '04*, pages 296–305, New York, NY, USA. ACM.
- Huang, J., Ertekin, S., and Giles, C. L. (2006). Efficient Name Disambiguation for Large-scale Databases. In *Proc. of the 10th European Conf. on PKDD*, pages 536–544, Berlin, Heidelberg. Springer-Verlag.
- Lee, D., On, B.-W., Kang, J., and Park, S. (2005). Effective and Scalable Solutions for Mixed and Split Citation Problems in Digital Libraries. In *Proc. of the 2nd int. workshop on Info. Quality in Info. Syst.*, pages 69–76.
- Oliveira, J. W. A. (2005). Uma estratégia para remoção de ambiguidades na identificação de autoria de objetos bibliográficos. Master's thesis, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.
- Santana, A. F., Gonçalves, M. A., Laender, A. H. F., and Ferreira, A. A. (2014). Combining domain-specific heuristics for author name disambiguation. In *Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on*, pages 173–182.
- Santana, A. F., Gonçalves, M. A., Laender, A. H. F., and Ferreira, A. A. (2015). On the combination of domain-specific heuristics for author name disambiguation: the nearest cluster method. *International Journal on Digital Libraries*, 16(3):229–246.
- Santana, A. F., Gonçalves, M. A., Laender, A. H. F., and Ferreira, A. A. (2016). Incremental author name disambiguation by exploiting domain-specific heuristics. *JASIST.*, (Aceito para publicação).
- Veloso, A., Ferreira, A. A., Gonçalves, M. A., Laender, A. H. F., and Meira, Jr., W. (2012). Cost-effective On-demand Associative Author Name Disambiguation. *Inf. Proc. & Manage.*, 48(4):680–697.

Multiple Parenting Relationships in Image Phylogeny: Tracking Down Forgeries and Their Creators Online*

Alberto Oliveira¹ and Anderson Rocha¹

¹ Institute of Computing – University of Campinas (UNICAMP)
Campinas – SP – Brazil

{alberto.oliveira, anderson.rocha}@ic.unicamp.br

Abstract. *Due to the large amount of images shared on the web, tracking the spread and evolution of their content have become an increasingly important problem. As an image might be a composition created through the combination of the semantic information existent in two or more source images, establishing a relationship between the sources and the composite is an ever-growing problem of interest. We name as Multiple Parenting Phylogeny the problem of identifying such relationships in a set containing near-duplicate subsets of source and composition images. To tackle this problem, this work presents a three-step solution: (1) separation of near-duplicate groups; (2) classification of the relations between the groups; and (3) identification of the images used to create the original composition. Furthermore, we extend upon this framework by introducing key improvements, such as better identification of when two images share content, and improved ways to compare this content. Evaluation of the proposed method is performed by means of quantitative metrics established for evaluating the accuracy in reconstructing phylogenies and finding multiple parenting relationships in the different datasets. Finally, we also analyze the results qualitatively, with images obtained from the web*

1. Introduction

Once a digital document is shared online, a common fate for it is to be copied, transformed, and re-shared, often with a completely different intention from the original. Multimedia files portraying the same semantic content, but diverging by minor image processing transformations, receive the name of *near duplicates*, and their detection and recognition (NDDR) has been greatly studied in the literature, often focused in the image domain [Zhao and Ngo 2009, Zhao et al. 2007].



Figure 1. Example of a set of **Semantically Similar Images (SSIs)** and **Near-Duplicate Images (NDIs)**. (a) and (b), are original Images. (c) and (d) are NDIs obtained from (a) while (e) is an NDI obtained from (b).

Such works, however, overlooked any causal relationships between near duplicate images. A parenting relationship exists between *Near-Duplicate Images*(NDIs) *A* and *B*,

*We thank the support of UNICAMP, CAPES DéjàVu, CNPq, FAPESP, and the European Union.

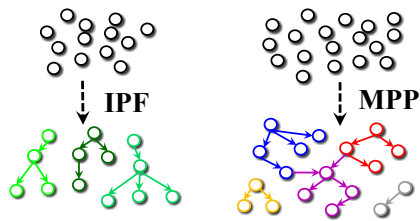


Figure 2. Image Phylogeny Forest (IPF) reconstructs a forest to represent the entire SSI set. Multiple Parenting Phylogeny (MPP) reconstructs multiple trees, finding, if existent, a relationship between a composition and the images used to create it.

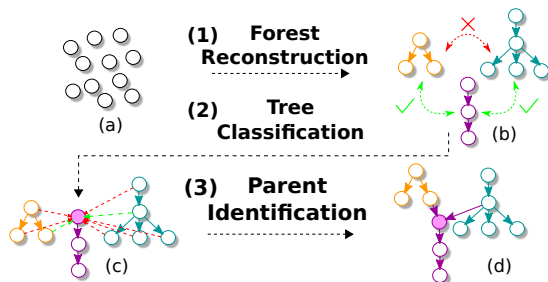


Figure 3. Multiple Parenting Phylogeny framework. (a) Set of images. (b) Searching for content relationships between IPTs in order to classify them. (c) The purple colored tree in the center, having a content relationship with the other two, is the composition tree. (d) We search for the host and alien parent of the composition root in the remaining IPTs.

when one was transformed generating the other, as Figure 1 illustrates. Recently, several works have been concerned with modeling such relationships in a set of NDIs, with the objective of better understanding how content evolves when shared and remixed on the internet [Kennedy and Chang 2008, De Rosa et al. 2010, Dias et al. 2010, Dias et al. 2012]. The work of Dias et al. [Dias et al. 2010, Dias et al. 2012] introduced the *Image Phylogeny Tree* (IPT), a directed graph representing all the parenting relationships between images of an NDI set. An asymmetric *dissimilarity* measure is employed to estimate the likelihood that an image A is parent of an image B . Once computed for all pairs of images, in both directions, the IPT is reconstructed by using a minimum spanning tree algorithm adapted to directed graphs. The root of the IPT is expected to be the original image that spawned the set.

Dias et al. [Dias et al. 2013] later expanded their work for a more generic set of *Semantically Similar Images* (SSIs). Two images are SSIs if both depict the same scene, but were not necessarily generated from the same source image. An example of the occurrence of this scenario is when two pictures are taken from the same scene, but using different cameras, as Figure 1 depicts. Costa et al. [Costa et al. 2014] further introduced several new IPF reconstruction approaches.

The aforementioned works, however, considered that a parenting relationship could only exist between two NDIs. A common scenario on the web is the creation of new images through the combination of existing ones. Splicings, montages, and mosaics are possible ways to create *compositions* by using already existing content. The parenting relationship existent between compositions and the source images used to create them gave rise to a new problem, which we refer to as *Multiple Parenting Phylogeny* (MPP). Given a set of images of varied content, the objective of MPP is to (1) identify the different phylogenies existent and (2) find the multiple parenting relationships, if any. Figure 2 contrasts the objectives of MPP and IPF. Applications of MPP range from forensic and copyright enforcement, by providing proof that an image is an forgery or created from protected content, as well as the tracking of how viral contents evolve and are created on the internet.

In this work, we tackle the MPP problem, considering *splicing* compositions, created by inserting an object extracted from an *alien* image to the background provided by a

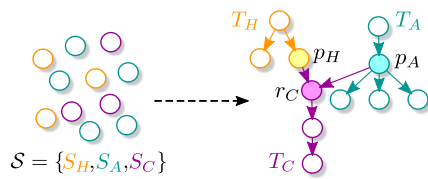


Figure 4. A multiple parenting scenario for splicing compositions. The studied set of images comprises three subsets of host, alien and composition NDIs.

host image. Given a set containing alien, host, and composition NDIs, we aim at identifying the IPTs of each of those three subsets, as well as the multiple parenting relationship existent between the subsets. For that, we introduce a three step framework: (1) Forest Reconstruction, (2) Tree Classification, and (3) Parent Identification. Figure 3 illustrates the proposed framework. In addition to the proposed framework, another contribution of this work is the creation of three datasets, comprising of compositions and their source images: We have also laid out an evaluation plan for this scenario, proposing several metrics to quantitatively assess the quality of the MPP approach. The following sections further detail this work’s contributions.

2. Multiple Parenting Phylogeny Framework

Within the Multiple Parenting Phylogeny problem, we focus our interest on *splicing*-type compositions, created by combining the object of an *alien* image with the background of a *host* image. Our test scenario is one in which three groups of NDIs exist, containing host (S_H), alien (S_A) and composition (S_C) NDIs. Each set forms an IPT, T_H , T_A , and T_C , respectively, and the root of T_C , or r_C , was created by combining a host parent p_H from T_H with an alien parent p_A from T_A . This scenario is illustrated by Figure 4. Our objectives within MPP are twofold: (1) reconstruct T_H , T_A , and T_C , and (2) identify r_C , p_A , and p_H . For this, we proposed the framework depicted by Figure 3. Next, we detail each step of the framework.

2.1. Forest Reconstruction

Before finding the MPP relationships, it is important to identify and separate which images belong to the host, alien and composition set of NDIs. Dias et al. [Dias et al. 2013] proposed the *Automatic Oriented Kruskal* (AOK) algorithm for Image Phylogeny Forests arguing that the dissimilarity between SSIs from different sources was significantly larger than the dissimilarity from NDIs. This allowed them to compute an adaptive threshold from the dissimilarity values, in order to identify the different IPTs of the forest. Costa et al. [Costa et al. 2014] later proposed the *Extended Automatic Optimum Branching* (E-AOB), employing the Optimum Branching algorithm in a similar fashion.

We compared both algorithms as NDI group separation tools in the MPP scenario. The rationale is that, if both algorithms work well in the SSI scenario, whereby images have strongly related content, it should also work well in the MPP scenario, whereby images have either strongly related (compositions and hosts), weakly related (compositions and aliens) or unrelated (hosts and aliens) content. Moreover, employing an IPF algorithm allows us to integrate MPP to the already existent Image Phylogeny frameworks. Our results have shown that, in fact, both algorithms perform really well in the MPP scenario, with a particular highlight to the performance in root identification, which was around 80% for both. Although E-AOB outperformed AOK in the controlled datasets, we observed that in the web-scenario AOK performed better, thus pointing out that the best choice for algorithm is largely dependent on the type of problem faced.

2.2. Tree Classification

Once NDIs are separated into trees, we need to identify which is the host, alien, and composition trees. To do this, we took advantage of the content relationships existent between images of different trees. We define two images as having a content relationship when some of their content is shared. A composition and a host share the background, a composition and an alien share the spliced object, while a host and an alien have no content relationship at all. Taking advantage of this information, we used randomly selected images from each IPT, and using the content relationships between each pair of them, find out which tree is most likely to contain composition images.

To discover if two images share content, we start by matching local features between them. Shared content means that parts from both images are the same, and thus many local matches between the pair should conform to the same rigid transformation. Therefore, it is possible to cluster matches based on their spatial transformation, and if a big enough cluster is observed, we assume that it is likely that the matched images share content. Additionally, by examining the area covered by the clustered matches, it is possible to reason that the shared content is the whole background, or only a small object. By registering this information when looking for shared content between images from different IPTs, we can use it after identifying which is the composition IPT to classify the remaining IPTs into host or alien.

When analyzing the complete test cases (no unrelated nodes and no missing nodes), the classification scheme has yielded very good results, with composition IPT classification accuracy of over 85% for the hardest, professionally made splicings, and over 90% classification accuracy for both host and alien IPTs for all datasets. Those results, depend strongly on a good IPF reconstruction.

2.3. Parent Identification

After classifying the IPTs, the last step of the framework consists in identifying which nodes participated in the composition process. The composition root (r_C) was created by combining one of the host nodes (p_H) with one of the alien nodes (p_A). Finding r_C comes directly from correctly reconstructing the IPF and classifying the composition IPT, as r_C is the root of that tree. For all datasets, in the complete scenario, the r_C identification accuracy ranged from 73% to 81%, considering all datasets.

Once we identify r_C , we compare it against all the candidate nodes from the host and alien IPTs in order to find the two nodes most likely to be p_H and p_A . For this, we developed the *Local Dissimilarity*, an approach to compare images considering only their shared content. First, we detect the shared content between a pair of candidate images, by clustering their local feature matches. Then, only the region inside the convex hull of the clustered matching features is used for comparison. The remainder comparison is similar to that of the original dissimilarity procedure, including registration, color adjustment and compression adjustment. The host and alien image with smaller local dissimilarity to the r_C are pointed as p_H and p_A . In the hardest, professionally made dataset, our p_H and p_A was at least 72% and 53%, respectively. It is significantly harder to identify p_A , as the region it covers is much smaller and harder to locate.

3. Datasets

For validation we built two datasets containing compositions and the source images used to create them. The first was an *Automatic Splicing Dataset*, which is further divided into two different pasting processes: *Direct Pasting* and *Poisson Blending*. In the former, we paste the alien object onto the host image as is, with no image processing techniques employed. In the later, the pasting process uses Poisson Blending to better blend the pasted object and the background. This dataset uses images from popular image retrieval and image segmentation datasets, and employs the same procedure from to create phylogeny trees. Once an IPT is generated, in the same fashion employed by Dias et al. [Dias et al. 2010, Dias et al. 2012], for the alien and host images, two nodes were randomly picked to generate the composition.

We also created a *Professional Splicing Dataset*, for which we hired a professional artist to create compositions intended to fool human viewers. Additionally, the host and alien images are all high resolution (obtained through image sharing websites).

Once the compositions were created, we generated three sets of test cases. The first comprises complete IPTs of host, alien, and composition images. The second, is the expanded set, adding unrelated images to the complete set. The last, is the missing nodes scenario, with removal of randomly chosen nodes, as well as nodes involved in the composition process (compositon root, alien parent, and host parent). Both expanded and missing nodes scenario were only created for the Professional Splicing Dataset.

4. Web Scenario

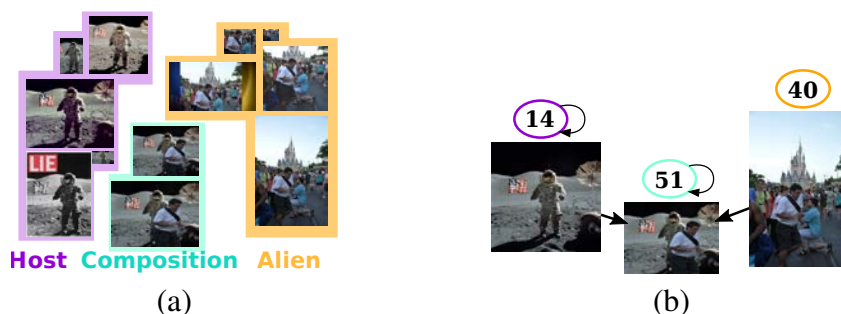


Figure 5. (a) Web scenario example images. (b) Multiple parents on web scenario. The alien parent is scaled down.

Besides the controlled scenarios mentioned before, we also validated our approach on a test case obtained from the web. This test case contained a total of 55 images, divided in 25 host NDIs, 25 alien NDIs, and 5 composition NDIs. Figure 5 illustrates some examples. Correct separation of trees is one of the biggest hurdles for MPP, as incorrectly joining the host and composition IPTs, or separating some of the IPTs in many smaller IPTs, might make it difficult to later classify the IPFs. In this regard, contrary to the controlled scenario, AOK performed better at IPF reconstruction than E-AOB, as it ended up finding a smaller number of IPTs. However, both algorithms separated the IPTs very well, with no images from different groups in the same IPT.

The classification of IPTs also proceeded correctly for the composition IPT, showing the robustness of the method. When classifying the remainder IPTs, only a single of them (composed of a single node) was mis-classified as alien, when it was a host. This

image in particular was heavily modified, thus presenting a challenge to classification. Finally, the images identified as p_H and p_A are shown by Figure 5. Although we have no way to confirm those are in fact the correct host and alien parent, they are reasonable choices, as no content present in the composition image is missing in them, which would indicate an incorrect result.

5. Conclusions and Future Research Directions

In this work, we have dealt with the Multiple Parenting Phylogeny problem, which rises when images are created by combining the content of other images. We proposed a 3-step framework for this scenario, and extensively tested it on controlled scenarios of varied difficulty, plus in one uncontrolled, web scenario. Our results show that we can, not only detect montages and compositions online, but also pinpoint the images used to create them. This is a prime step toward empowering forgery detection systems with capabilities of tracking the forgery creators other than just detecting such forgeries.

However, there are still some untied knots that need to be properly taken care of, which includes alternative forms for separating images from different NDI groups, as well as the design of better IPF reconstruction algorithms, crucial for a correct identification of the relationships between IPTs. Finally, it is also important to refine the ways that we detect and extract regions that contain similar or equal content between different images, as well as judging when shared content is not present.

6. Publications and Relevant Production

1. de Oliveira, A. A., Ferrara, P., De Rosa, A., Piva, A., Barni, M., Goldenstein, S., Dias, Z., and Rocha, A. (2015). Multiple parenting phylogeny relationships in digital images. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11(2), 328-343. (**Impact Factor**: 2.408 Journal Citation Reports).
2. de Oliveira, A. A., Ferrara, P., De Rosa, A., Piva, A., Barni, M., Goldenstein, S., Dias, Z., and Rocha, A. (2014). Multiple parenting identification in image phylogeny. *IEEE Intl. Conference on In Image Processing (ICIP)*, (pp. 5347-5351). IEEE.
3. Costa, Filipe, de Oliveira, A. A., Ferrara, P., Goldenstein, S., Dias, Z., and Rocha, A. New Dissimilarity Measures for Image Phylogeny Reconstruction. *Elsevier Journal of Visual Communication and Image (JVCI)*, submitted, 2015. (**Impact Factor**: 1.218 Journal Citation Reports).

References

- Costa, F., Oikawa, M., Dias, Z., Goldenstein, S., and Rezende de Rocha, A. (2014). Image phylogeny forests reconstruction. *IEEE Transactions on Information Forensics and Security (TIFS)*, 9(10):1533–1546.
- De Rosa, A., Uccheddu, F., Costanzo, A., Piva, A., and Barni, M. (2010). Exploring image dependencies: a new challenge in image forensics. In *SPIE Media Forensics and Security II*, page 75410.
- Dias, Z., Rocha, A., and Goldenstein, S. (2010). First steps toward image phylogeny. In *IEEE Intl. Workshop on Information Forensics and Security (WIFS)*, pages 1–6.
- Dias, Z., Rocha, A., and Goldenstein, S. (2012). Image phylogeny by minimal spanning trees. *IEEE Transactions on Information Forensics and Security (TIFS)*, 7(2):774–788.
- Dias, Z., Rocha, A., and Goldenstein, S. (2013). Toward image phylogeny forests: Automatically recovering semantically similar image relationships. *Elsevier Forensic Science Intl.*, 231(1-3):178–189.
- Kennedy, L. and Chang, S.-F. (2008). Internet image archaeology: Automatically tracing the manipulation history of photographs on the web. In *ACM Intl. Conference on Multimedia (ACM-MM)*, pages 349–358.
- Zhao, W.-L. and Ngo, C.-W. (2009). Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. *IEEE Transactions on Image Processing (TIP)*, 18(2):412–423.
- Zhao, W.-L., Ngo, C.-W., Tan, H.-K., and Wu, X. (2007). Near-duplicate keyframe identification with interest point matching and pattern learning. *IEEE Transactions on Multimedia (MM)*, 9(5):1037–1048.

Quality of Service Provisioning in LTE-EPON Integrated Networks

Carlos A. Astudillo, Nelson L. S. da Fonseca (Advisor),
Juliana Freitag Borin (Co-advisor)

¹Institute of Computing – State University of Campinas (UNICAMP)
Avenida Albert Einstein, 1251 – 13.083-852 – Campinas – SP – Brazil

castudillo@lrc.ic.unicamp.br, {nfonseca, juliana}@ic.unicamp.br

Abstract. *Passive optical networking (PON) has been recognized as a solution for backhauling next generation cellular networks. In this new scenario, QoS provisioning is an important challenge. This article proposes a Quality of Service (QoS) provisioning framework for LTE with EPON-based backhauling, which improves performance by taking advantage of information available from the backhaul as well as from the radio interface. It also introduces a QoS-aware LTE scheduler that dynamically adjusts users' priorities to increase the overall throughput. Experimental results show how the proposals can improve resource utilization and quality of service provisioning.*

1. Introduction

In the past few years, users' demand of bandwidth and the need to meet strict Quality of Service (QoS) requirements in cellular networks have increased significantly. This scenario is driven by [Cisco 2014]: (i) the number of mobile users and mobile devices, such as tablets, laptops and smartphones, which have exponentially grown in the past years and such trend will continue for next years; (ii) the amount of video traffic, which has reached 70% of the mobile Internet traffic; and (iii) the changes mobile Internet have introduced in the habits of users who are accessing the Internet seamlessly to complete multiple daily activities.

This situation has led Mobile Network Operators (MNOs) to implement the Long Term Evolution (LTE) technologies and to significantly increase the number of base stations. Since LTE base stations, evolved NodeBs (eNBs) in the LTE jargon, can offer peak downlink rates of 100 Mb/s in LTE and 1 Gb/s in LTE-Advanced as well as half of these values in the uplink direction, they will inject a large amount of traffic into the Mobile Backhaul (MBH) network. Although the cost of the future MBH will be high, it can be reduced by using already deployed Fiber to Anywhere (FTTx) systems based on Passive Optical Networks (PONs), in which Ethernet PON (EPON) is the dominant technology in the market. Motivated by this, an attractive solution is to use EPONs as backhaul for the LTE technology.

In this new scenario, an LTE base station is a subscriber of the PON, leading to a new network architecture, called LTE-EPON, in which the Optical Network Unit (ONU) and the eNB can be integrated in a single device, called ONU-eNB. Since the ONU-eNB competes for bandwidth with other ONUs, the bandwidth given to the ONU-eNB can be lower than that needed to support the mobile users in the LTE network. Given this

limitation in the available backhaul capacity, the support of QoS requirements of mobile users can be jeopardized.

This article presents the work in the dissertation [Astudillo 2015], which addresses the problem of QoS provisioning in LTE-EPON integrated networks and introduces the following original contributions:

- A novel QoS management framework for LTE/EPON networks. This framework is the first to describe how the ONU-eNB devices can improve performance by taking advantage of information sharing between EPON and LTE technologies.
- A new class of LTE schedulers, named *backhaul-aware schedulers*, that allocates radio resources considering information not only from the radio interface but also from the backhaul. The first *backhaul-aware scheduler* ever proposed was introduced in the dissertation.
- An uplink packet scheduler for QoS provisioning in LTE networks. The original contribution of this scheduler is the dynamic prioritization of users to increase the overall throughput of the system.

These contributions advance the state of the art in the design of integrated fiber-wireless (FIWI) access networks, which is expected to be the predominant technology for next generation Internet access networks. This technology takes advantage of both the mobility and ubiquity of the wireless systems and the high capacity of the optical networks. The relevance of the original contributions in this paper is, therefore, to contribute to the evolution of Internet access technology which is so fundamental to our daily life. Currently, the bandwidth bottleneck of the Internet is located at the access network, which prevents Internet users to have greater access to high demanding applications and services based on video. We postulate that the impact of these contributions is to support the increasing demand for mobile Internet access, by meeting QoS requirements of the different users' applications. If the MNOs do not cope with the increase of number of users (as well as their demand), users' experience in the "connected world" will be quite compromised. The solutions for these problems proposed in [Astudillo 2015] provide less resource utilization at reduced costs since they use under-provisioned backhaul links and radio resources efficiently.

2. Related Work

This section briefly reviews the literature related to QoS provisioning in LTE networks and LTE-PON integrated networks, including the use of backhaul information in LTE networks. Existing QoS-aware schedulers for LTE networks [Anas et al. 2008, Delgado and Jaumard 2010, Marwat et al. 2014] prioritize Guaranteed Bit Rate (GBR) over non-GBR (nGBR) bearers by adopting different priority queues for each type of bearer. This can yield unnecessary loss of nGBR requests, decreasing the total throughput of the system. Moreover, all of these schedulers assume that the GBR parameter should be guaranteed to *all* traffic flows, which violates the LTE specification since non-GBR bearers have no rate guarantees.

On the other hand, few QoS provisioning schemes for cellular networks using PON-based backhaul can be found in the literature. Chung *et al.* [Chung et al. 2008] proposed a scheme for the two-upstream-wavelength PON (2W-PON) architecture in which real-time and control packets are prioritized by transmitting them on one wavelength

while the low priority traffic is sent on another wavelength. However, 2W-PON architecture is not common in current deployed networks. Some QoS mapping schemes between LTE and PON have been proposed [Lim et al. 2012], [Chung et al. 2008], but with these approaches, the traffic in the EPON network can impact on the QoS provisioning of mobile users. The impact of available backhaul link capacity on performance of LTE networks has received little attention in the literature. In [Ghimire and Rosenberg 2014], the effect of limited backhaul capacity on the scheduling performance is studied. However, scheduling mechanisms that can deal with under-provisioned backhaul links are not reported in the literature.

In summary, there is no other work in the literature that proposes the use of backhaul information in LTE schedulers. Moreover, none of the above mentioned papers addresses the issue of coping with receiving less than the requested bandwidth. In addition, the dynamic prioritization of bearers to improve throughput was not explored before.

3. Proposed Scheduler for QoS Provisioning in LTE Networks

Although the focus of the present work is the QoS provisioning problem in LTE-EPON integrated networks, the LTE packet scheduler is essential for guaranteeing QoS requirements to mobile users and for optimally using scarce radio resources. The dissertation [Astudillo 2015] proposed the Z-Based QoS Scheduler (ZBQoS), a standard-compliant LTE uplink scheduler that provides QoS guarantees and improves overall network throughput. The ZBQoS scheduler dynamically modifies the priority of mobile users to favor those with strict QoS requirements over other users, when there are requests with delays close to the deadlines. However, when real-time traffic can be delayed (*i.e.*, a request can wait a while to be served, but still respecting their QoS requirements), non-real-time traffic can be served in order to increase the network throughput. Section 3.2 in [Astudillo 2015] gives a detailed description of the ZBQoS algorithm and its scheduling metrics which allow the dynamic prioritization of bearers.

We used the LTE-Sim simulator version 4.0 to evaluate the performance of the ZBQoS scheduler. The performance of the proposed LTE uplink scheduler was compared to that of a recently proposed scheduler called Bandwidth and QoS Aware (BQA) [Marwat et al. 2014], which is not standard compliant. The simulation scenarios are composed of a single cell, one eNB and several User Equipment (UE) devices (varying from 5 to 150, with increments of 5). Users are uniformly distributed, transmitting Voice-over-IP (VoIP), video or Constant Bit Rate (CBR) traffic. VoIP and video traffic are transmitted using GBR bearers and CBR (best effort) traffic uses nGBR bearers. Table 3.1 and 3.2 in [Astudillo 2015] give a complete description of simulation parameters, traffic models, QoS requirements and assumptions employed in the simulations.

The ZBQoS scheduler produces lower Packet Loss Ratio (PLR) values for CBR traffic when the network is *overloaded* but slightly higher PLR values when the network is *heavily loaded* (Fig. 1a). However, the lower PLR given by the BQA scheduler in *heavy-load* regime is achieved at the expense of the QoS provisioning for video traffic (Fig. 1b). Although the maximum acceptable PLR value for video traffic is 1%, the values produced by BQA scheduler increase with the traffic load, reaching 15% under *heavy load*. Thus, only the proposed scheduler can provide QoS guarantees to real-time traffic under all traffic conditions. Subsection 3.4.2 in [Astudillo 2015] presents additional results.

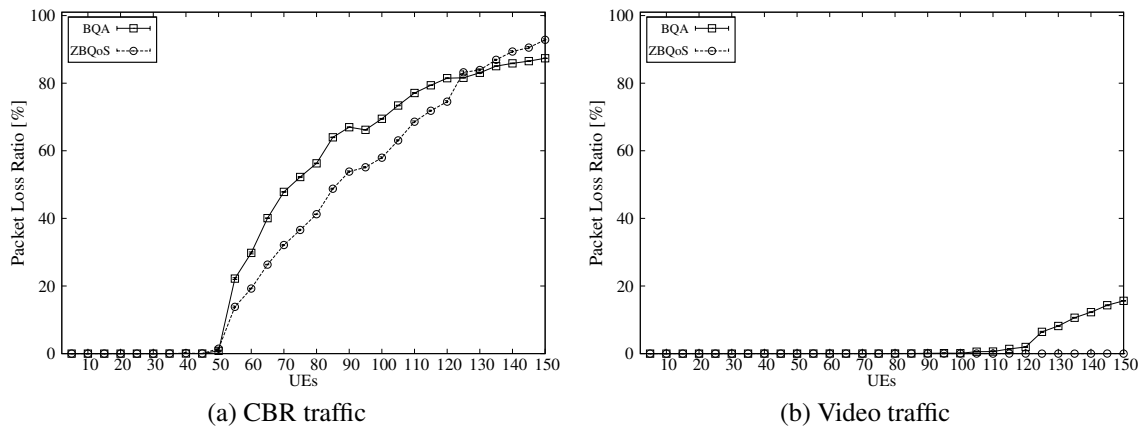


Figure 1. Performance of the ZBQoS scheduler

4. Proposed Framework for QoS Provisioning in LTE-EPON Integrated Networks

The previous section proposed a novel LTE uplink scheduler for QoS provisioning, which performs dynamic prioritization of mobile users to improve performance. However, the proposed scheduler and those existing in the literature do not take into account the state of the backhaul link to allocate radio resources, which can potentially degrade QoS provisioning for mobile users in overload conditions as well as waste both radio and backhaul resources. To cope with the variability of the backhaul link capacity and provide QoS requirements to mobile users in an integrated LTE-EPON network, this section introduces a *framework* that improves QoS provisioning for mobile users of LTE networks by taking advantage of the capacity variability of EPON-based backhaul links.

The *framework* provides a *method* for modifying existing schedulers so that the backhaul load can be considered to allocate radio resources. It also includes the functional architecture of the ONU-eNB integrated device and a QoS mapping between LTE and EPON networks. This *framework* can be rapidly adopted in current and future deployments, capitalizing in the already deployed PON systems.

The proposed *method* is based on detecting when the ONU-eNB receives less bandwidth than the amount required to support its traffic. When the backhaul link is not congested, the LTE scheduler can prioritize users as usual, properly using radio and backhaul resources. However, when a “deficit” of bandwidth is detected, the scheduler should reduce the amount of traffic scheduled to nGBR bearers. In this way, the ONU-eNB buffer size can be controlled to prevent waste of both air interface and backhaul resources. If large amount of nGBR traffic arrives at the ONU-eNB buffer during congestion in the backhaul link, the delay of packets in the buffer will increase (undesirable effect in the network). Moreover, the buffer can overflow, thus wasting radio resources used to carry the dropped packets. The proposed *method* and the other components of the framework are described in [Astudillo 2015, Section 4.3].

Applying the *method* described above, modifications to the ZBQoS scheduler were introduced so that it can be aware of backhaul capacity. As a result, the Hybrid Z-Based QoS Scheduler (HZBQoS) was proposed. This scheduler represents the first mechanism in a new class of schedulers for LTE networks that we named *backhaul-aware schedulers*. The modifications introduced are described in [Astudillo 2015, Section 4.4].

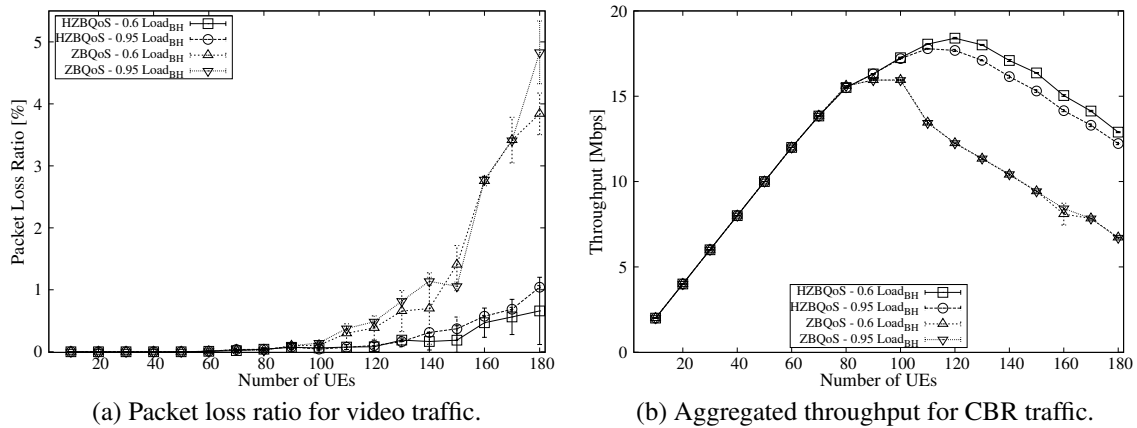


Figure 2. Performance of the HZBQoS scheduler

A simulator for LTE-EPON integrated networks was developed to assess the ability of the proposed framework to support QoS requirements of the LTE users and to improve performance under variable traffic conditions in both LTE and EPON network. We compared the performance of the scheduler modified with the proposed framework (HZBQoS) to that of the unmodified scheduler (ZBQoS).

The simulated scenarios use an EPON network with tree topology and optical channel capacity of 1 Gb/s. This network is composed of 1 Optical Line Terminal (OLT), 1 ONU-eNB and 31 conventional ONUs. The OLT employs the Interleaved Polling with Adaptive Cycle Time (IPACT) protocol with the limited scheduling policy [Kramer et al. 2002] for inter-ONU scheduling. Two distinct loads in the backhaul were used: 0.6 and 0.95 of the maximum EPON capacity, which correspond to a *lightly loaded* and a *heavily loaded* backhaul scenario, respectively. The LTE network part is composed of a single cell (served by the ONU-eNB device) and several users (varying from 10 to 180, with increments of 10). Table 4.1 and 4.2 in [Astudillo 2015] contain the settings in these simulations. The assumptions made in the evaluation of the ZBQoS scheduler hold here.

The HZBQoS scheduler is able to provide low packet loss ratio to video traffic under all traffic conditions in both the backhaul and the LTE network (Fig. 2a), which does not happen with the service provided by the ZBQoS scheduler. The PLR produced by ZBQoS surpasses 1% and increases with the traffic load, reaching 5% under heavy load.

The aggregated throughput of CBR users (Fig. 2b) shows the decrease of CBR traffic to meet the QoS requirements of real-time traffic in overloaded scenarios. However, peak throughput under the ZBQoS scheduler is achieved with 90 UEs while under the HZBQoS scheduler it is achieved with 120 UEs. Moreover, the throughput achieved by ZBQoS is independent of the backhaul load condition. This fact evinces the lack of capacity of the ZBQoS scheduler to deal with capacity variation of the backhaul link. The HZBQoS scheduler promotes a better use of network resources in scenarios with high load in both the backhaul and the mobile network, providing throughput 40% higher than that given by the ZBQoS scheduler. When Figures 2a and 2b are analyzed together, the benefits the proposed framework brings are clear since the HZBQoS scheduler yields higher aggregated throughput while generating PLR values below 1% under highly loaded scenarios. More results are presented in [Astudillo 2015, Section 4.5].

5. Publications

Results in the dissertation [Astudillo 2015] were reported in four papers. Results on LTE uplink scheduling were published in the IEEE Globecom 2013 (Qualis A1) and the WRA 2014. Results on QoS provisioning in LTE-EPON networks were published in the IEEE WCNC 2014 (Qualis A2) and in the IEEE Wireless Communications (Qualis A1). In the 2014 Journal Citation Reports, this latter publication obtained the highest Impact Factor (6.524) in Computer Science, Telecommunications, and Electrical & Electronic Engineering categories [Chen 2014]. In addition, we are preparing an article extending the results of the LTE scheduler that will be submitted to an international journal. Full references of the published papers can be found in Section 1.2 of the dissertation.

6. Conclusion

This article summarized the contribution of the master's thesis in [Astudillo 2015], which studied the QoS provisioning problem in LTE-EPON integrated networks. We introduced two novel uplink schedulers and a framework for QoS provisioning in these networks. The dynamic prioritization of requests and the consideration of backhaul information in radio resource allocation are original contributions of this work and they showed to increase the utilization of the network resources, while guaranteeing QoS requirements.

References

- [Anas et al. 2008] Anas, M., Rosa, C., Calabrese, F., Pedersen, K., and Mogensen, P. (2008). Combined admission control and scheduling for QoS differentiation in LTE uplink. In *VTC 2008-Fall*, pages 1–5.
- [Astudillo 2015] Astudillo, C. A. (2015). Provisão de qualidade de serviço em redes integradas LTE-EPON. Master thesis, LRC, Institute of Computing, UNICAMP.
- [Chen 2014] Chen, H. (2014). Impact factor and open call articles [message from the editor-in-chief]. *IEEE Wireless Commun.*, 21(4):2–4.
- [Chung et al. 2008] Chung, J., Park, J., Choi, B., Park, Y., Yoo, J., and Kim, B.-W. (2008). Application of 2W-PON to the backhaul of a wireless network. *J. Optical Networking*, 7:905.
- [Cisco 2014] Cisco (2014). The zettabyte era: Trends and analysis. White Paper.
- [Delgado and Jaumard 2010] Delgado, O. and Jaumard, B. (2010). Scheduling and resource allocation for multiclass services in LTE uplink systems. In *WiMob*, pages 355–360.
- [Ghimire and Rosenberg 2014] Ghimire, J. and Rosenberg, C. (2014). Impact of limited backhaul capacity on user scheduling in heterogeneous networks. In *WCNC 2014*, pages 2480–2485.
- [Kramer et al. 2002] Kramer, G., Mukherjee, B., and Pesavento, G. (2002). IPACT a dynamic protocol for an Ethernet PON (EPON). *IEEE Commun. Mag.*, 40(2):74–80.
- [Lim et al. 2012] Lim, W., Milosavljevic, M., Kourtessis, P., and John, M. (2012). QoS mapping for LTE backhauling over OFDMA-PONs. In *ICTON*, pages 1–4.
- [Marwat et al. 2014] Marwat, S., Weerawardane, T., Zaki, Y., Goerg, C., and Timm-Giel, A. (2014). Analysis of radio resource allocation in LTE uplink. *Wireless Personal Communications*, 79(3):2305–2322.

Efficient High-Dimensional Filtering for Image and Video Processing

Eduardo S. L. Gastal¹, Manuel M. Oliveira¹ (Advisor)

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)

{eslgastal,oliveira}@inf.ufrgs.br

Abstract. *High-dimensional filters are a fundamental building block for several applications, having recently received considerable attention from the research community. Unfortunately, naive implementations of such an important class of filters are too slow for many practical uses. This dissertation describes three novel approaches to efficiently perform high-dimensional filtering with linear cost in both the number of pixels and in the dimensionality of the space in which the filters operate. Our filters address the main limitations of previous techniques, in addition to providing the fastest performance (both on CPU and GPU) for a variety of real-world applications.*

1. Introduction

Filtering is arguably the single most important operation in image and video processing. In particular, *high-dimensional filters* are a fundamental building block for several applications, including upsampling [Kopf et al. 2007], stylization [Winnemöller et al. 2006], spatio-temporal filtering [Lang et al. 2012], denoising [Buades et al. 2005], detail manipulation [Gastal and Oliveira 2012], tone mapping [Fattal 2009], filtering of photon maps for rendering [Bauszat et al. 2011], and recoloring [Gastal and Oliveira 2011].

High-dimensional filters can be classified as *Euclidean* or *geodesic*, according to how they compute the distances between samples. The main difference between the two groups is the filter behavior near strong discontinuities (commonly called edges) in the signal. In general, Euclidean filters allow for samples belonging to different sides of a discontinuity to be combined, while geodesic filters do not. Thus, each filter type provides best results for different applications.

Due to their wide applicability, several high-dimensional and related filters have been proposed—the most popular one being the bilateral filter [Smith and Brady 1997]. Unfortunately, naive implementations of such an important class of filters are too slow for many practical uses, specially in light of the ever increasing resolution of digitally captured images, and the fact that some applications may require working with spaces containing hundreds of dimensions. As a result, several techniques have been developed that either try to accelerate high-dimensional filters, or introduce alternative ways of performing similar filtering operations on images and videos. While they clearly improve performance, these solutions still face limitations such as not being sufficiently fast for real-time applications [Adams et al. 2010], restricting filtering to certain scales [Fattal 2009], and only handling grayscale images [Yang et al. 2009], among others.

2. Contributions of this Dissertation

This dissertation describes three novel approaches to efficiently perform high-dimensional filtering, capable of achieving real-time performance: the **domain transform** for geodesic

response, the **adaptive manifolds** for Euclidean response, and a mathematical formulation for recursive filtering of **non-uniformly sampled signals**. With these frameworks, we propose several filters that address the main limitations of previous techniques, in addition to providing the fastest performance (both on CPU and GPU) for a variety of real-world applications (Figures 1 to 3). This efficiency comes from their *linear cost* in both the number of pixels and in the dimensionality of the space in which the filters operate. The contributions of this work include:

- The *first geodesic high-dimensional filter* that simultaneously exhibits the following properties: it supports a continuum of scales; its processing time is linear in the number of pixels and independent of the filter parameters, allowing for real-time computations; it correctly handles color images; and it offers control over the kernel's shape.
- The *first Euclidean high-dimensional filter* with linear complexity in both the number of pixels and in the dimensionality of the space in which the filter operates. We also give the first demonstration of a single-pass *hybrid Euclidean-geodesic filter*.
- A mathematical formulation for applying recursive digital filters to non-uniformly sampled signals, whose time complexity is linear in both the number of samples being filtered and in the order of the filter. This formulation enables, for the first time, *geodesic edge-aware* evaluation of *arbitrary* recursive infinite impulse response filters (not only low-pass), allowing for practically unlimited control over the shape of the filtering kernel.

2.1. Impact in the Area

This thesis received the inaugural ACM SIGGRAPH *Outstanding Doctoral Dissertation Award* “in recognition of the potential impact and the notable contributions” (<http://s2016.siggraph.org/content/acm-siggraph-awards>). This is the most prestigious award for a dissertation in the field of Computer Graphics and Interactive Techniques. Furthermore, our research has resulted in two publications at ACM SIGGRAPH (2011 and 2012), and a publication at Eurographics (2015). It is important to note that ACM SIGGRAPH is the world's biggest and most important conference of Computer Graphics, whose proceedings are published as a special edition of *ACM Transactions on Graphics*, the journal with the highest impact index in the area. Similarly, Eurographics is the most important computer graphics conference in Europe, and its proceedings are published as a special edition of the *Computer Graphics Forum* journal.

Our filters and algorithms have been well received by the international community. Our latest publication [Gastal and Oliveira 2015] received the Eurographics 2015 *Best Paper Award Honorable Mention*, conceded by the European Association for Computer Graphics. Furthermore, the papers describing our filters, despite of being recently published, have already received a few hundred citations, according to Google Scholar. In particular, our filters have already been put to use by researchers at Adobe, NVIDIA, Google, and Disney, in applications such as color grading [Bonneel et al. 2013], enforcing temporal consistency in video effects [Lang et al. 2012], synthetic defocus for smartphone cameras [Barron et al. 2015], and viewfinder editing for digital cameras [Baek et al. 2013]. Perhaps a better indication of this interest in our algorithms is the fact that third-party implementations of the Domain-Transform and Adaptive-Manifold filters have been included in [OpenCV 2016], a widely used library in computer vision and image processing.

In conclusion, the filters introduced by this dissertation provide a valuable tool for

the image and video processing, computer graphics, computer vision, and computational photography communities.

3. Overview of our High-Dimensional Filters

Due to space restrictions, this article cannot present all the details of the developed filters and algorithms. The complete text of the dissertation is available at <http://hdl.handle.net/10183/118258>, and the homepages of our three publications [Gastal and Oliveira 2011, Gastal and Oliveira 2012, Gastal and Oliveira 2015] may be found at <http://inf.ufrgs.br/~eslgastal>. We suggest that the readers watch the videos available in these homepages for more intuitive descriptions and examples of applications of the developed techniques.

Geodesic Filtering In [Gastal and Oliveira 2011], we introduced the *Domain Transform*, a state-of-the-art technique for geodesic filtering of images and videos. The innovation behind this method is the use of a dimensionality reduction strategy for efficient implementation of high-dimensional filters. Its inner workings are based on the interpretation of images as manifolds embedded in high-dimensional spaces. For instance, an RGB color image is a 2D manifold embedded in 5D space: x, y, r, g, b ; and a manifold is simply the generalization of a surface to high-dimensional spaces. Our idea is to map curves on this surface to a lower-dimensional space, in a way that the geodesic distances between points on the curve is preserved. For this, we derived an isometric (distance-preserving) domain transformation $t(x)$, which maps image-space pixel positions to a new space which preserves the original distances measured in high-dimensional space:

$$t(x) = \int_0^x 1 + \frac{\sigma_s}{\sigma_r} \sum_{k=1}^{d_{\mathcal{R}}} \left| \frac{d}{du} f_{[k]}(u) \right| du. \quad (1)$$

In [Gastal and Oliveira 2011], we present all the necessary steps for the derivation of this formula, along with mathematical proofs of its properties. The variables σ_s and σ_r are parameters of the filter, while $f(u)$ represents the image being filtered. Equation (1) allowed us to demonstrate real-time edge-preserving color filtering of high-definition videos. Furthermore, ours is the first geodesic filtering technique capable of working on color images at arbitrary scales in real time, without resorting to subsampling or quantization. Figure 1 illustrates the flexibility of our domain transform by using it to obtain various effects from the original photograph in Figure 1(a).

Euclidean Filtering In [Gastal and Oliveira 2012], we introduced the *Adaptive Manifolds*, a state-of-the-art technique for Euclidean filtering of images and videos. The central idea of this technique consists in subdividing the high-dimensional space into low-dimensional manifolds which adapt to the signal being filtered. Filtering is then performed independently and efficiently *on top of* each low-dimensional manifold, and the results interpolated to approximate the desired filter. Mathematically, a naive *quadratic-cost* (thus inefficient) Euclidean filter is given by $g_i = \sum_{p_j \in \mathcal{S}} \phi_{\Sigma}(\hat{p}_i - \hat{p}_j) f_j$, where ϕ is a Gaussian function, f_j are the colors of the pixels in the input image, g_i are the colors of the output pixels, and \hat{p}_i are the high-dimensional positions of the pixels. We accelerate this filter by evaluating its response on a reduced set of sampling points, and using these values to interpolate the filter's response at all input pixels. Our filter is thus defined by:

$$g_i = \sum_{k=1}^K w_{ki} \sum_{p_j \in \mathcal{S}} \phi_{\Sigma_S}(\hat{p}_i - \hat{p}_j) \phi_{\Sigma_{\mathcal{R}/2}}(\eta_{ki} - f_j) f_j,$$

where w_{ki} are weights derived from the Gauss-Hermite quadrature. With this formulation, we present a filter which is faster and requires less memory than previous approaches, being the first high-dimensional Euclidean filter with *linear cost* in both the number of pixels and in the dimensionality of the space. In [Gastal and Oliveira 2012], we present a derivation for the equations that define our method, providing a theoretical justification for the technique and for its properties. Furthermore, the flexibility of our approach allows for the first demonstration of a hybrid Euclidean-geodesic filter that runs in a single pass. Figure 2 shows examples of Euclidean-filtering applications using the adaptive manifolds.

Non-Uniform Filtering More recently [Gastal and Oliveira 2015], we introduced a mathematical formulation for applying recursive digital filters to *non-uniformly sampled signals*. This formulation enables, for the first time, geodesic edge-aware evaluation of *arbitrary* recursive infinite impulse response digital filters (not only low-pass), which allows practically unlimited control over the shape of the filtering kernel. For this, we decompose a P -th order filter into a set of 1st-order ones, and we then derive the equations for the individual 1st-order filters in a non-uniform domain. Traditionally, a 1st-order recursive filter (with standard uniform sampling) is defined by the equation $g[k] = a f[k] + b g[k-1]$, where a and b are the coefficients of the filter, $f[k]$ are the input samples, and $g[k]$ the output ones. In [Gastal and Oliveira 2015], we extend this equation to support non-uniform sampling, which results in the following expression:

$$g[k] = a f[k] + b^{\Delta t_k} g[k-1] + \left(\frac{b^{\Delta t_k} - 1}{r_0 \Delta t_k} - r_1 b \right) f[k] - \left(\frac{b^{\Delta t_k} - 1}{r_0 \Delta t_k} - r_1 b^{\Delta t_k} \right) f[k-1],$$

where r_0 and r_1 are constants, and $\{\Delta t_k\}$ is the set containing the non-uniform distances between samples of the signal. This formulation allows for greater control over the desired responses of the filters being used, which can then be adapted to specific applications. As examples, we provide the first demonstrations of a *variety of high-dimensional geodesic filters*, including Gaussian, Laplacian of the Gaussian, and low/high/band-pass Butterworth and Cauer filters, among others. Figure 3 illustrates some of the possible detail-enhancement effects generated by such responses. By allowing the use of arbitrary filters, our method enables the creation of a new rich variety of effects for image and video applications, many of which were not previously within the reach of previous techniques.

References

- Adams, A., Baek, J., and Davis, M. A. (2010). Fast high-dimensional filtering using the permutohedral lattice. *Computer Graphics Forum*, 29(2):753–762.
- Baek, J., Pajak, D., Kim, K., Pulli, K., and Levoy, M. (2013). Wysiwyg computational photography via viewfinder editing. *ACM TOG*, 32(6):198:1–198:10.
- Barron, J. T., Adams, A., Shih, Y., and Hernández, C. (2015). Fast Bilateral-Space Stereo for Synthetic Defocus. In *IEEE CVPR*, pages 4466–4474.
- Bauszat, P., Eisemann, M., and Magnor, M. (2011). Guided image filtering for interactive high-quality global illumination. *Computer Graphics Forum*, 30(4):1361–1368.
- Bonneel, N., Sunkavalli, K., Paris, S., and Pfister, H. (2013). Example-based video color grading. *ACM TOG*, 32(4):39:1–39:12.
- Buades, A., Coll, B., and Morel, J. (2005). A non-local algorithm for image denoising. In *IEEE CVPR*, volume 2, pages 60–65.

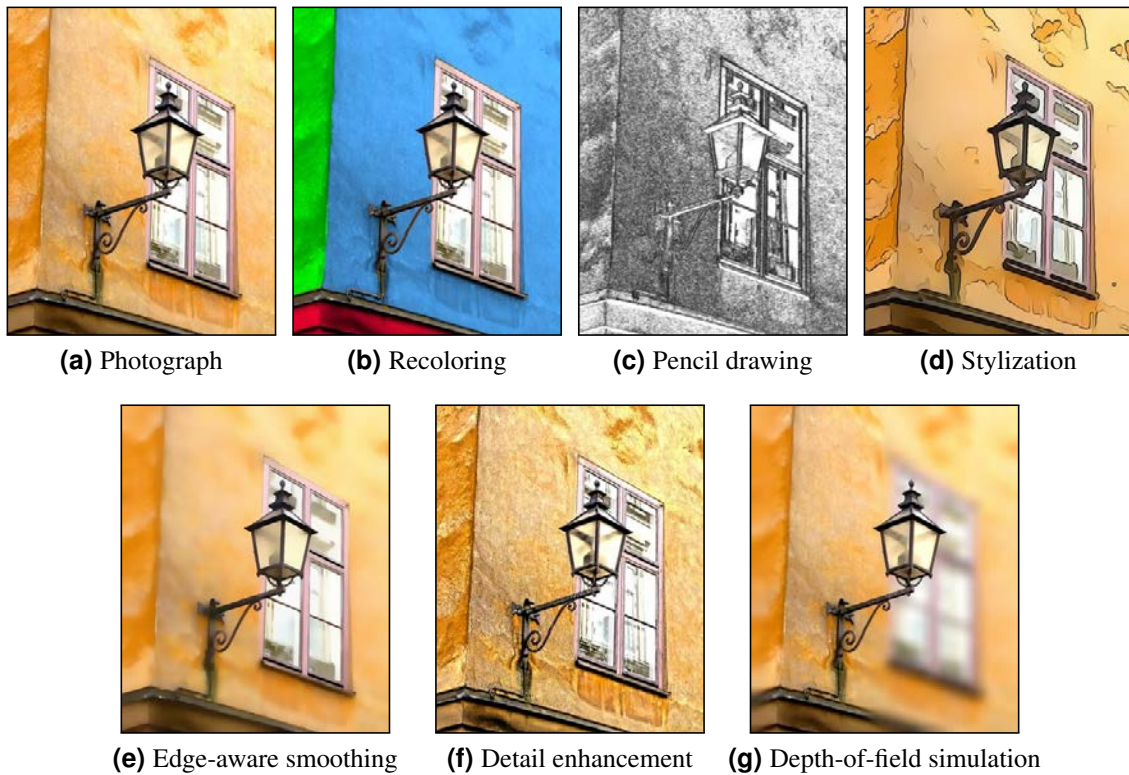


Figure 1. A variety of effects illustrating the versatility of our *domain transform* [Gastal and Oliveira 2011] when used to filter the photograph in (a).

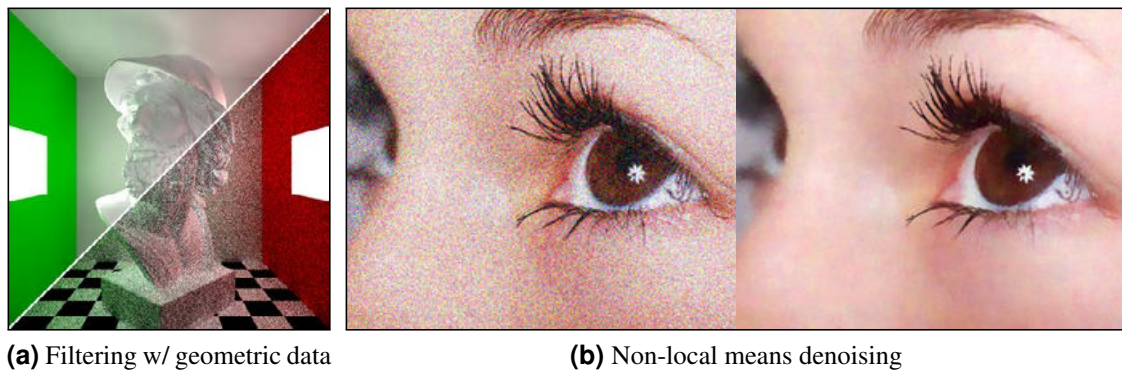


Figure 2. Examples of filtering results produced with our adaptive-manifold filter [Gastal and Oliveira 2012]. (a) Filtering (performed in 8 dimensions) of a noisy undersampled image generated using path tracing. The split rendition compares the input (bottom right) and the filtered result (top left). Note the smoothness of the shading. (b) Denoising of natural images using non-local-means (performed in 27 dimensions). Notice the noise reduction while retaining fine details.

Fattal, R. (2009). Edge-avoiding wavelets and their applications. *ACM TOG*, 28(3):22.

Gastal, E. S. L. and Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. *ACM TOG*, 30(4):69:1–69:12. Proc. of SIGGRAPH 2011.

Gastal, E. S. L. and Oliveira, M. M. (2012). Adaptive manifolds for real-time high-dimensional filtering. *ACM TOG*, 31(4):33:1–33:13. Proc. of SIGGRAPH 2012.

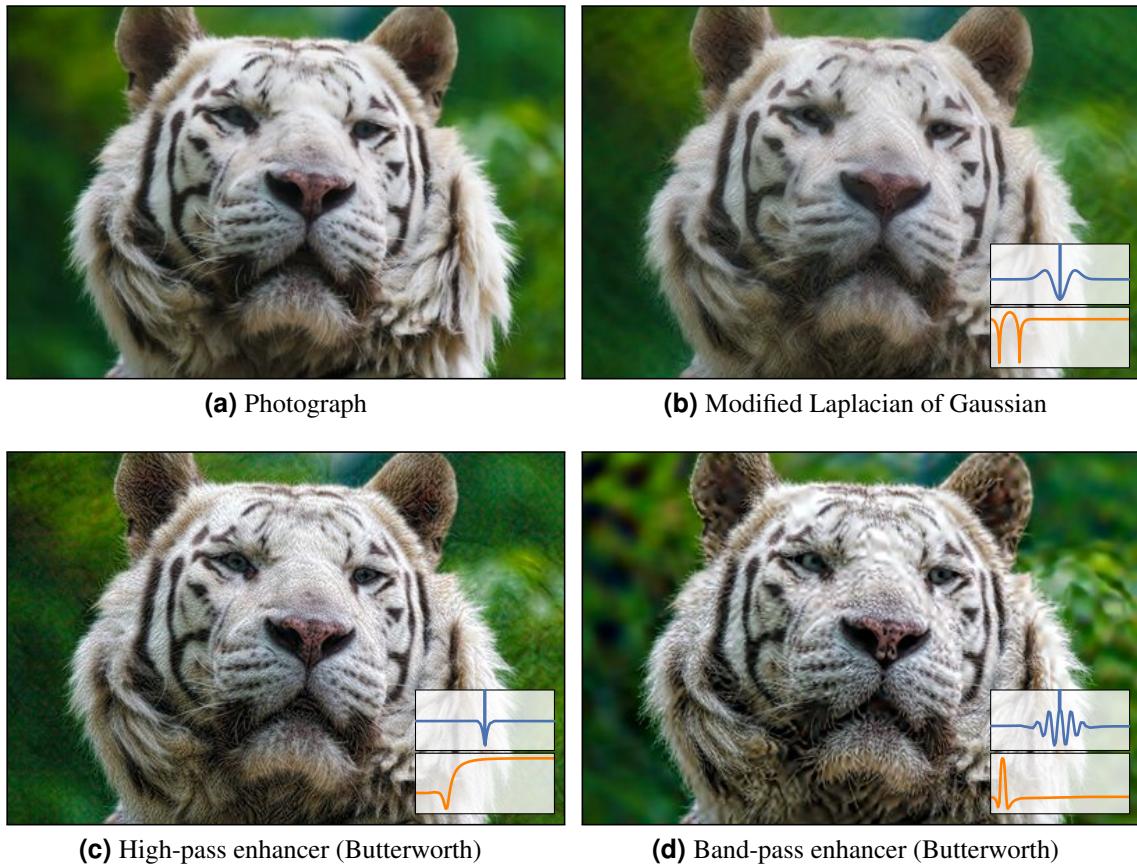


Figure 3. A variety of high-order recursive filters applied by our method for non-uniform filtering [Gastal and Oliveira 2015] to the photograph in (a). The resulting filters preserve the image structure and do not introduce visual artifacts such as halos around objects. The graphs in the insets show the filter's impulse response in blue, and its frequency response (Bode magnitude plot) in orange.

- Gastal, E. S. L. and Oliveira, M. M. (2015). High-order recursive filtering of non-uniformly sampled signals for image and video processing. *Computer Graphics Forum*, 34(2):81–93. Proc. of Eurographics 2015.
- Kopf, J., Cohen, M. F., Lischinski, D., and Uyttendaele, M. (2007). Joint bilateral upsampling. *ACM TOG*, 26:96:1–96:5.
- Lang, M., Wang, O., Aydin, T., Smolic, A., and Gross, M. (2012). Practical temporal consistency for image-based graphics applications. *ACM TOG*, 31(4):34:1–34:8.
- OpenCV (2016). Open Source Computer Vision library 3.0. <http://opencv.org>.
- Smith, S. M. and Brady, J. M. (1997). SUSAN – a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78.
- Winnemöller, H., Olsen, S. C., and Gooch, B. (2006). Real-time video abstraction. *ACM TOG*, 25(3):1226.
- Yang, Q., Tan, K. H., and Ahuja, N. (2009). Real-time O(1) bilateral filtering. In *CVPR*, pages 557–564.

Algoritmos Paralelos Exatos e Otimizações para Alinhamento de Sequências Biológicas Longas em Plataformas de Alto Desempenho

Autor: Edans Flávio de Oliveira Sandes¹, Orientadora: Alba Cristina M. A. de Melo¹

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Caixa Postal 4466 – 70910-900 – Brasília – DF – Brasil

{edans, albamm}@cic.unb.br

1. Introdução

A comparação de sequências biológicas é uma das operações mais básicas e relevantes na Bioinformática, sendo amplamente utilizada para determinar o grau de similaridade entre as sequências [Mount 2004] e inferir características comuns entre espécies [Durbin et al. 2002]. Essa análise é de extrema importância, pois permite, dentre outros benefícios, identificar genes que causam doenças e determinar eventos evolucionários entre organismos. O resultado de uma operação de comparação de sequências biológicas pode ser (a) um escore que indica a similaridade entre as mesmas ou (b) o escore e o alinhamento, onde uma sequência (ou parte dela) é colocada sobre a outra (ou parte dela), de maneira a evidenciar as regiões de similaridades/diferenças [Mount 2004].

O alinhamento de duas sequências biológicas é então definido como um pareamento entre os caracteres das sequências, com a possível inserção de espaços (*gaps*) entre os caracteres. Chamamos de *matches* os pareamentos entre caracteres iguais e de *mismatches* os pareamentos entre caracteres diferentes. A Figura 1 apresenta um exemplo de alinhamento entre duas sequências, onde os *matches* estão representados pelo símbolo ‘:’, *mismatches* pelo símbolo ‘.’ e *gaps* pelo símbolo ‘-’.

```
GCTCACGCCGGTAGTCCCAGCACAGAGGGAG---GAGGCGAACGTATCACCTGAGGTC-----
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
-----GCCTGTAATCCC-GCACTTTGGGAGGCCGAGGTGGGCGCATCAC--GAGGTCAGCGCGAAG
```

Figura 1. Exemplo de alinhamento entre duas sequências.

O problema do alinhamento ótimo de sequências visa encontrar, entre todas as possibilidades de pareamento entre duas sequências, um alinhamento cujo escore seja o maior possível de acordo com o tipo de alinhamento e os parâmetros de pontuação para os *matches*, *mismatches* e *gaps*. Dentre os tipos de alinhamento, o alinhamento *global* considera todos os caracteres de ambas as sequências, o alinhamento *semi-global* permite ignorar o início ou o final de no mínimo uma das sequências e o alinhamento *local* considera *substrings* das sequências. Para a pontuação de *gap*, os modelos mais comuns são o *linear gap* (penalidade proporcional ao comprimento do *gap*) e o *affine gap* (a abertura do *gap* possui uma penalidade adicional) [Durbin et al. 2002].

2. Motivação

Dentre os algoritmos exatos propostos para resolver o problema de alinhamento ótimo de duas sequências biológicas, podemos citar o Needleman-Wunsch

[Needleman and Wunsch 1970] para alinhamento global ótimo, Smith-Waterman [Smith and Waterman 1981] para alinhamento local ótimo e o Gotoh [Gotoh 1982] para alinhamento global ótimo com o modelo de *affine gap*. Os algoritmos exatos que computam o alinhamento ótimo de sequências biológicas calculam, em geral, uma ou mais matrizes de programação dinâmica de tamanho $m \times n$, onde m e n são os tamanhos das sequências comparadas. Sendo assim, esses algoritmos demandam alto poder de processamento e uma grande quantidade de memória. Por este motivo, o uso de algoritmos exatos foi considerado inviável por muito tempo e, com isso, algoritmos heurísticos surgiram para acelerar o procedimento de comparação de sequências, embora sem garantir a produção do resultado ótimo.

O uso de memória dos algoritmos exatos também é um fator limitante, pois a complexidade quadrática de espaço ($O(mn)$) restringe bastante o tamanho das sequências comparadas. Este fato fica evidente quando comparamos sequências muito longas. Por exemplo, para alinhar sequências de um milhão de pares de bases (MBP) precisaríamos de mais de um terabyte de memória. Myers e Miller [Myers and Miller 1988] utilizaram as ideias de Hirschberg [Hirschberg 1975] para reduzir o uso de memória por meio de técnicas de dividir para conquistar, reduzindo o uso de memória para complexidade linear. Ao se utilizar esse tipo de técnica, o tempo de processamento duplica no pior caso.

Para aumentar o desempenho dos algoritmos exatos e, conseqüentemente, reduzir o tempo necessário para encontrar alinhamentos ótimos, utilizam-se de técnicas de paralelismo. Para comparar o desempenho das implementações paralelas de algoritmos de comparação de sequências, convencionou-se o uso da métrica CUPS (*Cells Updated Per Second*), que é calculada dividindo-se o tamanho da matriz de programação dinâmica ($m \times n$) pelo tempo de execução em segundos.

Ao iniciar a presente tese em 2011, o melhor desempenho para comparação de sequências biológicas com algoritmos exatos era de 243 GCUPS em ASIC (*Application Specific Integrated Circuit*) [Sarkar et al. 2010] e o desempenho máximo em GPUs (*Graphics Processing Units*) era de 29,7 GCUPS [Liu et al. 2010]. Com 29,7 GCUPS, é impraticável comparar sequências de DNA muito longas, tais como 230 MBP \times 230 MBP, pois a comparação demoraria cerca de 20 dias. Com o desempenho de 243 GCUPS em projeto ASIC, a mesma comparação demoraria menos tempo, mas ainda levaria cerca de 2 dias e meio. Adicionalmente, a maioria das implementações impõe um limite para o tamanho da sequência de busca, o que significa que elas não são capazes de alinhar sequências muito longas. Por exemplo, a maioria das implementações não aceita sequências maiores que 60 mil caracteres, sendo que alguns cromossomos humanos possuem mais de 200 milhões de caracteres.

A principal motivação desta tese é, então, de evoluir o estado da arte de forma que o alinhamento ótimo de duas sequências longas de DNA possa ser executado em tempo viável, permitindo que cromossomos completos sejam comparados em poucas horas ou até mesmo em minutos. Visto que as ferramentas existentes não são capazes de produzir em pouco tempo os alinhamentos ótimos de sequências com mais de 200 milhões de bases, os biólogos ficam limitados ao uso de métodos heurísticos tanto para comparação como para a geração do alinhamento de sequências longas. As ferramentas propostas por meio dessa tese poderão ser utilizadas por pesquisadores para complementar as análises já efetuadas na literatura, mas considerando métodos exatos em vez de heurísticos.

3. Objetivos

A presente tese de Doutorado possui o objetivo geral de desenvolver algoritmos e otimizações que permitam que o alinhamento ótimo de sequências muito longas de DNA seja obtido em tempo reduzido em plataformas de alto desempenho. Consideramos que o processamento é realizado em tempo reduzido se duas sequências na ordem de 200 milhões de pares de base forem alinhadas em menos de duas horas, algo ainda não visto na literatura com métodos exatos.

A principal plataforma escolhida para o desenvolvimento desta tese foi a arquitetura CUDA (*Compute Unified Device Architecture*), das placas de processamento gráfico da NVIDIA. Entretanto, ao longo da pesquisa, o requisito de heterogeneidade foi levado em consideração, de forma que outras plataformas pudessem ser utilizadas para acelerar ainda mais o processamento dos algoritmos envolvidos.

Os objetivos específicos da tese estão elencados a seguir:

- Desenvolver e avaliar de algoritmos paralelos que permitam a recuperação eficiente de alinhamentos ótimos de sequências de DNA longas em GPUs;
- Propor e avaliar teoricamente de método para redução do espaço de busca dos algoritmos propostos, sem prejuízo do resultado ótimo;
- Propor, implementar e avaliar estratégia para execução com múltiplas GPUs dos algoritmos propostos;
- Propor uma arquitetura de *software* que permita a execução dos algoritmos propostos em plataformas homogêneas ou heterogêneas compostas por *multicores* ou GPUs, recuperando alinhamentos ótimos locais, globais e semiglobais;

4. Contribuições

Como contribuição desta tese, algoritmos paralelos e otimizações foram desenvolvidos para permitir a recuperação de alinhamentos ótimos entre sequências de DNA. Esses novos algoritmos foram implementados no *software* CUDAlign, que foi evoluído em diversas versões incrementais. A primeira versão, CUDAlign 1.0, foi desenvolvida na dissertação de mestrado [Sandes 2011] do mesmo autor desta tese. A seguir listamos as evoluções de cada uma das versões propostas no escopo desta tese:

- CUDAlign 2.0 [Sandes and Melo 2011]: versão capaz de recuperar alinhamentos locais ótimos entre cromossomos completos em uma GPU utilizando seis estágios (Figura 2). Isso foi possível devido às otimizações propostas nessa tese (*matching* baseado em objetivo, execução ortogonal e divisão balanceada). Atingiu-se 23,1 GCUPS em uma GPU;
- CUDAlign 2.1 [Sandes and Melo 2013]: versão que propõe a otimização *Block Pruning* para alinhamentos locais ótimos, permitindo acelerar o cálculo da matriz de programação dinâmica em mais de 50%, para sequências similares. Nesta otimização, descartam-se áreas da matriz de programação dinâmica que não contribuem para a obtenção do alinhamento ótimo, sem prejuízo do resultado ótimo. Atingiu-se 50,7 GCUPS em uma GPU;
- CUDAlign 3.0 [Sandes et al. 2014b][Sandes et al. 2014a]: versão capaz de comparar sequências longas de DNA em *clusters* com múltiplas GPUs (homogêneas e heterogêneas), permitindo comparar sequências de até 249 milhões de pares de base (MBP). Atingiu-se 1,73 TCUPS (Trilhões de CUPS) com 64 GPUs;

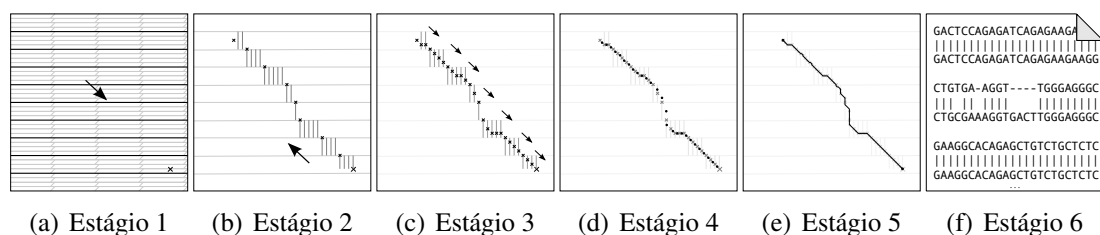


Figura 2. Execução do CUDAAlign 2.0 dividida em seis estágios.

- CUDAAlign 4.0 [Sandes et al. 2016b]: versão que obtém o alinhamento completo em múltiplas GPUs de maneira eficiente. Por meio do mecanismo de *traceback* especulativo proposto, foi possível obter o alinhamento completo de todos os cromossomos homólogos entre o homem e o chimpanzé, totalizando mais de 500 trilhões de células processadas. Atingiu-se 10,37 TCUPS com 384 GPUs.

Adicionalmente, as seguintes contribuições foram apresentadas nesta tese.

- Análise teórica da otimização *Block Pruning*: a análise matemática do *Block Pruning* foi feita para identificar quais os aspectos que contribuem ou prejudicam a eficácia deste método;
- Método de balanceamento dinâmico de carga [Sandes et al. 2014c]: para que ambientes distribuídos e não dedicados sejam utilizados para execução do CUDAAlign, é necessário que haja um balanceamento dinâmico de carga eficiente. Nesta tese, propusemos um método de balanceamento de carga baseado em agentes, sem necessidade de um elemento central coordenador. Este método foi testado em ambiente simulado, mostrando ser bastante efetivo para o uso proposto;
- Arquitetura MASA [Sandes et al. 2016c]: A arquitetura MASA (*Multi-platform Architecture for Sequence Aligners*) foi projetada para simplificar a portabilidade do CUDAAlign para outras arquiteturas. O MASA suporta alinhamentos ótimos locais, globais e semi-globais. Nesta tese, a arquitetura MASA foi aplicada em 4 plataformas distintas: GPU (CUDA), CPU (OpenMP e OmpSs) e Intel Phi (OpenMP). Como contribuição indireta desta tese, a arquitetura MASA foi aplicada na arquitetura OpenCL no escopo da dissertação de mestrado do aluno Marco Antônio C. de Figueiredo Jr. [de Figueiredo Jr. 2015], do mesmo grupo de trabalho do autor desta tese. Neste trabalho, atingiu-se 179,2 GCUPS em uma única GPU da AMD [de Figueiredo Jr. et al. 2015].

A Tabela 1 apresenta um resumo com o desempenho máximo obtido nos testes de cada uma das versões. Com os resultados apresentados, conseguiu-se evoluir o estado da arte em dois aspectos: a) o maior desempenho conhecido para alinhamento ótimo de seqüências biológicas tornou-se 10,37 TCUPS com 384 GPUs; b) até onde sabemos, a maior seqüência alinhada com métodos exatos passou a ser de 249 MBP.

4.1. Artigos publicados ou aceitos para publicação

O autor desta tese publicou 8 artigos dentro do escopo da tese, todos eles como autor principal. Destes, 5 foram artigos completos em periódicos internacionais, 2 artigos em conferências internacionais e 1 artigo resumido em conferência internacional. As qualificações dos artigos publicados encontram-se abaixo.

Tabela 1. Desempenho das versões do CUDAlign propostas nesta tese

Versão	Saída	Desempenho	Ambiente	Tam. Máx.
CUDAlign 2.0	Alinhamento	23,1 GCUPS	1 × GTX 285	47 MBP
CUDAlign 2.1	Alinhamento	50,7 GCUPS	1 × GTX 560	59 MBP
CUDAlign 3.0	Escore	1,73 TCUPS	64 × M2090	249 MBP
CUDAlign 4.0	Alinhamento	10,37 TCUPS	384 × M2090	249 MBP

Periódicos internacionais:

- [Sandes and Melo 2013]: IEEE Transactions on Parallel and Distributed Systems (TPDS) – Capes Qualis CC - A1 (JCR 2,173).
- [Sandes and Melo 2013]: Expert Systems with Applications (ESWA) – Capes Qualis CC - A1 (JCR 1,965)
- [Sandes et al. 2016c]: ACM Transactions on Parallel Computing (TOPC) – (periódico novo, ainda sem JCR)
- [Sandes et al. 2016b]: IEEE Transactions on Parallel and Distributed Systems (TPDS) – Capes Qualis CC - A1 (JCR 2,173): *Paper aceito, DOI=10.1109/TPDS.2016.2515597, aguardando publicação.*
- [Sandes et al. 2016a]: ACM Computing Surveys: Capes Qualis CC - A1 (JCR 4,043): *Paper aceito, aguardando publicação.*

Conferências internacionais:

- [Sandes and Melo 2011]: IEEE International Parallel & Distributed Processing Symposium (IPDPS) – Capes Qualis CC - A1.
- [Sandes et al. 2014b]: IEEE/ACM Symposium on Cluster, Cloud and Grid Computing (CCGrid) – Capes Qualis CC - A1.
- [Sandes et al. 2014a]: Symposium on Principles and Practice of Parallel Programming (PPoPP) – Capes Qualis CC - A2: (*artigo resumido*).

Referências

- de Figueiredo Jr., M. A. C. (2015). MASA-OpenCL: Comparação Paralela de Sequências Biológicas Longas em GPU. Master's thesis, Universidade de Brasília, Brasília, Brasil.
- de Figueiredo Jr., M. A. C., Sandes, E. F. O., and Melo, A. C. M. A. (2015). Parallel megabase dna sequence comparison with opencl. In *22st International Conference on High Performance Computing, HiPC*, pages 436–445.
- Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (2002). *Biological sequence analysis*. Cambridge University Press.
- Gotoh, O. (1982). An improved algorithm for matching biological sequences. *Journal of Molecular Biology*, 162(3):705–708.
- Hirschberg, D. S. (1975). A linear space algorithm for computing maximal common subsequences. *Communications of the ACM*, 18(6):341–343.
- Liu, Y., Schmidt, B., and Maskell, D. (2010). CUDASW++2.0: enhanced Smith-Waterman protein database search on CUDA-enabled GPUs based on SIMD and virtualized simd abstractions. *BMC Research Notes*, 3(1):93.

- Mount, D. M. (2004). *Bioinformatics - sequence and genome analysis (2. ed.)*. Cold Spring Harbor Laboratory Press.
- Myers, E. W. and Miller, W. (1988). Optimal alignments in linear space. *Computer applications in the Biosciences*, 4(1):11–17.
- Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453.
- Sandes, E. F. O. (2011). Comparação Paralela de Sequências Biológicas Longas utilizando Unidades de Processamento Gráfico (GPUs). Master's thesis, Universidade de Brasília, Brasília, Brasil.
- Sandes, E. F. O., Boukerche, A., and Melo, A. C. M. A. (2016a). Parallel Exact Pairwise Biological Sequence Comparison: Algorithms, Platforms and Classification. *ACM Computing Surveys (accepted)*.
- Sandes, E. F. O. and Melo, A. C. M. A. (2011). Smith-Waterman alignment of huge sequences with GPU in linear space. In *IEEE International Parallel Distributed Processing Symposium*, pages 1199–1211.
- Sandes, E. F. O. and Melo, A. C. M. A. (2013). Retrieving smith-waterman alignments with optimizations for megabase biological sequences using gpu. *IEEE Transactions on Parallel and Distributed Systems*, 24(5):1009–1021.
- Sandes, E. F. O., Miranda, G., , Melo, A. C. M. A., Martorell, X., and Ayguadé, E. (2014a). Fine-grain parallel megabase sequence comparison with multiple heterogeneous GPUs. In *Proceedings of the 19th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, PPOPP '14, pages 383–384 (short paper).
- Sandes, E. F. O., Miranda, G., Martorell, X., Ayguadé, E., Teodoro, G., and Melo, A. C. M. A. (2016b). CUDAlign 4.0: Incremental Speculative Traceback for Exact Chromosome-Wide Alignment in GPU Clusters. *IEEE Transactions on Parallel and Distributed Systems*, PP(99):1–1.
- Sandes, E. F. O., Miranda, G., Martorell, X., Ayguadé, E., Teodoro, G., and Melo, A. C. M. A. (2016c). MASA: a multiplatform architecture for sequence aligners with block pruning. *ACM Transactions on Parallel Computing*, 2(4):28:1–28:31.
- Sandes, E. F. O., Miranda, G., Melo, A. C. M. A., Martorell, X., and Ayguade, E. (2014b). CUDAlign 3.0: Parallel Biological Sequence Comparison in Large GPU Clusters. In *IEEE/ACM Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, pages 160–169.
- Sandes, E. F. O., Ralha, C. G., and Melo, A. C. M. A. (2014c). An agent-based solution for dynamic multi-node wavefront balancing in biological sequence comparison. *Expert Systems with Applications*, 41(10):4929 – 4938.
- Sarkar, S., Kulkarni, G., Pande, P., and Kalyanaraman, A. (2010). Network-on-chip hardware accelerators for biological sequence alignment. *IEEE Transactions on Computers*, 59(1):29–41.
- Smith, T. F. and Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197.

Content Delivery in Vehicular Ad Hoc Networks

Fabício A. Silva¹,
Linnyer B. Ruiz² (Advisor), Antonio A. F. Loureiro¹ (Co-Advisor)

¹Departamento de Ciência da Computação (UFMG)

²Departamento de Informática (UEM)

Abstract. *Vehicular Ad Hoc Network (VANET) applications are emerging to reality with the objective of making traffic safer, less congested, more informative, and enjoyable. With the advance of such applications, comes the need for content delivery solutions, since the efficient delivery of content is a fundamental requirement for most vehicular network applications. The main objective of this thesis is to investigate how existing concepts can be adapted and applied to VANETs and to propose content delivery solutions for this kind of network.*

1. Introduction

Vehicular Ad Hoc Networks (VANETs) are migrating from theory to practice mainly due to the great interest of vehicle's manufacturers to provide new on-road services to their clients [Zeadally et al. 2012]. This kind of network consists of vehicles with on-board wireless communication facilities that are able to establish ad hoc communication with their peers as well as with infrastructure stations [Wang and Li 2009]. VANET applications are emerging to reality with the objective of making traffic safer, less congested, more informative, and enjoyable [Lee et al. 2014, Costa-Montenegro et al. 2012]. With the advance of such applications, comes the need for content delivery solutions, since the efficient delivery of content is a fundamental requirement for most vehicular network applications, such as information, advertisement, and entertainment systems. Nevertheless, the VANET particular characteristics make the task of delivering content easier said than done [Gerla et al. 2014].

Usually, two approaches have been used to deliver content in the traditional Internet [Passarella 2012]: Content Delivery Networks (CDN) and Peer-to-Peer (P2P). Several characteristics of VANETs and their applications suggest that pure CDN and P2P models, as originally conceived for the Internet, are not suitable for them. First, many applications are referred to as push-based, meaning that content should be pushed to the clients even in the absence of a request, like an accident notification, for example. In addition, contacts in VANETs are intermittent, making the establishment and maintenance of end-to-end links very difficult. Furthermore, the deployment of surrogate servers in a large-scale urban scenario is a costly, time-consuming task. When it comes to content, several entities are potential sources, such as vehicles, Intelligent Transportation System (ITS) solutions, Wireless Sensor Networks (WSNs), mobile users, among others. Moreover, content sources and clients move in considerable speed, causing constant changes in the network topology. Finally, content in many applications is location- and time-dependent, meaning it is valid only inside a region of interest (RoI) during a given period.

1.1. Objective and Contributions

Given the need for more efficient vehicular content delivery solutions, the hypothesis of the thesis is defined as: the adoption of a hybrid model that inherits and adapts concepts

from both CDN and P2P will improve the performance of VANET applications in terms of content availability, quality of experience, and delivery cost. Therefore, the main objective is to investigate how CDN and P2P concepts can be adapted and applied to VANETs, forming a Vehicular Content Delivery Network (VCDN), and propose solutions that validate the aforementioned hypothesis.

The main contributions of the thesis are summarized as follows:

1. A novel content delivery model, called VCDN, that inherits and adapts concepts from CDN and P2P, and a framework to help application designers to define the most appropriate architecture and techniques to be adopted by their applications;
2. Two innovative solutions based on our proposed VCDN model developed to make content available in city-wide and region-wide scenarios. These solutions, in addition to validate our hypothesis, advance the state-of-the-art since they are based on novel methods that consider origin-destination points as input;
3. Characterization results that lead to important insights on how to select appropriate replica vehicles in content delivery applications;
4. A survey of VANET content delivery solutions found in the literature organized according to their techniques and architectures.

2. Background

In Chapter 2 of the thesis, we present an in-depth survey of the literature in terms of content delivery concepts applied to VANETs. Basically, there are two major challenges when designing content delivery solutions [Peng 2004]. The first one is the issue of selecting good places to replicate content. The second challenge is related to the discovery of the most appropriate server, either origin or surrogate, and the content delivery itself. These issues are the two fundamental building blocks that compose a content delivery system, and are even more challenging in such dynamic environments as VANETs.

In the thesis, we classify the studies from the literature under their solution basis and architecture perspectives, as shown in Figure 1. This detailed survey guided us to a promising challenge in the content delivery for vehicular networks, which is the proposing of a novel model that considers the VANET characteristics.

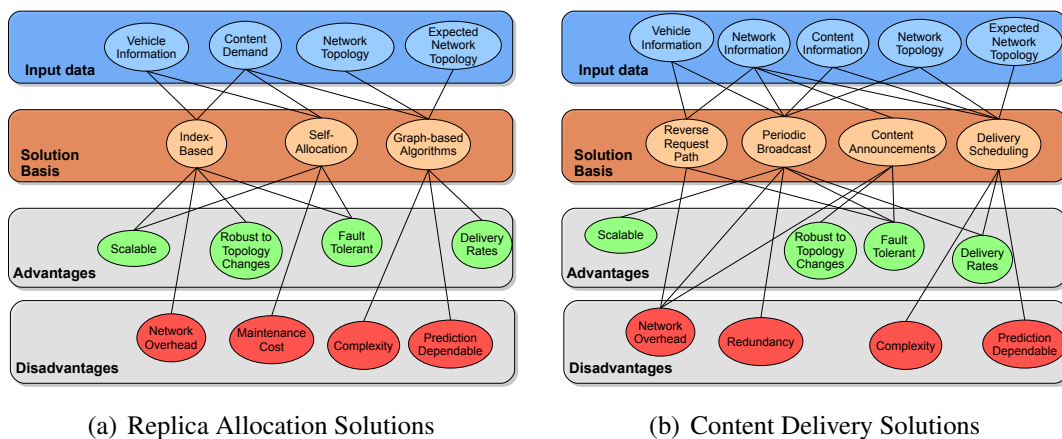


Figure 1. Replica allocation and content delivery solutions classification.

This is the first contribution of the thesis reported in the following publications:

1. [Journal A1, Impact Factor: 3.37] **Fabrcio A. Silva**; Azzedine Boukerche; Thais R. M. Braga Silva; Linnyer B. Ruiz; Eduardo Cerqueira; Antonio A. F. Loureiro, "Vehicular Networks: A New Challenge for Content Delivery-based Applications". *ACM Computing Surveys* (To Appear)
2. [Conference] **Fabrcio A. Silva**; Azzedine Boukerche; Thais R. M. B. Silva; Linnyer B. Ruiz; Eduardo Cerqueira; Antonio A. F. Loureiro, "Content Replication and Delivery in Vehicular Networks". *4th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, (2014):127-132.

3. Vehicular Content Delivery Model

As observed in the results of the survey of the state-of-the-art, there is no model to help designers when developing VANET applications with content delivery demand. Therefore, in Chapter 3 of the thesis we propose a novel model called Vehicular Content Delivery Network (VCDN), where aspects of CDN and P2P are integrated, adapted, and extended to VANETs.

The original Content Delivery Network (CDN) and Peer-to-Peer (P2P) solutions have advantages and disadvantages. The main advantage of the pure CDN approach, considering the VANET context, is the infrastructured network comprised of the surrogate servers, leading to highly content availability. However, there is a high cost, in terms of money and time, to deploy and maintain the surrogate servers considering the large-scale scenario of VANETs. Furthermore, this approach is not fault-tolerant since the failure of a surrogate server may cause disconnections and, consequently, uncovered regions.

On the other hand, the main advantage of such pure P2P approach is its fault-tolerance, since the failure of a provider is not perceived given the existence of other servers nearby. Nevertheless, there is a high cost in searching for content in terms of delay and network overhead, given the highly dynamic topologies posed by VANETs and the poor content availability that may occur.

Our proposed model takes advantage of the benefits from both CDN and P2P approaches, and extends some of their concepts considering the VANET's characteristics. On one hand, we exploit the idea of replicating content on surrogate servers running on Roadside Units (RSUs) to increase content availability. Furthermore, we extend such concept by allowing content to be replicated also in moving vehicles as well, which increases content availability and fault-tolerance. On the other hand, P2P concepts for distributed content discovery and delivery are exploited, together with incentive mechanisms that benefit cooperative peers. To reduce the overhead and time to find a content in the network, our model takes advantage of the infrastructured stations to track the content availability nearby and, then, to indicate potential providers to clients. To help the content replication and delivery tasks, our model includes a mobility management service that monitors mobility patterns and supports the decision of content replication, discovery, and delivery.

Figure 2 presents the VCDN system architecture and components. The VCDN components encompass all aspects of the overall content delivery problem for VANETs. In summary, our VCDN model is scalable, fault-tolerant, mobility-aware, and works even in the absence of infrastructure stations.

This is the second contribution of the thesis reported in the following publications:

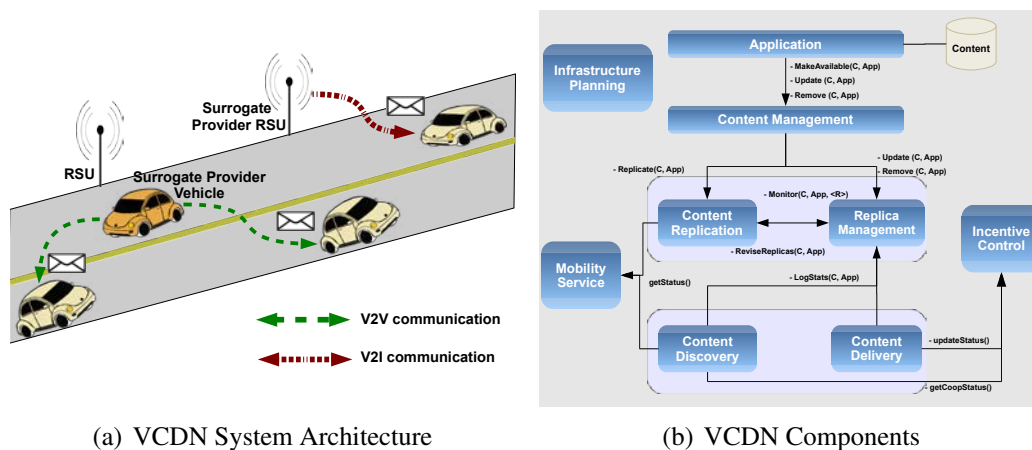


Figure 2. VCDN System Architecture and Components.

3. [Journal A1, Impact Factor: 5.41] **Fabrcio A. Silva**; A. Boukerche; Thais R. M. B. Silva; Linnyer B. Ruiz; E. Cerqueira; Antonio A. F. Loureiro, "Vehicular Content Delivery Networks". *IEEE Wireless Communications Magazine* (To Appear)
4. [Conference B2] **Fabrcio A. Silva**; Thais R. M. Braga Silva; Linnyer B. Ruiz; Eduardo Cerqueira; Antonio A. F. Loureiro, "Um modelo hbrido para entrega de contedo em redes veiculares". *Simpósio Brasileiro de Redes de Computadores*

4. Evaluation

We adopt the proposed model in real VANET applications with the objective of validating its benefits. However, before that, we present vehicular mobility characterization results to give insights on how to implement the VCDN building blocks.

In Chapter 4 of the thesis, we characterize a realistic, large-scale vehicular mobility trace under two perspectives: graph-based and origin-destination-based. In summary, the results reveal that some vehicles are more likely to be better replicas than others. Moreover, it was possible to infer a model to represent macroscopic aspects related to origin-destination (O-D) points. Given the difficulty to obtain a complete contact graph of a VANET, we argue that O-D points, which are easily obtained from navigation systems, can be useful for helping the design of content delivery solutions.

Therefore, in Chapter 5 of the thesis we implement the proposed VCDN model for two different applications: city-wide and region-wide. In the first, the content should be delivered to all vehicles in the network. In the second, on the other hand, the content should be delivered only to vehicles that are inside a region of interest during the content lifetime. Both solutions rely on the characterization results obtained in terms of O-D points.

We compare our solutions with others from the literature under important metrics to content delivery applications, such as coverage, the time a vehicle has to wait before being covered, redundancy, number of messages, among others. Figure 3 illustrates some of the results for the region-wide scenario. In summary, our solution (GO-DCR) leads to a higher coverage and a lower time to be covered (i.e., vehicles receive content earlier than in the baseline solutions). In addition, our solution leads to lower communication cost since it requires less vehicle-to-infrastructure (V2I) messages.

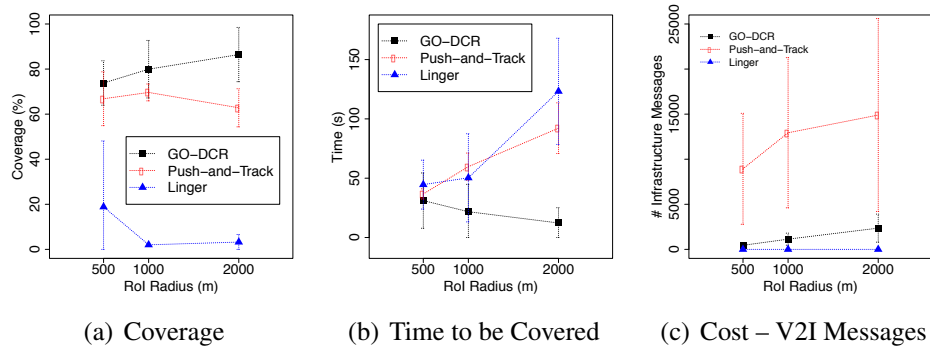


Figure 3. Sample results of the region-wide scenario.

In general, for both scenarios, our VCDN model was able to improve the content availability while reducing the communication cost. In addition to validate our proposed VCDN model, the contribution of this chapter also lies in the advance in the state-of-the-art in concerns with content replication and delivery solutions for VANET applications.

The results of this part of the thesis were reported in the following publications:

5. [Journal B1, Impact Factor: 1.94] **Fabrício A. Silva**; Azzedine Boukerche; Thais R. M. B. Silva; Linnyer B. Ruiz; Antonio A. F. Loureiro, "Geo-localized Content Availability in VANETs". *Ad Hoc Networks (Elsevier)*, 36(2):425-434, Jan. 2016.
6. [Journal A1, Impact Factor: 2.64] **Fabrício A. Silva**; Azzedine Boukerche; Thais R. M. B. Silva; Fabrício Benevenuto; Linnyer B. Ruiz; Antonio A. F. Loureiro, "ODCRep: Origin-Destination-based Content Replication for Vehicular Networks". *IEEE Transactions on Vehicular Technology* 64(12):1-12, Dec. 2015.
7. [Journal A2, Impact Factor: 1.28] **Fabrício A. Silva**; A. Boukerche; Thais R. M. B. Silva; Linnyer B. Ruiz; Antonio A. F. Loureiro, "A novel macroscopic mobility model for vehicular networks". *Computer Networks*, 79(C):188-202, March 2015.
8. [Conference B1] **Fabrício A. Silva**; Clayson Celes; Azzedine Boukerche; Linnyer B. Ruiz; Antonio A. F. Loureiro, "Filling the Gaps of Vehicular Mobility Traces". *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*,(2015):47-54.
9. [Conference A2] **Fabrício A. Silva**; Azzedine Boukerche; Thais R. M. B. Silva; Linnyer B. Ruiz; Antonio A. F. Loureiro, "Geo-localized Content Replication for Vehicular Ad-hoc Networks". *IEEE Symposium on Computers and Communications (ISCC)*,(2015):695-700.
10. [Conference A2] **Fabrício A. Silva**; Linnyer B. Ruiz; Antonio A. F. Loureiro, "Content Replication in Mobile Vehicular Ad-hoc Networks". *IEEE Mobile Data Management (MDM)*,(2015):26-29.
11. [Conference A2] **Fabrício A. Silva**; Thais R. M. B. Silva; Fabrício Benevenuto; Linnyer B. Ruiz; Antonio A. F. Loureiro, "Improving information dissemination in vehicular networks by selecting appropriate disseminators". *28th IEEE International Conference on Advanced Information Networking and Applications (AINA)*,(2014):681-688.
12. [Conference A1] **Fabrício A. Silva**; Thais R. M. B. Silva; Linnyer B. Ruiz; Antonio A. F. Loureiro; Rafael Vicente, "On the improvement of vehicular macro-

- scopic mobility models". *IEEE 78th Vehicular Technology Conference (VTC-Fall)*, (2013):1-5.
13. [Conference A1] **Fabrcio A. Silva**; Thais R. M. B. Silva; Linnyer B. Ruiz; Antonio A. F. Loureiro, "ConProVA: A Smart Context Provisioning Middleware for VANET Applications". *IEEE 77th Vehicular Technology Conference (VTC-Spring)*, (2013):1-5.
 14. [Conference B2] **Fabrcio A. Silva**; Thais M. B. Silva; Linnyer B. Ruiz; Antonio A. F. Loureiro, "Um Middleware para provisionamento de contextos para redes veiculares". *Simpósio Brasileiro de Redes de Computadores*, (2013):615-628.

5. Final Remarks and Impact

In the thesis, we explored the field of content delivery for vehicular ad hoc networks (VANETs), with the hypothesis that concepts inherited from Content Delivery Network (CDN) and Peer-to-Peer (P2P) and adapted to VANET applications would improve their performance. Given that, the objective was to investigate how those concepts could be applied to VANETs, and propose solutions that validate this hypothesis.

The impact of the thesis can be observed by the relevant results reported in very qualified journal and conference publications, as presented in the previous sections. All such publications are expected to contribute significantly with other researchers in this field. We highlight the work to appear in the *ACM Computing Surveys* (Impact Factor 3.37), which demonstrates our familiarity with the state-of-the-art in the area. We also state the importance of the other three journal papers published in the *Ad Hoc Networks* (Impact Factor 1.94), *IEEE Transactions on Vehicular Technology* (Impact Factor 2.64) and *Computer Networks* (Impact Factor 1.28), which reinforce that our work has been well accepted by the research community. Finally, we should mention the last journal paper that describes the overall VCDN model to appear in the *IEEE Wireless Communications Magazine* (Impact Factor: 5.41).

References

- Costa-Montenegro, E., Quiñoy-García, F., González-Castaño, F., and Gil-Castineira, F. (2012). Vehicular entertainment systems: Mobile application enhancement in networked infrastructures. *IEEE Vehicular Technology Magazine*, 7(3):73–79.
- Gerla, M., Wu, C., Pau, G., and Zhu, X. (2014). Content distribution in VANETs. *Vehicular Communications*, 1(1):3–12.
- Lee, E., Lee, E.-k., and Gerla, M. (2014). Vehicular Cloud Networking: Architecture and Design Principles. *IEEE Communications Magazine*, (2):148–155.
- Passarella, A. (2012). A survey on content-centric technologies for the current internet: Cdn and p2p solutions. *Computer Communications*, 35(1):1–32.
- Peng, G. (2004). CDN: Content Distribution Network. *arXiv preprint cs/0411069*.
- Wang, Y. and Li, F. (2009). Vehicular ad hoc networks. In *Guide to wireless ad hoc networks*, pages 503–525. Springer.
- Zeadally, S., Hunt, R., Chen, Y.-S., Irwin, A., and Hassan, A. (2012). Vehicular ad hoc networks (VANETS): status, results, and challenges. *Telecommunication Systems*, 50(4):217–241.

Técnicas de Projeção para Identificação de Grupos e Comparação de Dados Multidimensionais Usando Diferentes Medidas de Similaridade

Paulo Joia¹, Luis Gustavo Nonato¹

¹Instituto de Ciências Matemáticas e de Computação (ICMC) – USP, São Carlos, Brasil

{pjoia, gnonato}@icmc.usp.br

Abstract. *This work explores the potential of the projection techniques to solve problems related to: **clustering** and **similarity search** in multidimensional data. For clustering data, a local and interactive projection technique capable of projecting data with effective preservation of distances was developed and also a new clustering method, which operates in the visual space, ensuring that clusters are not fragmented during the visualization. For the similarity search, we build a family of class-specific metrics and the fuzzy set theory was used to estimate a degree of uncertainty that is embedded in the metric, increasing its precision. The results confirm the effectiveness of the developed techniques, which represent significant contributions for this investigation area.*

Resumo. *Este trabalho explora o potencial das técnicas de projeção para resolver problemas relacionados à: **identificação de agrupamentos** e **busca por similaridade** em dados multidimensionais. Para identificação de agrupamentos foi desenvolvida uma técnica de projeção local e interativa capaz de projetar dados com ótima preservação de distâncias, além de um novo método para identificação de agrupamentos, o qual opera no espaço visual, garantindo que os grupos obtidos não fiquem fragmentados durante a visualização. Para as buscas por similaridade em dados multidimensionais, uma família de métricas baseada em classes foi construída e a teoria dos conjuntos fuzzy foi usada para estimar um valor de incerteza que é transferido para a métrica, aumentando sua precisão. Os resultados confirmam a efetividade das técnicas desenvolvidas, as quais representam significativa contribuição nesta área de investigação.*

1. Introdução

Visualização de informação desempenha um papel importante na organização e exploração de dados multidimensionais, graças à capacidade de percepção visual do sistema cognitivo humano, assim como técnicas de projeção, as quais reduzem a dimensionalidade do conjunto de dados permitindo visualizar informações muitas vezes ocultas na alta dimensão. Este trabalho explora o potencial das técnicas de visualização de informação com ênfase em projeção para auxiliar na *identificação de agrupamentos* e *busca por similaridade em dados multidimensionais*.

1.1. Motivação

Consideráveis avanços têm sido observados nas técnicas de projeção nos últimos anos, com aplicações em diferentes domínios. Algumas até propõem soluções para os problemas de identificação de agrupamentos e busca por similaridade em dados multidimensionais, contudo, longe da solução ideal. Por exemplo, identificar grupos é uma tarefa

complexa e na maioria das vezes os grupos obtidos não correspondem à verdadeira natureza dos dados, normalmente organizados com base na geometria. Quanto à busca por similaridade, existem várias abordagens, todavia, poucas empregam medidas de similaridade realmente aptas a discriminar objetos de acordo com as classes existentes.

2. Contribuições

A tese de doutorado aqui apresentada, produziu resultados significativos na área de visualização de informação, com ênfase em projeção multidimensional. Resultados que podem ser comprovados pelas novas técnicas desenvolvidas e respectivas metodologias empregadas no desenvolvimento, conforme sumarizado abaixo. Para um exame mais detalhado sobre cada técnica, incluindo sua formulação matemática, algoritmo, resultados e comparações com outras técnicas, recomendamos que o leitor consulte [Joia 2015].

2.1. Técnica de Projeção Local Interativa

Local Affine Multidimensional Projection (LAMP) [Joia et al. 2011a] permite manipular pontos de controle no espaço visual de modo a organizá-los, possibilitando ao usuário guiar a projeção, porém, com uma grande vantagem sobre as demais técnicas que se apoiam em subconjunto de amostras: requer um número muito reduzido de pontos de controle como entrada, tornando-se ideal para aplicações interativas. Tem formulação matemática baseada em mapeamentos ortogonais, garantindo ótima preservação de distâncias durante a projeção multidimensional e, não depende de grafos de vizinhança para construir o mapeamento. É altamente precisa, com baixo custo computacional, apropriada para aplicações interativas envolvendo grandes volumes de dados. LAMP está atualmente entre as técnicas do estado da arte em relação à preservação de distâncias e eficiência computacional, além de permitir projetar dados explorando tanto relações globais como locais entre instâncias, de maneira efetiva. A Figura 1 mostra o potencial da LAMP ao estabelecer uma correlação visual entre conjuntos de dados, a princípio, sem qualquer conexão.

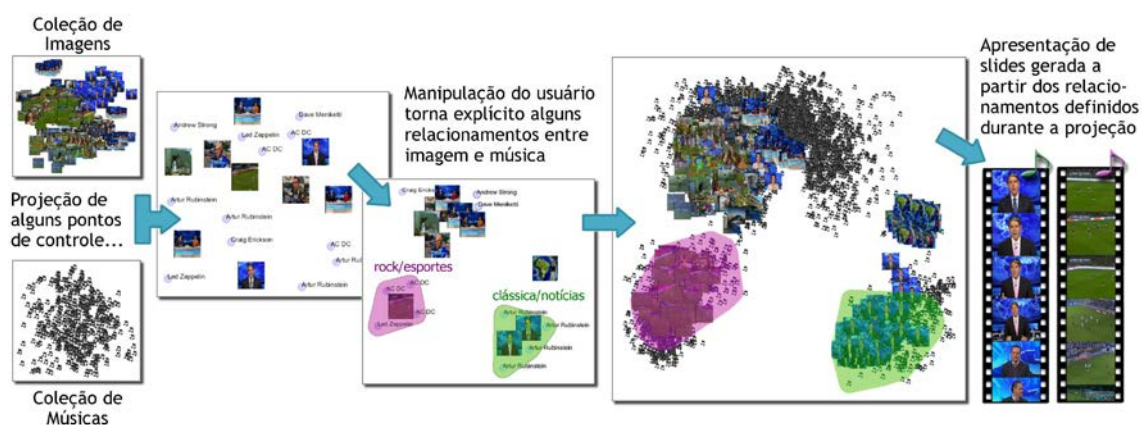


Figura 1. Utilizando a LAMP para correlacionar dados de diferentes naturezas. Inicialmente, uma projeção é criada para cada conjunto de dados, a partir de algumas amostras. A correlação entre as amostras é definida pelo usuário, agrupando objetos no espaço visual (imagens e músicas). Em seguida, os dados são projetados segundo as associações criadas pelo usuário. Por fim, as listas de objetos associados são usadas para criar uma apresentação de slides onde imagens e músicas são reproduzidas de forma sincronizada.

2.2. Método para Identificação de Grupos com Base em Projeção

Column Selection Method (CSM) [Joia et al. 2015], um método de visualização apoiado em projeção multidimensional que permite agrupar dados. CSM opera no espaço visual, garantindo que os grupos obtidos não fiquem fragmentados durante a visualização. É orientado por um mecanismo de amostragem determinístico, capaz de identificar instâncias representativas que correspondem a um certo padrão nos dados. O mecanismo de amostragem é baseado em decomposição matricial (SVD) e capaz de operar mesmo em conjuntos de dados desbalanceados. Além de identificar instâncias representativas, o mecanismo de amostragem pode ser ajustado para identificar os atributos mais relevantes de cada agrupamento obtido. Portanto, em um único *framework*, três tarefas são contempladas: *amostragem de dados*, *detecção de agrupamentos* e *seleção de atributos*. A Figura 2 ilustra a metáfora visual utilizada pela CSM para representar grupos e atributos, por meio de superfícies e nuvens de palavras.

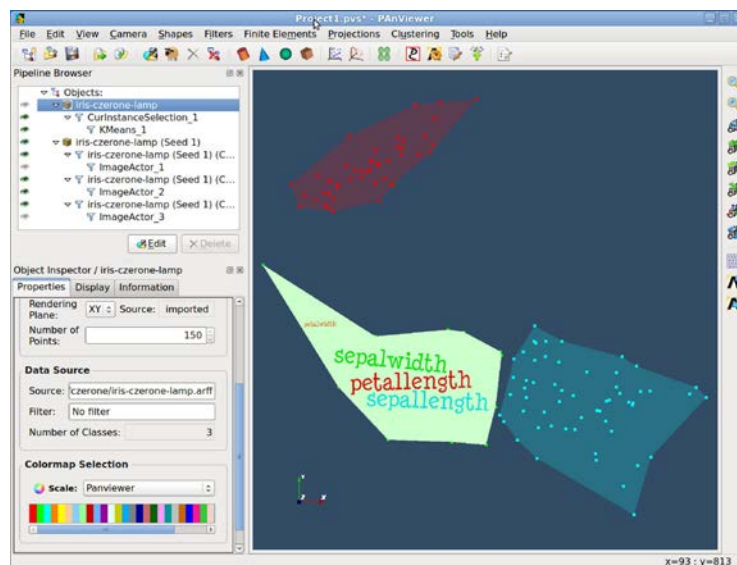


Figura 2. Metáfora visual utilizada pela CSM para representar grupos e atributos.

2.3. Família de Métricas Classes-Específicas

Muitas técnicas propõem medidas de similaridade para comparar dados multidimensionais, mas nenhuma diretamente relacionada às classes de objetos existentes no conjunto de dados. A *Class-Specific Multidimensional Projection* (CSMP) [Joia et al. 2012] é uma técnica de projeção baseada em uma família de métricas específicas por classe para projetar e comparar dados multidimensionais. As métricas são obtidas pela seleção dos atributos que melhor representam cada classe do conjunto de dados, de modo a minimizar a dissimilaridade entre pares de objetos pertencentes à mesma classe e, ao mesmo tempo, maximizá-la para objetos pertencentes a classes distintas. As métricas classes-específicas são avaliadas no contexto de recuperação de imagens por conteúdo para encontrar imagens similares a uma dada imagem de consulta. A lista de imagens similares pode ser retornada pelo sistema ou selecionada diretamente pelo usuário, a partir do *layout* da projeção, conforme exemplificado na Figura 3.

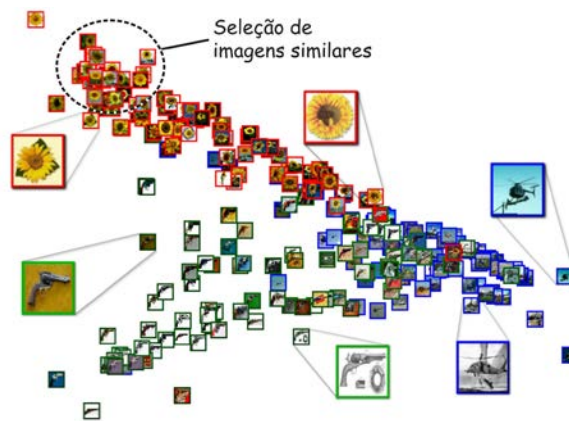


Figura 3. Seleção de imagens similares com o uso da CSMP.

2.4. Cálculo de Incerteza na Família de Métricas Classes-Específicas

Com o intuito de aumentar a precisão da família de métricas classes-específicas empregada na CSMP, uma nova técnica denominada *Class-Specific with Weight Image Retrieval* (CSWIRe) foi desenvolvida. Nesta abordagem, o usuário constrói um modelo a partir de um subconjunto de imagens, denominado “*modelo de classes*”. A seguir, um classificador é aplicado sobre este modelo, retornando as melhores características e pesos que representam cada classe do modelo. Utilizando a teoria dos conjuntos *fuzzy*, um valor de incerteza é então calculado e associado à resposta do classificador para derivar uma família de métricas classes-específicas com pesos utilizada para comparar imagens com maior precisão. A Figura 4 ilustra o processo de recuperação de imagens da CSWIRe, utilizando uma interface gráfica apropriada.

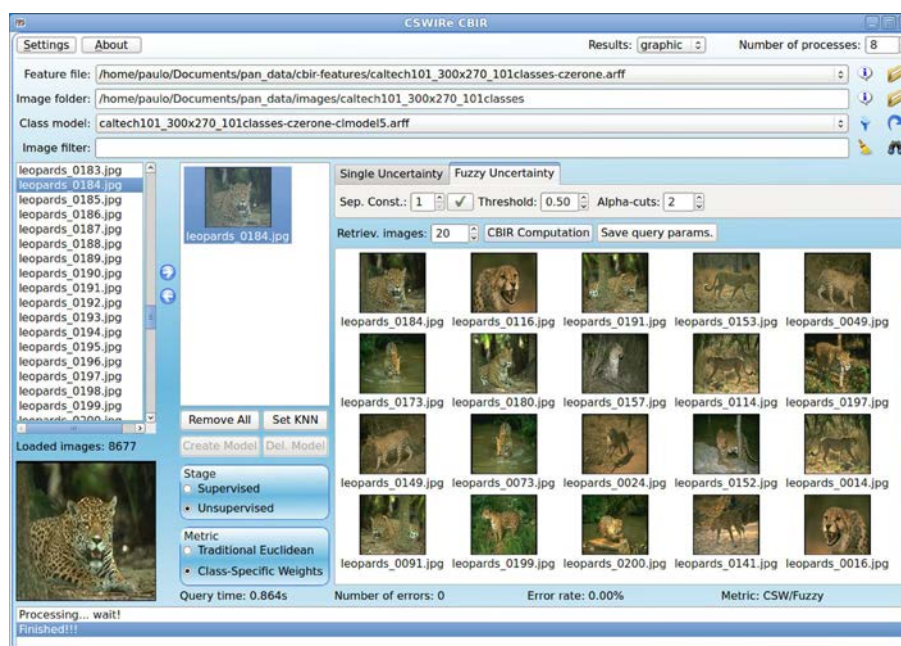


Figura 4. Interface gráfica da CSWIRe, mostrando o processo de recuperação de imagens por conteúdo.

3. Outras Produções Relevantes

Além das contribuições apresentadas anteriormente, merece destaque o trabalho desenvolvido em colaboração com o grupo de pesquisa da *Universidade de Calgary*¹, visando a exploração de espaços multidimensionais via projeção inversa, com uso da técnica intitulada *inverse-LAMP* (iLAMP). Esta técnica executa a projeção inversa através de mapeamentos locais afins que preservam a distância entre as novas amostras de modo preciso, já que ela segue os mesmos preceitos da LAMP. Desse modo, o usuário pode interativamente criar instâncias no conjunto de dados original, gerando assim, dados multidimensionais sintéticos além dos já existentes na disposição inicial. Para maiores detalhes sobre iLAMP, consulte [Dos Santos Amorim et al. 2012].

Para atender os requisitos do projeto, foram desenvolvidas algumas ferramentas computacionais. *Projection Analyzer* (PAN), um conjunto de bibliotecas de alto desempenho em ANSI C (versão inicial disponível em <http://sites.google.com/site/paulojoiafilho/tools>). Também foi implementado um módulo em Python para facilitar a execução das tarefas, que além de reutilizar o código em C, permite integração com pacotes de matemática numérica conhecidos. As interfaces gráficas mostradas nas Figuras 2 e 4, por exemplo, foram desenvolvidas a partir destas ferramentas.

4. Publicações, Prêmios e Estado da Arte

A técnica LAMP [Joia et al. 2011a] destaca-se como uma das técnicas de projeção mais precisas da atualidade. O recente trabalho apresentado por [Fadel et al. 2015] comprova este fato ao comparar várias técnicas de projeção (*stress* e tempo computacional), posicionando a LAMP como uma das técnicas de projeção do *estado da arte* em relação à preservação de distâncias e custo computacional. Além de um número considerável de citações em trabalhos científicos, LAMP também recebeu menção honrosa na *Conferência IEEE Information Visualization*, uma das principais dessa área de pesquisa do mundo (mais informações em <http://www.usp.br/agen/?p=79480>).

4.1. Lista de Publicações

1. [Joia et al. 2011a] “Local Affine Multidimensional Projection”, *IEEE Transactions on Visualization and Computer Graphics* (**Qualis A1**).
2. [Joia et al. 2011b] “Projection-based Image Retrieval using Class-Specific Metrics”, *24th SIBGRAPI*, IEEE Computer Society (**Qualis B1**).
3. [Joia et al. 2012] “Class-Specific Metrics for Multidimensional Data Projection Applied to CBIR”, *The Visual Computer Journal* (**Qualis B1**).
4. [Dos Santos Amorim et al. 2012] “iLAMP: Exploring High-Dimensional Spacing through Backward Multidimensional Projection”, *IEEE VAST'12* (**Qualis B1**).
5. [Casaca et al. 2013] “Spectral Image Segmentation Using Image Decomposition and Inner Product-Based Metric”, *Journal of Mathematical Imaging and Vision* (**Qualis A2**).
6. [Joia et al. 2015] “Uncovering Representative Groups in Multidimensional Projections”, *Computer Graphics Forum* (EuroVis'15) (**Qualis A2**).
7. Joia, P.; da Silva, S. F.; Batista, J.; Nonato, L. G. “Class-specific metrics with weights and uncertainty modeling using fuzzy sets applied to content-based image retrieval”, *Expert Systems with Applications*, Elsevier. Submetido, em processo de revisão (**Qualis A1**).

¹ *Interactive Reservoir Modeling and Visualization Group*, Universidade de Calgary, Alberta, Canadá.

4.2. Prêmios

1. [Joia et al. 2011a] *Local Affine Multidimensional Projection*. Honorable Mention Award - IEEE InfoVis 2011 (<http://www.cad.zju.edu.cn/home/vag/~cwf/vispapers/infovis2011.html>).
2. [Joia et al. 2011b] *Projection-based Image Retrieval using Class-Specific Metrics*. Best Paper Award - Sibgrapi 2011 (<https://www.computer.org/csdl/proceedings/sibgrapi/2011/4548/00/4548z022.pdf>).

5. Conclusões

Este trabalho apresentou soluções para problemas de *identificação de agrupamentos e busca por similaridade em dados multidimensionais*, utilizando técnicas de projeção. Para cada problema, duas novas técnicas foram desenvolvidas, atingindo resultados expressivos em cada uma delas, uma das quais, inclusive, equipara-se ao estado da arte em termos de precisão e eficiência computacional (LAMP).

Agradecimentos

Agradecemos ao Instituto de Ciências Matemáticas e de Computação (ICMC/USP) pelo suporte ao projeto e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela apoio financeiro (Processo #2010/07367-9).

Referências

- Casaca, W., Paiva, A., Gomez-Nieto, E., Joia, P., and Nonato, L. G. (2013). Spectral Image Segmentation Using Image Decomposition and Inner Product-Based Metric. *Journal of Mathematical Imaging and Vision*, 45(3):227–238.
- Dos Santos Amorim, E. P., Brazil, E. V., Daniels, J., Joia, P., Nonato, L. G., and Sousa, M. C. (2012). iLAMP: Exploring High-Dimensional Spacing through Backward Multidimensional Projection. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 53–62. IEEE.
- Fadel, S. G., Fatore, F. M., Duarte, F. S. L. G., and Paulovich, F. V. (2015). LoCH: A neighborhood-based multidimensional projection technique for high-dimensional sparse spaces. *Neurocomputing*, 150(Part B):546–556.
- Joia, P. (2015). *Técnicas de projeção para identificação de grupos e comparação de dados multidimensionais usando diferentes medidas de similaridade*. Tese de Doutorado, Universidade de São Paulo - USP, São Carlos, SP. Disp. em: <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-29032016-143247>>. Acesso em: 25 maio 2016.
- Joia, P., Coimbra, D., Cuminato, J. A., Paulovich, F. V., and Nonato, L. G. (2011a). Local Affine Multidimensional Projection. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2563–2571.
- Joia, P., Gomez-Nieto, E., Batista Neto, J., Casaca, W., Botelho, G., Paiva, A., and Gustavo Nonato, L. (2012). Class-specific metrics for multidimensional data projection applied to CBIR. *The Visual Computer*, 28(10):1027–1037.
- Joia, P., Gomez Nieto, E., Botelho, G., Batista Neto, J., Paiva, A., and Nonato, L. G. (2011b). Projection-based Image Retrieval using Class-Specific Metrics. In Lewiner, T. and Torres, R., editors, *24th SIBGRAPI*, pages 125–132, Maceio, AL. IEEE Computer Society.
- Joia, P., Petronetto, F., and Nonato, L. G. (2015). Uncovering Representative Groups in Multidimensional Projections. *Computer Graphics Forum*, 34(3):281–290.

Electromigration Aware Cell Design

Gracieli Posser¹, Ricardo Reis¹, Sachin S. Sapatnekar²

¹ Universidade Federal do Rio Grande do Sul (UFRGS) - PPGC/PGMicro
Porto Alegre, RS - Brazil

²University of Minnesota - ECE Department - Minneapolis, MN 55455

{gposser, reis}@inf.ufrgs.br, sachin@umn.edu

Abstract. *Electromigration (EM) in on-chip metal interconnects is a critical reliability failure mechanism in nanometer-scale technologies. This work addresses the problem of EM on signal interconnects and on Vdd and Vss rails within a standard cell. An approach for modeling and efficient characterization of cell-internal EM is developed, incorporating Joule heating effects. We also present a graph-based algorithm that computes the currents when the pin position is moved avoiding a new characterization for each pin position and consequently reducing considerable the characterization time. We use the cell lifetime analysis to determine the lifetime of large benchmark circuits, and show that these circuit lifetimes can be improved up to 80.95% by avoiding the critical output, Vdd, and Vss pin positions of the cells, using minor layout modifications.*

1. Introduction

Electromigration (EM) is a major source of failure in on-chip wires and vias, and is becoming a progressively increasing concern as feature sizes shrink [Lienig 2013]. EM is initiated by current flow through metal wires and may cause open-circuit failures over time in copper interconnects. As consequence, circuits can stop work or work incorrectly, producing undesired results. Therefore, reduce the EM effects and increase the circuit lifetime is becoming a great and important challenge. We do not expect acquire an electronic product that will live just for a couple of months.

Traditionally, EM has been a significant concern in global power delivery networks, which largely experience unidirectional current flow. Recently, two new issues have emerged. First, EM analysis can no longer be restricted just to global wires. Traditional EM analysis has focused on higher metal layers, but with shrinking wire dimensions and increasing currents, the current densities in lower metal layers are also now in the range where EM effects are manifested. EM effects are visible at current densities of about $1\text{MA}/\text{cm}^2$, and such current densities are seen in the internal metal wires of standard cells, resulting in cell-internal signal EM [Jain and Jain 2012]. These high current densities arise because local interconnect wires within standard cells typically use low wire widths to ensure compact cell layouts. In short metal wires, such effects were traditionally thought to be offset by Blech length considerations, but such effects do not help protect intra-cell wires in designs at deeply scaled technology nodes. Second, EM has become increasingly important in signal wires, where the direction of current flow is bidirectional. This is due to increased current densities, whose impact on EM is amplified by Joule heating effects [Lee 2012], since EM depends exponentially on temperature.

Therefore, the current that flows through these wires to charge/discharge the output load can be large enough to create significant EM effects over the lifetime of the chip.

Intra-cell power networks are also associated with EM concerns. In going down to deeply scaled technology nodes, the current through the power rails of the cells has remained roughly constant while the cross-sectional area of power rails has decreased, causing the current density in power rails to increase [Wang et al. 2014]. Moreover, the power rails are generally subjected to a unidirectional current flow, referred as DC electromigration, which acts more aggressively in causing electromigration.

In the cell library used in this work, we can see high current densities on the Vdd and Vss power rails as well as on signal wires, reducing the lifetime of the cells. For example, we compute signal wires in an INV_X4 cell to have an effective average current density of 1.8 MA/cm^2 at 2GHz, while power wires have an effective current density of 2.15 MA/cm^2 in a 22nm technology. This switching rate is very realistic, and can be seen in, for example, clock buffers in almost any modern design.

While the cell-internal signal EM problem has been described in industry publications such as [Jain and Jain 2012], its efficient analysis is an open problem. The solution presented in this work is the first to analyze the EM effects inside of the logical cells in a circuit. Due to innovation and importance of this solution it has a patent application [Sapatnekar et al. 2015]. Moreover, the novelty and importance of this work can also be seen in a journal publication [Posser et al. 2016] and important conferences in the field [Posser et al. 2014a], [Posser et al. 2014b], [Posser et al. 2015c], [Posser et al. 2015a] and also a best paper in a regional conference [Posser et al. 2015b].

In this work, we study the problem of systematically analyzing cell-internal signal EM due to both AC EM on signal wires and DC EM on the Vdd and Vss rails of the cells. We devise a solution that facilitates the analysis and optimization of cell-internal signal EM for a standard cell library based design. We first develop an approach to efficiently characterize cell-internal EM over all output, Vdd, and Vss pin locations within a cell, incorporating Joule heating effects into our analysis. EM is modeled using the well-known Black's equation [Black 1969]. We then formulate the pin optimization problem that chooses cell output pins during place-and-route so as to maximize the lifetime of large benchmark circuits.

2. Problem Motivation

We motivate the problem using the INV_X4 (inverter with size 4) cell, shown in Fig. 1(a), from the 45nm NANGATE library [nan]. The input signal A is connected to the polysilicon structure. The layout uses four parallel transistors for the pull-up (poly over p-diffusion, upper half of the figure) and four for the pull-down (poly over n-diffusion, lower half of the figure), and the output signal can be tapped along the H-shaped metal net in the center of the cell. are numbered 1 through 7, and the edges of the structure are labeled e_1 through e_6 , as shown in the figure. Since the four PMOS transistors are all identical, by symmetry, the currents injected at nodes 1 and 5 are equal; similarly, the NMOS-injected currents at nodes 3 and 7 are equal [Posser et al. 2016].

Let us first consider cell-internal signal EM. When the output pin is at node 4, the charge/discharge current is as shown in Fig. 1(b). Moving the pin changes the current

distribution in e_1 – e_6 . If the pin is at node 3 (Fig. 1(c)), since the rise and fall discharge currents have similar values, the charging current in edge e_2 is about $2\times$ larger than the earlier case, while the discharging current is about the same (with opposite direction). The larger peak current leads to a stronger net electron wind that causes EM, resulting in a larger *effective average current*, and therefore, a lower lifetime. Based on exact parasitic extraction of the layout, fed to SPICE (thus including short-circuit and leakage currents), the average effective EM current through e_2 is $1.17\times$ larger than when the pin is at node 4. Accounting for Joule heating, this results in a 19% lifetime reduction. For the Vdd and Vss pins, a similar effect occurs when the pin position changes.

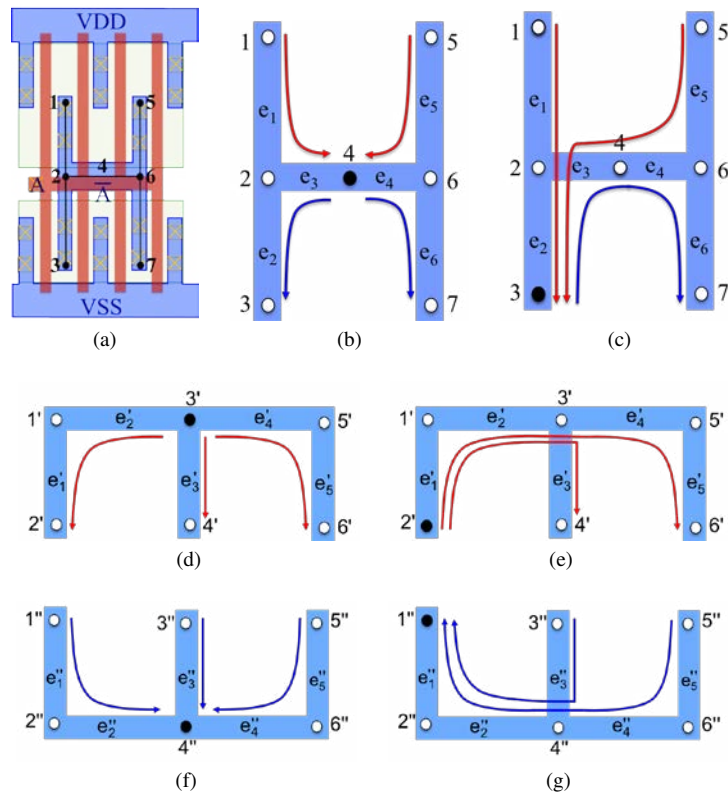


Figure 1. (a) The layout and output pin position options for INV_X4. Charge/discharge currents when the output pin is at (b) node 4 and (c) node 3. The red [blue] lines represent rise [fall] currents. (d) The Vdd pin position options for INV_X4 and the currents when the Vdd pin is at node 3' and (e) node 2'. (f) The Vss pin position options for INV_X4 and the currents when the Vss pin is at node 4'' and (g) node 1''.

Next, we consider EM on the supply wires. Fig. 1(d) and (e) represent the Vdd rail, where the Vdd pin can be placed on the nodes numbered 1' through 6'. Fig. 1(d) shows how the charge current is flowing through the edges when the Vdd pin is placed at node 3'. We can see that the current flows are symmetric for this pin position. Since the edge e'_3 supplies two transistors, as shown in Fig. 1(a), the current flowing through e'_3 is larger than the current flowing through the other edges, which each supply just one transistor. Thus, the edge e'_3 is the critical edge when the Vdd pin is placed at node 3'. Fig. 1(e) shows the current flowing through the edges when the Vdd pin is placed at node 2'. In this case, the current flowing through edge e'_1 supplies three of the four transistors, is $3\times$ larger than the current flowing through this same edge when the pin is at node 3'.

Thus, this is the critical edge for this pin position, reducing the lifetime of the cell by $2\times$ compared with the lifetime when the pin is placed at node $3'$. Similarly, the Vss rail of the INV_X4 cell is represented in Figs. 1(f) and (g). The Vss pin can be placed on the numbered nodes $1''$ through $6''$, and the currents being discharged through the edges by the Vss pin placed at node $4''$ are shown in Fig. 1(f). Using a similar argument as for the Vdd case, moving the pin from node $4''$ in Fig. 1(f) to pin $1''$ in Fig. 1(g) changes the critical edge from e_3'' to e_1'' , and the lifetime again degrades by about $2\times$.

3. Current Calculation

The evaluation of EM TTF (time to failure) requires a characterization of (a) the average and RMS currents through a Vdd/Vss line and (b) the average currents, I_{avg}^r and I_{avg}^f and the RMS current I_{rms} . All of these parameters are both dependent on the pin position, as demonstrated in Fig. 1, and an obvious approach would be to enumerate the characterization over all possible combinations. Thereby, we develop a graph-based method for determining this redirection, and an algebra for computing I_{avg} and I_{rms} for each pin position based on the values from the reference case. Thereby, we developed an efficient graph-based algorithm to speed up the characterization of cell-internal over all output, Vdd, and Vss pin locations within a cell. This algorithm estimates the currents when the pin position is moved from a reference pin position to other one. In this way, a new characterization for each pin position is avoided, reducing the characterization time. Our algorithm produces an average error of just 0.53% compared to SPICE simulation.

4. Results

A method for optimizing the output, Vdd and Vss pin placement of the cells and consequently to optimize the circuit lifetime using minor layout modifications is proposed, just restricting the pin positions to the ones that reduces EM effects. Table 1 presents the results for a set of ITC'99 and ISCAS'89 benchmarks circuits mapped to our set of characterized cells and placed-and-routed. The results for Vdd and Vss pin placement are suppressed because the page limit, they can be seen in the Ph.D. dissertation. . For each benchmark the number of combinational cells, the clock period, total power consumption (leakage and switching power), area of core and total wirelength (WL) are presented, as reported by Encounter. These results correspond to a post place-and-route layout with no EM awareness, and the gap between the best and worst TTF values indicates how much the lifetime can be improved. The number of critical nets corresponds to the nets that violate the Joule heating constraint, and the number of critical cells corresponds to the cells that have pin positions that correspond to lifetimes below the best TTF. Interestingly, these numbers are both small, implying that large improvements to the lifetime can be obtained through a few small changes to the layout. Note that the best TTF values are in the range required for many modern applications (e.g., mobile devices) with short TTF specs of 3 – 4 years. The lifetime of a circuit can be improved by up to 62.50% by altering the pin position of a few cells.

We now redo the routing step to guarantee that the best TTF in Table 1 can be met by outlawing all pin positions whose TTF is worse than the best TTF or with a Joule heating violation. Since the best TTF was computed by choosing the best pin position for each cell the circuit lifetime will be significantly enhanced. (Note that by the definition of best TTF, each cell is guaranteed to have at least one allowable pin).

Table 1. Cell-internal EM analysis for a set of benchmark circuits.

Circuit	# of comb. cells	Period (ns)	Power (mW)	Area of core (μm^2)	Total wire length (μm)	Worst TTF (years)	Best TTF (years)	TTF Improv.	# of crit. nets	# of crit. cells
b05	859	0.544	0.551	504	2682.50	4.07	6.53	37.59%	-	4
b07	461	0.306	0.352	317	1426.87	3.81	5.25	27.43%	-	3
b11	821	0.384	0.460	471	2439.83	2.75	5.82	52.80%	1	5
b12	1217	0.282	0.810	824	4236.15	3.13	3.14	0.15%	3	1
b13	340	0.208	0.467	272	1272.99	3.89	6.05	35.70%	1	7
s5378	1219	0.299	0.679	890	6418.27	2.74	3.59	23.67%	2	1
s9234	1044	0.373	0.584	849	4873.30	2.73	3.48	21.39%	-	1
s13207	1401	0.720	1.063	1733	7146.48	4.94	13.18	62.50%	-	7
s38417	10068	0.493	8.836	7959	46419.93	3.43	5.77	40.51%	2	6

After these new constraints are imposed on the pin positions, the router makes incremental changes to some interconnect routes. Table 2 shows the results after physical synthesis considering the best pin positions, i.e., for each cell, we disallow EM-unsafe pin positions. Thus, we see that the circuit lifetime is improved up to 62.50% while keeping the delay, area and power of the circuit unchanged, and with marginal changes ($\leq 0.15\%$) to the total wirelength.

Table 2. Performance impact of EM-aware physical synthesis using pin optimization.

Circuit	Period (ns)	Δ Period (%)	Power (mW)	Area (μm^2)	WL (μm)	Δ WL (%)
b05	0.544	-	0.551	504	2682.6	0.00
b07	0.306	-	0.353	317	1428.5	0.12
b11	0.384	-	0.460	471	2443.5	0.15
b12	0.280	-0.89	0.808	824	4112.8	-2.91
b13	0.208	-	0.467	272	1273.5	0.04
s5378	0.299	-	0.679	890	6422.2	0.06
s9234	0.373	-	0.584	849	4873.4	0.00
s13207	0.720	-	1.063	1733	7146.6	0.02
s38417	0.493	-	8.836	7959	46420.2	0.00

Runtime: As previously cited, the circuit analysis is executed by Encounter tool and the runtime for each benchmark is less than 40s. The critical pin positions for each circuit are reported in under 1s.

5. Conclusion

To our best knowledge, this is the first work that address the EM by optimizing the output, Vdd and Vss pin positions in cells improving the circuit lifetime. To optimize the TTF of the circuits just the LEF file of the critical instances is changed to avoid the critical pin positions. The cell layout is not changed. The circuit lifetime could be improved up to 62.50% at the same area, delay, and power. This is because changing the pin positions affects very marginally the routing. This lifetime improvement is achieved just avoiding the critical output pin positions of the cells [Posser et al. 2016], [Posser et al. 2014a], [Sapatnekar et al. 2015].

A lifetime improvement of 78.54% is achieved avoiding the critical Vdd pin positions and 89.89% avoiding the critical Vss pin positions. When the output, Vdd, and Vss pin positions are all optimized simultaneously, the TTF of the circuit could be improved up to 80.95% (from 1 year to 5.25 years). Considering the largest and smallest

lifetimes over all pin candidates for a set of cells, a lifetime of a cell can be improved up to $76\times$ by the output pin placement. Some cells present a small TTF improvement when the output pin position changes, because the current densities through the edges are almost unchanged when the pin position changes. For these cells, layout optimizations are suggested to improve their lifetime [Posser et al. 2015a]. At circuit level, we present an analysis of the EM effects on different metal layers and different wire lengths for signal wires (nets) that connect cells [Posser et al. 2014b].

References

- NanGate 45nm Open Cell Library. <http://www.nangate.com>.
- Black, J. R. (1969). Electromigration – a brief survey and some recent results. *IEEE Transactions on Electron Devices*, ED-16:338–347.
- Jain, P. and Jain, A. (2012). Accurate current estimation for interconnect reliability analysis. *IEEE Transactions on VLSI Systems*, 20(9):1634–1644.
- Lee, K.-D. (2012). Electromigration recovery and short lead effect under bipolar- and unipolar-pulse current. In *Proceedings of the IEEE International Reliability Physics Symposium*, pages 6.B.3.1–6.B.3.4.
- Lienig, J. (2013). Electromigration and its impact on physical design in future technologies. In *ACM International Symposium on Physical Design*, pages 33–40.
- Posser, G., de Paris, L., Mishra, V., Jain, P., Reis, R., and Sapatnekar, S. S. (2015a). Reducing the signal electromigration effects on different logic gates by cell layout optimization. In *2015 IEEE 6th Latin American Symposium on Circuits Systems (LAS-CAS)*, pages 1–4.
- Posser, G., Mishra, V., Jain, P., Reis, R., and Sapatnekar, S. S. (2014a). A systematic approach for analyzing and optimizing cell-internal signal electromigration. In *Proceedings of the International Conference on Computer-Aided Design, ICCAD '14*.
- Posser, G., Mishra, V., Jain, P., Reis, R., and Sapatnekar, S. S. (2015b). Analyzing and optimizing cell-internal signal electromigration. In *South Symposium on Microelectronics, SIM 2015*, Santa Maria, RS.
- Posser, G., Mishra, V., Jain, P., Reis, R., and Sapatnekar, S. S. (2015c). Impact on performance, power, area and wirelength using electromigration-aware cells. In *IEEE International Conference on Electronics, Circuits and Systems (ICECS)*.
- Posser, G., Mishra, V., Jain, P., Reis, R., and Sapatnekar, S. S. (2016). Cell-internal electromigration: Analysis and pin placement based optimization. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 35(2):220–231.
- Posser, G., Mishra, V., Reis, R., and Sapatnekar, S. S. (2014b). Analyzing the electromigration effects on different metal layers and different wire lengths. In *IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pages 682–685.
- Sapatnekar, S. S., Mishra, V., Jain, P., Posser, G., and Reis, R. (2015). Cell-level signal electromigration. US Patent Application number 20150347665.
- Wang, C.-H., Tam, K.-H., and Chen, H.-Y. (2014). Automatic place and route method for electromigration tolerant power distribution. US Patent 8,694,945.

Algorithms for Sorting by Reversals or Transpositions, with Application to Genome Rearrangement

Gustavo Rodrigues Galvão (Author)¹ and Zanoni Dias (Supervisor)¹

¹Institute of Computing
University of Campinas (Unicamp) – Campinas, SP – Brazil

{ggalvao, zanoni}@ic.unicamp.br

Abstract. *The problem of finding the minimum sequence of rearrangements that transforms one genome into another is a well-studied problem that finds application in comparative genomics. Representing genomes as permutations, in which genes appear as elements, that problem can be reduced to the combinatorial problem of sorting a permutation using a minimum number of rearrangements. Such combinatorial problem varies according to the types of rearrangements considered. The PhD thesis summarized in this paper presents exact, approximation, and heuristic algorithms for solving variants of the permutation sorting problem involving two types of rearrangements: reversals and transpositions.*

1. Introduction

One of the challenges of modern science is to understand how species evolve. As evolution can be viewed as a branching process, whereby new species arise from changes occurring in living organisms, the study of the evolutionary history of a group of species is commonly made by analyzing trees whose nodes represent species and edges represent evolutionary relationships. Since these relationships are referred to as phylogeny, such trees are called phylogenetic trees.

Phylogenies can be inferred from different kinds of data, from geographic and ecological, through behavioral, morphological, and metabolic, to molecular data, such as DNA. Molecular data have the advantage of being exact and reproducible, at least within experimental error, not to mention fairly easy to obtain [Gascuel 2005, Chapter 12]. Distance-based methods form one of the three large groups of methods to infer phylogenetic trees from molecular or sequence data [Lemey et al. 2009, Chapter 5]. Such methods proceed in two steps. First, the evolutionary distance is computed for every sequence pair and this information is stored in a matrix of pairwise distances. Then, a phylogenetic tree is constructed from this matrix using a specific algorithm, such as *Neighbor-Joining* [Saitou and Nei 1987]. Note that, in order to complete the first step, we need some method to estimate the evolutionary distance between a sequence pair. Assuming the sequence data correspond to complete genomes, we can resort to the genome rearrangement approach [Fertin et al. 2009] in order to estimate the evolutionary distance.

Using the genome rearrangement approach, one estimates the evolutionary distance between two genomes by finding the rearrangement distance between them, which is the length of the shortest sequence of rearrangement operations that transforms one genome into the other. Assuming genomes consist of a single chromosome, share the same set of genes, and contain no duplicated genes, we can represent them as permutations of integers, where each integer corresponds to a gene. If, besides the order, the

orientation of the genes is also considered, then each integer has a sign, $+$ or $-$, and the permutation is called a signed permutation. Similarly, we also refer to a permutation as an unsigned permutation when its elements do not have signs. Moreover, if the genomes are circular, then the permutations are also circular; otherwise, they are linear.

By representing genomes as permutations, the problem of finding the shortest sequence of operations that transforms one genome into another can be reduced to the combinatorial problem of computing the minimum number of operations necessary to transform one permutation into another. By algebraic properties of permutations, this problem can be equivalently stated as the problem of computing the minimum number of operations necessary to transform one permutation into the identity permutation ($+1 +2 \dots +n$). This problem is commonly referred to as the permutation sorting problem or as the rearrangement sorting problem.

Depending on the operations allowed to sort a permutation, we have a different variant of the permutation sorting problem. The PhD thesis defended by the author [Galvão 2015] focus on solving variants that take into account two types of operations: reversals and transpositions. A reversal is responsible for reversing the order and flipping the signs of a sequence of elements within a permutation, while a transposition is responsible for switching the location of two contiguous portions of a permutation. They are the most often considered operations for rearrangement-based phylogenetic reconstruction.

2. Contributions and Organization

The thesis consists of a collection of 5 articles that were published in peer-reviewed journals and conference proceedings in the course of the PhD. Specifically, the thesis contains 7 chapters: an introduction followed by 5 chapters (each one corresponding to one article) and a conclusion. The following paragraphs summarize the contents of the intermediary chapters, highlighting the main contributions.

Chapter 2 corresponds to an article [Galvão and Dias 2014b] published in *ACM Journal of Experimental Algorithmics*. In this chapter, we present a tool, called GRAAu, to audit algorithms for permutation sorting problems. The audit consists in comparing, for all permutations of up to a given size, the distance outputted by a given algorithm with the related rearrangement distance, and then producing statistics that can be used to analyze the performance of this algorithm. We also present tightness results for some approximation algorithms regarding two variants of the permutation sorting problem: the problem of sorting by prefix reversals and the problem of sorting by prefix transpositions. Part of the results presented in this chapter were also presented in the Master thesis [Galvão 2012] of the author. Chapter 2 serves as a prelude to the other chapters. For instance, Section 2.2 introduces the basic concepts used throughout the thesis and provides a literature review of several variants of the permutation sorting problem.

Chapter 3 corresponds to an article [Dias et al. 2014] published in *Journal of Bioinformatics and Computational Biology*. In this chapter, we present a general heuristic for permutation sorting problems. The heuristic works by iteratively improving an initial solution produced by other algorithm. In each step, it makes a local change within a sliding window, which moves across the solution. The main idea employed by the heuristic is to transform the sliding window into a small instance of the permutation sorting problem

in such a way that an optimal solution for that instance can be retrieved from a solution database. To evaluate the heuristic, we applied it to the solutions provided by 23 approximation algorithms. The performance of the heuristic varied considerably depending on the algorithm that produced the initial solutions: it ranged from almost 5% of improvement to near 100%. The observed variation is mainly due to the quality of the initial solutions: the closer they are to the optimal solution, the more difficult it is to improve them.

Chapter 4 corresponds to an article [Galvão and Dias 2014] published in *Journal of Universal Computer Science*. The best known algorithms for the problem of sorting by transpositions are based on a standard tool for tackling permutation sorting problems, the cycle graph. In an attempt to bypass it, a few researches proposed algorithms based on alternative tools. In Chapter 4, we address three of these algorithms: a 2.25-approximation algorithm proposed by Walter et al. (2000), a 3-approximation algorithm proposed by Benoît-Gagné and Hamel (2007), and a heuristic proposed by Guyer et al. (1997). On the theoretical side, we close a missing gap on the proof of the approximation ratio of Benoît-Gagné and Hamel’s algorithm [Benoît-Gagné and Hamel 2007] and we demonstrate a way to run their algorithm in $O(n \log n)$ time. Moreover, we propose a minor adaptation to Guyer, Heath, and Vergara’s heuristic [Guyer et al. 1997] that allow us to prove an approximation bound of 3. On the evaluation side, we present experimental data indicating that Walter, Dias, and Meidanis’ algorithm [Walter et al. 2000] is the best of the algorithms based on alternative approaches and that it is the only one comparable to the algorithms based on the cycle graph.

Chapter 5 corresponds to an article [Galvão et al. 2015b] published in *Algorithms for Molecular Biology*. In this chapter, we investigate the problem of sorting a signed permutation by short operations (*i. e.* operations involving just few genes). The biological relevance of this problem is grounded on the assumption that rearrangement events affecting large portions of a genome are less likely to occur. In the past, corroborating evidence has emerged, that is, separate sets of observations have shown the prevalence and significance of short reversals in the evolution of bacterial genomes [Dalevi et al. 2002, Lefebvre et al. 2003] and lower eukaryote genomes [McLysaght et al. 2000, Seoighe et al. 2000]. This fact, together with the realization that signed permutations constitute a more biologically relevant model for genomes, motivated us to investigate the problem of sorting a signed permutation by short operations.

In preliminary work, we [Galvão and Dias 2014a] investigated the problem of sorting a signed permutation by reversals of length at most 3 and presented three approximation algorithms, the best one having an approximation factor of 9. In Chapter 5, we not only present an approximation algorithm with a better approximation factor, but also consider other variants of the problem. More precisely, we study four variants of the permutation sorting problem: (i) the problem of sorting a signed permutation by reversals of length at most 2, (ii) the problem of sorting a signed permutation by reversals of length at most 3, (iii) the problem of sorting a signed permutation by reversals and transpositions of length at most 2, and (iv) the problem of sorting a signed permutation by reversals and transpositions of length at most 3. We present polynomial-time solutions for problems (i) and (iii), a 5-approximation for problem (ii), and a 3-approximation for problem (iv). Moreover, we show that the expected approximation factor of the 5-approximation algo-

rithm is not greater than 3 for random signed permutations with more than 12 elements. Finally, we present experimental results that show that the approximation factors of the approximation algorithms cannot be smaller than 3. In particular, this means that the approximation factor of the 3-approximation algorithm is tight.

Chapter 6 corresponds to an article [Galvão et al. 2015a] published in the proceedings of the *11th International Symposium on Bioinformatics Research and Applications*. In this chapter, we consider the problem of sorting a circular permutation by reversals of length at most 2. Polynomial-time solutions for the unsigned version of this problem were known, but the signed version remained open. In Chapter 6, we present the first polynomial-time solution for the signed version of the problem. Moreover, we perform an experiment to infer distances and phylogenies for published *Yersinia* genomes and compare the results with the phylogenies presented in previous works [Darling et al. 2008, Egri-Nagy et al. 2014].

Recently, the full version of the conference paper corresponding to Chapter 6 has been accepted for publication in *IEEE/ACM Transactions on Computational Biology and Bioinformatics* [Galvão et al. 2016]. The first relevant difference between the journal version and the conference version of the paper is that the journal version contains expanded explanations and expositions of the methods. In fact, the theoretical sections have been rewritten, at least to some extent, to provide more details. In particular, the journal version uses some of the formalism introduced by Meidanis et al. (2000) and Solomon et al. (2003) for dealing with the problem of sorting circular permutations by reversals. The second relevant difference is that the journal version shows additional experimental results: in addition to the experiment with *Yersinia* genomes, we also performed an experiment to infer distances and phylogenies of γ -proteobacterial genomes. Lastly, the third relevant difference is that the journal version presents a web tool for rearrangement-based phylogenetic inference using short operations. The aim of this tool is to facilitate phylogenetic studies based on the methods proposed in the paper. It is available at:

<http://mirza.ic.unicamp.br:8080/shortphy>.

3. Conclusion

In this paper, we presented a summary of the PhD thesis defended by the author [Galvão 2015]. The thesis is focused on solving variants of the permutation sorting problem that involves reversals or transpositions. The main contributions of the thesis include: i) a general heuristic that can be used to improve the solutions provided by any non-optimal algorithm for the permutation sorting problem; ii) a theoretical and experimental investigation of three algorithms based on alternative approaches for the problem of sorting by transpositions; iii) exact and approximation algorithms for the problem of sorting signed permutations by short operations; and iv) a web tool for rearrangement-based phylogenetic inference using short operations. All contributions were published in peer-reviewed international journals, such as described in Table 1.

Table 1. Overview of the bibliographic production. The third column shows the 2014 JCR Impact Factor of the journals, while the fourth column shows the 2014 Qualis classification in Computer Science. The journal *Algorithms for Molecular Biology* does not have a 2014 (nor a 2013) Qualis classification in Computer Science, hence we are showing the 2012 Qualis classification.

Chap.	Journal	JCR	Qualis
2	ACM Journal of Experimental Algorithmics	–	B4
3	Journal of Bioinformatics and Computational Biology	0.78	B1
4	Journal of Universal Computer Science	0.46	B1
5	Algorithms for Molecular Biology	1.46	A1
6	IEEE/ACM Transactions on Computational Biology and Bioinformatics	1.43	B1

References

- Benoît-Gagné, M. and Hamel, S. (2007). A new and faster method of sorting by transpositions. In *Proceedings of the 18th Annual Symposium on Combinatorial Pattern Matching (CPM'2007)*, volume 4580 of *Lecture Notes in Computer Science*, pages 131–141, London, Ontario, Canada. Springer-Verlag.
- Dalevi, D. A., Eriksen, N., Eriksson, K., and Andersson, S. G. E. (2002). Measuring genome divergence in bacteria: A case study using chlamydian data. *Journal of Molecular Evolution*, 55(1):24–36.
- Darling, A. E., Miklós, I., and Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. *PLoS Genetics*, 4(7):e1000128.
- Dias, U., Galvão, G. R., Lintzmayer, C. N., and Dias, Z. (2014). A general heuristic for genome rearrangement problems. *Journal of Bioinformatics and Computational Biology*, 12(3):1450012.
- Egri-Nagy, A., Gebhardt, V., Tanaka, M. M., and Francis, A. R. (2014). Group-theoretic models of the inversion process in bacterial genomes. *Journal of Mathematical Biology*, 69(1):243–265.
- Fertin, G., Labarre, A., Rusu, I., Tannier, E., and Vialette, S. (2009). *Combinatorics of Genome Rearrangements*. The MIT Press, Cambridge, MA, USA.
- Galvão, G. R. (2012). *Uma Ferramenta de Auditoria para Algoritmos de Rearranjo de Genomas*. Master's thesis, University of Campinas. In Portuguese.
- Galvão, G. R. (2015). *Algorithms for Sorting by Reversals or Transpositions, with Application to Genome Rearrangement*. PhD thesis, University of Campinas.
- Galvão, G. R., Baudet, C., and Dias, Z. (2015a). Sorting signed circular permutations by super short reversals. In *Proceedings of the 11th International Symposium on Bioinformatics Research and Applications (ISBRA'2015)*, volume 9096 of *Lecture Notes in Computer Science*, pages 272–283. Springer International Publishing.
- Galvão, G. R., Baudet, C., and Dias, Z. (2016). Sorting circular permutations by super short reversals. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. To appear.

- Galvão, G. R. and Dias, Z. (2014a). Approximation algorithms for sorting by signed short reversals. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB'2014)*, pages 360–369, Newport Beach, California, USA. ACM Press.
- Galvão, G. R. and Dias, Z. (2014b). An audit tool for genome rearrangement algorithms. *ACM Journal of Experimental Algorithmics*, 19:1.1–1.34.
- Galvão, G. R., Lee, O., and Dias, Z. (2015b). Sorting signed permutations by short operations. *Algorithms for Molecular Biology*, 10(12).
- Galvão, G. R. and Dias, Z. (2014). On alternative approaches for approximating the transposition distance. *Journal of Universal Computer Science*, 20(9):1259–1283.
- Gascuel, O. (2005). *Mathematics of Evolution and Phylogeny*. Oxford University Press, Inc., New York, NY, USA.
- Guyer, S. A., Heath, L. S., and Vergara, J. P. C. (1997). Subsequence and run heuristics for sorting by transpositions. Technical Report TR-97-20, Virginia Polytechnic Institute & State University.
- Lefebvre, J. F., El-Mabrouk, N., Tillier, E., and Sankoff, D. (2003). Detection and validation of single gene inversions. *Bioinformatics*, 19(suppl 1):i190–i196.
- Lemey, P., Salemi, M., and Vandamme, A. (2009). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press, Cambridge, UK.
- McLysaght, A., Seoighe, C., and Wolfe, K. H. (2000). High frequency of inversions during eukaryote gene order evolution. In Sankoff, D. and Nadeau, J. H., editors, *Comparative Genomics*, volume 1 of *Computational Biology*, pages 47–58. Springer Netherlands.
- Meidanis, J., Walter, M. E. M. T., and Dias, Z. (2000). Reversal distance of signed circular chromosomes. Technical Report IC-00-23, Institute of Computing, University of Campinas.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(1):406–425.
- Seoighe, C., Federspiel, N., Jones, T., Hansen, N., Bivolarovic, V., Surzycki, R., Tamse, R., Komp, C., Huizar, L., Davis, R. W., Scherer, S., Tait, E., Shaw, D. J., Harris, D., Murphy, L., Oliver, K., Taylor, K., Rajandream, M. A., Barrell, B. G., and Wolfe, K. H. (2000). Prevalence of small inversions in yeast gene order evolution. *Proceedings of the National Academy of Sciences USA*, 97(26):14433–14437.
- Solomon, A., Sutcliffe, P., and Lister, R. (2003). Sorting circular permutations by reversal. In Dehne, F., Sack, J.-R., and Smid, M., editors, *Algorithms and Data Structures*, volume 2748 of *LNCS*, pages 319–328. Springer Berlin Heidelberg.
- Walter, M. E. M. T., Dias, Z., and Meidanis, J. (2000). A new approach for approximating the transposition distance. In *Proceedings of the Seventh International Symposium on String Processing Information Retrieval (SPIRE'2000)*, pages 199–208, Washington, DC, USA. IEEE Computer Society.

XXXV CTIC - Concurso de Trabalhos de Iniciação Científica

Apresentação

O Concurso de Trabalhos de Iniciação Científica (CTIC) é um evento anual da Sociedade Brasileira de Computação que ocorre durante o Congresso da Sociedade Brasileira de Computação (CSBC). O CTIC tem como objetivo incentivar alunos de graduação em Computação a produzirem textos científicos originais sobre trabalhos realizados por eles, em temas relacionados à área.

Participam do CTIC 2016 estudantes matriculados, ou que tenham sido diplomados após 30 de julho de 2015, em Instituições de Ensino Superior brasileiras. Trabalhos integrantes de projetos tiveram no conteúdo do artigo a identificação, de forma clara, os limites de desenvolvimento, como o trabalho se relaciona às demais atividades do projeto e quais foram efetivamente as ações e a contribuição do aluno de graduação ou grupo de alunos de graduação.

Em sua 35ª edição, o CTIC teve um total de 48 submissões, sendo 10 delas aceitas para publicação, gerando uma taxa de aceitação de 21%. Dentre as 10 submissões aceitas, foram selecionadas para apresentação 3 propostas para fins de premiação. O processo de seleção dos artigos foi realizado em duas etapas. Na primeira etapa, os artigos selecionados foram revisados por, pelo menos, três especialistas. Os dez primeiros artigos mais bem avaliados, considerando diferentes critérios, foram selecionados para publicação nos Anais do CSBC. Na segunda etapa, foram selecionados três artigos para apresentação oral no CTIC 2016, e premiação. Para fins de premiação, a qualidade da apresentação oral também é considerada. Da mesma forma que em anos anteriores, trabalhos não apresentados no CTIC 2016 não concorrem ao prêmio.

Avelino F. Zorzo (PUCRS)

Carina F. Dorneles (UFSC)

Comitê de Organização

Comitê de Programa

Adriano Veloso (UFMG)
Artur Ziviani (LNCC)
Bernadette Loscio (UFPE)
Carina F. Dorneles (UFSC) - Coordenadora
Cristiano Cervi (UPF)
Deise Saccol (UFSM)
Denio Duarte (UFFS)
Eduardo Borges (FURG)
Everton Cavalcante (UFRN)
Francieli Boito (UFSC)
Genaina Rodrigues (UNB)
Gisele Pappa (UFMG)
Giseli Lopes (UFRJ)
Isabela Gasparini (UDESC)
José Maria David (UFJF)
Jose Rodrigues Jr (USP)
Juliano Wickboldt (UFRGS)
Kalinka Castelo Branco (USP)
Leandro Wives (UFRGS)
Lucia Giraffa (PUC-RS)
Maria Augusta Nunes (UFSE)
Mario Dantas (UFSC)
Márcio Castro (UFSC)
Mirella Moro (UFMG)
Renata Galante (UFRGS)
Rodrigo Veras (UFPI)
Ronaldo Mello (UFSC)
Ronaldo Ferreira (HP)
Valéria C. Times (UFPE)

Revisores Externos

André Prisco Vargas (FURG)
Cristiane Aparecida Lana (USP)
Cristiano Vasconcellos (UDESC)
Francisco Vieira de Souza (UFPI)
Juliana Vizzoto (UFSM)
Mariele de Almeida Lanes (FURG)
Michele Brandão (UFMG)
Tiago Volpato (USP)

Trabalhos aceitos

*** Artigos finalistas do CTIC**

Blum axioms and nondeterministic computation of functions *

Tiago Royer (Universidade de São Paulo)

Jerusa Marchi (Universidade Federal de Santa Catarina)

Conectando Opiniões a Opinadores: Um estudo de caso sobre protestos políticos no Brasil *

Ramon Vieira (Universidade Federal de São João del Rei)

Alan Neves (Universidade Federal de São João del Rei)

Fernando Mourão (Universidade Federal de São João del Rei)

Leonardo Rocha (Universidade Federal de São João Del Rei)

Arquitetura e Implementação do Sistema Móvel de Assistência ao Idoso *

Matheus Costa Stutzel (Universidade Estadual do Rio de Janeiro)

Alexandre Sztajnberg (Universidade do Estado do Rio de Janeiro)

Sobre quadrados mágicos, grafos graciosos e colorações especiais: teoremas, algoritmos e aplicações

Victoria Patricia Aires (Universidade Federal do Amazonas)

Rosiane de Freitas (Universidade Federal do Amazonas)

Proposta de framework para condução da pesquisa científica na Ciência da Computação

Nicole Davila (Universidade Feevale)

Adriana dos Reis (Universidade Feevale)

Using Active Mediators and Passive Extractors Inside Materialized Data Integration Systems

Paulo Cardoso (Universidade Federal de Santa Maria)

Flavio Franzin (Universidade Federal de Santa Maria)

Sergio Mergen (Universidade Federal de Santa Maria)

Reconhecimento de sinais estáticos de LIBRAS com Support Vector Machines usando Kinect

Leonardo Perdomo (Centro Universitário La Salle)

Mozart Siqueira (Centro Universitário La Salle)

Desmistificando o uso de Matrix Factorization em domínios de recomendação

Nicollas Silva (Universidade Federal de São João Del Rei)

Leonardo Rocha (Universidade Federal de São João Del Rei)

Fernando Mourão (Universidade Federal de São João Del Rei)

D-STHARK: Avaliando Escalonadores Dinâmicos de Tarefas em Arquiteturas - Híbridas Simuladas

Savyo Machado (Universidade Federal de São João del-Rei)

Danilo Amaral (Universidade Federal de São João del-Rei)

Guilherme Andrade (Universidade Federal de Minas Gerais)

Fernando Mourão (Universidade Federal de São João Del Rei)

Leonardo Rocha (Universidade Federal de São João Del Rei)

GPU Accelerated Data Indexing for Hierarchical Density-Based Clustering

Danilo Amaral (Universidade Federal de São João del-Rei)

Savyo Machado (Universidade Federal de São João del-Rei)

Guilherme Andrade (Universidade Federal de Minas Gerais)

Fernando Mourão (Universidade Federal de São João Del Rei)

Leonardo Rocha (Universidade Federal de São João Del Rei)

Blum axioms and nondeterministic computation of functions

Tiago Royer¹, Jerusa Marchi¹

¹ Universidade Federal de Santa Catarina — Departamento de Informática e Estatística

royertiago@gmail.com, jerusa.marchi@ufsc.br

Abstract. *In his doctoral thesis, Manuel Blum proposed two axioms for complexity measures that allows us to talk about complexity in an axiomatic manner. His axioms does not even specify the machine model — it just requires it to satisfy some properties. Blum axioms, however, are defined in the context of function computation. This restriction is easy to implement with deterministic machines, since there is only one output for a given input, but how can a nondeterministic Turing machine compute a function? This paper surveys techniques to associate nondeterministic machines with functions and analyze how they interact with computational complexity.*

1. Introduction

In Theory of Computation, we usually use languages to mathematically model problems in the real world. Decision problems (“yes/no”) are mapped to languages in a very natural way, by just putting every “yes” instance in the language, and leaving the rest out. Search problems usually are rewritten as a decision problem, and then this problem is converted to a language. For instance, the task of finding a satisfying assignment for a Boolean formula is reinterpreted as the task of deciding whether such an assignment exists, and this task is then converted to a language — in this example we have SAT, the Boolean satisfiability problem. This does the trick when it comes to proving that something is hard; for instance, if we show that the decision problem is NP-hard or undecidable, then intuitively the corresponding search problem must be at least as hard. Therefore, concepts like “decidable”, “NP-complete”, “polynomial-time decidable” arise naturally in the context of decision problems.

However, even from the theoretical standpoint, it is useful to extend these concepts to functions. For instance, the concept of polynomial-time computable functions are required to define Karp reductions [Arora and Barak 2009, p. 42]. For single-tape deterministic Turing machines, this definition is easy to extend: A function $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$ is said to be *polynomial-time computable* if there is a Turing machine M and a polynomial p such that, when given the string x as input, M halts with $f(x)$ in its tape within $p(|x|)$ steps.

Most extensions to this basic model, like the use of several tracks, or several tapes, or multidimensional tapes, can be easily incorporated in the definition of polynomial-time computable. Except nondeterminism. We hit a wall right in the start: how does a nondeterministic machine computes a function in the first place?

This paper presents several attempts to establish how a nondeterministic Turing machine could compute a function, and to extend the concept of “polynomial-time computable” under each definition. Section 2 sets reasonableness criteria to both the definition

of function computation and complexity of the computation. Section 3, which contains the several definitions, presents and criticizes Hopcroft and Ullman's and Goldreich's definitions (sections 3.1 and 3.2), proposes one possible definition that meet the quality standards (section 3.3), and shows other two definitions, by Krentel (section 3.4) and Valiant (section 3.5), that, although were not proposed in the context of general nondeterministic computation of functions, also meet the quality standards and sidesteps the problems with the proposed definition (presented in section 3.3.1). The development of this project is mentioned in section 5, after the concluding remarks (section 4).

2. Our approach: Gödel numberings and Blum axioms

In this paper, we will try to associate nondeterministic computation with partial recursive functions, in some well-behaved manner, and preserving the apparent¹ exponential speed-up present in nondeterministic deciders. The concept of Gödel numberings (section 2.1) captures the notion of “well-behaved”. The Blum axioms (section 2.2) capture the notions of computational complexity. We thus will demand the definitions to satisfy the requirements of Gödel numberings and Blum axioms.

We are restricting ourselves to using single-valued functions, but there are alternative approaches. Complexity classes like NPVM (see, for example, the paper of Selman [Selman 1994, p. 359]) are defined using *multivalued functions*, which are allowed to return several values for a single input. Another approach is to use *function problems* associated to problems in NP, where the machine is required to return any certificate for the given instance, or to reject the input if it is not in the language [Papadimitriou 1994, p. 229]. We will not consider these approaches here.

2.1. Reasonableness criterion: acceptable Gödel numberings

One of the most important theoretical results concerning Turing machines is the existence of undecidable problems. Namely, the *halting problem* (the task of deciding whether a given Turing machine will halt on a given input) cannot be solved by Turing machines [Arora and Barak 2009, p. 23]. The formalization (and proof) of this fact requires the definition of some sort of *encoding*; since Turing machines can only reason about strings, we need somehow to encode Turing machines into strings, to be able to pose the halting problem as a language question.

Each Turing machine can be associated to the corresponding partial recursive function it computes. There are several ways to encode Turing machines as strings, but what is most important about them is that they allow us to manipulate these partial recursive functions *indirectly* — partial recursive functions are (potentially) infinite objects, so we cannot write them down on a Turing machine tape, but we *can* write the encoding of a Turing machine that compute these functions.

Therefore, these encodings provide a way to associate a string (which is a finite, manipulable object) with a partial recursive function (which is an infinite, mathematical, “untouchable” object). Encodings are *enumerations* of all partial recursive functions.

¹It is apparent in the sense that, if proven would show $P \neq NP$, and if disproved (showing a polynomial slowdown is the best we can do) would show $P = NP$. Although most researchers expect the former to be the case [Gasarch 2012, p. 54], as Papadimitriou noted [Papadimitriou 1994, p. 412], in the absence of a proof that $P \neq NP$ we should not be too emphatic in stating the simulation *require* (as opposed to *seemingly require*) exponential slowdown.

One important feature about the standard encodings of deterministic Turing machines is the Universal Turing Machine Theorem: the existence of a partial recursive function $U : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}^*$ such that, if w encodes the machine M , then $U(w, x)$ is the result of running the machine M on x . A machine that computes U is called *universal Turing machine*.

The concept of *acceptable Gödel numbering* ([Rogers 1987, p. 41], [Blum 1967, p. 324]) encompasses the existence of universal machines and a little more. We will use it as our reasonableness criteria to our definitions.

Definition 1. Let \mathcal{P} be the set of all partial recursive functions. An *acceptable Gödel numbering* is a function $\phi : \{0, 1\}^* \rightarrow \mathcal{P}$, that associates each string (or program²) $w \in \{0, 1\}^*$ to a function $\phi_w \in \mathcal{P}$, that satisfies

1. ϕ is surjective; that is, every partial recursive function $f \in \mathcal{P}$ has a program $w \in \{0, 1\}^*$ such that $\phi_w = f$;
2. There is a partial recursive function $U : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}^*$ such that, for every w and x , $U(w, x)$ is defined if and only if $\phi_w(x)$ is defined, and, in this case,

$$U(w, x) = \phi_w(x);$$

3. There is a total recursive function $\sigma : \{0, 1\}^* \times \{0, 1\}^* \rightarrow \{0, 1\}^*$ such that, for every y , $\phi_w(x, y)$ is defined if and only if $\phi_{\sigma(w, x)}(y)$ is defined, and, in this case,

$$\phi_w(x, y) = \phi_{\sigma(w, x)}(y).$$

This concept captures the intuitive notion of “well-behaved numbering”.

Condition 1 guarantees that the image of ϕ is all of \mathcal{P} , so that the numbering neither “forgets” a partial recursive function nor generate a function that is not partial recursive. Condition 2 is the universal Turing machine theorem.

Condition 3 is the “little more” we mentioned earlier. It is known as the S_{mn} theorem [Rogers 1987, p. 24]. Essentially, given a partial recursive function of two variables, we can obtain a partial recursive function of one variable by fixing the first argument. The function σ provides a systematic way of doing this: given a description w of a partial recursive function of two variables and the value x to be fixed as the first variable, $\sigma(w, x)$ is a machine that computes this new partial recursive function.

Example 2. Any encoding of deterministic Turing machines as a binary string yields an acceptable numbering of the recursive functions.

Example 3. Any programming language can be understood as an acceptable Gödel numbering. For example, if we restrict a C program to perform input and output only using the standard input and standard output (that is, forbid interactions with the user, file reading, GUIs, access to system clock, etc.), the resulting program will map a binary input to a binary output, characterizing a partial recursive function. Thus, we can see a C compiler as an implementation of an acceptable Gödel numbering.³

²In texts like Rogers’ [Rogers 1987], the partial recursive functions have the naturals as domain and codomain (that is, they are of the form $f : \mathbb{N} \rightarrow \mathbb{N}$), and Gödel numberings associates natural numbers with recursive functions. In this paper we will work with binary strings instead of numbers, which will simplify the definition of complexity classes and allows us to think the string w as a *program* for ϕ_w (see example 3).

³Note we are ignoring here issues like compilation and run-time errors. These can be dealt with as in the case of Turing machines: any invalid program will signify some fixed partial recursive function (say, the function that is defined nowhere); and any invalid step in computation makes the function to be not defined on that input.

One important theorem that can be proven using acceptable Gödel numberings alone is the recursion theorem [Rogers 1987, p. 181]. It states that, if ϕ is any acceptable Gödel numbering and f is any total recursive function, then there is some program w_0 such that $\phi_{w_0} = \phi_{f(w_0)}$. That is, if f is any systematic transformation on programs, there is a program w (a fixed point for f) whose meaning under ϕ is unchanged. The recursion theorem can be used, for example, to show the existence of quines (programs whose output are their own source codes) in any Turing-complete programming language: choose f to be the function that, given a program w , returns another program $f(w)$ that prints the string w when run.⁴ Then, by the recursion theorem, there is some program w_0 that is equivalent to its transformed version $f(w_0)$; thus, w_0 already writes the string w_0 , its own source code. Therefore, any programming language has quines [Kozen 2006, p. 227].

2.2. Efficiency criterion: Blum axioms

The *complexity* of a computation is how much of a resource that is invested in that computation [Hopcroft and Ullman 1979, p. 285]. For each model of computation and each resource under that model, we can establish a *complexity measure* concerning that resource. This section is devoted to formalizing this notion. As we did with the machine encodings, we will impose some restrictions on what can be a complexity measure to be able to manipulate it (at least indirectly). In our case, we will use Blum axioms [Blum 1967, p. 324].

Definition 4. Given an acceptable Gödel numbering ϕ , a *complexity measure* for ϕ is a function $\Phi : \Sigma^* \times \Sigma^* \rightarrow \mathbb{N}$ of two variables that satisfies [Blum 1967, p. 324]:

1. For every w and x , $\phi_w(x)$ exists if and only if $\Phi(w, x)$ exists; and
2. For every string w, x and every natural number k , the predicate “ $\Phi(w, x) = k$?” is decidable.

$\Phi(w, x)$ is the complexity of computing $\phi_w(x)$ using the program w . The first axiom says that it only makes sense to talk about the complexity of a computation that ends. The second axiom gives minimum tools to manipulate Φ indirectly, in the same manner we require ϕ to be an *acceptable* Gödel numbering.

Example 5. The standard measures of time and space can be constructed over the acceptable numbering of example 2. They are, respectively, the number of moves and tape cells scanned before halting. Leave the complexity undefined if the machine does not halt — this satisfies the first axiom. The predicate of the second axiom is simple for time complexity; for space complexity, we must keep the whole history of computation to make sure the machine does not loop in a limited amount of space (because the complexity is not defined in this case).

Definition 6. Given a complexity measure Φ for an acceptable Gödel numbering ϕ and a total recursive function $f : \mathbb{N} \rightarrow \mathbb{N}$, define the *complexity class* $\mathcal{C}_\Phi(f)$ by ([Kozen 2006, p. 232])

$$\mathcal{C}_\Phi(f) = \{\phi_w \mid \Phi(w, x) \leq f(|x|) \text{ almost everywhere}^5\}.$$

⁴For Turing machines, for instance, we can encode the bits of w in the transition table of $f(w)$.

⁵*Almost everywhere* means that the inequality $\Phi_i(x) \leq f(|x|)$ holds for all but a finite number of different x .

This formalizes the notion of complexity class. There are some important theorems concerning complexity classes under the Blum axioms; we mention only the Union Theorem [Kozen 2006, p. 234]. It states that, if $\{f_i\}$ is any recursive list of increasing functions such that $f_i(n) \leq f_{i+1}(n)$ for all x (that is, each f_i is greater than the previous), then there is some function g such that

$$\mathcal{C}_\Phi(g) = \bigcup_{i \in \mathbb{N}} \mathcal{C}_\Phi(f_i).$$

That is, the class $\mathcal{C}_\Phi(g)$ contains *exactly* all functions present in the classes $\mathcal{C}_\Phi(f_i)$. Choosing Φ as the time complexity and $f_i(n) = n^i$ (the polynomial functions), the union in the right is exactly the class P, the problems solvable in polynomial time. By the Union Theorem, P is $\mathcal{C}_\Phi(g)$ for some recursive function g , so, even though P does not have an easily specifiable bounding function, such function exists nevertheless.

3. Nondeterministic computation of functions

This section surveys several approaches for defining nondeterministic computation of functions.

Sections 2.1 and 2.2 introduced the concept of acceptable Gödel numbering, the Blum axioms, and its complexity classes. As we want to regard these concepts as “quality requirements” for the definitions, we will analyze each of them to see whether they define acceptable Gödel numberings, the analogous time complexity satisfies Blum axioms, and that the characteristic function⁶ of the Boolean satisfiability problem can be solved in “polynomial time” according to that complexity measure. (This last requirement expresses that the definition preserves the exponential speed-up that nondeterminism gives to deciders.)

3.1. Hopcroft-Ullman’s definition

Definition 7 (Hopcroft and Ullman’s definition⁷). If w is an encoding for the Turing machine M , we say that $\phi_w(x) = y$ if and only if, when processing x , there is some branch of M that halts with y in the tape, and there is no branch that halts with some $z \neq x$ in the tape.

The problem with their definition is that $\phi_w(x)$ is allowed to be defined, even if some branch of computation does not halt. This allows us to solve the complement of the halting problem.

Proposition 8. Define the partial function $f : \Sigma^* \times \Sigma^* \rightarrow \{0, 1\}$ by

$$f(w, x) = \begin{cases} 1, & \text{if the machine } w \text{ does not halt on } x. \\ \text{undefined,} & \text{if } w \text{ halts on } x. \end{cases}$$

This function can be computed by a nondeterministic Turing machine under Hopcroft and Ullman’s definition.

⁶The characteristic function of a set A is the function 1_A defined to be 1 for $x \in A$ and 0 for $x \notin A$.

⁷Hopcroft and Ullman’s original definition [Hopcroft and Ullman 1979, p. 313] was defined in the context of computation of integer functions. We are rephrasing here in terms of strings, but keeping the relation they imposed on the execution branches.

Proof. On input (w, x) , create two branches of computation. On the first, write 1 on the tape and halt. On the second, simulate the universal Turing machine U in the input, and if U halts, write 0 on the tape and halt too.

If w halts on x , there will be two halting branches of computation, each writing a different value in the tape, so, by definition, the function is not defined on this input. If w never halts, then only the first branch will halt (and with 1 written on the tape), so the function is defined on this input and its value is 1. \square

Therefore, under Hopcroft and Ullman’s definition, we can compute some non-computable functions, violating the requirement 1 of Gödel numberings.

We can try to fix this definition by forcing all branches to halt; but then, as all branches are required to return the same value, the machine will be (almost) deterministic. Therefore, if there is a machine M that computes the characteristic function of the satisfiability problem in nondeterministic polynomial time, we could simulate M on a given input, choosing (say) always the first option when confronted with nondeterminism. If this branch of computation returns 1, then every branch returns 1 and the input is satisfiable; if this branch returns 0, every branch returns 0 and the input is unsatisfiable. We thus could solve SAT in *deterministic* polynomial time. So, with this restriction, we lose the apparent exponential speed-up in computation time we have when using nondeterminism.

3.2. Goldreich’s definition

Definition 9 (Goldreich’s definition). Let \perp be some special symbol not in $\{0, 1\}^*$. (This symbol will represents “don’t know”.) A nondeterministic machine M computes the function f if, when processing the input x , both the following conditions hold [Goldreich 2008, p. 168]:

- Every branch of M halts and outputs either $f(x)$ or \perp .
- At least one branch of M halts with $f(x)$ on the tape.

The extra symbol \perp sidesteps the problems of a branch looping forever. This allows us to mechanically convert nondeterministic machines under Goldreich’s definition to deterministic machines via simulation — the deterministic machine just need to simulate all branches until completion, to actually be sure every branch halts; if some branch do not halt, then the simulating machine will not halt either, but the function is not defined in this case, so this behavior is correct.

Thus, we only enumerate computable functions. To show the universal machine theorem and the S_{mn} , we can simply first convert the machine in question to a deterministic machine and use their theorems; thus, this definition yields an acceptable Gödel numbering. And, by counting the number of steps of the deepest branch, we have a Blum complexity measure.

So, Goldreich’s definition defines an acceptable Gödel numbering and we can form a complexity measure that satisfies Blum axioms. But, again, we have trouble with the “exponential speed-up” requirement. For instance, a machine trying to solve the satisfiability problem would correctly return 1 for a satisfiable instance, but no branch can write a 0 alone because it cannot be sure that instance is unsatisfiable — so, the function would be undefined for unsatisfiable formulas.

That is, unless $\text{NP} = \text{coNP}$.

Proposition 10. If there is a nondeterministic Turing machine that computes in polynomial time, according to Goldreich’s definition, the characteristic function of the satisfiability problem, then $NP = coNP$.⁸

Proof. Suppose M is the machine that computes SAT’s characteristic function, under Goldreich’s definition, in polynomial time.

The characteristic function of SAT is a total function, and its only outputs are 0 and 1. Therefore, in every computation of M , there is at least one branch that writes something different than \perp on the tape, and whatever it writes is the correct answer. So, if we convert this to a nondeterministic decider and invert the output (branches that write 0 will accept the input and vice-versa; branches that write \perp always reject), any satisfiable formula will be rejected, because no branch of M ever writes 0 on the tape for these formulas; and any unsatisfiable formula will be accepted, because at least one branch of M writes 0 for these formulas. Thus, we can decide \overline{SAT} , the complement of SAT.

Since \overline{SAT} is coNP-complete, the existence of such a machine M would show that $coNP \subseteq NP$, and this implies that $coNP = NP$.⁹ \square

Therefore, even though Goldreich’s definition yields an acceptable Gödel numbering and a Blum complexity measure, it also have trouble in transposing the exponential speed-up we have using nondeterminism.

3.3. Proposed definition

Analyzing the problems with the first two definitions, we know the nondeterministic machine must be allowed to return several values (one for each branch) and somehow pick only one to be the value of the function.

If M is any deterministic decider for the language L , we can create a machine that computes the characteristic function of L by running M and returning 1 if M accepted and 0 if it rejected.

If we apply this transformation to a nondeterministic machine that recognizes the Boolean satisfiability problem, then, when running this machine, we have three possible results.

- For tautological formulas, the set of possible answers is only $\{1\}$.
- For contradictory formulas, the set of answers is $\{0\}$.
- For satisfiable formulas that are not tautological, we have both values: $\{0, 1\}$.

We want the return value for all satisfiable formulas to be 1 and for unsatisfiable formulas to be 0. Note that this corresponds exactly to the maximum value of each set; so, our definition of nondeterministic computation of functions will preserve exactly this behavior.

Definition 11 (Proposed definition). Let M be a nondeterministic Turing machine, and x an input. If every branch of M halts when processing x , the value of the function

⁸ $NP = coNP$ implies that $NP = PH$; that is, the polynomial hierarchy collapses to the first level [Kozen 2006, p. 280].

⁹To see why $coNP \subseteq NP$ implies $coNP = NP$, pick a language $L \in NP$. Its complement \overline{L} is in coNP, by the definition of coNP. But by hypothesis, \overline{L} is in NP, so, by the definition of coNP, its complement, $\overline{\overline{L}} = L$ is in coNP, thus showing $coNP = NP$.

computed by M on x is the lexicographically maximum between all strings written in the computation branches. If some branch does not halt, leave the function undefined in x .

The same reasoning of Goldreich's definition applies here; therefore, we have an acceptable Gödel numbering, and counting steps of the deepest branch (as in Goldreich's definition) yields a Blum complexity measure.

And, using exactly the algorithm mentioned above, we can compute the characteristic function of the satisfiability problem under linear time; thus, this definition meets all the requirements proposed in the beginning of section 3.

3.3.1. Extending NP-completeness

The last definition provides an extension of the class NP to function computation, so the next step is extend the concept of NP-completeness.

Call FNP the class of all functions computable in polynomial time, under our definition¹⁰. A language L is NP-complete if both $L \in \text{NP}$ and every language in NP reduces to L ; that is, for every language $L' \in \text{NP}$, there is a polynomial-time computable function f such that, for every x ,

$$x \in L' \text{ if and only if } f(x) \in L.$$

(We are using Karp reductions here [Arora and Barak 2009, p. 42].) If we rephrase in terms of the characteristic functions 1_L and $1_{L'}$, of L and L' , respectively, we have $1_{L'}(x) = 1_L(f(x))$ for all x . We will generalize specifically this equation to define FNP-completeness.

Definition 12. A function f is FNP-complete if $f \in \text{FNP}$ and, for every function g in FNP, there is a polynomial-time computable function h such that

$$g(x) = f(h(x))$$

for every $x \in \{0, 1\}^*$.

We can construct a FNP-complete function based on the halting problem: define $f : \{0, 1\}^* \times \mathbb{N} \rightarrow \{0, 1\}^*$ such that $f(w, n)$ is the lexicographically greatest value written by any branch of w after running for n steps. Such functions are FNP-complete because they simulate Turing machines directly. However, this definition is very rigid and allow for few functions to be FNP-complete; the requirement of directly returning the output of the function g above restricts the class of FNP-complete functions to functions that perform simulations.

3.4. Krentel's OptP class

The next two authors were not specifically concerned with nondeterministic computation of functions, but rather in generalizing the NP class for functions. Therefore, the corresponding notion of NP-completeness behave better than our proposed generalization.

Krentel's definition, in particular, are very similar to ours.

¹⁰Note that this definition is different from the one given by Papadimitriou [Papadimitriou 1994, p 229].

Definition 13. A function $f : \{0, 1\}^* \rightarrow \mathbb{N}$ is in OptP if there is some nondeterministic Turing machine M such that [Krentel 1988, p. 493]:

- For every input, every branch of M halts within a polynomial number of steps and writes in its tape a number in binary; and
- Either, for all x , the largest number written by M on x is $f(x)$, or, for all x , the smallest number written by M on x is $f(x)$.

Therefore, Krentel’s OptP class contains the optimization problems that can be “solved” in polynomial time by nondeterministic Turing machines. (Note we must always take the maximum value, or always take the minimum value.) The definition of OptP-completeness, however, is significantly different.

Definition 14. A function f is OptP-complete if $f \in \text{OptP}$ and, for every function $g \in \text{OptP}$, there are two functions $T_1 : \{0, 1\}^* \rightarrow \{0, 1\}^*$ and $T_2 : \{0, 1\}^* \times \mathbb{N} \rightarrow \{0, 1\}^*$, both computable in deterministic polynomial time, such that, for all x ,

$$g(x) = T_2(x, f(T_1(x))).$$

That is, besides the preprocessing function T_1 , we are allowed to make a post-processing which have access both to the input x and to the “reduced output” $f(T_1(x))$. Under this definition, the traveling salesperson problem and 0 – 1 integer linear programming are both OptP-complete [Krentel 1988, p 495].

3.5. Valiant’s #P class

Definition 15 (Valiant’s definition). A nondeterministic machine M computes a function $f : \{0, 1\}^* \rightarrow \mathbb{N}$ if, for every input x , M halts on every branch¹¹ and the number of accepting branches of computation is $f(x)$ [Valiant 1979, p. 191].

The class #P is the set of functions computed in polynomial time, under Valiant’s definition. #P-completeness is defined through oracles; that is, a function f is #P-complete if $f \in \#P$ and $\#P \subseteq \text{FP}^f$ (that is, every function of #P can be computed by a deterministic machine with access to an oracle that computes f) [Valiant 1979, p. 191].

Besides problems like counting the number of satisfying assignments for a given formula, Valiant also proves less trivial problems are #P-complete, like the task of computing the permanent of a matrix [Valiant 1979, p. 194]. Therefore, this definition is also flexible, like Krentel’s.

Krentel note that both his and Valiant’s definition arise from applying an associative operator to all values returned in the branches of computation — maximum/minimum in Krentel’s case, and cardinality (counting) in Valiant’s case. Therefore, using other associative operations yield alternative definitions of nondeterministic function computation [Krentel 1988, p. 493].

4. Concluding Remarks

As we have seen, it is possible to associate nondeterministic Turing machines with function computation, although the association is not so straightforward as with deterministic machines.

¹¹Valiant did not include the requirement of halting in his definition, but we will include it to avoid having the same problems of Hopcroft and Ullman’s definition.

Through the use of the notions of acceptable Gödel numberings and Blum axioms, we formalized the notion of “reasonable” definition for nondeterministic function computation, and by consequence we formalized what is a “generalization of NP to functions”. Krentel’s and Valiant’s definitions, besides generalizing NP, in particular, also allow for a flexible generalization of NP-completeness — that is, they allow for interesting problems to be “functionally NP-complete”, under each definition.

5. Project Development

Tiago Royer studied this problem of associating a nondeterministic machine with a function in his undergraduate thesis¹², under the supervision of Jerusa Marchi. After finding Hopcroft and Ullman’s definition (section 3.1) and noting it yields a computation that is very close to be deterministic, he devised the definition of section 3.3 (using the concept of acceptable Gödel numberings and Blum axioms to be sure his definition would not be too unreasonable). As noted in section 3.3.1, his generalization to NP and NP-completeness are very rigid.

Krentel’s and Valiant’s definitions (which were discovered later in the project), although not created specifically to associate nondeterministic machine with functions, provides a more elegant generalization of NP and NP-completeness; this paper summarize all these findings under the light of Gödel numberings and Blum axioms.

References

- Arora, S. and Barak, B. (2009). *Computational Complexity - A Modern Approach*. Cambridge University Press.
- Blum, M. (1967). A machine-independent theory of the complexity of recursive functions. *J. ACM*, 14(2):322–336.
- Gasarch, W. I. (2012). Guest column: the second P =? NP poll. *SIGACT News*, 43(2):53–77.
- Goldreich, O. (2008). *Computational complexity - a conceptual perspective*. Cambridge University Press.
- Hopcroft, J. E. and Ullman, J. D. (1979). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley.
- Kozen, D. (2006). *Theory of Computation*. Texts in Computer Science. Springer.
- Krentel, M. W. (1988). The complexity of optimization problems. *Journal of Computer and System Sciences*, 36(3):490 – 509.
- Papadimitriou, C. H. (1994). *Computational complexity*. Addison-Wesley.
- Rogers, Jr., H. (1987). *Theory of recursive functions and effective computability (Reprint from 1967)*. MIT Press, Cambridge, MA, USA.
- Selman, A. L. (1994). A taxonomy of complexity classes of functions. *Journal of Computer and System Sciences*, 48(2):357–381.
- Valiant, L. G. (1979). The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201.

¹²The project was developed while he was an undergraduate student at Federal University of Santa Catarina (UFSC). Currently, he is pursuing his master’s degree at the University of São Paulo (USP).

Conectando Opiniões a Opinadores: Um estudo de caso sobre protestos políticos no Brasil*

Ramon Vieira , Alan Neves , Fernando Mourão , Leonardo Rocha

DCOMP/UFSJ - São João del-Rei, MG , Brasil

{ramonv, aneves, fhmourao, lcrocha}@ufsj.edu.br

***Abstract.** Sentiment analysis (SA) on Social Media content, as well as the Influential Users Detection (IUD), also called opinion-leaders, provide valuable information for many applications. Despite the intrinsic relation between opinions and opinion-leaders, most of the recent works focus exclusively on one of these two tasks. By empirical assessments on a data sample of tweets about the Brazilian president, this work demonstrates the potential benefits of combining SA Methods with IUD ones. In our analysis, we identified distinct behaviors of opinion propagation and demonstrated that the collective opinion may be accurately estimated by using a few opinion-leaders.*

***Resumo.** Análise de Sentimento (AS) sobre conteúdo de Mídias Sociais, bem como a Identificação de Usuários Influentes (IUI), também chamados de opinadores, provêm informações valiosas atualmente. Apesar da intrínseca relação entre opiniões e opinadores, muitos dos trabalhos recentes focam exclusivamente em uma das duas tarefas. Por meio de avaliações empíricas em uma amostra de dados de tweets relacionada à presidente do Brasil, nesse trabalho apresentamos os potenciais benefícios de se combinar métodos de AS com os de IUI. Em nossas análises, identificamos comportamentos distintos de propagação de opiniões e demonstramos que a opinião coletiva pode ser estimada com precisão utilizando tweets relacionados a poucos opinadores.*

1. Introdução

Mídias Sociais vêm se consolidando como um importante ambiente em que pessoas publicam suas opiniões sobre variados assuntos na WEB. Além disso, aplicações de mídias sociais se tornaram decisivas nos processos de tomada de decisão dos usuários. De fato, existe um número crescente de usuários a procura de *reviews* e recomendações sobre produtos e serviços em interações sociais online antes de efetuarem uma escolha [Edelman 2010]. Nesse contexto, compreender e modelar apropriadamente opiniões predominantes de uma dada população (e.g., a opinião coletiva) a respeito de um dado tema (e.g., produto, serviço, etc.), bem como identificar o subconjunto de pessoas formadoras de opinião (opinadores), os quais são capazes de persuadir as outras em relação a um determinado tema, surgiram como problemas relevantes para diversas aplicações, tais como recomendação, publicidade, avaliação de marcas, entre outras.

Apesar dos conceitos de opinião e opinadores apresentarem uma relação intrínseca de difícil dissociação, principalmente quando falamos de dados oriundos de

*Esse trabalho foi parcialmente financiado por CNPq, CAPES, FINER, Fapemig, e INWEB.

redes sociais, em Ciência da Computação esses dois conceitos vem sendo tratados e abordados distintamente. Grande parte das pesquisas foca em identificar quais são as opiniões preponderantes (i.e. opinião coletiva) ou quais são os principais opinadores responsáveis pela propagação dessas opiniões. Enquanto as opiniões são avaliadas por técnicas de Análise de Sentimento (AS) [Zhao et al. 2012, Rocha et al. 2015], sem necessariamente identificar quem são os responsáveis pelas opiniões analisadas, os opinadores são determinados por técnicas de identificação de usuários influentes (IUI) [Ilyas and Radha 2011, Lee et al. 2010, Silva et al. 2013, Page et al. 1999, Neves et al. 2015], sem que o teor de suas opiniões seja considerado. Uma análise combinada, além de melhorar a identificação adequada tanto das opiniões quanto dos opinadores, pode fornecer maior conhecimento sobre o processo de difusão de informação na Web. Apesar da relevância, encontramos poucos trabalhos na literatura focados em combinar AS e IUI.

Nesse trabalho, apresentamos um estudo preliminar que avalia duas questões principais a respeito da combinação entre AS e IUI. Primeiramente, visamos identificar quais são os principais fatores que podem afetar a eficácia e a relevância da combinação dessas técnicas em cenários reais. Em segundo lugar, pretendemos mostrar alguns potenciais benefícios ao realizar essa combinação. Neste sentido, propomos uma metodologia de quatro passos aplicável a cenários distintos. No primeiro passo, avaliamos as características do domínio que podem afetar AS e IUI, por exemplo, a dinamicidade temporal [Mourão et al. 2008]. O segundo passo refere-se às limitações das técnicas de AS e IUI, por exemplo, sensibilidade quanto a amostra de dados. No terceiro, avaliamos o quão correlacionadas estão AS e IUI. Por fim, propomos duas estratégias diretas para demonstrar o potencial em se combinar AS e IUI. A primeira visa qualificar as opiniões propagadas por cada opinador, enquanto a segunda estratégia pretende estimar a opinião coletiva através de uma análise utilizando apenas alguns opinadores.

Para avaliar a metodologia proposta, conduzimos uma análise empírica em uma amostra de dados do *Twitter*, dada a sua relevância no processo de difusão de informação na Web. Essa amostra corresponde a *tweets* em português relacionados à presidente do Brasil, Dilma Rousseff, postados durante protestos políticos em abril de 2015. Em nossas análises, adotamos as técnicas propostas em [Rocha et al. 2015] e [Neves et al. 2015] para identificar a opinião coletiva e os opinadores, respectivamente. Nossos experimentos indicaram que IUI é mais sensível ao tamanho da amostra do que AS. Além disso, identificamos que técnicas tradicionais de IUI não são capazes de identificar opinadores cujos sentimentos diferem da opinião coletiva. Por outro lado, observamos que a opinião coletiva pode ser estimada com precisão avaliando *posts* de um pequeno número de opinadores. Esses resultados apontam direcionamentos de pesquisa relevantes para a área, evidenciando abordagens novas e promissoras para tratar AS e IUI na Web.

Todas as implementações e execuções de experimentos foram realizadas pelo aluno Ramon Vieira, sob a orientação do professor Leonardo Rocha. A concepção da metodologia bem como as análises de todos os resultados foram feitas em conjunto, aluno e professor, com a colaboração do professor Fernando Mourão. Além disso, esse trabalho contou com o colaboração do aluno Alan Neves, que nos auxiliou na adaptação da técnica de Identificação de Usuários Influentes para o cenário avaliado nesse trabalho.

2. Trabalhos Relacionados

Existem basicamente duas abordagens para técnicas de AS, as supervisionadas e não-supervisionadas. Técnicas supervisionadas normalmente são adaptações de algoritmos de Aprendizado de Máquina tradicionais para aprender as classes relacionadas a sentimento em aplicações de Mídias Sociais. Esses cenários são, em sua maioria, constituídos de textos curtos, assim o desafio é construir modelos apropriados utilizando poucas informações [Brody and Diakopoulos 2011, Zhao et al. 2012, Hu et al. 2013]. Dada a dificuldade em se obter um conjunto de treinamento, métodos não-supervisionados estão assumindo um importante papel na busca de efetivas e eficientes abordagens para AS em conteúdos de mídias sociais. Grande parte desses métodos são baseados em lexicons e possuem dois passos distintos. No primeiro, a consolidação de um lexicon é realizada [Rocha et al. 2015]. Baseado em tal lexicon, o segundo passo foca na identificação do sentimento de cada postagem distinta [O'Connor et al. 2010]. A maioria dos trabalhos existentes focam em análises a nível de documento. O sentimento coletivo é derivado da agregação dos sentimentos de cada documento. Diferentemente, em [Rocha et al. 2015], os autores propõem um novo método que determina o sentimento coletivo diretamente, pela análise de um enorme grafo de termos, construído de acordo com a coocorrência de termos em cada documento.

Técnicas de IUI podem ser divididas em três grupos principais. O primeiro grupo consiste de estratégias que levam em conta a estrutura de redes estabelecidas através do relacionamento entre usuários. Nessa classe temos o trabalho que utiliza o algoritmo de *PageRank*TM para calcular uma pontuação de influência para cada usuário, considerando apenas relacionamentos e a propagação na rede [Page et al. 1999]. Outra estratégia desta classe é o PCC [Ilyas and Radha 2011], o qual usa uma métrica baseada em centralidade para determinar vizinhanças influentes em uma rede. O segundo grupo compreende estratégias que exploram o conteúdo e o fluxo de informação para determinar os opinadores. Neste grupo, estão presentes as estratégias ProfileRank [Silva et al. 2013] e Leitores Efetivos [Lee et al. 2010]. Enquanto o ProfileRank modela a difusão de informação considerando apenas a ordem temporal na qual as mensagens são propagadas em uma rede social, Leitores Efetivos avalia a difusão de informação como um efeito em cascata que tópicos têm sobre usuários. A terceira classe corresponde a estratégias focadas em sumarizações estatísticas de logs de atividade dos usuários. Estas estratégias objetivam determinar uma pontuação de influência para cada usuário de acordo com alguns de seus atributos, tais como número de seguidores e número de *posts* propagados. O principal obstáculo desses trabalhos existentes na literatura é que não há consenso entre eles [Neves et al. 2015].

Não identificamos muitos trabalhos que combinassem AS com IUI. Em [Bigonha et al. 2010] os autores apresentam uma métrica para determinar usuários influentes, baseado em três tipos de informação: a rede de menções, a polaridade de conteúdo dos tweets publicados (os tweets foram classificados individualmente e manualmente), e a qualidade destes tweets. Assim, torna-se possível determinar usuários influentes com viés negativos e positivos. Em [Bae and Lee 2012], é apresentada uma análise relacionando usuários famosos e influentes no Twitter (e.g., Barack Obama e Britney Spears) com suas audiências positivas e negativas, determinando assim seu grau de aprovação. Entretanto, questões tais como se é possível derivar sentimento coletivo pela análise de um pequeno número de opinadores ou quão forte é a correlação entre sentimento coletivo e os opinadores permanecem negligenciada.

3. Metodologia

Nessa seção, apresentamos uma metodologia de quatro passos para quantificar fatores distintos que afetam a análise combinada entre AS e IUI em domínios reais, bem como os potenciais benefícios dessa análise. Basicamente, a metodologia proposta leva em consideração a dinâmica temporal do domínio, a sensibilidade à amostragem dos métodos e a reciprocidade observada entre a opinião coletiva e as opiniões propagadas pelos usuários influentes (opinadores). É importante deixar claro que não estamos assumindo uma avaliação fechada e completa para todas as questões existentes. Nossa proposta é estabelecer direções de pesquisa promissoras para a área.

3.1. Análise de Dinâmica Temporal

A principal característica que afeta a análise combinada entre AS e IUI é a dinâmica temporal inerente ao domínio de análise. A premissa é que opiniões e opinadores temporalmente não alinhados em relação às oscilações observadas não estão diretamente correlacionados. De fato, sempre que a opinião coletiva oscila mais rapidamente do que o subconjunto de opinadores, ou vice-versa, pode não ser possível relacionar ambos. Por exemplo, em cenários nos quais o conjunto de opinadores não está adequadamente consolidado, devido à alta dinamicidade temporal, pode ser desafiador, ou mesmo impraticável, estimar o sentimento propagado por eles. Dessa maneira, esse passo objetiva mensurar o quanto dinâmico são as opiniões predominantes e os opinadores em um domínio.

A respeito da opinião coletiva, mensuramos sua dinâmica temporal como segue. Primeiro, derivamos a opinião coletiva O da amostra de dados inteira D , usando um método de AS existente na literatura. Especificamente, adotamos neste trabalho o método SACI [Rocha et al. 2015]. O SACI é relevante para o nosso objetivo uma vez que ele foi originalmente proposto para estimar, de forma eficiente, o sentimento coletivo em amostras de dados, ao invés de agregar o sentimento derivado para cada documento individual. Além disso, os autores demonstraram que o SACI é mais efetivo em estimar a opinião coletiva do que métodos de AS baseados em agregação. O SACI representa O como uma distribuição de probabilidades entre as classes de sentimento positiva, negativa e neutra. Dessa maneira, dividimos D entre unidades temporais de mesmo tamanho (e.g. dias, semanas, meses). Então, estimamos a opinião coletiva O_t usando apenas os *posts* pertencentes a cada unidade temporal distinta t . Finalmente, realizamos uma inspeção visual sobre as distribuições derivadas. Quanto mais dinâmico for um domínio, mais diferentes são as opiniões estimadas em unidades temporais distintas.

Por sua vez, a dinâmica temporal dos opinadores foi medida como segue. Primeiro, identificamos a lista ordenada L de top-k opinadores em D utilizando algum método de IUI. Especificamente, usamos aquele apresentado em [Neves et al. 2015], uma estratégia de *meta-learning* baseada em PCA que combina linearmente informações ortogonais exploradas por distintos métodos de IUI. Denominamos esse método de PCA-IUI. Escolhemos o PCA-IUI uma vez que ele combina estratégias não-consensuais em uma única, capaz de capturar perspectivas distintas. Novamente, consideramos unidades temporais distintas em D e derivamos uma lista distinta L_t de top-k opinadores para cada unidade de tempo t . Em seguida, comparamos cada lista L_t com L usando a correlação generalizada de Kendall's tau

3.2. Análise de Sensibilidade à Amostragem

Além de características do domínio, limitações inerentes aos métodos selecionados para realizar AS coletiva e a IUI podem afetar as nossas análises. Atenção especial deve ser dada para a sensibilidade desses métodos ao tamanho da amostra. É bem conhecido o impacto do tamanho da amostra no processo de aprendizado em áreas distintas [Zhao et al. 2012, Mourão et al. 2008]. A premissa, nesse caso, é que a amostra de dados disponível para análise é suficiente para prover um processo de aprendizagem adequado. Claramente, este requisito depende do método específico usado, bem como da própria tarefa de aprendizado. O desafio, entretanto, é como quantificar tal sensibilidade.

A fim de resolver essa questão, contrastaremos os resultados dos métodos de AS e IUI quando aplicados a amostras aleatórias de tamanhos distintos. Considere novamente a amostra original de dados D . Suponha que comecemos com uma amostra pequena S_i composta por X posts distintos aleatoriamente escolhidos de D , onde $X \ll |D|$. Começamos a análise gerando novas amostras S_{i+1} ao adicionar a S_i outros X posts presentes no conjunto $D - S_i$, aleatoriamente escolhidos sem repetição. Em cada amostra S_i aplicamos o SACI e o PCA-IUI contrastando os resultados com aqueles obtidos sob todo conjunto D . Para AS realizamos este contraste ao comparar as distribuições distintas por meio de uma métrica de aproximação de erros, a Distância Euclidiana. Quanto menor a Distância Euclidiana entre duas distribuições, maior a concordância entre elas. Para a IUI, usamos a correlação Kendall's tau, conforme previamente mencionado. Com o objetivo de obtermos robustez estatística, repetimos este processo 10 vezes e consideramos para avaliação os valores de média, bem como o desvio padrão.

3.3. Análise de Reciprocidade

A terceira questão se refere à reciprocidade entre a opinião coletiva de uma população e as opiniões propagadas pelos opinadores. Consideramos que essas duas informações são recíprocas sempre que a opinião coletiva corresponder à opinião predominante entre os opinadores, e vice-versa. Esta é uma informação relevante uma vez que listas top-k de opinadores podem não ser representativas em alguns domínios, emitindo opiniões não absorvidas ou propagadas por toda a população. De forma a realizar essa análise, dividimos a amostra de dados entre três subcoleções disjuntas D_c , de acordo com a classe de sentimento c (i.e., positivo, negativo e neutro) estimado para cada documento. Nesse caso, adaptamos o SACI para realizar análise individual seguindo as sugestões do artigo original. Em seguida, derivamos uma lista L_c de top-k opinadores usando apenas os posts e usuários presentes em cada subcoleção D_c . Por fim, comparamos cada uma das listas L_c com os top-k opinadores L derivados de D , usando a correlação Kendall's tau generalizada com penalidade $p = 0$. Quanto maior a correlação entre L_c (relacionada a classe predominante c) e L , mais recíprocas serão a opinião coletiva e a opinião dos opinadores.

3.4. Análise de Ganho de Informação

Por fim, propomos dois experimentos para avaliar os potenciais benefícios de combinar AS com IUI. No primeiro experimento visamos identificar comportamentos distintos dos opinadores em relação aos tipos de opinião comumente propagadas. Nesse sentido, primeiro aplicamos o PCA-IUI na amostra de dados total D , identificando uma lista ordenada L . Em seguida, para cada usuário influente distinto, aplicamos o SACI em seu conjunto de posts, derivando uma distribuição de sentimento para esse usuário. Finalmente,

agrupamos esses opinadores de acordo com as suas respectivas distribuições de sentimento. Cada grupo representa um padrão de comportamento distinto. Esse tipo de análise permite qualificar o impacto de opinadores no processo de difusão de informação na WEB.

O segundo experimento objetiva verificar se a opinião de poucos opinadores é suficiente para estimar a opinião coletiva da população inteira. Isso é relevante uma vez que poderia facilitar uma AS mais acurada em muitos domínios em que não é possível se obter uma amostragem representativas da população. Assim, derivamos novamente a lista L de top- k opinadores considerando a amostra de dados completa D . Em seguida, aplicamos o SACI no subconjunto de *posts* publicados somente pelos opinadores identificados, derivando uma distribuição de sentimentos para esse subconjunto. Por fim, calculamos a Distância Euclidiana entre essa distribuição de sentimento e a distribuição de sentimento derivada para D . Repetimos esse processo considerando valores distintos de k .

4. Estudo de caso

Essa seção visa validar a metodologia proposta e apontar evidências da relevância deste estudo em domínios reais. Nesse sentido, aplicamos a metodologia em uma amostra de dados coletada do *Twitter*. Iniciamos a discussão descrevendo a amostra de dados coletada bem como as configurações do ambiente de execução. Em seguida, apresentamos os resultados relacionados a cada passo da metodologia, juntamente com as principais conclusões e implicações para a área.

4.1. Coleção de Dados e Configurações do Ambiente

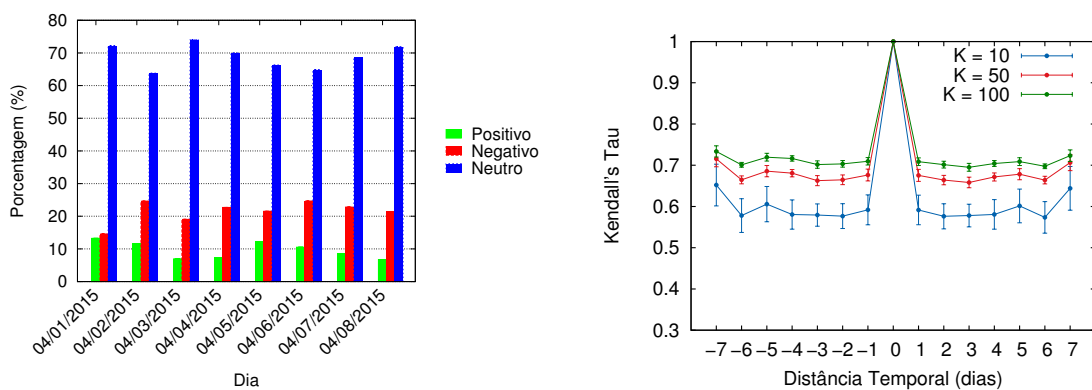
Utilizamos uma amostra de dados coletada do *Twitter*, que corresponde a *tweets* relacionado aos protestos políticos em 2015 em relação a presidente do Brasil, Dilma Rousseff. A coleta foi feita por meio da *API de Streaming* do *Twitter* utilizando as palavras-chave ‘Dilma’ e ‘Dilma Rousseff’. Removemos de cada tweet a pontuação, caracteres especiais, padrões repetitivos, bem como convertemos todas as letras para minúsculo. Além disso, removemos palavras com menos de três letras ou mais de 12 letras, URLs e menções a usuários do *Twitter*. Esses dados foram coletados entre **01-04-2015 e 08-04-2015** e contém **225,885 tweets**. Utilizando a API REST coletamos também informações sobre os **61,294 usuários** que publicaram os tweets coletados, tais como seguidores e amigos.

Conforme mencionado, em nossos experimentos adotamos para AS o algoritmo SACI [Rocha et al. 2015]. SACI tem quatro parâmetros relacionados à construção do *lexicon*, *suporte mínimo*, *confiança*, *número de sementes* e *distância máxima de propagação*, que foram fixados como 3, 0.8, 100 e 3, respectivamente. No passo de análise, o SACI tem o parâmetro *raio máximo de transformação*, definido como 4. Todos esses valores correspondem às configurações com os melhores resultados reportados pelos autores. No que se refere a IUI, adotamos a PCA-IUI [Neves et al. 2015], que corresponde a uma estratégia de *meta-learning* baseada em PCA que combina sete estratégias: PCC (número máximo de autovetores=100), Pagerank (erro máximo entre iterações consecutivas=0.0000001 e $p=0.85$), Profilerank (erro máximo entre iterações consecutivas=0.0000001), Closeness, Betweenness, Número de Retweets e Número de Seguidores.

4.2. Análise de Dinâmica Temporal

A fim de se avaliar a dinâmica temporal em nossa amostra de dados, a dividimos em dias, definindo oito subcoleções disjuntas. Começando as nossas análises pela Análise

de Sentimento, a figura 1 (a) retrata as distribuições de sentimento derivadas pelo SACI ao longo dos dias. Observamos distribuições bastante similares ao longo dos dias, com a classe neutra apresentando cerca de 65%-70% de porcentagem de ocorrência, seguido das classes negativa e positiva, respectivamente. Uma exceção é observada em 1º de abril, quando a porcentagem de *posts* negativos foi quase a mesma da porcentagem de positivos. A explicação para esse comportamento é que se trata do Dia da Mentira, resultando em muitos *posts* irônicos. Além disso, ao aplicarmos o SACI na amostra de dados completa, descobrimos que a distribuição de sentimento é de 63.70% de *posts* neutros, 23.91% de negativos e 12.39% de positivos. Esse resultado mostra que o sentimento coletivo estimado para qualquer dia é similar ao sentimento de todo o período de tempo avaliado, o que retrata o domínio avaliado como estável em relação à opinião coletiva.



(a) Opinião Coletiva Estimada ao longo do Tempo. (b) Opinadores Identificados ao longo do Tempo.

Figura 1. Dinâmica Temporal da amostra de dados coletadas a partir de Twitter.

A figura 1 (b) apresenta a dinâmica temporal dos opinadores. Neste experimento, avaliamos a similaridade dos top- k opinadores ao longo do tempo considerando valores distintos de k (e.g., $k = 10$, $k = 50$ e $k = 100$). Também plotamos os valores de Kendall's tau entre cada par de dias t_i e t_j usando a distância temporal entre eles (i.e., a distância é igual a $t_j - t_i$). Cada ponto neste gráfico representa, portanto, o valor de correlação média para todos os pares de dias cuja distância temporal é a mesma. Primeiramente, observamos que listas top- k pertencentes a dias distintos apresentam valores de correlação acima de 60% e quase inalterados conforme a distância temporal aumenta. Isto demonstra que este domínio também é estável em relação ao conjunto de opinadores ao longo do tempo, apresentando poucas alterações nas listas. Também observamos que quanto maior o tamanho das listas, maior é a correlação média das listas top- k distintas. Conforme esperado, pequenas alterações em listas pequenas de opinadores produzem grandes impactos nas correlações inter-listas ao longo do tempo, tal como observado para $k = 10$.

4.3. Análise de Sensibilidade à Amostragem

Em relação a sensibilidade dos métodos quanto à amostragem, as figuras 2 (a) e (b) apresentam os resultados para as tarefas de AS e IUI, respectivamente. Novamente, os valores plotados correspondem à média de 10 execuções. Observamos que, na amostra de dados avaliada, AS foi menos sensível ao tamanho da amostra do que a IUI. De fato, foi possível aproximar a opinião coletiva usando apenas 200 *tweets* (distância euclidiana menor que 0,05). Isso significa que usando apenas 0,08% da coleção foi possível aproximar a distribuição de sentimento alcançada usando a coleção inteira, superestimando ou subestimando a classe majoritária por 5% apenas. Por outro lado,

foi necessário considerar grandes amostras para identificar opinadores. Por exemplo, atingimos valores de correlação maiores do que 0,7 usando pelo menos 50.000 *tweets* (22% da amostra de dados). Esses resultados apontam que amostras pequenas de dados podem não ser suficientes para identificar efetivamente opinadores em alguns domínios.

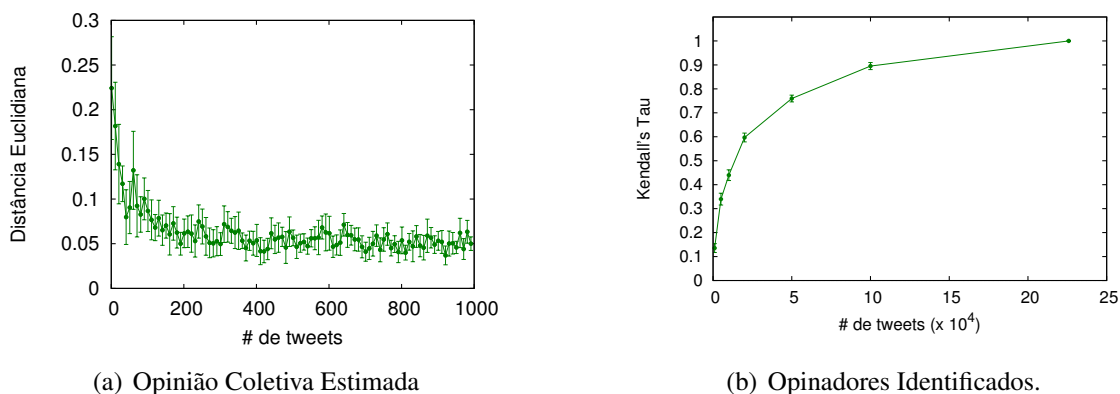


Figura 2. Sensibilidade a Amostragem do SACI (a) e PCA-IUI (b).

4.4. Análise de Reciprocidade

Avaliamos a reciprocidade entre a opinião coletiva e a opinião predominante dos opinadores variando o valor de k nas listas top- k de 5 até 50, tal como mostrado na figura 3. Nela podemos ver que a maioria dos top- k opinadores da amostra de dados total também são os opinadores que propagam a opinião coletiva predominante (a classe neutra, no caso), exibindo correlações inter-listas de cerca de 85% desconsiderando o valor de k . Por outro lado, opinadores que propaguem sentimento positivo em nossas amostras possuem o menor grau de correlação com os top- k opinadores na coleção, evidenciando uma baixa interseção dessas duas listas. Assim, a opinião predominante propagada pelos top- k opinadores em nossos dados é igual a opinião coletiva predominante.

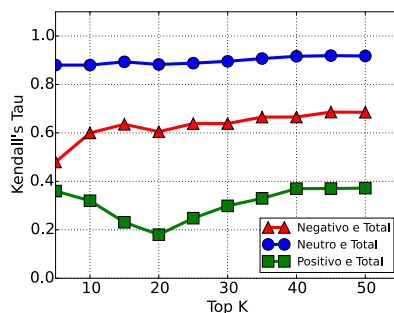


Figura 3. Análise de Reciprocidade entre SACI e PCA-IUI.

Esses resultados nos mostram que métodos de IUI estado-da-arte podem não identificar corretamente usuários individuais que agem como opinadores. Como a identificação de opinadores é enviesada para a opinião predominante e tal opinião é a classe neutra em vários domínios, os opinadores identificados são basicamente jornais ou outras mídias de comunicação. Além disso, é quase impossível identificar opinadores que propaguem opiniões que diverjam da predominante. Por exemplo, apesar de uma companhia estar interessada em identificar os opinadores que não gostaram de um determinado produto, os métodos de IUI atuais podem não ser capaz de encontrá-los. Filtrar *posts* negativos e aplicar os métodos de IUI nos dados resultantes pode não funcionar devido à sensibilidade à amostragem.

4.5. Análises de Ganhos da Combinação de AS e IUI

Por fim, avaliamos os potenciais benefícios de se combinar AS e IUI em nossa amostra de dados. Primeiramente, pretendemos identificar comportamentos distintos de opinadores em relação aos tipos de opinião geralmente propagadas por eles. A figura 4 (a) apresenta os resultados relacionados a este experimento. Para esse gráfico, usamos os top-200 opinadores identificados pelo PCA-IUI. Identificamos seis grupos com comportamentos distintos. O grupo 1 (G1) corresponde a usuários com uma posição negativa bem definida em relação à presidente do Brasil, enquanto o grupo 2 (G2) compreende usuários com uma posição positiva. O grupo 3 (G3) é composto por usuários que apoiam a presidente em alguns aspectos mas não em outros. Por sua vez, os grupos 4, 5 e 6 (G4, G5 e G6) representam usuários neutros, que publicam fatos em sua maioria, ora com uma tendência negativa (G5) e ora positiva (G6). Além disso, esses resultados mostram que AS pode melhorar a IUI, ao extrair informações úteis para o entendimento de como as opiniões se propagam.

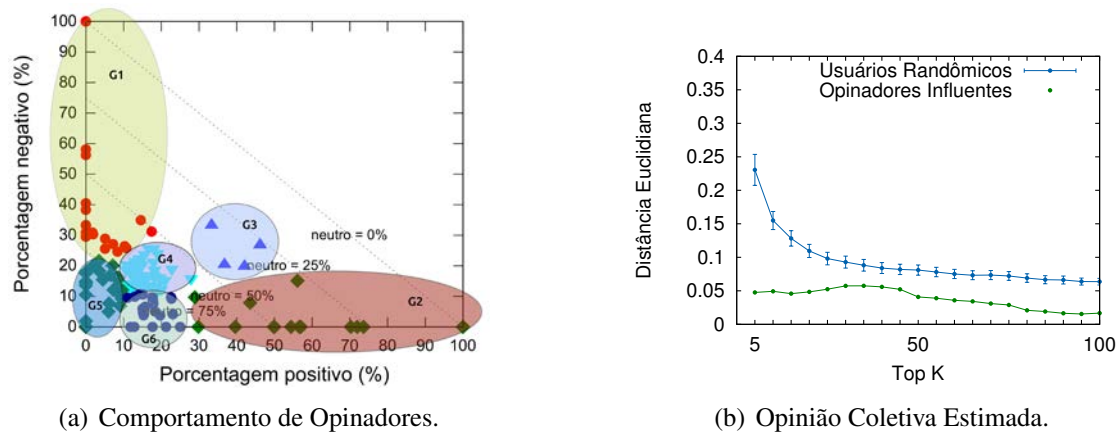


Figura 4. Análises dos benefícios de combinar AS e IUI.

No segundo experimento, verificamos se a opinião de poucos opinadores é suficiente para estimar a opinião coletiva de uma população inteira. A figura 4 (b) apresenta os resultados quando variamos os top- k opinadores usados no experimento entre $k = 5$ e $k = 100$. O gráfico também mostra os resultados do mesmo experimento quando usamos k usuários aleatoriamente escolhidos da amostra de dados. Cada ponto i da curva representa a distância euclidiana média da posição i em 100 execuções, bem como o desvio padrão. Observamos que a opinião definida pelos top- k opinadores está mais próxima da opinião coletiva do que a opinião relacionada a usuários aleatórios. Observe que usar poucos opinadores (e.g., os top-5) é suficiente para produzir uma aproximação melhor do que 100 usuários aleatórios. Além disso, o erro de aproximação (i.e., a distância euclidiana) convergiu rapidamente para ambos conjuntos de usuários. Esses resultados demonstram como técnicas de AS podem se beneficiar dos resultados da IUI.

5. Conclusões e trabalhos futuros

Nesse trabalho apresentamos uma metodologia para quantificar fatores que possam afetar avaliações combinadas de Análise de Sentimento (AS) coletiva e Identificação de Usuários Influentes (IUI) em domínios reais, bem como seus potenciais benefícios. Avaliamos nossa metodologia utilizando postagens relacionadas aos protestos políticos de abril de 2015. Em nossas avaliações identificamos um cenário de estabilidade temporal tanto para as opiniões coletivas quanto para opinadores e que o método de

Arquitetura e Implementação do Sistema Móvel de Assistência ao Idoso

Matheus Costa Stutzel¹, Alexandre Sztajnberg^{1,2,3}

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ

¹Bacharelado em Ciência da Computação - Instituto de Matemática e Estatística

²Pós-Graduação em Ciências Computacionais - Instituto de Matemática e Estatística

³Pós-Graduação em Eletrônica - Faculdade de Engenharia

matheusstutzel@yahoo.com.br, alexszt@ime.uerj.br

Abstract. *SMAI provides an infrastructure to monitor elderly patients with chronic degenerative disease that present functional loss. It is composed of Android applications and Web Service that provide authentication procedures and access to the database. The architecture and implementation of the system are presented and discussed. SMAI has been developed under the strict guidance of a health team and is being assessed in a quasi-randomized clinical trial with 30 +30 patients accompanied by NAI/UERJ.*

Resumo. *O SMAI oferece uma infraestrutura para o monitoramento de pacientes idosos com doença crônico-degenerativa que apresentem perda funcional. O sistema é composto de aplicações para o Android OS e de um conjunto de Web Services que provêm serviços de autenticação e de acesso ao banco de dados. A arquitetura e a implementação do sistema são discutidas. O SMAI foi desenvolvido segundo a orientação estrita da equipe de saúde e está sendo avaliado em um ensaio clínico quasi-randomizado com 30 + 30 pacientes acompanhados pelo NAI/UERJ.*

1. Introdução

O projeto do SMAI, Sistema Móvel de Assistência ao Idoso, envolve duas equipes: de computação, o LCC - Laboratório de Ciência da Computação e de saúde, o NAI – Núcleo de Assistência ao Idoso, ambos da UERJ. O público alvo do sistema é composto de idosos, com algum grau de dependência, e seus cuidadores. A equipe do NAI tem experiência no monitoramento e acompanhamento deste grupo, empregando os procedimentos e as práticas disponíveis.

O SMAI tem como objetivos (i) tornar a comunicação do cuidador com a equipe de saúde mais ágil; (ii) reduzir o estresse do cuidador; (iii) facilitar o cuidado e suporte do idoso nas suas atividades de vida diária; (iv) prover à equipe médica informações sobre o estado do paciente com maior frequência, de forma organizada, facilitando a tomada de decisões.

Como preparação, foi realizado um levantamento sobre o uso de recursos de computação e telefonia móvel, aplicando-se questionários ao grupo acompanhado pelo NAI. Os resultados obtidos constituíram importante informação para direcionar a abordagem proposta para o SMAI, desde a escolha pelo sistema Android [Google, 2003], até o desenvolvimento de interfaces interativas.

O sistema é composto de 2 aplicações: uma delas utilizada pelo profissional de saúde, otimizada para a utilização em um *tablet*, e a outra desenvolvida para *smartphone*, dedicada ao cuidador ou paciente. Um serviço de suporte também foi desenvolvido para dar acesso ao banco de dados, e inclui diversas ferramentas para auxiliar no monitoramento dos pacientes.

O presente artigo tem foco nos aspectos técnicos da arquitetura e implementação do SMAI. A arquitetura foi desenvolvida de maneira modular, facilitando a adição de novas funções e simplificando o reuso de módulos personalizados. Esta abordagem permite que a aplicação receba atualizações, mesmo durante o seu uso. As aplicações foram implementadas considerando as limitações de hardware dos *tablets* e *smartphones*. Aspectos de avaliação e usabilidade, também importantes no projeto, estão sendo desenvolvidos em conjunto com a equipe do NAI [Stutzel, 2016].

O aluno Matheus Stutzel, sob orientação do professor Alexandre Sztajnberg, participa do projeto desde o desenvolvimento do primeiro protótipo. É responsável pelo levantamento de requisitos junto à equipe de saúde e desenvolveu os sistemas descritos neste artigo. Em 2015 passou a coordenar outros alunos que entraram na equipe. Tem também acompanhado o processo de registro do sistema no INPI.

O restante do texto está estruturado da seguinte forma. A Seção 2 apresenta trabalhos relacionados. A Seção 3 apresenta uma visão geral da arquitetura do sistema proposto e a Seção 4 discute sua arquitetura. A Seção 5 aborda a implementação das aplicações desenvolvidas. A Seção 6 resume as avaliações do SMAI. Concluindo o texto, a Seção 7 traz as últimas considerações.

2. Trabalhos relacionados

O uso da telemedicina no controle de doenças crônicas tem sido discutido nos últimos anos, demonstrando benefícios para o paciente e para os sistemas de saúde [Dorsey, 2013], [Celler, 2006], [Jaglal, 2013], [Finkelstein, 2006]. Alguns deles se concentram em sistemas de notificação e lembretes aos pacientes, com o objetivo de melhorar a aderência ao tratamento [Fisher, 2006], [Keranen, 2013].

As atividades de desenvolvimento e pesquisa em novas tecnologias na área da telemedicina no Brasil estão começando a emergir. Considerando aplicações de telessaúde, algumas propostas já podem ser elencadas, como [Sanchez, 2012].

Em [Nicholas, 2007] são avaliados vários aspectos da tecnologia para e-Health, incluindo o impacto para o paciente, para o médico e para os serviços de saúde já estabelecidos. Um dos problemas recorrentes é o abandono do uso do sistema, por várias razões: falta de incentivo, falta de interesse e a não percepção de benefícios.

A revisão apresentada em [Mosa, 2012] classifica aplicações para *smartphones* apresentadas em 2894 artigos. A maior parte das aplicações está relacionada à diagnóstico de doenças ou calculadoras médicas. Apenas 15 das aplicações pesquisadas estão relacionadas ao monitoramento de pacientes com doenças crônicas. Destas 15 aplicações poucas se preocupam em tornar lúdico o uso da aplicação. No SMAI procuramos adicionar aspectos lúdicos e características de *gamificação* ao aplicativo destinado aos pacientes/cuidadores como forma de atrair o usuário ao uso contínuo do sistema (mais detalhes são discutidos em [Stutzel, 2016]).

3. SMAI - Sistema Móvel de Assistência ao Idoso

O SMAI é composto de duas aplicações: paciente/cuidador e médico, e um serviço de suporte, baseado em Web Service, hospedado na nuvem (Figura 1)

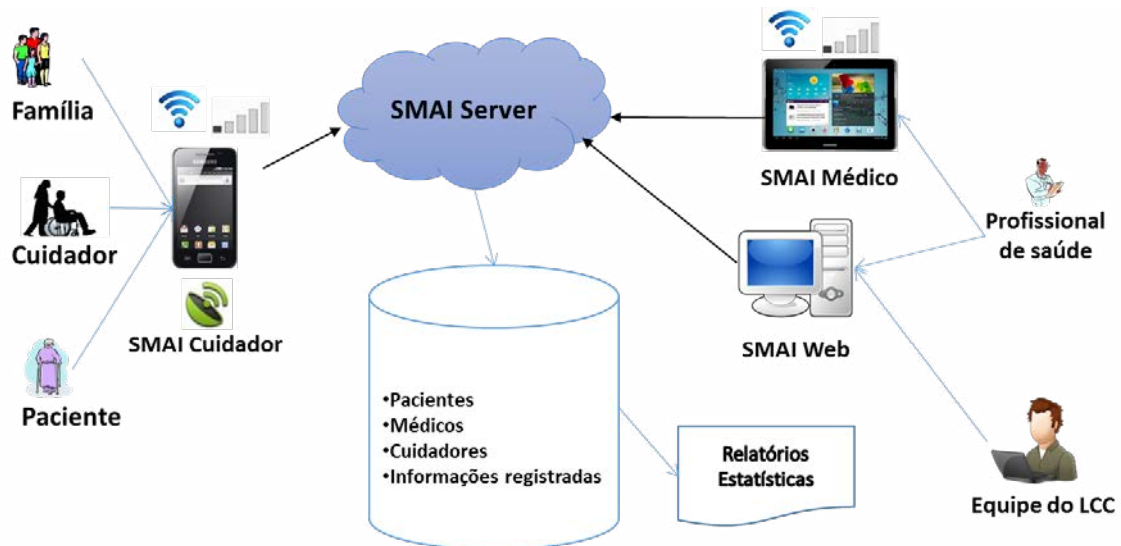


Figura 1. Estrutura geral do SMAI

Através de aparelhos *smartphone* uma aplicação (*app*), chamada SMAI Cuidador, permite ao cuidador/familiar, ou ao próprio paciente, enviar e receber informações e notificações para/da equipe de saúde:

- lembretes sobre uso de medicação e consultas agendadas;
- informações de medidas realizadas para acompanhar pressão arterial e glicemia;
- relatório diário com informações sobre o comportamento, rotina alimentar, evacuação e micção do paciente e relatório semanal sobre o próprio cuidador.
- recebimento de mensagens de suporte ao cuidador nas orientações de cuidado;
- envio periódico da localização do paciente, usando uma combinação de possibilidades como o GPS e sistemas de localização baseado na rede.

A aplicação utilizada pelo cuidador/familiar tem o objetivo de auxiliar seu dia a dia e facilitar o envio de informações para o NAI. A aplicação quando utilizada pelo paciente tem um número reduzido de opções, cujo objetivo é receber deste paciente informações sobre seu estado geral, mantendo uma conexão “virtual” com o NAI. Este versionamento da aplicação é realizado utilizando a ferramenta de *flavors* disponibilizada pela IDE utilizada, o Android Studio.

Os profissionais da saúde têm acesso ao sistema através de um *tablet*. A aplicação SMAI Médico apresenta a visualização de todos os pacientes de forma estruturada, e permite o monitoramento e a comunicação individual ou com o grupo de pacientes. Entre as funções estão:

- visualizar alarmes e informações atualizadas dos pacientes, em grupo;
- visualizar detalhes de cada paciente, incluindo a sua localização atual e dos últimos dias, permitindo verificar sua mobilidade;
- envio de agendamento de consulta e lembretes de medicamentos, além de mensagens para usuários individuais ou grupos de pacientes;

- visualizar os relatórios relacionados aos pacientes e o relatório do cuidador.

A troca de informações entre as aplicações dos pacientes e da equipe médica é feita através de um serviço desenvolvido para o sistema, o SMAI Server. Este serviço permite o armazenamento das informações geradas pelos usuários, além de realizar a autenticação e o controle de acesso aos dados dos pacientes. Integrada ao SMAI Server também é disponibilizada uma visualização alternativa das informações através de uma página Web, o SMAI Web, que permite verificar a utilização do sistema, relatórios de erros e o uso da rede dos dispositivos móveis na transmissão de dados das aplicações.

4. Arquitetura

A arquitetura do SMAI é baseada no padrão *Model-View-Controller* (MVC) [Krasner, 1988]. O sistema foi planejado para que cada categoria de informação (fisiológica, estado, relatório) possuísse seu próprio módulo de entrada. Desta maneira, a alteração ou adição de uma nova categoria pode ser realizada de maneira simples.

O sistema foi desenvolvido para ser utilizado em *smartphones* e *tablets* Android e faz uso de diversos serviços disponibilizados pelo sistema operacional. A integração entre os módulos da aplicação e o Android é realizada através das classes *Super**, com destaque para a *SuperActivity* (Figura 2).

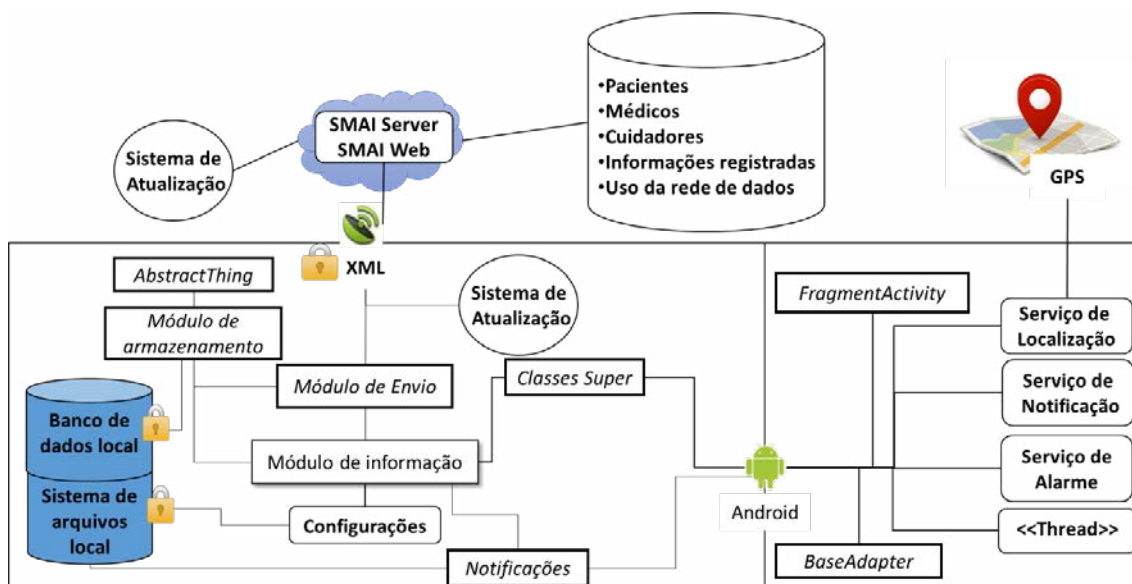


Figura 2. Arquitetura do SMAI Cuidador e SMAI Médico

Cada módulo de informação possui as interfaces necessárias para a entrada dos dados ou para a exibição dos mesmos, de acordo com os requisitos da aplicação. Por exemplo, o módulo responsável pelos registros de dor possui a entrada de dados no SMAI Cuidador e a exibição no SMAI Médico, já no caso dos remédios a equipe de saúde fica responsável pela entrada dos dados, enquanto o cuidador pode apenas visualizar os remédios atuais e confirmar se eles foram tomados.

O Módulo de Envio é responsável por registrar/obter novos dados consumindo serviços do SMAI Server, de acordo com a interação do usuário ou em períodos preestabelecidos. Para a transmissão dos dados é realizada uma conversão da informação de acordo com as especificações da respectiva classe *AbstractThing*.

O armazenamento dos dados das aplicações é realizado utilizando o banco de dados local através do Módulo de Armazenamento. Entretanto as configurações e as notificações são armazenadas utilizando o sistema de arquivos local de forma a se utilizar os serviços disponibilizados pelo Android, tal como o *FileObserver*, que dispara eventos de acordo com a criação/alteração de determinados arquivos/diretórios.

A arquitetura do SMAI oferece a possibilidade da reutilização dos módulos, bem como a fácil adição de novos componentes. Um Sistema de Atualização especialmente desenvolvido permite que novas versões da aplicação, com melhorias e correções, sejam disponibilizadas, sem que haja interrupção no serviço.

5. Implementação

Os elementos da arquitetura do SMAI foram implementados para avaliação. As aplicações SMAI Cuidador e SMAI Médico foram desenvolvidas em Java para Android, compatível com a versão 4.0 e superiores, enquanto o SMAI Server utiliza o GlassFish [Oracle, 2010] como servidor de aplicação e seus Web Services são disponibilizados por uma API RESTful sobre HTTPS. A implementação SMAI Server encontra-se em sua 2ª versão, enquanto as aplicações Android encontram-se na 5ª versão.

As funcionalidades, painéis de informação e notificações foram implementadas de acordo com a orientação e requisitos da equipe de saúde do NAI, seguindo as práticas e padrões de projeto recomendados pela Google.

A interface com o usuário foi desenvolvida de modo a se manter simples e a possibilitar que em poucos cliques o usuário fosse capaz de registrar as informações desejadas, sejam, por exemplo, registros de pressão arterial ou a confirmação de que um remédio foi tomado (Figura 3). Assim que o usuário seleciona a opção “enviar”, os dados são criptografados e armazenados localmente, em seguida é iniciado o Módulo de Envio, responsável por ler essas informações em disco, autenticar a aplicação no servidor e realizar a transmissão dos dados, utilizando XML, baseado nas classes *AbstractThing*. Em caso de falha o processo é reiniciado, caso contrário o arquivo local é apagado. Atualmente a autenticação e o envio são realizados utilizando chamadas RESTful no SMAI Server e toda a comunicação é realizada através de HTTPS.

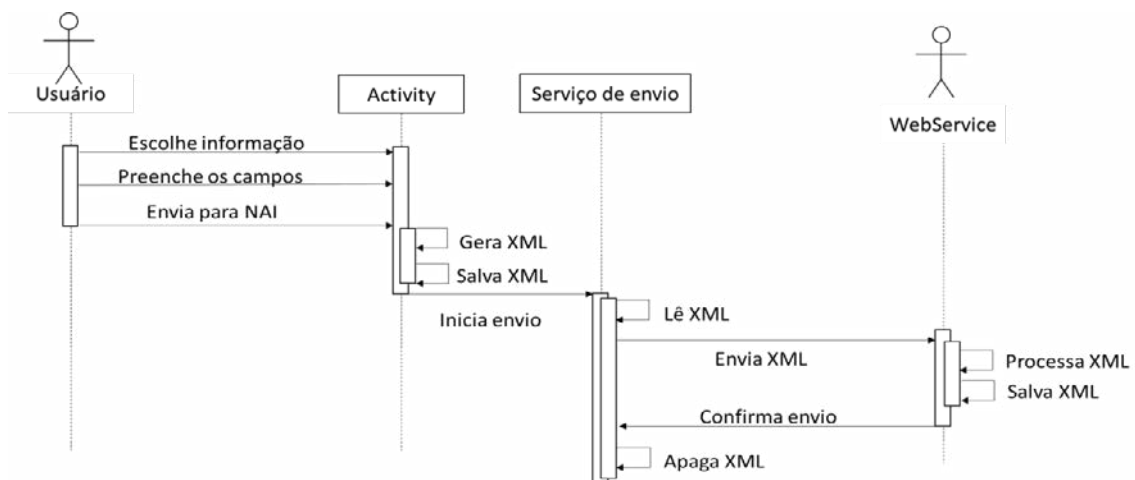


Figura 3 – Sequência do registro de uma informação no sistema

A autenticação do paciente é realizada na primeira inicialização das aplicações, quando as credenciais do usuário são solicitadas, verificadas e armazenadas, de maneira segura, localmente, dispensando que o usuário se identifique a cada ação no sistema. O acesso do profissional de saúde aos dados dos pacientes é controlado por mecanismos de permissão e compartilhamento autorizado. Os agendamentos, lembretes de medicações e notificações para os pacientes têm associada a informação do profissional de saúde responsável.

As notificações no SMAI Cuidador e Médico utilizam um mecanismo especialmente adaptado. No momento em que o usuário precisa ser notificado, o sistema executa três ações: (i) a notificação padrão do Android é acionada; (ii) uma notificação “flutuante” é ativada - um ícone que é exibido “sobrepondo” uma parte do aplicativo que o usuário estiver utilizando – tornando-se um aviso insistente para o cuidador realizar alguma tarefa no sistema e, (iii) também é armazenado localmente um registro, criptografado, contendo informações sobre a notificação. Este registro permite que o SMAI gerencie as notificações. Por exemplo, se o tipo de notificação demanda do cuidador uma confirmação, esta notificação é apagada apenas quando esta confirmação é feita. Em outros casos, como por exemplo, o Relatório Diário, que gera uma notificação todos os dias, se o cuidador não fizer o relatório num dia, o sistema evita "acumular" os Relatórios. Assim, uma notificação antiga é substituída pela atual, evitando sobrecarregar o cuidador com dois relatórios no mesmo dia. Observamos que este foi mais um dos requisitos específicos da equipe de saúde.

5.1. Aplicação do cuidador (SMAI Cuidador)

A aplicação do cuidador possui um menu principal, com módulos de entrada de informações, por exemplo *Dor*, *Pressão Arterial* e *Temperatura* (Figura 4). Alguns módulos são cientes de contexto, acionando ações específicas. Por exemplo, se a temperatura ultrapassar $37,8^{\circ}$ é solicitada a informação sobre tosse. Estes limites são determinados pela equipe de saúde. O usuário possui ainda, a possibilidade de acionar o Alarme NAI, para que um profissional de saúde seja notificado e entre em contato com o paciente de acordo com o protocolo adotado. Ao ser carregado, o programa também cria *threads* separadas para os serviços de segundo plano, por exemplo, o envio da localização do paciente e a verificação de atualizações.

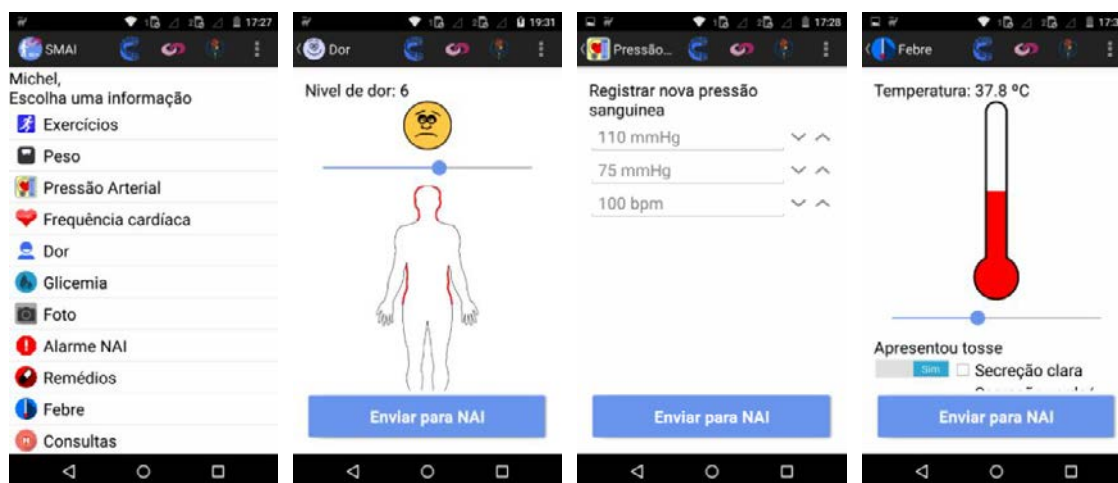


Figura 4. Tela principal e envio de informações

O mecanismo de *Alarm* do Android é usado para agendar determinados eventos recorrentes, como, por exemplo, diariamente uma notificação é gerada às 20:00 como lembrete para o cuidador preencher um relatório (Figura 5). O usuário pode preencher os itens de acordo com sua percepção de importância ou indicação da equipe de saúde. Este é um dos módulos desenvolvidos especificamente para o conjunto de pacientes monitorados e aderentes à prática adotada pelo cuidador. Já existe uma rotina para que informações do dia a dia do paciente sejam anotadas em uma planilha em papel, de forma que a equipe de saúde possa acompanhar a evolução do paciente e intercorrências. O relatório através do SMAI objetiva tornar esta rotina mais ágil para o cuidador e proporcionar um acompanhamento mais próximo da equipe de saúde, já que os relatórios ficam disponíveis imediatamente.

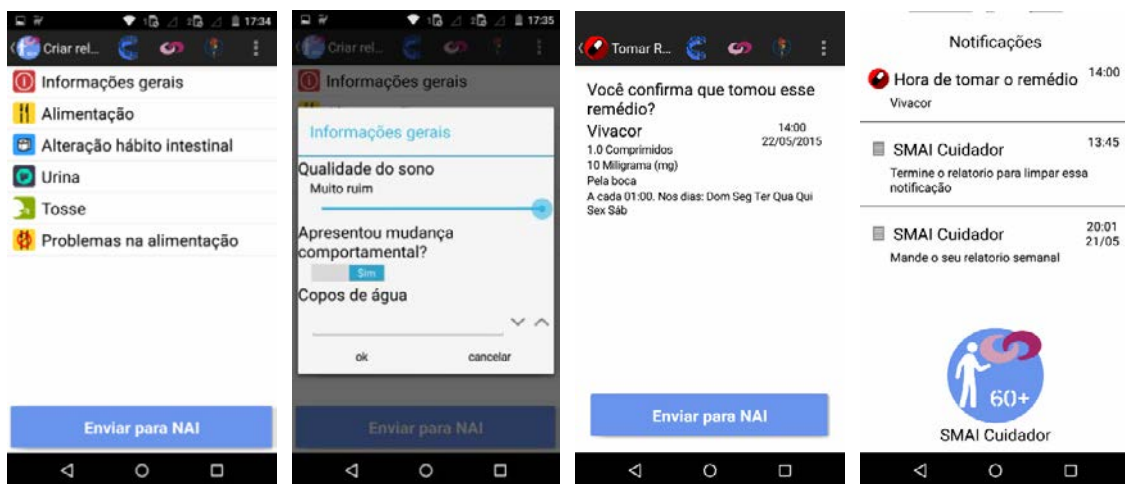


Figura 5. Relatórios, notificações e lembretes

Outras notificações, como a ingestão de remédios também são enviadas como lembrete (Figura 5). A confirmação das mesmas fica registrada para a equipe de saúde.

5.2. Aplicação do profissional de saúde (SMAI Médico)

A arquitetura do SMAI Médico se assemelha ao SMAI Cuidador (Figura 2), com pontos diferentes na estrutura da implementação, como interface com o usuário e a utilização da API de mapas da Google, além do gerenciamento de vários pacientes.

A aplicação tem como base uma lista de pacientes monitorados, e um mecanismo automático de atualização de informações que traz do SMAI Server as últimas informações enviadas. Esta lista é persistida em um banco de dados local, e exibida na forma de um Painel de Controle (Figura 6(a)). Desta forma, é minimizada a necessidade de se obter os dados do servidor a cada interação do profissional de saúde.

Quando um cuidador aciona o Alarme NAI, o paciente associado é exibido com a borda em vermelho e o profissional de saúde é notificado. Este alarme é atualizado a cada 2 minutos, com isso ele é recebido rapidamente pela equipe de saúde, tornando menor o tempo de resposta ao paciente. O sistema também utiliza esta atualização para verificar no servidor a necessidade de atualizar as demais informações, economizando assim a utilização dos recursos. Caso haja novas informações disponíveis o sistema faz o download completo e de forma automática, porém em um intervalo de tempo maior. A qualquer momento a equipe de saúde pode solicitar os dados mais recentes, seja de um determinado paciente ou de todo o grupo.

Quando um paciente é selecionado, um menu similar ao do SMAI Cuidador também é apresentado (Figura 6(b)), com opções adicionais, como o gráfico das medidas fisiológicas transmitidas e o registro de localização do usuário (Figura 7).

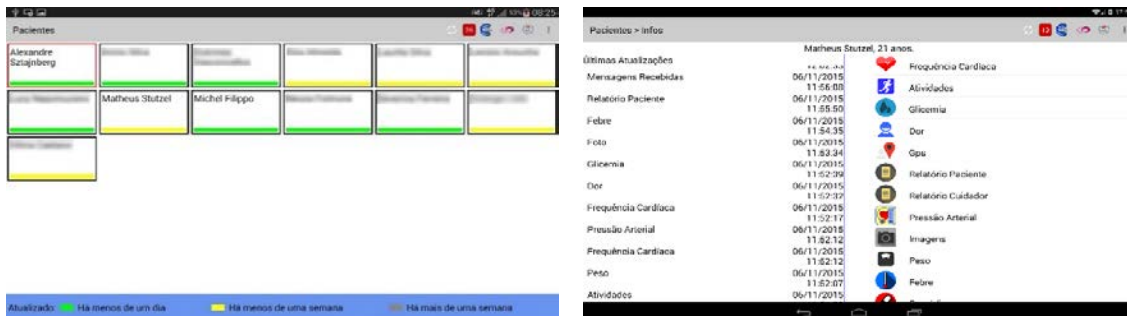


Figura 6 (a) Painel de controle, (b) detalhe de um paciente



Figura 7 (a) Detalhes do histórico de dor; (b) Localização do paciente

Algumas opções permitem que se listem os itens enviados pelo paciente, na forma de relatórios ou que se adicionem informações para serem notificados para o paciente. Este é o caso dos Medicamentos, que podem ser adicionados/alterados/descontinuados para cada paciente. Além disso, é possível acompanhar o registro dos medicamentos que tiveram sua notificação confirmada por um cuidador.

Para algumas informações fisiológicas, como a Pressão Arterial, o profissional de saúde pode individualizar os parâmetros para cada paciente. Deste modo, se uma informação deste tipo estiver fora dos limites estipulados individualizados, uma notificação é gerada pelo SMAI Médico para alertar o profissional.

5.3. Servidor de suporte (SMAI Server) e Interface Web (SMAI Web)

O SMAI Server é responsável por prover toda a infraestrutura para o armazenamento, autenticação, controle de acesso e transmissão de dados para as aplicações Android.

O SMAI Web foi desenvolvido com o objetivo de prover o suporte à equipe de saúde e aos pacientes e é composto por módulos que possibilitam monitorar as informações registradas, a quantidade de dados usados para a transmissão destas informações e gerar relatórios de falhas no sistema.

Nas primeiras etapas do desenvolvimento o sistema fez uso do Microsoft Health Vault [Microsoft, 2007]. Atualmente é utilizado um servidor de aplicação independente. Este servidor é composto por aplicações Java, que oferecem seus serviços através de Web Service e por uma interface Web. Para oferecer a compatibilidade entre versões diferentes das aplicações, a API utilizada na etapa anterior foi reproduzida na nova versão.

Através dos dados obtidos pelo SMAI Web (Figura 8), é possível confirmar, por exemplo, o aumento do consumo de dados ao se lançar uma nova versão do sistema. É importante destacar que houve um aumento no consumo de dados no mês de Fevereiro de 2016 devido a migração do sistema para o SMAI Server. Outro ponto importante a ser observado é que a transmissão de dados pelos módulos iniciados a partir das notificações é maior (Relatórios e Remédios tomados).

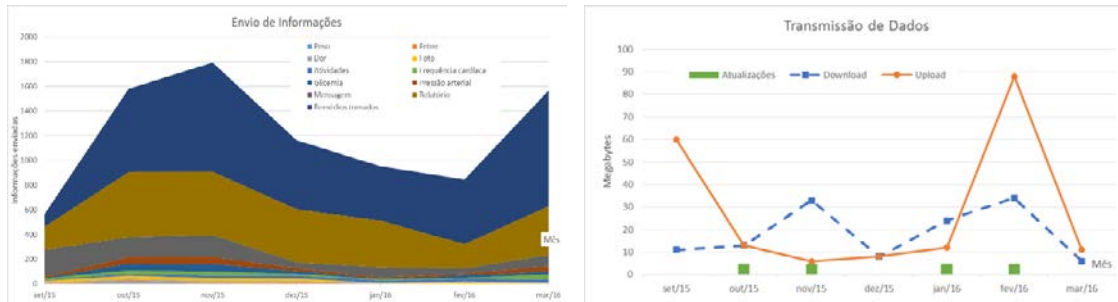


Figura 8. (a) Quantidade de informações enviadas (b) consumo de dados

Atualmente o SMAI consome, por usuário, uma média de 28 MBytes de upload e 18 MBytes de download por mês. O sistema registra uma média de 1200 informações inseridas pelos usuários e mais de 8900 localizações registradas automaticamente.

6. Avaliação

Estudo piloto. No início de 2015 o protótipo do SMAI passou por uma primeira avaliação. Um grupo de 5 cuidadores recebeu um aparelho com a versão SMAI instalada. O sistema deveria ser usado por 15 dias, da melhor forma, sem compromisso de fornecer dados reais. Uma reunião foi realizada para a entrega dos aparelhos e um treinamento simples no o uso das principais funções, *hands-on*, foi oferecido pela equipe de Computação. Este treinamento durou em média 40 minutos.

Durante a reunião marcada para a devolução dos aparelhos, foi realizada uma pesquisa informal de opinião com os cuidadores com destaque para os seguintes pontos: (i) todos os cuidadores faziam uso contínuo do sistema; (ii) alguns cuidadores não entenderam completamente algumas opções; (iii) sugestões foram registradas no sentido de mudar algumas opções ou acrescentar a possibilidade de informar detalhes.

Teste clínico. Um teste clínico *quasi-randomizado* está sendo realizado para avaliar o sistema. 30 cuidadores receberam um aparelho Motorola Moto G com um plano de dados de 50 Mbytes para utilizar o SMAI Cuidador e um grupo de controle de 30 cuidadores é acompanhado com questionários por telefone. A duração prevista é de nove meses. Reuniões de grupo focal são realizadas periodicamente, além das consultas ambulatoriais programadas. O retorno dos cuidadores tem sido muito positivo.

7. Conclusão

O SMAI foi desenvolvido com uma arquitetura modular, visando facilitar o acréscimo de novas funções e o lançamento de novas versões, o que foi mais frequente no início de sua avaliação. A implementação desta arquitetura foi realizada de acordo com os requisitos obtidos com os profissionais de saúde do NAI e a partir dos padrões sugeridos pela Google. Os usuários tem acesso ao sistema através de *smartphones* e *tablets* e a interface foi desenvolvida de modo a simplificar a interação com o aplicativo.

O SMAI conta ainda com um Web Service, que realiza a comunicação entre o aplicativo dos cuidadores e a aplicação da equipe de saúde. Além disso, o serviço também disponibiliza uma página Web para a equipe de desenvolvimento verificar a utilização do sistema, relatórios de erro e o consumo de dados das aplicações.

Vários aprimoramentos tem sido incluídos por demanda da equipe de saúde e dos próprios cuidadores. Também está sendo desenvolvida uma versão das aplicações voltada para o sistema móvel iOS, principalmente para atender a equipe de saúde.

Agradecimentos. Agradecemos o auxílio financeiro da FAPERJ e CNPq e o apoio do InovUERJ no processo de registro do software junto ao INPI e nos pedidos de patente.

8. Referências

- Celler, B. G., Basilakis, J., Budge, M. e Lovel, N. H. (2006) “A clinical monitoring and management system for residential aged care facilities”, 28th IEEE EMBS Annual Intl. Conferences, pp. 3301-3304, New York, EUA.
- Dorsey, E. R. (2013) “Randomized controlled trial of “virtual houses calls” for Parkinson disease”, JAMA Neurology, vol. 70, no. 5, pp. 565- 70.
- Finkelstein, S. M., Speedie, S. M. e Potthoff, S. (2006) “Home telehealth improves clinical outcomes at lower cost for home healthcare”, Telemed. Jnl. and E-health, vol. 12, no. 2, pp. 128-38.
- Fisher, G. S. (2006) “Brief Communication: Electronic Reminds to Patients Within an Interactive Patients”, Telemed. Jnl. and e-Health, vol.12, no. 2, pp. 128-38.
- Google, “Android”, <http://www.android.com/about/>
- Jaglal, S. B., Haroun, V. A., Salbach, N. M., Hawker, G., Voth, J. e Lou, W. (2013) “Increasing access to chronic disease self-management programs in rural and remote communities using telehealth”, Telemed. Jnl. and e-Health, vol.19, no.6, pp.467-74.
- Krasner, G. E., Pope, S. T. (1988) “A description of the model-view-controller user interface paradigm in the smalltalk-80 system”, Journal of object oriented programming 1.3, pp. 26-49.
- Keranen, T. e Liikkanen, S. (2013) “Medication reminder service for mobile phones: an open feasibility study in patients with Parkinson's disease”, Telemed. Jnl. and e-Health, vol.19, no. 11, pp. 888-90.
- Microsoft, “HealthVault”, <https://www.healthvault.com/br/en>
- Mosa, A. S. M., Yoo, I. e Sheets, L. (2012) “A Systematic Review of Healthcare Applications for Smartphone”, BioMed Central – Medical Informatics and Decision Making, vol. 12, pp. 67.
- Nicholas, D., Huntington, P. e Jamali, H. (2007) Digital Health Information for the Consumer: Evidence and Policy Implications, Ashgate Publishing, Ltd., England.
- Oracle, “GlassFish”, <http://www.oracle.com/technetwork/middleware/glassfish/overview/index.html>
- Sanches, L. M., Alves, D. S., Lopes, M. H. e Novaes, M. A. (2012) “The practice of telehealth by nurses: an experience in primary healthcare in Brazil”, Telemed. Jnl. and e-Health, vol. 18, no.9, pp. 679-83.
- Stutzel, M. C, Filippo, M. P., Sztajnberg, A., Brittes, A., Motta, L. B. (2016) “SMAI – Mobile System for Elderly Monitoring”, 4th IEEE SeGAH International Conference on Serious Games and Applications for Health, Orlando, EUA.

Sobre quadrados mágicos, grafos graciosos e colorações especiais: teoremas, algoritmos e aplicações*

Victória Aires, Rosiane de Freitas

Instituto de Computação – Universidade Federal do Amazonas (UFAM)
69077-000 – Manaus – AM – Brasil

{vpsa,rosiane}@icompu.ufam.edu.br

Abstract. *In this study, graph labeling problems were investigated. This consists of assigning labels, usually integers, on vertices or edges, or both, to respect some constraints. Labeled graphs include the magical graphs, graceful graphs and colorings with distances. Magic squares can be modeled as complete bipartite graphs, which in turn can have a graceful labeling. Such labeling can be seen as a particular case of graph coloring under distance constraints. These correlations were explored in this work, being rewritten in details the proof of the main theorems, as well as implementations of algorithms for recognition and determination of feasible solutions - exact and greedy heuristics, using the DIMACS benchmark. NP-completeness proof for coloring with equal distances is given.*

Resumo. *Neste trabalho foram investigados problemas de rotulação em grafos, que consistem em atribuir rótulos, geralmente números inteiros, a vértices ou arestas, ou ambos, de modo a respeitar alguma restrição. Grafos rotulados incluem os grafos mágicos, grafos graciosos e colorações com restrições de distâncias. Quadrados mágicos podem ser modelados como grafos bipartidos completos, que por sua vez podem ter uma rotulação graciosa. Tal rotulação pode ser vista como um caso particular de coloração de vértices que respeita certa restrição de distância. Estas correlações foram exploradas nesta pesquisa, sendo reescritas em detalhes as provas dos principais teoremas, bem como implementados algoritmos de reconhecimento e determinação de soluções factíveis - exato e heurísticas gulosas, com uso do benchmark DIMACS. A prova da NP-completude para coloração com distâncias iguais é apresentada.*

1. Introdução

A área de teoria dos grafos estuda as relações entre os elementos de um conjunto, em uma estrutura matemática principal conhecida por **grafos**. Deste modo, é dado um grafo $G = (V, E)$, onde V é o conjunto de **vértices**, que representam os objetos, e onde E o conjunto de **arestas**, representando os relacionamentos entre eles. Dependendo do problema em estudo, podem ser acrescentadas outras características a um grafo. A classe de **grafos rotulados** reúne aqueles que possuem rótulos nas arestas, vértices ou ambos, que geralmente são valores numéricos que podem representar pesos ou distâncias [Feofiloff et al. 2011].

*Apoiado pelo Programa Institucional de Bolsas de Iniciação Científica - PIBIC (UFAM e CNPq).

Neste trabalho, foram investigados alguns problemas pertencentes a esta classe, sendo: **grafos mágicos**, **grafos graciosos** e problemas de **coloração com restrições de distâncias**. Em uma rotulação mágica, existe uma soma constante associada a cada vértice. Já em uma rotulação graciosa, as arestas devem receber uma sequência de números naturais como rótulos [Watson 1972]. Uma outra rotulação conhecida é a chamada coloração de vértices, onde os rótulos são interpretados como cores e vértices adjacentes devem ter cores diferentes [Kosowski and Manuszewski 2004]. Aqui, o foco está numa variação deste problema, acrescentando uma restrição relacionada a rótulos nas arestas. Deste modo, os objetivos da pesquisa são: explorar as correlações entre os problemas, reescrever os principais teoremas, compreender propriedades identificadas anteriormente e auxiliar na proposição de novas, além de implementar algoritmos para rotulação graciosa de algumas classes específicas e coloração utilizando heurísticas gulosas.

Este artigo está organizado do seguinte modo. Na Seção 2 é apresentado o problema do quadrado mágico e sua modelagem teórica em grafos, envolvendo rotulação mágica e aplicações. A Seção 3 traz os aspectos referentes aos grafos graciosos, suas características, teoremas e algoritmos. Já na Seção 4, são abordadas variações de problemas de coloração clássica, mas além da rotulação de vértices, envolvem restrições de distância vistas como um tipo de rotulação de arestas. A Seção 5 sumariza as contribuições do trabalho. Por fim, na Seção 6, são feitas as considerações finais do artigo.

2. Quadrado Mágico: modelagem em teoria dos grafos

Um quadrado mágico é uma tabela de números, onde a soma dos elementos de uma linha, coluna e diagonais é constante [Arsie 2010]. Não há repetição de números. A constante é denotada por S_n e chamada de **constante mágica**. Esta é dada por $S_n = \frac{(1+n^2) \cdot n}{2}$. Assim, a soma varia de acordo com a ordem n do quadrado.

Existe uma classe de grafos rotulados útil na modelagem do quadrado mágico: os **grafos mágicos**. Um grafo é dito mágico se as arestas podem ser rotuladas com inteiros não negativos de modo que a soma dos rótulos das arestas incidentes a um vértice é a mesma, independente da escolha do vértice [Arnold 2012]. Pode-se verificar que ao representar um quadrado mágico em grafos, obtêm-se uma configuração semelhante à da Figura 1, ou seja, uma rotulação mágica. Cada vértice representa uma linha ou coluna e a soma dos rótulos das arestas associadas é constante, para todos os vértices. Como já comentado, um grafo mágico é um tipo de grafo rotulado que admite uma rotulação mágica, assim como os grafos graciosos, que são aqueles que possuem uma rotulação graciosa, tal como abordado na próxima seção.

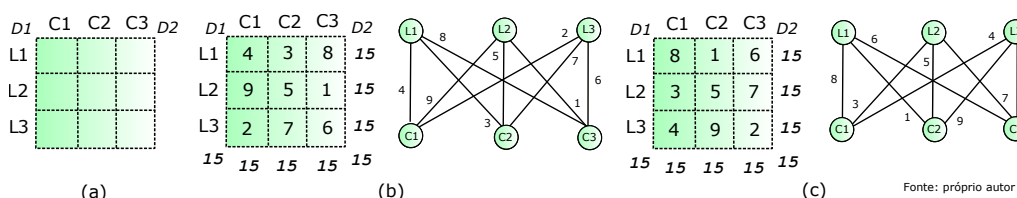


Figura 1. Grafo mágico representando um quadrado mágico 3×3 . Em (a), uma instância inicial não preenchida. Em (b) e (c), configurações que satisfazem a propriedade de quadrado mágico. A soma dos rótulos das arestas adjacentes a cada vértice é sempre 15, a constante mágica para ordem 3.

Se o objetivo for implementar um algoritmo para calcular todos os quadrados

mágicos possíveis para uma dada ordem n , uma possibilidade seria o uso da técnica de **enumeração explícita (*backtracking*)**, ou seja, testar todas as combinações possíveis. No entanto, esta alternativa é muito custosa computacionalmente, uma vez que para um quadrado de ordem n , é possível preencher o quadrado com n^2 números que podem ser dispostos de maneiras diferentes até satisfazer a restrição do quadrado mágico. Além disso, há um algoritmo polinomial que utiliza cálculos matemáticos simples para gerar apenas uma solução para um quadrado mágico de ordem n .

2.1. Aplicações dos quadrados mágicos

Os quadrados mágicos aparecem na literatura relacionados a outros problemas, como jogos que também tem como objetivo preencher um *grid* de números, como o **quadrado latino**. Criado por Leonhard Euler, matemático também interessado no estudo dos quadrados mágicos, o jogo consiste em um quadrado com n^2 elementos que devem ser dispostos de forma que cada linha e cada coluna possua cada um dos n números diferentes, sem repetição [dos Santos et al. 2007]. O jogo **Sudoku**, por exemplo, é um quadrado latino com uma restrição a mais: além das linhas e colunas, existem subdivisões do *grid*, e nelas também não pode haver repetição de números [Arsie 2010].

Uma outra aplicação interessante é no **problema do passeio do rei**. Nele, o objetivo é encontrar um passeio do rei em todas as casas de um tabuleiro de xadrez. Ele só pode passar em cada casa apenas uma vez e se move para apenas uma casa em qualquer direção. Em uma das soluções, demonstrada na Figura 2, o passeio feito pelo rei resulta em um quadrado mágico 8×8 [Gherzi 1921]. Os números do quadrado mágico indicam a ordem em que cada casa foi visitada. As linhas, colunas e diagonais do quadrado mágico feito pelo passeio do rei somam 260, que é o valor da constante mágica para o tamanho 8.

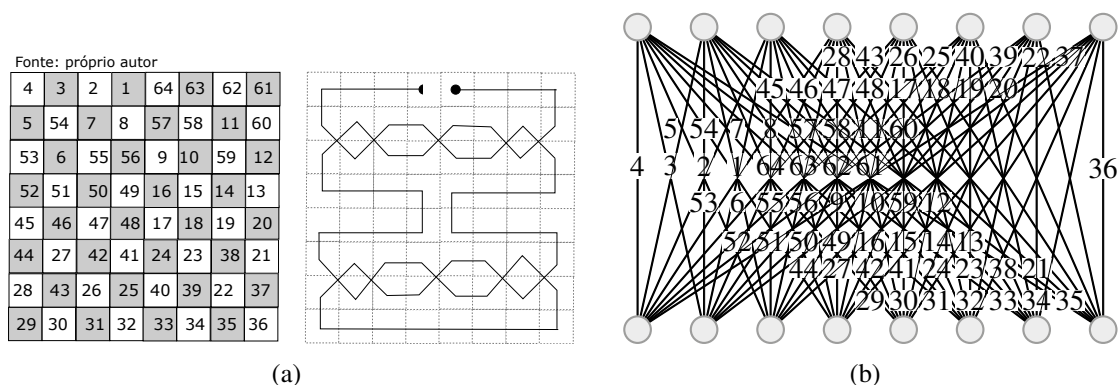


Figura 2. Uma solução para o passeio do rei que resulta em um quadrado mágico. Em (a), têm-se o quadrado mágico 8×8 representando a ordem de visita para cada casa, onde a constante mágica é 260 e, ao lado, o caminho percorrido pelo rei: iniciando pelo meio círculo e terminando no círculo. Em (b), o grafo bipartido e mágico $K_{8,8}$ correspondente.

Além destas aplicações, existem outras mais peculiares. Os quadrados mágicos de ordem 3 até 9 representam os planetas astrológicos [Drury 1992], uma ligação com misticismo e astrologia. Também são relacionados à música: alguns músicos associaram os números do quadrado mágico a notas musicais obtendo uma base para a composição de suas peças [Roberts 2015].

3. Grafos Graciosos

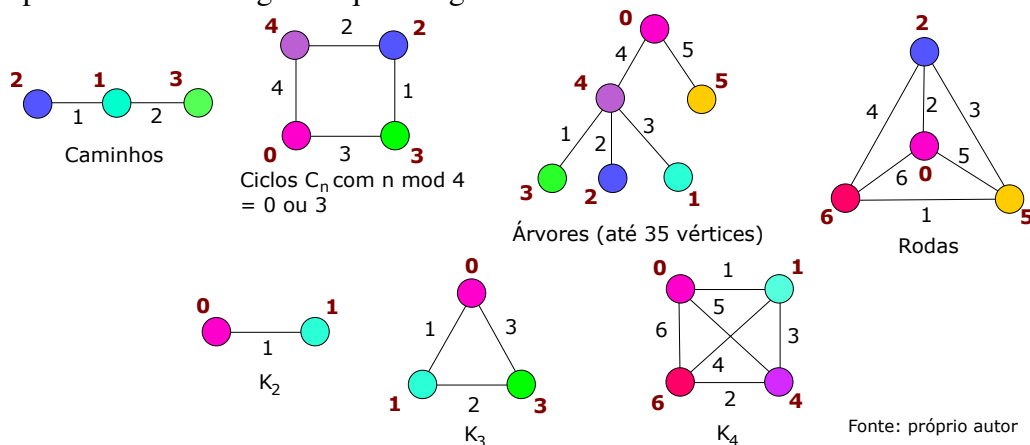
A classe de grafos rotulados reúne diversas subclasses. O que as diferencia são as condições para atribuição dos rótulos e a propriedade adquirida após a rotulação. Dentre as rotulações existentes, além da rotulação mágica vista na seção anterior, existe a **rotulação graciosa** [Gallian 2015]. O artigo de Rosa sobre grafos graciosos [Rosa 1967] é considerado um marco na introdução do conceito de grafos rotulados.

Em uma rotulação graciosa, dado um grafo $G = (V, E)$, deseja-se atribuir um rótulo $f(v)$ para cada vértice v . Além disso, cada aresta uv deve ter como rótulo $|f(u) - f(v)|$. Todas as arestas devem possuir um rótulo diferente e, ao final, estes rótulos devem ser uma sequência de números naturais. Um grafo que pode receber uma rotulação graciosa é dito **gracioso**. Estes grafos despertam muito interesse na literatura, em parte devido à Conjectura 1, a seguir.

Conjectura 1. *Toda árvore é um grafo gracioso.*

Esta conjectura foi proposta por Anton Kotzig e Gerhard Ringel e citada por Rosa em seu artigo original [Rosa 1967]. Embora seja simples, ela permanece em aberto: sua validade só foi provada, através de computador, para árvores de até 35 vértices [Fang 2010].

Além disso, outros estudos mostraram que algumas classes de grafos sempre podem receber uma rotulação graciosa, por exemplo: grafos lagarta, os grafos completos K_2 , K_3 e K_4 , os bipartidos completos e os caminhos [Golomb 1974]. A Figura 3 contém exemplos de classes de grafos que são graciosos.



Fonte: próprio autor

Figura 3. Algumas classes de grafos graciosos. Os números em negrito são rótulos dos vértices. Os demais são rótulos das arestas.

3.1. Teoremas relacionados a grafos graciosos

A seguir, serão reescritas as provas de alguns teoremas que mostram a validade da rotulação graciosa para algumas classes de grafos. São elas: grafos caminhos, estrela e bipartido completo.

Teorema 1. *Todo grafo caminho é um grafo gracioso [Watson 1972].*

Demonstração. Dado um grafo caminho, o número de arestas $m = n - 1$, onde n é o número de vértices. O grafo pode sempre receber uma rotulação graciosa como descrita a seguir.

Inicia-se pelo vértice em uma das extremidades. Ele é rotulado com o inteiro 0. O vértice adjacente a ele recebe $n - 1$ como rótulo. O próximo adjacente não rotulado recebe 1, o próximo $n - 2$ e segue desta forma, alternando entre vértices que recebem rótulos incrementados de 1 e os que recebem rótulos decrementados de 1.

Considerando dois casos: n par e n ímpar. Em ambos, $k = \lfloor \frac{n}{2} \rfloor$. Para o caso par, os rótulos das arestas, começando pela aresta mais à esquerda, são: $|(n - 1) - 0|, |(n - 1) - 1|, |(n - 2) - 1|, \dots, |(n - k) - (k - 1)| = n - 1, n - 2, n - 3, \dots, 1$. Similarmemente, no caso ímpar, os valores das arestas, começando à esquerda, são $n - 1, n - 2, n - 3, \dots, 1$. Todos os valores entre 1 e m são usados, como exemplificado na Figura 4. Logo, esta rotulação é graciosa. Portanto, grafos caminhos são graciosos. \square

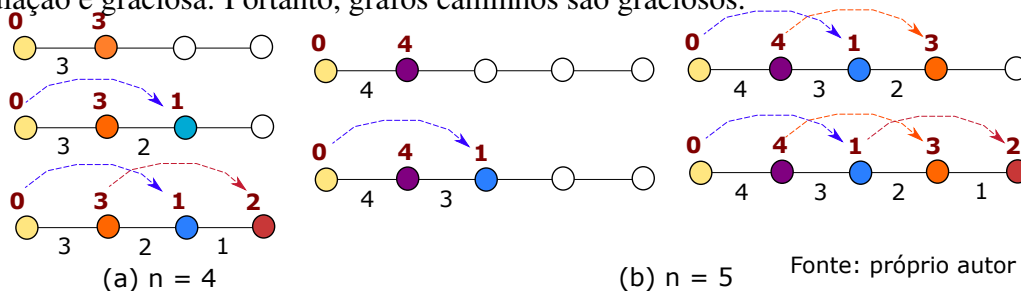


Figura 4. Exemplos de rotulação graciosa para grafos caminhos, como definido na prova do Teorema 1. Em (a) n é par e em (b), n é ímpar. A rotulação obtida é graciosa para ambos.

Teorema 2. *Grafos estrela são graciosos.*

Demonstração. É trivial verificar que um grafo estrela é sempre gracioso. Basta atribuir o rótulo 0 para o vértice central e os números de 1 a m para os demais vértices, sem repetição. Assim, as arestas recebem o mesmo rótulo dos vértices das pontas da estrela, respeitando a rotulação graciosa. \square

Teorema 3. *Grafos bipartidos completos são graciosos [Beavers 2001].*

Demonstração. Dado um grafo bipartido completo $K_{a,b}$. Considerando os conjuntos A e B , contendo a e b vértices, respectivamente. Os vértices do conjunto A são rotulados com $0, 1, 2, \dots, a - 1$. Já os do conjunto B recebem os números $a, 2a, \dots, ba$ como rótulos. Deste modo, todo inteiro de 1 até ab possui uma única representação: a diferença absoluta entre o rótulo de um vértice em A e um vértice em B , como exemplificado na Figura 5. \square

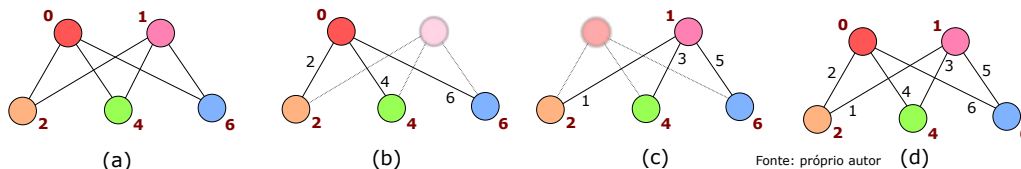


Figura 5. Exemplo da rotulação descrita no Teorema 3. Em (a), os vértices recebem a rotulação. Em (b) e (c), observam-se os rótulos das arestas incidentes a cada vértice superior. Em (d), vê-se que cada inteiro de 1 a m possui representação única.

3.2. Algoritmos para rotulação graciosa de classes específicas de grafos

Os teoremas anteriores possuem algo em comum: todos eles apresentam um modo de rotular graciosamente os tipos específicos de grafos. Deste modo, a ideia principal de cada teorema foi utilizada pela aluna para elaborar algoritmos de rotulação, apresentados nos Algoritmos 1, 2 e 3, a seguir. Cada um deles utiliza os teoremas descritos para rotular grafos caminhos, estrela e bipartidos.

Algoritmo 1: Rotulação graciosa do grafo caminho ($P = (V, E)$)

```

início
  Dados: n, cont1, cont2
  n  $\leftarrow$   $|V(P)|$ ; cont1  $\leftarrow$  0; cont2  $\leftarrow$  n - 1;
  para cada  $v_i \in V(P)$  faça
    se  $i \bmod 2 = 0$  então
      | rotulo( $v_i$ )  $\leftarrow$  cont1; cont1  $\leftarrow$  cont1 + 1;
    fim se
    senão
      | rotulo( $v_i$ )  $\leftarrow$  cont2; cont2  $\leftarrow$  cont2 - 1;
    fim se
  fim para cada
fim

```

Algoritmo 2: Rotulação graciosa do grafo estrela ($S = (V, E)$)

```

início
  para cada  $v_i \in V(S)$  faça
    | rotulo( $v_i$ )  $\leftarrow$   $i$ ;
  fim para cada
fim

```

Algoritmo 3: Rotulação graciosa do grafo bipartido completo ($K = (A + B, E)$)

```

início
  Dados: cont1, cont2
  cont1  $\leftarrow$  0; cont2  $\leftarrow$  1;
  para cada  $v_i \in A(K)$  faça
    | rotulo( $v_i$ )  $\leftarrow$  cont1; cont1++;
  fim para cada
  para cada  $v_j \in B(K)$  faça
    | rotulo( $v_j$ )  $\leftarrow$  cont2 *  $|A(K)|$ ; cont2++;
  fim para cada
fim

```

3.3. Aplicações de grafos graciosos

Os grafos graciosos possuem algumas aplicações práticas interessantes. Uma delas é **modelagem de redes sem fio de telecomunicações** [Nishad 2012]. Considerando que uma rede possui $n + 1$ centrais, estas podem ser consideradas vértices. Entre elas, há conexões, que podem ser vistas como arestas. Se cada central for rotulada com um inteiro entre 0 e n , as conexões são rotuladas com a diferença entre os rótulos das centrais conectadas. Deste modo, cada conexão adquire um rótulo distinto, exatamente como em um grafo gracioso. Isto permite, por exemplo, a execução de um algoritmo simples quando uma conexão é perdida. Como cada aresta é rotulada com a diferença entre os rótulos de duas centrais, é fácil detectar quais são as afetadas e corrigir o problema.

Outra aplicação envolvendo grafos graciosos são as chamadas **régua de Golomb** [Beavers 2001]. Para tanto, são usados grafos graciosos e variações, como os chamados semi graciosos e quase graciosos, em que menos restrições são aplicadas.

Em um grafo rotulado graciosamente (ou de acordo com uma variação), os rótulos dos vértices são assinalados em uma régua, como uma marca. Os rótulos das arestas são as distâncias entre as marcas correspondentes. O objetivo é não haver uma mesma distância entre duas marcas quaisquer, como mostra o exemplo da Figura 6. Estas régua são úteis para criação de códigos auto-ortogonais (usados em *lasers* leitores de discos), reduzir ambiguidade em cristalografia de raios-X e outros usos em teoria da codificação.

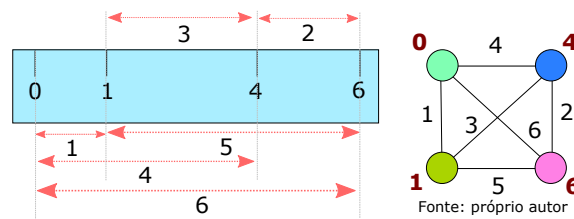


Figura 6. Exemplo de uma régua de Golomb e o grafo gracioso correspondente a ela.

4. Problemas de coloração com restrições de distâncias

O problema de atribuir rótulos aos vértices, como nos grafos graciosos, é análogo à atribuição de cores, como no problema da **coloração** de vértices. Para identificar tal relação, basta considerar cores como inteiros não negativos. Em uma coloração clássica, o objetivo é atribuir cores aos vértices de forma a vértices adjacentes possuem cores diferentes [Feofiloff et al. 2011]. Os problemas estudados neste trabalho, por outro lado, abordam casos especiais relacionados aos valores das distâncias e tipo de (des)igualdade envolvida.

Tal classe de problemas tem sido denominada de **coloração com distâncias**, como resultado do trabalho do doutorando Bruno Raphael Dias, aluno da orientadora desta iniciação científica, Rosiane de Freitas, em conjunto com a mesma. À presente aluna, coube a tarefa de compreender os aspectos já definidos e auxiliar na caracterização de novos casos. Portanto, nesta seção serão relatados os pontos de contribuição do projeto de iniciação científica neste contexto maior.

A teoria de **geometria de distâncias** é aplicada na caracterização de tal classe de problemas, onde é fornecido um grafo simples $G = (V, E, d)$, sendo $d: E \rightarrow \mathbb{R}^+$, e deve-se encontrar a projeção $x: V \rightarrow \mathbb{R}^3$ tal que $\|x(i) - x(j)\| = d_{ij}$ para toda aresta $(i, j) \in E$. Ao considerar o espaço como sendo \mathbb{R}^1 , obtêm-se a coloração com distâncias. Foram estudadas duas variações, onde a diferença reside na restrição imposta: **distâncias iguais**, onde $|x(i) - x(j)| = d_{ij}$ e **distâncias mínimas**, em que $|x(i) - x(j)| \geq d_{ij}$.

4.1. Teoremas de factibilidade e prova de NP-completude

Em seu projeto de pesquisa, o aluno Bruno Dias determinou duas propriedades de factibilidade para o problema de coloração com distâncias iguais. Na primeira, levou-se em conta que todas as distâncias são iguais a uma constante. Na segunda, essa restrição não existe: as distâncias podem assumir qualquer valor inteiro não negativo. Estas

propriedades são descritas nos Teoremas 4 e 5 a seguir.

Teorema 4. *Dado um grafo G , há solução para a coloração com distâncias iguais a uma constante se e somente se G é bipartido.*

Teorema 5. *Se um grafo G é uma árvore, então G sempre admite solução para o caso de coloração com distâncias iguais.*

Além disso, há uma correlação interessante: se um grafo G é gracioso, então sempre admite pelo menos uma solução para a coloração com distâncias iguais. Basta fazer os rótulos das arestas tornarem-se restrições de igualdade e os rótulos dos vértices tornam-se uma solução viável.

O estudo das propriedades identificadas permitiu que a autora auxiliasse na prova de NP-completude para a coloração com distâncias iguais. Para a prova, considera-se a versão de decisão do problema: dado um grafo, o objetivo é decidir se ele possui solução para o caso de distâncias iguais.

O problema pertence a NP, com verificação simples. Para isto, basta elaborar um algoritmo que recebe como entrada um grafo $G = (V, E, d)$ onde cada vértice $v \in V$ recebeu uma cor $x(v)$. Verifica-se, então, se G recebeu uma solução válida tomando cada aresta $(i, j) \in E$ e checando se $|x(i) - x(j)| = d_{ij}$. Se nenhuma restrição de distância foi violada, então G recebeu uma solução positiva para o problema. Como esta verificação ocorre em tempo $O(|E|)$, ela é linear. Portanto, a coloração com distâncias iguais está em NP.

A prova de NP-dificuldade reduz o problema da partição (NP-completo, vide [Garey and Johnson 1979]) para a coloração com distâncias iguais. Considerando uma instância para partição, consistindo em um conjunto M de r inteiros, ou seja, $M = \{m_1, m_2, \dots, m_r\}$. A partir desta instância, é construído um grafo ciclo G com $|V| = |E| = r$, levando em conta o seguinte: $V = \{i_0, i_1, \dots, i_{r-1}\}$, $E = \{(i_b, i_{b+1 \bmod r}) \mid 0 \leq b \leq r\}$ e $d_{i_b, i_{b+1 \bmod r}} = m_{b+1} (\forall 0 \leq b \leq r)$. Assim, seja uma rotulação de vértices x . Se ela for válida, podem ser definidos dois conjuntos: $S_1 = \{m_{b+1} \mid x(i_b) < x(i_{b+1 \bmod r})\}$ e $S_2 = \{m_{b+1} \mid x(i_b) > x(i_{b+1 \bmod r})\}$.

Deste modo, têm-se que S_1 e S_2 formam uma partição de M , onde a soma dos elementos em S_1 é igual à soma dos elementos em S_2 . Isto é, se o grafo G admite uma rotulação válida, então M têm uma resposta positiva e vice-versa. Como esta redução pode ser feita em tempo $O(r)$, a coloração com distâncias iguais é NP-difícil. Sendo assim, conclui-se que decidir se um grafo possui solução para a **coloração com distâncias iguais é NP-completo**.

4.2. Algoritmos gulosos para problemas de minimização

Além do estudo teórico, foram implementados algoritmos heurísticos para a coloração com distâncias mínimas, ou seja, restrição maior ou igual. A escolha da ordem em que os vértices serão coloridos é importante para a heurística utilizada, chamada **gulosa**. Nesta estratégia, dado um grafo G e uma sequência de vértices $V = \{v_1, v_2, \dots, v_k\}$, cada vértice recebe a menor cor possível [Kosowski and Manuszewski 2004]. Este método não garante que o mínimo de cores será utilizado, sendo muito afetado pela escolha dos vértices a serem coloridos. Para a implementação, foram escolhidas três **estratégias gulosas**:

ordem arbitrária; em ordem decrescente de acordo com o grau; e ordenação decrescente segundo o grau de saturação.

Será apresentado apenas o pseudocódigo do algoritmo que utiliza ordem do **grau de saturação** ou DSATUR (Degree of Saturation). O grau de saturação de um vértice é a quantidade de vértices adjacentes a ele que já receberam uma cor. O Algoritmo 4 demonstra como o algoritmo guloso DSATUR funciona para coloração com distâncias iguais. Os algoritmos foram executados utilizando como *benchmark* as instâncias GEOM, da DIMACS, criadas por Michael Trick e disponíveis na página do mesmo [Trick 2004].

Algoritmo 4: Coloração gulosa DSATUR ($G = (V,E)$)

início

 Ordenar os vértices $V(G)$ em ordem decrescente de grau;

 Vértice de maior grau recebe a cor 1;

enquanto \exists *vértice descolorido* **faça**

$v \leftarrow$ maior grau saturacao (G);

se \exists *mais de um vértice com grau de saturação igual a v* **então**

 | Escolhe o vértice de maior grau;

fim se

 Colore o vértice com a menor cor disponível.

fim enquanto

fim

5. Contribuições do trabalho

Os pontos a se destacar neste projeto de pesquisa são:

- realização de uma ampla revisão da literatura sobre o problema do quadrado mágico e classes específicas de grafos rotulados, com ênfase nos grafos mágicos, grafos graciosos e colorações com distâncias;
- aplicação dos conceitos de teoria dos grafos para a compreensão e reescrita das provas de alguns teoremas importantes;
- elaboração e implementação de algoritmos polinomiais para reconhecimento de classes de grafos graciosos e para determinação de soluções para coloração com distâncias, usando heurísticas gulosa;
- uso de *benchmark* consolidado na literatura, DIMACS, com instâncias de coloração com restrições de distância;
- colaboração com a tese de Doutorado do aluno Bruno Dias, no intuito de compreender resultados identificados anteriormente, o que possibilitou, principalmente, alguns resultados adicionais como a elaboração de prova de NP-completude para um problema estudado.

6. Considerações finais

Este projeto de pesquisa, apoiado pelo Programa Institucional de Bolsas de Iniciação Científica (PIBIC – UFAM e CNPq), no período de 2015/2016 abordou aspectos de grafos rotulados a partir do estudo do quadrado mágicos, grafos graciosos e colorações especiais.

Foram reescritas provas de teoremas, elaborada uma prova de NP-completude, bem como adaptados e elaborados algoritmos exatos e heurísticos, com uso de *benchmark* da literatura. A experiência do projeto de pesquisa foi muito interessante para a

aluna, permitindo um ótimo contato com aspectos teóricos da computação. A elaboração de revisão da literatura, aperfeiçoamento do método científico e estudo de diversas propriedades de grafos foi, com certeza, essencial para a formação e interesse por pesquisa. Como projetos futuros, pretende-se continuar investigando os problemas de coloração com distâncias, analisando outras restrições e os relacionando aos grafos graciosos.

Referências

- Arnold, F. (2012). Totally magic graphs - a complete search on small graphs. Master's thesis, Clausthal University of Technology.
- Arsie, K. C. (2010). Jogos sudoku e quadrado mágico. Master's thesis, Universidade Federal do Paraná.
- Beavers, B. (2001). Golomb rulers and graceful labeling. *Loisiana State University*.
- Dias, B. R. C., de Freitas Rodrigues, R., and Filho, N. M. (2012). Alocação de canais em redes celulares sem fio: algoritmos e modelos teóricos em grafos e escalonamento. In *SBPO - Simpósio Brasileiro de Pesquisas Operacionais*.
- dos Santos, C. P., Neto, J. P., and Silva, J. N. (2007). *Os quadrados latinos + Puzzle Hexágono Mágico*. Edimpresa.
- Drury, N. (1992). *Dictionary of Mysticism and the Esoteric Traditions*. Bridport, Dorset: Prism Press.
- Fang, W. (2010). A computational approach to the graceful tree conjecture.
- Feofiloff, P., Kohayakawa, Y., and Wakabayashi, Y. (2011). *Uma Introdução Sucinta à Teoria dos Grafos*. Universidade de São Paulo.
- Gallian, J. A. (2015). A dynamic survey of graph labeling. *The Electronic Journal of Combinatorics*.
- Garey, M. and Johnson, D. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.
- Gherzi, I. I. (1921). *Matematica Dilettevole e Curiosa*. Ulrico Hoepli, Milan, Italy.
- Golomb, S. W. (1974). The largest graceful subgraph of the complete graph. *The American Mathematical Monthly*, 81(5):499–501.
- Kosowski, A. and Manuszewski, K. (2004). Classical coloring of graphs. *Contemporary Mathematics*.
- Nishad, T. M. (2012). Application of strong graphs in wireless networks. *International Journal of Scientific and Engineering Research*.
- Roberts, G. E. (2015). Composing with numbers: Sir peter maxwell davies and magic squares. *Math, Music and Identity*.
- Rosa, A. (1967). On certain valuations of the vertices of a graph. *Theory of Graphs (Internat. Symposium, Rome, July 1966)*.
- Trick, M. (2004). Graph coloring and its generalizations. <http://mat.gsia.cmu.edu/COLOR03/>. Acesso em: 04/2016.
- Watson, R. L. (1972). A survey on the graceful labeling of graphs. Master's thesis, Roanoke College.

Proposta de *framework* para condução da pesquisa científica na Ciência da Computação

Nicole C. Davila, Adriana N. dos Reis

Instituto de Ciências Exatas e da Terra – Universidade Feevale
ERS-239, 2755 – 93525-075 – Novo Hamburgo – RS – Brasil

{nicoledavilaa, adriana.anreis}@gmail.com

Abstract. *Scientific research is fundamental in the student's academic life of Science Computer. However, many students feel unmotivated to the research. The methodological approach and the theoretical foundation are necessary for scientific rigor, but they are far from the practice experienced by some of these students. The purpose of this research, in progress, consists of theoretical research about scientific methodology combined with the best practices of Scrum and Design Science Research, in order to build a support framework for the conduct of Science Computer research that makes transparent and pleasant process to student.*

Resumo. *A pesquisa científica tem papel fundamental na vida acadêmica do estudante de Ciência da Computação. Contudo, muitos alunos sentem-se desmotivados com a pesquisa, que surge somente com o Trabalho de Conclusão de Curso. O aporte metodológico e o embasamento teórico necessários para o rigor científico estão distantes da realidade prática vivida por alguns destes alunos. Este artigo apresenta o trabalho, em andamento, de levantamento teórico sobre a metodologia científica, as boas práticas do Scrum e a metodologia Design Science Research, aliando as mesmas em uma proposta, parcialmente construída, de um framework de suporte para a condução da pesquisa científica na Ciência da Computação.*

1. Introdução

A pesquisa científica pode ser considerada uma forma de construção de conhecimento, conduzindo o indivíduo na identificação de um problema e nas etapas que levam a um resultado para atender o mesmo (PRODANOV; FREITAS, 2013). Para Reis (2010), a pesquisa é a atividade propulsora que leva a intervir, crítica e criativamente, na realidade. Logo, a pesquisa tem grande valor para a sociedade, contribuindo para o desenvolvimento e o avanço nas mais diversas áreas.

Considerada uma prática pedagógica para conduzir o processo de ensino-aprendizagem (REIS, 2010), a pesquisa científica faz parte da realidade acadêmica. Na graduação, no formato de Trabalho de Conclusão de Curso (TCC), a pesquisa acaba causando em alguns estudantes a sensação de incômodo, conforme caracteriza Gonsalves (2007). Tal estudante presume que será uma tarefa difícil de ser executada, não conseguindo visualizar o que precisa ser feito. Appolinário (2012) tem uma visão semelhante, considerando que o estudante não consegue desfrutar da produção científica

e, com isso, considera o ato de pesquisar uma atividade “intolerável”. O curso de Ciência da Computação não está isento dessa realidade.

Appolinário (2012) questiona-se o que seria um cientista ao abordar uma visão geral da ciência. Para o autor o cientista é alguém que busca, através do método científico, novos conhecimentos, não relacionando este processo a determinada área. Partindo de tal afirmação, seria válido concluir, então, que o curso de Ciência da Computação trabalhará fortemente o método e a pesquisa científica no seu decorrer, preparando o estudante para o trabalho com a ciência. Essa realidade, entretanto, é questionável.

A proximidade do estudante com a pesquisa científica, apesar de muitas vezes ser conturbada, é essencial para a formação dos futuros cientistas da computação. Logo, é válido buscar meios que tornem os caminhos da pesquisa mais aprazíveis. Gonsalves (2007) coloca que pesquisar não é uma tarefa fácil, envolve muito trabalho. Nesse contexto esta investigação levanta o seguinte questionamento: é possível construir um *framework* para a condução da pesquisa científica na Ciência da Computação que contribua para a organização, transparência e produtividade do trabalho, aproximando o estudante da pesquisa?

Iniciado em março de 2016, com o envolvimento exclusivo dos autores deste artigo, este trabalho adotou o método *Design Science Research* para a condução da pesquisa, uma vez que o objetivo da mesma é construir um artefato, um *framework*. Identificado o problema e a relevância a ele associada, foi iniciada a revisão da literatura. Três tópicos para guiar a revisão foram definidos: a pesquisa científica, pela forte relação com o tema da pesquisa; a *Design Science Research*, por se tratar de um método voltado para a aplicação prática do conhecimento; e o Scrum, por oferecer um *framework* para a condução de projetos visando a transparência, inspeção e adaptação. O resultado é a contribuição para a pesquisa de autores como Appolinário (2012), Prodanov e Freitas (2013), Gonsalves (2007), Matias-Pereira (2012), Dresch et al. (2015), Bax (2014), Sutherland (2014), entre outros.

Na próxima seção o problema da investigação é contextualizado, abordando a realidade dos cursos de Ciência da Computação no Brasil. A seção 3, por sua vez, apresenta a revisão bibliográfica, conforme tópicos definidos para guiar a pesquisa: pesquisa científica, *Design Science Research* e Scrum. Na seção seguinte, os resultados parciais obtidos nesta primeira etapa da pesquisa são discutidos, relacionando o aporte teórico em uma proposta inicial de *framework* para a condução da pesquisa científica na Ciência da Computação. Por fim, na seção 5, são expostas as considerações finais.

2. A realidade dos cursos de Ciência da Computação

Futuros cientistas da computação têm uma forte relação com a tecnologia. Em 2012, o Ministério da Educação (MEC) publicou as Diretrizes Curriculares Nacionais (MEC, 2012) para graduações em Computação, entre as quais é indicado o perfil, competências e habilidades esperadas dos egressos. É a partir de tais diretrizes que as instituições elaboram as matrizes curriculares de seus cursos.

Conforme as diretrizes citadas, o egresso de um curso de bacharelado em Ciência da Computação deve ter competências e habilidades para, entre outros itens,

identificar e gerenciar riscos; especificar, desenvolver e manter sistemas de computação; reconhecer a importância do pensamento computacional; e gerenciar projetos de desenvolvimento de sistemas computacionais, por exemplo.

De forma explícita nenhuma das competências e habilidades está relacionada à pesquisa científica ou uso do método científico. Trata-se muito de gerência, análise, aplicação de princípios entre as diretrizes. É possível considerar, então, que o perfil esperado pelo MEC (2012) do egresso do bacharelado em Ciência da Computação terá habilidades e competências práticas.

A Tecnologia, segundo Wazlawick (2010), pode ser caracterizada como a aplicação de conhecimentos em atividades práticas. Logo, considerando as diretrizes do MEC (2012), podemos supor que os cursos da Computação oferecerão disciplinas voltadas para a aplicação prática do conhecimento, com forte relação com a Tecnologia. Contudo, uma vez que a pesquisa requer um trabalho teórico de revisão bibliográfica e adesão de métodos que justifiquem as escolhas e resultados, sua utilização como meio para a construção do conhecimento ficaria à escolha de cada docente de tais cursos.

O Trabalho de Curso, ou Trabalho de Conclusão de Curso, também é abordado pelas diretrizes publicadas pelo MEC (2012). Segundo o relatório há predominantemente duas funções nas quais os egressos de cursos de Ciência da Computação atuarão: na realização de atividades de processos ou na transformação de tais processos, desenvolvendo novas tecnologias para tal. Visto isso, o Ministério da Educação recomenda que o curso orientado à formação para atuar na transformação de processos ofereça, então, o espaço para desenvolver e defender o Trabalho de Curso. O texto publicado pelo MEC (2012) indica que, neste trabalho, o estudante aplique no desenvolvimento de aplicações os conhecimentos adquiridos em seu estado da arte.

Ao se deparar com o trabalho de conclusão e a necessidade de escolher um tema no qual trabalhar, o aluno da Ciência da Computação tende a escolher um dos assuntos práticos vistos ao longo da graduação. Seja para resolver um problema da empresa na qual trabalha ou aplicar algum conhecimento prático, a pesquisa científica terá caráter prático. Ou seja, o resultado do trabalho será um produto, uma ferramenta, uma solução. Bax (2014) coloca que, por não contribuir para o avanço do conhecimento científico, neste tipo de pesquisa, o rigor científico não é reconhecido.

A seguir será apresentada a revisão bibliográfica, que contribui para a base teórica da pesquisa. Tal embasamento será utilizado para a construção da proposta de *framework* para a condução da pesquisa que aproxime a realidade de aplicação prática do conhecimento do rigor da investigação científica.

3. Revisão bibliográfica

Neste tópico será apresentada a revisão bibliográfica realizada sobre pesquisa científica, *Design Science Research* e Scrum, assuntos utilizados como base teórica da investigação.

3.1. Pesquisa científica

A pesquisa faz parte do cotidiano das pessoas, sejam elas estudantes, profissionais ou qualquer indivíduo que deseja saber algo. Considerada uma forma de construir

conhecimento, a pesquisa é um conjunto de ações para encontrar respostas para determinadas questões (PRODANOV; FREITAS, 2013). Ou seja, no sentido mais amplo, pesquisar é buscar informações que auxiliem a chegar a determinado objetivo.

No âmbito científico, contudo, a pesquisa deve apresentar rigor (DRESCH et al., 2015), utilizando-se de métodos e processos (DIEHL, 2004) que garantam a confiabilidade dos resultados apresentados. O método, segundo Appolinário (2012), está presente em todos os âmbitos da experiência humana como processo organizado, lógico e sistemático. Logo, método pode ser entendido como um conjunto de procedimentos adotados para alcançar um objetivo.

O método científico, por sua vez, pode ser entendido como o caminho adotado pelo pesquisador para a produção de conhecimento (MATIAS-PEREIRA, 2012). Segundo Prodanov e Freitas (2013), no passado, muitos pensadores tentaram definir um método científico que pudesse ser utilizado por todas as ciências, gerando linhas de pensamento distintas. Contudo, Matias-Pereira (2012) afirma que não existe um único método científico, sendo o mesmo relacionado com a natureza da investigação.

Uma vez que o método está relacionado com a natureza da pesquisa, então primeiro é necessário que o pesquisador defina o objetivo da mesma (APPOLINÁRIO, 2012). A metodologia da pesquisa, segundo Matias-Pereira (2012), seria o emprego de tal método, de procedimentos e técnicas que variam de acordo com a ciência. Segundo Correia (2009), a metodologia depende ainda do tipo de pesquisa que será adotada para alcançar o objetivo, podendo ser bibliográfica, documental, estudo de caso, empírico-analítica ou pesquisa de campo, por exemplo.

3.1.1. Papéis da pesquisa

No âmbito da pesquisa científica têm-se dois papéis fundamentais: o pesquisador e o orientador. O professor orientador, segundo Almeida (2011), tem entre suas atribuições ajudar no esclarecimento de dúvidas, indicar leituras e sugerir correções no trabalho que está sendo desenvolvido. O pesquisador, ou orientando, é aquele que realiza o trabalho, realizando leituras, levantamentos, análises e escrita – Almeida (2011) ressalta, ainda, que o autor do trabalho é o orientando. O orientador, segundo o autor, tem “o papel de mostrar o caminho para que o aluno chegue à produção de um bom trabalho científico”.

Os encontros entre orientador e orientando têm como objetivo a troca de informações, a inspeção do que foi realizado e o planejamento dos próximos passos. Em geral, a periodicidade deste encontro é uma combinação entre os envolvidos, assim como a modalidade (presencial ou à distância).

3.2. Pesquisa na Ciência da Computação

Segundo Fonseca Filho (2007), não é possível listar claramente a história da Computação em uma sequência de invenções, nomes e datas para definir a evolução dos computadores. Contudo, o autor ressalta que no século XX, do ponto de vista científico, ocorreu uma aceleração tecnológica sem precedentes, uma revolução apoiada principalmente na evolução dos computadores.

A Computação, portanto, pode ser considerada uma área relativamente nova e em expansão. Contudo, para Wazlawick (2008), único autor brasileiro de livro

específico sobre o tema, isso não justifica que o método científico em tal área não seja bem definido. Ou seja, para o autor há a necessidade de desenvolver trabalhos na área da computação com melhor embasamento metodológico, adequado à realidade da área.

Wazlawick (2008) apresenta alguns estilos de pesquisa em Ciência da Computação, considerando o grau de amadurecimento na subárea. Segundo o autor, a classificação apresentada baseia-se em uma discussão ocorrida em lista da SBC (Sociedade Brasileira de Computação) há cerca de 10 anos. Os estilos de pesquisa seriam: apresentação de um produto, apresentação de algo diferente, apresentação de algo presumivelmente melhor, apresentação de algo reconhecidamente melhor e apresentação de uma prova.

Os estilos apresentados por Wazlawick (2008) não tratam apenas de cursos da graduação, mas também de mestrados e doutorados na área da Computação. Porém, para o autor, é aceitável que um trabalho de graduação trabalhe com objetivos técnicos no trabalho da pesquisa científica. Ou seja, é aceito que os futuros cientistas da computação demonstrem, no Trabalho de Conclusão, os conceitos aprendidos e como aplicá-los na prática no desenvolvimento de um sistema. Essa característica do curso pode ser observada nos estilos de pesquisa citados por Wazlawick (2008), sendo atendida pelo método *Design Science Research*, o qual será abordado no próximo tópico.

3.3. *Design Science Research*

Design Science, ou ciência do projeto, surgiu como alternativa para pesquisas científicas que não são atendidas de forma satisfatória pelas ciências tradicionais, que buscam “explorar, descrever, explicar e, quando possível prever” (DRESCH et al., 2015). Segundo Aken (2004), o objetivo da *Design Science* é trabalhar com o desenvolvimento de novos conhecimentos para a construção de artefatos, solução de problemas ou melhorias. Dresch et al. (2015), contudo, coloca que as ciências tradicionais e a *Design Science* se complementam, sendo possível utilizar ambas em um projeto.

A *Design Science Research* (DSR) é o método de pesquisa dentro da *Design Science* que indica o caminho para a condução da investigação. Segundo Bax (2014), envolve a construção, validação e avaliação dos artefatos gerados pela pesquisa. Para Dresch et al. (2015), as principais atividades da DSR são definir um problema, sugerir formas de contorná-lo, desenvolver uma solução e avaliá-la, concluindo a pesquisa.

O artefato citado na bibliografia da *Design Science* pode ser entendido como algo artificial, como um *software*, por exemplo. Há alguns tipos de produtos gerados pela DSR, uma classificação feita considerando os artefatos gerados. Segundo March e Smith (1995 apud DRESCH et al., 2015) os tipos são: constructos, ou elementos conceituais; modelos, um conjunto de proposições ou declarações que expressam as relações entre os constructos; métodos, um conjunto de procedimentos; instanciações, execução de um artefato; e, por fim, as *design propositions*, que tratam de um modelo genérico a ser utilizado como solução para uma determinada classe de problemas.

3.3.1. Condução da pesquisa utilizando *Design Science Research*

A condução de uma pesquisa que adote o método *Design Science Research* deve atender o rigor e a relevância de uma investigação científica. Para Dresch et al. (2015), a condução da DSR parte de um ambiente em que existirão necessidades que justifiquem

a relevância do trabalho. O método, por sua vez, guiará a construção de teorias e artefatos para tais necessidades. Esses artefatos podem ser avaliados e refinados através de procedimentos, gerando uma base de conhecimento (DRESCH et al., 2015).

Outra abordagem para a condução da pesquisa utilizando a DSR, semelhante em alguns pontos com a visão apresentada por Dresch et al. (2015), é apresentada por Vaishnavi e Kuechler (2004). Para os autores, a condução da pesquisa também tem início no problema, na conscientização sobre o mesmo. Na sequência serão propostas soluções para o problema considerando o conhecimento já existente na área relacionada e, então, uma delas será selecionada para o desenvolvimento, avaliação e, por fim, a conclusão da investigação.

Os autores Vaishnavi e Kuechler (2004) ainda indicam que as etapas de desenvolvimento e avaliação podem ser realizadas diversas vezes durante a pesquisa. Tais ciclos iterativos geram contribuições para as bases de conhecimento, tanto sobre os procedimentos adotados, como para o refinamento do artefato que será gerado (Dresch et al., 2015). Com o intuito de buscar boas práticas que contribuam na gerência deste trabalho, o próximo tópico abordará o Scrum, modelo ágil de gerência de projetos.

3.4. Scrum

O Scrum é um conjunto de boas práticas criado para ser uma forma mais rápida, eficaz e confiável de gerenciar projetos de *software*, conforme Sutherland (2014). Segundo o autor, o Scrum não é uma metodologia de trabalho, mas um *framework* que surgiu para substituir métodos tradicionais. Apesar de usualmente ser utilizado no desenvolvimento de *software*, não há na literatura restrição ou adaptação prevista quanto ao uso do Scrum em outros tipos de projeto. Ou seja, cabe a cada organização adaptar o Scrum ao seu modo de trabalho.

O Scrum propõe boas práticas baseadas em alguns dos princípios do Manifesto Ágil¹, como, por exemplo, a entrega contínua, adiantada e frequente do produto; receptividade à mudança de requisitos; e reflexão periódica sobre o processo e como melhorá-lo. Segundo Schwaber e Sutherland (2013), considerados criadores do Scrum, esse modelo é baseado em três pilares: transparência, de modo que os envolvidos tenham visão dos aspectos significativos do processo; inspeção, para que participantes detectem variações no processo; e, por fim, a adaptação, que ocorrerá como resultado da inspeção e evitará desvios do projeto.

Para Cohn (2006), o *Time Scrum* tem papel fundamental no sucesso de um projeto. Segundo o autor é necessário o comprometimento de todos com o objetivo, de modo que todos trabalhem por ele. Os *Times Scrum*, segundo Schwaber e Sutherland (2013), devem ser multifuncionais e auto-organizáveis, de modo que não dependam de pessoas externas à equipe.

O *Time Scrum* será formado pelo Dono do Produto (*Product Owner*), que definirá o que será desenvolvido e terá relação com o cliente do produto; pelo Mestre Scrum (*Scrum Master*), que tem a função de orientar o time com relação às práticas do

¹ Publicado pela Agile Alliance em 2001 - <http://www.agilemanifesto.org/iso/ptbr/principles.html>.

Scrum e remover impedimentos; e pelo time de desenvolvimento, responsável por transformar os requisitos em um incremento do produto.

Um elemento importante no Scrum é o *Sprint*, o ciclo de trabalho do time. Os *Sprints* são iterativos e incrementais, buscando aperfeiçoar a previsibilidade e o controle de riscos (SCHWABER; SUTHERLAND, 2013). Ao iniciar um *Sprint* é realizada uma Reunião de Planejamento (*Sprint Planning*), na qual o *Time Scrum* define o que será desenvolvido no próximo ciclo. No final de um *Sprint* o *Time Scrum* volta a se reunir para avaliar o processo (Reunião de Retrospectiva ou *Sprint Retrospective*) e revisar o produto (Reunião de Revisão ou *Sprint Review*). Durante o ciclo o time de desenvolvimento realiza Reuniões Diárias (*Daily Meeting Scrum*), na qual cada integrante informa aos demais o que fez no dia anterior, quais os impedimentos enfrentados e o que pretende fazer.

4. Resultados parciais

Esta pesquisa se propõe construir um *framework* para a condução da pesquisa científica na Ciência da Computação que contribua tanto para a organização, transparência e produtividade do trabalho, como também para tornar o caminho da investigação mais agradável, aproximando o estudante da pesquisa. Identificado o problema da pesquisa, a investigação prosseguiu com consultas às bases de conhecimento, utilizando como meio a revisão da literatura. Para guiar tais consultas foram definidos três tópicos: a pesquisa científica, a DSR e o Scrum.

Enquanto a pesquisa guia o futuro cientista da computação na construção de novos conhecimentos, a DSR oferece o aporte metodológico para a geração de artefatos. O Scrum, por sua vez, contribui com boas práticas visando à produtividade, transparência, inspeção e adaptação, fatores também importantes na pesquisa científica. As consultas realizadas às bases de conhecimentos, neste momento, buscaram identificar elementos e conceitos relacionados com a condução do trabalho.

O resultado da revisão sistemática da literatura foi utilizado, então, como aporte para iniciar a construção da proposta de *framework* para a condução da pesquisa científica. Nos próximos tópicos serão apresentadas as semelhanças entre os assuntos pesquisados, assim como a proposta de condução elaborada.

4.1. Pesquisa científica e o Scrum

Um trabalho científico terá dois papéis: o orientador e o orientando. Enquanto o primeiro atua guiando o pesquisador em sua produção científica, o segundo é quem executa o trabalho de fato. Os papéis do Scrum são semelhantes: o Mestre Scrum será aquele que guia o time na aplicação do modelo e o Time Desenvolvimento será responsável pela produção dos incrementos do artefato.

O Scrum prescreve também o papel de Dono do Produto, aquele que será responsável por manter o *backlog* do produto ordenado e priorizado. Na pesquisa científica da Ciência da Computação tem-se, frequentemente, a aplicação prática de conhecimento na construção de um artefato como proposta de solução. Logo, o pesquisador definirá uma série de requisitos que deverão ser atendidos para que a proposta seja considerada “pronta”. Tais requisitos compõem a fila de necessidades do artefato, que o Scrum define como *backlog* do produto. No caso da pesquisa, entretanto,

tanto orientador como orientando compartilham o papel de Dono do Produto: O orientador sugere a prioridade, porém é o orientando, o autor do trabalho, quem ordenará o *backlog*.

O *framework* Scrum também define determinados eventos, visando minimizar reuniões ou encontros não previstos. Da mesma forma, orientador e orientando definem encontros periódicos para avaliar o que foi realizado e combinar os próximos passos. No objetivo desse encontro é possível identificar características dos eventos Scrum: reunião de planejamento, retrospectiva e reunião de revisão.

4.2. Design Science Research e o Scrum

O método *Design Science Research* propõe ciclos de desenvolvimento e avaliação do artefato. Se considerado o artefato como incremento do produto final gerado pelo projeto, então tais ciclos podem ser considerados os *Sprints* descritos no modelo Scrum.

A avaliação do artefato indicada pela *Design Science Research* é fator essencial para a pesquisa científica, visto que gera contribuições para a base de conhecimento. Tais contribuições partem tanto da avaliação dos processos e métodos como do artefato desenvolvido. De forma semelhante, as reuniões de Retrospectiva e Revisão do Scrum buscam avaliar o processo do time e revisar o incremento gerado no produto.

4.3. Proposta de condução da pesquisa com Scrum e Design science research

Ao adotar a *Design Science Research* como método científico, este trabalho se propõe a elaborar propostas de artefatos como solução. Este artigo apresenta a construção da primeira proposta, explorando uma nova possibilidade de condução da pesquisa científica no cenário da Ciência da Computação, considerando ainda a relevância e rigor necessários para o carácter científico.

Na Figura 1 é apresentada uma proposta de condução da pesquisa científica, a qual foi elaborada a partir de uma representação das etapas da DSR apresentadas por Dresch et al. (2015), que se mostrou mais detalhada frente a outras da literatura. Foram relacionados conceitos do Scrum e da DSR, método que prevê o rigor científico e a geração de contribuições para bases de conhecimento.

Quando a proposta do artefato é elaborada são definidos requisitos, necessidades que devem ser incrementadas no artefato para que o mesmo seja considerado “pronto”. Os requisitos formam o *backlog* do artefato. As etapas de desenvolvimento e avaliação trabalharão em cima de tal *backlog*, o qual será constantemente ordenado e priorizado pelo orientando e orientador.

Para que os incrementos sejam produzidos dentro do cronograma da pesquisa científica podem ser definidos ciclos de trabalho, os *Sprints*. No início de cada ciclo o professor orientador e o pesquisador encontram-se para planejar o *Sprint* seguinte, definindo o que será incrementado no artefato e como o mesmo será avaliado. Tais definições são essenciais, pois, ao final do ciclo, ambos se reúnem novamente para avaliar o processo e os procedimentos adotados, a reunião de retrospectiva do Scrum. Neste momento também é realizada a reunião de revisão do artefato, verificando o que foi incrementado e se está de acordo com o projeto.

Os resultados desses encontros ao longo do desenvolvimento oferecem transparência ao processo, permitindo a inspeção e a adaptação quando necessário. Por outro lado, as reuniões também geram, de forma parcial e contínua, aprendizagens. Serão tais aprendizagens que, ao finalizar o desenvolvimento do artefato, serão utilizadas para gerar contribuições nas bases de conhecimento.

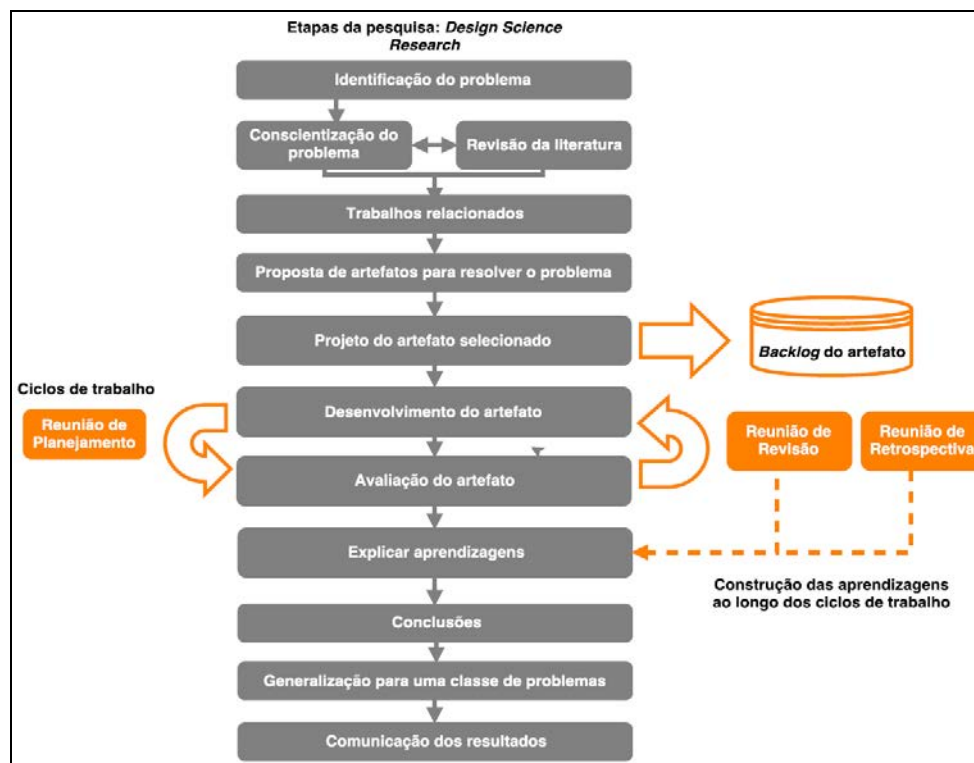


Figura 1. Proposta de condução da pesquisa científica

5. Considerações finais

A proposta de um *framework* para a condução da pesquisa científica na Ciência da Computação é um desafio. Não basta apenas reunir conceitos e elementos descritos na literatura, é necessário que eles possam ser aplicados na prática. O rigor também precisa estar presente, pois é essencial na construção do conhecimento científico.

A construção da proposta apresentada neste artigo é apenas o primeiro passo. Porém, o caminho é promissor para fins da área de Ciência da Computação, uma vez que associa abordagens teóricas e práticas. Tratando desde o fundamento científico até o monitoramento da condução da investigação, espera-se que o *framework* em construção contribua para um processo transparente e aprazível ao futuro cientista da computação.

Referências

- Almeida, M. S. (2011) "Elaboração de projeto, TCC, dissertação e tese: uma abordagem simples, prática e objetiva". Atlas, São Paulo/SP.
- Appolinário, F. (2012) "Metodologia da ciência: filosofia e prática da pesquisa". Cengage Learning, São Paulo/SP.

- Bax, M. P. (2014) “*Design science: filosofia da pesquisa em ciência da informação e tecnologia*”. In: XV ENANCIB (Encontro Nacional de Pesquisa em Ciência da Informação). ‘Além das Nuvens: Expandindo as Fronteiras da Ciência da Informação’. Belo Horizonte/MG.
- Cohn, M. (2006) “*Agile Estimating and Planning*”. Prentice Hall, New Jersey.
- Correia, W. (2009) “TCC não é um bicho-de-sete-cabeças”. Ciência Moderna, Rio de Janeiro/RJ.
- Diehl, A. A. (2004) “Pesquisas em ciências sociais aplicadas: métodos e técnicas”. Prentice Hall, São Paulo/SP.
- Dresch, A.; Lacerda, D. P.; Antunes Jr., J. A. V. (2015) “*Design science research: método de pesquisa para avanço da ciência e tecnologia*”. Bookman, Porto Alegre/RS.
- Fonseca Filho, C. (2007) “História da computação: O Caminho do Pensamento e da Tecnologia”. EDIPUCRS, Porto Alegre/RS.
- Gonsalves, E. P. (2007) “Conversas sobre a iniciação à pesquisa científica”. Alínea, Campinas/SP.
- Matias-Pereira, J. (2012) “Manual de metodologia da pesquisa científica”. Atlas, São Paulo/SP.
- MEC, Ministério da Educação. (2012) “Diretrizes Curriculares Nacionais para os cursos de graduação em Computação”. Disponível em: http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=11205-pces136-11-pdf&category_slug=julho-2012-pdf&Itemid=30192. Acesso em: Abril de 2016.
- Prodanov, C. C.; Freitas, E. C. de. (2013) “Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico”. Feevale, Novo Hamburgo/RS.
- Reis, L. G. (2010) “Produção de monografia da teoria à prática: o método educar pela pesquisa (MEP)”. Senac, Brasília/DF.
- Sutherland, J. (2014) “Scrum: a arte de fazer o dobro do trabalho na metade do tempo”. LeYa, São Paulo/SP.
- Schwaber, K.; Sutherland, J. (2016) “Guia do Scrum™”. Disponível em: <http://www.scrumguides.org/docs/scrumguide/v1/Scrum-Guide-Portuguese-BR.pdf>. Acesso em: Abril de 2016.
- Vaishnavi, V.; Kuechler, B. (2004) “*Design Science Research in Information Systems*”. Disponível em: <http://desrist.org/design-research-in-information-systems/>. Acesso em: Abril de 2016.
- Wazlawick, R. S. (2008) “Metodologia de pesquisa para ciência da computação”. Elsevier, Rio de Janeiro/RJ.
- Wazlawick, R. S. (2010) “Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico”. Revista de Sistemas de Informação da FSMA, Macaé/RJ.

Using Active Mediators and Passive Extractors Inside Materialized Data Integration Systems

Paulo V. M. Cardoso, Flavio F. Franzin,
Sergio L. S. Mergen

¹Curso de Ciência da Computação – Universidade Federal de Santa Maria (UFSM)
Santa Maria – RS – Brazil

{pcardoso, ffranzin, mergen}@inf.ufsm.br

Abstract. *Materialized data integration architectures are generally composed by active extractors and passive mediators, where the extractors have the goal of forwarding to the mediator all identified objects that may be of interest. This one hand communication road may lead to the difficulty of identifying relevant objects (specially if the source is not well structured) and a waste of energy analyzing irrelevant objects and possibly extracting them. Also, there is a processing overhead associated with the deduplication of objects that are already mapped. In this paper we propose an extension of the integration architecture that allows mediators to play a more active role, guiding passive extractors as to what objects need to be extracted and how they can be identified. We present a case study that shows how passive and active extractors can coexist under the same data integration system.*

1. Introduction

Materialized integration systems are built on top of architectures that allow generally three things: extracting data from the sources, mapping the extracted data into a unified (global) schema and finally accessing the materialized data from the unified schema to answer queries[Anter et al. 2016].

The extractors (in their many forms) have the goal of finding objects. The identified objects are send to a component called mediator. If more than one data source publishes overlapping information about an object, the mediator is responsible for merging the data so that each object is represented only once in the unified schema. What is worth noting here is that, during the extraction phase, there is typically a one way communication between the source and the mediator, where the source forwards all (possibly) relevant information about objects it was able to find. In other words, the extractors play active roles while the mediator is entirely passive. There are some caveats to this approach:

- If the source is semi-structured or unstructured, it may be hard to identify objects within the contents of a data source. For instance, suppose the system integrates information about artists and the supported extractor collects artists names from wikipedia articles. There are many named entities inside texts written in natural language. It is cumbersome for a extractor to locate a value inside an article and to conclude it is indeed not only a person's name but the name of an artist.

- The extractor may visit many sources that are not relevant in the context of the integration system. In the best case scenario, there is a waste of energy processing such sources. In the worst case scenario, it may lead to the mapping of irrelevant objects. For instance, the extractor of artists name may (unnecessarily) harvest the whole wikipedia to find what it was looking for.
- The system needs to check if an incoming object already exists in the unified schema, so that the two versions are merged into one. This matching process can be very time-consuming if all mapped objects need to be verified.

There are many proposals aimed at handling these issues: the extraction of named entities is by itself a research topic inside the natural language processing area. Focused crawlers are studied as a way to restrict the search to relevant sources only. Object matching and string matching are handful techniques to allow fast and precise object deduplication. However, considering these ideas are based on heuristics, there is always a chance of failure.

To minimize this sort of problems, in this paper we propose an extension of the general extraction architecture of a materialized data integration system by introducing the concepts of active mediators and passive extractors. Instead of having a mediator that merely receives information from the sources, it is given the freedom to define which object to look for, based on the objects already stored in the repository. Once the extractor receives that information, it proceeded with the extraction of information that complement the existing object. The complemented object is then send back to the mediator.

This paper is organized as follows: Section 2 shows how the extension can be accommodated in a general architecture for data extraction. It also shows how the active and passive extractors can coexists and how one can benefit from the other. Section 3 presents a case study regarding extracting data into a biological collaborative system. Both active and passive extractors are implemented. Section 4 presents the results obtained when executing the case study. Section 5 discusses related work. Finally, Section 6 brings our concluding remarks.

2. Supporting Active Mediators and Passive Extractors

Figure 1 shows a general data extraction architecture modified to support active mediators and passive extractors. We only care about materialized data integration systems, where data from the sources are fed into a global repository. The solid rectangle surrounds the traditional parts whereas the dotted rectangle surrounds the proposed extension.

An extractor is labeled as active if it receives no information from the mediator about the objects to extract. Its purpose is to identify all information it deems to be relevant. The information is than shaped in the form of objects represented in a standardized canonical format. Once the passive mediator receives the collected objects, it needs to store them. Also, the mediator needs to check if the object already exists in the repository. If so, it needs to merge the overlapping objects into one.

On the other hand, an extractor labeled as passive receives information from the mediator about the object to extract. The information (a message) can be anything that helps the extractor locate relevant data sources or parts of a data source that contain relevant information. When the extractor is passive, its purpose is to identify information

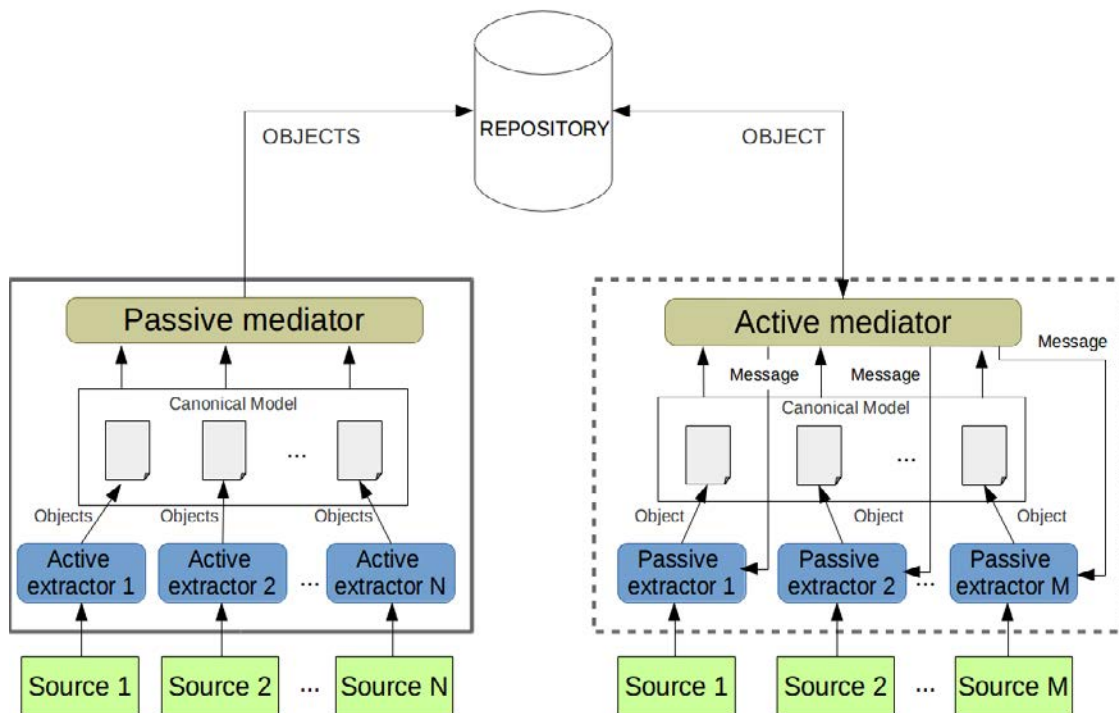


Figure 1. Data Extraction Architecture: distinction between active/passive extractors and active/passive mediators

about a single object, the one the mediator is asking for. The extracted object is described in the canonical format and send back to the mediator. Once the active mediator receives the collected object, it needs to store it in the repository, merging the preexisting object with the collected one.

Algorithm 2.1 shows the interaction between the active mediator and the passive extractor. The mediator prepares a message for each object in the repository. The extractor receives the message and use it as a guide to locate data sources and find the desired object.

```

EXTRACT(repository)
1: for each object  $O$  from the repository do
2:    $message \leftarrow extract\_message(O)$ 
3:    $O_{new} \leftarrow run\_passive\_extractor(message)$ 
4:    $merge(O, O_{new})$ 
5: end for

```

Algorithm 2.1: COMMUNICATION BETWEEN AN ACTIVE MEDIATOR AND A PASSIVE EXTRACTOR

Active and passive extractors can live together inside the same system. Active extractors can be used to create an initial representation of an object. Once the object exists in the global repository, passive extractors can be triggered to complement the object.

Passive extractors lack the ability to find new objects. However, they are more suited if the purpose is to enrich an existing object with new information, since it is more

capable to spot the parts of the source that describe the object, specially if the data source is not well structured. Besides, since a passive extractor knows what to look for, it is able to focus on data sources that are indeed relevant, instead of striving across unfitting sources. Finally, active mediators have no need to check if the object exists in the repository. It already knows the object exists, and keeps the older version ready to merge once the extractor finishes collecting data.

It is important to notice that we are not defining a particular format (or content) of the message, and neither a particular merging strategy. Our intention here is to define the foundations of the architecture and leave the details to real implementations. In the next section we discuss a case study where this architecture is implemented.

3. Case Study: BioID

BioID is a collaborative system developed by the students that authored this paper. The system is targeted at species labeling and classification. Generally speaking, it allows users to index life forms and their properties, such as habitat, location and ranks (e.g. species, genus, family, class, kingdom). The properties are expansive, so users can enrich the description of an object by informing additional properties.

Figure 2 shows a snapshot of the system where a species is wrapped inside a block. The block contains information such as common name, images, properties and comments. The user is allowed to add new information, but cannot edit information registered by other users. An overview of the main features of BioID is presented in [Cardoso et al. 2015].



Pipile jacutinga Created by bioid


+ Mais fotos

Reino	Animalia
Filo	Chordata
Classe	Aves
Ordem	Galliformes
Familia	Cracidae
Genero	Pipile
Especie	jacutinga

Figure 2. BioID's block with a life form information

According to [Mora et al. 2011], there are approximately 8,7 million species on the planet, and only 1,2 million that has a taxonomic classification. Also, there are large digital repositories that publishes information about classified species, such as *Cites*¹ and *CalPhotos*². However, none of them allows collaborative working.

¹(<https://www.cites.org/>)

²(<http://calphotos.berkeley.edu/>)

Given that a lot of information is already available online, we decided to create active and passive extractors to populate the BioID repository. The purpose of the extractors is to locate objects concerning life forms (animals, to be precise) and represent each one of them in a canonical format supported by the BioID system. The format of an object is as described in Definition 1.

Definition 1 (*Object*) An object O in the accepted canonical format is a tuple $\{P, I\}$. Element P is a list of properties $\{p_1, p_2, p_3, \dots, p_n\}$, where $p_i = \{key, value\}$. Element I is a list of images $\{i_1, i_2, i_3, \dots, i_j\}$, where $i_j = \{uri\}$. The key, value and uri elements are literals.

The *key* and *value* pair refers to the name of a property and its corresponding value, respectively. The *uri* refers to the location where a image was found. Also, we use the term 'incoming object' to refer to an extracted object that is not yet processed by the mediator. Conversely, the term 'stored object' is used to refer to an object already stored in the repository.

Two active and three passive extractors were implemented by the students who authored this paper. One of the active extractors collects objects from *Cites*. This data source publishes a CSV file with rank properties of catalogued animals. The second active extractor collects objects from *CalPhotos*, a source that publishes a CSV file with species common names. Since the sources are structured, it was relatively straightforward to create the extractors.

All of the passive extractors process HTML data from the following web pages: *Google Images*, *CalPhotos* and *Wikipedia*. The first two were given the task to complement objects with images. On the other hand, the *Wikipedia* variant needed to complement an object with images and properties related to ranks, according to the biological classification of a life form. For instance, 'species' and 'genus' are possible ranks to extract.

Recall that passive extractors need a message to guide the extraction. Given a stored object O in the repository, and its list of properties $P \in O$, the message is created as the value of a property $p \in P$, such that $p.key = 'species'$. If the object has no property whose key is species, the message is not created and the extraction for the given object is aborted. For instance, possible values for the message are 'H. sapiens' and 'E. coli'. The reasoning behind this heuristic is that the species property serves to uniquely identify a life form. Therefore, its value is useful for searching additional information about an object.

Once the message is generated, the extractors use it as described in the following steps:

1. A *url* is accessed. The *url* is formed by a fixed part and a dynamic part formed by the message. For instance, the Portuguese wikipedia extractor uses `https://pt.wikipedia.org/wiki/` as the fixed part.
2. The part of the web page where relevant data can be found is detected. For *Google Images* and *CalPhotos*, the relevant parts (images) are those that appear enclosed by the *img* tag. For *Wikipedia*, the relevant parts (properties and images) can be found inside a *div* tag named *infobox*.
3. The relevant information is described as an object in the canonical format. When multiple images exist inside the web page, only the first one is extracted (the first

img tag).

4. The retrieved object is merged with the stored one. Next section explains how the merging is performed.

3.1. Identification and Merging of Duplicated Objects

Definition 2 presents a notion of duplicated properties and images, which is important for the identification and merging of duplicated objects.

Definition 2 (*Duplicated Property and Duplicated Image*) Two properties p_i and p_j are duplicated if $p_i.key = p_j.key$. Similarly, two images i_i and i_j are duplicated if $i_i.uri = i_j.uri$.

Duplicated objects are naturally identified within active mediators, as Algorithm 2.1 showed. On the other hand, passive mediators need to check if an incoming object O_i has a duplicated object O_s in the repository, where $P_i \in O_i$ and $P_s \in O_s$. We do this by checking if properties $p_i \in P_i$ and $p_s \in P_s$ are duplicated, having 'species' as key and the exact same value. In other words, objects of the same species are inferred to be duplicated.

The merging of two duplicated objects O_i (incoming object) and O_s (stored object) follows the same reasoning for active and passive mediators, as Table 1 describes:

image $i_i \in O_i$ and image $i_s \in O_s$	property $p_i \in O_i$ and property $p_s \in O_s$
If i_i is not duplicated in O_s <ul style="list-style-type: none"> • i_i is added to O_s. 	If p_i is not duplicated in O_s <ul style="list-style-type: none"> • p_i is added to O_s.
If i_s and i_i are duplicated <ul style="list-style-type: none"> • no action is performed. 	If p_s and p_i are duplicated <ul style="list-style-type: none"> • $p_s.value = \text{concat}(p_s.value, ', ', p_i.value)$

Table 1. Rules for Merging Properties and Images of Duplicated Objects

We observe that this strategy is clearly limited, as we do not take into account the similarity of elements. For instance, the properties with keys *color* and *colour* would be considered unlike. Also, the above rules may end up concatenating equal values (e.g. $color = red, red$). We are aware of the limitations, that we leave as future work. We remark that our intention in this paper is to demonstrate how active and passive mediators can be used in combination, regardless of their particular weaknesses.

4. Experimental Results

This Section presents how effective the proposed extraction methods are for finding relevant objects (in the case of the active extractors) and finding complementary information about objects (in the case of passive extractors).

Figure 3 shows the number of animals found by the active extractors along with the relative proportion with respect to the number of species that have a biological classification. As shown, the extractors found 25.820 objects from the two visited data sources. After the object deduplication and merging, 24.680 distinct animals were indexed. This is approximately 1,43% of the total of classified species. If only animals are considered (the actual target of the extractors), the percentage grows to 2,58%.

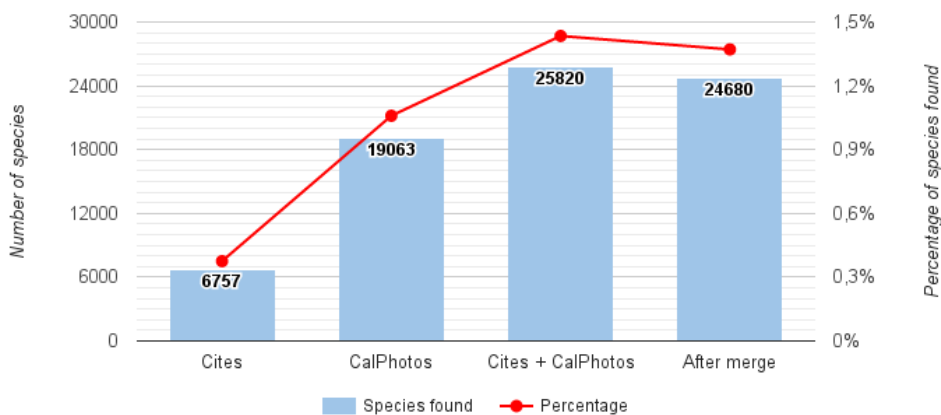


Figure 3. Animals found by the implemented active extractors.

Figure 4 shows how the extractors enrich objects with rank properties. The total amount of properties was computed as the product of the number of indexed objects and the seven rank properties (Kingdom, phylum, class, order, family, genus and species). The percentage value is the ratio between the amount of properties found and the total.

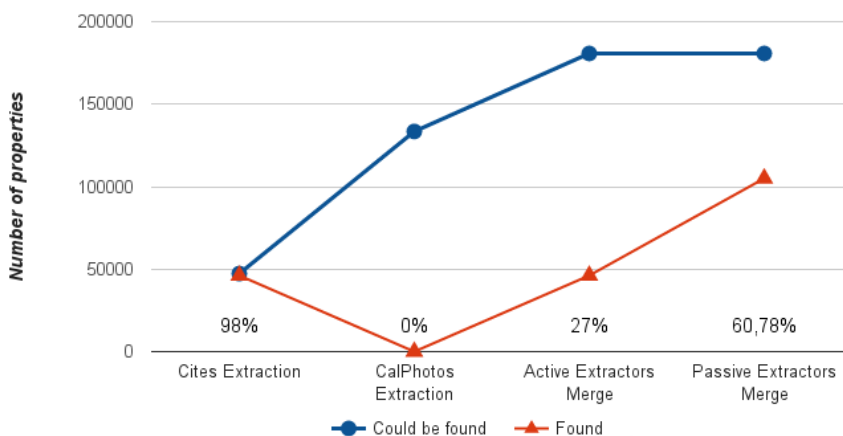


Figure 4. Number or rank properties found by active and passive extractors

The *CalPhotos* active extractor was unable to find the rank properties (only the common name was found). Conversely, the *Cites* active extractor found 46.310 properties out of 47.299, reaching 6,83 properties per animal (a 98% recall). After the deduplication and merge, we reached an average of 1,87 properties per animal (a 27% recall). After merging the indexed objects with the results coming from the *Wikipedia* passive extractor, we reached an average of 4,25 properties per animal (a 60,78% recall).

Another interesting usage of the passive extractors was to complement the indexed objects with images related to the corresponding animal. Table 2 shows the achieved

results. As we can see, *Google Images* was able to find images for almost every indexed animal, whereas wikipedia had images for approximately half the objects. Summing up, the average number of images per animal is greater than two. Only 47 animals remained without images.

Table 2. Images found by the passive extractors

Data source	Found Images	Images per animal
<i>Google Images</i>	24644	0,999
<i>CalPhotos</i>	18952	0,768
<i>Wikipedia</i>	13439	0,545
Total	57035	2,311

To conclude, we evaluate how reliable are the images found by the passive extractors. To help in the evaluation, we developed a tool that exhibits images for randomly chosen indexed animals and lets users indicate if the image really correspond to the animal (hit or not hit). The more hits obtained, the more precise are the passive extractors. Four settings were used, with fifty animals each, considering images extracted from i) *Google Images*, ii) *CalPhotos*, iii) *Wikipedia* and iv) all of them together. Two specialists participated of this experiment (one professor and one PhD student from areas related to biological research).

The results are presented in Figure 5. The hit and not hit values are an average of the number of hits and not hits from the two participants, respectively. As we can see, most of the sampled images are rightfully linked to an animal. Precision varies from 86% (*CalPhotos*) to 96% (*Google Images*). The extraction of the non corresponding images were due to the usage of unreliable data sources (e.g. *Google Images*). In other cases, the incorrect extraction was a symptom of the passive extractors deficiencies. The correction of the extraction algorithm (and possibly the adjustment of the message send by the mediator) is left as future work.

5. Related Work

The idea of helping extractors do their job is not really new. There are many possible ways of leveraging extraction methods. Here we discuss a few of them, namely focus crawling, wrapper induction and user-assisted extraction.

Focus crawlers are aimed at finding relevant sources (generally web pages) based on previously crawled relevant sources or a list of seeds provided by a user. There are several alternatives to select which pages to visit. Recent proposals achieve this by a diversity of machine learning algorithms that classify the sources as relevant or irrelevant [Safran et al. 2012, Dong and Hussain 2014].

Wrapper induction is a research area where the goal is to automatically create (or adjust) wrappers so they are able to extract new information from data sources. Previous knowledge is important to achieve this task. For instance, [Senellart et al. 2008] proposes a way to automatically create wrappers for the hidden web (data that sits behind HTML forms). An important part of the work is the extraction of data from the result pages. This

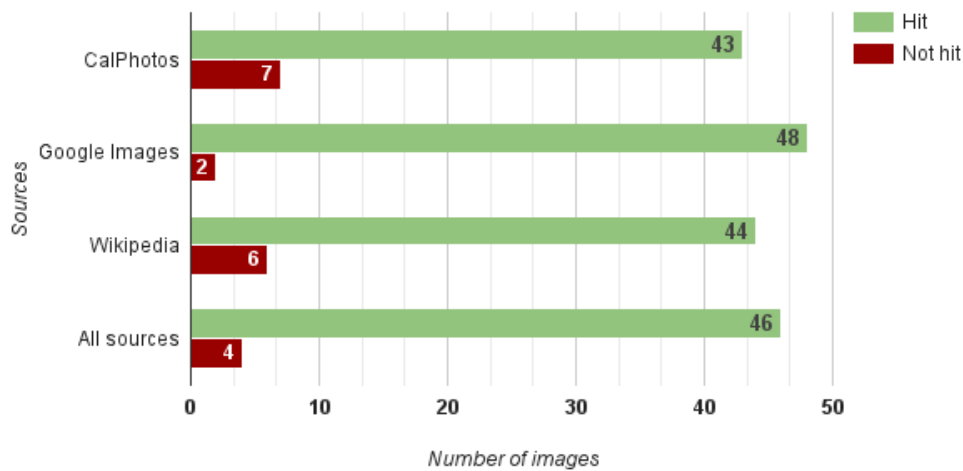


Figure 5. Precision of the image passive extractors.

is done by a supervised technique based on annotated pages used as training. A more recent approach proposes a wrapper induction technique that automatically identifies news articles from web sites. The idea is to use a list of news sites as input and to identify patterns among them. The wrapper then extracts news by seeking these patterns in other web sites [Xiang et al. 2015].

Finally, user-assisted extraction relies on information provided by the user to guide the extraction. In [Meng et al. 2003], the user accesses a tool to create mappings between an HTML page and a predefined schema. The mappings are transformed into extraction rules represented in XQuery. During extraction, the xquery is executed, transforming the data source into an XML that satisfies the schema. The identification of an object is done by unique elements of the schema. Another related approach is proposed by [Baumgartner et al. 2001]. In this case, the extraction rules are defined in a language called Elog.

Even though all of the above receive external help, they cannot be classified as passive extractors (as we propose) because their main purpose is to find new objects instead of complementing the ones that already exist. Besides, they are tailored for very specific goals: finding relevant data sources (focus crawling), finding objects within a data source (wrapper induction) and uniquely identifying an object as a way to allow deduplication (user assisted extraction). They are unfitted to perform the three things altogether. On the other hand, passive extractors are more naturally suitable for the mentioned tasks, since they focus on single objects at a time.

6. Conclusion

This paper presents an extension for extraction architectures used by materialized data integration systems. The extension classifies extractors and mediators as passive or active. In the active extractor/passive mediator combination, the extractors look for relevant objects and the mediators are responsible for checking whether the identified objects need to

be merged with already stored objects. Conversely, in the passive extractor/active mediator combination, the extractor receives a message from the mediator asking to complement information about a stored object.

An actual implementation of the architecture was also provided, with the purpose of extracting objects related to animals in order to feed the repository of bioID, a collaborative system for biological classification. The results showed that it is rather straightforward to create passive extractors that find relevant data sources, find objects within data sources and easily identify object duplication.

In comparison with active extraction approaches, the message-driven alternative has the potential to produce extractors that are less complex and more effective. Of course, active extractors are still of great importance. After all, it takes active extractors to feed the repository at the first place. What we state is that passive and active extractors combined constitute a more powerful solution, as passive extractors are able to fill the gaps left by ordinary extraction methods.

References

- Anter, S., Zellou, A., and Idri, A. (2016). Retrieving and materializing data in hybrid mediators. *International Journal of Applied Engineering Research*, 11(3):2128–2134.
- Baumgartner, R., Flesca, S., and Gottlob, G. (2001). Visual web information extraction with lixto. In *VLDB*, volume 1, pages 119–128.
- Cardoso, P., Peripolli, G., Franzin, F., and Mergen, S. (2015). Ambiente colaborativo para identificação de espécies. In *Anais do XIII Simpósio de Informática (SIRC)*, pages 68–73. Centro Universitário Franciscano.
- Dong, H. and Hussain, F. K. (2014). Self-adaptive semantic focused crawler for mining services information discovery. *Industrial Informatics, IEEE Transactions on*, 10(2):1616–1626.
- Meng, X., Wang, H., Hu, D., and Li, C. (2003). A supervised visual wrapper generator for web-data extraction. In *Computer Software and Applications Conference, 2003. COMPSAC 2003. Proceedings. 27th Annual International*, pages 657–662. IEEE.
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., and Worm, B. (2011). How many species are there on earth and in the ocean? *PLoS Biol*, 9(8):e1001127.
- Safran, M. S., Althagafi, A., and Che, D. (2012). Improving relevance prediction for focused web crawlers. In *Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on*, pages 161–166. IEEE.
- Senellart, P., Mittal, A., Muschick, D., Gilleron, R., and Tommasi, M. (2008). Automatic wrapper induction from hidden-web sources with domain knowledge. In *Proceedings of the 10th ACM workshop on Web information and data management*, pages 9–16. ACM.
- Xiang, Z.-L., Yu, X.-R., and Kang, D.-K. (2015). Wrapper induction of news information for feeding to social networking service on smartphone. In *Advanced Communication Technology (ICACT), 2015 17th International Conference on*, pages 292–295. IEEE.

Reconhecimento de sinais estáticos de LIBRAS com *Support Vector Machines* usando Kinect

Leonardo Perdomo¹, Mozart Lemos de Siqueira¹

¹Curso de Ciência da Computação – Centro Universitário La Salle (UNILASALLE)
Caixa Postal 2.288 – 92.010-000 – Canoas – RS – Brazil

leonardo.perdomo.edu@outlook.com, mozarts@unilasalle.edu.br

Abstract. *This paper presents the author experiments with his own developed prototype aiming computer recognition of the manual alphabet static signs from the Brazilian Sign Language (LIBRAS), captured with Microsoft Kinect depth sensor, using image pattern recognition techniques along with Multiclass Support Vector Machines (SVM) classifiers. The prototype results and an efficiency analysis with execution time measures are presented. The device practical distance limits interval (0.8m to 2.5m) was considered.*

Resumo. *Este artigo apresenta a experimentação realizada pelo autor em um protótipo desenvolvido pelo próprio para o reconhecimento computacional dos sinais estáticos do alfabeto manual da Língua Brasileira de Sinais (LIBRAS), capturados através do sensor de profundidade do Microsoft Kinect, utilizando técnicas de reconhecimento de padrões em imagens com classificação por Support Vector Machines (SVM) em uma abordagem multiclasse. São apresentados os resultados do protótipo e uma análise de eficiência em medições de tempo de execução e acerto no reconhecimento de sinais. Foi considerado o intervalo de distância dentro dos limites práticos (0,8m à 2,5m) do near-mode do dispositivo.*

1. Introdução

O desafio do reconhecimento das línguas de sinais por sistemas computacionais envolve a compreensão de sua organização estrutural através de estudos linguísticos, como os realizados em [Brito 1995] e [de Quadros 1997], que descrevem a composição da Língua Brasileira de Sinais (LIBRAS) em relação à *American Sign Language* (ASL) e às línguas faladas, também observando sua independência e originalidade sociocultural, a exemplo das diferenças gestuais existentes por regiões do Brasil.

Baseando-se nas diferentes abordagens adotadas em alguns dos trabalhos publicados sobre o tema [de Souza et al. 2012b], [de Souza et al. 2012a], [Zhu e Wong 2012] e [de Souza e Pizzolato 2013], este estudo apresenta o desenvolvimento de um protótipo para reconhecimento de 20 letras do alfabeto manual de LIBRAS (Figura 1a), escolhidas em razão da ausência de movimento na realização de seus respectivos gestos. A captura do gesto ocorre pelo sensor de profundidade do Microsoft Kinect (Figura 1b), seguida da segmentação e extração de descritores da imagem, posteriormente classificados por *Support Vector Machines* (SVM). Também são apresentadas medições de eficiência do protótipo, analisando os resultados obtidos em relação ao observado em outros estudos publicados. A concepção, implementação, experimentação e análise de resultados do trabalho foram realizadas pelo aluno Leonardo Perdomo, sob orientação do professor Mozart Lemos de Siqueira.

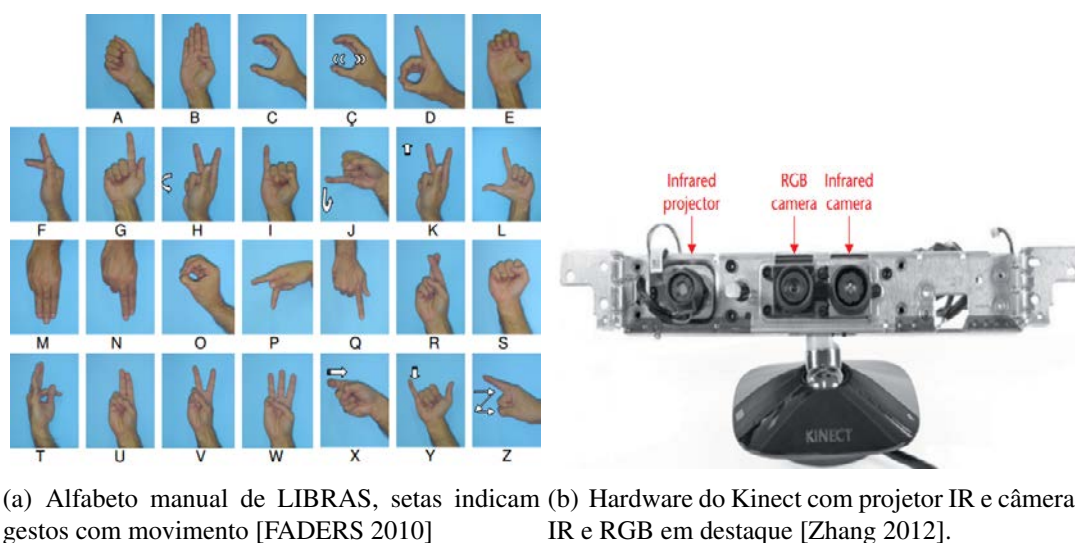


Figura 1. Alfabeto manual de LIBRAS e o dispositivo Kinect utilizado na captura de movimento do *console* de videogame Xbox 360.

No decorrer da seção 2 são apresentados outros trabalhos publicados sobre o tema estudado e seus respectivos resultados. Em seguida, na seção 3, é descrita a metodologia adotada para o desenvolvimento e análise do protótipo, disponível em código-aberto¹, para reconhecimento em tempo real dos sinais estáticos do alfabeto manual de LIBRAS, utilizando classificação por SVM de descritores SURF extraídos da captura de profundidade do Kinect. Os resultados de medições de eficiência do protótipo em percentuais médios de acerto e tempo de execução, considerando três implementações de *kernel* SVM (linear, polinomia e radial) no intervalo de distância dentro dos limites práticos em *near-mode* (0.8m à 2.5m) do dispositivo são apresentados na seção 4, com a exposição na seção 5 das conclusões obtidas pelo estudo realizado.

2. Trabalhos Relacionados

Os meios utilizados para o reconhecimento computacional de gestos das línguas de sinais podem envolver, conforme análise realizada por [Mitra e Acharya 2007], a coleta de dados por meio de sensores equipados em luvas, ou a utilização de câmeras em conjunto com técnicas para análise de imagens. Este levantamento também identifica a adoção, por trabalhos que abrangem este tema, da classificação dos gestos em sinais estáticos ou dinâmicos, definidos pela ocorrência ou não de movimento após sua postura inicial. Para um aprofundamento mais específico do tema, este estudo limitou-se a investigar referências que utilizaram técnicas de visão computacional para o reconhecimento do alfabeto e de palavras das línguas de sinais. Os trabalhos analisados apresentam resultados significativos, com limitações quanto à iluminação do ambiente quando no uso de câmeras RGB convencionais [Carneiro et al. 2009], [Pizzolato et al. 2010], relativamente superadas com a adoção de dispositivos com sensores de profundidade (RGB-D), como o Microsoft Kinect [Zafrulla et al. 2011], [Zhu e Wong 2012], [de Souza e Pizzolato 2013], [Almeida et al. 2014], [Pedersoli et al. 2014], [Rioux-Maldague e Giguere 2014], [Dong et al. 2015].

¹Disponível em https://github.com/lperdomo/reconhecedor_sinal_estatico

Entre os trabalhos analisados, [de Souza et al. 2012b], [de Souza et al. 2012a] destacam-se pela compreensiva análise de técnicas para o reconhecimento de gestos, verificando, por exemplo, uma superioridade de *Support Vector Machines* (SVM) com *kernels* não-lineares (polinomial quadrático e radial) no melhor resultado encontrado em relação às *Feedforward Neural Networks* (FNN). Ambos os trabalhos citados também apontam a necessidade de treinamentos significativamente menores com uso de classificadores por SVM em relação às FNNs. Aprofundando os estudos realizados apenas com câmeras RGB, [de Souza e Pizzolato 2013] investigam a adição do Kinect, e de um modelo de etapas para classificação de postura e sequência de movimento de sinais dinâmicos, apresentado em [Pizzolato et al. 2010]. Os autores verificam uma superioridade substancial na aprendizagem discriminativa com o classificador de sequência *Hidden Conditional Random Fields* (HCRF) em relação à generativa de *Hidden Markov Models* (HMM).

No uso do Kinect para captura de gestos, [Zhu e Wong 2012] destacam-se ao investigar uma implementação, com extração de descritores SIFT da *American Sign Language* (ASL) e classificação por SVM, considerando 5 distâncias do usuário em relação ao Kinect (0, 6m, 0, 9m, 1, 2m, 1, 5m e 1, 8m). Os autores verificam uma redução substancial da qualidade dos descritores de profundidade após 1, 5m que inviabilizou resultados satisfatórios, optando então pelo uso em conjunto com a câmera RGB do dispositivo.

Entre as publicações mais recentes analisadas, vale destacar a implementação em código aberto de [Pedersoli et al. 2014], para reconhecimento de sinais dinâmicos da ASL usando SVM e HMM. Em [Almeida et al. 2014] é implementada a extração de múltiplas características (velocidade, distância entre pixels, descritores SURF, entre outras), classificando gestos dinâmicos apenas com o uso de SVM. Distinguindo-se na abordagem em relação aos demais estudos observados, [Rioux-Maldague e Giguere 2014] utilizaram *Deep Belief Network* (DBN) com *Restricted Boltzmann Machines* (RBM). Já [Dong et al. 2015] optaram pela utilização de uma luva colorida para segmentação da estrutura das mãos, aplicando então um classificador em *Random Forest* (RF).

Este estudo se assemelha ao realizado por [Zhu e Wong 2012] quanto à utilização de técnicas para extração de descritores e classificação com SVM, avaliando a eficiência do reconhecimento em relação à distância do Kinect, porém distinguindo-se ao considerar somente descritores SURF da captura de profundidade. Também há semelhança na verificação de resultados entre os *kernels* SVM linear, polinomial e radial realizada em [de Souza et al. 2012a], [de Souza et al. 2012b] e [de Souza e Pizzolato 2013].

3. Metodologia

Para desenvolvimento e experimentação do protótipo reconhecedor de sinais estáticos do alfabeto manual da Língua Brasileira de Sinais (LIBRAS), foi utilizado um computador Intel I5 4690 3.5GHz com ASUS Z-97 PRO, 8Gb de RAM e placa gráfica GeForce GTX760, em conjunto com um dispositivo Microsoft Kinect for Xbox 360. A linguagem de programação C++ foi utilizada em ambiente Windows 8.1 64bits, com IDE Visual C++ 2010 em conjunto com a biblioteca OpenCV 2.4.10, integrada às bibliotecas OpenNI 1.5.4.0 32bits e NITE 1.5.2.21 32bits, acessando os recursos do Kinect pelo driver PrimeSense 5.1.2.1 32bits. A descrição sobre o desenvolvimento do protótipo é apresentada na subseção 3.1, enquanto a experimentação para investigação da eficiência do mesmo, em uma abordagem quantitativa, é apresentada na subseção 3.2.

3.1. Desenvolvimento do Protótipo

Para contornar as restrições impostas do modelo Kinect *for* Xbox 360, escolhido pelo custo significativamente menor em relação ao modelo de desenvolvimento, foram utilizadas as bibliotecas OpenCV, NITE e OpenNI com o *driver* PrimeSense, devido à integração entre as mesmas [Cruz et al. 2012], além desta última possuir a biblioteca SURF integrada e implementação SVM *1-vs-1* baseada em LIBSVM [OpenCV 2011]. Desta forma o mapeamento de disparidade capturado pelo sensor de profundidade do Kinect em *near-mode* foi realizado sem uso da sua biblioteca SDK. Já o acesso às estruturas do esqueleto do usuário em nodos foi obtido por meio da biblioteca SkeletonSensor, desenvolvida por [Johnson et al. 2012].

Para segmentação, a área quadrada de 180x180 pixels centralizada no nodo do mão direita foi arbitrariamente escolhida como Região de Interesse (ROI), baseando-se na observação do espaço ocupado na distância de 0,8m em relação ao Kinect. O fundo então é removido com base na informação de profundidade do nodo da mão, enquanto, para remoção do antebraço, inicialmente é aplicado *thresholding* com método de Otsu [Otsu 1979], utilizando-o na identificação do polígono que abrange mão e antebraço. Neste polígono é realizada a busca pelo maior círculo interno, pressupondo a palma da mão ao considerar a situação ideal onde o usuário não utilize vestimentas espessas no antebraço que comportariam círculos internos maiores. A informação de profundidade no centro do círculo identificado é utilizada para remoção de regiões mais distantes abaixo do mesmo, na tentativa de remover o antebraço e preservar estruturas da mão para sinais com postura dos dedos apontada para baixo. Em seguida o algoritmo de Canny [Canny 1986] é usado para detecção das estruturas internas da mão entre as diferentes profundidades detectadas, como pode ser observado na Figura 2.

Após a segmentação das características da mão, a biblioteca SURF é utilizada para extração dos descritores de cada *frame* capturado. Estes são utilizados para composição de histogramas, medindo as ocorrências de descritores extraídos em relação aos previamente definidos em um vocabulário visual. Este vocabulário é composto por descritores agrupados e otimizados por gestos, através do algoritmo *k-means*, também usado para este fim em outros estudos [Csurka et al. 2004], [Jiang et al. 2007]. Por fim, os histogramas são encaminhados às três SVMs distintas, com *kernels* linear, polinomial e radial, para treinamento ou classificação entre as 20 letras definidas do alfabeto manual de LIBRAS.

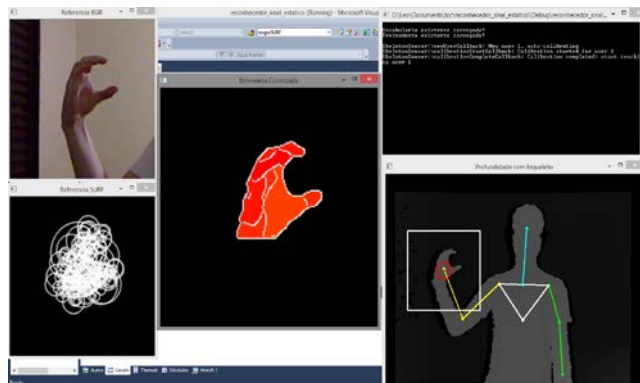


Figura 2. Protótipo desenvolvido em funcionamento, demonstrando o esqueleto detectado na cena, com a mão segmentada e os descritores SURF encontrados.

3.2. Experimentação do protótipo

Considerando os limites práticos do *near-mode* do Kinect ($0,8m$ à $2,5m$) [Microsoft 2015], foram definidos intervalos de $0,2m$, inicialmente compreendendo as distâncias entre $0,8m$ e $2,4m$ para avaliação do protótipo, com o dispositivo a $1,28m$ de altura. Após as primeiras observações, optou-se pela remoção das distâncias de $2,2m$ e $2,4m$, devido à extração insuficiente de descritores em razão da resolução pequena de captura, somada à alta incidência de ruído (dispersão de IR). Também optou-se pela composição do vocabulário e da primeira avaliação do protótipo com amostras capturadas a $1m$, devido à incidência de ruído ocorrida em $0,8m$ (limite prático do *near-mode*).

Baseando-se na quantidade adotada por [Zhu e Wong 2012], os treinamentos foram realizados capturando 40 amostras por letra do conjunto definido (20 letras) à cada distância treinada, usando a mão direita do autor. O conjunto de amostras, em formato PNG, foi fornecido às três SVMs utilizando o processo de treinamento e otimização existente na biblioteca OpenCV. Os parâmetros para cada *kernel* foram estimados por buscas em grade com validação cruzada *k-fold*, com $k = 10$ (padrão da biblioteca), obtendo os parâmetros $C = 3,125 * 10^2$ (coeficiente de distinção das classes) para os três *kernels*, $\gamma = 0,50625 + 9 * 10^{-11}$ para o polinomial e o radial, além de grau $d = 0,07 + 7 * 10^{-16}$ e $coef0 = 0,1 + 1 * 10^{-15}$ para o polinomial. As validações utilizaram o mesmo critério adotado nos treinamentos, totalizando 5600 amostras, também gravadas em formato PNG.

4. Resultados

Em sua primeira avaliação com treinamentos a $1m$, o protótipo obteve percentual médio de acerto entre os três *kernels* de $81,75\% \pm 0,5885\%$, conforme Figura 3. Treinamentos adicionais em $0,8m$, $1,2m$, $1,4m$ e $1,6m$, resultaram em um aumento significativo de acertos em distâncias até $1,4m$. A interrupção de treinamentos em $1,6m$ ocorreu pela insuficiente diminuição do índice de dispersão ($CV = 36,8089\%$), com pequeno aumento no acerto desta distância (aprox. $10,58\%$), de forma semelhante ao verificado por [Zhu e Wong 2012] em descritores de profundidade devido à baixa resolução após $1,5m$.

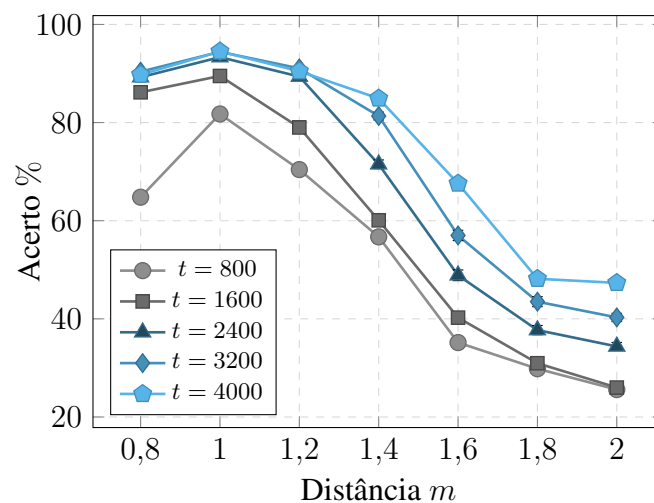


Figura 3. Percentuais médios de acerto dos *kernels* SVM em relação à quantidade de treinamento t realizada.

Os resultados individuais de cada *kernel* são observados na Figura 4a, que demonstra a ocorrência na distância de $1m$ dos maiores percentuais médios de acerto em relação às demais distâncias, com o *kernel* linear apresentando o valor de $94,5\% \pm 0,5496\%$, o polinomial, $94,375\% \pm 0,5417\%$ e o radial, $94,625\% \pm 0,5221\%$, considerando um intervalo de confiança (IC) de 95%. Comparativamente, os *kernels* demonstram grande proximidade entre seus resultados percentuais de acerto, nesta e em outras distâncias, não sendo possível determinar a superioridade de um em relação aos demais por este critério. É importante destacar que estes resultados, utilizando apenas sensores de profundidade na captura, são semelhantes aos valores obtidos com câmeras RGB por [de Souza et al. 2012a] e com ambos (RGB-D) em [Zhu e Wong 2012].

A ocorrência dos melhores resultados na distância de $1m$ tem como principal motivo a composição do vocabulário visual por amostras capturadas nesta distância. Como resultado, descritores capturados na mesma posição possuem maior semelhança aos utilizados nos histogramas representativos de cada letra, em relação às capturas realizadas em outras distâncias. O critério adotado para este procedimento, conforme citado na seção 3, envolve a preferência por uma posição mais próxima ao dispositivo, proporcionando a extração de descritores mais precisos, em uma imagem com maior resolução. A distância de $0,8m$ foi considerada inadequada devido à observação de uma maior ocorrência de ruído, ocasionado por tratar-se da limitação prática do *near-mode* do dispositivo.

A investigação mais aprofundada da diminuição do percentual médio de acertos entre as distâncias $1,2m$ e $1,6m$, apresentada na Figura 4b, verificou a relação ao comportamento observado após $1,5m$ por [Zhu e Wong 2012] em descritores de profundidade obtidos em captura com o Kinect. Por meio de validações adicionais nas distâncias de $1,3m$ e $1,5m$, seguindo os mesmos critérios adotados anteriormente nas demais validações, demonstrou a diminuição gradual da precisão do protótipo até $1,5m$, seguida de uma queda brusca de aproximadamente 11,5%, para o *kernel* linear, 11,75%, para o polinomial, e 12,625% para o radial.

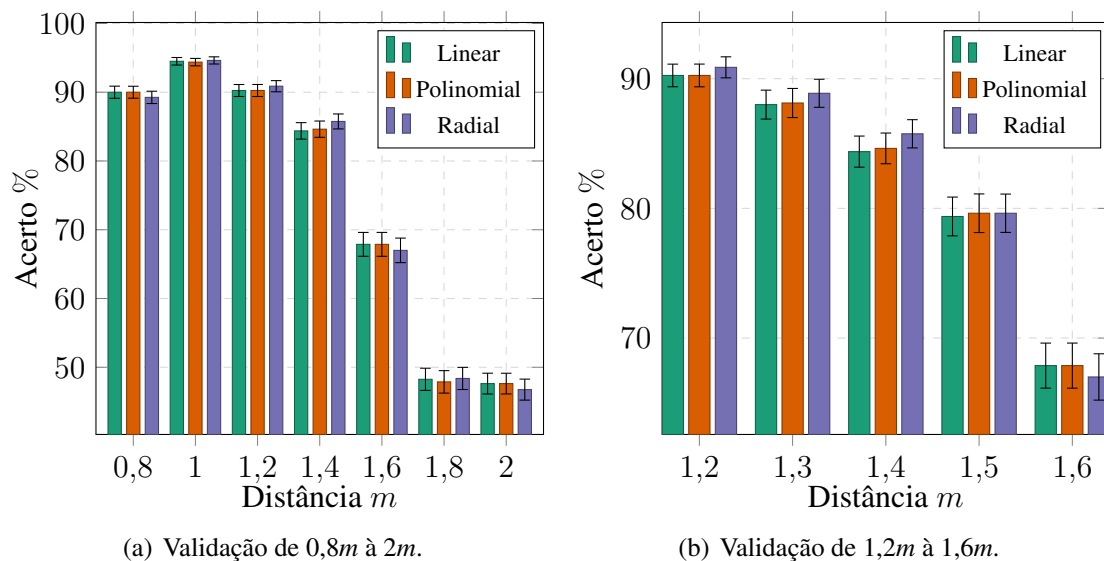
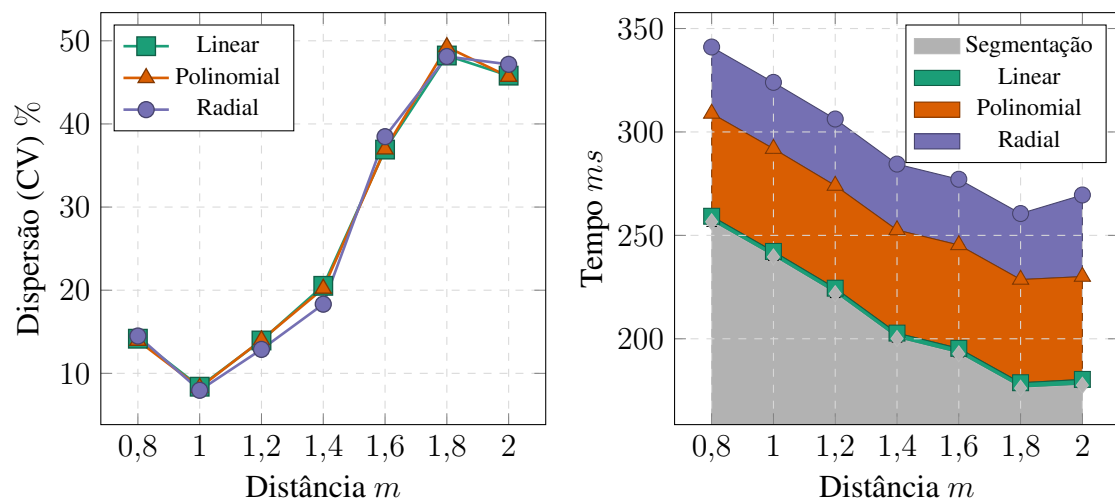


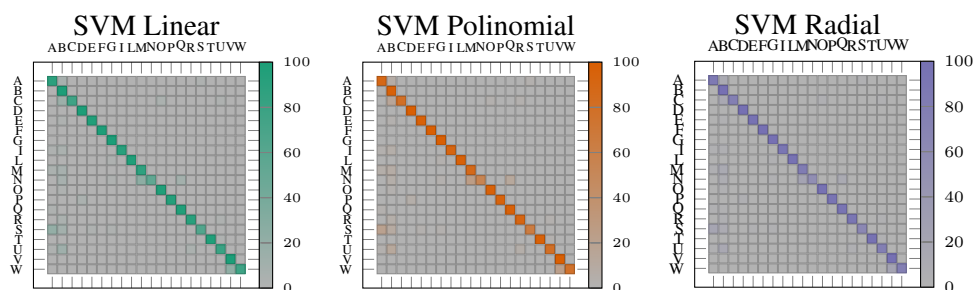
Figura 4. Percentuais médios de acerto por *kernel* SVM em validação, considerando IC de 95%.

Ao investigar a dispersão dos percentuais de acerto através do coeficiente de variação (CV), é observada uma maior homogeneidade na distância $1m$, com o *kernel* linear apresentando $8,3928\%$, o polinomial $8,2827\%$ e o radial $7,9619\%$. A Figura 5a permite afirmar que, entre as distâncias até $1,4m$ e independentemente do *kernel*, os resultados apresentam-se de forma homogênea, com dispersão de até 20% , em relação à heterogeneidade devido à queda brusca de qualidade na captura a partir de $1,6m$. As ocorrências de erros no reconhecimento de cada letra na distância de $1m$, apresentadas na Figura 5c, demonstram confusão especialmente entre as letras N e M, e, em menor intensidade, entre as letras A e S. Este problema ocorre pela semelhança entre estes gestos, com diferença apenas na posição de um dedo que exige uma captura de maior precisão.

Observando os resultados das medidas de desempenho por tempo de execução, apresentado na Figura 5b, é verificada a distinção entre o tempo decorrido para classificação por cada *kernel*, com a implementação linear tomando apenas $2,6681 \pm 0,0026ms$, em relação aos $52,4260 \pm 0,0088ms$ e $85,4959 \pm 0,0702ms$ (IC de 95%) das implementações polinomial e radial, respectivamente. As variações no tempo total do processo de reconhecimento ocorrem no tempo de segmentação, devido relação da proximidade com a quantidade de características segmentadas e extraídas no *frame*, diminuindo gradualmente com o aumento da distância.



(a) Dispersão (CV) por *kernel* SVM entre $0,8m$ e $2m$ (b) Tempo de execução por etapas entre $0,8m$ e $2m$



(c) Matrices de confusão por *kernel* SVM dos resultados à $1m$ do dispositivo.

Figura 5. Análise de dispersão (CV), tempo de execução e matrizes de confusão por *kernel* com ocorrências entre classes na distância de $1m$.

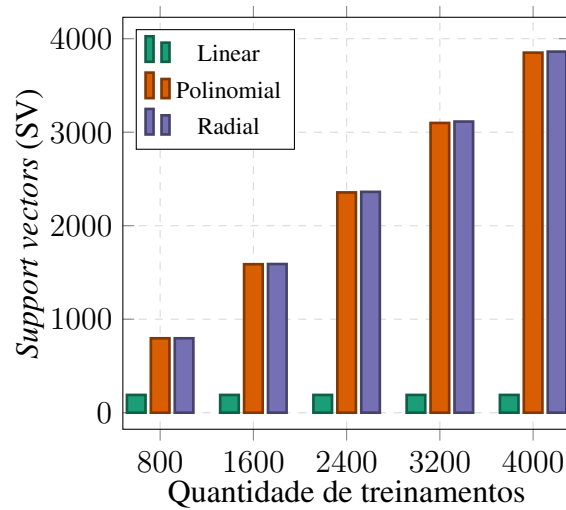


Figura 6. Support vectors (SV) por kernel SVM após treinamento e otimização.

Considerando as diferenças de desempenho observadas entre os três *kernels* quanto ao tempo de execução do processo de classificação, optou-se pela verificação da quantidade de *Support Vectors* (SVs) encontrados após os processos de treinamento e otimização realizados, existentes para a implementação SVM da biblioteca OpenCV. Como apontado por [de Souza et al. 2012a], a quantidade de SVs impacta no desempenho do modelo, podendo não produzir uma boa generalização em quantidade insuficiente, ou *overfitting* se em excesso. Baseando-se nisto, é possível observar na Figura 6 que o *kernel* linear manteve sua quantidade de SVs (190) constante, após sucessivas adições na quantidade de treinamentos. Já para os *kernels* não-lineares há um aumento progressivo de seus SVs, com uma pequena e crescente diferença entre ambos, a ponto de, na última adição de treinamentos, o *kernel* radial possuir 11 SVs adicionais em relação ao polinomial. Estas diferenças indicam a otimização obtida em cada SVM, justificando o melhor desempenho em tempo de execução do *kernel* linear em relação aos não-lineares.

5. Conclusão

O reconhecimento de sinais estáticos de LIBRAS proposto foi implementado utilizando *Support Vector Machines* (SVM) para classificação de descritores extraídos após segmentação da captura de profundidade do Kinect. As avaliações realizadas com o protótipo apresentaram maiores percentuais médios de acerto na distância de 1m, com $94,5\% \pm 0,5496\%$ para o *kernel* linear, $94,375\% \pm 0,5417\%$ para o polinomial e $94,625\% \pm 0,5221\%$ para o radial, considerando um intervalo de confiança (IC) de 95%. Estes valores percentuais não são suficientemente distintos para determinar a abordagem SVM com maior precisão, sendo então investigado o desempenho em tempo de execução de classificação. Nesta perspectiva, a implementação linear obteve a maior eficiência considerando seus resultados alcançados com tempo de execução consumindo em média $2,6681 \pm 0,0026ms$ em relação aos $52,4260 \pm 0,0088ms$ e $85,4959 \pm 0,0702ms$ das implementações polinomial e radial, respectivamente. Este melhor desempenho pode ser justificado através de uma generalização otimizada obtida em treinamento, com apenas 190 *Support Vectors* (SVs) em relação aos respectivos 3851 SVs e 3862 SVs de ambas implementações não-lineares.

Através da avaliação do protótipo considerando as distâncias do usuário em relação ao dispositivo, foi possível reproduzir a queda drástica em percentuais de acerto após 1,5m observada por [Zhu e Wong 2012], ocorrida devido à baixa resolução da câmera de profundidade do Kinect. Já a superioridade dos resultados apresentados na distância de 1m, em comparação às demais posições, justifica-se pela sua utilização na captura de amostras para composição dos histogramas de descritores do vocabulário visual. É importante destacar que os valores obtidos nesta distância são semelhantes aos resultados encontrados em [de Souza et al. 2012a] e [Zhu e Wong 2012], considerando que o primeiro utiliza câmera RGB, e o segundo RGB-D do Kinect, enquanto este estudo apenas utilizou a captura do sensor de profundidade.

Por fim, sugere-se uma investigação deste protótipo com participação de voluntários, a fim de observar seu desempenho utilizando amostras com mãos de diferentes formatos. Outra sugestão consiste na experimentação com outros dispositivos, como o Kinect 2, em busca de uma maior precisão na captura, possivelmente superando as limitações verificadas no Kinect neste estudo. Também recomenda-se experimentar novas abordagens, utilizando outras técnicas de segmentação, extração e classificação a fim de observar possíveis vantagens para precisão e desempenho em sua utilização.

Referências

- Almeida, S. G. M., Guimarães, F. G., e Ramírez, J. A. (2014). Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors. *Expert Systems with Applications*, 41(16):7259–7271.
- Brito, L. F. (1995). *Por uma Gramática de Línguas de Sinais*. Tempo Brasileiro, 1 edition.
- Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698.
- Carneiro, A. T. S., Cortez, P. C., e Costa, R. C. S. (2009). Reconhecimento de Gestos da LIBRAS com Classificadores Neurais a partir dos Momentos Invariantes de Hu. In *Anais do 1º Congresso Regional de Design de Interação - Interaction South America 09*, pages 193–198.
- Cruz, L., Lucio, D., e L.Velho (2012). Kinect and RGBD Images: Challenges and Applications. In *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pages 36–49.
- Csurka, G., Dance, C. R., Fan, L., Willamowski, J., e Bray, C. (2004). Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22.
- de Quadros, R. M. (1997). *Educação de surdos - A aquisição da linguagem*. Artmed.
- de Souza, C. R. e Pizzolato, E. B. (2013). Sign Language Recognition with Support Vector Machines and Hidden Conditional Random Fields: Going from Fingerspelling to Natural Articulated Words. In *Proceedings of the 9th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM 2013)*, pages 84–98. Springer.
- de Souza, C. R., Pizzolato, E. B., e dos Santos Anjo, M. (2012a). Fingerspelling Recognition with Support Vector Machines and Hidden Conditional Random Fields: A

- Comparison with Neural Networks and Hidden Markov Models. In *Proceedings of the 13th Ibero-American Conference on Artificial Intelligence (IBERAMIA'12)*, pages 561–570. Springer.
- de Souza, C. R., Pizzolato, E. B., e dos Santos Anjo, M. (2012b). Recognizing Static Signs from the Brazilian Sign Language: Comparing Large-Margin Decision Directed Acyclic Graphs, Voting Support Vector Machines and Artificial Neural Networks.
- Dong, C., Leu, M. C., e Yin, Z. (2015). American Sign Language Alphabet Recognition Using Microsoft Kinect. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 44–52.
- FADERS (2010). *Mini dicionário*. Fundação de Articulação and Desenvolvimento de Políticas Públicas para Pessoas com Deficiências and Altas Habilidades no Rio Grande do Sul (FADERS).
- Jiang, Y.-G., Ngo, C.-W., e Yang, J. (2007). Towards optimal bag-of-features for object categorization and semantic video retrieval. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval, CIVR '07*, pages 494–501. ACM.
- Johnson, G. P., Abram, G. D., Westing, B., Navrátil, P., e Gaither, K. (2012). Display-Cluster: An Interactive Visualization Environment for Tiled Displays. In *2012 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 239–247.
- Microsoft (2015). Kinect for Windows Programming Guide.
- Mitra, S. e Acharya, T. (2007). Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3):311–324.
- OpenCV (2011). *OpenCV 2.4 documentation*. OpenCV Foundation.
- Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66.
- Pedersoli, F., Benini, S., Adami, N., e Leonardi, R. (2014). XKin: an open source framework for hand pose and gesture recognition using kinect. In *The Visual Computer*, volume 30, pages 1107–1122. Springer.
- Pizzolato, E. B., dos Santos Anjo, M., e Pedroso, G. (2010). Automatic Recognition of Finger Spelling for LIBRAS Based on a Two-Layer Architecture. In *Proceedings of the 2010 ACM Symposium on Applied Computing (SAC '10)*, pages 969–973.
- Rioux-Maldague, L. e Giguere, P. (2014). Sign Language Fingerspelling Classification from Depth and Color Images Using a Deep Belief Network. In *2014 Canadian Conference on Computer and Robot Vision (CRV)*, pages 92–97.
- Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., e Presti, P. (2011). American Sign Language Recognition with the Kinect. In *Proceedings of the 13th International Conference on Multimodal Interaction (ICMI'11)*, pages 279–286.
- Zhang, Z. (2012). Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia*, 19(2):4–12.
- Zhu, X. e Wong, K. K. (2012). Single-frame hand gesture recognition using color and depth kernel descriptors. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, pages 2989–2992. IEEE.

Desmistificando o uso de *Matrix Factorization* em domínios de recomendação

Nícollas Silva , Leonardo Rocha , Fernando Mourão

DCOMP/UFSJ - São João del Rei, Minas Gerais, Brasil

{nicollasilva, lcrocha, fhmourao}@ufsj.edu.br

Abstract. *Despite the popularity of Matrix factorization methods (MF) in Recommender Systems (RSs), it is not clear the practical distinction among them in real scenarios. In this context, this work organizes existing methods, proposing a taxonomic organization for the area, and characterizes their performance, correlating the application success with domain characteristics. Aiming at the practical use of this research, we present a characterization framework for MF methods applied to RSs, useful for contrasting the performance of different methods. Analyses of actual collections have shown that recommendations generated via SVD tend to focus on non-popular items, while recommendations issued by PCA focuses on popular items.*

Resumo. *Apesar da popularidade de métodos de Matrix Factorization (MF) em Sistemas de Recomendação (SsR), não é clara a distinção prática entre eles em domínios reais. Neste contexto, este trabalho organiza métodos existentes, propondo uma organização taxonômica para a área, e caracteriza seus desempenhos, correlacionando seus sucessos da aplicação com características de cada domínio. Visando o uso prático desta pesquisa, apresentamos um arcabouço para caracterização de métodos de MF aplicados a SsR, útil para se contrastar o desempenho de distintos métodos. Análises sobre coleções reais demonstraram que recomendações geradas via SVD tendem a focar em itens não populares, enquanto recomendações oriundas do PCA focam em itens populares.*

1. Introdução

Atualmente, Sistemas de Recomendação (SsR) constituem uma das mais importantes ferramentas utilizadas para auxiliar usuários, de maneira personalizada, na tomada de decisão em variados domínios reais [Koren et al. 2009]. Dentre as diversas propostas existentes na literatura, particular atenção é dada a SsR baseados em Fatoração de Matrizes (i.e., *matrix Factorization* - MF), dada a elegante formalização que MF proporciona ao problema de recomendação, bem como a alta eficiência frequentemente observada em sua aplicação neste tipo de problema. De fato, os principais recomendadores estado-da-arte atualmente utilizam em alguma etapa do processo de recomendação um método baseado em MF [Sharifi et al. 2013, Adomavicius and Tuzhilin 2005]. A maioria destes SsR correspondem a adaptações para os domínios de recomendação dos tradicionais métodos de Singular Value Decomposition (SVD) e o Principal Component Analysis (PCA).

Apesar da crescente popularidade destes métodos, os trabalhos atuais não se preocupam em entender e contrastar o real impacto de cada adaptação sobre a tarefa de

recomendação. Perguntas tais como: *Quais as diferenças e semelhanças entre as diversas propostas de adaptação? Diferenças conceituais determinam diferenças práticas quanto ao desempenho destes SsR? Quais domínios são mais apropriados para cada tipo de técnica?* apesar de pertinentes são negligenciadas na literatura. Responder tais perguntas pode nos permitir gerar recomendações com melhor qualidade. Assim, este trabalho propõe revisar, caracterizar, organizar e distinguir distintos métodos de recomendação baseadas em MF. Além disso, almejamos correlacionar características dos domínios de análise com a qualidade das recomendações geradas por cada método avaliado.

Neste sentido, primeiramente, consolidamos uma taxonomia de classificação para diversos métodos de recomendação baseados em MF, com base em características comumente exploradas por estes métodos. Tal taxonomia considera cinco dimensões, associadas a passos distintos do processo de recomendação, identificadas através de análises não-automáticas sobre diversos métodos existentes. Estas dimensões levam em conta a tarefa que está sendo realizada pela técnica, a forma de modelagem dos dados de entrada, o método de fatoração utilizado, a estratégia de recomendação e a fórmula de predição utilizada. Cabe ressaltar que a taxonomia proposta é não-fechada e extensível, uma vez que não engloba todas as características potencialmente relevantes, nem todos os métodos existentes. Em seguida, organizamos 16 SsR distintos baseados em MF de acordo com a taxonomia proposta, enfatizando diferenças e semelhanças entre estes. Dessa forma, observamos mais facilmente que métodos tradicionalmente apresentados como distintos, tais como *SimpleSVD* e *ConventionalSVD*, são na verdade idênticos. Ou ainda, constatamos que a diferença entre os métodos *NormalizedSVD* e *SVDP* restringe-se apenas à fórmula de modelagem dos dados utilizada. De maneira geral, por meio da taxonomia proposta, este trabalho auxilia pesquisadores da área a facilmente visualizarem conceitos e semelhanças entre técnicas, que são apresentadas de maneira confusa na literatura.

Além de organizar qualitativamente os diversos SsR baseados em MF usando suas premissas e abordagens, propomos uma metodologia de caracterização capaz de quantificar o comportamento de cada método, bem como correlacionar indicadores de qualidade de SsR com algumas características das coleções de entrada. Consequentemente, visamos entender melhor as características, pontos fortes e fracos de cada método, bem como o domínio mais adequado para a aplicação destes. Por fim, consolidamos um arcabouço para caracterização aplicável a distintos domínios (e.g., músicas, vídeos ou notícias). Para tal, construímos quatro módulos de análise distintos, baseado nas etapas da metodologia de caracterização. A partir deste arcabouço é possível, facilmente, quantificar as principais diferenças práticas entre os distintos SsR baseados em MF. De fato, avaliações em domínios reais sugerem que técnicas baseadas em SVD e PCA recomendam itens distintos e que combiná-las em uma técnica híbrida pode melhorar significativamente os resultados encontrados. Além disso, observamos que características do domínio e individuais de cada usuário podem afetar o desempenho do SVD e PCA em recomendação.

Enfatizamos que todas as implementações, execuções de experimentos e análises de resultados foram realizadas pelo aluno Nícollas Silva, sob a orientação dos professores Fernando Mourão e Leonardo Rocha. A construção da metodologia de caracterização foi conduzida em conjunto, aluno e professores.

2. Trabalhos Relacionados

Métodos de MF oriundos da Álgebra Linear (AL) e da Estatística (ES) têm se tornado cruciais para o sucesso da tarefa de recomendação em variados domínios [Yin and Peng 2012, Sarwar et al. 2000]. Estes métodos apresentam alta eficácia, especialmente, quando combinados a estratégias de Filtragem Colaborativa, uma vez que eles conseguem através da redução de dimensionalidade atenuar o problema da esparsidade comum a domínios de recomendação. Além da recorrente eficácia, a utilização de MF em recomendação permite uma elegante formalização do problema e requer apenas uma simples transformação dos dados para um formato numérico matricial.

Dessa forma, um número crescente de trabalhos vem utilizando métodos de MF baseados em SVD e PCA nas tarefas de recomendação. Por exemplo, em [Yin and Peng 2012], é proposta uma técnica de predição de *ratings* denominada *SVDP* (*SVD prediction algorithm*), que considera os fatores latentes resultantes da aplicação do SVD para relacionar itens a usuários. Em [Sharifi et al. 2013], é apresentado o *Simple SVD*, que implementa o SVD básico na tarefa de predição, o *Normalized SVD*, que normaliza os *ratings* de entrada antes de aplicar a predição, além de algumas variações da modelagem dos dados de entrada, tais como *SimpleSVD*, *MIDItemSVD*, *MIDUserSVD* e *MIDTotalSVD*. As técnicas *Conventional SVD* e *NSVD* são descritas em [Koren 2008]. Neste mesmo trabalho é apresentada uma das técnicas mais eficientes da literatura, o *Assymmetric SVD*, uma extensão da técnica do *SVD++*, proposta em [Bell and Koren 2007]. Ambas são técnicas de vizinhança inspiradas no SVD, que consideram os fatores latentes resultantes da aplicação do SVD relacionados aos usuários e aos itens. Outra técnica inspirada em SVD e de grande relevância na literatura é o *FunkSVD*, proposta em [Funk 2006] e descrita em [Paterrek 2007] como *RSVD* (*Regularized SVD*).

Em outro conjunto de abordagens, como em [Yin and Peng 2012], os autores focam em técnicas híbridas no intuito de se correlacionar as vantagens da aplicação de SVD e PCA com técnicas de Aprendizado de Máquina clássicas, tais como relacionados à tarefa de classificação. Entre elas, destacam-se as técnicas *PCA-KNN* (*Hybrid algorithm based on PCA and KNN*) e *SVD-KNN* (*Hybrid algorithm based on SVD and KNN*), que foram propostas na tentativa de se combinar as vantagens do PCA e SVD com o KNN (*k-Nearest Neighbour*). Outras técnicas híbridas eficientes na tarefa de recomendação que devem ser destacadas são a *Eigentaste*, proposta em [Goldberg et al. 2001], e a *Neighborhood Model*, proposta em [Koren 2010], também inspiradas em métodos de MF.

A grande diversidade de propostas para aplicação de MF em SsR, porém, faz surgir um desafio para a área. Em geral, os estudos não se preocupam em entender as diferenças teóricas e práticas entre as distintas propostas. Pesquisas recentes focam em propor novas estratégias, negligenciando o estudo de domínios e características que tornam as técnicas existentes diferentes. Dessa forma, questões como: *Qual a diferença prática de se aplicar SVD ou PCA em SsR?*, permanecem em aberto, dificultando a consolidação do conhecimento ao longo do tempo. De fato, SVD e PCA são confundidos inclusive em outros domínios além de recomendação, como mostramos no trabalho premiado no CTIC 2015 [Silva et al. 2015]. Este trabalho representa assim uma extensão do CTIC 2015, visando distinguir conceitos usualmente confusos quando aplicamos MF em SsR.

3. Taxonomia para SsR baseados em MF

Nesta seção, apresentamos uma taxonomia para SsR baseados em métodos de MF. Nosso intuito é criar uma classificação para as técnicas mais relevantes da literatura atual, com base em características comuns a estas técnicas. A partir de uma análise não-automática sobre os principais trabalhos da área, identificamos cinco dimensões de classificação comumente associadas a distintos passos do processo de recomendação. Enfatizamos que a taxonomia proposta é não-fechada, uma vez que não engloba todas as características potencialmente relevantes, nem todos os métodos existentes. Porém, tal taxonomia é facilmente extensível a qualquer outra característica ou método. Considerando o escopo deste trabalho, avaliamos cada método com base nas seguintes dimensões:

1. **Tipo de tarefa:** refere-se ao tipo de tarefa de recomendação que está sendo realizada, a qual pode ser a tarefa de **predição**, quando a técnica prediz os valores dos *ratings* para alguns itens, ou mesmo a tarefa **Top-N**, quando a técnica recomenda N itens aos usuários.
2. **Forma de imputação:** analisamos como cada uma das técnicas lida com os *ratings* desconhecidos no conjunto de dados, onde na sua maioria são preenchidos com valores: **zeros**, mediana dos *ratings* dados pelos usuários (**MID Users**), média dos *ratings* recebidos por cada item (**MID Items**), média total dos *ratings* do conjunto (**MID Total**) ou mesmo desconsiderando estes itens não classificados.
3. **Método de fatoração:** refere-se a cada um dos métodos, podendo ser, por exemplo, o SVD, PCA de Covariância ou Correlação. Embora existam outros métodos de fatoração aplicados a recomendação, nossas análises restringem-se apenas à comparação destes três métodos.
4. **Estratégia de recomendação:** compreende a estratégia de exploração do método de fatoração no processo de recomendação. As técnicas que se baseiam nas propriedades algébricas do SVD e PCA, usando-os diretamente para a tarefa de predição de *ratings*, são classificadas como *Conventional Strategy*. Por outro lado, as técnicas que tem como estratégia, adaptar o modelo de predição de *ratings* de SVD e PCA para a aplicação de outros métodos, são classificadas como *Adaptation Strategy*. Por sua vez, as técnicas que utilizam SVD e PCA apenas como modelos para refinar os dados de entrada, no intuito de se aplicar outros métodos de recomendação, são classificadas como *Model Refinement Strategy*.
5. **Fórmula de predição:** representa a função de geração de *score* ou *ratings*, baseada nas funções mais utilizadas na literatura.

Dada a relevância prática dos métodos de Fatoração de Matrizes em domínios de SsR, uma importante contribuição deste trabalho consiste na organização das técnicas mais relevantes da literatura. Basicamente, aplicamos a taxonomia proposta, nas técnicas citadas nos trabalhos relacionados. A tabela 1 sumariza todas nossas análises, classificando cada técnica quanto ao grupo em que a técnica se encaixa, o método que é utilizado, a forma com que o conjunto de dados é tratado, a regra de predição dos *ratings* e o tipo de dados que são considerados. Note que, as técnicas que se encaixam no grupo de *Model Refinement Strategy* não seguem as regras de predição específicas, comuns nos outros modelos, uma vez que SVD e PCA são aplicados apenas para refinar a modelagem dos dados e não para gerar as predições propriamente ditas.

Ressaltamos que, por meio dessa taxonomia, podemos analisar a semelhança entre as técnicas já existentes. As técnicas intituladas *SimpleSVD* e *ConventionalSVD* são semelhantes, enquanto que, a técnica *Asymmetric SVD* é uma extensão da técnica *SVD++*. Da mesma forma, nota-se que o *Normalized SVD* consiste nos mesmos passos do *SVDP*, porém normalizando o conjunto de entrada através da técnica *z-score*. Entender

tais semelhanças é uma importante contribuição, auxiliando pesquisadores da área a compreender melhor conceitos e métodos apresentados de forma confusa na literatura.

Tabela 1. Classificação das técnicas mais relevantes da literatura atual seguindo a taxonomia proposta por este trabalho.

	Técnica	Fatoração	Imputação	Fórmula de predição
Conventional	PureSVD *	SVD	zeros	$r_{ui} = r_u \cdot Q \cdot q_i^T$
	SVDP	PCA-Corr	zeros	$r_{ui} = \overline{b_{ui}} + \sigma_i(Z_d) \cdot u_k$
	Simple SVD	SVD	zeros	$r_{ui} = \overline{b_{ui}} + r_u \cdot Q \cdot q_i^T$
	Conventional SVD	SVD	zeros	$r_{ui} = \overline{b_{ui}} + r_u \cdot Q \cdot q_i^T$
	Normalized SVD	PCA-Corr	zeros (z-score)	$r_{ui} = \overline{b_{ui}} + \sigma_i(Z_d) \cdot u_k$
	MIDItem SVD	SVD	MID Items	$r_{ui} = \overline{b_{ui}} + r_u \cdot Q \cdot q_i^T$
	MIDUser SVD	SVD	MID Users	$r_{ui} = r_u \cdot Q \cdot q_i^T$
	MIDTotal SVD	SVD	MID Total	$r_{ui} = r_u \cdot Q \cdot q_i^T$
NSVD	SVD	ratings existentes	$r_{ui} = \overline{b_{ui}} + q_i^T \left(\sum_{j \in R(u)} r_j \cdot R(u) ^{-\frac{1}{2}} \right)$	
Adaptation	SVD ++	SVD	valores binários	$r_{ui} = \overline{b_{ui}} + q_i^T \left(p_u + \sum_{j \in N(u)} y_j \cdot N(u) ^{-\frac{1}{2}} \right)$
	Asymmetric SVD	SVD	ratings existentes ou não	$r_{ui} = \overline{b_{ui}} + q_i^T \left(\sum_{j \in R(u)} (r_{uj} - b_{uj}) x_j \cdot R(u) ^{-\frac{1}{2}} + \sum_{j \in N(u)} y_j \cdot N(u) ^{-\frac{1}{2}} \right)$
	Funk SVD	SVD	ratings existentes	técnica de regressão linear
Refinement	SVD-KNN	SVD	zeros	-
	PCA-KNN	PCA-Corr	zeros (z-score)	-
	Eigentaste	PCA-Corr	zeros (z-score)	-
	Neighbor Model	SVD	zeros ou somente os itens com ratings	-

* representa técnicas que pertencem a tarefa Top-N, enquanto as demais pertencem a tarefa de predição.

4. Sobre as diferenças práticas entre SsR baseados em MF

Nesta seção, apresentamos uma metodologia e um arcabouço para caracterização do comportamento dos métodos de Fatoração de Matrizes aplicados a recomendação. Por meio deste arcabouço propomos formas de, facilmente, quantificarmos as principais diferenças entre as distintas técnicas baseadas em MF. Consequentemente, visamos entender melhor as características, pontos fortes e fraquezas de cada método, bem como o domínio mais adequado para a aplicação destes.

4.1. Metodologia de Caracterização

De forma a quantificar os impactos da aplicação de métodos de MF sobre a tarefa de recomendação, propomos uma metodologia de caracterização. Nossa metodologia estabelece métricas e procedimentos que nos permitem correlacionar características inerentes às coleções de entrada com indicadores de qualidade na aplicação destes métodos. Um uso prático dessa caracterização seria a identificação de padrões relevantes e recorrentes que nos permitam prever o desempenho dos métodos de MF de acordo com características de cada domínio. Nossa metodologia consiste nos quatro passos descritos abaixo. A tabela 2 apresenta uma sumarização dos indicadores de qualidade avaliados, características selecionadas, bem como as principais métricas de correlação utilizadas.

1. **Seleção e implementação dos métodos de Fatoração de Matrizes:** baseados na nossa organização taxonômica, o primeiro passo consiste em selecionar e implementar técnicas oriundas de categorias distintas de a forma a quantificar o desempenho de cada uma na tarefa de recomendação;
2. **Seleção das características de análise das coleções:** Selecionamos neste passo quatro características comumente relevantes em domínios de recomendação;

3. **Seleção dos indicadores de qualidade:** selecionamos três métricas bem difundidas no domínio de recomendação, capazes de avaliar a qualidade das recomendações geradas, considerando diferentes perspectivas de qualidade;
4. **Análise de correlação entre características e indicadores:** no intuito de se identificar quais características de cada domínio foram relevantes para o desempenho das técnicas implementadas, correlacionamos as características analisadas com os resultados obtidos pelos indicadores de qualidade.

Tabela 2. Características e indicadores de qualidade utilizados na metodologia.

Indicadores	<ol style="list-style-type: none"> 1. Precision: representa a probabilidade que um item selecionado seja relevante e é definida como a razão dos itens relevantes selecionados pelo número de itens recomendados. 2. Recall: definida como a razão entre o número de itens relevantes selecionados pelo número de itens relevantes existentes. <i>Recall</i> representa a probabilidade de um item relevante ser selecionado. 3. Taxa de Acertos: representam o número de acertos do recomendador. Simplesmente verificamos quantos itens da lista Top-N, gerada pelo recomendador, foram de fato <i>consumidos</i> pelos usuários de acordo com as informações retidas no conjunto teste.
Características da Coleção	<ol style="list-style-type: none"> 1. Popularidade dos itens: De maneira simples, para cada item i da coleção, calculamos quantos usuários distintos consumiram-no. Em seguida, ordenamos decrescentemente os valores obtidos construindo uma distribuição de popularidade. Essa distribuição, quando correlacionada às técnicas, permite analisar se os itens recomendados estão relacionados à distribuição <i>long tail</i> ou não. 2. Histórico de consumo dos usuários: Para cada usuário u da coleção, mensuramos a quantidade de itens distintos que este usuário consumiu. Novamente, ordenamos decrescentemente os valores obtidos, agora resultando em uma distribuição de histórico de consumo. Tal distribuição nos permite demonstrar se o número de acertos da técnica está ou não relacionado à assiduidade de cada usuário dentro do conjunto de dados. 3. Nota média de cada usuário: Para cada usuário u da coleção, calculamos a média dos <i>ratings</i> atribuídos por u, simplesmente somando os valores existentes e dividindo pela quantidade de itens avaliados por u. No fim, ordenamos decrescentemente os valores de média, obtendo uma distribuição de média que representa todos os usuários da coleção. Com isso, podemos verificar se o desempenho das técnicas, estão relacionadas às notas atribuídas pelos usuários. 4. Variância média das notas do usuário: Para cada item i da coleção, calculamos a variância dos valores de <i>ratings</i> atribuídos a i por todos os usuários que o consumiram. Posteriormente, ordenamos decrescentemente os valores obtidos, a fim de construir uma distribuição de variância dos <i>ratings</i> da coleção. Com essa distribuição, podemos verificar se o desempenho das técnicas, estão relacionados ou não com a variância das notas atribuídas pelos usuários.
Correlação	<ol style="list-style-type: none"> 1. Correlação de Spearman: teste não-paramétrico, que não exige nenhum pressuposto de distribuição normal, que aplica uma métrica de correlação de <i>rankings</i> ou "postos". 2. Correlação de Kendall-tau: métrica de correlação de <i>rankings</i> mais sensível ao número de observações existentes.

4.2. Ferramenta de Caracterização

Com base na metodologia de caracterização apresentada, consolidamos um arcabouço capaz de realizar o estudo comparativo dos métodos de MF na tarefa de recomendação. Novamente, este trabalho foca especificamente nos métodos de MF denominados SVD e PCA. Motivados pela confusão dos conceitos de SVD e PCA no domínio de recomendação, vamos avaliar as técnicas que se encontram no grupo de *Conventional Strategy*, uma vez que estas técnicas são a base para os outros grupos. Além disso, limitamos nossas análises à tarefa Top-N, dada sua relevância prática em domínios reais. Para tal, selecionamos a técnica de *PureSVD* e implementamos uma adaptação deste algoritmo para a utilização do PCA, sendo chamado aqui de *PurePCA*. O arcabouço proposto é organizado em quatro módulos de análise distintos:

1. **Módulo de Recomendação:** consiste na aplicação das técnicas de MF selecionadas, sobre o conjunto treino gerado com 70% dos usuários de cada coleção, no intuito de gerar uma lista de recomendação para cada usuário;
2. **Módulo de Extração de Características:** extrai as características levantadas na metodologia sobre o conjunto de treino, gerando as distribuições de análise;
3. **Módulo de Avaliação:** responsável por avaliar a qualidade da recomendação, com base no conjunto teste existente, através dos indicadores de qualidade selecionados na metodologia. Para este trabalho, definimos como um item relevante ao usuário, um item que possui nota média superior a nota média dos *ratings* atribuídos por este;
4. **Módulo de Correlacionamento:** responsável por correlacionar as características da coleção com os indicadores de qualidade, através das métricas de correlação selecionadas.

5. Estudo de Caso

Nesta seção, apresentamos um estudo de caso sobre as técnicas de recomendação baseadas em SVD e PCA quando aplicadas a coleções relevantes da literatura. Basicamente, visamos responder as perguntas pertinentes levantadas por este trabalho por meio da aplicação do arcabouço proposto em coleções reais.

5.1. Coleções de Dados

Utilizamos em nossas análises os conjuntos de dados *MovieLens 100k* e *MovieLens 1M*¹ que foram consolidados e tornados públicos pelo grupo de pesquisa *GroupLens*. A tabela 3 sumariza as principais informações sobre cada coleção. Ambos são conjuntos de dados explícitos sobre recomendação de filmes, em que cada usuário assinala uma avaliação (*ratings*) entre 1 e 5 a cada filme.

Tabela 3. Informações sobre as bases de dados utilizadas.

Base	Usuários	Itens	Esparsidade
<i>MovieLens 100k</i>	943	1676	93,67%
<i>MovieLens 1M</i>	6040	3952	95,82%

5.2. Análise de Resultados

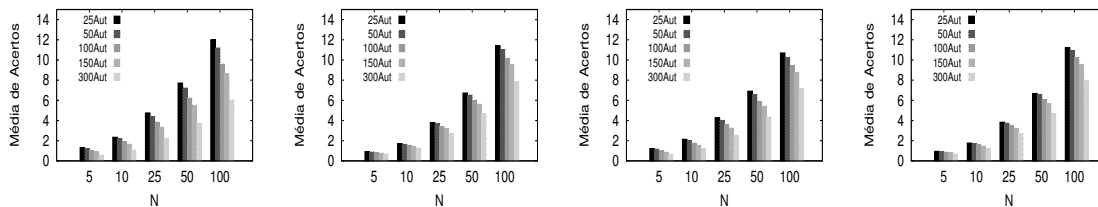
Motivados a caracterizar as distinções entre SVD e PCA quando aplicados em técnicas de recomendação, consolidamos um projeto de análise experimental que está relacionado aos Módulos de Avaliação e Correlacionamento, propostos pelo nosso arcabouço. Este projeto experimental é dividido em quatro etapas: (1) análise da extração de características das coleções de dados, onde avaliamos as distribuições de ranking geradas; (2) análise individual do comportamento de cada um dos métodos, onde nos preocupamos em calcular a taxa de acertos e os valores de *precision* e *recall* na aplicação destes; (3) análise pareada dos métodos, onde contrastamos os itens recomendados por cada um dos métodos; e (4) aplicação de um projeto fatorial 2^k , que nos permite verificar os fatores mais importantes que podem afetar o desempenho de métodos baseados em SVD e PCA.

Na primeira etapa de nosso projeto experimental, após extrair as características de cada coleção e plotar as distribuições de ranking, verificamos que o consumo dos usuários está relacionado a popularidade dos filmes em ambas coleções. Notamos também que os

¹Disponíveis em: <http://grouplens.org/datasets/>

filmes mais populares são também aqueles que possuem muitas opiniões distintas, sugerindo que a aplicação de PCA tende a obter melhor resultado nestes cenários, visto que os autovetores do PCA priorizam as dimensões com maiores variâncias nos dados. Em contrapartida, notamos que os filmes com notas médias maiores são os mais relevantes em ambas coleções, o que nos permite dizer, que a aplicação de SVD é mais recomendada, visto que os autovetores de SVD tendem a apontar para dimensões de maior média.

Na segunda etapa, aplicamos os indicadores de qualidade levantados em cada coleção e para cada método, a fim de se caracterizar o desempenho do SVD e PCA de maneira individual. Nesta análise, variamos o número de autovetores relativos ao processo de decomposição de ambos considerando os valores [25, 50, 100, 150, 300]. De maneira análoga, variamos a quantidade de itens recomendadas na tarefa Top-N em [5, 10, 25, 50, 100]. Com base na taxa de acertos e nos valores de *precision* e *recall* calculados para estas configurações, conforme mostra a figura 1 observamos que utilizar apenas 25 autovetores eleva a taxa de acertos para ambas as técnicas, e portanto, será o parâmetro que utilizaremos em nossas análises seguintes.



(a) PureSVD - ML-100k (b) PureSVD - ML-1M (c) PurePCA - ML-100k (d) PurePCA - ML-1M

Figura 1. Influência dos parâmetros no comportamento de cada método.

Na etapa de análise pareada dos métodos, ao considerar os indicadores de qualidade resultantes do processo de recomendação Top-N, identificamos uma similaridade entre o desempenho do *PureSVD* e *PurePCA*. Sabe-se que existe uma diferença prática na aplicação dos dois métodos, porém é necessário mensurar as diferenças entre ambos métodos na tarefa de recomendação. Dessa forma, analisando as distribuições de similaridade entre as listas de recomendação geradas por cada método para todos os usuários, observamos que de 30% a 60% dessas recomendações são idênticas para a maioria dos usuários. Com isso, ainda temos uma porcentagem significativa (i.e., $\geq 40\%$) de recomendações que são geradas por apenas um dos métodos. De maneira complementar, verificamos se a taxa média de acertos relativa aos métodos de *PureSVD* e *PurePCA* estão relacionadas aos itens recomendados por ambos, ou se estão relacionadas aos itens diferentes recomendados por cada técnica. Neste intuito, dividimos cada lista de recomendação gerada em três grupos e comparamos a média de acertos de cada um deles, com base na média de acertos existente sobre toda a lista de recomendação.

- **Grupo 1:** grupo de itens idênticos em ambas as listas de recomendações;
- **Grupo 2:** grupo de itens existentes apenas na lista gerada por *PureSVD*;
- **Grupo 3:** grupo de itens existentes apenas na lista gerada por *PurePCA*.

A tabela 4 demonstra os resultados encontrados para cada um dos grupos, evidenciando que a média de acertos não está relacionada apenas aos itens recomendados por ambos, mas também aos itens recomendados apenas por uma das técnicas. Observe que estes itens recomendados apenas por um método (grupos 2 e 3) apresentam uma

média de acerto significativa, quando comparada à média de acertos total de cada uma das técnicas. Em outras palavras, não só existe diferenças ao aplicar SVD e PCA como base para técnicas de recomendação, mas também, combiná-los em uma técnica híbrida, pode trazer resultados promissores para este contexto de análise.

Tabela 4. Média de acertos para cada um dos grupos de análise.

Conjunto	Grupo 1	Grupo 2	Grupo 3	PureSVD	PurePCA
<i>MovieLens 100k</i>	1,0858	1,2535	1,0520	2,3393	2,1378
<i>MovieLens 1M</i>	0,8079	0,8919	0,9551	1,6998	1,7630

Por fim, na tentativa de se identificar quais fatores influenciam o sucesso de uso do SVD e PCA em recomendação, avaliamos a correlação existente entre os indicadores de qualidade e as características das coleções utilizando um projeto fatorial 2^k . Utilizamos as métricas de correlação propostas, capazes de enfatizar melhor a relação existente entre os fatores de análise. Especificamente, correlacionamos os indicadores de *precision* e *recall* com cada característica extraída, conforme proposto pelo Módulo de Correlação do nosso arcabouço. Esta análise nos permitiu dizer que a taxa de acertos está relacionada ao consumo dos usuários, devido ao aprendizado supervisionado dos métodos. Em uma análise mais detalhada, onde analisamos a popularidade de cada item da lista de recomendação gerada por *PureSVD* e *PurePCA*, notamos que apesar de pequena diferença de correlação entre média de popularidade e os valores de *recall*, para *PureSVD* e *PurePCA* (i.e., como enfatizado pelo tom de cinza mais claro), quando analisado qual dos métodos recomenda itens mais populares, notamos uma diferença entre eles. Nestes cenários, a técnica de *PurePCA* tende a recomendar itens mais populares que *PureSVD*, uma vez que itens mais populares apresentam uma maior variância das notas, o que influencia na aplicação do PCA.

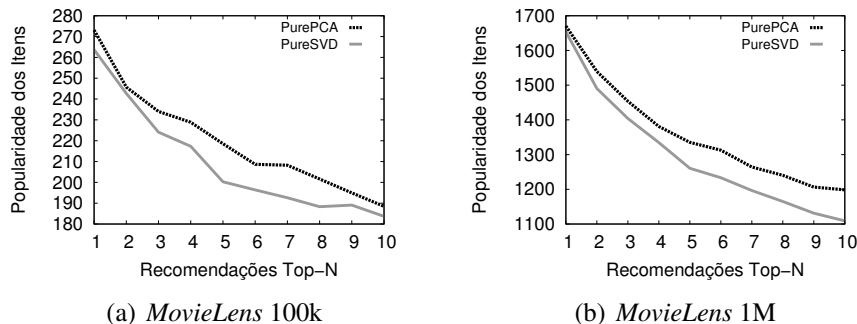


Figura 2. Popularidade média por posição no Top-N.

6. Conclusões e trabalhos futuros

Neste trabalho, revisamos, caracterizamos, organizamos e comparamos distintos métodos de recomendação baseadas em MF. Para tanto, primeiramente, propomos uma taxonomia para métodos de MF baseada em cinco dimensões associadas aos passos do processo de recomendação. Em seguida, classificamos 16 técnicas mais relevantes da literatura através dessa taxonomia, no intuito de auxiliar pesquisadores da área a entender conceitos e semelhanças entre elas. Com esta classificação fomos capazes de analisar as semelhanças entre as técnicas já existentes, bem como, entender melhor as modelagens existentes na aplicação de métodos de fatoração de matrizes. Posteriormente, apresentamos uma metodologia de caracterização e a consolidamos em um arcabouço capaz de

quantificar o comportamento e as diferenças de SVD e PCA quando aplicados à tarefa de recomendação. Avaliações em domínios reais, a partir do arcabouço proposto, sugerem que técnicas baseadas em SVD e PCA recomendam itens distintos e que combiná-las em uma técnica híbrida pode melhorar significativamente os resultados encontrados. Além disso, observamos que características do domínio e individuais de cada usuário podem afetar o desempenho do SVD e PCA em recomendação. Dessa forma, a definição de um projeto experimental apropriado, que nos permitirá validar cada padrão encontrado, é uma direção de pesquisa futura. A avaliação de outros métodos de MF, por meio de nosso arcabouço, representa outra importante direção de pesquisa.

Referências

- Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749.
- Bell, R. M. and Koren, Y. (2007). Lessons from the netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79.
- Funk, S. (2006). Netflix update: Try this at home (december 2006). *URL <http://sifter.org/~simon/journal/20061211.html>*.
- Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4(2):133–151.
- Koren, Y. (2008). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM.
- Koren, Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 4(1):1.
- Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, (8):30–37.
- Paterek, A. (2007). Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of KDD cup and workshop*, volume 2007, pages 5–8.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2000). Application of dimensionality reduction in recommender system—a case study. Technical report, DTIC Document.
- Sharifi, Z., Rezghi, M., and Nasiri, M. (2013). New algorithm for recommender systems based on singular value decomposition method. In *Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference on*, pages 86–91. IEEE.
- Silva, N., Neves, A., Rocha, R. C., and Mourão, F. (2015). Sobre as diferenças de aplicação entre svd e pca: Um estudo pragmático. In *Anais do XXXV Congresso da Sociedade Brasileira de Computação*.
- Yin, C.-X. and Peng, Q.-K. (2012). A careful assessment of recommendation algorithms related to dimension reduction techniques. *Knowledge-Based Systems*, 27:407–423.

D-STHARK: Avaliando Escalonadores Dinâmicos de Tarefas em Arquiteturas Híbridas Simuladas*

**Sávyo Toledo , Danilo Melo , Guilherme Andrade ,
Fernando Mourão , Leonardo Rocha**

DCOMP/UFSJ - São João del-Rei, MG , Brasil

{savyomachado,daniloamaral,gandrade,fhmourao,lcrocha}@uufs.j.edu.br

Abstract. *Aiming to exploit better new computing architectures, with several Processing Units (PUs), such as CPUs, GPUs and Intel Xeon Phi, recent works focus on proposing novel runtime environments that offer a variety of methods for scheduling tasks dynamically on different PUs. A main limitation of such proposals refers to the constrained system configurations to evaluate them. In this work we present D-STHARK (Dynamic Scheduling of Tasks in Hybrid Simulated ARchitectures), a tool that provides a complete simulated execution environment that allows evaluating dynamic scheduling strategies on simulated applications and hybrid architectures. Adopting D-STHARK, we simulated some experiments presented in recent works, achieving the same conclusions.*

Resumo. *Para melhor explorar novas arquiteturas computacionais, composta por diversas Unidades de Processamento (UPs), tais como CPUs, GPUs e Intel Xeon Phi, estudos recentes focam em propor ambientes de execução que oferecem métodos de escalonamento dinâmico para diferentes UPs. A principal limitação dessas propostas são as restrições de configuração de sistema para avaliá-las. Nesse trabalho apresentamos o D-STHARK, (Dynamic Scheduling of Tasks in Hybrid Simulated ARchitectures), uma ferramenta que provê um ambiente completo de simulação de arquiteturas híbridas para avaliação de estratégias de escalonamento dinâmico. Simulamos com o D-STHARK experimentos apresentados em trabalhos recentes alcançando as mesmas conclusões.*

1. Introdução

O crescimento contínuo de dados, associados a processamentos mais sofisticados em diferentes áreas de conhecimento, tem impulsionado avanços significantes nas arquiteturas de computação, refletindo em sistemas de armazenamentos mais eficientes, como também o uso de diferentes tipos de unidades de processamento (UPs), denominadas de arquiteturas híbridas. Um exemplo dessas novas arquiteturas são computadores com múltiplos processadores (arquiteturas multicore) e diferentes coprocessadores, tais como unidades de processamento gráfico (Graphic Processing Unit - GPUs) [Fatica and Luebke 2007] e o Intel Xeon Phi (MIC) [Jeffers and Reinders 2013]. Dessa forma, tem se tornado essencial que aplicações oriundas de diferentes domínios sejam capazes de explorar, de forma coordenada e eficiente, todas as UPs disponíveis, aproveitando ao máximo as suas capacidades de processamento.

*Esse trabalho foi parcialmente financiado por CNPq, CAPES, FINER, Fapemig, e INWEB.

Observamos na literatura várias propostas de ambientes de execução com o objetivo de tornar mais transparente o uso de diferentes UPs para os desenvolvedores [Augonnet et al. 2010]. Entre os principais métodos fornecidos por estes ambientes, destacam-se os escalonadores de tarefas, responsáveis pela distribuição adequada das várias tarefas que compõem uma aplicação entre as UPs. Escalonadores podem ser: (1) estáticos, em que as características das tarefas e as capacidades das UPS são avaliadas em um estágio de pré-processamento, usando informação globais da aplicação; e (2) dinâmicos, em que a avaliação é feita em tempo de execução, considerando uma visão limitada da execução da aplicação, sendo portanto, um cenário mais desafiador.

Embora existam diferentes propostas de escalonadores dinâmicos na literatura, a maioria deles é avaliada em configurações de sistemas reais restritas, compostos por um número reduzido de UPs (i.e, alguns núcleos de CPU, uma ou duas GPUs e / ou um ou dois MICs) devido a elevados custos associados à criação de ambientes mais completos. Por exemplo, o preço de um coprocessador Intel Xeon Phi 7120P Henhexaconta-Core é aproximadamente R\$20.000,00. Como consequência, as conclusões obtidas pela avaliação da escalabilidade destas propostas podem ser limitadas. Neste contexto, um ambiente simulado em que seja possível configurar diferentes arquiteturas híbridas, composto por um número ilimitado de UPs, a fim de avaliar estratégias de escalonamento dinâmico, torna-se uma contribuição importante, sendo esse o foco do trabalho.

Mais especificamente, neste artigo, apresentamos *D-STHARK* (**D**ynamic **S**cheduling of **T**asks in **H**ybrid **S**imulated **A**Rchitectures), cujo objetivo é fornecer um ambiente de execução simulado completo que permite avaliar estratégias de escalonamento dinâmico de tarefas em aplicações simuladas. Além disso, o *D-STHARK* permite ao usuário simular arquiteturas híbridas, variando os tipos e número de UPs (CPUs, GPUs e MICs), bem como simular diferentes aplicações, compostas por diversas tarefas de características distintas (e.g., dependências entre tarefas, volume de dados a ser manipulado, etc.). Além de fornecer diferentes estratégias de escalonamento dinâmico, [Andrade et al. 2014b, Andrade et al. 2014a], o *D-STHARK* permite que novas estratégias sejam adicionadas por meio de uma API. Ressaltamos ainda que não encontramos na literatura propostas de simuladores com objetivos similares aos nossos.

No intuito de avaliar nossa ferramenta, simulamos a execução de uma aplicação de análise de imagens utilizada em estudos da morfologia de câncer no cérebro [Andrade et al. 2014a], considerando algumas estratégias de escalonamento dinâmico propostas em [Andrade et al. 2014b] e avaliando diferentes configurações de arquitetura. Em um primeiro momento, demonstramos que o *D-STHARK* foi capaz de apresentar os mesmos resultados encontrados nos trabalhos originais, demonstrando assim sua efetividade. Apresentamos também uma análise complementar das estratégias de escalonamento usando configurações de arquitetura não avaliadas no trabalho original, evidenciando a utilidade do *D-STHARK* em fornecer análises mais amplas.

A implementação da ferramenta e execuções de experimentos foram realizadas pelo aluno Sávyo Toledo, sob a orientação do professor Leonardo Rocha. As análises dos resultados foram feitas com a colaboração do professor Fernando Mourão. Esse trabalho também contou com a colaboração do aluno Danilo Melo na inserção das estratégias de escalonamento, bem como do aluno de pós-graduação Guilherme Andrade na compreensão e adaptação da aplicação avaliada.

2. Trabalhos Relacionados

O uso eficiente de sistemas híbridos equipados com CPUs e aceleradores é um problema desafiador que exige a implementação de aplicações otimizadas para múltiplos processadores e escalonamento de tarefas entre dispositivos heterogêneos. Recentemente, técnicas de compilação [Ravi et al. 2014], bibliotecas de domínio específico [Bradski 2000], e, sobretudo, ambientes de execução [Rossbach et al. 2011, Augonnet et al. 2010, Bueno et al. 2012, Ravi et al. 2014] têm sido propostos para reduzir o esforço de programação relacionados à portabilidade de aplicações para esses sistemas.

Plataformas de execução distribuída entre CPU, GPU e MIC têm sido alvo de vários projetos recentes [Augonnet et al. 2010, Bueno et al. 2012, Ravi et al. 2014, Lima et al. 2013]. Em [Ravi et al. 2014], é proposto um tradutor de operações genéricas de redução, baseado em técnicas de compilação, para sistemas híbridos CPU-GPU. Em [Bueno et al. 2012], os autores propõem o OmpSs, um modelo de programação paralela para aplicações de fluxo-dados que permite paralelização de códigos por meio do compilador a partir de anotações feitas no código pelo programador. Em [Augonnet et al. 2010], é apresentado o StarPU, um ambiente em que as tarefas de uma aplicação são modeladas por meio de Grafo Direcionado Acíclico. XKaapi [Lima et al. 2013] é outro ambiente de execução que possibilita a execução cooperativa em CPU-GPU-MIC, na qual as operações possuem diversas implementações focadas em diferentes dispositivos de computação.

Esforços recentes em ambientes de execução têm dado atenção especial à exploração dos chamados escalonadores dinâmicos que distribuem, em tempo de execução, as tarefas de uma dada aplicação entre as diferentes UPs disponíveis. Basicamente, estes escalonadores realizam uma avaliação em tempo de execução das características de cada tarefa e UP. Em seguida, com base nessa avaliação, determinam a UP mais adequada para executar cada tarefa. O desafio, nesse caso, é a forma de maximizar as oportunidades de paralelização com base apenas em um conhecimento local e limitado do conjunto de tarefas que compõem a aplicação. Escalonadores devem evitar eventos que comprometam a distribuição adequada de tarefas, tais como a sobrecarga de uma UP, escolha de um UP que não seja a mais adequada para executar uma determinada tarefa, e até mesmo transferências excessivas de dados entre memórias não compartilhadas de UPs distintas. Existem várias propostas de escalonadores dinâmicos na literatura [Andrade et al. 2014b]. A principal limitação dessas propostas refere-se às restrições quanto a configurações de sistemas híbridos reais para avaliação e ajustes dessas estratégias. Em razão do custo de se criar ambientes mais completos e diversificados, grande parte destas avaliações é feita considerando-se apenas uma arquitetura em particular.

3. D-STHARK

Nesta seção, apresentamos *D-STHARK* dividindo a descrição em três partes. Primeiro, apresentamos seus principais componentes. Em seguida, descrevemos como esses componentes funcionam e interagem uns com outros para executar uma simulação. Por fim, apresentamos a API que permite aos usuários definir e inserir distintas estratégias de escalonamento. O *D-STHARK* está disponível em <https://github.com/SavyoToledo/D-STHARK>.

3.1. Componentes do *D-STHARK*

Basicamente, *D-STHARK* é composto de quatro componentes principais: (1) configuração ambiente; (2) criação de tarefas; (3) definição da estratégia de escalonamento; (4)

execução experimental e avaliação. Os usuários podem ajustar a simulação de acordo com seus objetivos, avaliando cada cenário em particular mais próximo de condições reais.

1. Configurações do Ambiente. *D-STHARK* permite aos usuários criarem arquiteturas híbridas usando três tipos de UPs: CPU, GPU e MIC. Cada uma destas pode ser instanciada diversas vezes. Por exemplo, podemos criar um ambiente com 6 CPUs, 2 GPUs e 2 MICs. Além disso, para cada coprocessador (i.e., GPU e MIC) é possível definir a largura da banda de transferência de dados, que corresponde à quantidade de dados que pode ser transferida entre o coprocessador e a CPU a cada unidade de tempo. Esta configuração torna as simulações ainda mais próximas de cenários reais. Além disso, os usuários podem salvar essas configurações para usar em outras simulações.

2. Criação de tarefas. *D-STHARK* simula cada aplicação como um conjunto de tarefas distintas e suas dependências. Por sua vez, cada tarefa é definida por três características: (1) tipo de tarefa, (2) taxa de erro (3) e tamanho da carga de trabalho. O speedup de uma UP pode variar bastante de acordo com as operações consideradas. Além disso, o desempenho relativo entre as UPs varia de acordo com a operação executadas. Consequentemente, certas UPs serão mais eficientes para determinados tipos de operações. Sendo assim, (1) *Tipo de Tarefa* representa o comportamento que um grupo de tarefas semelhantes geralmente apresenta em termos de tempo de execução para cada tipo de UP. Todas as tarefas que pertencem a um mesmo tipo terão um tempo de execução específico para cada UP, o qual é informado pelo usuário. A fim de permitir que a simulação seja o mais real possível, os usuários podem definir a *Taxa de Erro* (2), um intervalo (i.e., o valor mínimo e máximo de erro) que corresponde ao quanto se espera que o tempo de execução de uma tarefa pode variar. Depois de definir essas duas características, o próximo passo é criar as tarefas propriamente ditas. *D-STHARK* permite aos usuários inserir todas as tarefas que compõem a aplicação principal a ser simulada, onde cada tarefa deve pertencer a um dos tipos de tarefas acima referidos. De acordo com o tipo de tarefa e a taxa de erro, um tempo aleatório é definido para cada tarefa para as diferentes UPs. Finalmente, a terceira característica exigida pelo *D-STHARK* é o *Tamanho da Carga de Trabalho* (3) em MB. Essa característica é relevante, uma vez que muitas estratégias de escalonamento consideram o tamanho dos dados e o número de operações a serem executadas em cada UP para estimar o custo de comunicação. Por questões de simplicidade, consideramos que o custo de transferência varia linearmente com o tamanho dos dados. A fim de identificar cada tarefa criada, os usuários podem atribuir um identificador único (ID) para cada tarefa no *D-STHARK*. Usando esses IDs, os usuários podem designar dependências entre tarefas distintas que compõem uma aplicação, como ocorre em cenários reais. Por exemplo, a dependência de uma tarefa t_1 e uma t_2 significa que t_2 deve ser executado antes de t_1 . Mais uma vez, os usuários podem salvar todas essas configurações.

3. Definição das Estratégias de Escalonamento. Atualmente, a *D-STHARK* tem quatro estratégias implementadas: (1) FCFS (*First Come First Served*); (2) HEFT (*Heterogeneous Earliest Finish Time*); e (3) HEFT-DA (*Heterogeneous Earliest Finish Time Data-Aware*); e (4) SEQ (Sequencial), uma serialização simples de todas as tarefas a serem executadas. Enquanto a primeira estratégia é muito tradicional em trabalhos relacionados a escalonadores, a segunda e terceira estratégias foram propostas e extensivamente avaliadas em [Andrade et al. 2014a]. A última estratégia é utilizada como uma

linha de base para contrastar a diferença de desempenho das estratégias. Os usuários podem selecionar diferentes escalonadores para compor uma única simulação e, neste caso, *D-STHARK* executa individualmente cada escalonador selecionado, permitindo aos usuários comparar a performance alcançada por eles.

4. Execução Experimental e Avaliação. Nesta etapa, os usuários analisam e confirmam todas as configurações. Em seguida o *D-STHARK* começa a executar a simulação. A ferramenta também permite definir quantas vezes cada simulação será executada. Os resultados são relatados como a média destas execuções. O desvio padrão também é apresentado. Durante a execução de simulação, *D-STHARK* exibe uma visualização do *log* de execução detalhando a simulação e mostrando qual tarefa está sendo executada por cada UP em cada momento. Esta visualização de *log* pode ser usada para depurar os escalonadores, verificando se eles estão funcionando como esperado. Ao final da simulação, *D-STHARK* apresenta os resultados detalhados para cada escalonador simulado, o quais podem ser exportados para um arquivo de saída. Mais especificamente, nossa ferramenta mostra (1) Speedup alcançados; (2) histograma com a distribuição de tarefas entre as UPs; e (3) porcentagem de processamento realizado por cada UP.

3.2. Processo de Execução de Simulações

Para realizar a simulação propriamente dita, o *D-STHARK* cria uma thread¹ distinta, nomeada **Worker Thread**, para simular cada UP. Além disso, uma thread *principal* é instanciada para ler as configurações da tarefa (rotina *GetTask*) e enviá-la para execução (rotina *SubmitTask*). Por fim, a thread principal cria uma fila para cada **Worker Thread** que será gerida de acordo com a estratégia de escalonamento definida na configuração.

O gerenciamento de tarefas é executado simultaneamente com as execuções de tarefas, como em ambientes de escalonamento dinâmicos reais. Depois de receber uma tarefa, cada **Worker Thread** verifica suas dependências. Se a mesma possuir dependências, será inserida na **Stuck Task List**, que armazena todas as tarefas que ainda não tiveram suas dependências executadas. Caso contrário, será enviada para o método **PushTask** que insere a tarefa em uma das filas, de acordo com a política de escalonamento.

Quando uma **Worker Thread** específica fica ociosa, ela busca uma nova tarefa a ser executada utilizando o método *PopTask*, definida de acordo com cada estratégia de escalonamento dinâmico. Como mencionado acima, cada tarefa tem um tempo de execução específico definido para cada UP. Assim, por meio de um *sleep* o **Worker Thread** simula o tempo de execução na UP correspondente. Quando uma tarefa simula o tratamento de dados, uma segunda chamada ao método *sleep* é feita para simular o tempo de transferência de dados da memória do CPU para a memória do coprocessador. Esse tempo é calculado de acordo com a largura da banda definida, bem como o tamanho da carga de trabalho. Por exemplo, a PCI-Express 16x bus oferece uma velocidade de transferência de 4GB por segundo. Assim, para cada MegaByte de dados da tarefa, realizamos um *sleep* de 0.000244141s, aproximando-se das condições reais de execução. Finalizando o processo de execução, a **Worker Thread** verifica a **Stuck Task List** e remove todas as dependências associadas à tarefa que acabou de ser executada. Em seguida, a **thread principal** começa a buscar tarefas da **Stuck Task List** com todas as

¹Implementamos as threads usando a biblioteca pThread padrão.

suas dependências resolvidas. Desta forma, garantimos que as tarefas serão executadas na ordem correta. A Figura 1 ilustra o processo acima descrito.

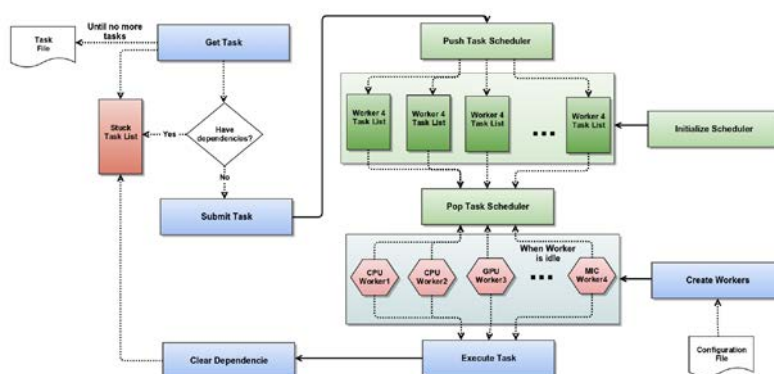


Figura 1. Processo de Execução de Simulações

3.3. API do Escalonador Dinâmico

O *D-STHARK* permite que os usuários incluam as suas próprias estratégias de escalonamento. Para este propósito, nossa ferramenta oferece uma API que especifica uma diretriz a ser seguida a fim de tornar essas novas estratégias compatíveis com o sistema. Basicamente, todos os novos escalonadores devem ser implementados na linguagem C. Essas implementações podem usar diferentes rotinas e nomes de arquivos, que devem ser carregados no *D-STHARK* através da GUI disponível. No entanto, o código-fonte deve ter um arquivo nomeado *scheduler.c*, nomeado, com quatro rotinas obrigatórias:

1. **InitializeScheduler**: responsável por iniciar componentes do escalonador. Ela também cria as filas de tarefas, associando-as com a respectiva UP.
2. **PushTask**: define qual estratégia será usada para escalonar as tarefas.
3. **PopTask**: esta rotina é chamada por cada **Worker Thread**, sempre que esta estiver ociosa, definindo a próxima tarefa a ser executada.
4. **DestroyScheduler**: responsável por finalizar o escalonador. Nesta rotina a memória usada pelo escalonador deve ser liberada.

4. Avaliação Experimental

Nesta seção, apresentamos os experimentos realizados para avaliar *D-STHARK*, que consistem em simular o mesmo cenário originalmente apresentado em [Andrade et al. 2014a], em que os autores avaliaram diferentes estratégias de escalonamento dinâmico para uma aplicação real [Teodoro et al. 2014]. Nossas avaliações contrastam os resultados simulados contra os originais, considerando ambos os tempos de execução e distribuição de tarefas entre UPs. Além disso, simulamos também o uso de mais coprocessadores do que inicialmente relatado para avaliar a utilidade do *D-STHARK* em fornecer análises mais amplas.

4.1. Configuração da simulação

4.1.1. Aplicação da Simulação

Em nossos experimentos, simulamos uma aplicação relacionada a estudos de câncer no cérebro [Teodoro et al. 2014], que visa encontrar melhores classificações para os tumores

usando *Whole Tissue Slide Images (WSIs)*. Esta aplicação também particiona cada WSI em vários pedaços *tiles* de imagem que podem ser analisados de forma independente. Há muitas fases nesta análise de imagens, mas as que demandam a maior tempo de execução são a de segmentação (*segmentation*) e computação de características (*features computation*). Estas duas fases têm sido o foco de otimização de execução em máquinas híbridas em grande escala [Andrade et al. 2014a]. Cada pedaço de imagem é submetido a diferentes operações, formando um fluxo de execução como mostrado na Figura 2.

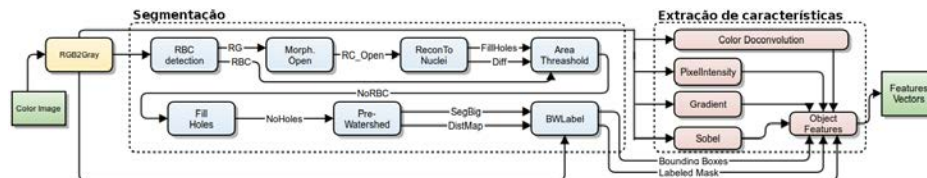


Figura 2. Fluxo de Execução da Aplicação

No *D-STHARK*, instanciamos um **tipo de tarefa** para cada operação, representados como retângulos coloridos na figura acima. O desempenho relativo (i.e., tempo específico) para cada UP em cada tipo de tarefa foi definido com base nos valores originalmente reportados em [Teodoro et al. 2014]. A taxa de erro foi configurada como zero, seguindo os experimentos originais [Andrade et al. 2014a]. O processamento de vários pedaços de imagem é um problema embaraçosamente paralelo, uma vez que cada um deles determina um fluxo de execução diferente. Assim, instanciamos uma tarefa individual para cada operação distinta que é aplicada em cada pedaço de imagem. Além disso, inserimos as dependências entre tarefas, como mostrado no fluxo de execução, para simular a ordem correta. Para ilustrar esse processo, considere as operações *Morph.Open* e *Recon.Nuclei* aplicadas a dois pedaços de imagem. Primeiro, criamos dois tipos de tarefa, *Tipo1* (i.e., *Morph.Open*) e *Tipo2* (i.e., *Recon.Nuclei*). Então, *D-STHARK* determina automaticamente os tempos de execução nas UPs para cada tipo de tarefa, com base nas configurações previamente definidas. Em seguida, são instanciadas as tarefas relacionadas a cada pedaço da imagem: (a) t_1 do *Tipo1* para o primeiro pedaço; (b) t_2 do *Tipo2* para o primeiro pedaço; (c) t_3 do *Tipo1* para o segundo pedaço; e (d) t_4 do *Tipo2* para o primeiro pedaço. Por fim, adicionamos as dependências de t_2 para t_1 e de t_4 para t_3 , garantindo assim a ordem correta da execução.

Simulamos 800 pedaços imagem, o que gera 10.400 tarefas para execução. Além disso, no artigo original, os autores relataram que cada pedaço apresenta uma resolução de $4K \times 4K$ pixels. Considerando uma representação de 256 cores, cada pedaço de imagem tem um tamanho de 15, $25MB$, que foi a carga de trabalho instanciada para cada tarefa.

4.1.2. Escalonadores Simulados

As características da aplicação discutida acima induzem comportamentos bastante semelhantes em duas estratégias (i.e., HEFT e HEFT-DA). Assim restringimos nossa análise a duas estratégias avaliadas em [Andrade et al. 2014a]:

- **FCFS:** é uma estratégia de escalonamento baseada em uma fila global. Quando uma **Worker Thread** está ociosa, ela busca a primeira tarefa na inicio da fila.
- **HEFT:** Mantém uma fila de tarefas para cada UP. A distribuição de tarefas entre as UPs é definida de acordo com as suas capacidades de processamento. Mais

especificamente, o escalonador mantém um histórico dos tempos de execução e, assim, atribui uma tarefa a uma UP específica que minimize a Equação 1.

$$\min_{P_i}(Avail(P_i) + Est_{P_i}(T)) \quad (1)$$

P_i é a UP a ser avaliada; $Avail(P_i)$ representa a quantidade de tempo que leva para P_i processar todas as tarefas que lhe são atribuídas; e Est_{P_i} denota o tempo estimado que P_i leva para executar uma tarefa específica T .

4.1.3. Arquiteturas Simuladas

Basicamente, nossas análises consideram três arquiteturas híbridas distintas [Andrade et al. 2014a]: (1) **CPU-GPU**: 15 núcleos de CPU e 1 GPU; (2) **CPU-MIC**: 15 núcleos de CPU e 1 MIC; e (3) **CPU-GPU-MIC**: 14 núcleos de CPU, 1 GPU, e 1 MIC. Avaliamos todos os escalonadores simulados em cada uma dessas arquiteturas. Além disso, simulamos outros cenários, considerando mais coprocessadores do que relatado em [Andrade et al. 2014a], no intuito de evidenciar a utilidade do *D-STHARK* em fornecer análises mais amplas. Em todas essas análises, definimos a largura de banda como *4GB*, como uma PCI-Express 16x usada nos experimentos originais.

4.2. Análise dos Resultados

Primeiramente, analisamos o quão semelhante foi a distribuição de tarefas definida por cada estratégia de escalonamento na arquitetura simulada com as distribuições observadas na arquitetura real. Em seguida, investigamos o quão similar estão os tempos de execução de nossas simulações para os tempos medidos em uma arquitetura real. Para estas duas questões, consideramos a arquitetura de **CPU-GPU-MIC**, como avaliado em [Andrade et al. 2014a]. Finalmente, estendemos a avaliação original para outros tipos de arquiteturas, a fim de chegar a conclusões mais amplas. Ressaltamos também que todos os tempos discutidos abaixo representam a média de 10 execuções distintas.

4.2.1. Distribuição de Tarefas

Em [Andrade et al. 2014a], os autores avaliaram o impacto do número de pedaços de imagem distintas simultaneamente processadas por cada *Worker Thread*, concluindo que o desempenho geral pode ser melhorado utilizando altos níveis de concorrência (ou seja, 55 pedaços de imagem). Em nossas análises, os escalonadores simulados funcionaram sem limitações sobre este número, apresentando os resultados encontrados pelos autores. A Figura 3 apresenta os resultados do *D-STHARK*. Em primeiro lugar, observa-se que o escalonador FCFS para as UPs é quase o mesmo para todas as operações, demonstrando que o FCFS não é capaz de tirar pleno proveito sobre a variabilidade de desempenho entre as operações. Por outro lado, podemos notar que HEFT tem priorizado o uso de núcleos de CPU para a maioria das tarefas, independentemente do desempenho das outras UPs. Conforme apresentado em [Teodoro et al. 2014], algumas operações (e.g., *PreWaterShad*, *ReconNuclei* e *RBC*), com elevados custos computacionais, podem diminuir o tempo de execução usando coprocessadores. Na verdade, HEFT foi capaz de identificar essas características, reduzindo o tempo total de execução. Uma discussão mais detalhada destes resultados pode ser encontrada no trabalho original [Andrade et al. 2014a].

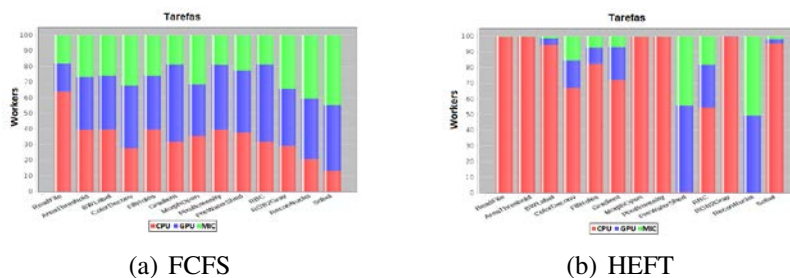


Figura 3. Distribuição de tarefas executadas na Simulação entre as UPs.

4.2.2. Tempos de Execução

Os tempos de execução dos escalonadores são apresentados na Figura 4 (a) e (b), que correspondem aos resultados simulados e reais, respectivamente. Mais uma vez, os resultados obtidos utilizando *D-STHARK* são bastante semelhantes aos obtidos usando uma arquitetura real sobre a qual FCFS atinge o pior desempenho entre todas as configurações. Como relatado em [Andrade et al. 2014a], é importante salientar que a utilização da GPU produz sempre um bom desempenho. Por exemplo, o melhor tempo de execução CPU-GPU é de cerca de 1,26× mais rápido do que o melhor tempo de execução CPU-MIC. Este resultado nos motiva a realizar os experimentos descritos a seguir, a fim de demonstrar a aplicabilidade da nossa ferramenta.

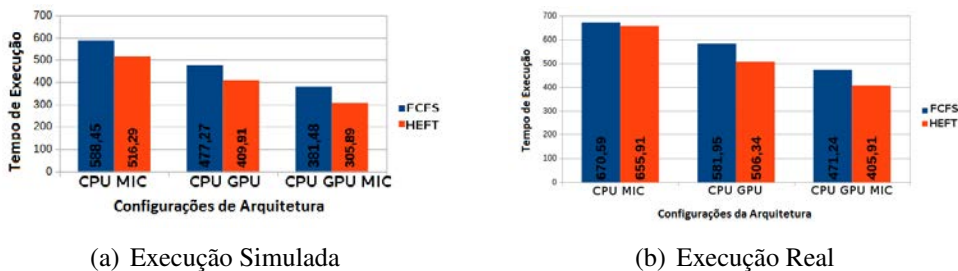


Figura 4. Desempenho dos Escalonadores para Diferentes Arquiteturas Híbridas.

4.2.3. Aumento do número de coprocessadores

Apesar de nas conclusões do trabalho [Andrade et al. 2014a] os autores mencionarem que aumentar o número de coprocessadores pode melhorar o desempenho da aplicação, este fato não foi verificado. Nesta seção, fornecemos esta análise, configurando diferentes arquiteturas no *D-STHARK*. Usando apenas o escalonador HEFT, diminuímos o número de núcleos de CPU e aumentamos o número de coprocessadores. A Tabela 1 apresenta os resultados dessa análise. Podemos notar que, aumentando apenas um coprocessador MIC é possível remover 2 núcleos de CPU. De modo semelhante, aumentando uma GPU é possível remover 4 núcleos de CPU, ambos sem degradar o desempenho do sistema. Estes resultados demonstram que podemos reduzir o consumo de energia (reduzindo núcleos de CPU e aumentando os coprocessadores), com o mesmo desempenho.

Configuração dos Processadores	Tempo(s) do escalonador HEFT
14 CPUs + 1 GPU + 1 MIC	305.892
11 CPUs + 1 GPU + 2 MIC	309.537
9 CPUs + 2 GPU + 1 MIC	301.079
4 CPUs + 2 GPU + 2 MIC	302.646

Tabela 1. Simulando Arquiteturas variando o número de coprocessadores.

5. Conclusões e Trabalhos Futuros

Neste trabalho, apresentamos *D-STHARK* (Dynamic Scheduling of Tasks in Hybrid Simulated ARchitectures), cujo objetivo é fornecer um ambiente simulado completo de arquiteturas híbridas (CPUs, GPUs e MICs) que permite avaliar estratégias de escalonamento dinâmico de aplicações. Avaliamos nossa ferramenta simulando uma aplicação real [Teodoro et al. 2014], bem como a arquitetura e métodos de escalonamento apresentados em [Andrade et al. 2014a]. Os resultados e conclusões obtidos com *D-STHARK* foram os mesmos originalmente relatados, mostrando a efetividade da nossa proposta. Além disso, realizamos um experimento simulando configurações de arquiteturas compostas de mais coprocessadores, as quais não foram avaliadas no trabalho original devido à falta de uma arquitetura real. Nossos resultados alcançaram as mesmas conclusões levantadas no trabalho original de que é possível reduzir o consumo de energia, mantendo o desempenho, utilizando mais coprocessadores. Como trabalho futuro, visamos eliminar algumas simplificações adotadas, incluir outras estratégias de escalonamento, bem como avaliar tarefas dinâmicas e aplicações não embarçosamente paralelas.

Referências

- Andrade, G., Ferreira, R., Teodoro, G., da Rocha, L. C., Saltz, J. H., and Kurç, T. M. (2014a). Efficient execution of microscopy image analysis on cpu, gpu, and MIC equipped cluster systems. In *26th IEEE SBAC-PAD Paris, France*.
- Andrade, G., Ramos, G., Madeira, D., Sachetto, R., Clua, E., Ferreira, R., and Rocha, L. (2014b). Efficient dynamic scheduling of heterogeneous applications in hybrid architectures. In *Proceedings of ACM SAC*, pages 866–871.
- Augonnet, C., Thibault, S., and Namyst, R. (2010). StarPU: a Runtime System for Scheduling Tasks over Accelerator-Based Multicore Machines. Rapport de recherche RR-7240, INRIA.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Bueno, J., Planas, J., Duran, A., Badia, R., Martorell, X., Ayguade, E., and Labarta, J. (2012). Productive Programming of GPU Clusters with OmpSs. In *26th IEEE IPDPS*.
- Fatica, M. and Luebke, D. (2007). High performance computing with CUDA. Supercomputing 2007 tutorial. In *Supercomputing 2007 tutorial notes*.
- Jeffers, J. and Reinders, J. (2013). *Intel Xeon Phi Coprocessor High Performance Programming*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition.
- Lima, J. V. F., Broquedis, F., Gautier, T., and Raffin, B. (2013). Preliminary experiments with xkaapi on intel xeon phi coprocessor. In *25th IEEE SBAC-PAD*, pages 105–112.
- Ravi, V., Ma, W., Chiu, D., and Agrawal, G. (2014). Compiler and runtime support for enabling generalized reduction computations on heterogeneous parallel configurations. In *ACM Supercomputing*.
- Roszbach, C. J., Currey, J., Silberstein, M., Ray, B., and Witchel, E. (2011). PTask: operating system abstractions to manage GPUs as compute devices. In *ACM SOSP*.
- Teodoro, G., Kurc, T., Kong, J., Cooper, L., and Saltz, J. (2014). Comparative Performance Analysis of Intel Xeon Phi, GPU, and CPU: A Case Study from Microscopy Image Analysis. In *IPDPS*.

Paralelização da Indexação de Dados no OPTICS via GPU*

Danilo Melo , Sávyo Toledo , Guilherme Andrade ,
Fernando Mourão , Leonardo Rocha

DCOMP/UFSJ - São João del-Rei, MG , Brasil

{daniloamaral,savyomachado,gandrade,fhmourao,lcrocha}@uufs.j.edu.br

Abstract. *Organizing the large volume of data existing in several domains has become one of the biggest problems in Computer Science. Among the different strategies propose to deal efficiently and effectively with this purpose, we highlight those related to density-based clustering strategies, such as DBSCAN and OPTICS. These strategies allow defining clusters of arbitrary shape, dealing with the presence of data noise robustly. In this work, we present a new approach to make OPTICS feasible based on data indexing strategy using graphs, reducing considerably its complexity. Exploring some parallelization opportunities using GPUs, we show in our experiments that our proposal can be over 200x faster than its sequential version using CPU.*

Resumo. *Organizar o grande volume de dados que vem sendo gerado recentemente tornou-se um dos grandes problemas na computação. Dentre as estratégias que se propõem a lidar com esse problema, destacam-se os de agrupamento baseado em densidade, tais como DBSCAN e OPTICS. Essas técnicas permitem agrupamento de dados em formato arbitrário lidando com ruídos de forma robusta. Nesse trabalho apresentamos uma nova abordagem do OPTICS baseada em uma estratégia de indexação de dados por meio de grafos, reduzindo consideravelmente a complexidade do algoritmo. Explorando algumas oportunidades de paralelização utilizando GPUs, mostramos que a nossa proposta é até 200x mais rápida do que sua versão sequencial usando CPU.*

1. Introdução

Observamos recentemente um aumento significativo no volume de dados que vem sendo gerado. Organizar esses dados e encontrar informações úteis para os usuários tornou-se um dos maiores desafios em Ciência da Computação. A todo o momento surgem novos modelos e algoritmos que sejam capazes de lidar com esses dados de forma eficiente (tempo de resposta e uso adequado dos recursos computacionais) e eficaz (qualidade da resposta). Dentre esses trabalhos, destacam-se os relacionados a agrupamento, que consiste em organizar grandes conjuntos de objetos em diferentes grupos (i.e., *clusters*) de acordo com alguma métrica de similaridade determinada [Jain et al. 1999].

Embora existam muitos algoritmos de agrupamento propostos na literatura, a efetividade da maioria deles é bastante limitada em função da existência de parâmetros de entrada muito sensíveis. Pequenas mudanças nesses parâmetros leva a resultados bastante distintos [Aggarwal and Reddy 2013]. Diante desses desafios, estratégias de agrupamento

*Esse trabalho foi parcialmente financiado por CNPq, CAPES, FINER, Fapemig, e INWEB.

baseadas em densidade [Ester et al. 1996] têm se apresentado como boas alternativas, uma vez que são capazes de formar grupos em formato arbitrário, além de lidar com a presença de ruídos nos dados de forma robusta. A principal ideia por trás dessas propostas é que *clusters* são áreas com alta densidade (muitos objetos) separadas por áreas de baixa densidade (poucos objetos). Uma área é considerada densa se existirem mais objetos que $MinPts$ com uma distância entre eles menor que ϵ' . No entanto, escolher os valores adequados para $MinPts$ e, especialmente, ϵ' ainda não é trivial. Em [Ankerst et al. 1999], é apresentado o OPTICS, um algoritmo capaz de lidar com esses problemas. Ao invés de definir diretamente os *clusters*, o OPTICS produz uma ordenação dos objetos (pontos) de acordo com uma estrutura denominada *reachability*. Iniciando com um valor alto de ϵ , essa estrutura é capaz de representar diferentes densidades dentro do conjunto de dados, sendo que os *clusters* podem ser extraídos utilizando diferentes valores de $\epsilon' \leq \epsilon$.

Como o OPTICS é baseado em distância, em que é necessário calcular a distância entre todos pares de objetos, ele apresenta problemas de eficiência, principalmente quando aplicado em cenários grandes e complexos. Observamos na literatura algumas propostas que visam tornar a utilização desse tipo de algoritmo computacionalmente viável, seja por meio de indexação de dados [Christen 2012], seja utilizando paralelização em diferentes unidades de processamento [Patwary et al. 2013]. Com relação a estratégias de paralelização, propostas que utilizam placas aceleradoras gráficas (GPUs) vêm recebendo merecido destaque, uma vez que são capazes de proporcionar níveis mais elevados de paralelismo, associados a baixos consumos de energia [Andrade et al. 2013].

Nesse trabalho, apresentamos uma nova abordagem para tornar o OPTICS computacionalmente viável baseada em uma estratégia de indexação de dados paralelizada em GPU. Similarmente ao proposto em [Patwary et al. 2013], nossa abordagem representa os dados originais como um grafo $G(V, E)$, onde V são os objetos e E as arestas que ligam objetos que estejam a uma distância menor que ϵ . Para representar esse grafo, adaptamos a estrutura de dados compacta METIS [Karypis and Kumar 1998] utilizando três vetores: Va representa os vértices (o índice representa o vértice e cada posição armazena seu grau e a posição no vetor Ea onde se inicia sua lista de adjacência); Ea_n armazena os vértices na lista de adjacência; e Ea_d armazena o peso da aresta (distância). A construção dessa estrutura é feita completamente em paralelo em GPU e, no final, a lista de adjacência de cada vértice também é ordenada em paralelo. A partir dessa estrutura, a execução do OPTICS se torna muito rápida, uma vez que as operações originalmente mais caras são reduzidas para $O(1)$. Sendo assim, a complexidade do OPTICS passa a ser dominada pelo tempo de construção de uma *heap*, cujo custo é $O(E * \log V)$. Em nossos experimentos de avaliação, demonstramos que essa abordagem resulta em um algoritmo extremamente eficiente, alcançando *Speedup* superiores a $200\times$ utilizando apenas uma GPU.

A implementação do algoritmo e execuções de experimentos foram realizadas pelo aluno Danilo Melo, sob a orientação do professor Leonardo Rocha. As análises dos resultados foram feitas com a colaboração do professor Fernando Mourão. Esse trabalho também contou com a colaboração do aluno Sávyo Toledo na implementação da versão sequencial do algoritmo, bem como do aluno de pós-graduação Guilherme Andrade na geração e validação dos conjuntos de dados utilizados nos experimentos de avaliação.

2. Trabalhos Relacionados

Agrupamento de dados consiste, basicamente, em organizar dados em grupos semanticamente consistentes, baseado em alguma métrica de similaridade previamente definida [Jain et al. 1999]. Vários são os desafios relacionados ao uso dessas técnicas, tais como configurar corretamente os parâmetros de entrada, escolher adequadamente uma métrica de similaridade, bem como trabalhar com grandes volumes de dados. Encontramos na literatura uma grande variedade de algoritmos que visam lidar com esses desafios [Aggarwal and Reddy 2013], que vão desde técnicas simples, porém amplamente utilizadas, como o k-means [Jain et al. 1999], até técnicas mais elaboradas e focadas em contextos específicos, como agrupamento de subespaços [Agrawal et al. 1998]. Um conjunto de técnicas de agrupamento que vem recebendo grande atenção é aquele relacionado com o agrupamento baseado em densidade [Ester et al. 1996, Ankerst et al. 1999]. Tais técnicas são distinguidas pela facilidade de execução e pela aplicabilidade em contextos diferentes. Além disso, essas técnicas não exigem que número de *clusters* seja definido como parâmetro de entrada, como é feito por diversas técnicas.

O algoritmo de agrupamento baseado em densidade mais referenciado na literatura é o DBSCAN [Ester et al. 1996]. Seu funcionamento é baseado no cálculo de proximidade entre cada par de objetos, a qual é definida de acordo com a métrica de similaridade adotada. Os objetos são agrupados a partir de um raio mínimo de proximidade, definido como parâmetro de entrada. O principal problema associado ao DBSCAN é que os *clusters* podem apresentar densidades diferentes, sendo necessário executar o DBSCAN para diferentes raios mínimos de proximidade. O principal algoritmo proposto para resolver este problema é o OPTICS [Ankerst et al. 1999], que ao contrário da maioria das técnicas mencionadas, não produz diretamente *clusters*. Baseado em um valor elevado do raio mínimo de proximidade (ϵ), o OPTICS define uma ordenação dos objetos da coleção que pode representar diferentes *clusters*. A partir deste conjunto de dados ordenado, é possível extrair *clusters* usando diferentes valores de $\epsilon' \leq \epsilon$.

Apesar de encontrar *clusters* de densidades diferentes com apenas uma execução, o OPTICS ainda é uma proposta baseada no cálculo das distâncias entre todos os pares de objetos de uma coleção, assim como o DBSCAN, o que é computacionalmente muito caro. Uma das estratégias mais utilizadas para melhorar o desempenho destes algoritmos é a indexação dos dados [Christen 2012, Karypis and Kumar 1998]. Outra estratégia adotada para tornar esses algoritmos capazes de serem aplicados em cenários de grandes volume de dados é a paralelização dos mesmos utilizando múltiplas unidades de processamento [Patwary et al. 2013] e, mais recentemente, utilizando placas aceleradoras gráficas (GPU) [Andrade et al. 2013]. Nesse trabalho apresentamos uma nova e eficiente abordagem do OPTICS baseada na indexação de dados paralelizada em GPU.

3. Indexação de Dados em GPU para o OPTICS

3.1. OPTICS: Uma visão geral

Como já mencionado, o OPTICS não produz diretamente os *clusters*, mas uma ordenação dos objetos que pode representar diferentes *clusters*. A partir de um raio de proximidade mínimo ϵ , essa ordenação encapsula as informações de todos os *clusters* que podem ser gerados utilizando-se valores de $\epsilon' \leq \epsilon$. O OPTICS é baseado em três conceitos

principais: ϵ -neighborhood, core-distance e reachability-distance, conforme definido abaixo:

Definição 3.1. (ϵ -neighborhood): A vizinhança de um objeto p é o conjunto de objetos s em que $distance(p, s) \leq \epsilon$:

$$N_\epsilon(p) = \{s \mid distance(p, s) \leq \epsilon\}$$

Definição 3.2. (core-distance de p): Tomando $MinPts$ como um número natural e $MinPts$ -distance(p) como a distância de p para seu $MinPts$ vizinho mais próximo, a core-distance de p é definida abaixo:

$$coreDist_{\epsilon, MinPts}(p) = \begin{cases} Undefined, & \text{se } |N_\epsilon(p)| < MinPts, \\ MinPts\text{-distance}(p), & \text{caso contrário.} \end{cases}$$

Definição 3.3. (reachability-distance): A reachability-distance de um objeto p para um objeto o é definido:

$$reachDist_{\epsilon, MinPts}(p, o) = \begin{cases} Undefined, & \text{se } N_\epsilon(o) < MinPts, \\ \max(coreDist(o), distance(o, p)), & \text{caso contrário.} \end{cases}$$

Intuitivamente, a core-distance corresponde à menor distância ϵ' entre p e um objeto em ϵ -neighborhood, ou *Undefined* se p não é core. A reachability-distance de um objeto p para outro objeto o é a menor distância tal que p é alcançável a partir de o se o é core. Todos os objetos começam com a reachability-distance como *Undefined*.

Algorithm 1 Algoritmo OPTICS

```

function OPTICS(Dataset,  $\epsilon$ , MinPts)
  for all ( $p \in \text{Dataset}$ ) and ( $p$ .processed  $\neq$  True) do
     $N = \text{neighbors}(p, \epsilon)$ 
     $p$ .processed = True
    output  $p$ 
     $p$ .setCoreDist( $p, \epsilon, \text{MinPts}$ )
    if  $p$ .coreDist  $\neq$  UNDEFINED then
       $Seeds = \text{empty priority queue}$ 
      Update( $N, p, Seeds$ )
      for all  $q \in Seeds$  do
         $N' = \text{neighbors}(q, \epsilon)$ 
         $q$ .processed = True
        output  $q$ 
         $q$ .setCoreDist( $q, \epsilon, \text{MinPts}$ )
        if  $q$ .coreDist  $\neq$  UNDEFINED then
          Update( $N', q$ )
        end if
      end for
    end if
  end for
end function

function UPDATE( $N, p, Seeds$ )
   $cdist = p$ .coreDist
  for all ( $o \in N$ ) and ( $o$ .processed  $\neq$  True) do
     $newrdist = \max(cdist, distance(p, o))$ 
    if  $o$ .reachDist = UNDEFINED then
      //  $o$  is not in Seeds
       $o$ .reachDist =  $newrdist$ 
      insert( $o, Seeds$ )
    else
      if  $newrdist < o$ .reachDist then
         $o$ .reachDist =  $newrdist$ 
        // move-up on queue
        decrease( $o, Seeds$ )
      end if
    end if
  end for
end function

```

Conforme ilustrado no Algoritmo 1, o OPTICS mantém uma fila de prioridade chamada *Seeds* para expandir a ordenação dos objetos. A prioridade é definida de acordo com a reachability-distance dos objetos. Primeiro, ele seleciona aleatoriamente um objeto p que ainda não foi processado e determina sua core-distance. Se p é um objeto core ($coreDist_{\epsilon, MinPts}(p) \neq Undefined$), para cada vizinho $q \in N_\epsilon(p)$ é calculada uma nova reachability-distance. Se q não está em *Seeds*, ele é inserido no mesmo de acordo com sua reachability-distance. Se q já está em *Seeds* e sua reachability-distance atual é maior do que recentemente calculada, q é movido para cima na fila de prioridades *Seeds*. Todo este processo é repetido até que *Seeds* esteja vazia e não existam mais objetos a serem processados. Com base na saída do algoritmo OPTICS, podemos extrair clusters de densidades distintas $\epsilon' \leq \epsilon$ percorrendo a ordenação gerada, associando os objetos aos clusters de acordo com a reachability-distance e a core-distance dos objetos. Conforme

podemos observar, o tempo de execução do OPTICS é dominado pelo tempo de se obter os ϵ -neighborhood de cada objeto da coleção, cujo custo é $O(V^2)$. Propomos uma adaptação da estrutura METIS [Karypis and Kumar 1998] para contornar esse problema.

3.2. Implementação Sequencial

Em nossa proposta, representamos os dados como um grafo $G(V, E)$, onde V representa objetos de uma dada coleção e E as arestas que ligam os objetos que estejam dentro do raio mínimo (distância menor do que ϵ). Trata-se de um grafo ponderado em que os pesos representam a distância entre dois objetos. Esta distância pode ser calculada por métricas de similaridade (e.g., Similaridade de Cosseno) ou distância (e.g., Distância Euclidiana), de acordo com cada cenário. Como nosso objetivo principal é a paralelização do algoritmo usando GPUs, onde existe uma grande limitação da memória disponível, escolhemos representar o grafo usando uma lista de adjacência compacta, adaptando a estrutura METIS [Karypis and Kumar 1998]. Para isso, usamos três vetores: Va que representa os vértices; Ea_n que armazena a lista de adjacência de cada vértice; e Ea_d que armazena o peso das arestas. Em Va o índice i representa o vértice e cada posição do vetor guarda dois valores: o número total de vértices adjacentes de i (grau); e a posição no vetor Ea onde se inicia sua lista de adjacência.

Ilustramos esta estrutura de dados por meio da Figura 1. Por exemplo, para encontrar a lista de adjacência do objeto com identificação 0, primeiro devemos obter os valores armazenados na posição 0 de Va . O primeiro valor representa o grau do vértice (neste caso 2) e a segunda o índice em que a sua lista de adjacência começa em Ea_n (índice 0). Em seguida, podemos obter a lista de adjacência de 0 visitando duas posições em Ea_n a partir da posição 0. Também é possível obter a distância entre os objetos acessando Ea_d . Neste caso, obtemos os objetos 2 e 3, que têm distância 0.5 e 1.75 para o objeto 0, respectivamente. A complexidade de espaço desta estrutura de dados é $O(V + E)$.

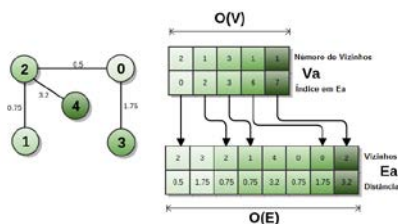


Figura 1. Estrutura de dados utilizada para representar o grafo.

No Algoritmo 2, ilustramos o processo de construção do grafo que irá representar o conjunto de dados. Como podemos observar, o algoritmo recebe como parâmetros de entrada o (*dataset*) (conjunto de dados), ϵ e $MinPts$, retornando o grafo resultante utilizando a estrutura acima descrita. Para cada objeto, primeiro devemos calcular a distância para os demais objetos da coleção. Se o valor calculado é inferior ao parâmetro de entrada ϵ , uma aresta é criada entre esses objetos, e o peso da aresta é armazenado em Ea_d . Após este primeiro passo, devemos ordenar pela distância a lista de adjacência de cada vértice. No caso da execução sequencial, adotamos o algoritmo QuickSort, que tem uma complexidade de $O(n \log n)$. Essa etapa de ordenação é muito importante, uma vez que reduz o custo de se encontrar a *core distance* a $O(1)$ (i.e., determinar o $MinPts^{ésimo}$ vizinho).

Depois de indexar os dados, o passo seguinte é aplicar o OPTICS, conforme apresentado no Algoritmo 1. Neste caso, as funções para determinar o total de vizinhos de um objeto p , bem como a sua *core-distance*, são $O(1)$. Em nosso caso, adotamos uma estrutura *heap* para representar a fila de prioridade *Seeds*. Portanto, a complexidade de tempo do OPTICS passa a ser dominada pelo processo de construção e manutenção da *heap* dada por $O(E * \log V)$. Por outro lado, o processo de indexação tem uma complexidade de tempo de $O(V^2)$, uma vez que, no pior dos casos, será necessário uma comparação entre cada par de objetos. Embora existam técnicas mais elaboradas de indexação que visam reduzir essa complexidade, nossa escolha foi baseada nas oportunidades de paralelização que podem ser exploradas usando GPUs, conforme veremos na próxima seção.

Algorithm 2 Algoritmo para a Construção do Grafo

```

function MAKEGRAPH(MinPts,  $\epsilon$ , dataset, Graph)
  for all  $p \in \text{data.Set}$  do
    for all  $q \in \text{dataset}$  do
      if  $\text{proximity}(p, q) \leq \epsilon$  then
        InsertEdge( $p, q$ , Graph)
      end if
    end for
    dataset = dataset -  $p$ 
    SortAdjacentList( $p$ )
  end for
end function

```

3.3. Implementação paralela

Conforme discutido na seção anterior, o OPTICS é dividido em duas fases principais: (1) construção do grafo e (2) execução propriamente dita do OPTICS. Nesse trabalho nos concentramos na paralelização da primeira etapa, por três razões principais. Em primeiro lugar, a complexidade de construção do grafo é maior que a execução do OPTICS quando o mesmo utiliza nossa estrutura. Em segundo lugar, essa etapa de construção possui muitas oportunidades de paralelização. Por fim, o processo de paralelização do OPTICS usando a estrutura de grafo proposta é limitado à paralelização da estrutura *heap*, a qual possui muitas seções críticas. Como veremos na seção 4, o processamento do OPTICS em si corresponde a menos de 1% do tempo total de execução.

Como descrito na Seção 3.2, o processo de construção do grafo envolve o preenchimento dos vetores Va e Ea (Ea_n e Ea_d) com complexidade de espaço de $O(V)$ e $O(E)$, e está dividido em quatro passos básicos: (1) cálculo do primeiro valor de Va ; (2) cálculo do segundo valor de Va ; (3) construção de Ea ; e, finalmente (4) ordenação das listas de adjacências de Ea_n pela distância armazenada em Ea_d . Todas essas etapas foram paralelizadas usando GPUs, conforme descrito abaixo e resumido na Figura 2.

- **Cálculo do grau dos vértices:** Para cada vértice, calculamos o número total de vértices adjacentes. Nesta etapa, podemos usar os vários núcleos da GPU para processar múltiplos vértices em paralelo. Nossa estratégia paralela usando GPU atribui uma *thread* (núcleo) para cada vértice, isto é, cada entrada do vetor Va . Cada núcleo da GPU, então, irá contar quantos vértices adjacentes têm o ponto sob a sua responsabilidade, preenchendo o primeiro valor no vetor Va . Como podemos ver, não há nenhuma dependência ou comunicação entre as tarefas paralelas, ou seja, trata-se de um problema embaraçosamente paralelo. Assim, a complexidade computacional pode ser reduzida de $O(V^2)$ para $O(V)$.

- Cálculo do índice:** O segundo valor em Va corresponde ao índice inicial em Ea contendo a lista de adjacência de um vértice particular. O cálculo desse valor depende do índice inicial da lista de adjacência e do grau do vértice anterior. Por exemplo, o índice inicial para o vértice 0 é 0, uma vez que ele é o primeiro vértice. Para o vértice 1, o índice inicial é dado pela soma entre o índice do vértice anterior (0) com seu respectivo grau, já calculado no passo anterior (2). Conforme podemos perceber, temos uma dependência de dados onde o próximo vértice depende do cálculo dos vértices anteriores. Este é um problema que pode ser eficientemente resolvido em paralelo por meio de um método de **exclusive_scan** [Blelloch 1990]. Dessa forma, utilizamos a biblioteca *thrust*, distribuída como parte do CUDA SDK, que fornece, entre outros algoritmos, uma aplicação **exclusive_scan** otimizada adequada para o nosso problema.
- Montagem das listas de adjacência:** Estando o vetor Va completamente preenchido, ou seja, para cada vértice conhecemos seu grau e o índice inicial de sua lista de adjacência, calculados nos dois passos anteriores, podemos agora efetivamente montar a lista de adjacência compacta, representada por Ea . Seguindo a lógica do primeiro passo, atribuímos uma *thread* de GPU para cada vértice. Cada uma dessas *threads* estará responsável por preencher em Ea toda a lista de adjacência de um vértice em particular. A lista de adjacência de cada vértice inicia-se no índice referenciado no segundo valor do índice de Va e possui um *offset* igual ao grau de cada vértice armazenado no segundo valor de Va .
- Ordenação das listas de adjacências:** Tendo o vetor Ea_n e Ea_d completamente preenchidos, podemos agora simplesmente ordenar cada lista adjacente apresentada em Ea . Seguindo a lógica da terceira etapa, atribuímos uma *thread* de GPU para cada vértice. Cada uma delas ordenará a lista de adjacência do vértice atribuído a ela. Ao invés de usar o Quicksort, adotamos o SelectionSort, uma vez que a complexidade deste algoritmo é sempre $O(n^2)$, no pior caso, melhor caso e caso médio, evitando assim um desbalanceamento de carga de trabalho entre as *threads*, o que é normalmente muito crítico no caso de paralelização em GPUs.

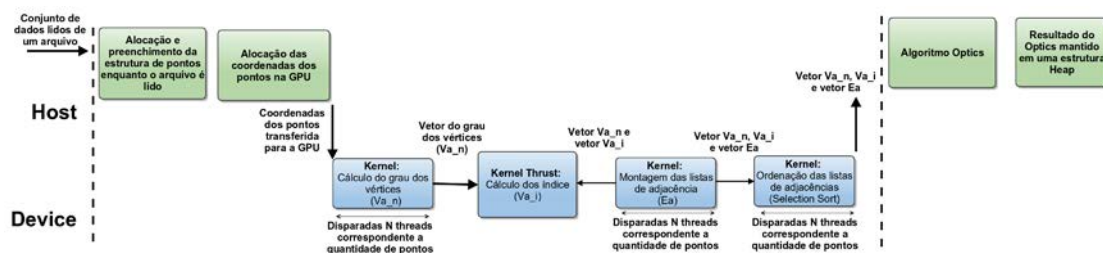


Figura 2. Ilustração do cálculo e transferência dos dados.

4. Avaliação Experimental

4.1. Amostra de dados

Para avaliar a estratégia proposta, geramos várias coleções de dados de diferentes tamanhos, variando entre 5.000 e 700.000 o número total de objetos de entrada em um espaço bidimensional (dois atributos). Para cada coleção, extraímos os tempos de execução referentes à construção do grafo, processo de ordenação das listas de adjacências,

processamento do OPTICS em si, bem como o tempo total gasto pelo algoritmo, tanto para a versão serial implementada em CPU quanto para a versão em GPU proposta nesse trabalho. A partir destes dados, avaliamos e comparamos o Speedup alcançado pela nossa abordagem utilizando GPU em cada parte do algoritmo, bem como na aplicação como um todo. Em todos os nossos testes usamos um conjunto de 20 *clusters Gaussianos* gerados aleatoriamente. Além disso, os parâmetros de entrada são fixos para todos os testes sendo $Min.Pts = 4$ e $\epsilon = 0.05$. A escolha de toda esta configuração das coleções de teste baseou-se no trabalho apresentado em [Böhm et al. 2009], em que os autores propõem e avaliam uma versão paralela do DBSCAN.

Todas as implementações foram feitas em C e C para CUDA (nVidia), e todos os experimentos foram realizadas utilizando uma máquina 4.1.13 GNU/Linux, com 32 GB de memória e um processador Intel Core processador i7-4930K de 3.40 GHz. Utilizamos também uma GPU Tesla K40c, com 2880 CUDA núcleos organizados em 15 multiprocessadores, com 12 GB de memória. A métrica de distância adotada foi a Distância Euclidiana. Nas subseções seguintes, primeiramente, apresentamos uma análise do perfil de execução do algoritmo a fim de verificar quais são as fases computacionalmente mais custosas para, assim, justificar as escolhas de paralelização descritas. Em seguida, apresentamos os resultados obtidos no processo de construção do grafo, e finalmente no algoritmo como um todo, apresentando o Speedup e os ganhos alcançados.

4.2. Execução do Profiling

A fim de avaliar quais etapas do algoritmo são computacionalmente mais custosas, medimos o tempo de execução para a construção do grafo, para o processo de ordenação das listas de adjacências, para o processamento do OPTICS, além do tempo total gasto pelo algoritmo em CPU. A Figura 3 mostra o tempos de execução de cada etapa para todos os conjuntos de dados testados. Observamos que o tempo de construção do grafo domina o tempo de execução do algoritmo para todos os conjuntos de dados testados, sendo 99,97% do tempo total para uma entrada de 700.000 objetos. Este resultado confirma que a paralelização da construção do grafo foi uma boa escolha. A etapa de ordenação também é uma operação importante, pois reduz a complexidade de algumas operações do OPTICS em si para $O(1)$. Apesar de termos paralelizado este passo, o mesmo representa uma fração pequena do tempo total. Assim, nossas análises foram concentradas no Speedup alcançado na construção do grafo e no tempo de execução total do algoritmo.

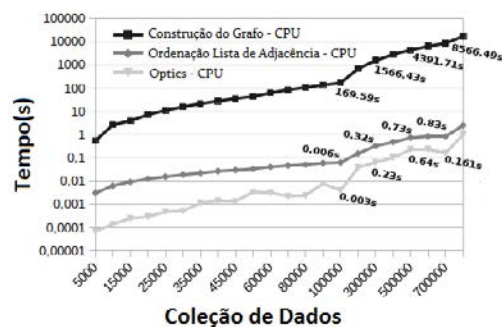


Figura 3. Tempo de execução de cada etapa na CPU

4.3. Avaliação da construção do grafo

Conforme descrito anteriormente, a construção de grafo apresenta a maior complexidade de tempo ($O(V^2)$), exigindo, assim, a maior parte do tempo de processamento de todo o

algoritmo. Na Figura 4, apresentamos os resultados obtidos com o paralelização em GPU deste passo. Como pode ser visto, existe uma redução significativa no tempo de execução desta etapa. O Speedup aumenta significativamente até $N = 100,000$. Para os valores de N superior a 100.000 o crescimento é menor, estabilizando por volta de $N = 600,000$ com um Speedup de $214\times$. Esta estabilização pode ser explicada pelo fato do aumento do tamanho de entrada representar também um aumentado no número de blocos de *threads* da GPU. Assim, a sobrecarga gerada para escalar essas *threads* torna-se um fator limitante, estagnando o Speedup a partir desse ponto. De qualquer maneira, um Speedup de $214\times$ usando apenas uma GPU pode ser considerado um resultado contundente.

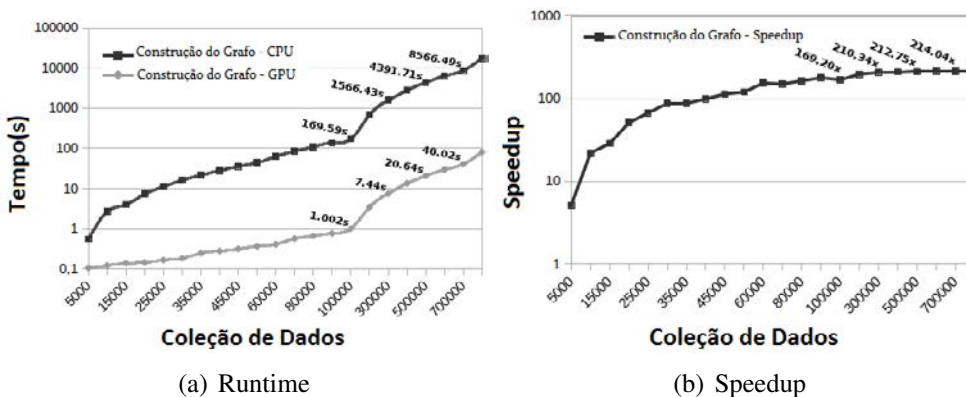


Figura 4. Tempo de execução e Speedup da Construção do Grafo

4.4. Avaliação do tempo total

Na Figura 5 apresentamos os resultados relacionados ao tempo total de execução. Avaliando o tempo de execução alcançado por nosso algoritmo, podemos ver que o Speedup máximo atingido foi de $211\times$, diminuindo o tempo de execução de 8.568,75s na CPU para 40,59s na GPU, com 700,000 pontos. Dessa forma, mesmo com uma implementação multicore deste algoritmo que apresentasse um Speedup linear, seriam necessários mais de 200 CPUs para atingir o mesmo resultado alcançado pela implementação em GPU. Se tomarmos em consideração o custo, os nossos resultados são ainda mais impressionantes. Enquanto uma GPU, como o que foi utilizada nos experimentos, custa cerca de US\$2.000.00, a obtenção de uma configuração com 200 núcleos de CPU custaria um valor consideravelmente superior.

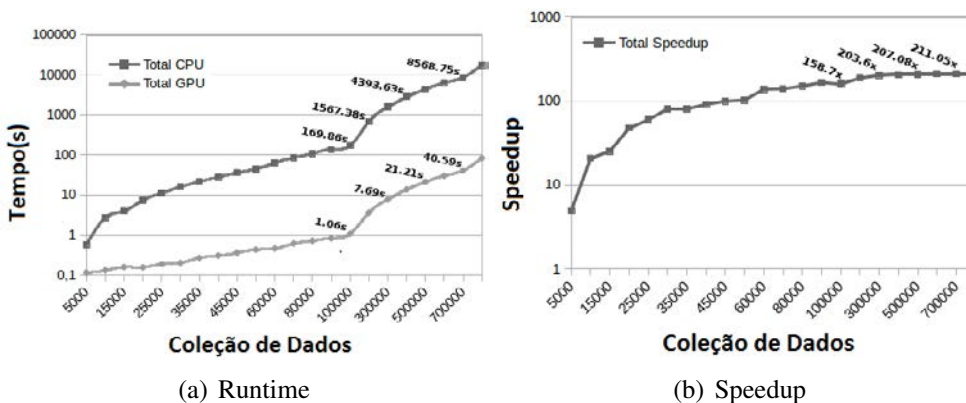


Figura 5. Tempo de execução e Speedup Total

5. Conclusão e Trabalhos Futuros

Nesse trabalho, apresentamos uma nova abordagem para tornar o OPTICS computacionalmente viável baseada em uma estratégia de indexação de dados paralelizada em GPU. Representamos coleções de dados como grafo $G(V, E)$, onde V representa os objetos e E as arestas que ligam os objetos que têm distâncias entre si menores do que um raio ϵ . Construímos uma estrutura de dados adaptada da METIS [Karypis and Kumar 1998], adotando uma estratégia de paralelização para GPUs. Utilizando essa representação, a complexidade do OPTICS é reduzida para $O(E * \log V)$. A fim de avaliarmos empiricamente nossa proposta, desenvolvemos vários conjuntos de testes onde o número de objetos a serem agrupados foi variado entre 5.000 e 700.000 objetos num espaço bidimensional. Em seguida, medimos o tempo de execução do OPTICS, tanto para uma implementação serial em CPU quanto para nossa abordagem em GPU. Nessa avaliação, mostramos que nossa proposta é até $200\times$ mais rápida do que sua versão sequencial usando CPU, usando apenas uma única GPU. Como trabalho futuro, temos proposta de novas estratégias de paralelização do OPTICS em si utilizando GPUs. Pretendemos também estender essas estratégias para usar várias GPUs, bem como avaliar todas as propostas sobre cenários de dados reais, onde o volume é muito grande e os dados são representados por várias dimensões.

Referências

- Aggarwal, C. C. and Reddy, C. K. (2013). *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC, 1st edition.
- Agrawal, R., Gehrke, J., Gunopulos, D., and Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. In *ACM SIGMOD*.
- Andrade, G., Ramos, G., Madeira, D., Sachetto, R., Ferreira, R., and Rocha, L. (2013). G-dbscan: A gpu accelerated algorithm for density-based clustering. ICCS.
- Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). Optics: Ordering points to identify the clustering structure. In *ACM SIGMOD*.
- Blelloch, G. E. (1990). Prefix sums and their applications. Technical Report CMU-CS-90-190, School of Computer Science, Carnegie Mellon University.
- Böhm, C., Noll, R., Plant, C., and Wackersreuther, B. (2009). Density-based clustering using graphics processors. In *ACM CIKM*.
- Christen, P. (2012). A survey of indexing techniques for scalable record linkage and deduplication. *IEEE Transactions on Knowledge and Data Engineering*.
- Ester, M., Kriegel, H., S, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323.
- Karypis, G. and Kumar, V. (1998). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.*, 20(1):359–392.
- Patwary, M. A., Palsetia, D., Agrawal, A., Liao, W.-k., Manne, F., and Choudhary, A. (2013). Scalable parallel optics data clustering using graph algorithmic techniques. In *Proc. of ACM SC*.

DesafIE! - 5º Workshop de Desafios da Computação Aplicada à Educação

Apresentação

O DesafIE! é um evento da Comissão Especial de Informática na Educação (CEIE) da SBC que ocorre desde 2012 e está voltado principalmente a um dos Grandes Desafios científicos da computação apontados pela SBC para a década (2006-2016) que é o “Acesso participativo e universal do cidadão brasileiro ao conhecimento”. Nesta quinta edição, o tema do evento é: “Desafios da Informática na Educação: Considerando fatores sócio-afetivos do estudante”. O DesafIE!2016 busca promover um debate sobre que tipo de tecnologias a comunidade de pesquisa em Informática na Educação busca desenvolver e como elas podem, e se devem, auxiliar na formação mais integral dos indivíduos. Dessa forma, essa edição busca promover uma reflexão sobre o papel da comunidade de Informática na Educação como propositora de soluções computacionais que busquem, além do acesso ao conhecimento e desenvolvimento de habilidades cognitivas, desenvolver habilidades meta-cognitivas e sócio-emocionais para uma formação mais integral dos indivíduos. Algumas questões que norteiam as reflexões dessa edição são: Como as novas tecnologias educacionais propostas estão auxiliando na formação do perfil desejado do cidadão contemporâneo? As tecnologias devem focar apenas no acesso à informação e ao desenvolvimento de habilidades cognitivas? Como as tecnologias podem auxiliar no desenvolvimento de competências meta-cognitivas? Pode, e deve, a tecnologia auxiliar no desenvolvimento de outras habilidades que não cognitivas, como habilidades sócio-emocionais?

A submissão de artigos ao DesafIE foi realizada em duas etapas. Na primeira etapa, foram recebidos 56 resumos de até 1.000 palavras. Estes resumos foram avaliados por três revisores. Foram aceitos 28 resumos que realmente abordavam desafios das pesquisas em informática na educação. A segunda etapa envolveu o envio da versão completa daqueles resumos aceitos na primeira fase. Dos 28 artigos, 13 artigos completos de até 10 páginas foram aceitos. Estas submissões também foram avaliadas por três membros do comitê de programa. A taxa final de aceitação do evento foi de 23%.

Coordenação Geral: Dra. Patrícia Jaques (UNISINOS)
Coordenação do Comitê de Programa: Dra. Maria Augusta Silveira Netto Nunes (UFS)

Comitê de Organização

Coordenação Geral

Dra. Patrícia Jaques (UNISINOS)

Coordenação Local

Dra. Liane Tarouco (UFRGS)

Dra. Marcia de Borba Campos (PUCRS)

Coordenação do Comitê de Programa

Dra. Maria Augusta Silveira Netto Nunes (UFS)

Comitê de Programa

Alberto Castro - Universidade Federal do Amazonas (UFAM)
Alex Sandro Gomes - Universidade Federal de Pernambuco (UFPE)
Anarosa Brandão - Universidade de São Paulo (USP)
André Raabe - Universidade do Vale do Itajaí (UNIVALI)
Andrey Pimentel – Universidade Federal do Paraná (UFPR)
Anita Fernandes - Universidade do Vale do Itajaí (UNIVALI)
Claudia Motta - Universidade Federal do Rio de Janeiro (UFRJ)
Crediné de Menezes - Universidade Federal do Rio Grande do Sul (UFRGS)
Davidson Cury - Universidade Federal do Espírito Santo (UES)
Débora Barbosa - Universidade Feevale (FEEVALE)
Dinani Gomes Amorim - Universidade do Estado da Bahia (UNEB) e FACAPE
Ecivaldo Matos - Universidade Federal da Bahia (UFBA)
Edilson Ferneda - Universidade Católica de Brasília (UCB)
Edson P. Pimentel - Universidade Federal do ABC (UFABC)
Ellen Francine Barbosa - Universidade de São Paulo (USP)
Fernanda Campos - Universidade Federal de Juiz de Fora (UFJF)
Ig Ibert Bittencourt - Universidade Federal de Alagoas (UFAL)
Ismar Frango Silveira - Universidade Cruzeiro do Sul
Itana Stiubiener - Universidade Federal do ABC (UFABC)
José Francisco Netto - Universidade Federal do Amazonas (UFAM)
Julia Marques Carvalho da Silva - Instituto Federal do Rio Grande do Sul (IFRS)
Leonardo Cunha de Miranda - Universidade Federal do Rio Grande do Norte (UFRN)
Liane Rockenbach Tarouco - Universidade Federal do Rio Grande do Sul (UFRGS)
Lucia Franco - Universidade Federal do ABC (UFABC)
Lucia Giraffa – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Luis Paulo Mercado - Universidade Federal de Alagoas (UFAL)
Marcos A. F. Borges - Universidade Estadual de Campinas (UNICAMP)
Maria Augusta Silveira Netto Nunes - Universidade Federal de Sergipe (UFS)
Maria da Graça Campos Pimentel - Universidade de São Paulo (USP)
Márcia C. Moraes - Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

Mirella Moro - Universidade Federal de Minas Gerais (UFMG)
Neide Santos – Universidade do Estado do Rio de Janeiro (UERJ)
Nizam Omar - Universidade Presbiteriana Mackenzie (Mackenzie)
Orivaldo de Lira Tavares – Universidade Federal do Espírito Santo (UFES)
Patrícia Tedesco - Universidade Federal de Pernambuco (UFPE)
Patricia Alejandra Behar - Universidade Federal do Rio Grande do Sul (UFRGS)
Patricia A. Jaques – Universidade do Vale do Rio dos Sinos (UNISINOS)
Patrick Brito - Universidade Federal de Alagoas (UFAL)
Paulo Cysneiros - Universidade Federal de Pernambuco (UFPE)
Raymundo Ferreira Fº - Instituto Federal de Educação, Ciência e Tecnologia Sul-rio-grandense (IFSUL)
Rejane Frozza - Universidade de Santa Cruz do Sul (UNISC)
Ricardo José Rocha Amorim - Universidade do Estado da Bahia (UNEB) e Faculdade de Ciências Aplicadas e Sociais de Petrolina (FACAPE)
Robinson Noronha - Universidade Tecnológica Federal do Paraná/Instituto Tecnológico de Aeronáutica (UTPR)
Romero Tori - Centro Universitário Senac / Universidade de São Paulo (USP)
Rosa Maria Costa - Universidade do Estado do Rio de Janeiro (UERJ)
Sean Siqueira - Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Valguima Odakura - Universidade Federal da Grande Dourados (UFGD)
Wilson Castello Branco Neto - Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina (IFSC)

Palestra

Technology support for learning social-emotional skills for interaction - the building blocks of CSCW

Geraldine Fitzpatrick (Vienna University of Technology, Vienna, Austria)

In Computer Supported Cooperative Work (CSCW) research, the focus is on enabling people to collaborate and communicate together via advances in mobile and social networking technologies. However, the basic social skills that are employed by people as part of these communications and collaborations are largely assumed and taken for granted. We are interested in how technologies can open up opportunities to support the basics of social interaction, such as connectedness, rapport and empathy, and to support the teaching and learning of social and emotional skills. Drawing on recent advances in Psychology, Social Neuroscience and Prevention Science, as well as ubiquitous computing, I explore the design of new technologies to support the more implicit aspects of social interaction and the learning of social-emotional skills, with a particular concern for interactions in everyday contexts.

Painel

Desafios da Pesquisa com Qualidade em Informática na Educação

Dr. Ig Ibert Bittencourt (UFAL)

Dra. Maria Augusta Silveira Netto Nunes (UFS)

Dr. Mariano Pimentel (UNIRIO)

Dr. Sean Siqueira (UNIRIO)

A tecnologia da informação é no século XXI um instrumento para promoção de desenvolvimento cognitivo dos indivíduos, ampliando o potencial humano. Assim, como a escrita levou a uma grande mudança na forma de comunicação e aprendizagem, o computador e suas tecnologias potencializam as formas de aprender, se comunicar e interagir. Nesse contexto, é essencial que a comunidade científica brasileira promova novas tecnologias educacionais inovadoras e adaptadas ao cenário nacional das escolas e universidades. Para que os pesquisadores apresentem resultados científicos inovadores e de reconhecida qualidade científica internacional, é primordial que os trabalhos sigam rigorosa metodologia científica, propondo questões de pesquisa para problemas originais e atuais e passando por um processo de avaliação empírica que busque objetivamente mostrar se a abordagem proposta promove os objetivos educacionais propostos. Esse painel visa discutir os desafios de desenvolver uma pesquisa aplicada à demandas reais de nossa comunidade e com qualidade em Informática na Educação no Brasil.

Trabalhos aceitos

Sistemas Tutores Afetivos: Estado da Arte e Desafios para o Reconhecimento de Emoções Relacionadas à Aprendizagem via Expressões Faciais

Ernani Gottardo (IFECT)

Andrey R. Pimentel (UFPR)

Uma Abordagem para a Recomendação Inteligente de Conteúdo Web na Criação de Materiais e Objetos de Aprendizagem

Clarivando F. Belizário Júnior (UFU)

Fabiano A. Dorça (UFU)

Concepção de Objetos Virtuais de Aprendizagem como uma Prática Multidisciplinar Proveniente de Pressupostos da Computação

Cristiane Ellwanger (UFRGS/URI)

Cristina Paludo Santos (URI)

Análise da Complexidade Ciclomática como Apoio ao Processo de Desenvolvimento do Pensamento Algorítmico

Vinicius Ajala (URI)

Denilson Rodrigues da Silva (URI)

Cristina Paludo Santos (URI)

Carlos O. Rolim (UFRGS)

Challenges of Games Virtualization Applied to Educational Games

Wilk Oliveira dos Santos (UFAL)

Clovis Gomes da Silva Junior (UPE)

Uma Abordagem para a Organização Automática de Objetos de Aprendizagem em Ambientes Educacionais Ubíquos com Base em Estilos de Aprendizagem

Miller M. Mendes (UFU)

Vitor C. de Carvalho (UFU)

Fabiano A. Dorça (UFU)

Renan G. Cattelan (UFU)

Computação aplicada à Educação Musical: Desafios e Perspectivas ao Planejamento Docente no Contexto da Educação Inclusiva

Amanda Melo (UNIPAMPA)

Carla Eugenia Lopardo (UNIPAMPA)

Guilherme Moreira de Melo (UNIPAMPA)

Discussão sobre a Adaptabilidade em Ambientes U-learning baseada em Estilos de Aprendizagem

Luma da Rocha Seixas (UFPE)

Ivanildo José de Melo Filho (IFPE)

Alex Sandro Gomes (UFPE)

As Tecnologias de Análise de Aprendizagem e os Desafios de Prever Desempenhos de Estudantes de Programação

Marcia G. de Oliveira (IFES)

Inteligência Artificial: um Desafio na Detecção Precoce de Indivíduos em Risco de Dislexia

Luis Fernando Orleans (UFRRJ)

Raimundo José Macário Costa (UFRRJ)

Sergio Manuel Serra da Cruz (UFRRJ)

Desafios na Construção de um Ambiente Adaptativo Baseado em Inteligências Múltiplas com apoio de Gamificação

Cleon X. Pereira Júnior (IFGO)

Marcia A. Fernandes (UFU)

O Processo de Autoavaliação em Ambientes Colaborativos e sua Relação com a Mineração de Dados Educacionais

Hugo Vieira Lucena de Souza (UFPE)

Ivanildo Melo Filho (IFPE)

Rodrigo Lins Rodrigues (UFRPE)

Alex Sandro Gomes (UFPE)

Fala, Leitura e Escrita em Aplicativos Móveis: Desafios de uma Abordagem Discursiva

Alline M. R. Kobayashi (UNICAMP)

Fernanda Freire (UNICAMP)

Sistemas Tutores Afetivos: Estado da Arte e Desafios para o Reconhecimento de Emoções Relacionadas à Aprendizagem via Expressões Faciais

Ernani Gottardo¹, Andrey Ricardo Pimentel²

¹Instituto Federal de Educação Ciência e Tecnologia do RS (IFRS)
Erechim – RS – Brasil

²Departamento de Informática - Universidade Federal do Paraná (UFPR)
Curitiba – PR – Brasil.

ernani.gottardo@erechim.ifrs.edu.br, andrey@inf.ufpr.edu.br

Abstract. *Given the inextricable link between learning and emotions, a growing body of research has investigated how to provide functionality for recognizing and adapting to learners' affective reactions in educational software. In this paper, techniques for recognition of basic emotions from facial expressions are presented. The results show that for some basic emotions with most prominent expressions, such as joy, the results are considered promising. However, some learning-centered emotions such as sadness are not easily recognized via facial expressions, needing improvements and adaptations to currently existing methods.*

Resumo. *Considerando a intrínseca relação entre aprendizagem e emoções, pesquisas recentes têm investigado como prover funcionalidades de reconhecimento e adaptação às reações afetivas dos aprendizes em softwares educacionais. Nesse trabalho, técnicas de reconhecimento de emoções básicas via expressões faciais são apresentadas, juntamente com alguns experimentos. Os resultados demonstram que para algumas emoções básicas com expressões mais destacadas, como alegria, os resultados são considerados promissores. Entretanto, algumas emoções relacionadas à aprendizagem como tristeza não são facilmente reconhecidas via expressões faciais, necessitando de avanços e adaptações aos métodos existentes atualmente.*

1. Introdução

Um desafio frequentemente tratado em pesquisas na comunidade científica de informática em educação refere-se ao de desenvolvimento de softwares educacionais que ofereçam um suporte mais efetivo para os estudantes, principalmente agregando funcionalidades adaptativas às necessidades individuais. Alguns ambientes como os Sistemas Tutores Inteligentes (STI) implementam funcionalidades de adaptação, porém, essas funcionalidades normalmente baseiam-se apenas em informações cognitivas, não levando em conta as reações emocionais dos estudantes (Khan et al., 2010; Baker et al., 2012).

Considerando-se a estreita relação entre aprendizagem e emoções, nos últimos anos, a investigação do impacto no processo de aprendizado de fatores como motivação e estados afetivos dos estudantes têm emergido como uma linha de pesquisa promissora (Grafsgaard et al., 2013). Como exemplo, Jaques et al. (2012) destacam que seria muito importante para um sistema detectar quando um estudante está frustrado a fim de encorajá-lo a continuar estudando e a realizar as suas atividades

Entretanto, Picard et al. (2014) observam que estamos ainda nos primeiros estágios do entendimento da complexidade e da relação entre afeto, aprendizagem e *feedback* e que essas questões permanecem um desafio em pesquisas envolvendo ambientes computacionais de aprendizagem adaptativos.

Para que se possa realizar qualquer ação de adaptação às reações afetivas dos estudantes é necessário primeiramente reconhecê-las com exatidão. Nesse sentido, um conjunto de pesquisas tem sido desenvolvido, utilizando técnicas que incluem desde sensores de baixo nível como sinais fisiológicos (e.g. batimentos cardíacos, condutividade da pele) até propostas que utilizam apenas dados de interação com o ambiente.

Uma abordagem que se destaca no processo de reconhecimento das emoções é a utilização de expressões faciais (Sarrafzadeh et al., 2008; Grafsgaard et al., 2013). Isso se deve ao fato de que (a) existe uma forte relação entre características faciais e estados afetivos, (b) as expressões faciais não dependem do ambiente, e (c) não requerem dispositivos caros ou altamente intrusivos para seu reconhecimento.

Nesse trabalho serão apresentadas as etapas e demonstrados os resultados de experimentos com a utilização de métodos de inferência de emoções básicas utilizando expressões faciais. Esses resultados servirão como base para a análise e discussão dos principais desafios e necessidades de avanços nessa área.

Na descrição dos resultados deste trabalho, este documento está estruturado em cinco seções. Na primeira seção apresentou-se a introdução do trabalho, destacando o contexto em que o trabalho está inserido. Na Seção 2, são descritos alguns trabalhos correlatos. Na seção 3 e 4, apresentam-se, respectivamente, detalhes e análise dos resultados dos experimentos realizados. O artigo é finalizado com as conclusões e perspectivas de continuidade deste trabalho.

2. Trabalhos Relacionados

Considerando a importância do desenvolvimento de funcionalidade adaptativas aos estados afetivos dos aprendizes, nos últimos anos observa-se o crescente desenvolvimento de pesquisas que aplicam conceitos oriundos da área conhecida como Computação Afetiva (em inglês *Affective Computing*) em ambientes de software educacionais (D'mello e Graesser, 2012).

Computação Afetiva é uma área multidisciplinar de pesquisa definida por Picard (1997) como “computação que está relacionada com, que surge de ou deliberadamente influencia emoções”(p. 3). A Computação Afetiva busca desenvolver métodos para dotar computadores de várias características sociais, tais como reconhecer emoções e responder apropriadamente a elas, expressar emoções, mostrar empatia, identificar personalidade, entre outros (Jaques et al., 2012).

Relativamente ao processo de reconhecimento de emoções, cabe ressaltar que emoções ou estados emocionais não são diretamente observáveis ou acessados. O que é revelado de forma voluntária ou involuntária pelas pessoas são padrões de expressões. Considerando esses padrões, pessoas ou sistemas podem aplicar técnicas para inferir ou estimar o estado emocional, considerando um determinado nível de erro ou incerteza (Picard, 1997).

A coleta dados para servir de base para a construção em ambientes adaptativos às reações afetivas dos estudantes têm sido realizada utilizando técnicas diversificadas que, frequentemente, requerem a utilização de dispositivos externos. Dentre esses dispositivos, pode-se citar sensor de postura e movimento (Grafsgaard et al., 2013), monitor de batimento cardíaco (Grafsgaard et al., 2013, Picard, 1997), sensores de gestos e expressões faciais (Alexander, 2007; Sarrafzadeh et al., 2008; Grafsgaard et al., 2013; D'mello e Graesser 2012), sensor de condutividade da pele e temperatura (Picard, 1997).

É importante observar que alguns dispositivos citados anteriormente (e.g. monitor cardíaco, sensor de condutividade de pele) podem ser considerados intrusivos, de alto custo ou inviáveis de utilização em larga escala. Além disso, muitos desses dispositivos não são uma solução satisfatória em cenários nos quais sensores externos podem distrair os usuários e interferir nas tarefas principais (D'mello et al, 2008).

Cada modalidade tem suas vantagens e desvantagens para sua utilização como fonte viável de inferência de emoções. Dentre os fatores que interferem na escolha destaca-se: i) a validade do sinal como uma forma natural de identificação de estados afetivos, ii) a confiabilidade do sinal em ambientes reais, iii) o tempo de obtenção e iv) custo e intrusividade do sinal (Calvo et al., 2010)

Nesse contexto, a utilização das expressões faciais para a inferência de estados afetivos é uma das técnicas que apresenta uma relação eficiente nos quatro aspectos citados anteriormente. Isso se deve ao fato de que a face desempenha um papel central no processo de comunicação e interação dos seres humanos, sendo uma das principais e mais aparentes formas de representar o que uma pessoa está tentando comunicar (Picard, 1997).

O método *Facial Action Coding System* (FACS) (Ekman, 1992) é a técnica mais utilizada em pesquisas que utilizam informações de expressões faciais (Picard, 2014). No FACS cada movimento independente da face é denominado *Action Unit*- AU (ex. movimento de sobrancelhas, pálpebras, lábios) que representam os movimentos dos músculos usados para gerar expressões. São definidos aproximadamente 58 AU's, cujas combinações resultam em um grande conjunto de expressões faciais possíveis. É importante observar que, segundo Ekman (1992) alguns desses músculos não podem ser manipulados conscientemente e somente são movimentados por emoções genuínas.

Dada sua abrangência, o métodos FACS tem sido útil para descoberta de movimentos faciais que são indicativos de estados afetivos (Bosch et al., 2014). Contudo, algumas limitações no modelo devem ser consideradas. Conforme observa (Picard, 1997) essa técnica falhará no reconhecimento do estado de alegria caso a pessoa que estiver experimentando alegria suprimir todas as expressões faciais. Outra limitação importante refere-se ao tempo necessário para identificação das expressões e também ao fato de que essa técnica foi desenvolvida originalmente para especialistas humanos codificar emoções baseadas em figuras estáticas.

A acurácia na inferência utilizando expressões faciais é variável entre os diversos estados afetivos. Conforme concluem (McDaniel2007 et al., 2007), encantamento é facilmente detectado por apresentar um grande conjunto de AUs. Por outro lado, frustração (e.g. meio sorriso) e aborrecimento mostraram-se difíceis de serem diferenciadas do estado ‘neutro’ por apresentarem poucas expressões destacadas.

Considerando as limitações do modelo FACS, pesquisas têm sido desenvolvidas para seu aprimoramento. O método denominado FACS+ (Essa et al., 1997) propõem a utilização de técnicas de processamento de fluxo óptico juntamente com um modelo dinâmico de movimento. Essa proposta também leva em consideração um componente temporal (não presente no modelo FACS) em que as ações faciais ocorrem em três fases distintas: aplicação, liberação e relaxamento.

Outras pesquisas promissoras na área de reconhecimento de expressões faciais têm demonstrado a viabilidade de distinguir sorrisos de frustração e satisfação (Hoque et al., 2012) ou expressão de aflição real e deliberada (Littlewort, et al., 2009).

Com o objetivo de analisar os resultados práticos do reconhecimento de emoções via expressões faciais, foi realizado um procedimento experimental que será descrito na sequência.

3.1 Descrição dos dados utilizado no experimento

Para a realização de experimentos utilizando o método de inferência e emoções via expressões faciais foi utilizado nesse trabalho o banco de dados “Cohn-Kanade *Expression Database*” (Kanade et al., 2000). Esse banco de dados foi projetado pelo Grupo de Análise Afetiva da Universidade de Pittsburg com o objetivo de propiciar o desenvolvimento de pesquisas envolvendo análise e síntese de automática de imagens faciais.

As expressões faciais constantes no banco de dados foram coletadas a partir de 210 adultos com idades entre 18 e 50 anos. Nesse grupo, 69% eram do sexo feminino, 31% masculinos, 81% Euro-Americanos, 13% Afro-Americanos e 6% de outros grupos.

Para a coleta das imagens, os participantes foram orientados pelos pesquisadores a realizar uma série de sequência de expressões faciais. Cada sequência deveria começar com uma expressão neutra e proceder para um pico da expressão solicitada (e.g. raiva). O pico de cada expressão para cada sequência foi então codificada utilizando o método FACS (Ekman, 1992) e também classificado em uma das 8 categorias a seguir: neutro, raiva, desprezo, desgosto, medo, alegria, tristeza e surpresa. Essas categorias são baseadas nas expressões básicas definidas em Ekman (1992).

É importante destacar que a classificação das emoções nas categorias descritas foi realizada considerando o que foi requisitado ao participante e não necessariamente o que esse participante expressou efetivamente. Esse método pode diferir em aparência e em tempo comparado com comportamentos espontâneos (Kanade et al., 2000). Destaca-se também que na obtenção das imagens questões de iluminação e ângulo das imagens foram aspectos controlados, utilizando-se lâmpadas de alta intensidade e também coletando imagens em ângulos de rotação de 30 graus.

No total o banco de dados disponibiliza 486 imagens, podendo ser utilizado para treinamento e testes de algoritmos de aprendizagem de máquina por disponibilizar

imagens com seus respectivos metadados relativos às Unidades de Ações (*Action Units*) e rótulos classificando cada imagem.

3.2 Descrição da metodologia do experimento

Para a obtenção de um modelo de inferência de emoções a partir de uma base de dados rotulada, diversas tarefas intermediárias precisam ser realizadas. Nesse artigo utilizou-se com referência a metodologia proposta nos trabalhos de Bartlett et al.(2006) e Littlewort et al. (2004). As principais etapas dessa metodologia envolvem:

1. Detecção e extração da face a partir das imagens do *corpus* sendo utilizado
2. Extração de atributos da face detectada utilizando Filtros Gabor
3. Treinamento de classificadores com base nos atributos extraídos pelos Filtros Gabor para construção de um modelo de inferência de emoções

Para a execução das etapas descritas foram feitos acima tomou-se como base a implementação de Mella e Bellavista (2016). A partir dessa solução base alguns ajustes de parâmetros e adaptações na implementação foram realizados com vistas a adequar às necessidades desse trabalho. Além disso, é importante destacar que todas as etapas utilizam de forma intensa funções e procedimentos disponíveis na plataforma OPENCV (*Open Source Computer Vision*), que é uma solução aberta e largamente utilizada na comunidade científica em aplicações que envolvam processamento de imagens.

Na primeira etapa, a detecção da face nas imagens foi realizada utilizando o método de detecção de objetos usando a técnica conhecida como “*Haar Feature-based Cascade Classifier*” (Viola e Jones, 2001). Esse método utiliza uma abordagem de aprendizagem de máquina na qual classificadores são treinados utilizando um conjunto de imagens positivo e negativo (i.e imagens contendo faces e outras não). Aplicando-se esse método sobre o banco de dados descrito, novas imagens foram então geradas contendo apenas a região da face, que é a área de interesse (*ROI – Region of Interest*).

A plataforma OPENCV provê modelos prontos para a detecção de faces como, por exemplo, o modelo “*haarcascade_frontal_face_defaul.xml*”. Entretanto existem outras opções como o modelo “*haarcascade_frontalface_cbcl.xml*” desenvolvido pelo *Center for Biological & Computational Learning* (CBCL) no MIT. Avaliando-se essas duas opções, no contexto deste trabalho decidiu-se pela primeira por apresentar melhores resultados na detecção de face com menos requisitos de qualidade de iluminação. Entende-se que em um contexto educacional real não se pode esperar ou garantir que as condições de iluminação sejam ideais. Entretanto, a segunda opção (CBCL), quando consegue detectar a face (i.e. boas condições de iluminação), apresenta maiores índices de acerto na classificação final das emoções.

Na segunda etapa, um conjunto de filtros Gabor disponíveis no OPENCV foram aplicados sobre as imagens resultantes da etapa 2 (região facial) para extração de características ou atributos das imagens. Esse tipo de filtro é frequentemente usado no processamento de imagens em tarefas como detecção de bordas, representação e discriminação de textura (Bartlett et al., 2006; Littlewort et al., 2004)

Finalmente, considerando os atributos obtidos pela aplicação dos filtros Gabor, técnicas de aprendizagem de máquinas supervisionadas podem ser utilizadas para o treinamento de classificadores e construção de modelos de inferência. Nesse estudo foram utilizados o classificador SVM (*Support Vector Machines*), considerando sua

eficácia e performance em cenários com um grande número de atributos, que nesse caso são gerados pelos filtros Gabor.

3.3 Aplicação da metodologia descrita e resultados do experimento

Seguindo-se as etapas da metodologia descrita anteriormente, um modelo de classificação para as 7 emoções consideradas (mais o estado neutro) foi gerado. Os principais parâmetros utilizados no estudo foram definidos experimentalmente, testando-se algumas opções e escolhendo a que obteve melhor resultado. O tamanho da imagem da face foi configurado como 48 x 48. Outros parâmetros relevantes referem-se aos filtros Gabor, para os quais foram utilizados os seguintes valores: Sigma=3, Lambda=5 e Theta=4.

Na Tabela 1 a seguir são descritos os resultados da aplicação do modelo de inferência de emoções utilizando como teste 189 das 486 imagens do banco de dados Cohn-Kanade, selecionadas aleatoriamente. Na tabela é apresentada a emoção rotulada no banco de dados, o número de imagens testada dessa emoção e o percentual de acerto do modelo de classificação. Na coluna observação são descritos os principais aspectos relativos aos erros de classificação.

Tabela 1 - Resultados da aplicação do modelo de inferência de emoções no banco de dados Cohn-Kanade

Emoção	Número de Imagens Testadas	Percentual de acerto	Observações sobre os erros de classificação
Tristeza	11	27%	55% foram incorretamente classificadas como Neutro
Neutro	99	95%	O percentual de erro ficou igualmente dividido entre surpresa, desgosto, desprezo e raiva
Desgosto	17	94%	As classificações incorretas foram todas como Neutro (6%)
Raiva	4	100%	O único caso com nenhum erro de classificação
Surpresa	27	93%	As classificações incorretas foram todas como Neutro (7%)
Medo	6	50%	O percentual de erro (50%) ficou igualmente distribuído entre neutro, desprezo e felicidade
Desprezo	7	0%	Nesse caso, as emoções foram classificadas como: Neutro 72%, Tristeza 14% e Raiva 14%
Felicidade	18	94%	As classificações incorretas foram todas como Medo (6%)

Com o objetivo de realizar um teste em um cenário real, com características semelhantes a um estudante usando um software em um ambiente educacional, foi aplicado o modelo gerado para realizar a inferência utilizando vídeo do autor do

trabalho obtido por uma webcam VGA integrada do notebook modelo HP 430. Nesse vídeo foram simuladas algumas expressões para verificação do processo de inferência.

Na Figura 1 são exibidos três quadros desse vídeo em escala de cinza com o respectivo resultado do modelo de inferência descrito na legenda. Nesses quadros considera-se, na avaliação do autor, que houve acerto no processo de classificação que é impresso pelo modelo de inferência na parte superior esquerda da imagem.

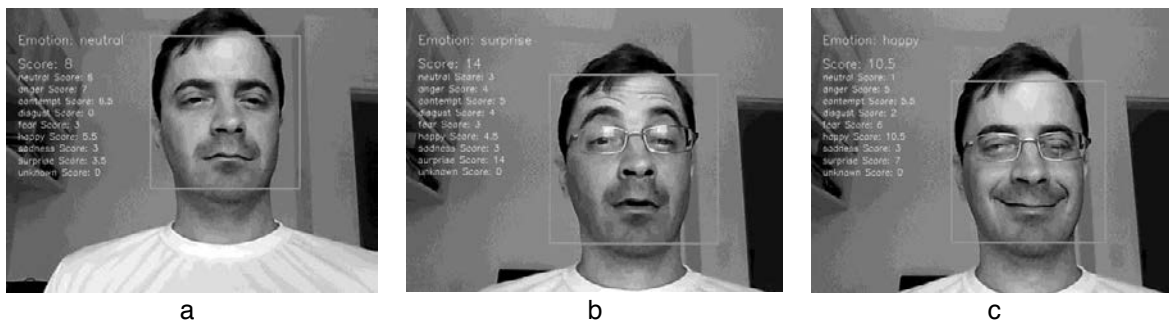


Figura 1 – Quadros do vídeo demonstrando experimento com o autor em que as imagens foram classificadas como: a) neutro, b) surpresa e c) felicidade

Na Figura 2 foram selecionados outros três quadros do mesmo vídeo em que, na avaliação do autor, foi considerado que houve erro no processo de classificação.

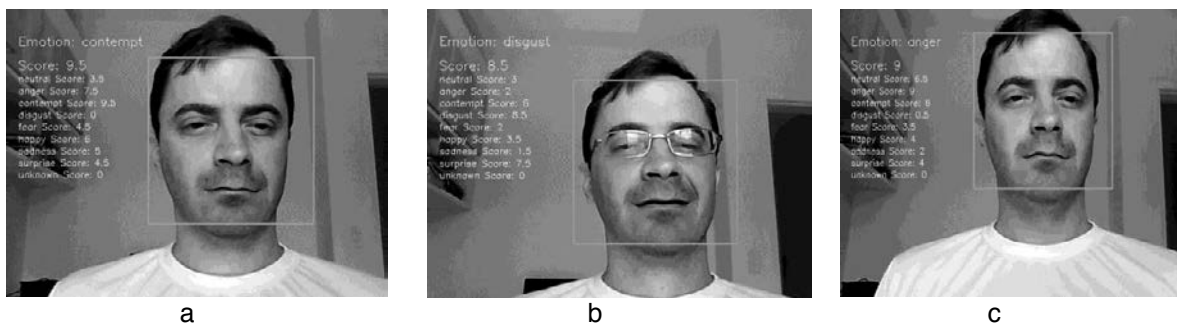


Figura 2 - Quadros do vídeo demonstrando experimento com o autor em que as imagens foram classificadas como: a) desprezo, b) desgosto e c) raiva

4. Análise dos Resultados

Analisando-se os resultados obtidos com o processo de teste com o banco de dados Cohn-Kanade apresentados na Tabela 1 pode-se perceber alguns comportamentos destacáveis. Primeiramente, observa-se que emoções como felicidade, raiva, surpresa e desgosto obtiveram índices elevados de acerto na classificação. Esses índices superam 90% de acurácia, e poderiam, se analisados de forma isolada, serem considerados excelentes.

Porém, índices extremamente menores foram obtidos com as emoções tristeza, medo, chegando a nenhum acerto para a emoção desprezo, demonstrando a dificuldade de inferência desse grupo de emoções, via expressões faciais. Uma observação específica poderia ser feita em relação à emoção Tristeza, que certamente representa uma importante situação de atenção, monitoramento e adaptação em um ambiente educacional. Porém, a classificação dessa emoção em 55% dos casos foi feita como Neutra.

Percebe-se que o grupo de emoções com acurácia superior apresenta expressões com características bastante específicas e altamente diferenciadas entre si e do estado neutro. Por outro lado, expressões como desgosto, tristeza e desprezo apresentam muitas semelhanças entre si e também com o estado neutro o que explica a maior dificuldade de classificação.

No experimento realizado com vídeos contendo expressões simuladas pelo próprio autor, conforme apresentado nas Figuras 1 e 2, os resultados foram analisados visualmente de forma manual pelo próprio autor. Através dessa análise, pode-se perceber que os resultados são muito similares aos obtidos no experimento com o banco de dados Cohn-Kanade, no qual as expressões mais destacadas (e.g raiva, surpresa felicidade) apresentam taxas de acertos superiores às demais.

Esse resultado tem sua relevância considerando-se que nesse segundo experimento as condições de captura das imagens são feitas em um ambiente natural, sem trabalhar aspectos como iluminação ou qualidade de equipamento. Outro aspecto a destacar é que com a aplicação da técnica de reconhecimento facial “*haarcascade_frontal_face_defaul.xml*” disponível no OPENCV conseguiu-se altas taxas de reconhecimento facial no vídeo. Testes com o modelo “*haarcascade_frontalface_cbcl1.xml*” demonstraram que os índices de detecção da face foram extremamente menores, devido principalmente a iluminação.

É importante destacar que no experimento com o banco de dados Cohn-Kanade a escolha aleatória de um grupo de imagens representa uma limitação. Essa escolha poderia ser implementada em versões futuras utilizando-se métodos como validação cruzada, por exemplo. No experimento com vídeos do próprio autor, a principal limitação refere-se à falta de índices globais de acurácia e a análise manual feito pelo autor que poderiam ser obtidos classificando-se todos os quadros por um especialista.

5. Conclusões e Trabalhos Futuros

Mesmo com os avanços alcançados nos últimos anos, diversos desafios ainda precisam ser vencidos para a disseminação de aplicações baseadas em técnicas da computação afetiva em ambientes educacionais. Segundo Van Den Broek (2012), as principais barreiras para o avanço da computação afetiva são: definições conceituais relacionadas às emoções, operacionalização baseada nas definições conceituais e mapeamento dos sinais em emoções.

Analisando-se os resultados desse trabalho, mesmo considerando algumas limitações nos experimentos, pode-se concluir que, levando-se em conta o contexto educacional, diversas questões se impõem: i) as emoções experimentadas por estudantes são menos destacadas ou intensas, tornando mais difícil sua identificação ou distinção; ii) o conjunto de emoções básicas não representa completamente as emoções relacionadas à aprendizagem como por exemplo frustração e aborrecimento; iii) as condições de captura de imagens (e.g. iluminação, posição da face) em um ambiente real de aprendizagem não são as mesmas que a utilizada para construção de bancos de dados em laboratório.

Diante disso, conclui-se que muitas barreiras precisam ser vencidas para a evolução do reconhecimento de emoções via expressões faciais aplicadas ao domínio educacional. Uma das necessidades fundamentais refere-se à construção e disponibilização de bancos de dados que tenha foco no conjunto de emoções

relacionadas ao contexto educacional. É fundamental também que esses conjuntos de dados sejam construídos utilizando ambientes mais próximos ao natural, como por exemplo, com estudantes utilizando um software educacional, utilizando técnicas consolidadas para a anotação da base. Além disso, a inclusão de registros detalhados relativos à interação do estudante com o ambiente seria de extrema valia para a construção de modelos cognitivo-afetivos de inferência de emoções.

Finalmente, é importante observar que teorias e pesquisadores da área de emoção e cognição concordam largamente que as emoções nos seres humanos baseiam-se em dois componentes básicos e interligados: 1) emoções são físicas, enfatizando seu componente corporal; e 2) emoções são cognitivas, ressaltando seu componente mental (Picard, 1997). Dessa maneira, considerar esses dois aspectos em conjunto representa ainda um desafio e, ao mesmo tempo, um caminho promissor em investigações futuras com vistas ao aprimoramento de sistemas tutores que reconheçam e adaptem-se de maneira mais afetiva às reações afetivas dos aprendizes.

Referências

- Alexander, S. T. V. (2007) “An affect-sensitive intelligent tutoring system with an animated pedagogical agent that adapts to student emotion like a human tutor”. Tese de Doutorado. Philosophy in Computer Science, Massey University, Albany, New Zealand.
- Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Alevan, V., Kusbit, G.W., Ocumpaugh, J., Rossi, L. (2012) “Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra”. Proceedings of the 5th International Conference on Educational Data Mining. pp. 126–133.
- Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J. (2006) . Fully automatic facial action recognition in spontaneous behavior. In in 7th International Conference on Automatic Face and Gesture Recognition, pages 223-230.
- Bosch, N., Chen, Y., and D’Mello, S. (2014). It’s written on your face: detecting affective states from facial expressions while learning computer programming. In Intelligent Tutoring Systems, pages 39–44. Springer.
- Calvo, R., D’Mello, S., et al. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on*, 1(1):18–37.
- D’mello, S.K., Craig, S.D., Witherspoon, A.W., Mcdaniel, B.T., Graesser, A.C. (2008) “Automatic Detection of Learner’s Affect from Conversational Cues”. *User Modeling and User-Adapted Interaction*, p. 45-80.
- D’mello, S.; Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, v. 22, n. 2, p. 145–157.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion* 6.3-4 169-200.
- Essa, I., Pentland, A. P., et al. (1997). Coding, analysis, interpretation, and recognition of facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):757–763.

- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., Lester, J. C. (2013) “Embodied Affect in Tutorial Dialogue: Student Gesture and Posture”. *Artificial Intelligence in Education, Lecture Notes in Computer Science* vol. 7926, p. 1-10.
- Hoque, M. E., McDuff, D. J., and Picard, R. W. (2012). Exploring temporal patterns in classifying frustrated and delighted smiles. *Affective Computing, IEEE Transactions on*, 3(3):323–334.
- Jaques, P. A., Nunes, M. A. S. N., Isotani, S., Bittencourt, I. (2012) “Computação Afetiva aplicada à Educação: Dotando Sistemas Tutores Inteligentes de Habilidades Sociais”. *Anais do XXXII Congresso da Sociedade Brasileira de Computação, DesafIE 2012*.
- Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. Paper presented at the Fourth IEEE International Conference on Automatic Face and Gesture Recognition.
- Khan, F. A., Graf, S., Weippl, E. R., Tjoa, A. M. (2010) “Identifying and Incorporating Affective States and Learning Styles in Web-based Learning Management Systems”. *Interaction Design and Architecture(s) Journal - IxD&A*, N. 9-10, p. 85-103.
- Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J., Movellan, J. (2004). Dynamics of facial expression extracted automatically from video. In *J. Image & Vision Computing*, pages 615-625.
- Littlewort, G. C., Bartlett, M. S., and Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, 27(12):1797–1803.
- McDaniel, B., D’Mello, S., King, B., Chipman, P., Tapp, K., and Graesser, A. (2007). Facial features for affective state detection in learning environments. In *Proceedings of the 29th Annual Cognitive Science Society*, pages 467–472. Cognitive Science Society Austin, TX.
- Mella, L., Bellavista, D. (2016). *Emotime: Recognizing emotional states in faces*. *Elaborazione Delle Immagini LM*. University of Bologna.
- Picard, R. W. (1997) *Affective computing*. Cambridge, Mass: MIT Press.
- Picard, R. W., Fedor, S., Ayzenberg, Y. (2014). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*.
- Sarrafzadeh, A., Alexander, S., Dadgostar, F., Fan, C., Bigdeli, A. (2008) “ ‘How do you know that I don’t understand?’ A look at the future of intelligent tutoring systems”. *Computers in Human Behavior* 24, p. 1342–1363.
- Van Den Broek, E. L. (2012). Affective computing: a reverence for a century of research. In *Cognitive Behavioural Systems*, pages 434–448. Springer.
- Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511). IEEE.

Uma Abordagem para a Recomendação Inteligente de Conteúdo Web na Criação de Materiais e Objetos de Aprendizagem

Clarivando F. Belizário Júnior, Fabiano A. Dorça

Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brasil

clarivando@ufu.br, fabianodor@ufu.br

Resumo. *A criação de cursos online em larga escala adaptáveis aos estilos cognitivos e de aprendizagem dos estudantes é um problema desafiante e pendente. Muitas abordagens tratam apenas da adaptação desses cursos, e as poucas que tratam da criação, exigem do usuário conhecimento técnico e de domínio do conhecimento na construção de objetos de aprendizagem. Neste artigo, uma nova abordagem é modelada para lidar com esses desafios. Nela, dado um programa de curso, suas seções são mapeadas em conceitos representados em ontologias descritas na linguagem OWL da Web Semântica. Esses conceitos são utilizados na busca inteligente por objetos de aprendizagem e por recursos web anotados em RDF, que podem ser utilizados na criação de objetos de aprendizagem.*

Abstract. *The creation of online adaptive courses, which consider cognitive and learning styles information about students, is a challenging problem. Many approaches consider only the adaptation of these courses, and the few that deal with the creation of them require tutor expertise and domain knowledge in the construction of learning objects. In this article, a new approach is modeled to deal with these challenges. In the proposed, given a course design, its sections are mapped to concepts represented in ontologies described in OWL language of the Semantic Web. These concepts are used in an intelligent search for web resources with RDF metadata which comply with the lesson domain, automating the authoring process of new learning objects.*

Palavras-chave: Web Semântica, Objetos de Aprendizagem, Busca Semântica

1. Introdução

Um dos maiores desafios da Educação é adaptar o processo de ensino-aprendizagem aos estilos cognitivo e de aprendizagem de cada aluno, pois no contexto da sala de aula, as estratégias de ensino são voltadas para a turma. Há poucos anos, a informática tornou-se uma grande aliada da Educação na solução desse desafio com a utilização de hipermídias adaptativas capazes de adaptar conteúdo educacional digital às preferências de aprendizagem de cada estudante. Essas hipermídias podem ser integradas a Sistemas de Gerenciamento de Aprendizagem, como o Moodle, permitindo a criação de cursos online.

Se as máquinas compreendessem a linguagem humana, esses cursos seriam criados de maneira totalmente automatizada com base num simples programa de curso.

Por outro lado, a criação desses cursos pelo usuário tende a ser um processo demorado que coloca o esforço no conhecimento técnico em vez do domínio de instrução. Além disso, a busca por recursos web para compor recursos educacionais poucas vezes retorna os melhores conteúdos. Esses são alguns dos desafios da computação que precisam ser solucionados para a efetivação do uso de tecnologias na Educação.

Para enfrentar esses desafios, propõe-se uma abordagem para modelar ferramentas de criação de cursos online usando a Web Semântica, exigindo, assim, o menor esforço possível do criador do curso. Embora a Web possa ter informações suficientes para compor um curso online, elas não são compreendidas pelas máquinas. A Web Semântica surgiu como uma extensão da Web tradicional para superar essa limitação permitindo que os computadores pudessem compreender as informações publicadas na internet.

No contexto deste trabalho, considera-se que um programa de curso é dividido em capítulos que por sua vez são divididos em seções. O material de aprendizagem é formado por um conjunto de OAs de tipos diferentes, por exemplo, exercício, questionário e conclusão. A ferramenta que implementar esta abordagem deve criar um material de aprendizagem para cada seção do programa de curso. Para isso, o primeiro passo é mapear cada seção do programa de curso para um conceito de alto nível, que pode ser decomposto em outros conceitos de nível inferior, representado na ontologia do domínio de conhecimento do curso. Esses conceitos de alto nível representam as metas de aprendizagem dos aprendizes no início do curso e devem ser utilizados na busca pelos recursos web que comporão o curso online. Selecionaram-se quatro trabalhos da literatura que utilizam abordagens diferentes para recomendar (buscar por) OAs usando ontologias. Essas abordagens, que consideram as competências e os estilos cognitivos dos aprendizes, serão implementadas e testadas para ser selecionada a melhor no contexto da abordagem proposta neste trabalho.

Como nem sempre o usuário encontra materiais e objetos de aprendizagem da Web prontos para compor seu curso, ele deve criá-los do zero. A abordagem proposta neste trabalho modela a criação de OAs para auxiliar o usuário na realização dessa tarefa. Para isso, há alguns métodos de buscas na Web Semântica, mas não há nenhum estudo comparativo da eficiência deles no contexto da criação de recursos educacionais. Portanto, comparar-se-ão esses métodos para escolher o mais eficiente no contexto educacional e, em seguida, utilizá-lo na criação de um algoritmo de busca semântica a ser utilizado na abordagem proposta. O algoritmo deve executar a busca em conteúdos web anotados na linguagem RDF (*Resource Description Framework*). O resultado desse algoritmo é um conjunto de recursos web apropriados para compor cada objeto de aprendizagem a ser criado pelo usuário. Há trabalhos recentes que permitem a geração semiautomática de metadados de OAs derivados de capturas multimídia (ARAÚJO et al., 2014) e conteúdos da Wiki (MENOLLI; MALUCELLI; REINEHR, 2011), mas não derivados de conteúdos web anotados em RDF. Portanto, na abordagem apresentada neste trabalho investigam-se maneiras de automatizar a criação de metadados de OAs no padrão IEEE-LOM derivados de recursos web anotados em RDF.

É importante ressaltar que novos conteúdos são publicados todos os dias na rede pelo grande público e, portanto, acredita-se que vá haver sempre a necessidade de criação de novos OAs e, conseqüentemente, a criação e atualização de ontologias para representarem novos conteúdos, por isso, a criação de OAs é um processo contínuo que

deve ser automatizado, dado que na maioria das vezes os objetos recomendados estarão ultrapassados.

Assim, na ferramenta web que implementar a abordagem proposta, o usuário entrará apenas com o programa de curso e informações de filtragem que julgar necessárias (como palavras-chave e os formatos dos OAs que comporão cada seção) e a ferramenta web usará ontologias para mapear cada seção do curso descrita em linguagem natural para conceitos de alto nível, os quais são utilizados na busca inteligente por OAs que poderão compor cada material de aprendizagem (seção) do curso. Caso o usuário julgue algum objeto inadequado, poderá criar seu próprio objeto de aprendizagem usando a recomendação inteligente de conteúdo realizada pelo algoritmo de busca semântica em arquivos RDF. Os OAs criados pelo usuário deverão ser adicionados a repositórios de OAs descritos no padrão IEEE-LOM para promover a reusabilidade.

Portanto, a abordagem proposta contribuirá para a criação de materiais e objetos de aprendizagem exigindo o menor esforço possível do usuário e permitirá a reusabilidade desses recursos educacionais por diversos profissionais da Educação. Ao suprir as expectativas desses profissionais, esta abordagem mostra-se uma solução promissora para a efetivação do uso de tecnologias na Educação.

2. Referencial Teórico

2.1. A Web Semântica

A Web Semântica (BERNERS-LEE et al., 2001) surgiu como uma extensão da Web tradicional para permitir a compreensão dos dados da internet por computadores. Diversos estudos estão sendo realizados para a consolidação das camadas RDF (*Resource Description Framework*), RDF-S e ontológica pertencentes à Web Semântica. A camada RDF pode ser considerada uma camada de base composta por triplas em XML (*Extensible Markup Language*) formadas por sujeito, predicado e objeto capazes de especificar a semântica de metadados facilitando a busca por recursos na Web.

A RDF-S (*RDF-Schema*) tem o objetivo de descrever propriedades e classes, bem como suas hierarquias. Essa camada é utilizada em conjunto com a RDF, visando ainda mais o enriquecimento semântico. A camada ontológica utiliza a linguagem OWL (*Web Ontology Language*) para instanciar ontologias na Web. Essa camada fornece vocabulário adicional com uma semântica formal para potencializar a interpretação por máquinas do conteúdo da Web.

De modo geral, essas camadas utilizam metadados (descritos em RDF ou OWL) para descrever os relacionamentos semânticos entre os conteúdos da Web. A linguagem SPARQL é a linguagem padrão da Web Semântica usada para efetuar consultas sobre gráficos RDF e OWL. Como essas consultas consideram a semântica dos conteúdos, elas são bem mais eficientes que as consultas realizadas na Web tradicional.

2.2. Padrões de metadados de objetos de aprendizagem

Os diversos conteúdos da Web Semântica podem ser usados na criação de objetos de aprendizagem. O padrão IEEE-LOM (DUVAL, 2002) é reconhecido internacionalmente por facilitar a busca e criação de OAs. O padrão IEEE-LOM é parte integrante dos padrões OBAA (VICCARI et al., 2010) e SCORM (ADL, 2001). O padrão de metadados OBAA estende o padrão IEEE-LOM e fornece interoperabilidade entre plataformas de hardware para lidar com o contexto educacional brasileiro. O SCORM (*Sharable Content Objects Reference Model*) é um padrão de referência que assegura a interoperabilidade entre diferentes sistemas de gerenciamento da aprendizagem. Os metadados de OAs descritos nesses padrões são armazenados em repositórios de OAs, permitindo a reusabilidade desses recursos educacionais pelas ferramentas de recomendação de OAs.

3. Trabalhos Relacionados

O uso de ontologias parece promissor no ensino, e há trabalhos recentes que as utilizam para a recomendação de OAs. Di Martino (2009a) utilizou técnicas semânticas combinadas com técnicas de casamento de esquemas (*schema matching*) para buscar por OAs relevantes (no padrão SCORM) usando ontologias de um dado domínio.

Castro e Satler (2010) utilizaram técnicas de mineração de dados no repositório de OAs AGORA para gerar regras (conhecimento) na linguagem SWRL da Web Semântica. Essas regras representadas em uma ontologia permitem classificar os OAs em ativos, passivos e muito passivos. Os autores mostram que esses resultados são importantes para o sequenciamento de OAs, e essas técnicas de mineração de dados podem ser melhoradas para a busca e rotulação de OAs.

Jácome Júnior, Mendes Neto e Da Silva (2012) apresentam uma abordagem baseada em algoritmo genético (AG) para a recomendação de objetos de aprendizagem sensível ao contexto do estudante. Agentes de software são combinados com ontologias de descrição de perfis (estático e dinâmicos) dos alunos e com o AG para a recomendação de conteúdos educacionais de maneira autônoma.

Gaeta et al. (2013) utilizaram ontologias para modelar domínios do conhecimento considerando duas relações principais entre conceitos. Na primeira relação, um conceito ou um grupo de conceitos faz parte de um outro conceito de alto nível, assim chamado por conter um ou mais conceitos. Na segunda relação, um conceito é requerido por outro, isso implica que o conceito requerido deve ser aprendido antes desse outro. Os conceitos de alto nível que forem a meta final do processo de aprendizagem devem ser identificados manualmente pelo usuário na ontologia. Então, um algoritmo analisa essa ontologia e o modelo cognitivo do estudante para gerar um caminho de aprendizagem formado pelos conceitos necessários para atingir as metas de aprendizagem. Em seguida, outro algoritmo cria uma sequência de OAs associada a esses conceitos. Esse algoritmo utiliza um algoritmo guloso para minimizar o número de OAs dessa sequência, e recomenda os OAs que melhor satisfazem as preferências de aprendizagem do estudante.

Na Tabela 1, são apresentadas as principais técnicas de busca e recomendação de OAs no contexto dos quatro trabalhos apresentados anteriormente. Testar-se-ão essas técnicas para escolher a melhor no contexto da abordagem proposta neste trabalho.

As técnicas baseada em *schema matching* e algoritmo genético foram testadas em metadados do padrão SCORM, enquanto que as técnicas baseadas em regras de inferência e no algoritmo guloso foram testadas em metadados do padrão IEEE-LOM.

Tabela 1: Técnicas de recomendação de OAs

Literatura correlata	Técnicas de recomendação de OAs	Padrões de metadados
Di Martino (2009b)	<i>Schema matching</i>	SCORM
Castro e Satler (2010)	Regras de inferência (criadas pela Mineração de Dados)	IEEE-LOM
Jácome Júnior, Mendes Neto e Da Silva (2012)	Algoritmo Genético	SCORM
Gaeta et al. (2013)	Algoritmo Guloso	IEEE-LOM

Sabe-se que muitas vezes os usuários não encontram os OAs adequados para compor seus cursos online. Nessas situações, eles criam seus próprios OAs usando recursos da Web. Porém, poucas vezes a Web tradicional retorna os recursos adequados para compor um dado objeto de aprendizagem. Para solucionar esse problema, trabalhos recentes utilizam a Web Semântica para recomendar recursos web descritos em RDF de maneira mais eficiente. Os algoritmos de busca na Web Semântica são de propósito geral e, portanto, podem ser testados em diversos contextos. Contudo, a literatura carece de trabalhos que avaliem e comparem esses algoritmos no contexto do *e-learning*. Para proceder com essa tarefa, implementar-se-ão e avaliar-se-ão os principais algoritmos de busca na Web Semântica, quais sejam: OntoRank, TripleRank, RareRank e Re-Rank semântico.

Em seguida, escolher-se-á o melhor no contexto da nova abordagem apresentada. O OntoRank (FININ et al., 2005) é o algoritmo usado pelo Swoogle (DING et al., 2004) na busca por conteúdo web anotado em RDF. Esse algoritmo é fundamentado no modelo *rational surfer*. Um agente racional navega de um Documento da Web Semântica (SWD) para outro com uma probabilidade constante ou salta para um SWD aleatório. Esse agente utiliza ontologias para compreender os termos dos documentos web visitados.

O algoritmo TripleRank (FRANZ et al., 2009) captura a semântica latente de dados da Web Semântica, por meio de métodos estatísticos, a fim de produzir descrições mais detalhadas desses dados.

O RareRank (WEI; BARNAGHI; BARGIELA, 2011) é um algoritmo utilizado para ranquear entidades em sistemas de busca semântica. Esse algoritmo simula o comportamento racional de pesquisadores humanos na busca por conteúdo da literatura científica.

O algoritmo Re-Rank semântico (WANG et al., 2011) utiliza similaridade semântica para melhorar os resultados da busca. Os resultados candidatos retornados por uma ferramenta de busca são ranqueados novamente (*re-rank*) comparando-os com a consulta.

No contexto da nova abordagem proposta, dentre esses algoritmos, o escolhido será utilizado na busca por recursos web que poderão compor OAs. Esse processo envolve a criação de metadados no padrão IEEE-LOM usando os metadados em RDF

dos recursos web. A literatura carece de trabalhos que tratam essa questão, que permanece em aberto. Resolver essa questão é importante porque a descrição de objetos em um padrão específico do contexto educacional permite a construção de ferramentas de busca direcionadas a esse contexto e, portanto, mais otimizadas. Portanto, investigar-se-ão possíveis mapeamentos entre metadados RDF e metadados descritos no padrão IEEE-LOM.

4. Abordagem Proposta

Propõe-se uma abordagem para modelar ferramentas de criação de cursos online usando ontologias da Web Semântica. Essas ontologias lidam com a linguagem natural dos programas de curso para identificar neles os conceitos que representarão os objetivos de aprendizagem. Esses conceitos são usados na busca por OAs ou, caso os objetos não sejam encontrados, na busca por recursos web anotados em RDF, que podem ser utilizados na criação de OAs.

A arquitetura dessa abordagem é apresentada na Figura 1. As seções do curso representam conceitos relevantes de um domínio do conhecimento representado por uma ontologia, a qual é utilizada para gerar um material de aprendizagem (conjunto de OAs) para cada seção do programa de curso. Um programa de curso é dividido em capítulos, os quais são divididos em seções. Cada material de aprendizagem é formado por um conjunto de OAs de tipos diferentes. Esses tipos são os mesmos utilizados no mecanismo flexível de recomendação de OAs proposto por Graf, Kinshuk e Ives (2010). Isso significa que trabalhos no estado da arte baseados nesses tipos de OAs, como a abordagem proposta por Silva e Dorça (2014), podem ser facilmente combinados com a nova abordagem proposta neste trabalho para se criar um sistema tutorial inteligente completo, com autoria de conteúdo (semi)automatizada e personalização de cursos online com base nos estilos de aprendizagem dos estudantes.

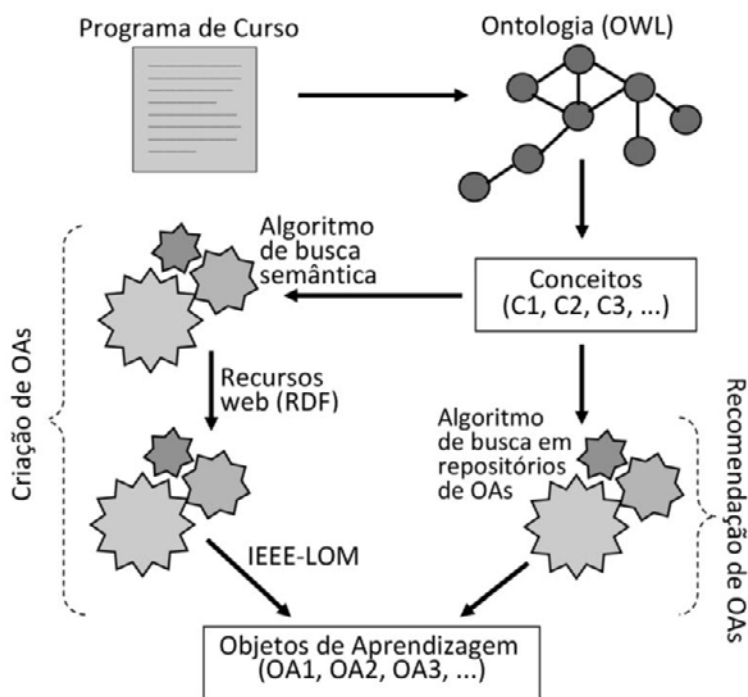


Figura 1: Modelo para a recomendação e criação de OAs

Um diferencial da abordagem apresentada é lidar com a linguagem natural dos programas de curso, dos quais são derivados os conceitos a serem aprendidos pelo estudante. Para isso, o primeiro passo é mapear a linguagem natural para conceitos representados em uma ontologia anotada em OWL, que é uma linguagem da Web Semântica que as máquinas entendem. Embora cada conceito seja parecido com a linguagem natural da sua seção de curso correspondente, acredita-se que esse mapeamento não seja simples, pois os computadores não compreendem a linguagem natural, por isso, usar-se-ão ontologias associadas a um domínio do conhecimento, capazes de associar um conceito de alto nível para cada seção do programa de curso. Utilizar-se-á a linguagem SWRL (*Semantic Web Rule Language*) para implementar as regras que farão a inferência desses conceitos usando a ontologia. Para recomendar OAs usando esses conceitos, testar-se-ão as quatro técnicas identificadas nos trabalhos relacionados a este. Em seguida, escolher-se-á a técnica mais eficiente no contexto da abordagem proposta neste trabalho.

Di Martino (2009a) propôs uma ferramenta que aplica técnicas de *schema matching* entre um esquema de origem e um esquema alvo. O esquema de origem é uma ontologia que modela um domínio do conhecimento e agrega exatamente os conceitos a serem aprendidos. O esquema alvo é um arquivo xml que representa os conceitos extraídos de repositórios de OAs. Os detalhes da criação desse segundo esquema são apresentados em Di Martino (2009b). Para implementar essa recomendação de OAs no contexto da abordagem deste trabalho, utilizar-se-ão os conceitos gerados com base no programa de curso para representar o esquema de origem, mantendo entre esses conceitos os mesmos relacionamentos presentes na ontologia usada para gerá-los. O esquema alvo será o mesmo que foi utilizado pelo autor. Assim, as técnicas de *schema matching* usadas pelo autor poderão ser testadas no escopo da abordagem deste trabalho.

Para testar as técnicas de mineração de dados (em repositórios do padrão IEEE-LOM) propostas por Castro e Satler (2010), utilizar-se-ão os conceitos gerados com base no programa de curso mantendo os mesmos relacionamentos da ontologia de domínio da qual esses conceitos foram derivados, formando assim, uma ontologia composta apenas pelos conceitos a serem aprendidos pelo estudante. As regras de inferência geradas pelas técnicas de mineração de dados a serem testadas serão inseridas nessa ontologia. Assim, essas técnicas serão avaliadas na recomendação de OAs. Investigar-se-ão trabalhos no estado da arte para filtrar esses OAs com base nos estilos de aprendizagem e cognitivo dos estudantes.

Jácome Júnior, Mendes Neto e Da Silva (2012) utilizaram um algoritmo genético (AG) para a recomendação de OAs do padrão SCORM com base nas características desses OAs relacionadas com o perfil do estudante representado por uma ontologia. No contexto da nova abordagem proposta, essa ontologia também modelará os conceitos a serem aprendidos pelo estudante. Assim, o AG a ser implementado será avaliado na recomendação de OAs com base nos estilos de aprendizagem e cognitivo dos aprendizes.

A última técnica a ser testada envolve a aplicação de um algoritmo guloso (GAETA et al., 2013) em uma sequência de OAs relacionados aos conceitos a serem aprendidos pelos estudantes. Esse algoritmo minimiza o número de OAs dessa sequência para recomendar os OAs que satisfazem as preferências de aprendizagem dos

estudantes. Assim como no trabalho de Gaeta et al. (2013), esses conceitos serão modelados por uma ontologia. Embora o algoritmo guloso não seja muito preciso, ele é rápido, por isso, é importante testá-lo no contexto da abordagem proposta.

Como nem sempre os OAs adequados são encontrados, a abordagem apresentada permite que o usuário crie seus próprios OAs usando os conteúdos da Web Semântica. Para isso, os mesmos conceitos utilizados na busca por OAs, são usados na busca por conteúdos web anotados em RDF, que servirão para compor os OAs. Testar-se-ão quatro algoritmos de busca diferentes, o OntoRank, o TripleRank, o RareRank e o Re-Rank semântico, para escolher o melhor no contexto do ensino-aprendizagem.

O conteúdo web, propriamente dito, retornado por esses algoritmos pode ser mapeado diretamente para o conteúdo de OAs, mas o mapeamento de metadados em RDF para padrões de OAs não é uma tarefa trivial, e a literatura carece de uma solução para essa questão. Portanto, investigar-se-ão possíveis mapeamentos entre metadados em RDF e metadados no padrão IEEE-LOM. Uma vez que esses mapeamentos tenham sido feitos para o padrão IEEE-LOM, outros padrões baseados no IEEE-LOM, como o OBAA e o SCORM, poderão se beneficiar dessa solução, promovendo a flexibilidade desta abordagem no uso de padrões de construção de OAs.

5. Considerações Finais

Um grande desafio da Educação é adaptar o processo de ensino-aprendizagem aos estilos cognitivo e de aprendizagem de cada estudante. No contexto da sala de aula, as estratégias de ensino são voltadas para a turma, pois parece impraticável que o professor atenda as especificidades de aprendizagem de cada aluno. Nesse aspecto, a Informática pode contribuir muito na Educação.

Neste artigo, apresenta-se uma pesquisa em andamento. Propõe-se uma nova abordagem, baseada nas tecnologias da Web Semântica, para a recomendação inteligente de conteúdo web na criação de materiais e objetos de aprendizagem. Além disso, mostra-se como um curso online pode ser criado usando OAs recomendados por buscas que usam ontologias. Mesmo que haja necessidade de criar novos objetos para esse curso, o usuário pode se beneficiar de uma série de conteúdos da Web Semântica, usando algoritmos semânticos que retornam os recursos web mais apropriados para compor esses novos OAs.

Os resultados preliminares das técnicas de busca estão sendo comparados para serem avaliados e discutidos, afim de implementar os algoritmos definitivos para a otimização da recomendação de OAs e da busca semântica por recursos web anotados em RDF. Então, esta abordagem poderá ser combinada com os trabalhos no estado da arte que recomendam tipos diferentes de OAs, definidos por Graf, Kinshuk e Ives (2010), para adaptar o processo de ensino aos estilos de aprendizagem e cognitivo de cada aprendiz.

Portanto, a recomendação e criação otimizadas de OAs modeladas por esta abordagem permitem que o usuário crie um curso online dando como entrada apenas o programa de curso. O usuário também pode entrar com informações opcionais usadas na filtragem de OAs ou criar seus próprios OAs, os quais podem ser utilizados por outros profissionais da Educação. Assim, esta abordagem mostra-se uma solução promissora para a efetivação do uso de tecnologias na Educação.

Agradecimentos

Os autores agradecem à Pró-Reitoria de Pesquisa e Pós-Graduação por todo o suporte prestado no desenvolvimento deste trabalho.

Referências

- ADVANCED DISTRIBUTED LEARNING INITIATIVE. Sharable Content Object Reference Model (SCORM™). **Advanced Distributed Learning**, 2001.
- ARAÚJO, R. D. et al. Autoria automática de objetos de aprendizagem a partir de captura multimídia e associação a estilos de aprendizagem. In: Anais do Simpósio Brasileiro de Informática na Educação, **Anais...**2014.
- BERNERS-LEE, T. et al. The semantic web. **Scientific american**, v. 284, n. 5, p. 28–37, 2001.
- CASTRO, C. V.; SATLER, M. F. Modelo ontológico para la secuenciación de objetos de aprendizaje. **Revista Iberoamericana de Tecnologías del/da Aprendizaje/ Aprendizagem**, p. 63, 2010.
- DI MARTINO, B. Semantic retrieval of learning objects with schema matching. **Journal of e-Learning and Knowledge Society**, v. 5, n. 3, 2009a.
- DI MARTINO, B. Semantic web services discovery based on structural ontology matching. **International Journal of Web and Grid Services**, v. 5, n. 1, p. 46–65, 2009b.
- DING, L. et al. Swoogle: a search and metadata engine for the semantic web. In: Proceedings of the thirteenth ACM international conference on Information and knowledge management, **Anais...ACM**, 2004.
- DUVAL, E. 1484.12. 1: IEEE Standard for Learning Object Metadata. **IEEE LTSC**, 2002.
- FININ, T. et al. Swoogle: Searching for knowledge on the Semantic Web. In: Proceedings of the National Conference on Artificial Intelligence, **Anais...Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005.**
- FRANZ, T. et al. **Triplerank: ranking semantic web data by tensor decomposition**. Berlin: Springer, 2009.
- GAETA, M. et al. An Approach To Personalized e-Learning. **Journal of Education, Informatics & Cybernetics**, v. 11, n. 1, 2013.
- GRAF, S.; KINSHUK; IVES, C. A Flexible Mechanism for Providing Adaptivity Based on Learning Styles in Learning Management Systems. In: 2010 IEEE 10th International Conference on Advanced Learning Technologies (ICALT), **Anais...** In: 2010 IEEE 10TH INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES (ICALT). jul. 2010.
- JÁCOME JÚNIOR, L.; MENDES NETO, F. M.; DA SILVA, L. C. N. Uma Abordagem Baseada em Algoritmo Genético para Recomendação de Objetos de Aprendizagem Sensível ao Contexto do Estudante. In: Anais do Simpósio Brasileiro de Informática na Educação, **Anais...**2012.

- MENOLLI, A.; MALUCELLI, A.; REINEHR, S. Criação Semi-Automática de Objetos de Aprendizagem a partir de Conteúdos da Wiki. In: Anais do Simpósio Brasileiro de Informática na Educação, **Anais...**2011.
- SILVA, D. H.; DORÇA, F. A. Uma Abordagem Automática para Personalização do Processo de Ensino Baseada em Estilos de Aprendizagem em Sistemas Adaptativos e Inteligentes para Educação a Distância. **Revista Brasileira de Informática na Educação**, v. 22, n. 2, p. 1–15, 2014.
- VICCARI, R. et al. The OBAA proposal for learning objects supported by agents. In: Procs. Of MASEIE Workshop–AAMAS, **Anais...**2010.
- WANG, R. et al. Re-ranking search results using semantic similarity. In: Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on, **Anais...**IEEE, 2011.
- WEI, W.; BARNAGHI, P.; BARGIELA, A. Rational research model for ranking semantic entities. **Information Sciences**, v. 181, n. 13, p. 2823–2840, 2011.

Concepção de Objetos Virtuais de Aprendizagem como uma Prática Multidisciplinar proveniente de Pressupostos da Computação

Cristiane Ellwanger, Cristina Paludo Santos

Universidade Regional Integrada do Alto Uruguai e das Missões (URI)
98.802-470 – Santo Ângelo – RS – Brasil

cristianeellwanger@gmail.com, paludo@san.uri.br

Abstract. *The conception of Virtual Learning Objects is similar the software development process, consequently knowledge derived from the field of computing can be used in its development. In this perspective, this article presents CAHES - a hybrid model structured by concepts arising out in the areas of Software Engineering, IHC and Decision-Making that establish as outline addressed to education professionals in the building of these objects.*

Resumo. *A concepção de Objetos Virtuais de Aprendizagem equipara-se ao processo de desenvolvimento de software, consequentemente conhecimentos advindos da área da computação podem ser empregados em seu desenvolvimento. Sob esta perspectiva, este artigo apresenta CAHES - um modelo híbrido estruturado por conceitos advindos das áreas de Engenharia de Software, IHC e Tomada de decisão e se estabelece como um arcabouço teórico-prático direcionado à auxiliar profissionais da educação na construção destes objetos.*

1. Introdução

A disseminação de tecnologias de comunicação e informação amplia as alternativas para a concepção de objetos virtuais de aprendizagem (OAs). No entanto, novas exigências surgem e contemplam a necessidade de se verificar não somente as tecnologias que dão suporte ao desenvolvimento de tais objetos, mas também o entendimento do contexto pedagógico em que os mesmos serão utilizados e em como associá-las a processos cognitivos dos alunos para apoiá-los em situações de aprendizagem.

Neste cenário, trabalhos da comunidade brasileira de Informática na Educação procuram contribuir na produção de OAs estabelecendo aspectos importantes como a sistematização de métodos e de técnicas de produção e de avaliação; preocupação com aspectos pedagógicos e técnicos; desenvolvimento de trabalho colaborativo, etc. Percebe-se, na revisão desses trabalhos, o desafio de lidar com a complexidade daquilo que se propõem a endereçar [Behar 2009].

Apesar das várias iniciativas ainda identifica-se uma série de dificuldades dos educadores no que se refere a quais direcionamentos seguir na concepção de OAs. Percebe-se também pouca autonomia de professores para a produção de conteúdo digital ou mesmo a produção com ênfase apenas em sua percepção de educador [Melo 2012]. Assim, a produção de OAs pode se tornar uma tarefa bastante complexa considerando que estes profissionais carecem de conhecimentos sobre modelos e métodos que podem ser aplicados para este fim.

Imerso neste contexto, este trabalho visa corroborar para a melhoria do processo de construção de objetos de aprendizagem, por meio de um modelo denominado CAHES. O modelo agrega em si perspectivas provenientes da Engenharia de Software (ES), da Interação Humano-Computador (IHC) e de Métodos e Ferramentas para a Tomada de Decisão (MFTD) se estabelecendo como um modelo híbrido dada a integração de áreas de conhecimento a ele vinculadas.

O modelo compreende diferentes fases que direcionam o processo de concepção e instigam os educadores a pensar em aspectos muitas vezes negligenciados na produção de OA. Uma descrição mais detalhada do modelo CAHES é apresentada nas seções subsequentes. A seção 2 apresenta os principais pressupostos teóricos que subsidiaram o desenvolvimento do modelo, bem como apresenta em linhas gerais alguns trabalhos relacionados. Na seção 3 são descritas as principais características do modelo CAHES. A seção 4 apresenta o estudo de caso realizado e, por fim, a seção 5 apresenta as considerações finais.

2. Pressupostos Teóricos e Trabalhos Relacionados

A seguir são apresentados os principais conceitos e trabalhos que subsidiaram o desenvolvimento do modelo CAHES.

2.1. A Engenharia de Software no desenvolvimento de Objetos de Aprendizagem

A produção de um OA é bastante complexa e vários fatores devem interagir de modo a atingir os objetivos desses produtos, tornando-se mandatário o uso de metodologias para organizar o processo de desenvolvimento [Steinmetz 2016]. Neste sentido a área de Engenharia de Software tem muito a contribuir, visto que conta com diferentes metodologias para o desenvolvimento de um software, dentre as quais destacam-se as metodologias ágeis que priorizam o equilíbrio entre a flexibilidade e a estrutura, sendo capazes de lidar com requisitos emergentes.

Esse tipo de metodologia se torna adequada para softwares de pequeno a médio porte que necessitam de constante modificação. Uma engenharia excessivamente demorada tornaria o produto final obsoleto já que seus requisitos seriam modificados num curto intervalo de tempo. Da mesma forma, uma abordagem com foco maior na concretização do software e menor na documentação, se torna mais adequada para o trabalho com objetos de aprendizagem.

Para a construção de OAs existem trabalhos realizados na tentativa de absorver o melhor de cada método a fim de priorizar o quesito de reusabilidade do objeto. Dentre os quais destaca-se neste trabalho o modelo GAIA PDOA [Pacheco et al, 2008] que atende aos conceitos e características que fundamentam a construção de um OA, empregando as melhores práticas das metodologias RUP (*Rational Unified Process*) e Scrum e seguindo preceitos do PMBOK (*Project Management Body of Knowledge*).

O ciclo de vida de tal modelo, utilizado como base para a concepção do modelo CAHES, estrutura-se em 4 fases específicas: (a) Iniciação – compreende o levantamento de requisitos com base no conteúdo pedagógico; (b) Planejamento - consiste nos processos de análise dos *sprints* e de prototipação do design e fatores ergonômicos; (c) Desenvolvimento – onde são realizadas a codificação, preparação dos fatores técnicos

(banco de dados e linguagens de programação), acompanhado de testes unitários e integração dos módulos e testes do código fonte em pequenas partes e, (e) Transição - consiste na apresentação do produto final ao *product owner*, que deverá fazer um teste piloto a fim de avaliar o OA.

2.2. IHC e sua Influência na Concepção de Objetos de Aprendizagem

Além dos aspectos e métodos advindos da área de Engenharia de Software que corroboram para o desenvolvimento de OAs, há que se considerar também os aspectos de interface. Para que um OA cumpra seu objetivo principal que é facilitar a relação ensino-aprendizagem de um conteúdo pedagógico, além do design instrucional, ele deve ter uma interface que propicie a aprendizagem. Logo, torna-se necessária a aplicação de conceitos relacionados com interface humano-computador.

Enquanto a engenharia de software foca no funcionamento dos processos do sistema, a IHC se volta para o que ocorre fora e através dele, ou seja, de que maneira essa interação pode ser projetada a fim de otimizar a utilização desse software. O usuário não deve ser obrigado a ajustar-se ao sistema, mas sim o contrário [Baranauskas 2003].

A maneira como o indivíduo interage com o computador e como este se adequa à interface leva a considerar aspectos cognitivos no momento do desenvolvimento de um software. Ao confrontar os recursos disponíveis atualmente, tais como, softwares e objetos educacionais, hipertextos, sites, ambientes educacionais, entre outros, com os princípios das teorias cognitivas, percebe-se em muitos casos um descompasso entre o que é apresentado ao educando e o processo cognitivo humano. Este descompasso se deve em muito dos casos, ao fato dos recursos tecnológicos chamarem a atenção pelo volume de informações, animações, sons e imagens, que apesar de tornar a interface e a interação mais atraentes não se preocupam com a sobrecarga cognitiva, o que pode vir a comprometer o aprendizado.

A teoria da carga cognitiva aporta contribuições que permitem melhorar a qualidade do projeto instrucional, ou seja, criar materiais educativos digitais mais adequados ao processo cognitivo humano, utilizando recursos de forma equilibrada e de maneira que venha a auxiliar o processo de aprendizagem do aluno. Esta teoria também reforça que os novos conhecimentos sejam apresentados aos alunos em quantidade alinhada ao seu processo cognitivo de modo que o processo de aprendizagem possa ocorrer de maneira natural e sem sobrecarga.

Considerando os estudos já realizados sobre aspectos como atenção, percepção, memória, aprendizado, leitura e fala, resolução de problemas entre outros, oriundos da Teoria da Cognição e abordados na engenharia cognitiva dentro da área de IHC [Baranauskas 2003][Preece 2013], vislumbra-se a possibilidade de integração destes na construção do modelo CAHES.

2.3. A Tomada de Decisão e o Método *Analytic Hierarchy Process* (AHP)

O processo de tomada de decisão pode ser definido como a forma que os indivíduos agem quando uma ação decisória é demandada, no qual se deve levar em consideração a análise e o julgamento das informações disponíveis, além de sua associação com o ambiente em que o indivíduo está inserido.

O processo de tomada de decisão não é simples, uma vez que a grande quantidade de informações exige um tratamento individual buscando transformá-las em conhecimento e, desta forma, utilizar esse conhecimento no processo pretendido. Neste cenário, a crescente compreensão das variações do comportamento humano e o avanço tecnológico que respalda e simula processos cognitivos melhoram, em muitas situações, a tomada de decisão.

Dentre os métodos que auxiliam a tomada de decisões em cenários complexos destaca-se o método AHP que transforma decisões complexas em decisões par a par, levando em conta aspectos qualitativos e quantitativos, ou seja, tanto aspectos matemáticos quanto psicológicos, de forma que decisões complexas que envolvam a preferência subjetiva do indivíduo possam ser bem estruturadas numa hierarquia de apoio à decisão [Saaty 2008].

Por meio desse método é possível analisar uma hierarquia de critérios transformando as comparações em valores numéricos que são processados e comparados. Da comparação se chega ao peso relativo e com este a probabilidade numérica, que demonstra a probabilidade de se atender à meta estabelecida. A partir da definição do objetivo geral se começa a construir a árvore hierárquica, colocando em cada nível de hierarquia os critérios a serem comparados entre si. Através da estrutura formada pela utilização do método, o decisor pode obter tanto uma visão específica de parte do problema como uma visão geral dependendo do modo como o problema é estruturado e como cada critério afeta a decisão.

No contexto educacional, pode-se entender a escolha pela metodologia de ensino e recursos pedagógico como um exemplo de decisões complexas a medida que a oferta de material cresce a todo momento. O peso de cada decisão e o foco no que se deseja atingir com cada escolha feita, é algo que requer um método de tomada de decisão capaz de gerar uma reflexão sobre os prováveis resultados.

2.4. Trabalhos Relacionados

Vários trabalhos são apresentados na literatura se direcionando a concepção de objetos virtuais de aprendizagem. Battistela e Wangenheim (2011) trazem a análise de seis ferramentas de autoria passíveis de criar objetos de aprendizagem no padrão SCORM (*Sharable Content Object Reference Model*) e as avalia de acordo com critérios de usabilidade e utilidade através de heurísticas. Como resultado, os autores apontam as ferramentas *eXe Learning* e a *CourseLab* com as melhores avaliações. Battistela (2009) amplia o escopo da pesquisa destacando a classificação de OAs. Apesar do modelo CAHES não recomendar uma ferramenta de autoria específica, tais trabalhos foram usados como base para constituir o estudo de caso em que o modelo foi aplicado.

Já, o trabalho de Pacheco (2008) traz a visão da engenharia de software voltada para a concepção dos objetos, analisando diversos modelos e buscando os pontos fortes de cada um, comparando-os e gerando o modelo GAIA PDOA (GAIA Processo para Desenvolvimento de Objetos de Aprendizagem), utilizado como base na concepção do CAHES. Porém, seu ciclo de vida contempla partes de desenvolvimento que não seriam necessariamente utilizadas com o uso de uma ferramenta de autoria, por isso o CAHES propõe uma simplificação dessas fases.

(d) A fase de transição muda o foco de apresentar o resultado ao *product owner*, visto que no CAHES o objeto foi construído pelo próprio. Assim, apresenta-se o objeto ao usuário final, ou seja, o educando.

No que diz respeito à área de IHC explorou-se as contribuições advindas dos estudos sobre processos cognitivos e suas influências na concepção de interfaces a fim de agregar ao método características importantes que devem ser considerados pelo educador na concepção de OAs. Quanto à tomada de decisão, utilizou-se o método AHP a fim de permitir que tanto decisões com poucas variáveis quanto decisões complexas, envolvendo a parte de planejamento, sejam contempladas pelo modelo. Desta forma, o usuário do modelo poderá também criar objetos mais complexos, que envolvam diversos requisitos pedagógicos e/ou processos cognitivos.

Assim como o modelo GAIA, o CAHES possui quatro fases distintas em seu ciclo de vida: Iniciação, Planejamento, Desenvolvimento e Transição. A ênfase neste modelo está na fase do planejamento, foco da reflexão sobre o objeto de aprendizagem final que se pretende criar. O ciclo de vida do modelo CAHES é apresentado na Figura 2.

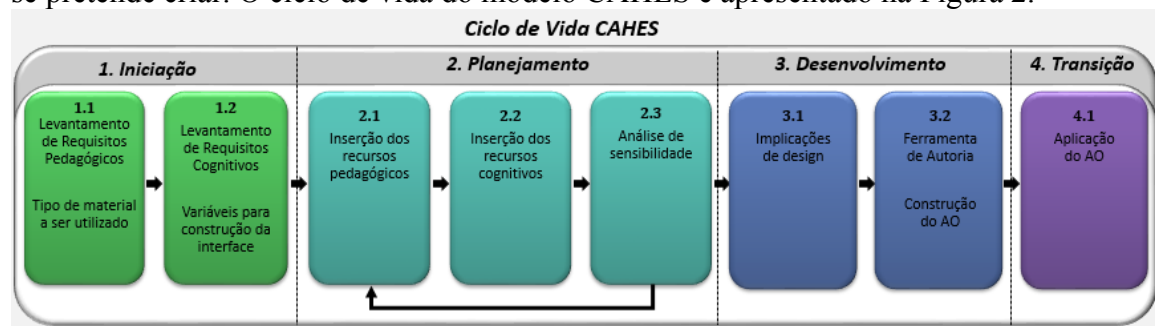


Figura 2. Ciclo de vida do Modelo CAHES

Na fase de **Iniciação** os recursos pedagógicos e requisitos cognitivos a serem considerados na concepção do AO devem ser definidos, servindo como subsídios para o processo de tomada de decisão. Como recursos pedagógicos, definidos na subfase 1.1, entende-se todo e qualquer tipo de material que o docente deseja utilizar dentro do seu OA. Exemplos incluem textos, vídeos, áudios, animações diversas, questões de múltipla escolha, etc. Já, a subfase 1.2 compreende a definição de quais e quantos requisitos cognitivos são relevantes na concepção do OA em questão com vistas a alcançar os objetivos almejados.

A quantidade de processos cognitivos a serem utilizados pode variar. Uma atenção especial nestas definições deve ser dada e direcionada, sobretudo, ao ambiente visual que será propiciado ao usuário final, já que as implicações dessas variáveis cognitivas afetam o design do objeto de aprendizagem e não o conteúdo pedagógico. A Tabela 1 apresenta uma síntese dos processos cognitivos considerados pelo CAHES e as implicações de cada processo na interface conforme Preece (2013).

A fase de **Planejamento** compreende o uso do AHP e a inserção dos recursos elicitados na fase de Iniciação. Nesta fase a inserção dos recursos pedagógicos é o início do processo de tomada de decisão com o método AHP (subfase 2.1). Nessa etapa, os recursos pedagógicos são avaliados no AHP e atribuído o seu respectivo valor nas comparações par a par. Esse valor pode ser atribuído considerando a importância do recurso mediante o objeto de aprendizagem como um todo ou a predominância do

recurso nesse objeto. O uso do método AHP neste contexto é pertinente, pois com a comparação em pares o usuário é instigado a avaliar a preferência que se tem por cada recurso pedagógico e processo cognitivo.

Tabela 1. Processos cognitivos e implicações na interface

PROCESSO COGNITIVO	IMPLICAÇÕES NA INTERFACE
ATENÇÃO - Consiste no processo de selecionar coisas em que se concentrar, em certo momento, dentre uma variedade de possibilidades disponíveis	Saliente informação necessária; técnicas como gráficos animados, cores, sublinhado, ordenação de itens, sequenciamento de informações, espaçamentos; limite a quantidade de informações na interface; opte pela simplicidade para a busca de informações;
PERCEPÇÃO - Refere-se a como a informação é assimilada pelos diferentes órgãos sensitivos e transformada em experiência.	Facilitar a identificação de significados em ícones e imagens; utilizar bordas e espaçamento para facilitar percepção e localização; Uso de sons audíveis e distinguíveis; Texto legível e distinguível do fundo;
MEMÓRIA – Envolve recordar vários tipos de conhecimento que nos permitem agir adequadamente	Os procedimentos devem ser simples para evitar sobrecarga de memória; Interfaces que promovam o reconhecimento ao invés da recordação; Utilização de categorias, cores, sinalizações, etc.;
APRENDIZAGEM – Como usar uma aplicação baseada em computador para entender um determinado tópico.	Incentive a exploração; restrinja e guie os usuários na seleção de ações; Ligação dinâmica entre representações concretas e conceitos abstratos;
LER, FALAR e OUVIR – Formas de processamento de linguagem	Opções de menu e instruções faladas devem ser mínimas; acentuar entonação das vozes geradas artificialmente; opção de ampliar o texto;
RESOLUÇÃO DE PROBLEMAS, PLANEJAMENTO, RACIOCÍNIO E TOMADA DE DECISÃO – Pensar sobre o que fazer, quais são as opções e quais as consequências de se realizar uma determinada ação.	Informações adicionais que fiquem escondidas, mas sejam fáceis de acessar; utilização de funções fáceis e simples de lembrar;

O mesmo processo deve ser realizado considerando as variáveis cognitivas e sua inserção no método AHP (subfase 2.2). Para cada processo cognitivo deve ser atribuído um peso, conforme a Escala de Saaty (2008), de acordo com o recurso pedagógico a qual ele está associado. A realização das ponderações de pesos e demais cálculos oriundos do método AHP para a tomada de decisão podem ser realizados com ou sem o auxílio de uma ferramenta.

Na subfase 2.3 da fase de Planejamento são verificados se os resultados obtidos são compatíveis com o esperado pelo educador ou se precisa de uma reavaliação. Essa etapa é fundamental para analisar o real impacto dos valores dados a cada recurso, pedagógico e cognitivo e como isso afetará a interface do objeto. Quando o resultado não é compatível com o esperado ou há muita discrepância de valores, o usuário deve repensar sua avaliação em relação a cada etapa do método, retornando ao início da fase de Planejamento. Com isso, as reflexões sobre os resultados ocorrem antes da fase de Desenvolvimento, evitando esforço desnecessário na implementação do AO e oportunizando a realização de ajustes, caso necessário.

A Figura 3 apresenta exemplos de gráficos gerados na fase de Planejamento. A interpretação dos gráficos pode variar, até mesmo porque as implicações de design

podem, muitas vezes, se complementar. Porém, ao se agregar um valor aos recursos pedagógicos e às variáveis cognitivas, define-se uma hierarquia mantendo no resultado final a importância de cada item analisado.

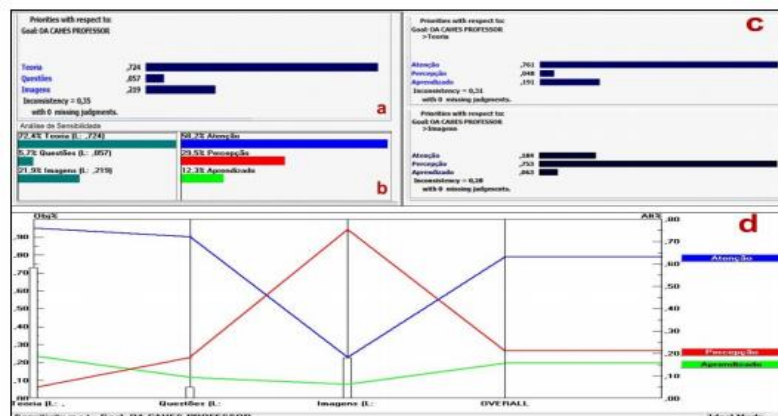


Figura 3. Gráficos gerados na fase de Planejamento

A terceira etapa compreende a fase de *Desenvolvimento* e envolve a construção do OA em si. Para o desenvolvimento consideram-se os resultados obtidos nas fases de Iniciação e, sobretudo, de Planejamento. É importante destacar que a aplicação do modelo CAHES não está limitada ao uso de uma ferramenta específica, ficando a critério do usuário a escolha da ferramenta. Por último, tem-se a fase de aplicação do modelo, denominada *Transição*, na qual se aplica o objeto de aprendizagem finalizado aos seus usuários alvo.

O que deve prevalecer no resultado do modelo são as características da variável cognitiva de maior destaque dentro da análise. O restante da execução do modelo é a criação do objeto em si, com a ferramenta de autoria escolhida pelo usuário e a aplicação do objeto em um contexto real de ensino.

4. Estudo de Caso

O cenário de aplicação do modelo contou com a colaboração de 10 educadores de diferentes disciplinas. No primeiro momento os professores foram incitados a criar OAs com base nos seus próprios conhecimentos. Já, no segundo, uma explanação sobre o modelo CAHES foi realizada e os educadores foram motivados a repensar o OA concebido considerando os pressupostos do modelo. Para demonstrar a aplicação do modelo são apresentados os OAs concebidos por um dos educadores que atua há mais de 20 anos no ensino das disciplinas de Ciências e Biologia.

Na fase de Iniciação o assunto-chave definido pelo docente foi “divisão celular” sendo definida a utilização dos recursos (1) Teoria, consistindo em explicação textual pura sobre o assunto; (2) Questões de múltipla escolha; (3) Imagens ilustrativas relacionadas ao tema. Quanto aos processos cognitivos o educador escolheu as variáveis (1) Atenção, (2) Percepção e (3) Aprendizado.

Na fase de Planejamento ocorre a classificação dos recursos em que se define a quantidade de cada recurso que se pretende colocar no OA, bem como a importância de cada um deles. Tal classificação é feita a partir da atribuição de pesos que serão utilizados

para os cálculos e comparações realizados pelo método AHP. Esta fase foi executada diretamente na ferramenta *Expert Choice 11*. A árvore hierárquica resultante da definição de recursos e processos cognitivos é apresentada da Figura 4.

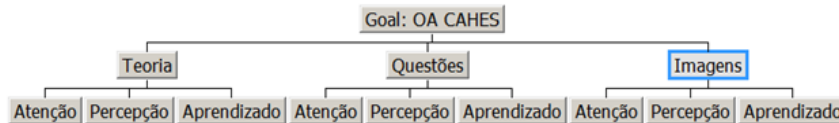


Figura 4. Árvore hierárquica do processo de tomada de decisão.

Com base nos pesos atribuídos é realizada uma comparação par a par resultando em dados que permitem avaliar o quanto um recurso deve ser aplicado na criação do OA. As variáveis cognitivas são comparadas em relação a cada recurso pedagógico sendo os resultados da comparação apresentado em forma de gráfico (Figura 5).

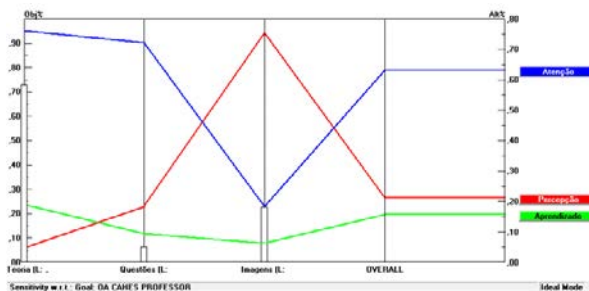


Figura 5. Gráfico de Performance

Para a fase de desenvolvimento, utilizou-se a ferramenta de autoria *CourseLab 2.4*. A Figura 6 apresenta os OAs criados pelo educador. Na OA apresentado na Figura 6(a), sem a aplicação do CAHES, não houve preocupação com a quantidade de informação na interface, baseando-se em excesso de informação em forma de texto sem destaque das palavras principais e apenas uma imagem para relacionar o conhecimento abstrato à representação concreta do mesmo. Já, o OA da Figura 6(b), concebido a partir da aplicação do modelo CAHES, emprega as recomendações apresentadas na Tabela 1, com palavras-chaves em destaque, informações divididas em diversas telas a fim de não sobrecarregar a interface, imagens que se relacionam ao conteúdo teórico, espaçamentos maiores, etc.



Figura 6. Objetos de Aprendizagem criados pelo educador

5. Considerações Finais e direcionamentos futuros

Um modelo de ciclo de vida com fases definidas pode ser adotado por profissionais de outras áreas para a concepção de objetos de aprendizagem, dada a sua facilidade de aplicação e entendimento. Associado a isso, a integração de qualidade aos objetos desenvolvidos, a partir da adoção de uma metodologia ágil como a *Scrum*, possibilita que conceitos provenientes da computação a engenharia de software, interação humano-computador e tomada de decisão possam ser integrados aos objetos desenvolvidos.

A utilização em conjunto desses conhecimentos mostrou-se não apenas um desafio, mas um interessante experimento de interdisciplinaridade, não somente por buscar atribuir uma melhor qualidade aos objetos de aprendizagem desenvolvidos, mas também no sentido de subsidiar profissionais da educação com conhecimentos e técnicas que, na maioria das vezes, não fazem parte do seu cotidiano.

Os trabalhos futuros remetem a se contemplar aspectos de acessibilidade ao modelo CAHES e na verificação do impacto dos processos cognitivos nas interfaces projetadas e sua viabilidade de aplicação no ensino fundamental, médio e superior.

Referências Bibliográficas

- Baranauskas, M. C. C.; Rocha, H. V. Design e avaliação de interfaces humano-computador. Campinas: NIED/UNICAMP, 2003.
- Barbosa, Simone Diniz Junqueira; Silva, Bruno Santana da. Interação Humano-Computador. Rio de Janeiro: Elsevier, 2010.
- Battistella, Paulo Eduardo; Wangenheim, Aldo Von. Avaliação de ferramentas de autoria gratuitas para produção de objetos de aprendizagem no padrão SCORM. Revista Brasileira de Informática na Educação, 2011.
- Battistella, Paulo Eduardo; Wangenheim, Aldo Von. Classificação de Objetos de Aprendizagem e análise de Ferramentas de Autoria. XX Simpósio Brasileiro de Informática na Educação, Florianópolis –SC, 2009.
- Behar, P. A.; Waquil, M. P. "Princípios da pesquisa científica para investigar ambientes virtuais de aprendizagem sob o ponto de vista do pensamento complexo." *Modelos pedagógicos de educação a distância*(2009): 146-178
- Melo, Amanda Meincke, J. F. Saldanha, and Maria Cristina Graeff Wernz. "Desafios à pesquisa em Computação em contexto educacional–qualidade no uso de objetos de aprendizagem em perspectiva." *Anais do Workshop de Desafios da Computação Aplicada à Educação*. 2012.
- Pacheco, Edson et al., GAIA PDOA – Processo Para Desenvolvimento de Objetos de Aprendizagem. (2012).
- Preece, J; Rogers, Y., Sharp, H. Design de Interação: Além da interação humano-computador. Porto Alegre: Bookman. (2013).
- Saaty, T. L. (2008). *Relative Measurement and its Generalization in Decision Making*. Madrid: Review of the Royal Spanish Academy of Sciences, Series A, Mathematics. Disponível em: <<http://www.rac.es/ficheros/doc/00576.pdf>>. Acesso em Outubro/2015.
- Steinmetz, Ernesto Henrique Radis. "A contribuição da arquitetura da informação na construção e utilização de ambientes informacionais colaborativos de ensino/aprendizagem." (2016).

Análise da Complexidade Ciclomática como Apoio ao Processo de Desenvolvimento do Pensamento Algorítmico

Vinicius Ajala, Denilson Rodrigues da Silva, Cristina Paludo Santos, Carlos Oberdan Rolim

Universidade Regional Integrada do Alto Uruguai e das Missões (URI)
98.802-470 – Santo ângelo – RS – Brasil

viniciusajala@outlook.com,
{deniro,paludo}@san.uri.br,oberdan@gmail.com

***Abstract.** One of the premises to establish the basis of algorithmic thinking is a concern for the best possible solution for a given problem. This paper proposes a descriptive computational model, which applies cyclomatic complexity to help students develop algorithmic solutions through demonstration of how far was their solution regarding the reference solution considered. As contributions of this study, we can highlight a tool designed from the model proposed with a view to strengthening teaching methods, theory sedimentation and achievement of better results in learning*

***Resumo.** Uma das premissas para se estabelecer a base do pensamento algorítmico é a preocupação pela melhor solução possível para determinado problema. O presente trabalho propõe um modelo computacional descritivo que aplica complexidade ciclomática para auxiliar o aluno a desenvolver soluções algorítmicas através da demonstração do quão distante esteve a sua solução com relação à solução considerada referencial. Como contribuições do trabalho pode-se destacar uma ferramenta concebida a partir do modelo proposto com vistas ao fortalecimento dos métodos de ensino e a sedimentação da teoria e obtenção de melhores resultados no aprendizado.*

1. Introdução

Os desafios da Informática na Educação são muitos, inclusive na própria formação de competências e habilidades de acadêmicos de cursos de computação [Bombasar 2015]. Dentre tais desafios destaca-se a constituição do pensamento computacional e algorítmico que, segundo Wing (2014) apresenta-se como uma das mais importantes contribuições da ciência da computação para o mundo, pois permite que os profissionais pensem nos problemas de forma analítica e desenvolvam soluções computacionais para os mais diversos domínios de aplicação.

Embora os princípios do pensamento computacional e algorítmico sejam explorados em diversas disciplinas curriculares na área de Computação [Silva 2015], sua maior dificuldade encontra-se no contato de acadêmicos em estágios iniciais de um curso de computação com o estudo de algoritmos, o raciocínio lógico e a programação. Logo, o desafio está em desenvolver modelos e ferramentas computacionais que auxiliem o processo de aprendizagem desde a fase inicial, contribuindo para despertar no aluno seu espírito crítico e estabelecendo estratégias para organizar as ideias, o

raciocínio e a representação simbólica, a fim de que o aluno incorpore esses hábitos e os utilize na resolução de problemas durante a trajetória acadêmica e, posteriormente, como profissional.

Nesse contexto, o presente trabalho propõe um modelo computacional descritivo que aplica complexidade ciclomática para auxiliar o aluno a desenvolver soluções algorítmicas através da demonstração do quão distante está a sua solução com relação à solução considerada referencial. A elaboração deste modelo computacional busca, de forma didática, tornar o ensino do conteúdo abordado mais prático e abrangente, de forma a despertar o interesse do aluno, o seu espírito de pesquisa e a busca de informações que possam torná-lo um profissional crítico e de opinião sólida. Como contribuições do trabalho pode-se destacar a ferramenta concebida a partir do modelo descritivo com vistas ao fortalecimento dos métodos de ensino e a sedimentação da teoria e obtenção de melhores resultados no aprendizado.

O trabalho está organizado da seguinte forma: a seção 2 apresenta os principais conceitos que nortearam o desenvolvimento do modelo, bem como alguns trabalhos relacionados que serviram de suporte para as decisões de projeto. A seção 3 apresenta o modelo descritivo proposto e a seção 4 expõem as principais características e interfaces da ferramenta concebida a partir do modelo. Por fim, a seção 5 apresenta as considerações finais e perspectivas de trabalhos futuros.

2. O Pensamento Algorítmico e a Complexidade Ciclomática

Um dos maiores desafios que se impõe ao aluno durante o processo de formação acadêmico profissional é o de deixar de ser um mero utilizador ou executor de determinados programas e passar a ser criador. Para tal, antes de ver o mundo como um conglomerado de desafios computacionais é necessário visualizar e assimilar os problemas como uma estrutura que pode ser solucionada, isto é, o aluno precisa desenvolver suas faculdades mentais de forma a tornar-se um indivíduo de pensamento crítico e sólido, para depois então dirimir os problemas da sociedade com sua especialidade. É neste contexto que surge a aptidão que é vista pelos pesquisadores de educação em computação como imprescindível – o pensamento algorítmico [Souza e Lencastre 2014].

Uma das premissas para se estabelecer o pensamento computacional é a preocupação pela melhor solução possível. Esta é a base do pensamento algorítmico que expressa soluções em diferentes passos de forma a encontrar uma saída mais eficiente e eficaz de resolver um problema [Resnick 2012]. O pensamento algorítmico é um ramo do pensamento computacional, sendo uma representação mais abstrata da forma de se pensar. Não é ensinado diretamente na graduação como uma área de estudo, mas sim de forma indireta quando se coloca o aluno em situações de aprendizagem de lógica de programação por meio de fatos do cotidiano, para depois formular uma solução através de procedimentos computacionais. Em outros termos, segundo Cuny (2010) este pensamento uma vez desenvolvido no sujeito, torna-o capaz de identificar problemas e solucioná-los de forma objetiva e eficaz de acordo com as necessidades dispostas.

2.1. Análise da Complexidade Ciclomática

A complexidade ciclomática é uma métrica cujo objetivo é medir a complexidade de determinado código fonte. Quanto maior o seu valor, maior a dificuldade de se entender, modificar e, conseqüentemente, testar o código fonte [McCabe 1976]. De forma a facilitar a visualização do fluxo de um código fonte, McCabe utiliza o conceito de Grafo de Fluxo de Controle como forma de representação, composto por vértices que fazem referência a blocos ou linhas de código fonte, e arestas que conectam os vértices formando caminhos do fluxo de execução [Allen 1970], identificando quais são os possíveis caminhos de execução de um algoritmo.

O grafo da complexidade ciclomática tem suas bases na teoria dos grafos e oferece uma métrica de software extremamente útil. Existem métodos, tanto analíticos quanto matemáticos para o cálculo da complexidade ciclomática [Pressman 2006]. Os métodos analíticos resultam em uma verificação visual mais detalhada, sendo grande base de apoio à didática. Já, o método matemático retorna um valor preciso, independente de análise visual, sendo o preferido para programas computacionais. Além disso, a complexidade ciclomática é usada para analisar aspectos de manutenibilidade e testabilidade de software. Em relação a manutenibilidade, o valor da complexidade gerado é analisado podendo assumir os valores baixo, médio ou alto, refletindo assim o nível de dificuldade na manutenção. Já quanto a testabilidade, faz-se uso de casos de teste conforme o valor ciclomático obtido [Sommerville 2003].

Neste trabalho, a complexidade ciclomática é empregada de forma inovadora no contexto de aprendizagem algorítmica, servindo de apoio no processo inicial de programação e constituição do pensamento algorítmico. Assim, a complexidade ciclomática como fator de qualidade na aprendizagem é a motivação deste trabalho a fim de iniciar uma proposta educacional de fortalecimento do pensamento algorítmico.

2.2. Trabalhos Relacionados

A complexidade ciclomática pode ser aplicada em diversos contextos. Cunha (2006) fez uso de sua aplicação como um instrumento para auxílio na testabilidade de banco de dados. Para tanto, implementou um sistema que calcula a complexidade ciclomática a partir de um código-fonte PL/SQL do Oracle. Seibt (2011) abordou o cálculo de métrica de software para códigos orientados a objetos codificados na linguagem Delphi. A ferramenta permite calcular métricas de projeto e de construção, como por exemplo, profundidade da árvore de herança. Faêda (2012) desenvolveu a ferramenta ComplexGraph, a fim de auxiliar na etapa de teste de software, em específico o teste unitário, gerando o grafo de fluxo de controle onde é possível identificar a quantidade de caminhos de execução contidos no código-fonte.

O modelo apresentado neste artigo assemelha-se em alguns pontos com os propostos por Cunha (2006), Seibt (2011) e Faêda (2012), principalmente no tratamento de leitura de códigos para cálculo da complexidade ciclomática. Nos três trabalhos o cálculo é efetuado extraindo de cada linha do código-fonte os *tokens*, que são uma seqüência de um ou mais caracteres contidos na gramática da linguagem. Para os *tokens* extraídos, são verificados se algum deles são palavras reservadas de estruturas que mudam o fluxo de execução, como por exemplo, um comando *if*. Caso existam estes

tipos de *tokens*, o sistema contabiliza a quantidade de suas ocorrências e retorna esta contagem como resposta do número ciclomático.

Entretanto, o modelo proposto difere dos demais trabalhos quanto ao foco e objetivos finais. Enquanto os citados propuseram ferramentas com o objetivo de mensurar códigos em uma linguagem específica por meio da métrica, o modelo proposto tem como objetivo despertar uma reflexão do aluno por intermédio de uma representação visual de diferentes soluções algorítmicas. Assim, o uso da métrica e seu valor ciclomático não buscam ser fatores decisivos na análise do aluno, mas sim o grafo de fluxo que é gerado.

3. Modelo Descritivo Proposto

Neste estudo a complexidade ciclomática é empregada de uma forma inovadora no contexto da aprendizagem algorítmica, servindo de apoio no processo inicial de programação e constituição do pensamento algorítmico. O trabalho atenta para o uso de indicadores quantitativos abordados para o aluno de forma qualitativa como, por exemplo, arestas e vértices representados por um grafo que visa refletir conceitos de qualidade de software na aprendizagem. Portanto, a complexidade ciclomática como fator de qualidade na aprendizagem é a motivação deste trabalho a fim de iniciar uma proposta educacional de fortalecimento do pensamento algorítmico.

O modelo proposto respalda-se nos princípios de praticidade e apoio à etapa de aprendizagem de programação básica, buscando refletir melhores soluções perante o aluno. De forma geral, a construção do modelo considera a ideia de que o aluno forneça a solução de um desafio a ele dado (algoritmo), que será analisado, considerando que exista uma base de conhecimento onde estarão os desafios e suas respectivas soluções (código-fonte). Após a análise é gerado o grafo ciclomático da solução do aluno e da solução armazenada na base de conhecimento, elucidando assim uma comparação gráfica que permite abstrair os conceitos de complexidade algorítmica.

Além disso, são retornados ao aluno os dois códigos (referencial e o do aluno) e alguns resultados quantitativos como a complexidade ciclomática, número de estruturas de seleção e repetição, que têm o propósito de auxiliar na compreensão das distinções, se houver, entre as soluções. A Figura 1 demonstra o ciclo de execução do modelo conforme Diagrama de Atividades da UML.

O modelo foi desenvolvido usando o *Model-view-controller* (MVC) - um padrão de arquitetura de software que separa a representação da informação da interação do aluno com ele. O modelo (*model*) consiste nos dados da aplicação. Uma visão (*view*) pode ser uma saída de representação de dados. É possível ter várias visões do mesmo dado, como um gráfico de barras para gerenciamento e uma visão tabular para contadores. O controlador (*controller*) faz a mediação da entrada, convertendo-a em comandos para o modelo ou visão. Justifica-se o uso deste padrão por questões de reusabilidade de código e separação de conceitos. Cabe ressaltar que o modelo pode ser implementado em uma linguagem estruturada, desde que seja feita uma adaptação das etapas do desenvolvimento e um estudo aprofundado para tratar cada componente do código.

Para a concepção do modelo as seguintes etapas fizeram-se necessárias: (a) definição de uma base de conhecimento; (b) extração de estruturas de desvio de fluxo do código-fonte; (c) definição dos componentes do grafo; (d) definição dos componentes externos para a geração dos grafos; (e) implementação do cálculo do número ciclômico e (f) comparação das soluções com retorno de indicadores ao aluno. Tais etapas compreendem:

a) Base de Conhecimento - Armazena uma coleção de códigos que se constitui em soluções apontadas na literatura como referências para determinados algoritmos. Esses algoritmos serão os mesmos que o aluno deverá solucionar, e partir do princípio que a solução esteja correta, isto é, as saídas do algoritmo sejam as desejadas respeitando as limitações impostas pelo enunciado, será feita uma comparação entre as duas soluções.

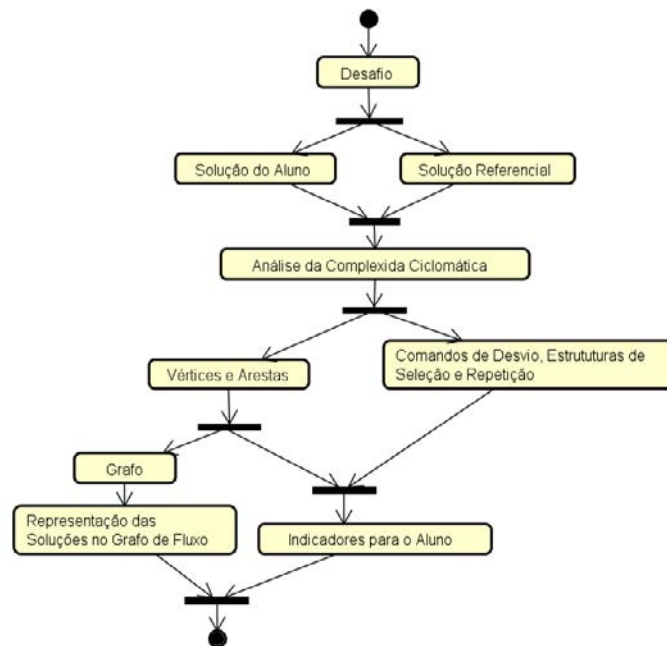


Figura 1. Diagrama de atividades do modelo

b) Extração das Estruturas de Desvio de Fluxo - Para construção do grafo da complexidade ciclômica e da resposta à solução do aluno faz-se necessário identificar nos códigos-fonte (do aluno e do referencial) as estruturas responsáveis por criar desvios de fluxo de execução do código. No modelo proposto foram consideradas as estruturas de repetição, constituídas de comandos como *for*, *while* e *do while*, e as estruturas condicionais, formadas por *if*, *else* e *case*.

c) Componentes para Geração do Grafo - Um dos requisitos funcionais mais importantes deste modelo é a geração do grafo da complexidade ciclômica. Assim, para geração do grafo utilizam-se os comandos de desvio extraídos do código, que passam a representar os vértices. Alguns comandos geram dois vértices, um com a sua denominação e outro acrescido da palavra *end* para simbolizar o fim da execução do comando de desvio. Os componentes necessários para formar o grafo são os vértices, as arestas e as coordenadas dos vértices. Cada comando de desvio é representado de forma

distinta no grafo, como mostra a Figura 2. No grafo os vértices representam os comandos de desvio e as arestas o fluxo de execução do código-fonte.

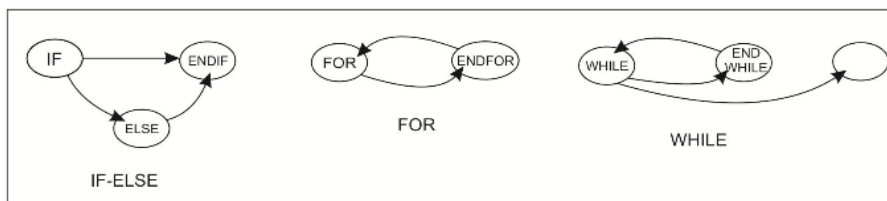


Figura 2. Notação do grafo da complexidade ciclomática

Uma vez encontrados os vértices, é possível definir as arestas do grafo. As arestas utilizadas no grafo são direcionadas, ou seja, representam o sentido do fluxo de execução do código. Além disso, as arestas dizem em quais vértices um determinado vértice irá se conectar. A Figura 3 apresenta a relação entre vértices e arestas, por exemplo, o comando *if* que consta na lista de vértices possui os números 6 e 2, isso significa que o comando irá ligar-se a sua esquerda ao vértice contido na posição 6 da lista e, a sua direita ao vértice da posição 2.

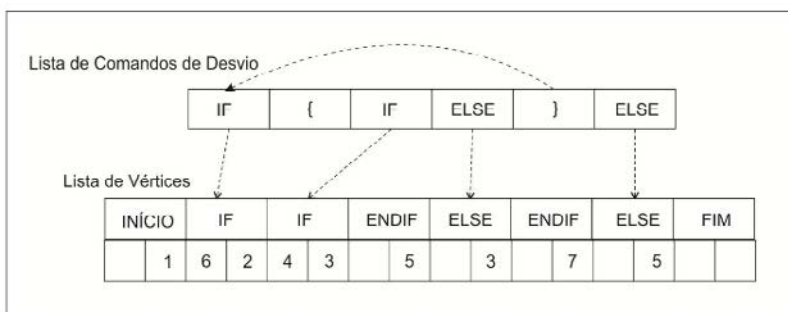


Figura 3. Relação entre vértices e arestas

d) Geração do Grafo - A etapa de geração do grafo da complexidade ciclomática se torna possível em virtude das informações extraídas do código-fonte. Com as informações dos vértices, arestas e as coordenadas x e y, que servem para garantir que um vértice não sobreponha outro, é possível plotar o grafo. O modelo foi concebido para a linguagem Java, uma vez que faz uso de uma biblioteca externa desta linguagem para criar os grafos.

O componente *mxGraph* foi utilizado por este constituir-se em uma biblioteca que permite a criação de forma ágil e prática de diferentes tipos de diagramas. A biblioteca também é eficiente quanto a personalização, permitindo mudar o layout do grafo, sendo possível alterar a fonte, a cor e o tamanho das bordas, a cor de fundo dos vértices, o formato do vértice, entre outros.

e) Cálculo da Complexidade Ciclométrica- Para o cálculo da complexidade ciclométrica utilizou-se a fórmula: $CC = A - V + 2$ [McCabe 1976], onde CC é o número ciclométrico, A é o número de arestas e V é o número de vértices do grafo. Este método foi escolhido por ser uma fórmula matemática, enquanto a maioria dos outros métodos são analíticos, assim facilitando sua implementação em um programa computacional e atendendo plenamente as solicitações do modelo.

f) Indicadores após Análise da Complexidade Ciclomática- Um dos principais objetivos ao desenvolver este modelo foi o de retornar ao aluno um resultado da comparação entre a solução do aluno e a solução referencial. Este resultado nada mais é que um indicador ao aluno que visa abstrair os conceitos de complexidade ciclomática ao nível de aprendizagem, isto é, de forma que o aluno aprenda a buscar uma solução mais eficiente do que a forma habitual de resolução.

3.1. Classes desenvolvidas

O modelo consiste em quatro classes principais: *CalculaCC*, *GeraEstrutura*, *AnalisaEstrutura* e *GeraGrafo*. A classe *GeraEstrutura* é responsável por varrer o código a procura de estruturas de desvio, *AnalisaEstrutura* é responsável por transformar as estruturas em vértices. As outras duas preocupadas na geração dos resultados, *GeraGrafo* implementa o componente que trata a geração dos grafos e *CalculaCC* gera os indicadores ao usuário. Também foram criadas classes intermediárias que auxiliam no processamento de informações, empregadas no pacote *Model*.

Na Figura 4 é representada por intermédio de um diagrama de blocos a estrutura principal do modelo, para as principais classes são detalhadas as instruções que compõem cada etapa. O relacionamento entre classe e instrução é representado por uma seta. Na validação deste modelo foi desenvolvida uma ferramenta computacional, usando da estrutura definida anteriormente para gerar os grafos e indicadores. As classes definidas anteriormente foram o ponto de partida do desenvolvimento, outras classes foram concebidas buscando tratar a interação com o usuário.

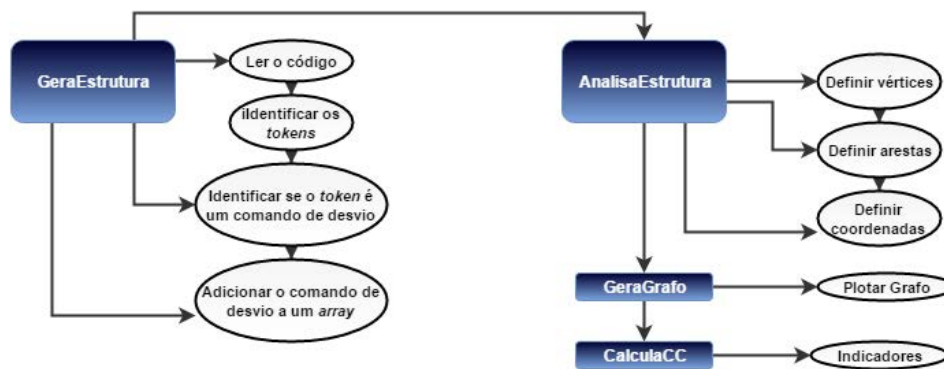


Figura 4. Representação estrutural do modelo

4. A Ferramenta CCEduc

As funcionalidades do modelo proposto foram materializadas na ferramenta CCEduc. Essa ferramenta foi codificada na linguagem Java e desenvolvida conforme o padrão de engenharia de software MVC. A Figura 5 apresenta o diagrama de classes utilizado. Do lado esquerdo estão as classes vinculadas ao Modelo conforme apresentado na seção 3.1 e do lado direito as classes relacionadas as funcionalidades da ferramenta em si.

Para fins de prova de conceito, optou-se por empregar o uso de códigos desenvolvidos na linguagem C, mas pode ser usado, com as devidas adaptações no modelo, para outras linguagens procedurais.

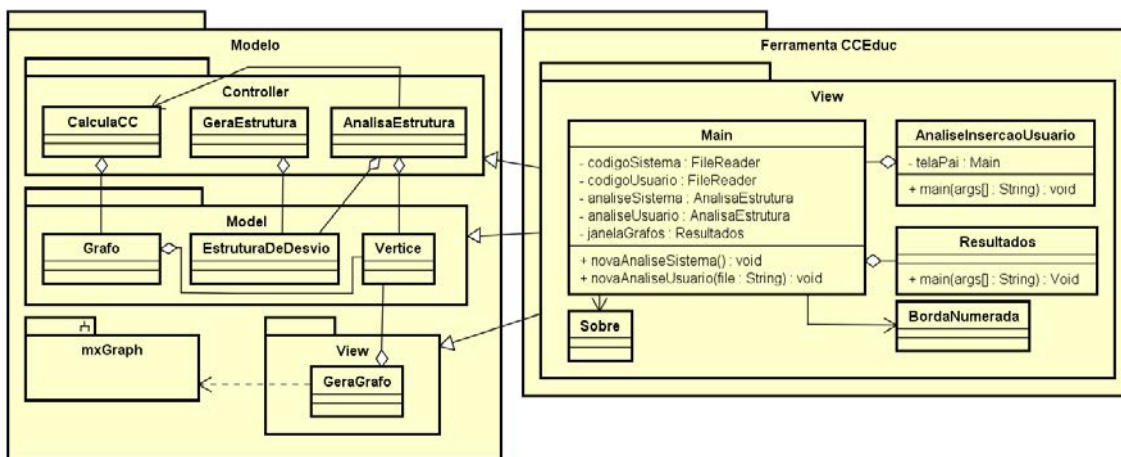


Figura 5. Estrutura de classes da ferramenta CCEduc

A Figura 6 apresenta a interface inicial fornecida para entrada da solução do aluno.

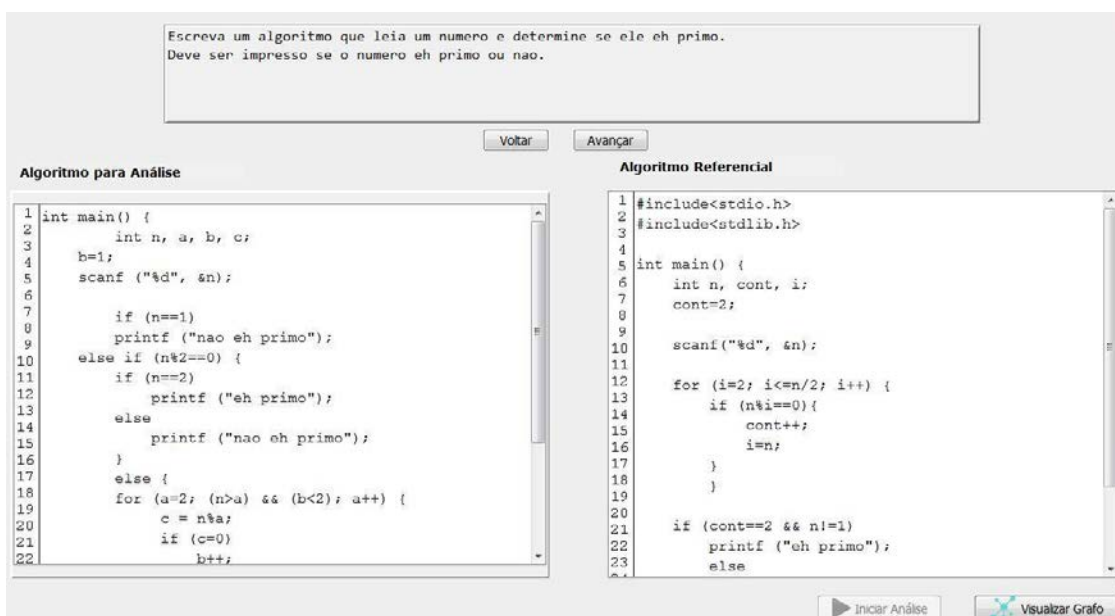


Figura 6. Interface inicial com os algoritmos a serem comparados

Uma vez que o aluno tenha fornecido a sua solução e clique no botão para iniciar a análise é apresentado para ele o algoritmo referencial. Em seguida é habilitado o botão para visualização do grafo de Complexidade Ciclomática. Quando clicado nesse botão é apresentada a tela da Figura 7. Nessa tela podem-se observar os resultados obtidos a partir da análise comparativa entre a solução do aluno e a solução referencial (retirada do livro “Algoritmos: Teoria e Prática”, Cormen 2004). A partir da análise da saída da ferramenta pode-se observar que existem diferenças entre a solução proposta pelo aluno e a solução considerada referencial. Ambas alcançam o mesmo objetivo. Porém, a solução proposta pelo aluno demanda maior esforço, usando de quatro comandos de desvio a mais do que a solução referencial.

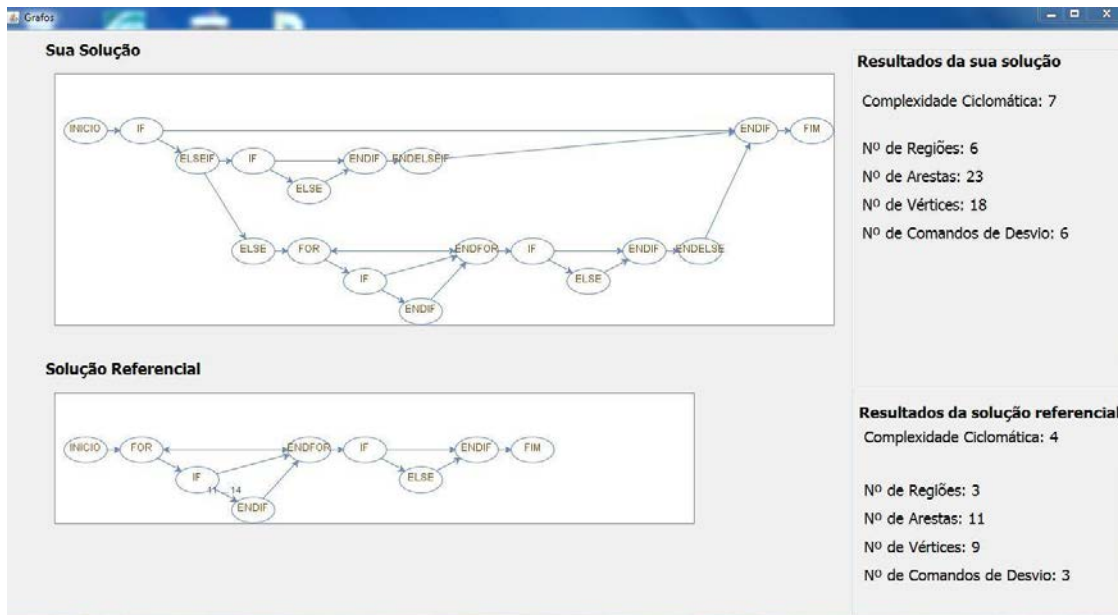


Figura 7. Resultados fornecidos pela ferramenta

Uma das principais vantagens de visualização dos grafos de fluxo e indicadores é facilitar o entendimento do código e identificar os caminhos de execução. Ao comparar as duas soluções o aluno perceberá que a sua solução quebra a execução do fluxo principal e a solução referencial a preserva. Assim os grafos e os indicadores propõem uma reflexão do aluno quanto a sua solução.

5. Considerações Finais

O modelo desenvolvido neste trabalho teve como proposta usar análise da complexidade ciclomática para comparar soluções algorítmicas, buscando uma reflexão maior do aluno iniciante em programação quanto ao seu desenvolvimento de soluções de problemas. Esta reflexão foi o ponto chave desta pesquisa, pois provoca o aluno a analisar por intermédio de dados visuais e quantitativos diferentes soluções, abstraindo conhecimentos avançados em complexidade computacional.

O modelo proposto foi materializado em uma ferramenta chamada de CCEduc. A avaliação dessa ferramenta foi realizada a partir da análise qualitativa em relação a 10 diferentes algoritmos. Em todos os testes executados obtiveram-se informações corretas em relação a análise ciclomática favorecendo o uso da ferramenta em contextos reais de ensino. Para aferir sobre o processo cognitivo da formação do pensamento algorítmico na prática, a ferramenta está sendo disponibilizada para uso de acadêmicos da disciplina de Algoritmos e Programação.

Como trabalhos futuros estão sendo realizadas adaptações em termos de interface visando contemplar recursos de acessibilidade. Além disso, há possibilidades de extensão do modelo de forma a permitir que a comparação ocorra não apenas entre o algoritmo do aluno e o algoritmo referencial, mas também pela comparação entre todos os algoritmos gerados pela turma de aprendizes.

Por fim, uma sugestão interessante seria a criação de um sistema onde não existam soluções referencias, e sim o próprio sistema enumera as melhores soluções submetidas pelos usuários, baseando-se na complexidade ciclomática das soluções. Dessa forma, enfatizando a aprendizagem pelas melhores e não melhores soluções. Esta abordagem usaria da estrutura do modelo apresentado neste trabalho, apenas introduzindo novas funções e buscando objetivos semelhantes. Seria considerada uma evolução do modelo.

Referências Bibliográficas

- Allen F. E. (1970). Control flow analysis, Proceedings of a symposium on Compileroptimization, Urbana-Champaign, Illinois, p. 1-19.
- Bombasar, James; Raabe, André; Miranda, Elisangela Maschio de; Santiago, Rafael (2015). Ferramentas para o Ensino-Aprendizagem do Pensamento Computacional: onde está Alan Turing? In: Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE), pg. 71-80. SBC: Maceio.
- Cormen, Thomas H. Cormen; Leiserson, Charles E.; Rivest, Ronald L.; et al. (2004). Algoritmos: Teoria e Prática. Ed: Campus.
- Cunha, Karine (2006). Software para Cálculo da Complexidade Ciclomática em Código-fonte PL/SQL. Universidade Regional De Blumenau. Blumenau - SC – Brasil.
- Cuny, J., Snyder, L., & Wing, J. (2010). Demystifying Computacional Thinking for Non Computer Scientists. work in progress. EUA.
- Faêda, Felipe Moreira (2012). COMPLEXGRAPH: Uma Ferramenta para Geração do Grafo da Complexidade Ciclomática com Foco em Teste de Software. Faculdade Governador Ozanam Coelho – FAGOC. MG – Brasil.
- McCabe, Thomas J. (1976). A Complexity Measure. Department of Defense, National Security Agency – EUA.
- Pressman, R. (2006). Engenharia de Software. 6. ed. Rio de Janeiro: McGraw-Hill.
- Resnick, M. Brennan (2012). New frameworks for studying and assessing the development of computational thinking. American Educational Research Association meeting. Vancouver – BC – Canadá.
- Seibt, Patrícia Ramos (2001). Ferramenta para Cálculo de Métrica em Softwares Orientados a Objetos. Universidade Regional de Blumenau. Blumenau – SC – Brasil
- Silva, T. R. da, Medeiros, T., Medeiros, H., Lopes, R., & Aranha, E. (2015). Ensino-aprendizagem de programação: uma revisão sistemática da literatura. *Revista Brasileira de Informática na Educação*, 23(01), 182.
- Sommerville (2003), I. Engenharia de software. 6. ed. São Paulo: Addison Wesley.
- Souza, J. F. de, & Coelho, S. A. (2015). Uma biblioteca gráfica para aprendizagem de algoritmos e estruturas de dados. *Revista Brasileira de Informática na Educação*, 23(01), 110.
- Wing, J. (2014). Computational Thinking Benefits Society. Social Issues in Computing. New York: Academic Press.

Challenges of Games Virtualization Applied to Educational Games

Wilk Oliveira dos Santos^{1,2}, Clovis Gomes da Silva Junior³

¹Computer Institute – Federal University of Alagoas (UFAL) – Maceió, AL – Brazil

²Department of Computer Science – University of Saskatchewan (U of S) – Saskatoon, SK – Canada

³University of Pernambuco (UPE) – Garanhuns, PE – Brazil

wos@ic.ufal.br, clovis.gomes@upe.br

Abstract. *Game Virtualization is a process of digital versions creation for traditional/physical games. In education, this process aims to create digital versions for traditional/ physical educational games, keeping psychological and pedagogical concepts from traditional version, as well as associate these concepts to contemporary game design elements. In this sense, this paper presents five challenges for Game Virtualization Applied to Educational Games: (i) design of methodologies for educational games virtualization, (i.i) design of methodologies for educational game evaluation, (i.ii) recommendation of resources for Educational Games Virtualization, (ii) multidisciplinary of Educational Games Virtualization, and (iii) professional training to Educational Games Virtualization.*

Resumo. *Virtualização de Jogos é um processo para criação de versões digitais para jogos tradicionais/ físicos. Em educação, este processo tem o objetivo de criar versões digitais para jogos educativos tradicionais/ físicos, mantendo os princípios pedagógicos e psicopedagógicos da versão tradicional do jogo, associando estes princípios às mecânicas contemporâneas de game design. Neste sentido, este artigo apresenta cinco desafios da Virtualização de Jogos aplicada aos Jogos Educativos: (i) criação de metodologias para virtualização de jogos digitais, (i.i) criação de metodologias específicas para avaliação de jogos, (i.ii) recomendação de recursos para Virtualização de Jogos Educativos, (ii) multidisciplinaridade da Virtualização de Jogos Educativos, e (iii) formação de profissionais para Virtualização de Jogos Educativos.*

1. Introduction

In the last decades, the international industry of game development has been increasing, becoming an important worldwide industry. According to a recent report of Entertainment Software Association (ESA 2015), only in the United States, there are approximately 155 million of players moving, in 2014, approximately \$ 22.41 billion

around the word. Additionally, another interesting recent study showed that 74% of K-8¹ teachers use digital games in the classroom (Lofgren 2015), enabling a considerable growth of the Educational Games industry in the last years.

In the academic context, a series of recent studies have been done in order to investigate effects of video games in different educational aspects and context (e.g. Anderson *et al.* (2010), Scoresby and Shelton (2011), Silva Neto *et al.* (2013) and others). These studies have highlighted a series of discussions related to video games positive effects in student's learning, such as: *game fullness*, motivation and others; and negative effects, such as: violence, social interaction absence and extreme appreciation of "digital" (digital activities conducted using some kind of technological device), in detriment to "traditional" (traditional (physical/ manual) activities).

In this context, recent studies addressing Game Virtualization, (e.g. Santos *et al.* (2015) and Oliveira and Silva Junior (2015)) in order to create digital versions to traditional/physical Educational Games, aim to keep the pedagogical and psychological aspects from traditional/physical game, and associate these aspects to contemporary game design elements. Like this, recent studies have showed a series of benefits of virtualized games, as well as highlighted a series of discussions related to design, application and evaluation of Educational Games developed based in Virtualization Games.

Thus, this paper proposes a series of different challenges for Games Virtualization applied to Educational Games: (i) Design of methodologies for educational games virtualization, (i.i) design of methodologies for educational game evaluation, (i.ii) recommendation of resources for Educational Games Virtualization, (ii) multidisciplinary of educational games virtualization, and (iii) professional training to Educational Games Virtualization. These challenges and some recommendations for future works in this field are presented and discussed in this paper.

2. Background

This work aims to propose a series of different challenges for Games Virtualization applied to Educational Games. Following, we present the main topics addressed in this study: Educational Games and Educational Games Virtualization, as well as some related works.

2.1. Educational Games

Over the last 20 years, computer games have been increasingly replaced the more traditional games as leisure activities, and have had a transformational impact on how we spend our leisure time (Connolly *et al.* 2012). According to these authors, Educational Games provide engaging activities and it seems like that, far from waning, interest in games for leisure is still growing.

¹ K-8 schools are schools in the United States that enroll students from kindergarten/pre-K (age 5) through 8th grade (up to age 14), combining the typical elementary school (K-5) and junior high or middle school (6-7-8).

At this point, in the last decades, Educational Games have been used in many teaching contexts of different topics as: Business, Geography, History, and others (Connolly *et al.* 2012), as well as studied in different perspectives as violence (Anderson and Gentile 2014), learning (Linehan *et al.* 2014), *playfulness* (Scoresby and Shelton 2011), and others.

According to recent studies, a series of empirical results related to games-based learning has shown that despite the overwhelming publicity given to the negative impact of games, like most technologies before them, computer games can have both positive and negative impacts (Connolly *et al.* 2012). These results are important to demonstrate not only the efficacy of games in educational context, but also to highlight the importance of new studies in different kinds of Educational Games.

2.2. Educational Games Virtualization

In the contemporary world, the idea of ‘virtualization’ is considered and studied by philosophy, psychology, physics, biotechnology, arts, and others (Kerimbayev 2015), increasing topics related to virtualization in different contexts as social life, industry and education. In this perspective, Educational Games Virtualization is a process to create digital versions for traditional/physical Educational Games before used to teaching of a determined topic. The main goals of Educational Games Virtualization are keep/preserve pedagogical and psychological aspects of the traditional/physical version of the educational games version and combine these aspects with the contemporary game design elements, in order to increase students’ learning.

Educational Games Virtualization can be used in different contexts and be support in different educational topics. According to Oliveira and Silva Junior (2015), the Educational Games Virtualization process should be start with the choice of the traditional educational game and end with the process of game evaluation, in order to identify if the educational game (in the digital version) keeps/ preserves the pedagogical and psychological aspects from traditional/physical game version, as well identify if the game can be increase students’ learning.

According to Santos *et al.* (2013) and Santos *et al.* (2015) this process involves a multidisciplinary team, with professionals from Computer Science (especially Human-Computer Interaction (HCI)), Pedagogy, Psychology and those from specific educational games application fields, for instance Mathematic, Language, Biology, and others, in order to auxiliary the Educational Game Virtualization process.

2.3. Related Works

In the last years, several studies have been done in order to develop and evaluate Educational Games based in the Educational Games Virtualization process. Generally, these studies are conducted in the context of mathematics education and aim to implement games capable to contribute to students’ learning in elementary school and high school.

Given the background, Santos *et al.* (2013) proposed a process of Virtualize Educational Games used in previous decades to teach mathematical concepts. With this proposal, authors summarize the importance of keeping pedagogical and psychological concepts of traditional/physical game version, as well as associate these concepts to

contemporary digital game elements. In this context, a series of performed studies aim to implement/virtualize Educational Games based on this process.

Togni *et al.* (2009) presented an Educational Game to mathematics teaching in elementary schools that was developed based in the Meaningful Learning Theory, and used with students in real context. However, they do not present concerns related to Computational and Pedagogical evaluation. Thusly, a similar work by Togni *et al.* (2009), Zorzal *et al.* (2008) showed a series of Educational Games virtualized using Augmented Reality (AR) resources, but do not presented concerns related to pedagogical and psychological aspects of games.

Recently, aiming to supply problems that were showed before, Santos *et al.* (2015) and Oliveira e Silva Junior (2015) proposed two Educational Games for mathematics teaching (*called* “Conquering with the Rest” and “Challenge with Sticks”) developed based on the Educational Games Virtualization. These works were developed in the Elementary School context and were evaluated with students and teachers.

The studies showed before do not presented concerns related to specific methodologies of game virtualization and game evaluation, as well as do not presented concerns related to pedagogical and psychological concepts maintenance and do not presented concerns related to the influences of the game in the students’ learning. Both studies before mentioned focused in mathematics teaching domain. The studies before showed contribute and reaffirm the importance of the challenges proposed in this paper.

3. Challenges of Virtualization Game in Educational Games

This section aims to present five challenges for Games Virtualization applied to Educational Games, however, some challenges also can be considered in the general Educational Games field. Figure 1 synthetize the challenges proposed, starting in the concept of Educational Game Virtualization.

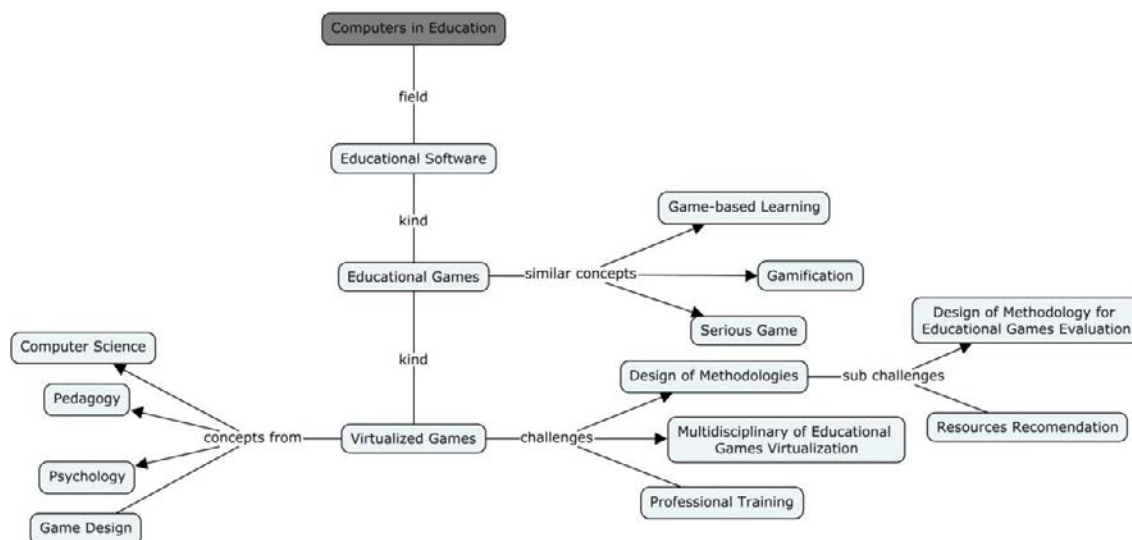


Figure 1: Conceptual map of challenges of Educational Games Virtualization

3.1. Challenge 1: Design of Methodologies for Educational Games Virtualization

In the last years, some studies have been made addressing Educational Games Design based on Games Virtualization (*e.g.* Santos *et al.* (2014) and Oliveira e Silva Junior (2015)). However, these studies present different methodologies and process for the development of games, or use others methodologies from traditional game design. On the other hand, Educational Games Virtualization, different from conventional Games Development, as well as the conventional Educational Games Development, use concepts from different fields as Pedagogy, Psychology, Computer Science and others domains related to the context of game application.

On this subject, we propose, as a big challenge for Games Virtualization applied to Educational Games, the development of methodologies for Educational Games Virtualization. Considering that the Virtualization of Educational Games uses several concepts regarding Pedagogy, Psychology and Computer Science (as well as specifically concepts of the area of game application, for instance: Mathematic, Geography, Biology and others), is important that the methodologies consider some steps, starting with the traditional game chosen in partnership with professionals of application area, and ending with the game evaluation in real scenario, with students and partnership with professionals of application area, in order to evaluate if the game maintain pedagogical and psychological concepts from the traditional game, as well as if contributes to students' learning.

Based on the challenge proposed, we present some directions for researchers and professionals: (i) design of methodologies to fill this gap, must be implemented with multidisciplinary team and collaboration of professionals of Game Design, Computer Science, Psychology and Pedagogy, in order to aggregate different knowledge areas and implement a methodology with potential scientific and industrial; (ii) methodologies hereafter proposed must be evaluated by different professionals in different academics and industrial scenarios to develop a methodology able to be used in both contexts; (iii) methodologies hereafter proposed should provide resources to virtualize Educational Games for different courses, not being addressed to specifically contexts.

3.1.1. Challenge 1.1: Design of Methodologies for Educational Game Evaluation

Recently, some studies have been done aiming to propose methodologies to Educational Games evaluation, for instance: Rocha and Araújo (2015) proposed an interactive methodology and integrative modes to develop serious games, integrating a holistic and multidisciplinary approach into the areas of simulation, game, learning, training and domain. This same study proposes a evaluate games model that is produced based in this methodology. From the basis, Barbosa *et al.* (2015) proposed a heuristic to Educational Games evaluation. Barbosa's methodology was based in a study with a series of existing heuristics, such as HEP, PLAY and *GameFlow*, and the Criteria for Designing Educational Computer Games from Nicola Whitton.

However, these methodologies and others similar methodologies (*e.g.* Medeiros (2015) and França e Silva (2014)) are focused in specifics fields, without elements with allowed use in different contexts, or yet, focus only in the game interface evaluation, in detriment of pedagogical elements, or focus only in pedagogical elements,

in detriment of game interface, hindering apply these methodologies in Educational Games developed based on the Games Virtualization. In order to propose a solution for the gap that was talked before, Santos *et al.* (2015) proposed a qualitative and quantitative approach to the interface and pedagogical evaluation in Educational Games, including Educational Games developed based on Game Virtualization. However, this approach do not presents concerns related with specifics points of Educational Virtualization Games, as the maintenance of Pedagogical and Psychological principles of traditional games.

In this context, we propose, as a big challenge for Games Virtualization applied to Educational Games, design of methodologies to Educational Games evaluation developed based on Games Virtualization. So, as the Games Virtualization has the goal of keeping the pedagogical and psychological concepts of the traditional game, as well as associate these concepts contemporary elements of game design, future methodologies proposals to evaluate the games developed by the virtualization process games should be able to identify whether the pedagogical and psychological aspects of the traditional game are actually kept in the digital game. Since also Educational Game Virtualization is a computational process to solve educational problems, future methodologies proposed must be able to evaluate interface concepts used in the game, as well as evaluate if the game contributes to students' learning.

Based on the challenge before proposed, we present some directions to researchers and professionals: (i) methodologies hereafter proposed should develop with professionals of different domains as: Pedagogy, Psychology and Computer Science (especially Software Engineering and HCI), in order to provide a methodology capable to evaluate games according to different criteria, from Pedagogy, Psychology and Computer Science (HCI); (ii) methodologies hereafter proposed should provide minimally the following resources: evaluate if the game keeps the Pedagogical and Psychological principles allied to contemporary game design elements, evaluate the game quantitatively and qualitatively in aspects of Computers Science (specially HCI) and Pedagogical aspects, as well as evaluate if the game contributes to students' learning.

3.1.2. Challenge 1.2: Recommendation of Resources for Educational Games Virtualization

In the last decades, with technology advent in several society contexts, a series of new technologies in the software development industry have been daily emerged. A series of new programming languages, tools, engines and others general resources emerged too. According to Ampatzoglou and Stamelos (2010), similar it happens in the industry of games that has a series of specifics programming languages, tools and general engines. In the Educational Games industry, this situation is highlighted because professionals need integration between technical and pedagogical contexts (Chaudy *et al.* 2014). On the other hand, few studies have been done to help professionals of educational game design.

In this regard, we propose, as a big challenge for Virtualization Game applied to Educational Game, recommendation of resources to Educational Games Virtualization process. This challenge refers to development of studies in order to

identify programming languages, tools, engines and general resources capable to help professionals of Educational Game Design to develop Educational Games based on the Educational Game Virtualization, this challenge is not an exclusive challenge on Educational Games Virtualization, but is a challenge that comprises all Educational Games development field.

Based on the challenge before proposed, we present some directions to researchers and professionals: (i) future studies aiming to identify programming languages, tools, engines and general resources should be developed with empirical metrics in order to provide resources to futures replications in different contexts; (ii) studies hereafter conducted to identify resources to development of Educational Games based in Virtualization Games should be developed with partnerships professionals of Science Computers and Education, in order to recommend resources capable to facilitate the process of development and maintenance of Educational Games (integration between technical and educational concepts).

3.2. Challenge 2: Multidisciplinary of Educative Games Virtualization

Recently, a series of studies addressing Educational Games Virtualization have been done. However, majority of studies have been carried out to Mathematic domain, especially considering studies before presented (see Related Works). On the other hand, considering the concept of Game Virtualization, many others domains use/used traditional games in classroom to increase students' learning in several activities, as well as to improve the involvement of the students, in domains for instance: Biology, Geography, Literature, and others (Connolly *et al.* 2012).

In this context, we propose, as a big challenge for Games Virtualization Applied to Educational Games, multidisciplinary development of educational games, based on Games Virtualization. Considering the majority studies in Educational Games Virtualization being developed in the context of Mathematic course, is recommended the development of others games in others domains, in order to evaluate the efficacy of Educational Game Virtualization in different contexts, as well as identify games' consequences in students in different situations and contexts.

Based on the challenge proposed before, we present some directions to researchers and professionals: (i) games hereafter virtualized should be developed and evaluated according to specifics methodologies of Educational Game Virtualization, in order to permit a robust evaluation in Computational and Pedagogical aspects, as well as permit future applications of works; (ii) game hereafter virtualized should be evaluated in different real contexts (with real classroom and real students).

3.3. Challenge 3: Professional Training to Educational Games Virtualization

Professional training has become a big challenge in the industry and education in the last years. In the past few years, various studies have been conducted to provide professional training resources in different domains (*e.g.* Rocha and Araújo (2015) and Gurinović *et al.* (2015)). In the game industry, it is no different, faced with a series of tools updates, devices, clients, among others, the industry of game design undergoes constant changes and need to invest a high value on training professionals, in order to maintain your competitiveness.

In the Educational Game industry also is no different, starting with the updates before showed, as well as increasing specifics points, for instance: multidisciplinary team, different evaluation process, integration with educational context, and others. Educational Game industry suffered even more with this updates and was necessary to invest annual big amounts with training of professionals. Thusly, we propose as a big challenge for Game Virtualization applied to Educational Games, professional training to Educational Games Virtualization.

Based on the challenge before proposed, we present some directions to researchers and professionals: (i) before the importance of professionals to development of educational resources, we consider necessary invest in the professional multidisciplinary capacitation and specialization, in order to form professionals capable to deal with different situations (in the industry and academy), as well as they will be able to deal with the integration of technology (game design approach), and education; (ii) futures research groups, companies and others professional groups interested in Educational Games Virtualization can invest in the integration academy-industry, in order to training professionals capable to deal with these different scenarios.

4. Concluding Remarks

In the past few years, Educational Game Virtualization have been addressed as a promising topic of Educational Game, especially to provide to students, pedagogical and psychological aspects of traditional/physical games, allied to modern game design approaches (e.g. Santos *et al.* (2015) and Oliveira and Silva Junior (2015)). However, before a series of studies recently conducted in Educational Virtualization Games, a series of challenges have been appeared. In this way, this paper proposed and discussed addressed a series of big challenges for the Game Virtualization applied to Educational Game Virtualization.

This paper proposed and discussed a series of challenges for Virtualization Games applied to Educational Games: (i) Design of methodologies for educational games virtualization, (i.i) design of methodologies for educational game evaluation, (i.ii) recommendation of resources for Educational Games Virtualization, (ii) multidisciplinary of Educational Games Virtualization, and (iii) professional training to Educational Games Virtualization.

Finally, as general recommendations, we highlight the importance of methodologies design to Educational Games development and evaluation, of educational games developed based on the process of Educational Games Virtualization, as well as the importance of evaluate these games in real and different contexts and invest in the training of professional to development of Educational Games based in the Educational Games Virtualization. Finally, as future work, we highlight the execution of State of the Art and Systematic Literature Review in order to obtain a general empirical overview in this field, as well as studies in order to propose specifically methodologies to Educational Games Virtualization.

References

Ampatzoglou, A., & Stamelos, I. (2010). Software engineering research for computer games: A systematic review. *Information and Software Technology*, 52(9), 888-901.

- Anderson, C. A., & Gentile, D. A. (2014). Violent Video Game effects on Aggressive thoughts, feelings, physiology, and Behavior. *Media Violence and Children: A Complete Guide for Parents and Professionals*, 229.
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: a meta-analytic review. *Psychological bulletin*, 136(2), 151.
- Babosa, M. B., Rêgo, A. B., & Medeiros, I. (2015). HEEG: Heuristic Evaluation for Educational Games. In: XIV Brazilian Symposium of Games and Digital Entertainment. Teresina-PI, Brazil.
- Chaudy, Y., Connolly, T. M., & Hainey, T. (2014). EngAGe: A link between Educational Games Developers and Educators. In: Games and Virtual Worlds for Serious Applications (VS-GAMES), 2014 6th International Conference on (pp. 1-7). IEEE.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661-686.
- Essential facts about the computer and video game industry: 2015 sales, demographic and usage data. Entertainment Software Association, 2015.
- França, R. S., & Silva, A. C. B. (2014). Avaliação de softwares educativos para o ensino de Língua Portuguesa. *Revista Brasileira de Informática na Educação*, 22(03), 23.
- Gurinović, M., Novaković, R., Šatalić, Z., Nikolić, M., Milešević, J., Ranić, M., & Glibetić, M. (2015). Professional training in nutrition in Central and Eastern Europe: current status and opportunities for capacity development. *Public health nutrition*, 18(02), 372-377.
- Kerimbayev, N. (2015). Virtual learning: Possibilities and realization. *Education and Information Technologies*, 1-13.
- Linehan, C., Bellord, G., Kirman, B., Morford, Z. H., & Roche, B. (2014). Learning curves: analysing pace and challenge in four successful puzzle games. In *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play* (pp. 181-190). ACM.
- Lofgren, K. (2015) Video Game Statistics & Trends Who's Playing What & Why?. Available in: <http://www.bigfishgames.com/blog/2015-global-video-game-stats-whos-playing-what-and-why/>. Accessed: 09/March/2016.
- Medeiros, J. F. (2015) Avaliação de Usabilidade e Jogabilidade: Métodos, Técnicas e Ferramentas. In: XIV Simpósio Brasileiro de Jogos e Entretenimento Digital. Teresina-PI, Brasil.
- Oliveira, W., Neto, S., da Silva Junior, C. G., & Bittencourt, I. I. (2015). Avaliação de Jogos Educativos: Uma Abordagem no Ensino de Matemática. In: *Anais do Simpósio Brasileiro de Informática na Educação* (Vol. 26, No. 1, p. 657).

- Oliveira, W., & da Silva Junior, C. G. (2015). Pesquisa, Desenvolvimento e Avaliação de um Jogo para o Ensino de Matemática, Baseado no Processo de Virtualização de Jogos. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação (Vol. 4, No. 1, p. 145).
- Rocha, R. (2015). Metodologia iterativa e modelos integradores para desenvolvimento de jogos sérios de treinamento e avaliação de desempenho humano. In Anais dos Workshops do Congresso Brasileiro de Informática na Educação (Vol. 4, No. 1, p. 13).
- Santos, W., Silva Neto, S. R., & Silva Junior, C. G. (2013). Uso de Games no ensino da Matemática. Uma proposta de virtualização dos jogos tradicionais, para uso como mecanismo de apoio ao processo de ensino e aprendizagem. In: Anais do Simpósio Hipertexto e Tecnologias na Educação, Recife-PE.
- Santos, W., O.; Souza, A., A.; da Silva, A., K., T.; Oliveira, M., L., S.; Silva Neto, S., R.; Rodrigues, A. N., Silva Junior, C. G. (2015) Development Process of an Educational Game: An experience in Brazil. In: XIV Brazilian Symposium of Games and Digital Entertainment. Teresina-PI, Brazil.
- Scoresby, J., & Shelton, B. E. (2011). Visual perspectives within educational computer games: effects on presence and flow within virtual immersive learning environments. *Instructional Science*, 39(3), 227-254.
- Rocha, R., & Araújo, R. B. (2015). Metodologia iterativa e modelos integradores para desenvolvimento de jogos sérios de treinamento e avaliação de desempenho humano. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação (Vol. 4, No. 1, p. 13).
- Scoresby, J., & Shelton, B. E. (2011). Visual perspectives within educational computer games: effects on presence and flow within virtual immersive learning environments. *Instructional Science*, 39(3), 227-254.
- Silva Neto, S. R., Santos, H. R., Souza A. A., Santos W. O. (2013) Jogos Educacionais como Ferramenta de Auxílio em Sala de Aula. In: Anais do Workshop de Informática na Escola (Vol. 1, No. 1, p. 130).
- Togni, A. C., Rehfeldt, M. J. H., Bersch, M. E., Thomas, M. I. P., Poletti, C. F., & Kronbauer, K. A. (2009). Piff geométrico: um objeto virtual de aprendizagem para o ensino de ciências exatas. *RENOTE*, 7(3), 45-52.
- Webster, D., & Celik, O. (2014). Systematic review of Kinect applications in elderly care and stroke rehabilitation. *Journal of neuroengineering and rehabilitation*, 11(1), 1.
- Zorzal, E. R., Kirner, C., Cardoso, A., Lamounier Jr, E., de Oliveira, M. R. F., & Silva, L. F. (2008). Ambientes Educacionais Colaborativos com Realidade Aumentada. *RENOTE*, 6(2).

Uma Abordagem Para a Organização Automática de Objetos de Aprendizagem em Ambientes Educacionais Ubíquos com Base em Estilos de Aprendizagem

Miller M. Mendes¹, Vitor C. de Carvalho¹, Fabiano A. Dorça¹, Renan G. Cattelan¹

¹ Faculdade de Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia - MG - Brasil.

miller@comp.ufu.br, vitor@mestrado.ufu.br

{renan, fabianodor}@ufu.br

Abstract. *Due to the large amount of learning objects generated from web systems, there is the need to arrange them so that they can be recommended efficiently. Therefore, this work proposes a content clustering for large learning object repositories to support and facilitate the content recommendation process in relation to student learning styles model. This proposal also takes into account the complementarity between the learning styles, where, if a student has a tendency to a learning style, he can also learn from other learning style, thus a feature can complement other to improve learning.*

Resumo. *Devido à grande quantidade de objetos de aprendizagem gerados a partir de um sistema web, tem-se a necessidade de organizá-los para que os mesmos possam ser recomendados de maneira eficiente. Para isso, este trabalho propõe um agrupamento de conteúdo em grandes repositórios para apoiar e facilitar o processo de recomendação de conteúdo em relação a estilos de aprendizagem de estudantes. Esta proposta leva também em consideração a complementação entre os estilos de aprendizagem, onde um aluno que possui tendência para um estilo de aprendizagem também pode aprender com outro estilo de aprendizagem, sendo assim uma característica pode complementar outra para melhorar o aprendizado.*

1. Introdução

A área educacional tem sido um domínio de interesse para o apoio de sistemas de computação. Diante disso, diversas pesquisas passaram a ser executadas com o intuito de desenvolver e analisar Ambientes Educacionais Ubíquos (AEUs) como facilitadores de aprendizagem [Levis and Barbosa 2008]. AEUs podem empregar Objetos de Aprendizagem (OAs) como fontes de conteúdo. OAs, deste modo, são classificados como entidades passíveis de utilização dentro do processo de ensino e aprendizagem.

Outro conceito importante para a criação de sistemas de computação para ambientes educacionais são nomeados de Estilo de Aprendizagem (EA). Os EAs estão relacionados às estratégias que um aluno tende a aplicar com frequência durante sua busca por conhecimento. [Felder and Silverman 1988] criaram um modelo específico para a aprendizagem de estudantes por meio de quatro dimensões: percepção, entrada, processamento e organização.

O primeiro passo em direção a um sistema capaz de recomendar automaticamente conteúdos de aprendizagem para estudantes é a identificação das melhores formas de organização dos OAs, de modo a permitir uma melhor qualidade na recomendação de conteúdo [Margarida et al. 2003].

Assim, utilizar algoritmos de agrupamento e classificação para organizar os OAs é uma abordagem promissora. Contudo, existem poucos trabalhos na literatura que utilizam abordagens aproveitando da clusterização dos OAs no contexto de ambientes educacionais ubíquos. Este trabalho vem suprir esta falta de abordagens com métodos de agrupamento e classificação para a organização dos OAs e analisar suas vantagens.

Uma relação entre metadados e o modelo [Felder and Silverman 1988] foi pesquisado por [Carvalho et al. 2014] apresentaram uma abordagem para a recomendação de OAs com base no mapeamento do padrão IEEE Learning Object Metadata (IEEE LOM) [IEEE 2002] com os EAs. Como resultado, foram apresentados uma quantidade mínima de campos que precisa ser utilizada para uma recomendação eficiente entre OAs.

As próximas páginas serão divididas em seções onde a Seção 2 apresenta trabalhos que estão relacionados ao tema em questão. Na Seção 3 são descritos, conceito inerente ao entendimento do trabalho. A Seção 4 descreve a abordagem utilizada e a Seção 5 mostra a conclusão sobre os estudos.

2. Trabalhos Relacionados

O modelo de alunos baseado nas dimensões de Felder e Silverman tem sido amplamente aplicado em diversos sistemas de recomendação encontrados na literatura. Como exemplo, Zaina et al. 2012 apresentaram a proposta E-LORS (e-Learning Object Recommendation), que se baseia nos conceitos de sistemas de recomendação, utilizando um algoritmo de filtragem baseado em conteúdo. Um requisito importante foi a utilização de OAs no padrão IEEE LOM sendo que o modelo de aluno adotado foi baseado nas dimensões de Felder e Silverman. A abordagem E-LORS não propõe a utilização de OAs agrupados como esta sendo proposto neste trabalho.

O trabalho apresentado por Paper and Moiss 2009 seleciona EAs utilizando a rede neural Self Organizing Map (SOM) [Kohonen 1998]. Neste trabalho a rede recebe diferentes padrões de aprendizagem do aluno obtidos a partir de um Learning Style Questionnaire (ILSQ) no padrão de Felder e Silverman. Foi implementado o sistema Educa onde existe a presença do instrutor e dos estudantes, o instrutor cria os OAs que são classificados dentro do modelo, criando assim uma pontuação para cada objeto. O método proposto não utiliza OAs no padrão IEEE LOM para representar estilos de aprendizagem.

Uma abordagem feita por Zapata and Menendez 2009 é uma proposta para apresentar os resultados da aplicação de uma metodologia de extração de conhecimento de OAs através da utilização de quatro fontes de dados: metadados, avaliações de qualidade pedagógica, perfis de usuário, e arquivos de log de OAs em um sistemas de gestão. Este trabalho não utiliza a relação entre EAs e OAs para obter resultados.

Em um modelo de recomendação Aguiar et al. 2015 utiliza objetos de aprendizagem no padrão IEEE LOM com o modelo de Felder e Silverman a partir da filtragem de OAs com base no conteúdo e posteriormente ordena os OAs com base nos EAs do aluno e sua popularidade, desta forma a ordenação baseada em conteúdo pode deixar que objetos

baseados em outros atributos fiquem com uma popularidade baixa. Na abordagem deste trabalho é considerado que todos atributos do OA sejam inerentes ao aprendizado.

Outra abordagem proposta por Casali et al. 2011 em encontrar OA mais adequados as suas necessidades e preferências com base em repositórios de aprendizagem distribuídos. OAs são recomendados seguindo uma abordagem baseada em regras, levando em consideração o perfil dos usuários.

Espera-se assim, que a proposta de agrupar OAs se torna uma forte contribuição para as técnicas de recomendação de conteúdo, melhorando a utilização de OAs em diferentes trabalhos, sendo uma melhoria na forma em que os objetos se organizam com base em sua similaridade.

3. Fundamentação Teórica

Nesta sessão, são apresentadas as definições e os conceitos necessários para o entendimento da abordagem proposta, além de citar trabalhos relacionados ao tema desta pesquisa. Serão apresentados os conceitos proposto por Felder and Silverman e o padrão utilizado para os objetos de aprendizagem. Posteriormente serão mostrados detalhes dos algoritmos utilizados no modelo.

3.1. Estilos de Aprendizagem Felder and Silverman

Os EAs proporcionam orientações para a individualização do ensino [Given 2002]. Até recentemente, o foco da investigação acerca dos estilos de aprendizagem centrou-se na sua aplicação ao processo de ensino e aprendizagem em cenários de sala de aula. Com o avanço nos estudos de facilitar a criação materiais educacionais multimídia e com o desenvolvimento das tecnologias baseadas na Web, a investigação acerca dos estilos de aprendizagem passou também a incluir contextos de ensino e aprendizagem em sistemas educacionais ubíquos [Levis and Barbosa 2008].

Para satisfazer a diferentes estilos é necessário que se utilize estratégias de ensino adequadas às várias perspectivas de aprendizagem. A partir desses conceitos o modelo possui quatro dimensões sendo: percepção, processamento, entrada e entendimento conceituadas a seguir respectivamente:

- Sensorial/Intuitivo: dentro do estilo sensorial é mais comum ter características praticas, concreto e exemplos, sendo que para os intuitivos é mais comum características teóricas e abstratas.
- Ativo/Reflexivo: dentro do estilo ativo é mais comum características ativas do aluno dentro do ensino, enquanto no reflexivo, o aluno é prefere materiais teóricos.
- Visual/Verbal: esta dimensão indica que um estudante pode ter preferência por informações transmitidas visualmente, sendo visual, ou informações com base em texto ou falas sendo assim verbal.
- Sequencial/Global: dentro do estilo sequencial a apresentação dos materiais deve proporcionar uma navegação preestabelecida enquanto que para a dimensão global a apresentação dos materiais deve possibilitar uma navegação por conta própria.

Este modelo é amplamente utilizado em vários trabalhos que utilizam estilos de aprendizagem [Branquinho et al. 2015, Pernas et al. 2009].

3.2. Objetos de Aprendizagem

OAs podem ser definidos como recursos digitais que são utilizados em diferentes ambientes de aprendizagem com uma base tecnológica com fins educacionais [da Silva et al. 2011]. Possuem características bem definidas conforme descrito por [Audino and da Silva Nascimento 2012] para que possam ser bem utilizados em áreas educacionais.

O modelo de dados IEEE LOM especifica quais aspectos de um OA devem ser descritos e são definidos vocabulários que pode ser utilizado para essas descrições [IEEE 2002]. Os metadados usados para descrever os OAs podem incluir informações educacionais, como forma de ensino, nível de escolaridade e pré-requisitos, além de informações a respeito do tipo de objeto, autor, proprietário entre outros.

Ele possui nove grupos distintos para representar os metadados: Geral para representar informações gerais, Ciclo de Vida onde representa histórico e estado atual do OA, Meta-metadata mostram informações sobre os metadados, Técnicos representam requisitos técnicos, Educacional possui informações relacionadas aos aspectos e educacionais, Direitos exibe os direitos, propriedade intelectual e condições de uso, Relação informa relacionamentos entre os OAs, Anotação mostra comentários sobre o uso educacional e Classificação onde possui o posicionamento do OA frente a sistemas de classificação pré-definido.

3.3. Recomendação de Objeto

Dentro os maiores objetivos de se utilizar OAs, a recomendação se destaca, pois permite que objetos sejam reutilizados para outras disciplinas diferentes das que foram criadas inicialmente, apresentados de forma personalizada aos estudantes. Outra característica dos objetos de aprendizagem é o fato de que pode ser disponibilizado através da internet, o que permite que um grande número de estudantes tenham acesso.

As recomendações são realizadas baseando-se em características que são relevantes ao escopo do sistema. Um sistema de recomendação com fins educacionais acompanha as ações do aluno quando este interage com determinados conteúdos para que seja possível posteriormente, sugerir materiais e serviços adequados as suas preferencias [Felfernig et al. 2007].

As recomendações devem fazer sugestões que se encaixam as expectativas dos usuários. Como por exemplo surgiram os sistemas de recomendação de [Cazella et al. 2010] em sistemas que visam auxiliar o usuário na busca e seleção de um conteúdo focado em seu perfil, funcionando como filtros de informação. Assim, o usuário terá como resultado de busca apenas os materiais mais próximos e relevantes.

3.4. Clusterização

Clusterização é a divisão de dados, com base na similaridade entre eles, em grupos disjuntos chamados *clusters*. Isso significa que dados em um mesmo cluster são mais similares do que dados pertencentes a clusters diferentes [Ochi et al. 2004]. O ato de agrupar os dados representa uma forma importante no aprendizado e compreensão.

Pode ser definido também como um problema de aprendizado não supervisionado, já que a estrutura dos dados e as propriedades que os tornam similares são desconhecidas. O objetivo da clusterização é encontrar uma organização válida e conveniente dos dados,

ao invés de separá-los em categorias como acontece no reconhecimento de padrões e na classificação de dados [Jain and Dubes 1988]. Na Figura 1 pode-se observar as etapas de um processo de clusterização.

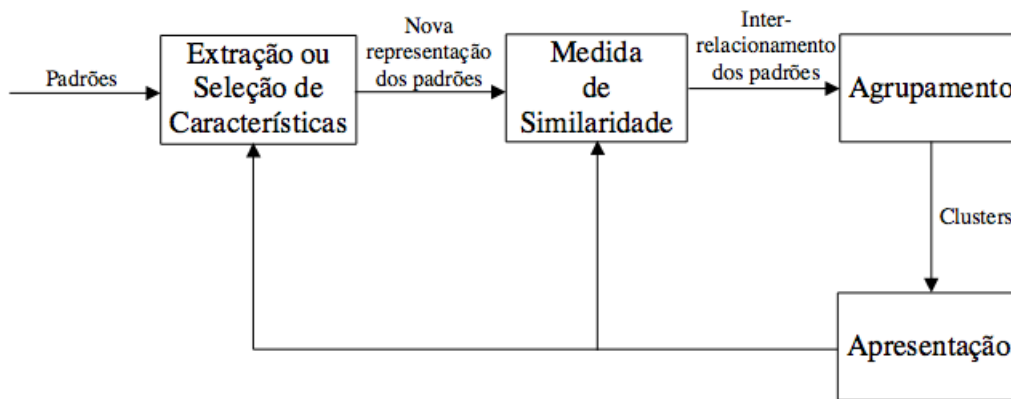


Figura 1. Etapas do processo de clusterização

As etapas apresentadas na Figura 1 são:

- **Seleção de Características:** a seleção de características é um processo de identificação do subconjunto mais efetivo dos atributos disponíveis para descrever cada padrão e extração de características é o uso de uma ou mais transformações nos atributos de entrada de modo a exaltar uma ou mais características dentre aquelas que estão presentes nos dados. Qualquer um ou ambas estas técnicas pode ser usado para obter um conjunto adequado de recursos para usar em agrupamentos
- **Medida de Similaridade:** para que a proximidade de dois dados, possa ser calculada, é necessário adotar alguma medida de similaridade entre eles. Existem diversas maneiras de quantificar a similaridade, ou dissimilaridade, entre pares de dados. A escolha da medida de similaridade adequada ao problema é fundamental para a clusterização dos dados.
- **Agrupamento:** define-se o modo de agrupamento dos dados, que pode ser realizado de diferentes maneiras. Os algoritmos de clusterização são classificados de acordo com as diferentes técnicas que empregam no agrupamento dos dados. Dentre esses algoritmos destacam-se os algoritmos de clusterização hierárquica e os algoritmos por particionamento. Na clusterização por particionamento o conjunto de dados é dividido em um número determinado de clusters uma única vez e no particionamento hierárquico são produzidas diversas partições do conjunto de dados com base na junção ou divisão dos clusters de acordo com a medida de similaridade.
- **Apresentação/Validação:** a qualidade dos clusters encontrados é avaliada sendo que essa validação pode ser feita com base em índices estatísticos ou através da comparação com outros algoritmos. Além disso, a análise dos resultados pode levar à redefinição dos atributos escolhidos e do cálculo de similaridade, definidos nas etapas anteriores.

Segue alguns algoritmos de clusterização utilizados para agrupar dados. O algoritmo K-means [MacQueen et al. 1967] é um algoritmo não supervisionado muito

utilizado em diversas áreas, sendo que a quantidade de grupos deve ser definida antecipadamente. O algoritmo funciona de maneira iterativa, onde centroides, criados inicialmente de maneira aleatória, são reajustados em direção ao centro dos grupos que são reorganizados a cada iteração.

A rede neural Self Organizing Map (SOM) [Kohonen 1998], que faz parte das redes neurais de aprendizado competitivo, é aquela na qual os neurônios de uma camada competem entre si pelo privilégio de permanecer ativos, tal que o neurônio com maior atividade seja o único a participar do processo de aprendizado. Na rede SOM é fornecido um vetor de entrada, no qual todos os neurônios a avaliam com uma função de avaliação distinta e aquele que obtiver a maior avaliação para a entrada tem seus pesos ajustados, levando a um neurônio vencedor. Esta característica faz com que grupos vizinhos no mapa discreto possuam padrões semelhantes no espaço de entrada.

Um método também conhecido é o Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [Ester et al. 1996] é um algoritmo baseado em densidade, que é significativamente efetivo para identificar agrupamentos de formato arbitrário e de diferentes tamanhos, separar os ruídos dos dados e detectar grupos padrões dentro do espaço de dados, sem qualquer informação preliminar sobre os grupos.

4. Abordagem Proposta

A abordagem neste trabalho é identificar grupos com características semelhantes entre os OAs, diante do conhecimento que um OA está classificado dentro das dimensões de Felder and Silverman 1988. É considerado que uma combinação de estilos de aprendizagem (CEA) é uma tupla formada por 4 EAs, cada um pertencente a uma dimensão de Felder e Silverman sendo então possível ter 16 combinações [Dorça et al. 2013].

Um estudante que tem uma tendência forte para a característica Visual não exclui o fato que também é possível o aprendizado com uma característica Reflexiva ou Verbal. Assim este trabalho tem como desafio exibir uma proposta para a vetorização dos OAs considerando seus metadados. Assim, podemos modelar os estudantes de forma probabilística e dinâmica, que estocasticamente se enquadram a estas categorias pedagógicas ao longo do processo de aprendizagem com maior ou menor probabilidade dentro das 16 categorias possíveis.

Tabela 1. Estudante

Estilos de Aprendizagem							
Processamento		Percepção		Entrada		Entendimento	
Sen.	Int.	Ati.	Ref.	Vis	Ver.	Seq.	Glob.
17.0	83.0	35.0	65.0	80.0	20.0	84.0	16.0

O aluno representado pelo modelo descrito na Tabela 1, possui preferências para os estilos: intuitivo, reflexivo, visual e sequencial. Neste contexto, o maior vantagem desta abordagem é a de considerar estocasticamente todas as combinações de EAs de acordo com os alunos supondo que os EAs podem estar incoerentes ou podem mudar ao longo do tempo [Dorça et al. 2013]. A proposta se baseia na teoria das cadeias de Markov (MC), que é um sistema matemático que modela as mudanças de estado entre um número finito

Tabela 2. Relação Objetos de Aprendizagem x Felder e Silverman

Structure	Interactivity Type	Learning Resource Type	Interactivity Level
Global	Ativo	Sensitivo	Reflexivo
collection networked hierarchical	active mixed	simulation, graphindex, table, figure, experiment, self assesement	verylow low medium
Sequencial	Reflexivo	Verbal	Ativo
linear atomic	expositive mixed	exercise, questionnaire, index, table, narrative text exam, problem statement, self assesement, lecture	medium high very high
		Intuitivo	
		diagram	
		Visual	
		visual, diagram, figure graph, experiment	
		Ativo	
		exercice, simulation questionnaire,exam experiment, problema statement self assesement	
		Reflexivo	
		diagram, figure, graph index, slide, table narrative text, lecture	

de possíveis estados, e eles são muitas vezes descrito por um grafo direcionado, onde as bordas são rotulados pelas probabilidades de ir de um estado para os outro [Meyn and Tweedie 2009].

A Tabela 2 representa um estudo detalhado foi realizado para relacionar os EAs de Felder e Silverman com os OAs no padrão IEEE LOM, que foi construída baseado no trabalho de Resende et al. 2014. Foi proposto um estudo dos conceitos envolvidos em ambas as teorias, a fim de pesquisar e definir os pontos em que os OAs e os EAs se relacionam. OS campos *Structure*, *Interactivity Type*, *Learning Resource Type* e *Interactivity Level* são campos que estão presentes em um OA, sendo que cada um deles possui um determinado valor, por exemplo *Structure:collection* que se encaixa no EA Global. Assim obteve como resultado os campos que o padrão IEEE LOM representam os EAs atendidos por determinado objeto.

Um exemplo de vetorização de um OA poderia ser representado com os atributos: *Structure* com atributo do tipo *linear* resultando em um ponto para Sequencial, *Interactivity Type* sendo *expositive* somando um ponto em Reflexivo, *Learning Resource Type* com o atributo *table* somando um ponto em Sensitivo, Verbal e Reflexivo e por final *Inte-*

ractivity Level com atributo *very high* atribuindo um ponto em Ativo. Somando os pontos atribuídos a cada EA obtem-se o a vetorização da Tabela 3 no qual soma-se os pontos obtidos em cada EA relacionando o atributo que o ativou.

Tabela 3. Vetorização de um Objeto de Aprendizagem

Objeto de Aprendizagem x Felder e Silverman							
Ati.	Ref.	Vis.	Ver.	Seq.	Glo.	Sen.	Int.
1	2	0	1	1	0	1	0

Esta vetorização possui grande importância para que possa ser utilizado os algoritmos de clusterização comentados na seção 3.4. A relevância da clusterização é alta dentro do contexto apresentado pois tem-se que um objeto de aprendizagem pode ser clusterizado baseados nos 16 cluster. Um exemplo de um cluster com base no vetor da Tabela 3 que o objeto vetorizado pertence ao cluster: Reflexivo, Verbal, Sequencial e Intuitivo.

Diante disto um OA que pode ser utilizado para recomendação de um aluno que seja Reflexivo também poderá ser recomendado para um aluno que tenha também um perfil Verbal pois as técnicas de clusterização podem criar agrupamentos para os OAs e identificar uma relação com outros grupos de OAs fazendo com que tenha uma recomendação mais dinâmica para um determinado aluno.

5. Conclusão

Utilizar da disponibilidade de OAs de acordo com o perfil de aprendizagem do estudante apoia o processo de ensino e aprendizagem, auxilia também o aluno, através de abordagem de conteúdos que estejam em conformidade com as suas preferencias de aprendizado.

Um ponto importante é a relação entre as preferencias definidas no modelo do aluno com as categorias descritivas dos objetos de aprendizagem do padrão LOM. O relacionamento entre os EAs e os metadados dos objetos de aprendizagem permite ter uma maior concordância entre o perfil de aprendizagem identificado no aluno e os objetos de aprendizagem.

Esta abordagem proposta possibilitará os agrupamentos de OAs em grandes repositórios utilizando diferentes técnicas de agrupamentos mencionadas, sendo que será levado em questão os EAs e os metadados dos OAs. Com a vetorização e clusterização dos OAs, espera-se que a criação do processo de recomendação de OAs considerando os EAs dos estudantes. Espera-se que esta proposta permita facilitar o processo de recomendação personalizada de conteúdo considerando EAs em ambientes educacionais.

Como trabalho futuro tem como objetivo a implementação dessas técnicas no sistema Classroom eXperience [Araújo et al. 2014] para validação do modelo com estudantes reais para que possa avaliar uma melhoria em como o aluno evolui com um aprendizado baseado em suas competências.

Agradecimentos

Agradecemos a Pró-Reitoria de Pesquisa e Pós-Graduação da Universidade Federal de Uberlândia(UFU) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior(CAPES) por todo o suporte prestado no desenvolvimento deste trabalho.

Referências

- Aguiar, J. J., Fachine¹, J. M., and Costa, E. B. (2015). Recomendação de objetos de aprendizagem baseada na popularidade dos objetos e nos estilos de aprendizagem dos alunos.
- Araújo, R. D., Brant-ribeiro, T., Freitas, R. S. D., Dorça, F. A., and Cattelan, R. G. (2014). Autoria automática de objetos de aprendizagem a partir de captura multimídia e associação a estilos de aprendizagem. (Cbie):229–238.
- Audino, D. F. and da Silva Nascimento, R. (2012). Objetos de aprendizagem—diálogos entre conceitos e uma nova proposição aplicada à educação. *Revista Contemporânea de Educação*, 5(10).
- Branquinho, A., Lopes, C., Dorça, F., Fernandes, M., Cattelan, R., and de Ávila, A. J. N. (2015). Aquisição automática de competências num ambiente educacional ubíquo. 26(1):1207.
- Carvalho, V. C., Dorça, F. A., Cattelan, R. G., and Araújo, R. D. (2014). Uma abordagem para recomendação automática e dinâmica de objetos de aprendizagem baseada em estilos de aprendizagem.
- Casali, A., Gerling, V., Deco, C., and Bender, C. (2011). A recommender system for learning objects personalized retrieval. *Educational Recommender Systems and Technologies: Practices and Challenges*, pages 182–210.
- Cazella, S. C., Nunes, M., and Reategui, E. (2010). A ciência da opinião: Estado da arte em sistemas de recomendação. *JAI Jorn. Atualização em Informática da SBC. Rio Janeiro, RJ PUC Rio*, pages 161–216.
- da Silva, E. L., Café, L., and Catapan, A. H. (2011). Os objetos educacionais, os metadados e os repositórios na sociedade da informação. *Ciência da Informação*, 39(3).
- Dorça, F. A., Lima, L. V., Fernandes, M. A., and Lopes, C. R. (2013). Automatic student modeling in adaptive educational systems through probabilistic learning style combinations: a qualitative comparison between two innovative stochastic approaches. *Journal of the Brazilian Computer Society*, 19(1):43–58.
- Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231.
- Felder, R. M. and Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education*, 78(7):674–681.
- Felfernig, A., Friedrich, G., and Schmidt-Thieme, L. (2007). Guest editors' introduction: Recommender systems. *IEEE Intelligent Systems*, (3):18–21.
- Given, B. (2002). The overlap between brain research and research on learning style. In *Learning Styles: Realibility & Validity, Proceedings of the 7 th Annual ELSIN Conference*, pages 173–178.
- IEEE (2002). IEEE Standard for Learning Object Metadata.
- Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.

- Kohonen, Teuvo, P. (1998). Self-organizing maps of symbol strings. *Neurocomputing*, 21(1):19–30.
- Levis, D. and Barbosa, J. L. V. (2008). Aperfeiçoamento Automático do Perfil do Aprendiz.
- MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.
- Margarida, L., Tarouco, R., and Fabre, M.-c. J. M. (2003). Reusabilidade de objetos educacionais. pages 1–11.
- Meyn, S. and Tweedie, R. L. (2009). Markov chains and stochastic stability. prologue by peter w. glynn. cambridge mathematical library.
- Ochi, L. S., Dias, C. R., and Soares, S. S. F. (2004). Clusterização em mineração de dados. *Instituto de Computação-Universidade Federal Fluminense-Niterói*.
- Paper, C. and Mois, J. (2009). A Kohonen Network for Modeling Students ' Learning Styles in Web 2 . 0 Collaborative Learning Systems. (NOVEMBER).
- Pernas, A. M., Gasparini, I., de Oliveira, J. P. M., and Pimenta, M. S. (2009). Um ambiente ead adaptativo considerando o contexto do usuário (position paper). pages 1151–1156.
- Resende, D. T., Dorça, F. A., Cattelan, R. G., and Rafael, D. (2014). Em direção à recuperação automática de objetos de aprendizagem em repositórios através da associação dos estilos de aprendizagem de estudantes com metadados no padrão IEEE-LOM. (Cbie):445–454.
- Zaina, L. A. M., Cardieri, M. A. C. A., Engenharia, F. D., and Sorocaba, D. (2012). e-LORS : Uma Abordagem para Recomendação de Objetos de Aprendizagem. 20.
- Zapata, A. and Menendez, V. H. (2009). Discovering Learning Objects Usability Characteristics.

Computação Aplicada à Educação Musical: desafios e perspectivas ao planejamento docente no contexto da Educação Inclusiva

Amanda Meincke Melo¹, Carla Eugenia Lopardo², Guilherme Moreira de Melo²

¹Universidade Federal do Pampa (Unipampa) – Campus Alegrete
Av. Tiarajú, 810 – Ibirapuitã – 97.546-550 – Alegrete – RS – Brasil

² Universidade Federal do Pampa (Unipampa) – Campus Bagé
Av. Maria Anunciação Gomes de Godoi, 1650 – Malafaia – 97.546-550 – Bagé – RS – Brasil

{amanda.melo, carla.lopardo}@unipampa.edu.br, sai.guilherme@gmail.com

Abstract. *The Music content is mandatory in Brazilian schools and should be taught in inclusive educational settings. However, there is lack of solutions to allow teachers of Music, with or without disability, identify and reuse didactic and musical material in their education planning in inclusive educational settings, which take students characteristics and needs into account as well as their own situation. Thus the effective use of Computer in Education is placed at risk. This paper aims at contributing to this scenario, pointing out ways to approach the challenge of effective use of Computer in Education, emphasizing teacher planning in Inclusive Music Education.*

Resumo. *O conteúdo de Música é obrigatório nas escolas brasileiras e deve ser oferecido em contexto educacional inclusivo. No entanto, faltam soluções que possibilitem ao professor de Música, com ou sem deficiência, identificar e reutilizar materiais didático-musicais para o seu planejamento em contextos educacionais inclusivos, que considerem as características e necessidades de seus alunos e a sua própria condição. Compromete-se, assim, a efetivação do uso de tecnologias digitais na Educação. Este artigo busca contribuir com esse cenário, apontando caminhos para lidar com o desafio da efetivação do uso das tecnologias digitais na educação, com ênfase no planejamento docente no contexto da Educação Musical Inclusiva.*

1. Introdução

O ensino de Música, desde 2008, é considerado conteúdo curricular obrigatório da Educação Básica. Portanto, deve-se articular à Educação Inclusiva, que considera a Educação Especial uma modalidade transversal a todos os níveis, etapas e modalidades de ensino. Nesse cenário as tecnologias digitais têm importante papel a desempenhar.

Uma série de contribuições da Computação existe para atender a demandas da Educação Musical. Entretanto, observa-se que a adoção das tecnologias digitais na Educação Musical por professores com deficiência está comprometida pela forma incipiente como requisitos de acessibilidade ainda têm sido contemplados – ou deixam de ser contemplados – na produção de materiais didático-musicais e na organização

desses materiais para que possam ser amplamente reutilizados. Além disso, aspectos específicos da Educação Musical podem ser mais bem explorados na catalogação de materiais didático-musicais de modo a favorecer um melhor aproveitamento por professores da área, entre eles professores com deficiência.

Vislumbra-se como desafio, portanto, a efetivação do uso de tecnologias na educação, respeitando-se a multiplicidade das diferenças entre os integrantes de uma comunidade escolar. Em particular, tem-se em vista a contribuição da Computação à Educação Musical no processo de planejamento didático, proporcionando que professores, com ou sem deficiência, possam organizar suas aulas considerando as capacidades e as necessidades de seus alunos.

O artigo está organizado como segue. A Seção 2 colabora ao delineamento do desafio do ponto de vista da Educação Musical. A Seção 3 apresenta contribuições da comunidade de Informática na Educação para o domínio da Educação Musical, apontando para a possibilidade de contribuições da subárea Metodologias e Técnicas da Computação. A Seção 4 sugere um caminho para vencer o desafio com propostas que envolvem a colaboração de áreas como Engenharia de Software, Interação Humano-Computador, Semiótica Organizacional, Desenho Universal, Tecnologia Assistiva e Objetos de Aprendizagem (OA). Finalmente, são apresentadas as considerações finais deste artigo.

2. A Educação e o Desafio

O parecer CNE/CEB nº 12/2013, aprovado em 4 de dezembro de 2013, o qual estabelece as Diretrizes Nacionais para a operacionalização do ensino de Música na Educação Básica, define as bases para a implementação da Lei 11.769/08 que torna obrigatório o ensino de Música nas escolas como conteúdo, não exclusivo, da disciplina Artes. Este documento destaca a importância da presença da música no ambiente escolar determinando o compromisso que cada uma das esferas envolvidas deverá assumir, especificando as competências de cada setor (Escolas, Secretarias de Educação, Instituições de Ensino Superior, Ministério de Educação, Conselhos de Educação) nas diversas etapas e modalidades da inserção da música. Ainda, ao referir-se sobre a importância da presença da música no ambiente escolar e seus benefícios, o parecer salienta que “existem ligações específicas entre o estudo de Música e a habilidade de manipular informação tanto na memória de trabalho (usada para pensar), como na memória de longa duração (usada para arquivar os conteúdos aprendidos, os métodos e a experiência)”, desta forma, “o estudo de Música impacta a aprendizagem de outras áreas do conhecimento, além de formar comportamentos de atenção que impulsionam e melhoram a cognição”, promovendo, assim, a formação integral da criança (p. 6).

Com base neste parecer, conceitos como educação integral, interdisciplinaridade e inovação surgem como desafios para a área da Educação Musical, pois, como mencionado no documento, ela atua diretamente no cérebro estimulando a capacidade de decorar, de imitar, de reproduzir, de coordenar, de associar, de identificar, dentre outros aspectos relacionados à inteligência musical, área de estudo da neurociência; além de desenvolver a atenção executiva, necessária para formar memórias de qualquer área do conhecimento formal e de suas metodologias, influenciando indiretamente outras áreas do conhecimento.

Além disso, a prática musical escolar contribui para o desenvolvimento do convívio social da criança e do jovem, estimulando a interação social e formação de identidade cultural, fortalecendo os vínculos entre os membros de uma comunidade. Uma das formas de comunicação que o cérebro humano utiliza para estabelecer relações com o mundo é a música, capaz de comunicar sentimentos e emoções que a fala não consegue transmitir. Portanto, a prática musical cria “condições especiais de comunicação, tais como aquelas entre e com as pessoas com deficiência (surdas, cegas, mudas, com síndromes, com alterações em seu desenvolvimento, patologias, paralisia cerebral, entre outros)” (p.6).

Nesse contexto, faz-se necessário remeter também à Política Nacional de Educação Especial na Perspectiva da Educação Inclusiva, de 2008, que objetiva “o acesso, a participação e a aprendizagem dos estudantes com deficiência, transtornos globais de desenvolvimento e altas habilidades/superdotação nas escolas regulares”. Reitera, assim, o direito à educação em todos os níveis de ensino, sendo a educação especial uma modalidade transversal a todos os níveis, etapas e modalidades da educação. A escola atual, portanto, ao transformar-se para promover a educação inclusiva, evidencia ao professor a necessidade de formação e capacitação que o habilite a desenvolver sua prática num espaço de reconhecimento e de respeito às diferenças.

Esta também é uma necessidade do professor de Música, que precisa adaptar sua prática e aprimorar sua capacitação para oferecer aos alunos o ensino de Música, visando a proporcionar experiências musicais significativas a todos, indistintamente. É preciso ir além, garantindo a participação de todos os implicados nesse processo (ex.: professores, gestores, profissionais da educação, familiares etc.), indo ao encontro do Decreto 6.949/2009, que promulga a Convenção Internacional sobre os Direitos das Pessoas com Deficiência e, mais recentemente, da Lei 13.146/2015, que institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência).

A revisão da literatura na área da Educação Musical Inclusiva evidencia que grande parte dos trabalhos é voltada ao estudo da integração do aluno com deficiência na Educação Básica através do desenvolvimento de projetos interdisciplinares, contemplando também a possibilidade de articulação com a musicoterapia como área que abraça aspectos relacionados tanto à psicologia quanto à medicina. Esses trabalhos sugerem a inserção de recursos de Tecnologia Assistiva (TA) viabilizando a inclusão da pessoa com deficiência física na Educação Musical [Louro 2003]. Algumas pesquisas relatam experiências de músicos cegos na procura por subsídios para o ensino de Música, evidenciando as diversas dificuldades com as quais se deparam diante do acesso a materiais didáticos e ao perceberem as lacunas existentes na própria formação [Oliveira e Reily, 2014]. Nesse contexto, a Musicografia Braille se apresenta como uma técnica muito utilizada entre alunos e professores com deficiência visual, mas nem sempre acessível a todos. Por esse motivo ela representa também um desafio enquanto ensino-aprendizagem da própria técnica, sendo necessária maior difusão, a adaptação de materiais e a implementação de cursos que evite o autodidatismo em alunos e professores [Bonilha 2010][Tudissaky 2015].

No âmbito dos *software* para Educação Musical Inclusiva estão alguns programas voltados ao ensino de Música derivado da Musicografia Braille. Um exemplo disso é o *software* MUSIBRAILLE, mas existe um grande problema para os deficientes

visuais: a escassa quantidade de obras transcritas para essa técnica [Souza 2013]. O *software* CROMOTMUSIC, por outro lado, explora o uso de mecanismos e dispositivos de vibração para a aprendizagem musical de crianças e jovens surdos, a partir de uma abordagem musicoterapêutica [Rodrigues 2015].

Além dos recursos mencionados, existem outras inovações que desafiam os limites da comunicação e expressão através da música. Entre os recursos criados pelo Drake Music (2016) – um centro de tecnologia musical da Inglaterra – estão o SOUNDBEAM 5 e o SKOOG. Enquanto o primeiro permite a expressão através do som usando apenas movimentos em interface com um controlador MIDI, o segundo apresenta uma interface musical – um cubo de espuma para plataforma iPad – que permite criar, interpretar e ouvir com acessibilidade e dinamismo expressivo-musical dados pelo seu Desenho Universal. Ambos foram testados com usuários com deficiência.

Ao analisar avanços nas pesquisas sobre Tecnologia Assistiva na área da Educação Musical, observa-se que os *software* voltados ao ensino de Música para pessoas com algum tipo de deficiência, seja esta física, visual, auditiva, intelectual ou múltipla, visam à aprendizagem de elementos da linguagem musical relacionados com a leitura e escrita de música e a apreciação musical, mas nenhum deles centra a sua atenção na relação e articulação entre os diferentes modos de fazer, ouvir e criar música a partir dos eixos organizacionais da área, nem nos processos de ensino-aprendizagem musical implícitos nessas práticas, necessários para o planejamento do professor ou aluno com algum tipo de deficiência. Apresenta-se, assim, uma situação que exige a existência de soluções que auxiliem na elaboração de materiais didáticos, na construção de planos de aula e na avaliação da prática musical em sala de aula, tanto para uso do professor como dos alunos inseridos em contexto educativo inclusivo. Esses aspectos se constituem atualmente como uma das principais lacunas que não foram vencidas pelos trabalhos visitados.

Para além da revisão de literatura apresentada, algumas questões, que partem da realidade em que está inserida uma estudante cega, matriculada Curso de Licenciatura em Música do Campus Bagé da Unipampa, também provocam a refletir sobre a efetivação do uso de tecnologias na educação no contexto da Educação Musical Inclusiva: *Como esta estudante, em seu campo de estágio, daria aula sem enxergar para alunos que enxergam? Como uma pessoa cega ministraria aulas de teoria musical? De percepção musical? De instrumentos? Quais ferramentas tecnológicas estariam disponíveis para auxiliar estudantes em seus estágios? Onde encontrar informações sobre as tecnologias existentes para auxiliar em suas tarefas de docente? Como organizar em uma ferramenta os conteúdos que envolvem a Educação Musical em seus diferentes eixos organizacionais, garantindo sua acessibilidade universal?*

Para o aluno estagiário ou o próprio professor de Música, com ou sem deficiência, ambos atuantes em sala de aula, portanto, é um desafio significativo pensar, criar, organizar, planejar e avaliar aulas, materiais e recursos didáticos em variados formatos que possibilitem conciliar suas capacidades com as capacidades e as necessidades da turma ou de alunos específicos.

Nesse cenário, as tecnologias digitais na educação desempenham papel fundamental e podem ser desenhadas para atender a todos na maior extensão possível,

de acordo com o Desenho Universal, ou desempenhar o papel de recurso de Tecnologia Assistiva, promovendo funcionalidades e a autonomia de pessoas com deficiência [Melo 2014].

3. A Computação e o Desafio

Barvinks e Odakura (2012), ao abordarem o papel da Informática na Educação no contexto dos desafios educacionais para o Século XXI, apresentam desafios em que a Computação tem um papel a desempenhar, entre eles a centralização de Objetos de Aprendizagem ou *links* a esses, para uso livre de escolas e alunos, de modo que possam enriquecer a aprendizagem e auxiliar na superação de barreiras na aprendizagem. Sugerem que estruturas existentes, a exemplo do repositório RIVED do MEC, poderiam ser mais intensivamente adotadas, cuidando-se a atribuição de autoria dos objetos depositados. Os atuais repositórios de Objetos de Aprendizagem, em âmbito nacional, entretanto, não consideram características de acessibilidade desejáveis à promoção da Educação Inclusiva ou mesmo a especificidade da Educação Musical. Esta, além da organização de seus conteúdos e atividades por níveis de ensino, adotam eixos como ouvir, interpretar e criar, relevantes à organização e à recuperação de materiais didático-musicais.

Ficheman *et al.* (2004) apresentam o portal EduMusical como um ambiente multimídia e interativo no qual alunos, professores e visitantes podem encontrar aplicativos para a aprendizagem musical. Fundamentam-se no modelo (T)EC(L)A, que trata do envolvimento direto com a música (ex.: Execução, Composição e Apreciação) e do conhecimento sobre a música (ex.: Técnica e Literatura). O portal, no entanto, foi implementado com a tecnologia *Flash*, sem evidências em sua página principal sobre a oferta de alternativas aderentes a padrões contemporâneos de acessibilidade, que favoreçam sua adoção irrestrita a pessoas com deficiência [Melo 2014].

Viana Junior e Castro-Filho (2005), baseados no Modelo Espiral de Desenvolvimento Musical e o Modelo C(L)A(S)P – ou (T)EC(L)A – recomendam um conjunto de categorias para a avaliação de *software* de Educação Musical, a saber: aquisição do conhecimento, processo de aprendizagem, aspectos auditivos, exemplos musicais, formas de intervenção e atividades musicais. Estas podem, ainda, serem subdivididas entre concepção tradicional e concepção construtivista, de acordo com a abordagem pedagógica adotada para o ensino da Música.

Ferreira *et al.* (2008) apresentam uma arquitetura para a recuperação personalizada de OA, ilustrando como exemplo de utilização uma aplicação voltada ao processo de aprendizagem de ritmos brasileiros. O enfoque está na adaptação de conteúdo, baseada nas interações aluno-sistema. Para isso, utilizam redes de Kohonen, ou mapas auto-organizáveis – um tipo de rede neural – em um protótipo desenvolvido com infraestrutura de *software* livre.

Pinhati e Siqueira (2015) propõem a Plataforma Mignone, que abrange diretrizes para a construção de OA e a especificação de um Ambiente Virtual de Aprendizagem (AVA) para facilitar a construção de Objetos de Aprendizagem e sua utilização em atividades de Educação Musical. O projeto da plataforma contempla especificidades do domínio da Educação Musical ao adotar como referência o modelo (T)EC(L)A para a

produção de OA e a organização de atividades de Educação Musical. Entretanto, os autores não deixam claro como requisitos de acessibilidade e de usabilidade estão contemplados na proposta.

Pode-se mencionar, ainda, soluções para o contexto da Educação Musical, a exemplo do *software* Zorelha voltado ao desenvolvimento da percepção musical por crianças entre 4 e 6 anos [Jesus *et al.* 2010]; a proposta de uma interface musical tangível com tecnologia de Realidade Aumentada (RA), que pode ser usada como compositor musical, jogo de memória musical ou jogo de tempo coincidente musical, que foi avaliada com pessoas com Distrofia Muscular Duchenne e crianças com Paralisia Cerebral [Corrêa *et al.* 2012]; e a modelagem de um *software* para a Educação Musical de forma colaborativa, com auxílio do *framework* LUCID, mediada pela plataforma Amadeus [Venega e Sousa 2012].

Em síntese, a análise de trabalhos publicados nos eventos da comunidade Brasileira de Informática na Educação como o DesafIE – Workshop de Desafios da Computação Aplicada à Educação, o SBIE – Simpósio Brasileiro de Informática na Educação e o WIE – Workshop de Informática na Escola, além da RBIE – Revista Brasileira de Informática na Educação, permite mencionar uma série de contribuições para a área da Educação Musical, incluindo a publicação na *web* de materiais didático-musicais, a organização de AVA, diretrizes para a produção de OA e a avaliação de *software*, a oferta de *software* para o ensino da Música, para o exercício da percepção musical, para a expressão e a criação musical, para a o processo de reabilitação, a formação de professores de Música, entre outros.

A efetivação do uso das tecnologias digitais na Educação para o planejamento docente no contexto da Educação Musical Inclusiva pode se beneficiar, portanto, da aplicação de Metodologias e Técnicas da Computação que contemplem adequadamente o domínio da Educação Musical, observem princípios de acessibilidade e contribuam à organização e à recuperação de materiais educacionais digitais considerando aspectos desse domínio e características de acessibilidade, além de padrões existentes para a organização e recuperação de materiais educacionais digitais. Envolve, deste modo, contribuições de áreas como Engenharia de Software, Interação Humano-Computador, Desenho Universal e Objetos de Aprendizagem.

4. Proposta: Um Caminho para Vencer o Desafio

Na atuação interdisciplinar e interprofissional entre as áreas da Educação Musical e da Computação, torna-se possível contribuir à efetivação do uso das tecnologias digitais na Educação para o planejamento docente no contexto da Educação Musical Inclusiva. Propõe-se, assim:

1. A construção de conhecimento sobre o domínio da Educação Musical, suas partes interessadas, além dos requisitos trazidos pela Educação Inclusiva, com auxílio de métodos e técnicas da Computação, integrando conhecimentos e procedimentos da Engenharia de Software, da Interação Humano-Computador e da Semiótica Organizacional [Melo *et al.* 2012][Melo 2014];
2. Identificação, classificação e catalogação de matérias didático-musicais, considerando-se aspectos como níveis de ensino, eixos da Educação Musical

[Colwell 1992], características de acessibilidade e padrões para a catalogação de Objetos de Aprendizagem [Braga 2014];

3. Incentivo à construção de materiais didático-musicais com Desenho Universal;
4. Incentivo à produção de recursos de Tecnologia Assistiva voltados à Educação Musical;
5. Desenvolvimento de repositórios de OA, com Desenho Universal, que contemplem adequadamente o domínio da Educação Musical, mas que também possam ser integrados a outros repositórios [Braga 2015][Melo 2014];
6. Desenvolvimento de espaços de colaboração, com Desenho Universal, para o fazer docente em Educação Musical, em rede, contemplando aspectos como planejamento, práticas didático-musicais e avaliação.

Desse modo, vislumbra-se um caminho possível para a Computação contribuir à Educação Musical no processo de planejamento didático, proporcionando que professores, com ou sem deficiência, possam organizar suas aulas considerando as capacidades e as necessidades de seus alunos em perspectiva inclusiva.

5. Considerações Finais

Este artigo problematizou o uso de tecnologias digitais no contexto da Educação Musical Inclusiva no processo de planejamento docente, que também deve considerar professores com deficiência como usuários dessas tecnologias. Uma série de soluções foi apresentada no contexto da Educação Musical, mas essas não contemplam adequadamente o planejamento docente amplamente acessível.

Metodologias e Técnicas da Computação podem ser colocadas a favor desse desafio, para a melhor compreensão do domínio da Educação Musical Inclusiva, para a organização e o desenvolvimento de materiais didático-musicais acessíveis, para apoiar professores no planejamento didático que contemple amplamente a todos, inclusive o próprio professor. Para isso, faz-se necessária a organização de iniciativas que alinhem a colaboração interdisciplinar e interprofissional entre as áreas da Educação Musical e da Computação.

Transcende-se, assim, à aplicação de normas técnicas de acessibilidade, envolvendo também o alinhamento de referenciais e de práticas de pesquisa, a análise e a revisão de práticas, a identificação e a avaliação de recursos, a concepção de novas tecnologias digitais, além da organização em rede de profissionais da Educação Musical.

Agradecimentos

Às acadêmicas Rosane de Fraga, da Licenciatura em Música, e Lisiane dos Santos da Rosa, do Bacharelado em Ciência da Computação, pelas importantes contribuições ao desenvolvimento do projeto de pesquisa, em andamento, “Educação Musical Inclusiva mediada por Tecnologias Digitais Acessíveis”.

Referências

Barvinski, C. A., Odakura, V. V. A. “Desafios educacionais para o Século XXI e o papel da Informática na Educação”, DesafIE 2012, SBC, Brasil.

- Bonilha, F. F. G. (2010), *Do toque ao som: o desafio da Musicografia Braille como um caminho para a educação musical inclusiva*, Unicamp (Tese de Doutorado).
- Braga, J. (2014), *Objetos de Aprendizagem volume 1 – introdução e fundamentos*, Editora UFABC.
- Braga, J. (2015), *Objetos de Aprendizagem volume 2 – metodologia de desenvolvimento*, Editora UFABC.
- Brasil (2008) “Lei n. 11.769, de 18 de Agosto de 2008”, http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/111769.htm
- Brasil (2009) “Decreto n. 6949, de 25 de Agosto de 2009”, http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/decreto/d6949.htm
- Brasil (2015) “Lei n. 13.146, de 6 de Julho de 2015”, http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13146.htm
- CNE/CEB (2013) “Diretrizes Nacionais para a operacionalização do ensino de Música na Educação Básica”, http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=14875-pceb012-13&category_slug=dezembro-2013-pdf&Itemid=30192
- Corrêa, A. G. D., Ficheman, I. K., Lopes, R. D. “O Fazer Musical de Pessoas com Deficiência: as novas tecnologias propiciando a inclusão”, SBIE 2012, SBC, Brasil.
- Colwell, R. (1992), *Handbook of research on music teaching and learning*, Schirmer.
- Drake Music (2016?), <http://www.drakemusic.org/>
- Ficheman, I. K., Lopes, R. D., Krüger, S. E., Bassani Neto, O. “PORTAL EDUMUSICAL: Telemática aplicada à Educação Musical”, SBIE 2004, SBC, Brasil.
- Jesus, E. A., Uriarte, M. Z., Raabe, A. L. A. (2010). Zorelha: um objeto de aprendizagem para auxiliar o desenvolvimento da percepção musical em crianças de 4 a 6 anos. In *Revista Brasileira de Informática na Educação*, p. 91–105. SBC.
- Louro, V. S. (2003), *As adaptações a favor da inclusão do portador de deficiência física na educação musical: um estudo de caso*, Unesp (Dissertação de Mestrado).
- MEC/SECADI (2008) “Política Nacional de Educação Especial na Perspectiva da Educação Inclusiva”, <http://portal.mec.gov.br/arquivos/pdf/politicaeducespecial.pdf>
- Melo, A. M. (2014) “Acessibilidade e Inclusão Digital em Contexto Educacional”, *Anais da 3ª Jornada de Atualização em Informática na Educação*, M. A. S. N. Nunes, E. M. Rocha, Brasil, Sociedade Brasileira de Computação, p. 1-41.
- Melo, A. M., Saldanha, J. F., Wernz, M. C. G. (2012) “Desafios à pesquisa em Computação em contexto educacional - qualidade no uso de objetos de aprendizagem em perspectiva”, *DesafIE 2012*, SBC, Brasil.
- Oliveira, L. A. C., Reily, L. H. (2014). Relatos de músicos cegos: subsídios para o ensino de música para alunos com deficiência visual. In *Revista Brasileira de Educação Especial*, p. 405–420.

- Pinhati, F. Siqueira, S. W. M. (2015). Plataforma Mignone: Ambiente Virtual de Aprendizagem e Objetos de Aprendizagem Especializados para a Educação Musical. In *Revista Brasileira de Informática na Educação*, p. 1–17. SBC.
- Rodrigues, I. O. (2015), Os efeitos da musicoterapia através do software Cromotmusic em aspectos sensoriais, emocionais e musicais de crianças e jovens surdos : ensaio controlado randomizado, UFRGS (Dissertação de Mestrado).
- Souza, L. S. (2013), O uso do software Musibraille no processo de ensino-aprendizagem musical de alunos com deficiência visual: da transcrição à utilização pedagógica, UFPB (Trabalho de Conclusão de Curso).
- Tudissaky, S. E. (2015), Ensino de música para pessoas com deficiência visual, Unesp (Dissertação de Mestrado).
- Venega, V. S., Sousa, W. P. “Modelagem Participativa de um Software de Ensino Musical”, DesafIE 2012, SBC, Brasil.
- Viana Júnior, G. S., Castro-Filho, J. A. “Avaliação de Software para o Ensino de Música: reconhecendo a singularidade do discurso musical”, WIE 2005, SBC, Brasil.

Discussão sobre a Adaptabilidade em Ambientes *U-learning* baseada em Estilos de Aprendizagem

Luma da Rocha Seixas¹, Ivanildo José de Melo Filho^{1,2}, Alex Sandro Gomes¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7.851 – 54.740-540 – Recife – PE – Brazil

²Instituto Federal de Pernambuco – Campus Belo Jardim (IFPE)

Av. Sebastião Rodrigues da Costa, S/N – CEP: 55155-730 – Belo Jardim – PE – Brazil

{lrs3,asg}@cin.ufpe.br, ivanildo.melo@belojardim.ifpe.edu.br

Abstract. *This article aims to present and discuss with this community the adaptability process in the U-Learning environments. Such environments need to consider the identification of user needs, mobile technologies and the different learning situations that may occur. It is known that different environments make use of different recognition strategies adaptation process based on profiles. The literature emphasizes the limiting nature of this kind of recognition in these environments, mainly in and not consider the different learning styles.*

Resumo. *Este artigo tem como objetivo apresentar e discutir com esta comunidade o processo de adaptabilidade nos ambientes de U-Learning. Tais ambientes precisam considerar a identificação das necessidades do usuário, tecnologias móveis bem como as diferentes situações de aprendizagem que podem ocorrer. É sabido que diferentes ambientes fazem uso de diferentes estratégias de reconhecimento de processo de adaptação baseado em perfis. A literatura enfatiza o caráter limitante deste tipo de reconhecimento nesses ambientes, principalmente e por não considerar os diferentes estilos de aprendizagem dos aprendizes.*

1. Introdução

A possibilidade do acesso instantâneo a informações e a conteúdos tem fomentado o uso de tecnologias, fazendo com que as práticas em sala de aula busquem diferentes maneiras de incentivar um aprendizado contextualizado, motivado por demandas que podem ser específicas em função do conteúdo, do objetivo, bem como, da sua localização ou pelo tempo que se pretende alcançar.

Pesquisas tem apontado fortemente para uma aprendizagem individualizada e colaborativa na educação escolar (KONG et al., 2014). As salas de aula presenciais mantêm a sua importância na interação e socialização dos aprendizes. Contudo, as situações de aprendizagem fora da sala de aula tem superado as expectativas propostas pelo ensino tradicional e, conseqüentemente desempenham um papel importante na construção do conhecimento dos aprendizes (KONG et al., 2014).

Chen e Li (2010) e Peng et al., (2008) evidenciam que os ambientes de aprendizagem ubíqua (*u-learning*) tem se posicionado como uma ferramenta com potencial no que diz respeito a adaptabilidade e flexibilidade para diferentes contextos de aprendizagem. Outras pesquisas relatam ainda significativa eficiência na aprendizagem nesses contextos (CHU et al., 2010; CHU, HWANG and TSAI, 2010; YANG, 2006). Barbosa (2007) enfatiza que nesses ambientes o modelo computacional deve atuar de forma invisível (*background*) e proativamente diante das necessidades dos aprendizes em suas atividades cotidianas, com base nas tecnologias, comunidades de aprendizagem e de interação social, nas redes de comunicação e nos demais serviços disponíveis.

Para Durán et al., (2014) ambientes *u-learning* incorporam um modelo de aprendizagem *ad-hoc*: aprendizagem adaptativa. Este modelo baseia-se na ideia de adaptar as metodologias de aprendizagem as formas de aprender dos alunos para proporcionar um conhecimento específico e personalizado quando necessário. Sendo assim, é possível compreender que a aprendizagem ubíqua busca combinar as vantagens da aprendizagem adaptativa, com os benefícios da computação ubíqua. E, ao contrário de modelos de aprendizagem desenvolvidos na última década e implementados por meio de uma arquitetura cliente-servidor, a aprendizagem ubíqua tem sido caracterizada por proporcionar maneiras intuitivas para identificar os conteúdos, serviços e colaboradores para uma aprendizagem no lugar e no momento adequado (BRANQUINHO et al., 2015).

No sentido de introduzir a adaptabilidade, iniciativas como Brusilovsky, Eklund e Schwarz (1998), Zakaria e Brailsford (1998) e Wu, Kort e Bra (2011) começaram a ser observadas nos últimos anos com o objetivo de personalizar o processo de aprendizagem. Em outras palavras, procuram proporcionar mecanismos personalizados para que o aprendiz possa selecionar caminhos dentro das possibilidades ofertadas pelo professor ou tutor (BRA e CALVI, 1998).

É preciso reconhecer que diversos são os sistemas que utilizam o perfil do aprendiz, porém, o uso deste para auxílio no processo de aprendizagem em um cenário ubíquo é recente (OGATA et al., 2005; BARBOSA, GEYER e BARBOSA, 2005; GASPERINI et al., 2004). Um ponto importante destacado por Aliança Neto, Neto, e Moreira (2014) refere-se a captura e manutenção eficaz do perfil do usuário, que de acordo com os autores não é uma atividade trivial. Sendo portanto, um dos problemas mais recorrentes neste processo a captura da informação.

Este artigo tem como propósito refletir a cerca de como ambientes de *u-learning* podem ser constituídos como uma alternativa para se adequar eficazmente às características dos aprendizes, principalmente em situações de contexto urbano. Dentro desse contexto de aprendizagem ubíqua, a adaptabilidade e a flexibilidade são pontos chaves para um melhor entendimento das reais necessidades dos aprendizes. Um ponto desafiador que permeia essa discussão concentra-se na percepção sobre o perfil dos aprendizes. Compreender o perfil do aprendiz, levando em consideração aspectos associados às suas características e o contexto ao qual ele encontra-se sinaliza como uma oportunidade a ser explorada nesse campo.

A organização desse documento é descrita a seguir: a seção 2 descreve trabalhos correlatos na comunidade científica brasileira. A seção 3 discute estudos que apontam o

uso de estilos de aprendizagem como alternativa para personalizar os ambientes. As considerações finais são sintetizadas na seção 4.

2. Trabalhos correlatos

Na aprendizagem ubíqua, a percepção sobre os elementos que envolvem o aprendiz transcende o virtual e alcança o mundo real (LEVIS et al., 2008). Um dos principais desafios nesse campo de pesquisa tem como ponto central o perfil do aprendiz ubíquo. Ao longo dos anos, pesquisas no âmbito nacional tem apresentado iniciativas com o suporte de ambientes de computação ubíqua, no intuito de promover uma aprendizagem personalizada. A seguir são apresentados 3 (três) ambientes de u-learning e suas estratégias de reconhecimento do perfil dos aprendizes.

- **LOCAL (*LO*cation and *CO*ntext *A*ware *L*earning)**

Barbosa et al., (2008) propõe um modelo que explora sistemas de localização e o suporte à contexto como instrumentos de apoio ao ensino. Os autores apresentam um modelo chamado de LOCAL (*LO*cation and *CO*ntext *A*ware *L*earning), trata-se de um sistema de localização que acompanha a mobilidade dos aprendizes e, baseado nas suas posições físicas, explora oportunidades educacionais.

Para identificar o perfil dos usuários, o modelo utiliza o padrão PAPI (*Public and Private Information for Learners*) (PAPI 2006). Basicamente, existem informações persistentes (Contato, Preferências e Interesses) que acompanham o aprendiz independentemente do contexto. Essas informações são armazenadas no Assistente Pessoal (AP) que acompanha o dispositivo móvel. Além disso, há ainda informações contextuais (Relacionamentos, Desempenho e Segurança), as quais estão ligadas aos contextos por onde o aprendiz se desloca.

O LOCAL gerencia ambos os tipos para exploração de oportunidades pedagógicas de acordo com o deslocamento dos alunos. O sistema de localização informa para o tutor, o contexto onde está o aprendiz. O professor e o tutor podem usar essa informação, aliada ao perfil do aprendiz, para determinação dos objetos relevantes no contexto. A proposta deste ambiente evidenciou fortes limitações no que tange às inferências pedagógicas do tutor.

Identificando esse problema, Levis et al., (2008) desenvolveu o PeLeP (*Pervasive Learning Profile*), cujo propósito foi refinar o rastreamento das atividades de aprendizagem no ambiente baseado em uma análise abrangente do perfil do aprendiz. Para tanto, é utilizado o histórico do aprendiz nos contextos do ambiente ubíquo, durante um determinado período de tempo. São armazenadas informações referentes à localização (contextos no qual o aprendiz esteve), recursos acessados (que objetos de aprendizagem foram acessados), dispositivos utilizados (registra dispositivos como *tablet*, PC, *notebook*) e aplicativos utilizados (quais programas foram acessados pelo aprendiz).

Ainda utilizando o LOCAL, Hahn e Barbosa (2008) introduziram uma arquitetura de assistente pessoal e modelado como um agente capaz de explorar automaticamente ambientes de educação ubíqua, com especial cuidado para a detecção e transição de ambientes. O modelo contribuiu na personalização dos processos pedagógicos individuais (através da recomendação de recursos contextualizados e

relevantes) e em grupo (através do estímulo livre à interação, proporcionado pela análise de similaridade). A principal contribuição do modelo proposto é a organização dos espaços de aprendizagem ubíqua, onde os agentes tem papel de destaque.

- **Global**

Explorando uma outra alternativa, Oliveira e Barbosa (2009) apresentaram um modelo de educação ubíqua descentralizado, baseado em sistemas multi-agentes, denominado Global. Este, disponibiliza agentes de *software* que executam tarefas comuns ao processo de aprendizagem ubíqua. Sendo sua customização direcionada para a criação de novos ambientes de aprendizagem ubíqua, através da extensão desses agentes ou adição de novos.

- **CoolEdu**

Trata-se de um ambiente descentralizado de educação ubíqua, Rabello et al., (2012) apresentam um modelo para auxiliar no processo de ensino e aprendizagem através do fomento à colaboração entre os aprendizes de forma contínua. Esta abordagem foi modelada usando agentes e empregando um modelo sócio-interacionista.

A estratégia para reconhecimento do perfil dos aprendizes ocorre por meio de categorias. Dentre elas, há a categoria *History* que agrupa as informações que o aprendiz está trabalhando, ou já trabalhou. Também é armazenado o contexto onde a dupla interagiu e uma nota, dada pelo próprio aprendiz, referente ao quanto foi útil essa interação. Estes dados podem mudar continuamente e servem como um *cache* do que está sendo ensinado e aprendido pelo aprendiz, e como um histórico do que já foi ensinado e aprendido.

- **UbiGroup**

Em Ferreira, Barbosa e Gluz (2013), os autores propõe um modelo de recomendação ubíqua de conteúdo educacional para grupos de aprendizes, o *UbiGroup*. Seu objetivo é recomendar Objetos de Aprendizagem (OA) considerando de forma integrada os perfis dos aprendizes e o contexto onde eles estão inseridos. Nesse caso, o modelo conta com um agente Gestor de Perfis (GP) que mantém os perfis dos aprendizes atualizados, realizando ainda um cálculo de similaridade entre eles.

Com base nesse perfil, contexto e a regra de recomendação contida no contexto, o agente define os filtros de pesquisa a serem utilizados e os envia para um agente Comunicador (CO) que efetua a comunicação com os repositórios de OAs. Este agente executa a consulta no repositório e retorna o resultado para o Agente Recomendador (RE). Após a execução das recomendações individuais para os perfis representativos, são unificadas em uma única lista, constituindo uma ordenação prévia de OAs a serem recomendados.

Os ambientes descritos acima, apresentam diferentes estratégias de reconhecimento do perfil dos aprendizes. É preciso reconhecer que mesmo diante das diferentes perspectivas sobre as quais os ambiente ubíquos são destinados, ainda não há uma compreensão cognitiva clara associada aos estilos de aprendizagem existentes. Este fato, gera uma limitação sobre a capacidade desses ambientes alinharem seus requisitos com a perspectiva pedagógica em oferecer um ambiente personalizado. Além de

ambientes de *u-learning*, outras estratégias buscam identificar o perfil do aprendiz ubíquo.

Em Mendonça et al., (2014) é apresentado um modelo de arquitetura de *software* que suporta atividades colaborativas de extensão e enriquecimento de conteúdo digital por meio de anotações e ranqueamento (classificação). Os elementos da arquitetura são capazes de prover informações para serviços de recomendação e personalização da informação no contexto da aplicação, unidos a uma infraestrutura dos ambientes ubíquos de captura e acesso. Esses visam proporcionar aos usuários a reconstituição da experiência vivida em sala de aula, complementando-a com a possibilidade de interações colaborativas e sociais.

Aliança Neto, Neto, e Moreira (2014) apresentam uma abordagem para a construção implícita de perfis de usuário de maneira genérica, independente do domínio de aplicação e facilmente adaptável a sistemas de recomendação de conteúdo. O perfil do usuário representa principalmente seus interesses, que na pesquisa são identificáveis a partir dos conteúdos dos sites acessados pelo estudante.

No momento em que um conteúdo (*site*, vídeo, rede social, etc.) é acessado, um componente analisador semântico identifica e isola as anotações e/ou marcações já existentes, como HTML e XML. O conteúdo restante é semanticamente anotado com conceitos provenientes das ontologias de domínio. Os dados das anotações semânticas do conteúdo sendo enriquecido são transmitidos para o indexador semântico, que seleciona dados importantes com base nos já cadastrados na ontologia de perfil onde os dados tratados são indexados e armazenados para constituir o perfil individual do usuário.

Em Branquinho et al., (2015), é apresentado uma proposta que tem como objetivo proporcionar um ambiente de computação ubíqua com uma ferramenta que possibilite prover ao usuário um mecanismo para reforço ou alcance de um conjunto de competências. A abordagem faz uso de objetos de aprendizagem adequados ao seu estilo de aprendizagem, que pode ser determinado automaticamente pelo sistema. Na solução do problema de sequenciamento de objetos considerou-se estratégias evolucionárias e uma técnica de busca e, pelos experimentos realizados, chegou-se a resultados melhores fazendo uso de um algoritmo evolucionário *steady-state*.

É possível observar que além das características nativas dos ambientes ubíquos, estratégias auxiliares são utilizadas no sentido de enriquecer o perfil do aprendiz. Contudo, em ambas as situações é negligenciado o alinhamento sobre a percepção pedagógica.

3. Estilos de Aprendizagem como Alternativa para Identificar o Perfil do Aprendiz

Hwang (2006) estabeleceu um conjunto de características de ambientes *u-learning* e que constituem também desafios para designers. Segundo ele, tais ambientes precisam ser: sensíveis ao contexto, fornecer suporte para personalização além de permitir uma aprendizagem no qual e ser capaz de se adaptar temas de conteúdo.

A adaptabilidade desempenha um papel importante na aprendizagem ubíqua, objetivando proporcionar aos aprendizes material de aprendizagem adaptado e

personalizado, atividades, e informações no lugar e momento certos (BRANQUINHO et al., 2015). Com isso, essa discussão propõe como alternativa para a identificação perfil do aprendiz ubíquo, o uso de estilos de aprendizagem, permitindo que o ambiente e os materiais nele apresentados se adequem ao perfil e propiciem, assim, um tratamento personalizado.

Existem na literatura diversos modelos de estilos de aprendizagem, contudo pesquisas como Kuljis e Liu (2005) e Carver et al., (1999) apontam que ao se tratar de ambientes educacionais adaptativos, o Modelo de Estilos de Aprendizagem Felder-Silverman (do inglês *Felder-Silverman Learning Styles Model - FLSM*) descrito em (Felder e Silverman, 1988) pode ser visto como o mais utilizado para classificação de estilos de aprendizagem.

De acordo com o modelo FLSM, o aluno pode ser caracterizado em quatro dimensões de estilos de aprendizagem. Estas dimensões podem ser identificadas respondendo-se à questionamentos sobre o tipo de informação que o aluno preferencialmente percebe, qual canal sensorial uma informação externa é mais eficientemente, como o aluno prefere processar a informação e como o aluno progride seu entendimento no assunto.

Foram encontradas na literatura algumas abordagens que utilizam o perfil do aluno para permitir uma personalização de ambientes ubíquos. A proposta dessa discussão é o uso de estilos de aprendizagem para auxiliar na personalização de ambientes de aprendizagem ubíqua.

Relacionar estilos de aprendizagem dos estudantes com ações instrucionais apropriadas é um fator importante ao estímulo da aprendizagem (CHU et al., 2010). Pesquisas têm apontado que a aprendizagem é facilitada se os métodos de ensino utilizados estiverem de acordo com os seus estilos de aprendizagem (GRAF e LIU, 2008; COFFIELD et al., 2009; HAIDER, SINHA e CHAUDHARY, 2010). Além disso, ao identificar o estilo de aprendizagem é possível se beneficiar, obtendo informações sobre como os alunos aprendem, o que lhes proporciona uma compreensão profunda e que tipo de materias frequentemente atendem às suas necessidades (GRAF et al., 2009).

Tal proposta, está alinhada à pesquisas como Graf et al., (2008), Kinshuk et al., (2009) e Graf et al., (2009), no qual identificam que a aprendizagem é facilitada se as estratégias pedagógicas estiverem de acordo com o estilo de aprendizagem do aprendiz, tornando o processo de aprendizagem efetivo e melhorando consideravelmente sua performance.

Propostas para detecção automática de estilos de aprendizagem (GRAF e VIOLA, 2009; LIMONGELLI et al., 2009) têm sido exploradas. Elas utilizam sistemas de inferência determinísticos baseados em padrões de comportamento pré-definidos do aprendiz, de forma que, através de seu comportamento e ações no sistema, possam inferir seus estilos de aprendizagem.

Contudo, tais abordagens ignoram importantes considerações relacionadas aos aspectos dinâmicos e não-determinísticos de estilos de aprendizagem e do comportamento do aprendiz. Nesse contexto, Dorça et al., (2014) apresentou uma abordagem automática, dinâmica e probabilística para modelagem de estilos de aprendizagem, avaliando e comparando através de experimentos, três diferentes políticas de aprendizagem por reforço.

Paredes et al., (2005) também propõem a utilização de modelos representados através de valores probabilísticos. Os autores apresentam a concepção e implementação de um sistema centralizado com uma abordagem baseada em múltiplos modelos para apoiar a personalização de ambientes de aprendizagem ubíqua. A arquitetura do sistema é baseada em informações sobre o aplicativo, características do usuário, comportamento e o ambiente no qual ele está inserido. Ao final, é gerado um estereótipo a partir das características estabelecidas no modelo.

Chiou e Tseng (2012) propõem uma navegação personalizada e sensível no contexto *u-learning* baseada em estilos de aprendizagem. Os autores utilizaram a Teoria da Orientação da Aprendizagem (*Learning Orientation Theory*) como estratégia para reconhecer os estilos de aprendizagem dos alunos. Tal teoria leva em consideração:

- i. A influência de um conjunto abrangente de fatores psicológicos que influenciam a aprendizagem;
- ii. O impacto das emoções e intenções na aprendizagem;
- iii. As relações entre os aprendizes nos ambientes de aprendizagem e a capacidade de aprendizagem;
- iv. A perspectiva abrangente para projetar ambientes para suporte à aprendizagem.

Existem estudos que utilizam técnicas de aprendizagem supervisionada de máquina, tais como redes neurais e redes Bayesianas (Carmona e Castillo, 2008; Zatarain-Cabada et al., 2009). A dificuldade dessas abordagens é a elevada complexidade de implementação e grande consumo de recursos computacionais de forma que a presença simultânea de vários estudantes pode comprometer, ou mesmo inviabilizar seu uso.

Apesar da literatura apresentar várias iniciativas no que diz respeito à identificação de perfis, ainda existem algumas lacunas a serem preenchidas para a melhor prática ensino. Faz-se necessário reforçar que em um cenário ubíquo, o perfil do aprendiz constitui-se um elemento fundamental, em função das inúmeras possibilidades de interatividade e as diversas tecnologias de acesso a informação. Desta forma, torna-se necessário que o ambiente conheça em profundidade o aprendiz e o auxilie a interagir com a diversidade de elementos do ambiente e oportunidades de interação à disposição.

4. Considerações Finais

Este trabalho, conforme descrito em sua introdução, procura evidenciar a importância de considerar a adaptabilidade nos ambientes de aprendizagem ubíquos. Para isso foram apresentados estudos que já apontam alternativas para a personalização desses ambientes, bem como, quais procedimentos foram utilizados.

Diversas iniciativas seja na construção de ambientes ou na proposição de estratégias têm oferecido diferentes caminhos para uma compreensão e atendimento às necessidades dos aprendizes. É importante destacar que cada iniciativa possui sua particularidade e nos casos descritos percebe-se um desalinhamento entre o que é desenvolvido com o potencial pedagógico que pode ser explorado.

Considerando que dentro do contexto de aprendizagem ubíqua a adaptabilidade e a flexibilidade são pontos chaves para um melhor entendimento das reais necessidades dos aprendizes, um ponto desafiador que permeia essa discussão concentra-se na percepção sobre o perfil dos aprendizes. Entre os pontos apresentados, compreender o perfil do aprendiz, levando em consideração aspectos complementares associados ao seu estilo de aprendizagem sinaliza como uma oportunidade a ser explorada neste campo.

Agradecimento

Luma da Rocha Seixas é bolsista de doutorado da Fundação de Amparo a Ciência e Tecnologia de Pernambuco – FACEPE – Processo N°. PBPG-0540-1.03/15.

Referências

- Barbosa, D. N. F. (2007). Um Modelo de Educação Ubíqua Orientado à Consciência do Contexto do Aprendiz. Tese de Doutorado, Unisinos.
- Barbosa, C.F.R., Geyer, J.L.V. Barbosa. (2005) GlobalEdu: An Architecture to Support Learning in a Pervasive Computing Environment. In IFIP Workshop on Educational Technology, Perth, páginas 1-10.
- Barbosa, J. L. V., Hahn, R. M., Rabello, S., Barbosa, D. F. N. (2008) LOCAL: a Model Geared Towards Ubiquitous Learning. In 39th ACM Technical Symposium on Computer Science Education (SIGCSE). New York, páginas 432-436, ACM Press
- Branquinho, A., Lopes, C., Dorça, F., Fernandes, M., Cattelan, R., and de Ávila, A. J. N. (2015). Aquisição Automática de Competências num Ambiente Educacional Ubíquo. In Anais do Simpósio Brasileiro de Informática na Educação (Vol. 26, No. 1, p. 1207).
- Bra, P. De and Calvi, L. AHA! (1998) An open adaptive hypermedia architecture. New Review of Hypermedia and Multimedia, vol. 4, no. 1, pp. 115–139, 1998.
- Brusilovsky, P., Eklund, J. and Schwarz, E. (1998). Web-based education for all: A tool for developing adaptive courseware. in Seventh International WWW Conference, pp. 291–300, Elsevier Science B.V.
- Carver, C. A., Howard, R.A. , and Lane W.D. (1999). Addressing different learningstyles through course hypermedia. IEEE Transactions on Education, 42(1), 33-38.
- Carmona, C. e Castillo, G. (2008). Designing a Dynamic Bayesian Network for Modeling Students Learning Styles. In Eighth IEEE International Conference on Advanced Learning Technologies, pages 346–350. IEEE.
- Chen, C.; Li, Y. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. Interactive Learning Environments, v. 18, n. 4, p. 341-364.
- Chiou, C. K., and Tseng, J. C. (2012). Design of a personalized navigation support system for context-aware ubiquitous learning environment. In Proceedings of the 2012 RecSys workshop on Personalizing the local mobile experience (pp. 1-6). ACM.

- de Aliança Neto, A. S., Neto, F. M. M., & Moreira, J. D. (2014). Uma Nova Abordagem para o Perfilamento de Usuários em Ambientes de Aprendizagem Ubíqua. In Anais do Simpósio Brasileiro de Informática na Educação (Vol. 25, No. 1, p. 1243).
- Dorça, F. A., Lima, L. V., Fernandes, M. A., and Lopes, C. R. (2014). Avaliação de Políticas de Aprendizagem por Reforço para Modelagem Automática e Dinâmica de Estilos de Aprendizagem: Uma Análise Experimental. *Revista de Informática Aplicada*, 9(2).
- Durán, Elena B., Margarita M. Álvarez, and Silvina I. Únzaga (2014). Ontological model-driven architecture for ubiquitous learning applications."Proceedings of the 7th Euro American Conference on Telematics and Information Systems. ACM.
- Felder, R. M.; Silverman, L. K. (1988) Learning and teaching styles in engineering education. *Engineering education*, v. 78, p. 674–681.
- Ferreira, L. G. A., Barbosa, J. L. V., and Gluz, J. C. (2013). Um Modelo de Recomendação Ubíqua de Conteúdo para Grupos de Aprendizes. In Anais do Simpósio Brasileiro de Informática na Educação (Vol. 24, No. 1, p. 697).
- Gasperini, M. A. I., Amaral, Pimenta, M. S., Oliveira, J. P. M. (2004) Navegação Adaptativa no Ambiente de EAD AdaptWeb: uso de XML na Adequação de Conteúdo e Navegação. In III Congresso Internacional de Ergonomia e Usabilidade Design de Interfaces e Interação Humano-Computador, páginas 124-130.
- Graf, S., Kinshuk, and Liu, T. (2009). Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach. *Journal of Educational Technology & Society*, 12(4):3–14.
- Graf, S., Liu, T.-C., and Kinshuk (2008). Interactions Between Students Learning Styles, Achievement and Behaviour in Mismatched Courses. In Proceedings of the International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2008), pages 223–230. IADIS International Conference.
- Graf, S.,; Viola, S. (2009). Automatic student modelling for detecting learning style preferences in learning management systems. Citeseer.
- Hahn, R. M., e Barbosa, J. L. V. (2008). Uma Arquitetura de Assistente Pessoal Orientada a Ambientes de Aprendizagem Ubíqua. In Anais do Simpósio Brasileiro de Informática na Educação (Vol. 1, No. 1, pp. 736-745).
- Hwang G. (2006)Criteria and Strategies of Ubiquitous Learning. Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC'06), IEEE.
- Kinshuk, Liu, T., and Graf, S. (2009). Coping with Mismatched Courses: Students' behaviour and performance in courses mismatched to their learning styles. *Educational Technology Research and Development*, 57(6):739–752.
- Kong, S. C., Chan, T. W., Griffin, P., Hoppe, U., Huang, R., Kinshuk, and Yu, S. (2014). E-learning in school education in the coming 10 years for developing 21st century skills: Critical research issues and policy implications. *Journal of Educational Technology & Society*, 17(1), 70-78.

- Kuljis, J.; Liu, F. (2005) A Comparison of Learning Style Theories on the Suitability for elearning. *Web Technologies, Applications, and Services*, p. 191-197, 2005.
- Levis, D., Barbosa, J. L., Pinto, S. C. C. S., and Barbosa, D. (2008). Aperfeiçoamento automático do perfil do aprendiz em ambientes de educação ubíqua. *Revista Brasileira de Informática na Educação*, 16(1), 29-41.
- Limongelli, C., Sciarrone, F., Temperini, M., and Vaste, G. (2009). Adaptive learning with the LS-plan system: a field evaluation. *Learning Technologies, IEEE Transactions on*, 2(3), 203-215.
- Mendonça, I. E. S., Araujo, R. D., Mendes, M. M., Brant-Ribeiro, T., Dorça, F. A., & Cattelan, R. G. (2014). Explorando Funcionalidades Sociais e Colaborativas em Ambientes Educacionais Ubíquos.
- Ogata, H., Yin, C., Yano, Y. (2005). JAPELAS: Support-ing Japanese Polite Expressions Learning Using PDA towards Ubiquitous Learning. *The Journal of Information and Systems in Education*. 2(1): 33-39,
- PAPI (2006). “Draft standard for learning technology. public and private information (papi) for learners (papi learner)”, August.
- Paredes, R. G. J., Ogata, H., Yano, Y. San Martin, G. A. (2005). A Multi-Model Approach for Supporting the Personalization of Ubiquitous Learning Applications. *International Workshop on Wireless and Mobile Technologies in Education (WMTE’05)*. IEEE.
- Peng, H.i; Chou, C.; Chang, C. (2008) From virtual environments to physical environments: Exploring interactivity in ubiquitous-learning systems. *Journal of Educational Technology & Society*, v. 11, n. 2, p. 54-66.
- Rabello, S., Barbosa, J. L., Oliveira, J., Wagner, A., Barbosa, D. N., and Bassani, P. B. (2012). Um Modelo para Colaboração em Ambientes Descentralizados de Educação Ubíqua. In *Anais do Simpósio Brasileiro de Informática na Educação (Vol. 23, No. 1)*.
- Tsai, Chin-Chung. (2010) Conceptions of learning versus conceptions of web-based learning: The differences revealed by college students. *Computers & Education*, v. 53, n. 4, p. 1092-1103.
- Wu, H., Kort, E. e Bra, P. de. (2001) Design issues for general-purpose adaptive hypermedia systems. in *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*, pp. 141–150, ACM.
- Zakaria, M. R. and Brailsford, T. J. (2002). User Modelling and Adaptive Educational Hypermedia Frameworks for Education. *New Review of Hypermedia and Multimedia*, vol. 1, no. 8, pp. 83–97,.
- Zatarain-Cabada, R., Barrón-Estrada, M., Zepeda-Sánchez, L., Sandoval, G., OsorioVelazquez, J., e Urias-Barrientos, J. (2009). A Kohonen Network for Modeling Students’ Learning Styles in Web 2.0 Collaborative Learning Systems. *MICAI 2009: Advances in Artificial Intelligence*, pages 512–520.

As Tecnologias de Análise de Aprendizagem e os Desafios de Prever Desempenhos de Estudantes de Programação

Márcia G. de Oliveira ¹

¹ Centro de Referência em Formação e EaD (Cefor)
Instituto Federal do Espírito Santo (IFES)
Vitória – ES – Brasil

clickmarcia@gmail.com

Abstract. *Predicting the future of the students' learning, anticipating the possibilities of failure to plan actions of reorientation of a learning process, is a challenge for computer science. Learning Analysis appears as a possibility to contemplate this challenge because it is an educational technique that aims to recognize profiles and learning trends from the collection and analysis of data in online environments. Knowing that the key processes of Learning Analysis are Selecting, Capturing, Adding and Reporting, Predicting, Using, Refining and Sharing, this work discusses the Predicting process, highlighting methodologies, technologies, practices, challenges and analysis of research paths of predictive learning analysis. Contemplating this discussion within the field of programming learning, this paper aims to present the art's state of predictive learning analysis, to propose a framework to predict programming performances and to point out ways to advance such research.*

Keywords: *Predictive Learning Analysis, Predicting of Performance, Programming Learning.*

Resumo. *Prever o futuro de aprendizagem de alunos, antecipando-se às possibilidades de fracasso escolar para planejar ações de reorientação do processo de aprendizagem, representa um desafio para as tecnologias do futuro. A Análise de Aprendizagem aparece como uma possibilidade de contemplar esse desafio, pois é uma técnica educacional que visa reconhecer perfis e tendências de aprendizagem a partir da coleta e da análise de dados de estudantes em ambientes online. Sabendo que os principais processos da Análise de Aprendizagem são Selecionar, Capturar, Agregar e Relatar, Prever, Usar, Refinar e Compartilhar, este trabalho traz à discussão o processo de Prever, destacando metodologias, tecnologias, práticas, desafios e caminhos de pesquisa da análise de aprendizagem preditiva. Contemplando essa discussão dentro do domínio da aprendizagem de programação, este trabalho tem como objetivos apresentar o estado da arte da análise de aprendizagem preditiva, propor um framework para previsão de desempenhos em programação e apontar caminhos para avançar nessas pesquisas.*

Palavras-chave: *Análise de Aprendizagem Preditiva, Previsão de Desempenhos, Aprendizagem de Programação.*

1. Introdução

A Análise de Aprendizagem é uma técnica educacional que visa reconhecer perfis e tendências de aprendizagem a partir da coleta e análise de dados estáticos e dinâmicos em ambientes *online* como, por exemplo, os ambientes virtuais de aprendizagem.

Os dados coletados e analisados a partir de interações em ambientes virtuais de aprendizagem possibilitam uma melhor compreensão e otimização de um processo de aprendizagem [Agudo-Peregrina et al. 2012]. Dessa forma, relacionando os dados em informações, a Análise de Aprendizagem busca, a partir destas, combinar um histórico de estados de aprendizagem de um estudante mapeados em desempenhos de atividades com seu atual estado de aprendizagem na tentativa de prever o que será útil para esse estudante no futuro [Lias and Elias 2011].

Atualmente a Análise de Aprendizagem tem despertado interesse em como os dados de interações de estudantes podem melhorar o ensino e a aprendizagem [Lias and Elias 2011]. No entanto, embora tenha havido um crescente interesse pela Análise de Aprendizagem e suas tecnologias adaptativas para melhorar o ensino e a aprendizagem, há poucos estudos que de fato detalhem resultados concretos [Johnson et al. 2016].

Os principais processos da Análise de Aprendizagem a partir de dados de interações em ambientes de aprendizagem são *Selecionar, Capturar, Agregar e Relatar, Prever, Usar, Refinar e Compartilhar* [Lias and Elias 2011]. Já as principais tecnologias utilizadas para o projeto de sistemas de Análise de Aprendizagem incluem mecanismos de Inteligência Artificial, *Data Mining*, Aprendizagem Adaptativa, *Inquiry*, Aprendizagem Baseada em Problemas, Modelagem Preditiva, *Clustering*, Mineração de Padrões, Mineração de Relacionamentos, Análise Sequencial, Predição de Desempenhos, Rastreamento de Conhecimento Bayesiano e Sequenciamento de Tarefas [Suchithra et al. 2015].

Neste trabalho, considerando o domínio da aprendizagem de programação, a ênfase será dada ao processo *Prever* e às tecnologias de predição de desempenhos como as Redes Neurais, as Árvores de Decisão, o *Clustering*, os Modelos de Regressão Linear e outras técnicas de *Data Mining* [Suchithra et al. 2015].

Embora haja algumas propostas de sistemas de previsão de desempenhos para o domínio da aprendizagem de programação [Watson et al. 2013], ainda há muitos desafios a serem vencidos, entre eles:

- Mapear perfis de aprendizagem de acordo com os desempenhos de estudantes em atividades de programação ao longo do tempo.
- Definir quais dados de entrada são altamente efetivos para sistemas de predição de desempenhos [Mat et al. 2013].
- Ter uma grande quantidade de informações de um histórico de desempenhos suficiente para prever com precisão desempenhos futuros.
- Tratar a variabilidade e a subjetividade da prática da programação que podem impactar em muitas variações de variáveis preditoras ao longo do tempo.
- Mapear perfis não apenas por conteúdos, mas por habilidades extraídas a partir de informações de códigos de programação desenvolvidos por alunos.

Com o propósito de apresentar o estado da arte da análise de aprendizagem preditiva, este trabalho destaca como as tecnologias dessa área evoluíram desde a década de

60 até a década atual. Além disso, são apontados os principais desafios e caminhos de pesquisa da Análise de Aprendizagem de Programação para os próximos anos.

Considerando esses desafios, este trabalho propõe um *framework* de previsão de desempenhos de estudantes de programação. Esse *framework* tem o objetivo de apresentar uma solução computacional que possibilite aos professores de programação ter uma visão dos estados de aprendizagem dos estudantes na prática da programação e realizar previsões de desempenhos futuros desses estudantes. Dessa forma, antecipando-se ao fracasso escolar, professores podem reorientar o ensino e realizar ações formativas para que os alunos melhorem seus desempenhos em programação.

Para apresentar as aplicações, os desafios, os caminhos de pesquisa e um *framework* da previsão de desempenhos de estudantes de programação, este trabalho está organizado conforme a ordem a seguir. Na Seção 2, é apresentada a análise de aprendizagem preditiva, a sua evolução e suas aplicações. Na Seção 3, é proposto um *framework* para a previsão de desempenhos de estudantes de programação. Na Seção 4, apontam-se os desafios da previsão de desempenhos no domínio da aprendizagem de programação e possíveis caminhos de pesquisa. Na Seção 5, conclui-se este trabalho com as considerações finais e propostas de trabalhos futuros.

2. A Análise de Aprendizagem Preditiva

A análise preditiva tem como objetivo descobrir padrões a partir de um histórico de dados para descobrir tendências e estimar valores que trazem à compreensão essas tendências. No caso da análise de aprendizagem preditiva, o objetivo é estimar valores que descrevem, a partir da análise de perfis de estudantes, um futuro de sucesso ou de fracasso escolar.

Uma estratégia específica de previsão de desempenhos de estudantes no domínio da programação é apresentada por [Watson et al. 2013]. Nesse trabalho, o método de predição de desempenhos de estudantes é baseado em dados de *logs* descrevendo vários aspectos de comportamentos de programação. Uma inovação desse método consiste em prever desempenhos baseando-se em como estudantes respondem a diferentes tipos de erros em comparação com seus pares [Watson et al. 2013].

A Tabela 1 apresenta a evolução de soluções propostas para análise de dados coletados e previsão de desempenhos de estudantes, especialmente no domínio da aprendizagem de programação, desde a década de 60 até a década de 90.

De acordo com a Tabela 1, as tendências de previsão de desempenhos em programação, na década de 60, estavam relacionadas a testes de aptidão que utilizavam técnicas de análise de regressão linear múltipla com os objetivos de identificar fatores relacionados a desempenhos, de selecionar profissionais e de reduzir custos de treinamento. Na década de 70, com o objetivo de avaliar competências específicas, houve uma maior preocupação de se analisar tanto as características dos programadores quanto dos programas utilizando, para isso, representações como as métricas de *software* [Curtis et al. 1979].

A partir da década de 80, a previsão de desempenhos de estudantes começa a ter uma perspectiva de reorientar o processo de aprendizagem de estudantes e, na década de 90, através do projeto de STIs (Sistemas Tutores Inteligentes) com funções de modelagem de estados de conhecimento, previsão de desempenhos e recomendação de sequências de atividades, visava-se monitorar e regular o processo de aprendizagem de estudantes para

Trabalhos de Previsão de Desempenhos de Estudantes - Características e Contribuições	
Década de 60	
[MCNAMARA and HUGHES 1961]	Apresenta uma revisão de pesquisas em seleção de programadores, de instrumentos de avaliação de performance e de pesquisa de variáveis preditoras mais relacionadas aos desempenhos de programadores.
[Badgley et al. 1962]	Relaciona atributos psicológicos, sociais e acadêmicos para prever sucesso ou fracasso de estudantes de medicina; Determina fatores úteis para modelos de predição; Analisa histórico de dados dos estudantes; Aplica técnica de análise de regressão linear; Identifica fatores de sucesso ou fracasso escolar; Desempenhos são preditos com 81% de acurácia.
[Bauer et al. 1968]	Aplica de testes para identificar aptidão em programação em processo seletivo; A finalidade é evitar desperdício de tempo e de custos em treinamentos; Aplica modelos de regressão linear múltipla com técnica <i>stepwise</i> para selecionar melhor combinação de variáveis para prever desempenhos; Compara a eficácia preditiva de vários testes; Os escores dos testes foram correlacionados com a nota final.
Década de 70	
[DeNELSKY and McKEE 1974]	Previsão de desempenhos com o objetivo de identificar competências específicas em programação; Criação de testes designados para medir aptidões específicas.
[Motley and Brooks 1977]	Previsão de taxas de erros de programação; Aplicação de análise de regressão linear múltipla a partir de características de programas e de variáveis de programadores.
[Curtis et al. 1979]	Investigação de características de software relacionadas à complexidade psicológica; Foco em validar o uso de métricas de complexidade de software para prever desempenhos de programadores; Reconhece como melhores preditores as métricas de <i>Halstead</i> , de <i>McCabe</i> e o número de linhas de código.
Década de 80	
[Hostetler 1983]	Previsão de aptidão em programação medindo certas habilidades cognitivas, traços de personalidade e histórico de desempenhos; O objetivo primário do estudo foi criar um modelo para prever sucesso em programação para reorientar estudantes de programação.
[Butcher and Muth 1985]	Previsão de desempenhos de calouros em cursos de Ciência da Computação a partir de informações sobre programas e históricos de desempenhos escolares junto com resultados de testes do <i>American College (Act)</i> .
[Werth 1986]	Estudos sobre a relação entre a nota do estudante no começo de um curso com sexo, idade, escola, performances acadêmicas anteriores, número de cursos de matemática e experiência de trabalho; Não há relação entre desempenhos com o tipo de personalidade; Relações mais significantes com nível de estudo, número de horas trabalhadas e número de aulas de matemática no ensino secundário.
Década de 90	
[Corbett and Anderson 1994]	Modelagem dos estados de conhecimento dos estudantes; O modelo de estudante ideal contém um conjunto de regras de produção ideal que forma o núcleo do Sistema Tutor Inteligente (STI); Esse modelo auxilia o tutor a monitorar o estado de conhecimento do aluno e a planejar sequências de atividades a serem recomendadas conforme necessidades do aprendiz; O modelo resultante prediz desempenhos com eficácia, prevê erros e, principalmente, possibilita a maioria dos estudantes alcançar altos níveis de desempenhos.

Tabela 1. A Evolução das Técnicas de Previsão de Desempenhos

que estes alcançassem êxitos de aprendizagem. Um exemplo de sistema tutor inteligente com essa perspectiva foi aplicado em um curso de programação introdutória com excelentes resultados de aprendizagem [Corbett and Anderson 1994].

A Tabela 2 apresenta as propostas mais recentes de previsão de desempenhos de estudantes, especialmente para o domínio da programação.

Trabalhos mais recentes (a partir de 2000)

[Minaei-Bidgoli et al. 2003]	Tecnologias educacionais da web para coleta de dados de estudantes; Coleta rotineira de dados em grandes quantidades por sistemas web; Aplicação de técnicas de <i>Data mining</i> ; Abordagem para classificar alunos afim de prever sua nota final a partir de características reconhecidas nos dados coletados; Combinação de vários classificadores para melhorar a eficácia de classificação; Considera peso de características e aplica um algoritmo genético para melhorar precisão de previsão; Visa identificar alunos em risco e permitir o instrutor orientá-los em tempo hábil.
[Kotsiantis et al. 2010]	Combinação de classificadores utilizando métodos de votação e algoritmos de aprendizagem <i>online</i> para previsão de desempenhos de estudantes em curso de informática a distância.
[Romero and Ventura 2010]	Apresenta as principais propostas de previsão de desempenhos utilizando técnicas de <i>data mining</i> .
[Romero et al. 2013]	Especifica ferramenta de mineração de dados em <i>Moodle</i> ; Compara experimentalmente várias técnicas de mineração de dados para previsão de desempenhos; Apresenta exemplos de modelos descobertos e explica que modelos de classificadores são apropriados para ambientes educacionais, mais eficazes e mais compreensíveis para os professores.
[Sorour et al. 2015]	Previsão de desempenhos de estudantes pelo métodos LDA (<i>Latent Dirichlet Allocation</i>) e SVM (<i>Support Vector Machine</i>); Melhores resultados de previsão de desempenhos em relação aos modelos LSA (<i>Latent Semantic Analysis</i>) e PLSA (<i>Probabilistic Latent Semantic Analysis</i>).
[Naser et al. 2015]	Um modelo de Redes Neurais Artificiais é utilizado para prever desempenhos de estudantes dos cursos de Engenharia e Tecnologia da Informação; Tem como objetivos identificar fatores que afetam desempenhos, converter esses fatores em formas apropriadas para uma codificação de sistemas adaptativos e prever desempenhos de um estudante a partir de dados desse estudante. São considerados fatores como desempenhos anteriores, escores em disciplinas de Matemática e gênero; O modelo foi desenvolvido e treinado abrangendo cinco gerações de graduados; O Modelo prevê corretamente o desempenho de mais de 80% dos estudantes.
[Qiu et al. 2016]	Modelagem e previsão de comportamentos de estudantes em MOOCS (<i>Massive Open Online Courses</i>) através da análise profunda de dados demográficos, padrões de atividades de aprendizagem em fóruns, vídeos e exercícios; Proposta de <i>LadFG</i> (<i>Dynamic Factor Graph</i>) para incorporar esses dados de estudantes; Duas tarefas de previsão: previsão de desempenhos em exercícios e previsão de receber certificação; Modelo flexível que pode ser aplicado a várias configurações.
[Strang 2016]	A contribuição metodológica é a combinação de métodos onde a hipótese quantitativa foi testada e seguida pela coleta de dados qualitativos na análise de textos e mais análise quantitativa para descobrir padrões em textos <i>online</i> ; Foco na análise de aprendizagem <i>online</i> ; Vários atributos importantes dos estudantes e suas atividades são examinadas para identificar o que melhor favorece a previsão de escores mais altos; Tem o propósito de explorar relações entre notas de estudantes e fatores de aprendizagem utilizando um amplo conjunto de dados de um curso; Estudos atuais apontam para a habilidade de prever resultados de aprendizagem a partir da análise de <i>Big Data</i> , mas poucos indicadores para isso têm sido de fato encontrados; Uso de preditores quantitativos e dados qualitativos extraídos de textos de estudantes; Quatro preditores de interação <i>online</i> são utilizados para explicar modelo de regressão.

Tabela 2. Estado da Arte da Previsão de Desempenhos

Atualmente, a análise de aprendizagem preditiva tem como tendências a coleta de dados em sistemas *web*, a análise de dados em *Big Data* utilizando técnicas de *data mining*, a combinação de técnicas de classificação para melhorar eficácia de previsão e a utilização de modelos de regressão linear múltipla para a estimação de desempenhos.

De acordo com a revisão do estado da arte da previsão de desempenhos no domínio da aprendizagem de programação, o *framework* proposto neste trabalho tem como objetivos específicos: analisar um histórico de códigos-fontes de programas desenvolvidos por estudantes em ambientes *online*, combinar características de códigos-fontes para modelar estados de aprendizagem ao longo de um curso gerando perfis, prever desempenhos de estudantes e realimentar estados de aprendizagem por recomendação de atividades de acordo com necessidades dos estudantes reconhecidas em seus perfis. O objetivo geral dessa proposta é, por conseguinte, alcançar êxitos coletivos de aprendizagem.

3. Um *Framework* para Previsão de Desempenhos em Programação

A proposta metodológica do *framework* de previsão de desempenhos deste trabalho aplica as ideias do STI de [Corbett and Anderson 1994] e do sistema de recomendação de atividades proposto por [Oliveira et al. 2013].

A Figura 1 apresenta o *framework* combinando as funcionalidades de mapeamento de perfis de estudantes, de previsão de desempenhos e de regulação da aprendizagem para o domínio da aprendizagem de programação.

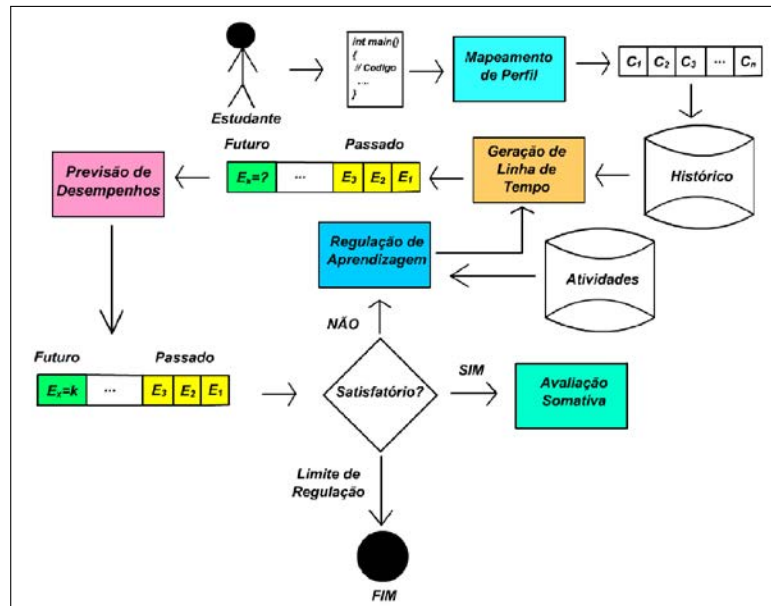


Figura 1. Framework de Previsão de Desempenhos

De acordo com a Figura 1, cada programa desenvolvido por um *Aluno* é mapeado em um perfil representado por um conjunto de componentes de habilidades que são variáveis de avaliação do domínio da programação [Oliveira et al. 2015b]. Em seguida, esse perfil é adicionado a um *histórico* de perfis do mesmo aluno ao longo de um curso. Para a *Geração de Linha do Tempo do Estudante*, cada perfil do histórico é mapeado em um estado de aprendizagem E_i , que pode ser o escore de cada exercício.

O estado de aprendizagem a ser predito é E_x . Para a *Previsão de Desempenhos*, é criado um modelo de regressão linear a partir dos perfis do *Histórico* e dos estados de aprendizagem E_i a eles associados. Em seguida, é predito um valor k para a variável independente E_x do modelo.

A partir do estado de aprendizagem predito, o processo de aprendizagem do aluno na prática da programação é reorientado. Se o estado for *satisfatório*, o aprendiz já poderá ser submetido a processos de *Avaliação Somativa* para qualificação de aprendizagem. Se não, ele é submetido a um processo de *Regulação da Aprendizagem* que consiste em ações de recomendação de atividades conforme dificuldades de aprendizagem. Uma nova linha do tempo é então criada a partir das novas atividades resolvidas e o estado de aprendizagem E_x é predito novamente.

O processo de *Regulação da Aprendizagem* se repete até que o estudante alcance um estado de aprendizagem *satisfatório*. No entanto, há um *limite de regulação* que in-

forma se a regulação possui possibilidades de êxito considerando os avanços em uma linha do tempo. Não havendo mais essas possibilidades depois de várias ações de regulação, o processo é encerrado e um relatório é gerado informando os estados de aprendizagem do aluno e as suas dificuldades para serem avaliadas individualmente pelo professor.

4. Desafios e Caminhos de Pesquisa para a Previsão de Desempenhos

Embora algumas soluções para previsão de desempenhos de estudantes já tenham sido desenvolvidas e possam ser estendidas para o domínio da programação, para um domínio tão complexo como este, há muitos desafios a serem vencidos entre os quais destacamos:

1. Assumindo uma representação vetorial do domínio a partir de códigos-fontes desenvolvidos por alunos, conforme [Oliveira et al. 2015b], como definir as variáveis preditoras, isto é, as dimensões ou *features*, que melhor representem um histórico de diferentes classes de exercícios de programação ao longo de um curso e sejam adequadas para formar um modelo de previsão de desempenhos?
2. Como tratar as variações de estilos nos códigos de programação de estudantes ao longo de uma linha do tempo?
3. Como estabelecer medições que garantam a confiabilidade dos sistemas de previsão de desempenhos de forma que motivem professores a utilizá-los? Será possível, a partir de um histórico de exercícios avaliados e mapeados em uma perspectiva multidimensional, prever o desempenho de um aluno na prova?
4. Como construir esse histórico de exercícios, uma vez que será necessária uma grande quantidade de amostras para cada aluno, sem demandar esforço em excesso de professores para corrigir todos esses exercícios?
5. No contexto de *MOOCs* (*Massive Open Online Courses - Cursos Online Abertos e Massivos*), como prever desempenhos de uma grande quantidade de alunos com a demanda de uma grande quantidade de exercícios pré-avaliados por professores?

Para vencer cada um desses desafios da previsão de desempenhos de estudantes de programação, apontamos os seguintes caminhos de pesquisa, respectivamente:

1. Na recomendação semi-automática de exercícios de programação para estudantes com dificuldades de aprendizagem, o trabalho de [Oliveira et al. 2013] representa o domínio da aprendizagem de programação a partir de informações de programas escritos em Linguagem C. Essas informações são mapeadas em vetores cujas dimensões são variáveis de avaliação de programação. Essa ideia pode ser, portanto, utilizada para representar perfis de alunos para prever seus desempenhos.
2. Uma proposta para tratar as variações em estilos de programação ao longo de um curso, é criar uma representação genérica dos exercícios de programação resolvidos pelos alunos através de métricas alternativas que quantifiquem padrões ou tendências de programação como esforço, complexidade, estilo e outras métricas [Hung et al. 1993].
3. Uma solução é identificar exercícios de programação resolvidos por alunos ao longo de um curso cujas soluções sejam estruturalmente semelhantes às soluções das questões de prova. Uma outra proposta seria comparar históricos de estudantes reformulando o problema de previsão em um problema de recomendação de filtragem colaborativa em que um aluno recebe a nota que outros alunos com históricos semelhantes ao seu receberiam. Ainda assim, para o domínio da programação, há altas variações no desenvolvimento de programas de aluno para aluno. Dessa forma, a confiabilidade de sistemas de previsão de desempenhos em programação ainda representa um grande desafio de pesquisa.

4. Nesse caso, sistemas de avaliação semi-automática de exercícios de programação poderiam ser utilizados em combinação com sistemas de previsão de desempenhos. Dessa forma, durante um curso, para cada exercício, apenas algumas amostras selecionadas seriam corrigidas por professores e as demais, por um sistema de avaliação semi-automática como o proposto por [Oliveira 2013] e melhorado por [Oliveira et al. 2015a]. Outras opções de avaliação automática de exercícios de programação são apresentadas por [Oliveira and Oliveira 2015].
5. No contexto de *MOOCs* [Pieterse 2013], com o apoio de sistemas de avaliação automática de exercícios de programação apontados no item anterior, a previsão de desempenhos pode tornar-se mais confiável uma vez que, havendo mais alunos, haverá mais históricos de exercícios e maior representação da diversidade de soluções desses exercícios. Nesse caso, o verdadeiro desafio a ser vencido é a redução de esforço de professores ao avaliar exemplos de soluções para cada um dos vários exercícios de programação aplicados com a finalidade de gerar o treino de sistemas de avaliação semi-automática. O ideal é, portanto, que, para cada exercício, o professor avalie o menor número possível de exemplos de soluções e que esses exemplos sejam os mais representativos de um conjunto com uma grande quantidade de amostras de soluções. O trabalho de [Oliveira et al. 2015a], por exemplo, é um passo inicial e relevante nesse caminho de pesquisa porque propõe uma estratégia de seleção de amostras e *features* de um conjunto de soluções de programação para formar o conjunto de treino de uma avaliador semi-automático.

5. Considerações Finais

Este trabalho apresentou um sumário do estado da arte da análise de aprendizagem preditiva, destacando a sua evolução, os métodos propostos, os desafios a serem vencidos e os caminhos de pesquisa a serem seguidos dentro do domínio da aprendizagem de programação. Além disso, propomos um *framework* para previsão de futuros desempenhos de estudantes a partir de um histórico de desempenhos em atividades de programação.

Para os próximos anos, os principais desafios da previsão de desempenhos em avaliações de programação são os seguintes: dispor de uma grande quantidade de informações a partir de um histórico de desempenhos suficiente para prever com precisão os desempenhos futuros e tratar a variabilidade e a subjetividade tão presentes no processo de construção de programas, o que pode impactar em ampla variabilidade nas variáveis preditoras de modelos de análise de aprendizagem preditiva.

A contribuição deste trabalho é promover a discussão da análise de aprendizagem preditiva na comunidade científica de interesse e incentivar, a partir deste estudo, o desenvolvimento de estratégias de previsão de desempenhos em programação com as finalidades de, antecipando-se a um futuro de fracasso escolar, reorientar práticas de ensino e remediar o processo de aprendizagem para que alunos melhorem seus desempenhos.

Referências

- Agudo-Peregrina, A., Hernández-García, A., and Iglesias-Pradas, S. (2012). Predicting academic performance with learning analytics in virtual learning environments: A comparative study of three interaction classifications. In *Computers in Education (SIIE), 2012 International Symposium on*, pages 1–6.
- Badgley, R. F., Hetherington, R. W., and Macleod, J. W. (1962). Social characteristics and prediction of academic performance of saskatchewan medical students. *Canadian Medical Association Journal*, 86(14):624.
- Bauer, R. et al. (1968). Predicting performance in a computer programming course.

- Butcher, D. F. and Muth, W. A. (1985). Predicting performance in an introductory computer science course. *Commun. ACM*, 28(3):263–268.
- Corbett, A. T. and Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278.
- Curtis, B., Sheppard, S. B., Milliman, P., Borst, M. A., and Love, T. (1979). Measuring the psychological complexity of software maintenance tasks with the halstead and mccabe metrics. *IEEE Trans. Softw. Eng.*, 5(2):96–104.
- DeNELSKY, G. Y. and McKEE, M. G. (1974). Prediction of computer programmer training and job performance using the aabp test1. *Personnel Psychology*, 27(1):129–137.
- Hostetler, T. R. (1983). Predicting student success in an introductory programming course. *SIGCSE Bull.*, 15(3):40–43.
- Hung, S.-l., Kwok, L.-f., and Chung, A. (1993). New metrics for automated programming assessment. In *Proceedings of the IFIP WG3.4/SEARCC (SRIG on Education and Training) Working Conference on Software Engineering Education*, pages 233–243, Amsterdam, The Netherlands, The Netherlands. North-Holland Publishing Co.
- Johnson, L., Becker, S. A., Cummins, M., Estrada, V., Freeman, A., and Hall, C. (2016). *NMC Horizon Report: 2016. Higher Education Edition. Learning Analytics and Adaptive Learning: Time-to-Adoption Horizon: One Year or Less*. The New Media Consortium, Austin, Texas.
- Kotsiantis, S., Patriarcheas, K., and Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students’ performance in distance education. *Knowledge-Based Systems*, 23(6):529–535.
- Lias, T. E. and Elias, T. (2011). Learning analytics: The definitions, the processes, and the potential.
- Mat, U. B., Buniyamin, N., Arsad, P. M., and Kassim, R. (2013). An overview of using academic analytics to predict and improve students’ achievement: A proposed proactive intelligent intervention. In *Engineering Education (ICEED), 2013 IEEE 5th Conference on*, pages 126–130.
- MCNAMARA, W. J. and HUGHES, J. L. (1961). A review of research on the selection of computer programmers. *Personnel Psychology*, 14(1):39–51.
- Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., and Punch, W. F. (2003). Predicting student performance: an application of data mining methods with an educational web-based system. In *Frontiers in Education, 2003. FIE 2003 33rd Annual*, volume 1, pages T2A–13.
- Motley, R. and Brooks, W. (1977). Statistical prediction of programming errors. Technical report, DTIC Document.
- Naser, S. A., Zaqout, I., Ghosh, M. A., Atallah, R., and Alajrami, E. (2015). Predicting student performance using artificial neural network: in the faculty of engineering and information technology. *International Journal of Hybrid Information Technology*, 8(2):221–228.
- Oliveira, M., Jiménez, N., Daher, P., and Oliveira, E. (2015a). Representação da diversidade de componentes latentes em exercícios de programação para classificação de

- perfis. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, page 1177.
- Oliveira, M., Nogueira, M. A., and Oliveira, E. (2015b). Sistema de Apoio à Prática Assistida de Programação por Execução em Massa e Análise de Programas. In *XXIII Workshop sobre Educação em Computação (WEI) - CSBC 2015*, Recife, PE. SBC.
- Oliveira, M. G. (2013). *Núcleos de Avaliações Diagnóstica e Formativa para Regulação da Aprendizagem de Programação*. Tese de doutorado, Universidade Federal do Espírito Santo.
- Oliveira, M. G., Ciarelli, P. M., and Oliveira, E. (2013). Recommendation of programming activities by multi-label classification for a formative assessment of students. *Expert Systems with Applications*, 40(16):6641–6651.
- Oliveira, M. G. and Oliveira, E. (2015). Abordagens, práticas e desafios da avaliação automática de exercícios de programação. In *CSBC 2015 - DesafIE 2015*, Recife.
- Pieterse, V. (2013). Automated assessment of programming assignments. In *Proceedings of the 3rd Computer Science Education Research Conference on Computer Science Education Research*, CSERC '13, pages 4:45–4:56, Open Univ., Heerlen, The Netherlands, The Netherlands. Open Universiteit, Heerlen.
- Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., and Xue, Y. (2016). Modeling and predicting learning behavior in moocs. *Science*, 45(50):55.
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., and Ventura, S. (2013). Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education*, 21(1):135–146.
- Romero, C. and Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6):601–618.
- Sorour, S. E., Goda, K., and Mine, T. (2015). Estimation of student performance by considering consecutive lessons. In *Advanced Applied Informatics (IIAI-AAI), 2015 IIAI 4th International Congress on*, pages 121–126. IEEE.
- Strang, K. D. (2016). Beyond engagement analytics: which online mixed-data factors predict student learning outcomes? *Education and Information Technologies*, pages 1–21.
- Suchithra, R., Vaidhehi, V., and Iyer, N. E. (2015). Survey of learning analytics based on purpose and techniques for improving student performance. *International Journal of Computer Applications*, 111(1).
- Watson, C., Li, F. W. B., and Godwin, J. L. (2013). Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In *2013 IEEE 13th International Conference on Advanced Learning Technologies*, pages 319–323.
- Werth, L. H. (1986). Predicting student performance in a beginning computer science class. In *Proceedings of the Seventeenth SIGCSE Technical Symposium on Computer Science Education*, SIGCSE86, pages 138–143, New York, NY, USA. ACM.

Inteligência Artificial: um desafio na detecção precoce de Indivíduos em Risco de Dislexia

Luís Fernando Orleans¹, Raimundo José Macário Costa², Sérgio Manoel Serra da Cruz²

¹Departamento de Ciência da Computação, Instituto Multidisciplinar – Universidade Federal Rural do Rio de Janeiro
Nova Iguaçu – RJ – Brasil

²Departamento de Matemática, Instituto de Ciências Exatas – Universidade Federal Rural do Rio de Janeiro
Seropédica – RJ – Brasil

{lforleans, macario, serra}@ufrrj.br

Abstract. *The acquired knowledge on dyslexia can be applied to build intelligent computing systems to be used by both education and health areas. Furthermore, those systems can help dyslexia diagnosis. The purpose of this paper is to point research directions and make considerations on the challenges related to Computational Neuroscience support on the Brazilian educational system. This work also presents a dyslexia based approach which is one of the Learning Disorders that has been attracting attention from both researchers from health and education areas. Finally, this research aims to investigate novel computational techniques to fulfill the lack of data gathering and exploratory analysis – both related to early detection of dyslexic individuals.*

Resumo. *O conhecimento adquirido sobre dislexia pode ser direcionado para a construção de um sistema computacional inteligente aplicado à área de educação e da área de saúde e aliada às novas formas de diagnóstico médico. O objetivo deste artigo é expor direções de pesquisa e considerações sobre os desafios referentes ao apoio computacional oferecido pela Neurociência Computacional na Educação Brasileira. O trabalho também apresenta uma abordagem relacionada com a dislexia, um dos Transtornos de Aprendizagem que tem despertado interesse nos pesquisadores, profissionais de saúde e educação. Esta pesquisa também visa investigar novas técnicas computacionais que venham a preencher a lacuna de ausência de instrumentos de coleta e análise exploratória de dados relacionados à detecção precoce de indivíduos disléxicos.*

1. Introdução

Há uma diversidade de recursos tecnológicos que podem ajudar pessoas com dificuldades de aprendizagem, sendo que conhecê-los apropriadamente é um desafio para os profissionais desta área de Tecnologia da Informação (TI). Dentre estas dificuldades, podemos exemplificar os seguintes transtornos de aprendizagem: déficit de atenção e hiperatividade (TDAH), dislexia, discalculia, dismusica, dislalia, entre outros. Se for particularizado para indivíduos disléxicos, as tecnologias que se utilizam para a detecção de indivíduos em risco de dislexia ainda são muito incipientes (SOUSA e SILVEIRA, 2015).

A detecção precoce das crianças em risco de dislexia é fundamental para a diminuição dos riscos de perda cognitiva e, futuramente, a lacuna escolar certamente será menor. Por outro lado, os baixos níveis sociais e as fragilidades dos vínculos familiares são possibilidades de fatores de risco, acarretando em perdas na linguagem, na memória e nas habilidades sociais, todas de extrema importância para o desenvolvimento cognitivo do indivíduo. Desta forma, com a identificação precoce de crianças em risco de dislexia, os fatores psicológicos e socioafetivos se beneficiam, uma vez que a detecção do problema irá minimizar os sentimentos de fracasso e o desânimo das crianças que enfrentam dificuldades no processo de aprendizagem (CARIDÁ e MENDES, 2012).

Neste sentido, a dislexia pode ser entendida como uma dificuldade enfrentada pelas crianças na idade escolar caracterizando-se como uma deficiência especial na aprendizagem. Esta é compreendida como uma alteração em um ou mais dos processos psicológicos básicos para a compreensão da linguagem escrita ou oral, que pode se manifestar como uma falta de habilidade para leitura, matemática, ou expressão escrita, além de problemas com a memória. Nessas condições podem estar incluídos uma falha de percepção, dano cerebral, disfunção cerebral mínima, dislexia e afasia do desenvolvimento. Não se aplica, assim, a crianças que apresentem deficiências visuais, auditivas, motoras, atraso mental, transtornos emocionais, situações de desvantagem ambiental, cultural ou econômica (SANTOS e MARINHO, 2007).

Dentre os problemas existentes acerca da aprendizagem, a dislexia destaca-se de outros transtornos, com os seus portadores podendo apresentar atrasos graves na leitura, na escrita e ortografia, assim como inversões de símbolos, apresentando natureza única e limitada do déficit fonológico (SHAYWITZ, 1999; MOUSINHO, 2003). Há registros da limitação médica pediátrica para constatar e avaliar os distúrbios de aprendizagem com vistas a um diagnóstico precoce adequado para as crianças com habilidades sociais e desenvolvimento motor apropriados até a terceira série do ensino fundamental.

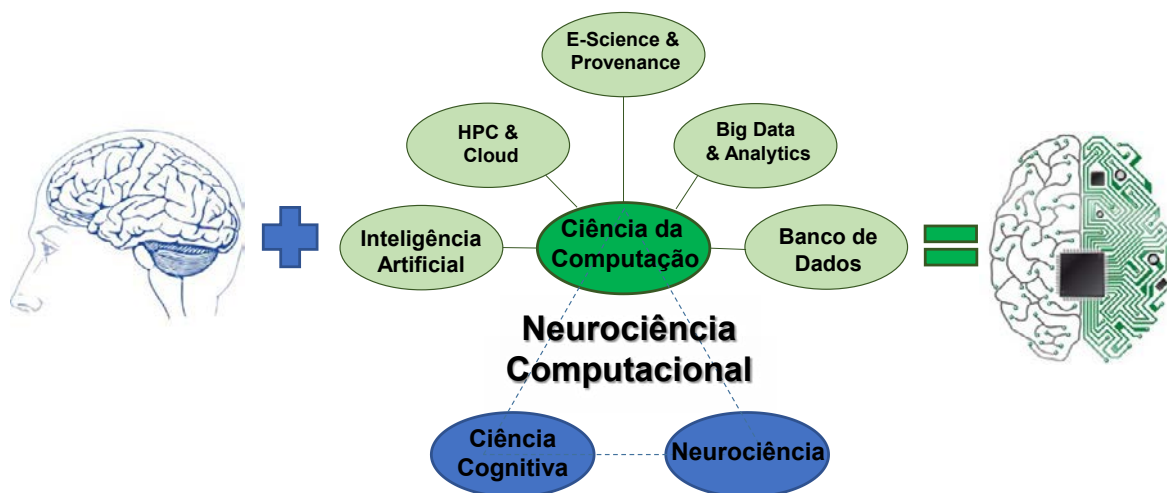


Figura 1. Três pilares na Neurociência Computacional e perspectivas de investigações tecnológicas apoiadas pela Ciência da Computação. Adaptado de Macário Costa et al (2014)

Um dos caminhos para encarar os desafios e propor soluções para o problema de detecção de indivíduos em risco de dislexia é a Neurociência Computacional (NC) (MACÁRIO COSTA et al, 2014). Esta é uma área essencialmente interdisciplinar e assim está apoiada em três pilares: Neurociência (das áreas de Medicina); Ciência Cognitiva (das áreas de Educação e Psicologia) e Ciência da Computação (áreas de Inteligência Artificial, Bancos de Dados, E-Science, Proveniência, Big Data, Nuvens, entre outros) (Figura 1). No entanto, uma das grandes dificuldades desta área é modelar (matemática e computacionalmente) um Transtorno de Aprendizagem, no caso a dislexia, identificar quais são os dados e variáveis mais relevantes, transcrevê-los para as soluções tecnológicas e avaliar se os resultados computacionais são significativos e válidos de acordo com os aspectos médico, ético e educacional.

O objetivo deste artigo é expor direções de pesquisa e considerações sobre os desafios referentes ao apoio computacional oferecido pela Neurociência Computacional na Educação Brasileira. O trabalho também apresenta uma linha de pesquisa relacionada com a dislexia, um dos Transtornos de Aprendizagem que tem despertado interesse nos pesquisadores, profissionais de saúde e educação, e a Computação.

Este artigo está dividido da seguinte forma: a seção 2 caracteriza o transtorno de dislexia, enquanto a seção 3 lista o estado da arte. A seção 4 aborda os desafios no desenvolvimento de sistemas computacionais para detecção precoce de indivíduos em risco de dislexia e, por fim, a seção 5 apresenta as oportunidades em aberto e as considerações finais.

2. Dificuldades de Aprendizagem e Dislexia

No âmbito da educação, novos estudos relacionam as dificuldades de aprendizagem enfrentadas por crianças e jovens com diversas técnicas computacionais. Tais questões têm relevância social e podem repercutir na evasão escolar, analfabetismo funcional e sucessivas reprovações. No início do processo de escolarização, a criança pode apresentar algumas dificuldades no aprendizado da leitura, escrita e cálculo. Convencionalmente, costumam-se dividir as dificuldades de aprendizagem em dois tipos: a) Dificuldades Escolares (DE) relacionadas a problemas de origem e ordem pedagógica e b) Distúrbios de Aprendizagem (DA) relacionados a uma disfunção no Sistema Nervoso Central (SNC), caracterizada por uma falha no processo de aquisição e/ou desenvolvimento das habilidades escolares (MACÁRIO COSTA et al. 2014b).

Os DA apresentam comprometimento em um ou mais componentes da linguagem. Trata-se de um transtorno caracterizado por desempenho substancialmente abaixo do esperado (nas áreas da leitura, escrita e matemática) tendo em vista a idade cronológica, medidas de inteligência e educação apropriada para a idade (DSM IV, 1995). Dentre os DA, a dislexia destaca-se de outros transtornos que incluem atrasos graves na leitura, na escrita e ortografia, assim como inversões de símbolos (Anexo I). O destaque se deve a sua natureza única e limitada do déficit fonológico (MOUSINHO, 2003; SHAYWITZ e SHAYWITZ, 1999).

Segundo a Associação Internacional de Dislexia, o transtorno da dislexia é caracterizado pelas dificuldades na correção e/ou fluência na leitura de palavras e por baixa competência leitora e ortográfica. Estas dificuldades resultam tipicamente de um déficit na componente fonológica da linguagem que é frequentemente imprevisível em relação a outras capacidades cognitivas e às condições educativas. Secundariamente, podem surgir dificuldades de compreensão leitora, experiência de leitura reduzida que podem impedir o desenvolvimento

do vocabulário e dos conhecimentos gerais (Associação Internacional de Dislexia, 2003, cit. por MACÁRIO COSTA et al, 2011).

Quanto ao processo utilizado para a realização do diagnóstico em sua forma tradicional, o responsável pelo paciente recebe um questionário a ser levado para casa e ser respondido. Ao término do preenchimento, o questionário é devolvido ao profissional fonoaudiólogo, que, por sua vez, o confere, certificando que todas as questões foram respondidas. Em muitas vezes, o respondente é questionado sobre algumas das respostas apresentadas. Em outras ocasiões, o profissional deve esclarecer perguntas não compreendidas pelo respondente. Finalmente, em um momento posterior, o questionário é tabulado para verificar o risco ou não de o indivíduo disléxico.

Porém, a confirmação final do diagnóstico de dislexia deverá ser feita por uma equipe multidisciplinar, que será responsável pela elaboração de um processo de avaliação minucioso e abrangente, também conhecido como avaliação multidisciplinar e de exclusão.

Hoje, para fazer uma avaliação de uma pessoa verifica-se uma demora média de três meses para o estabelecimento de um diagnóstico por um profissional qualificado. A partir do estabelecimento de um sistema inteligente pode ser possível que esta demora se reduza a poucas horas, oferecendo oportunidade de atendimento para mais pessoas em um curto espaço, de forma mais oportuna e eficiente.

3. Estado da Arte

O conhecimento adquirido na área de neurociência pode ser associado às ferramentas e técnicas computacionais para aperfeiçoar detecção precoce da dislexia. Para alcançarmos um patamar tecnológico compatível com as demandas multifacetadas dos estudos sobre a dislexia, deveremos considerar a estruturação de uma agenda de pesquisa com ênfase em novos modelos e incorporação de técnicas computacionais. Neste caso é possível verificar trabalhos que estão na fronteira da área:

- a) Modelagem e desenvolvimento de sistemas inteligentes e preditivos baseados em técnicas de Inteligência Artificial (MACÁRIO COSTA et al, 2007, 2008, 2009, 2010, 2011, 2013; ZAVALETA et al, 2012), os autores apresentam uma implementação de redes neurais para classificar probabilisticamente pacientes jovens e adultos com dislexia. Riveros et al (2005) propõe um sistema especialista voltado para um pré-diagnóstico de dislexia em crianças com vista ao desenvolvendo ferramentas lúdicas de avaliação.
- b) Desenvolvimento de artefatos de hardwares e softwares como tecnologias assistivas para contornar as dificuldades relacionadas à leitura. Atualmente, existem hardwares do tipo canetas que auxiliam os disléxicos na leitura e reconhecimento de textos. Estas tecnologias não substituem a intervenção direta do profissional de saúde, mas eles podem aumentar a eficácia das terapias (ZONG, 2014).

Essas tecnologias ajudam a aumentar a confiança do indivíduo nos ambientes residencial, sala de aula ou no trabalho. Atualmente já existem softwares no mercado (em língua inglesa) que visam auxiliar indivíduos comprovadamente disléxicos. Por exemplo, o Simon S.I.O, Sortegories, Wordmaker, todos disponíveis no sítio *Dyslexia Helper*¹, e

¹ <http://dyslexiahelp.umich.edu/tools/software-assistive-technology>

WebHelpDyslexia². O Simom S.I.O é um tutor pessoal que possibilita que indivíduos disléxicos adquiram as habilidades de leitura. O Sortegories é um plug-in Web que auxilia a construção de vocabulários a partir de fonemas básicos e a compreensão de leitura. O Wordmaker é voltado para estudantes que possuem limitadas habilidades de decodificação, fluência e compreensão de textos, e oferece uma abordagem sistemática e interativa para ensinar ortografia, leitura e escrita para indivíduos disléxicos. O WebHelpDyslexia é um software nacional que permite ao disléxico fazer as alterações necessárias (cor, espaçamento, tamanho de fonte) para visualizar e melhorar o entendimento de um texto sendo lido. O software também realça partes do texto, por meio de uma “régua de leitura”, o que reduz a distração.

4. Desafios Existentes

Dentro do contexto computacional, a Inteligência Artificial (IA) estuda processos cognitivos e tenta criar sistemas computacionais situando-se na fronteira entre diversas ciências tais como Psicologia, Filosofia, Linguística e Engenharia. A Inteligência Artificial pode desempenhar um papel importante na busca, elaboração e validação de modelos computacionais na construção de sistemas adaptativos e/ou evolutivos inteligentes (CARVALHO, 2005). É importante frisar a escassez de informações disponíveis sobre sistemas computacionais voltados para o rastreamento de pessoas com sinais de dislexia e outros transtornos. Dessa forma a lacuna de conhecimento a ser explorado justifica o investimento nacional em pesquisas e desenvolvimento de tecnologias.

De acordo com o exposto acima é admissível a criação de modelos computacionais inteligentes capazes de rastrear grupos de padrões por similaridades. Com as técnicas de Inteligência Artificial espera-se desenvolver um sistema computacional que seja capaz de rastrear pessoas em risco de dislexia.

Para alcançarmos um patamar tecnológico compatível com as demandas multifacetadas dos estudos do cérebro no Século XXI, deveremos considerar a estruturação de uma agenda de pesquisa com ênfase em *novos modelos* e incorporação de *técnicas computacionais (intensivas em dados) do E-science* (HEY *et al*, 2009) para o desenvolvimento de aplicações em Neurociência Computacional. Neste caso é possível avançar nas seguintes áreas:

- a) Desenvolvimento de sistemas inteligentes e preditivos baseados em técnicas de Inteligência Artificial (MACÁRIO COSTA *et al*, 2007, 2008, 2009, 2010, 2011, 2013; ZAVALETA *et al*, 2012) capazes de manipular grandes volumes de dados;
- b) Uso ambientes computacionais distribuídos de processamento de alto desempenho para apoiar simulações e experimentos *in silico* baseados em workflows científicos (DEELMAN *et al*, 2009) de simulações de modelos cerebrais (ABBOT, 2013; KUBILIUS, 2014; NIH, 2014; HBP, 2014).
- c) Adoção de técnicas de gestão de grandes volumes de dados biológicos semiestruturados e processamentos típicos de *Big Data* (DAVISON, 2010; BERMAN, 2011; ABBOTT, 2013), os projetos de neurociência computacional tendem mapear modelos cerebrais cada vez maiores, mais complexos e utilizar sensores e dados com diversos formatos (ZHONG *et al*. 2011; ZHONG, 2014);

² <http://ead.unifal-mg.edu.br/cead/?q=node/194>

- d) Incorporação de descritores de proveniência (CRUZ *et al*, 2009) e curadoria de dados e técnicas de gestão de conhecimento para ampliar a reprodutibilidade e a confiabilidade dos estudos em neurociência computacional (CHEN, ZHONG, LIANG, 2012; CICCARESE *et al*, 2013) estes tendem a ser conduzidos por times de pesquisa interdisciplinares e dispersos geográfica e temporalmente (CHEN, ZHONG, 2013);

O conhecimento adquirido na área de Neurociência Computacional pode ser associado às ferramentas e técnicas computacionais para aperfeiçoar as oportunidades de atuação sobre Transtornos de Aprendizagem.

5. Considerações Finais

Identificar precocemente crianças com dificuldades de leitura e aprendizagem torna-se prioridade diante da possibilidade de poder eliminar ou minimizar prejuízos escolares e sociais nas mesmas. Dessa maneira, as estratégias computacionais da Neurociência Computacional podem fazer toda a diferença nessa identificação.

Esta proposta de desafio visa investigar novas técnicas computacionais que venham a preencher a lacuna de ausência de instrumentos de coleta e análise exploratória de dados relacionados à detecção precoce de disléxicos em idade escolar no Brasil.

Agradecimentos

Os autores agradecem à FAPERJ e ao MEC/SeSU pelo apoio financeiro a pesquisa.

REFERÊNCIAS

- ABBOTT, A. (2013) “Neuroscience: Solving the brain”, In: Nature 499, pages. 272–274.
- BOULIC, R. and RENAULT, O. (1991) “3D Hierarchies for Animation”, In: New Trends in Animation and Visualization, Edited by Nadia Magnenat-Thalmann and Daniel Thalmann, John Wiley & Sons Ltd., England.
- CARIDÁ, Déborah Alcântara Prósperi; MENDES, Mônica Hoehne. “A importância do estímulo precoce em casos com risco para dislexia: um enfoque psicopedagógico”. Rev. Psicopedagogia 2012; 29(89): 226-35. Acesso em 15/04/2016.
- CARVALHO, L. A. V., 2005, Datamining – A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração. Rio de Janeiro: Editora Ciência Moderna Ltda.
- CICCARESE, P., SOILAND-REYES, S., BELHAJJAME, K., GRAY, A.J.G, GOBLE, C., Clark, T. (2013) “PAV ontology: provenance, authoring and versioning”, Journal of Biomedical Semantics 2013, 4:37.
- CHEN J. H., ZHONG, N. (2013) “Toward the Data-Brain driven systematic brain data analysis”. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(1), pages. 222-228.
- CHEN J. H., ZHONG, N., LIANG, P. P. (2012) “Data-Brain driven systematic human brain data analysis: A case study in numerical inductive reasoning centric investigation” Cognitive Systems Research, Elsevier, vol. 15-16, pages. 17-32.
- CRUZ, S.M.S, CAMPOS, M. L M., MATTOSO, M. (2009) “Towards a Taxonomy of Provenance in Scientific Workflow Management Systems”. SERVICES I 2009: 259-266.

- DAVISON, A.P. (2010) “Challenges and solutions in replicability and provenance tracking for simulation projects”. BMC Neuroscience 2010, 11(Suppl 1):P76.
- DEELMAN E, Gannon D, SHIELDS M, Taylor I (2009) “Workflows and e-Science: An overview of workflow system features and capabilities”, Future Generation Computer Systems 25(5):528–540.
- DUTRA, M. V. O.; SOUZA, M. N., 2001, Redes neurais artificiais e análise fatorial no diagnóstico do distúrbio neuropsicológico infantil. Memórias II Congresso Latinoamericano de Ingeniería Biomédica, Habana. La Habana, Cuba.
- DSM-IV, 1995, Manual diagnóstico e estatístico de transtornos mentais. Porto Alegre: Artes Médicas. Trad. da edição da APA. Diagnostic and Statistical Manual of Mental Disorders, 4 ed., 1994.
- GARCIA F.B. et al, 2008. Importância dos testes sorológicos de triagem e confirmatórios na detecção de doadores de sangue infectados pelo vírus da hepatite C. Revista brasileira de hematologia e hemoterapia. 2008;30(3):218-222. Acesso julho 2014.
- HBP (2014) “Human Brain Project”. <https://www.humanbrainproject.eu/>
- HEY, T., Tansley, S., Tolle, K (2009) “The Fourth Paradigm: Data-Intensive Scientific Discovery” Microsoft Press. 1st Edition.
- KUBILIUS, J. (2013) “A framework for streamlining research workflow in neuroscience and psychology”. Front. Neuroinform. 7: 52.
- LOPES CK, OLIVEIRA CI. A dislexia na ótica do psicopedagogo. 2007. Acesso em: 15/04/2016.
- MACÁRIO COSTA, R. J. et al.(2014) “Desafios e Oportunidades em Neurociência Computacional na Educação Brasileira”. In: Grandes Desafios da Computação no Brasil. 3º Seminário. SBC, 2014a.
- MACÁRIO COSTA, R. J. et al.(2014) “Usando Tecnologias de Neurociência Computacional na Educação Brasileira”. In: Conferência Internacional sobre Informática na Educação, 2014, Fortaleza, XIX TISE, 2014b.
- MACÁRIO COSTA, R. J. et al. Classificação de pacientes com transtorno de dislexia usando redes neurais artificiais. In: Congresso Nacional de Matemática Aplicada e Computacional, 2007, Florianópolis. XXX CNMAC. Florianópolis, 2007
- MACÁRIO COSTA, R. J. ; MOUSINHO, Renata ; VIDAL, L. A. . Redes Neurais: um instrumento no rastreio (screening) de pessoas com risco de transtorno específico de leitura. In: I Congresso Ibro/Larc de Neurociências da América Latina, Caribe e Península Ibérica, 2008, Búzios - RJ. Anais do I Congresso Ibro/Larc de Neurociências, 2008.
- MACÁRIO COSTA, R. J. ; MOUSINHO, Renata ; VIDAL, L. A. . Abordagem Computacional no Screening da Dislexia e do TDAH. In: XXXII Congresso Nacional de Matemática Aplicada e Computacional, 2009, Cuiabá - MT. Anais do CNMAC 2009A.
- MACÁRIO COSTA, R. J.; MOUSINHO, RENATA ; VIDAL, L. A. . Dislexia e Inteligência Computacional: Um sistema para rastrear (Screening) pessoas com sinais de transtorno de leitura. In: 2o. Congresso Internacional de Dislexia, 2009, São Paulo. Anais do II Congresso Internacional de Dislexia, 2009B.
- MACÁRIO COSTA, R. J. et al. Redes neuronais e transtornos de aprendizagem: rastreio de pessoas com dislexia. In: Simpósio Brasileiro de Informática na Educação, 2009,

- Florianópolis. Anais do Simpósio Brasileiro de Informática na Educação, 2009C. v. 20. p. 1-10.
- MACÁRIO COSTA, R. J. et al. Abordagem tecnológica para rastreamento de pessoas com dislexia. *Tecer* (Belo Horizonte), v. 4, p. 41-53, 2011A.
- MACÁRIO COSTA, R. J. Uma Estratégia computacional na detecção da dislexia. Rio de Janeiro: Tese – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2011B.
- MACÁRIO COSTA, R. J. et al. A Computational Approach for Screening Dyslexia.. In: CBMS 2013, 2013, Porto. 26th IEEE International Symposium on Computer-Based Medical Systems, 2013.
- MOUSINHO, R. (2003) “Desenvolvimento da Leitura, Escrita e seus Transtornos. In: Goldfeld, M. Fundamentos em Fonoaudiologia”. Guanabara Koogan. 2a edição.
- NIH (2014) “BRAIN 2025: A Scientific Vision”, <http://www.nih.gov/science/brain/2025/>.
- OLIVETTE, T. W. 2009. Diagnóstico Etiológico de AVC Isquêmico em Pacientes Joinvilenses utilizando Redes Neurais Artificiais. Joinville. SC.
- RIVEROS, L. J. M., SOARES, F. S. & MUNZLINGER, E. Validade de um Sistema Especialista para o Pré-Diagnóstico da Dislexia. XXV Congresso da Sociedade Brasileira de Computação. UNISINOS – São Leopoldo / RS. 22 a 29 de julho. 2005.
- SANTOS, C.F.L. e MARINHO, D.H., 2007, Reconhecimento e manejo do distúrbio de leitura na criança. Monografia. (Aperfeiçoamento/Especialização em Residência em Pediatria) - Hospital Universitário Evangélico de Curitiba.
- SHAYWITZ, S., 2006, Entendendo a dislexia: um novo e completo programa para todos os níveis de problemas de leitura; tradução Vinicius Figueira. – Porto Alegre : Artmed.
- SHAYWITZ, S.E. SHAYWITZ, B.A. (1999) “Dyslexia In: Swaiman KF, Ashwal S. Pediatric Neurology - Principal e Practice”, Connecticut. Ed. Mosby
- SOUSA, Leonardo Ramon Nunes de; Silveira, Ismar Frango. “Desafios das Interfaces para a Aprendizagem de Pessoas com Dispraxia”, 2015. DesafIE 2015 CSBC. Acesso em 15/04/2016.
- ZAVALETA, J. et al. (2012) “DysDTool: Uma Ferramenta Inteligente para Avaliação e Intervenção no Apoio ao Diagnóstico da Dislexia”. In: CSBC- XII Workshop de Informática Médica.
- ZHONG, N. et al. (2011) “Brain Informatics”. IEEE Intelligent Systems. September, pages 16-20.
- ZHONG, N (2014) “Research Issues and Challenges on Brain Informatics Towards Computing & Intelligence in the Big Data Era”. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, 18th PAKDD, 2014.

ANEXO I –Dificuldades Características da Dislexia



Desafios na construção de um ambiente adaptativo baseado em Inteligências Múltiplas com apoio de gamificação

Renan X. Pereira Júnior¹, Márcia A. Fernandes²

¹Instituto Federal Goiano (IF Goiano)
Campos Belos – GO – Brasil

²Computação – Universidade Federal de Uberlândia (UFU)
Uberlândia – MG – Brasil

renan.junior@ifgoiano.edu.br, marcia@ufu.br

The Challenges of Informatics in Education is the adaptation of learning environments and automatic modeling of the student. This work shows a proposal of a learning environment based on Multiple Intelligences contents and form the student model. A module of gamification is necessary to motivate the student to help in the choice of the best strategy for the effectuation of the learning. We believe that the creation of this environment can facilitate the process of distance education, helping in the reduction of evasion and low yield.

Os desafios apresentados pela Informática na Educação é a adaptação de ambientes de aprendizagem e a modelagem automática. Desta forma, este trabalho apresenta uma proposta de um ambiente baseado em Inteligências Múltiplas para a adaptação do conteúdo e formação do modelo do estudante. Um módulo de gamificação se faz necessário para motivar o aprendiz e o auxiliar na escolha do melhor cenário para a melhor efetivação da aprendizagem. Acredita-se que a criação de tal ambiente possa facilitar o processo de Educação a Distância, auxiliando na diminuição da evasão e do baixo rendimento.

1. Introdução

A tecnologia, atualmente, encontra-se presente e desafiadora em todas as modalidades de ensino. Nas modalidades à distância, ou semipresenciais, os desafios norteiam, principalmente, na escolha da ferramenta para a melhor abordagem do conteúdo didático. Pesquisadores da área de Informática na Educação (IE) vêm sendo instigados a propor sistemas educacionais adaptativos, isto é, aqueles que melhor se adaptem ao modelo do estudante, individualizando / personalizando assim o processo ensino / aprendizagem. Atualmente, é esperado que um ambiente virtual auxilie estudantes e professores na disponibilização e organização dos materiais de ensino para uma aprendizagem mais efetiva.

No Brasil, existem alguns programas que incentivam a Educação a distância (EaD). Através da Universidade Aberta do Brasil (UAB) e de outros programas, o governo vem incentivando, na modalidade a distância, a formação técnica e superior dos cidadãos. Tal alternativa vem ampliando nos últimos anos o maior acesso a educação por parte da população brasileira.

Apesar dos programas apresentarem grandes vantagens, estudos recentes ainda mostram um número considerável de evasões e reprovações nestas modalidades [Martins et al. 2016, Fernandes et al. 2014]. A falta de acesso aos ambientes virtuais e dificuldades na adaptação à metodologia são apresentados como fatores que levam à reprovação e à evasão dos aprendizes [Fernandes et al. 2014]. Desta forma, é possível crer que Sistemas dotados de inteligência ainda podem ser grandes aliados para mudarmos o cenário atual.

Pesquisadores têm proposto com mais frequência sistemas que empregam técnicas de Inteligência Artificial (IA) com o objetivo de identificar características específicas do estudante e, conseqüentemente, inferir ações pedagógicas apropriadas que melhoram a prática do ensino. Por exemplo, acredita-se que cada estudante tenha um Estilo de Aprendizagem (EA) e que, aplicadas em ambientes virtuais, podem favorecer a aprendizagem do sujeito [Graf et al. 2010].

Ainda no contexto da aprendizagem, a teoria das Inteligências Múltiplas (IM) pode desempenhar um papel semelhante ao esperado dos EA [Feldman et al. 2015]. Segundo Gardner [Gardner 1983], o ser humano possui inteligências, sendo que algumas delas podem destacar. [Barbosa et al. 2004], em sua pesquisa, conseguiu associar mídias com algumas inteligências, desta forma, é possível analisar situações em que Objetos de Aprendizagem (OA) sejam elaborados para coincidir com a inteligência em destaque, tornando assim a aprendizagem mais efetiva. Assim, um estudante que possui a inteligência lógico-matemática em alto grau, materiais que envolvem associação e raciocínio podem ser apresentados a este estudante tentando assim levar uma maior absorção.

Um desafio ainda existente na pesquisa são métodos para aferição de IM's na intenção de modelar o estudante. O questionário é o método mais tradicional e o principal utilizado para tal medição, embora Gardner [Gardner 1983] critique esta prática. O mesmo desafio ocorre para a aferição dos EAs, também sendo tradicionalmente medidos por questionários. Alguns trabalhos já apresentam uma evolução na estratégia de medição [Graf et al. 2010], porém ainda é um problema em aberto na pesquisa.

Por ser um problema em aberto, estratégias distintas devem ser adotadas visando melhores soluções. Desta maneira, a gamificação pode ser uma tentativa para auxiliar no processo. A gamificação, técnica utilizada nos últimos anos em IE, tem sido aliada para auxiliar na melhoria de ambientes de aprendizagem [Borges et al. 2013]. Definida como o uso de elementos de jogos em ambientes não *games* [Deterding et al. 2011a], espera-se neste trabalho aliar IM's e gamificação, na intenção de apresentar novas contribuições para a IE. Para tal, pretende-se discutir a possível colaboração das mecânicas de jogos para a identificação, de forma automática, das inteligências que se destacam no aprendiz. Também pretende-se verificar se, com a gamificação em um ambiente adaptativo baseado em IM's, é possível encontrar características (tais como engajamento e maximização do aprendizado por parte do estudante), que já foram vistas em outros trabalhos [Borges et al. 2013].

Este artigo está organizado como se segue. A seção 2 apresenta a adaptatividade, mostrando os tipos, métodos utilizados para a criação e as abordagens usadas para a modelagem do estudante. A seção 3 apresenta a gamificação no contexto educacional, mostrando as principais mecânicas de jogos e as vantagens em utilizá-las, ao final são

apresentados trabalhos brasileiros que utilizaram gamificação. A seção 4 descreve sobre o desafio na criação de um ambiente adaptativo baseado em inteligências múltiplas com uso de gamificação, apresentando as propostas trazidas até aqui. A seção 5 traz as considerações finais discutindo como o desafio proposto pode colaborar com a educação através da Informática.

2. A Educação e a Adaptatividade

Apesar da aprendizagem adaptativa conduzida por sistemas já apresentarem benefícios educacionais, ainda existem algumas lacunas a serem preenchidas para a melhor prática do ensino. Esta seção apresentará uma visão geral da adaptatividade no contexto educacional, associando os trabalhos já realizados junto com as teorias relacionadas à aprendizagem.

Quando se trata de adaptatividade no contexto da aprendizagem, existem três pilares a serem discutidos, segundo [Group et al. 2010], são eles:

- **Individualização:** São classificados neste item ambientes que esperam o mesmo objetivo de aprendizagem para diferentes estudantes, apresentando materiais distintos, adaptando ao perfil, respeitando o ritmo de cada aprendiz.
- **Diferenciação:** Neste caso, os objetivos de aprendizagem são os mesmos, porém os métodos ou a abordagem das instruções variam de acordo com a preferência do estudante.
- **Personalização:** Este item refere a ambientes que conseguem variar tanto os objetivos quanto os materiais que serão apresentados ao estudante de acordo com a classificação colocada a cada aprendiz.

Existem diferentes estratégias utilizadas por pesquisadores para a melhor individualização da aprendizagem, dentre elas as IM's. Segundo Gardner [Gardner 1983], o ser humano possui várias inteligências, sendo que, para determinadas áreas de atuação, algumas delas podem destacar enquanto outras não. De acordo com esta teoria, oito inteligências foram identificadas e são denominadas Linguística, musical, lógico-matemática, visual-espacial, corporal-cinestésica, interpessoal, intrapessoal e naturalista. Cogita-se ainda a possibilidade de uma nona inteligência (existencial), conforme Figura 1. Desta forma, para cada indivíduo, os níveis das IM's devem variar, tornando individual a forma de aprendizagem.

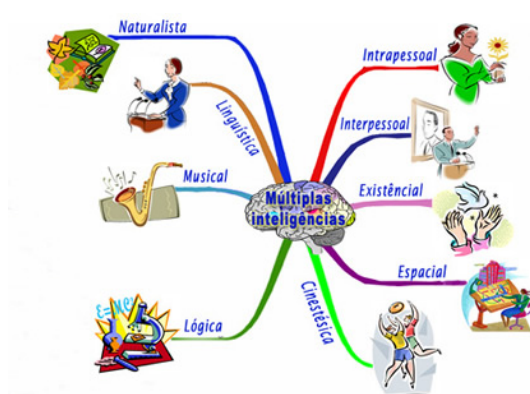


Figura 1. As Inteligências Múltiplas de Gardner [Barros 2015]

Outra estratégia também utilizada na adaptatividade em ambientes de aprendizagem são os EA. Segundo pesquisas [Feldman et al. 2015, Valaski et al. 2011], existem várias recomendações de EA, porém, os dois estilos mais utilizados são os propostos por [Felder and Silverman 1988] e [Kolb and Kolb 2005]. As IM's, se comparadas aos EAs, referem-se ao que o indivíduo é possível realizar, às habilidades, enquanto os EAs referem-se às preferências no uso das habilidades [Kelly and Tangney 2006].

No Brasil, alguns pesquisadores têm esforçado para a detecção dos EA's e recomendação de Objetos de Aprendizagem, gerando ambientes adaptativos. [Silva and Dorça 2014] apresentam uma abordagem automática para personalização do processo de ensino baseada nos EA de Felder e Silverman. [Mühlbeier and Mozzaquatro 2012] faz uma classificação do EA do estudante através de um questionário, podendo assim realizar a individualização do processo de ensino / aprendizagem. [de Moura and Fernandes 2012] apresentam uma proposta de um modelo computacional para a escolha de OAs baseado nos EA de Kolb e em IM's.

Dentre as oito IM's, Barbosa [Barbosa et al. 2004] cita que quatro delas (Linguístico-Verbal (LV), Lógico-Matemática (LM), Visual-Espacial (VE) e Cinestésico-Corporal (CC)) permitem associação com mídias tais como vídeos, apresentações e animações. Assim, para tais IM's é possível construir Objetos de Aprendizagem (OA) que forneçam conteúdos personalizados para o estudante. As três primeiras IM's (LV, LM e VE) também são apresentadas na construção do sistema adaptativo Educe [Kelly and Tangney 2006], que também apontam uma quarta, a Musical, pois acredita-se no alto poder emocional da música. Segundo Barbosa [Barbosa et al. 2004], o trabalho reduzido com IM's não traz prejuízos ao processo, pois as mesmas são independentes entre si.

Como já dito, IM's podem ser medidas através de questionários. Um questionário clássico para a medição possui 119 questões e provê um perfil descritivo das IM's do indivíduo [Feldman et al. 2015]. Outros instrumentos foram desenvolvidos tentando minimizar o número de questões e obter o mesmo resultado [Barbosa et al. 2004].

É possível notar que, na criação de ambientes adaptativos, há dois métodos de obtenção de informação para a modelagem do estudante. Segundo Graf [Graf et al. 2009], os métodos podem ser classificados como colaborativo e automático. Uma abordagem colaborativa permite que o aprendiz forneça informações de forma explícita, através de preenchimento de questionários, para a formação e atualização do seu modelo. Já na abordagem automática, através do comportamento e ações dos estudantes no sistema, o modelo é construído e atualizado de forma automática.

3. Gamificação no contexto educacional

Os sistemas dotados de inteligência desenvolvidos para o ensino procuram cada vez mais ferramentas que possam não somente enriquecer a forma como o conteúdo é apresentado, mas também fornecer um ambiente confortável para o estudante, levando estímulo e o tornando engajado no processo. Neste aspecto, a gamificação, definida como o uso de elementos de jogos em ambientes não *games* [Deterding et al. 2011b], vem sendo trabalhada nos últimos anos em ambientes educacionais.

Segundo Bunchball [Bunchball 2010], para a técnica de gamificação, algumas das mecânicas de jogos mais comuns são: pontos; níveis; desafios, troféus, emblemas / meda-

lhas e conquistas; bens virtuais; placar de classificação, ranking e tabela de pontuação. A Tabela 1 apresenta alguns exemplos de mecânica de jogos e suas respectivas explicações. Ainda segundo Bunchball [Bunchball 2010], as mecânicas de jogos impulsionam as pessoas pelas dinâmicas.

Tabela 1. Algumas mecânicas de jogos [Bunchball 2010]

Mecânicas de Jogos	
Tipos	Significado
Níveis	Indica que o usuário atingiu uma meta
Desafios	Representam missões para serem realizadas e posteriormente geram recompensas
Troféus, Emblemas/ Medalhas, Conquistas	Reconhecimentos visíveis que o usuário conquistou desafios ou atingiu níveis
Placar de Classificação, Ranking, Tabela de pontuação	Posiciona os usuários em relação aos demais
Recompensa	É dada ao usuário após a ocorrência de uma ação para motivá-lo a realizar aquilo novamente

Recompensa, status, realização, auto expressão, concorrência, altruísmo, etc., são os impulsionadores e gerados a partir de mecânicas de jogos. Por exemplo, a recompensa é dada ao usuário após a ocorrência de uma ação para motivá-lo a realizar aquela ação novamente. Em programas de fidelidade feitos por companhias aéreas, os pontos que são dados aos usuários através de parcerias e voos realizados, ao final são convertidos em passagens aéreas. Os usuários desta forma se sentem motivados em participar e se fidelizarem aos programas.

Em um ambiente educacional, o principal objetivo da gamificação tem sido a utilização de mecânicas de jogos para tornar o processo educativo mais interessante, iterativo e atrativo [Domínguez et al. 2013]. Borges [Borges et al. 2013], após realizar um mapeamento sistemático, verificou que, na educação, a gamificação tem atuado principalmente nas seguintes categorias: Aprimoramento de habilidades; Desafios; Guidelines; Engajamento; Maximização do aprendizado; Mudança de comportamento e socialização.

A gamificação, segundo alguns pesquisadores da área de educação, pode apresentar problemas. [Muntean 2011], por exemplo, comenta o risco de se ter um ambiente gamificado e o estudante ficar preso à necessidade de receber recompensas para executar alguma atividade. Neste caso, o desenvolvedor deve preocupar em controlar também o comportamento do estudante, para que a gamificação seja um acréscimo para a motivação e não um fator dependente.

No Brasil algumas pesquisas têm sido voltadas para o uso de gamificação em sistemas educacionais. [Ferreira et al. 2015] utiliza a gamificação como um módulo para enriquecer um ambiente educacional ubíquo e o tornar mais sociável. [da Rocha Seixas et al. 2014, da Rocha Seixas et al. 2016] apresenta a gamificação como estratégia de engajamento de estudantes do ensino fundamental. O trabalho também apresenta uma visão bem detalhada sobre o que pode ser tratado como engajamento. [Neto et al. 2015] faz uma análise do uso da gamificação em uma turma do primeiro ano do ensino médio e verifica uma melhora na aprendizagem. Nos anos anteriores do Desafie

também foi possível ver algumas propostas com uso de gamificação em Sistemas Tutores Inteligentes [Andrade et al. 2013, Andrade et al. 2014].

4. Proposta: Um caminho para Vencer o Desafio

Através das pesquisas apresentadas anteriormente foi possível ver alguns cenários já existentes e algumas questões em aberto para a Informática na Educação. O ensino orientado por sistemas para, principalmente, auxiliar na EaD é um contexto bem discutido mas que ainda apresentam lacunas a serem preenchidas, como, por exemplo, a modelagem automática do estudante, a motivação e o engajamento para a diminuição de evasão, o controle de acesso aos sistemas, etc.

Na intenção de contribuir com a IE, este trabalho tem como proposta o uso da gamificação em um ambiente de aprendizagem adaptativa baseado em Inteligências Múltiplas. Como as IM's são independentes entre si, a princípio pretende-se expor um cenário somente com as quatro inteligências que [Barbosa et al. 2004] associou a mídias.

O ambiente deve ser constituído de quatro cenários distintos, cada um formado de acordo com os OAs classificados para a respectiva IM. Por exemplo, para o cenário correspondente à inteligência Linguista-Verbal, vídeos, materiais para escrever e promoção do debate são itens que devem existir estimulando o aprendizado do estudante que possui um perfil verbal.

Por ter a intenção de criar uma abordagem automática, o aprendiz não fornecerá nenhuma informação inicial, sendo que a criação e atualização do modelo será realizada de acordo com a interação do usuário com o sistema. Neste caso, métodos probabilísticos devem ser utilizados para construir o modelo do estudante, isto deve ao fato de lidar com um desconhecido. Em Sistemas Tutores Inteligentes, o uso de Redes Bayesianas são bem aceitos como método probabilístico para modelar, por exemplo, o conhecimento do estudante [Vier et al. 2015].

O ambiente também deve ter a individualidade como critério de adaptatividade. Desta forma, os aprendizes terão o mesmo objetivo, porém o ritmo e os OAs fornecidos serão individuais. A gamificação fornecerá as informações necessárias para o estudante compreender o nível de aprendizagem em cada cenário. A princípio o estudante poderá escolher um cenário ou será selecionado aleatório, sem a necessidade de preenchimento de questionários, como já é de praxe em alguns ambientes adaptativos. Neste cenário o estudante terá materiais para aquisição de conhecimento e avaliações ao término de cada módulo para aferir a aprendizagem. A mudança de cenário poderá ocorrer sempre que o estudante achar viável. Agentes atuando junto às mecânicas de jogos ajudarão o mesmo a compreender em qual cenário está tendo mais fracasso ou sucesso.

Após um certo uso do sistema, já será possível formar um modelo, visualizar qual inteligência do estudante está sobrepondo e, desta forma, recompensa poderá ser dada para o mesmo manter no cenário que a aprendizagem apresenta mais sucesso. O modelo estará em constante atualização decorrente do tempo gasto, das interações e dos resultados nas avaliações.

A gamificação, neste caso, não ficará restrita somente ao auxílio na modelagem do estudante e recomendação do material de acordo com sua IM. Como já apresentado em outros ambientes de aprendizagem, não é esperado diferente em um sistema baseado em

inteligências múltiplas: É esperado que as mecânicas de jogos também possam estimular o engajamento e maximização do aprendizado, aprimorando assim as habilidades dos estudantes. Não pretende-se utilizar rankings, tabela de pontos e placar de classificação. A pontuação será individual, pois não tem a intenção de criar concorrência no ambiente.

5. Considerações Finais

Como apresentado no início deste artigo, a EaD no Brasil ainda enfrenta grandes problemas de evasão e grande queixa quanto à metodologia e aos ambientes virtuais de aprendizagem. A proposta apresentada aqui espera ajudar na solução deste problema atual. Com a criação de um ambiente com IM's e gamificação, pode possibilitar a motivação do estudante durante os estudos. O aprendiz poderá aprender com os materiais que mais possam despertar interesse e, através das mecânicas de jogo, se sentir engajado no processo educacional.

A individualidade na aprendizagem faz com que se crie um desafio durante a construção de sistemas educacionais. Com as propostas expostas anteriormente, espera-se um sistema de aprendizagem adaptativo, baseado em inteligências múltiplas, com uso de gamificação, para classificar estudantes de acordo com sua inteligência e também mantê-los motivados com uso de mecânicas de jogos.

Resumindo, os principais desafios discutidos são:

1. Determinar um modelo de estudante que seja capaz de obter informações relevantes sem o uso de questionários, utilizando a teoria de Inteligências Múltiplas;
2. Implementar atividades a serem realizadas através do computador que permitam manter a motivação do estudante, aumentar o aproveitamento no processo de ensino/aprendizagem e, principalmente, fornecer as informações que melhor identifiquem o perfil do estudante. Para tanto foi sugerida a gamificação;
3. Criar métodos probabilísticos para assistir a evolução do estudante; e
4. atualizar estas informações durante o processo a fim de acompanhar a evolução do estudante ou, até mesmo constatar que esta não ocorreu.

Trabalhos futuros tangem na integração do sistema construído a um Ambiente Virtual de Aprendizagem, de preferência o Moodle ¹, por ser mais utilizado em situações reais de ensino. Desta forma, permitirá maior número de acessos pelos estudantes. Também pretende-se criar módulos para tornar o ambiente colaborativo, talvez aproveitando as mecânicas de jogos, pois acredita-se que a colaboração possa enriquecer a aprendizagem e permitir a observação mais efetiva do comportamento do estudante.

Referências

- Andrade, F., Chalco, G., and Isotani, S. (2014). Em direção à gamificação de sistemas tutores inteligentes: Aplicando a teoria de fluxo Ótimo no design instrucional de cenários de aprendizagem colaborativa. In *Anais do DesafIE-IV Workshop de Desafios da Computação Aplicada à Educação*, volume 1.
- Andrade, F., Pedro, L., Lopes, A., Bittencourt, I., and Isotani, S. (2013). Desafio do uso de gamificação em sistemas tutores inteligentes baseados em web semântica. In *Anais do DesafIE-III Workshop de Desafios da Computação Aplicada à Educação*, volume 1.

¹Moodle: disponível em <http://moodle.com/>

- Barbosa, A. T. R. et al. (2004). *Mecanismo de adaptação baseado em redes neurais artificiais para sistemas hiperídia adaptativos*. PhD thesis, Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-graduação em Engenharia Elétrica.
- Barros, J. (2015). *Inteligências múltiplas – novo conceito em educação*.
- Borges, S. d. S., Reis, H. M., Durelli, V. H., Bittencourt, I. I., Jaques, P. A., and Isotani, S. (2013). Gamificação aplicada à educação: Um mapeamento sistemático. *Simpósio Brasileiro de Informática na Educação*, pages 234–243.
- Bunchball, I. (2010). Gamification 101: An introduction to the use of game dynamics to influence behavior [white paper].
- da Rocha Seixas, L., Gomes, A. S., and de Melo Filho, I. J. (2016). Effectiveness of gamification in the engagement of students. *Computers in Human Behavior*, 58:48–63.
- da Rocha Seixas, L., Gomes, A. S., Melo Filho, I. J., and Rodrigues, R. L. (2014). Gamificação como estratégia no engajamento de estudantes do ensino fundamental. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 25, page 559.
- de Moura, F. F. and Fernandes, M. A. (2012). Uma proposta de um modelo computacional que usa pso para a escolha de objetos de aprendizagem baseado na espiral de kolb e nas inteligências múltiplas. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 23.
- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011a). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM.
- Deterding, S., Khaled, R., Nacke, L. E., and Dixon, D. (2011b). Gamification: Toward a definition. In *CHI 2011 Gamification Workshop Proceedings*, pages 12–15.
- Domínguez, A., Saenz-de Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., and Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63:380–392.
- Felder, R. M. and Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering education*, 78(7):674–681.
- Feldman, J., Monteserin, A., and Amandi, A. (2015). Automatic detection of learning styles: state of the art. *Artificial Intelligence Review*, 44(2):157–186.
- Fernandes, J., da Silva Ferreira, A., de Oliveira Nascimento, D. C., Shimoda, E., and Teixeira, G. F. (2014). Identificação de fatores que influenciam na evasão em um curso superior de ensino a distância. *PerspectivasOnLine 2007-2010*, 4(16).
- Ferreira, H., Araújo, R. D., Souza, P., Chagas, S., Dorça, F., and Cattelan, R. (2015). Gamificação em ambientes educacionais ubíquos. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, page 509.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. NY: Basics.

- Graf, S. et al. (2009). Advanced adaptivity in learning management systems by considering learning styles. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 03*, pages 235–238. IEEE Computer Society.
- Graf, S., Ives, C., et al. (2010). A flexible mechanism for providing adaptivity based on learning styles in learning management systems. In *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on*, pages 30–34. IEEE.
- Group, N. E. T. P. T. W. et al. (2010). Transforming american education: Learning powered by technology.
- Kelly, D. and Tangney, B. (2006). Adapting to intelligence profile in an adaptive educational system. *Interacting with Computers*, 18(3):385–409.
- Kolb, A. Y. and Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of management learning & education*, 4(2):193–212.
- Martins, C. Z., de Lima Terçariol, A. A., and Gebran, R. A. (2016). Percepções do tutor: a evasão de alunos de curso de administração na modalidade a distância. *REVISTA ELETRÔNICA PESQUISEDUCA*, 7(14):470–492.
- Mühlbeier, A. R. and Mozzaquatro, P. M. (2012). Estilos e estratégias de aprendizagem personalizadas a alunos das modalidades presenciais e a distância. *Revista Brasileira de Informática na Educação*, 20:132.
- Muntean, C. I. (2011). Raising engagement in e-learning through gamification. In *Proc. 6th International Conference on Virtual Learning ICVL*, pages 323–329.
- Neto, A., da Silva, A. P., and Bittencourt, I. I. (2015). Uma análise do impacto da utilização de técnicas de gamificação como estratégia didática no aprendizado dos alunos. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, page 667.
- Silva, D. H. and Dorça, F. A. (2014). Uma abordagem automática para personalização do processo de ensino baseada em estilos de aprendizagem em sistemas adaptativos e inteligentes para educação a distância. *Revista Brasileira de Informática na Educação*, 22(02):1.
- Valaski, J., Malucelli, A., and Reinehr, S. (2011). Revisão dos modelos de estilos de aprendizagem aplicados à adaptação e personalização dos materiais de aprendizagem. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 1.
- Vier, J., Gluz, J., and Jaques, P. A. (2015). Empregando redes bayesianas para modular automaticamente o conhecimento dos alunos em lógica de programação. *Revista Brasileira de Informática na Educação*, 23(02):45.

O Processo de Autoavaliação em Ambientes Colaborativos e sua Relação com a Mineração de Dados Educacionais

Hugo Vieira Lucena de Souza¹, Ivanildo Melo Filho^{1,2},

Rodrigo Lins Rodrigues^{1,3}, Alex Sandro Gomes¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50.732-970 – Recife – PE – Brazil

²Instituto Federal de Pernambuco (IFPE) – Campus Belo Jardim
Caixa Postal 55154-065 – Belo Jardim – PE - Brazil

³Departamento de Educação - Universidade Federal Rural de Pernambuco (UFRPE) –
Caixa Postal 52171-900 – Recife – PE – Brazil

{hvls, asg}@cin.ufpe.br, ivanildo.melo@belojardim.ifpe.br,
rodrigo.linsrodrigues@ufrpe.br

***Abstract.** This paper presents a study on the different approaches of metacognitive phenomenon of self-assessment in learning and the prospects of possible relationships that may exist with the techniques proposed by Educational Data Mining. All sections are shown that develop research work that include the subject, some previous results obtained from these studies, and ultimately, the possible challenges are discussed to be found for future development of new solutions.*

***Resumo.** O presente trabalho apresenta um estudo realizado acerca das diferentes abordagens do fenômeno metacognitivo da autoavaliação da aprendizagem e as perspectivas de possíveis relações que possam existir com as técnicas propostas pela Mineração de Dados Educacionais. Nas seções são apresentados trabalhos que desenvolvem pesquisas que englobam o tema, alguns resultados prévios obtidos através destes trabalhos, e por fim, são discutidos os possíveis desafios a serem encontrados para um futuro desenvolvimento de novas soluções.*

1. Introdução

O avanço das tecnologias ao longo dos últimos anos tem se tornado um fator decisivo para a reorganização dos métodos da aprendizagem. O uso das novas tecnologias replica a concepção de cenários que englobam situações cada vez mais colaborativas, em que instrumentos como computadores, tablets e smartphones tornam-se fundamentais para o desenvolvimento de habilidades cognitivas e metacognitivas de professores, alunos e tutores nas mais diversas modalidades e abordagens do ensino (RAMOS et al.,2014).

Dentre alguns dos principais fenômenos metacognitivos, responsáveis por potencializar a efetividade da aprendizagem colaborativa está a autoavaliação (BOUD, 2007). Este fenômeno busca identificar quais características das pessoas expressam o

seu comportamento através da interação, sendo responsável principalmente por subsidiar melhorias perceptíveis no envolvimento do processo, no engajamento das atividades realizadas e na forma de entender a importância dos papéis que o professor, o tutor e o aluno exercem nas salas de aula. Entretanto, mensurar os índices de autoavaliação da aprendizagem não se torna uma atividade simples de ser realizada. Vários mecanismos propostos pela psicologia indicam que há limitações presentes, principalmente quando são considerados aspectos temporais (SCARTEZINI, 2014). Isto implica dizer que, o comportamento de uma pessoa adquire uma forma continuamente mutável e que os impactos podem ser percebidos à medida em que ocorre a evolução do curso em que ela participa.

Uma área que busca propor o entendimento de cenários da aprendizagem através da seleção, análise e tratamento dos dados é a Mineração de Dados Educacionais (EDM) (LINS et al., 2014). Os mecanismos propostos por ela oferecem totais condições para avaliar dados e informações provenientes de professores, tutores e alunos com o objetivo de prever eventuais situações que podem ocorrer em uma universidade ou escola utilizando a interação como fonte de captura. Portanto, o objetivo deste trabalho é apresentar quais são as possíveis relações que possam existir entre o processo de autoavaliação da aprendizagem desenvolvida em ambientes colaborativos através das técnicas da Mineração de Dados Educacionais. As seções descrevem os resultados publicados de outros trabalhos com o intuito de serem entendidos quais fatores colaboram para o desenvolvimento de soluções que possam surgir, além de destacar os novos desafios encontrados como perspectivas do futuro.

Este artigo está organizado da seguinte forma: a seção 2 descreve a autoavaliação em ambientes colaborativos. Na sequência são apresentadas as abordagens para mensuração da autoavaliação. A seção 4 traz as perspectivas da autoavaliação e o suporte do EDM nesses ambientes. Por fim, são apresentadas as considerações finais.

2. A Autoavaliação em Ambientes Colaborativos

O desenvolvimento do processo da aprendizagem é uma tarefa árdua a ser realizada por muitos professores, tutores ou alunos. Para Barros (2009) um ambiente colaborativo é aquele em que são utilizados mecanismos da tecnologia, dentre aplicativos e serviços que permitem que as pessoas se comuniquem de diversas formas, exercendo um papel fundamental durante a troca de informações na construção social do conhecimento.

Em cada cenário colaborativo percebem-se individualidades que variam mediante ao envolvimento no processo da aprendizagem. Dentre algumas individualidades mediadas pelos professores e tutores estão aquelas relacionadas às habilidades metacognitivas (BARROS, 2009). A interação é o primeiro fator determinante a ser observado quando é estabelecida uma linguagem entre os mesmos para que todos reconheçam a importância que exercem na evolução do processo. Outro fator determinante destacado por Barros (2009) é o entendimento da ideologia. A ideologia na aprendizagem colaborativa pode ser entendida como a forma em que as pessoas verificam suas capacidades e limitações, além da maneira que eles podem contribuir para o sucesso das atividades. O alinhamento da interação social com a ideologia depende da reflexão a ser realizada pelos professores, tutores e alunos no processo da aprendizagem, algo que pode ser obtido através das autoavaliações.

Para Boud (2007) a autoavaliação da aprendizagem pode ser entendida simplesmente quando os professores, tutores e alunos se perguntam: o que eu estou fazendo? Isto é suficiente? Isto está correto? O autor explica que este fenômeno metacognitivo, quando aplicado à aprendizagem, busca mensurar o grau de envolvimento da pessoa no processo, e as capacidades que ela tem de criar padrões ou critérios individuais de avaliação para realizar julgamentos que verificam se seus objetivos estabelecidos estão sendo atendidos ou não.

A autoavaliação da aprendizagem pode ser realizada em contextos que variam de acordo com a teoria da aprendizagem e a modalidade do ensino. Os contextos são situações que estabelecem os personagens, os agentes regulados, reguladores e os interlocutores. Kelvin (2012) cita que também se torna importante detectar as variáveis que inferem no processo para verificar deficiências que tenham sido encontradas. O autor apresenta alguns contextos comuns de serem deparados no dia-a-dia:

- **Autoavaliação do aluno regulada pelo professor:** é aquela em que o autoavaliado é o aluno e o interlocutor que supervisiona o processo é o professor. Neste caso, pode-se entender que o professor delimita quais aspectos dos alunos precisam ser corrigidos para replanejar suas aulas.
- **Autoavaliação do aluno interlocutada por outro aluno:** neste caso um aluno é o personagem que se autoavalia, mas o sem ter a regulação do seu colega. O colega serve apenas como um personagem que realiza uma interação simples, servindo como fonte para a criação dos padrões de autoavaliação.
- **Autoavaliação do aluno regulada por si próprio:** é aquela em que o personagem que se autoavalia e se regula são a mesma pessoa. Neste caso, o aluno reflete utilizando a sua própria figura, ações e características como evidências que porventura possam melhorar sua aprendizagem.

Kelvin (2012) cita que as autoavaliações podem ser realizadas em processos estruturados da aprendizagem formal ou em um contexto espontâneo da aprendizagem informal. A efetividade depende dos instrumentos utilizados para que os dados gerados pelos autoavaliados sejam armazenados e tratados. Uma modalidade de ensino com condições favoráveis que atende este objetivo é o ensino a distância (EAD), através do desenvolvimento de soluções que registram a rotina dos alunos nas salas de aula.

2.1. Abordagens de Autoavaliação da Aprendizagem em LMS

A expansão do ensino a distância tem sido algo que merece um certo destaque para a área da educação. Em um estudo realizado pela Associação Brasileira de Educação a Distância (ABED), apresentou-se uma perspectiva de crescimento no aumento nas matrículas em cursos de graduação que chegou à uma margem de 30,4% (FREITAS e FERREIRA, 2013). Isto significa dizer que cursos como pedagogia, administração e serviço social tiveram um aumento de matrículas de 34%, 27% e 8% respectivamente, contabilizando todas as matrículas realizadas nas modalidades presencial e EAD.

O comportamento dos alunos é algo complexo a ser personificado e autoavaliado em um ambiente colaborativo a distância, principalmente quando se trata de um *Learning Management System* (LMS). Segundo Chi (2009), o LMS engloba o uso de aplicações e serviços que disponibilizam para os professores e tutores a capacidade de

gerenciar as atividades realizadas pelos alunos, como também registrar os dados resultantes da interação com o sistema, tais como data, hora, tempo de acesso, quantidade de tarefas entregues, além dos dados sociais, como idade, sexo e faixa etária.

Uma das primeiras tentativas com resultados interessantes a serem discutidos nesta abordagem citada acima foi publicada por Brusilovsky e Sosnovsky (2005), quando os autores elaboraram exercícios de autoavaliação implementados através de um sistema, para verificar o grau de evolução da aprendizagem de seus alunos na linguagem de programação C. “*Os nossos estudos demonstram que quando QuizPACK é usado em um ambiente extra-classe de autoavaliação, o mesmo torna-se uma ferramenta de aprendizagem excepcional*” (Brusilovsky e Sosnovsky, 2005, p. 1).

O impacto percebido foi que “*os próprios alunos elogiaram o sistema como ferramenta de aprendizagem e demonstraram que sua utilização no modo de autoavaliação pode ser significativamente expandido motivando-os a praticá-los mais*” (Brusilovsky e Sosnovsky, 2005, p. 1). O estudo apontou que os fatores que delimitam uma autoavaliação qualitativa são os que ponderam parâmetros para que as atividades sejam controladas, senão a falta de engajamento afeta os resultados do processo.

Outra abordagem é proposta por HatziaPOSTOLOU e Paraskakis (2010) com a ideia de que o que potencializa a autoavaliação da aprendizagem através de um LMS é o desenvolvimento do *Design of the Student Feedback Space*. Os efeitos de uma nova forma de tornar o retorno da comunicação personalizável, com toques de humor, espontaneidade e informalidade são perceptíveis quando comparados aos métodos e técnicas tradicionais como fóruns, *chats* ou *e-mails*.

2.2. A Autoavaliação da Aprendizagem em um Contexto U-Learning

As mudanças constantes nos hábitos das pessoas, a temporalidade em que estas mudanças ocorrem e os crescentes avanços da tecnologia são fatores que influenciam a autoavaliação da aprendizagem. Muitas atividades exercidas pelos alunos ultrapassam os limites das universidades e escolas, tornando as responsabilidades coletivas em situações individuais, recomendadas por uma aprendizagem ubíqua (SOUSA, 2015).

A aprendizagem ubíqua é aquela em que as pessoas aprendem algo em qualquer lugar e a qualquer momento (SOUSA, 2015). De acordo com Sousa (2015), um dos principais desafios desta abordagem na área da educação é entender quais mecanismos são importantes para compreender o comportamento humano em cenários que utilizam aplicativos ou dispositivos móveis, e que não estão ao alcance de professores e tutores, com o intuito de propor formas de capturar os dados gerados para em seguida avaliá-los.

Guzmán et al., (2007) realizaram um estudo com 317 testes de autoavaliação da aprendizagem em alunos utilizando apenas tecnologias móveis. Os resultados apontam melhorias no aumento da permanência do tempo do aluno no aplicativo, além de uma mudança na postura dos mesmos estruturarem novas indagações. Algo interessante destacado pelos autores é que o formato das perguntas foi afetado, sendo identificada uma evolução na elaboração, que antes eram causais, e agora apresentavam características explicativas ou conclusivas.

A criação de uma situação-problema é uma das teses defendidas por Economides (2009) para a autoavaliação da aprendizagem em um cenário ubíquo. Em seu estudo ele

verificou que a criação de contextos de pessoas, organizados como espécies de perfis, gera condições favoráveis para mapear informações comuns ou restritas em que os professores e tutores poderiam atuar. Entretanto, o autor cita que é preciso ter uma certa atenção acerca dos fatores demográficos e sociais, principalmente aqueles que abrangem a cultura, a etnia e a experiência nas atividades desenvolvidas.

Uma pesquisa que tentou comprovar as relações existentes entre a autoavaliação da aprendizagem e computação ubíqua foi a criação da ferramenta *QuestInSitu* (SANTOS et al., 2011). A estratégia foi alinhar cenários cotidianos com pequenos testes com perguntas e respostas. Em um âmbito do professor, observou-se que os alunos melhoraram a comunicação, o trabalho em equipe e a orientação acerca dos objetivos reais das atividades. No âmbito das perguntas e respostas realizadas entre os alunos, percebeu-se que autoavaliação da aprendizagem com o aplicativo proporcionou melhorias em aspectos tais como atenção, motivação e curiosidade.

2.3. Entendendo a Autoavaliação da Aprendizagem em Ambientes *Blended Learning*

Os conceitos de aprendizagem em LMS e aprendizagem ubíqua expandem os horizontes em que autoavaliação pode ser implementada. As diretrizes propostas por estas linhas servem como requisito para a inserção de uma outra linha, conhecida como *Blended Learning*. Esta busca combinar métodos de aprendizagem tanto do ensino tradicional, quanto *online*, no intuito de integrar os mecanismos colaborativos necessários para torna-los adaptáveis aos cenários que possibilitam o desenvolvimento das competências e habilidades dos alunos (PEREIRA et al., 2007).

Em sua pesquisa acerca das estratégias *Blended Learning*, adaptadas com softwares e testes de autoavaliação de aprendizagem, em disciplinas presenciais de anatomia humana, Pereira et al., (2007) entenderam que cada grupo de alunos concentra características a serem descobertas, tratadas e melhoradas quando estão combinadas diferentes modalidades do ensino. Eles dizem que “*avaliar o conhecimento adquirido por meio de um teste de autoavaliação local, que incorporou um mock-up online de um teste de múltipla escolha similar no projeto, é um uma forma de estabelecer um parâmetro fundamental para acompanhar a evolução do grupo, quando forem adotadas diferentes formas de compor as notas de um exame* (PEREIRA et al., 2007, p.4);

Pereira et al., (2007) citam também que “*ao mesmo tempo, as disciplinas devem oferecer a realização dos exercícios, e atividades para autoavaliação, que permitam que o acompanhamento individual dos alunos subsidie o progresso coletivo durante o curso, desde que se tenha uma infraestrutura de T.I. que disponibilize o acesso aos materiais para complementar a aprendizagem*” (PEREIRA et al., 2007, p.6). A discussão evidencia uma visão acerca da importância de alinhar métodos de aprendizagem com o uso da tecnologia, mas sem descartar a responsabilidade de conhecer as barreiras de aprendizagem encontradas em cursos superiores.

Delialioglu e Yildirim (2007) investigaram que autoavaliação pode replicar na aprendizagem quando os métodos adaptados são desenvolvidos apenas em um ambiente *online*. O trabalho aponta que através análise de bases de dados virtuais, sistematizadas por uma plataforma, e consolidadas com uma alimentação substancial de dados, se torna possível mensurar questões como o desempenho e a efetividade da aprendizagem. Para

isto, é preciso identificar quais questões dos fenômenos podem ser representadas por construtos, para que em seguida, se escolham as técnicas adequadas predição dos dados.

3. Abordagens de Mensuração da Autoavaliação da Aprendizagem

A mensuração de um fenômeno metacognitivo é uma tarefa difícil de ser realizada. Vários aspectos como a participação, o engajamento, a regulação e a avaliação de uma situação em que a construção do conhecimento se torna um processo complexo de ser entendido, influenciam o que de fato será mensurado e como esta mensuração ocorrerá. Nas seções abaixo são discutidas abordagens qualitativas e quantitativas que buscam mensurar a autoavaliação da aprendizagem.

3.1. O Papel da Psicologia e seus Mecanismos de Mensurar a Autoavaliação da Aprendizagem

O comportamento das pessoas é algo subjetivo para ser mensurado. No campo da psicologia aplicada a educação estão propostos métodos que tentam entender quais características associáveis ao fenômeno poderão ser mensuradas em professores, tutores e alunos. Em um estudo que englobou a abordagem desta perspectiva, Nunes e Nobre (2013) montaram os perfis de atuação com o objetivo de entender quais são os impactos e as limitações que podem ser verificados quando os mesmos reconhecem ou não de forma correta suas atribuições em salas virtuais de ensino a distância.

Os autores explicam que o processo de autoavaliação da aprendizagem pode ser mensurado através de uma abordagem metodológica quantitativa. O método proposto pela psicologia orienta que inicialmente sejam utilizados instrumentos objetivos como questionários estruturados que buscam expressar as respostas dos entrevistados, ou a aplicação de testes de lógica, percepção ou raciocínio como o Teste de Raven¹, Teste de Rorschach² e o Teste de Bender³. Quando as respostas estão registradas, em uma segunda etapa, os mecanismos são alinhados ao método quando são estabelecidas as variáveis de investigação, as hipóteses, os estudos abrangentes, além das ferramentas de coleta e análise para o tratamento dos dados através de métodos da estatística.

Em uma análise de 934 resumos de painéis de três Congressos Nacionais de Avaliação Psicológica, realizados pelo Instituto Brasileiro de Avaliação Psicológica (IBAP) em 2003, 2005 e 2007, (Silva et al., 2007), constataram que a utilização desses procedimentos quantitativos para a mensuração de fenômenos metacognitivos nos eventos chegou a respectivos 74,2%, 49,4% e 64,4%. Isto implica dizer que, quando contabilizados os percentuais de todos os trabalhos, e comparados os métodos de mensuração quantitativos, aos qualitativos, os valores chegaram aos 62% para os métodos quantitativos e 16,4% para os métodos qualitativos. A discussão dos autores concluiu que apesar dos instrumentos citados no método quantitativo serem relevantes, por estabelecerem parâmetros seguros na mensuração de um fenômeno, a quantidade de instrumentos e as formas de captação de dados são limitadas. Eles evidenciam que “os

¹ O Teste *Raven* é um teste que busca aferir a capacidade progressiva de percepção e lógica do avaliado em relação aos eventos do universo (SCARTEZINI, 2014)

² O Teste *Rorschach* é um teste de percepção do avaliado acerca do entendimento dos fenômenos às imagens dispostas em borrões de tinta (SCARTEZINI, 2014).

³ O Teste de *Bender* é um teste que busca avaliar o raciocínio do avaliado através de estímulos desenvolvidos pelo campo visomotor (SCARTEZINI, 2014).

instrumentos objetivos foram os mais utilizados, porém com uma variabilidade o que não possibilitou que um destes instrumentos se destacasse dentre os demais” (SILVA et al., 2007, p. 12). Para os autores, *“talvez a baixa frequência total da maioria dos testes indica a dificuldade de dar continuidade às pesquisas de um único instrumento”* (SILVA et al., 2007, p. 12).

Scartezini (2014) cita que mensurar a autoavaliação através de um método quantitativo torna esta atividade um processo somatório e não finalizado. Isto significa dizer que a compreensão do fenômeno requer uma observação duradoura, entendendo o comportamento das pessoas com o objetivo de obter o conhecimento adequado de suas características pessoais, principalmente quando o público-alvo são grupos grandes de pessoas. Neste aspecto, a abordagem qualitativa da psicologia recomenda que esta perspectiva depende da organização de um espaço adequado para aprendizagem, devido aos recursos metodológicos são provenientes de múltiplas fontes de informação.

3.2. Estratégias da Mineração de Dados Educacionais na Busca da Medição da Autoavaliação da Aprendizagem

A concepção de atividades que propõem uma aprendizagem diferenciada atualmente requerer uma visão crítica acerca dos perfis socioculturais das pessoas. Para os alunos, as atividades se tornam responsabilidades obrigatórias de serem realizadas e em determinadas situações podem adquirir pesos de importância diferentes. Com a variação destes pesos, eles constituem características tão específicas que muitas vezes se tornam difíceis de serem mapeadas pelos professores ou tutores (RAMOS et al., 2014).

Diante desse contexto, mensurar a autoavaliação da aprendizagem de forma segura requer a utilização de mecanismos que captem evidências suficientes destas individualidades ao mesmo tempo que não as tornem confusas com as informações geradas por perfis tão heterogêneos. Imagine, por exemplo, que cada aluno tem um comportamento na plataforma aprendizagem, e que, este comportamento, infere diretamente no seu envolvimento no processo da aprendizagem durante a realização das tarefas, na participação dos chats e no cumprimento dos prazos (RAMOS et al., 2014).

Desta forma, a *Mineração de Dados Educacionais* busca propor os métodos para exploração de dados gerados em plataformas educacionais e que possam alinhar de maneira contínua essa relação (HEINER et al., 2007). Heiner et al., (2007) cita que a aplicação da EDM, sigla da disciplina, desenvolve ações direcionadas para entender os dados comportamentais da interação quando forem utilizadas as seguintes estratégias:

- **Definir de um problema a ser analisado:** compreende entender o que o usuário deseja descobrir acerca do fenômeno. Esta etapa se torna similar a de um projeto de software em que são coletados requisitos através da realização de entrevistas.
- **Adquirir e avaliar os dados do fenômeno:** consiste em identificar, formatar, gerar amostras, calibrar, testar e validar dados, que podem delimitar as características da consulta acerca do entendimento e descrição do fenômeno.
- **Prototipar os dados e gerar modelos:** consiste em criar modelos descritivos, metadados e protótipos de consultas. A prototipagem também pode estar alinhada com métodos de busca ou sistemas de recomendação para decisões.
- **Detectar os efeitos do desengajamento:** consiste em verificar as causas e efeitos da participação com empenho dos cursistas relacionando os dados.

- **Medir os efeitos das intervenções individuais:** consiste em avaliar os efeitos de pequenas mudanças nas entregas de atividades, prazos e eventos quando aplicados nas pessoas de maneira direta ou através de terceiros.
- **Prever o desempenho e o comportamento dos alunos:** o desempenho é uma virtude que reflete nos resultados. Desta forma, sua análise e o monitoramento são estratégias sugeridas pela EDM.
- **Avaliar o esforço de aprendizagem dos alunos:** o esforço é relativo a análise de uma série de aspectos que influenciam os alunos. Estes podem variar dos mais simples como a inclusão digital aos críticos como a economia de um local.

Além destas técnicas, a EDM recomenda a montagem de outros cenários quando incluídos outros atores, como os responsáveis pelo suporte do sistema ou funcionários de polos de ensino. É importante que as entidades participativas, os papéis e as responsabilidades estejam bem representadas, pois, esse objetivo só pode ser alcançado com a integração consolidada das pessoas, dos processos e das tecnologias.

4. Perspectivas da Autoavaliação da Aprendizagem com o Uso da EDM em Ambientes Colaborativos

Os estudos que potencializam a aplicação de técnicas de mineração de dados educacionais em fenômenos cognitivos têm ganhado uma visibilidade cada vez mais consolidada através de novas pesquisas. Santana et al., (2014) abordaram em seu estudo uma contribuição relevante do uso das técnicas proporcionadas pela EDM na avaliação de perfis de alunos no ambiente virtual Moodle. Os autores dimensionaram os perfis com a criação de atributos de maneira que fossem detectados os dados que afetariam no desempenho dos alunos através do número total de interações com ambiente através da quantidade de acessos, interação com materiais da disciplina, interações com vídeo-aulas e o número total de acesso ao fórum.

Nos resultados, a discussão levantada pelos autores foi a de que se tornou possível obter limiares que tem como intuito realizar a previsão dos índices de aprovação dos alunos. A contribuição da pesquisa é de extrema relevância, ao mesmo tempo que desperta a perspectiva para um novo desafio que é o de entender como a utilização de dados dos alunos, coletados de ambientes colaborativos e analisados com técnicas propostas pela EDM poderia contribuir para a representação adequada de um fenômeno cognitivo ou metacognitivo de maneira clara e segura? Sabe-se que o comportamento dos alunos replica padrões subjetivos, complexos a serem descritos, mas, como os dados de interação, típicos do seu comportamento em uma plataforma de aprendizagem colaborativa, poderiam ser associados ao fenômeno da autoavaliação para melhorar efetivamente o processo da aprendizagem?

Rodrigues et al., (2014) cita em um estudo acerca do avanço das pesquisas da EDM no Brasil que a dificuldade de entender os dados está limitada anteriormente ao entendimento do problema, para viabilizar a escolha e uso das técnicas adequadas da EDM. Portanto, a proposição de um framework, software ou sistema computacional que estabelecesse a capacidade de representar um cenário bem descrito, em qual abordagem da aprendizagem a autoavaliação seria realizada, quem seriam os participantes (professores, tutores ou alunos) através de um fluxo decisório sistematizado e aliado aos mecanismos que processam os dados de interação do aluno em ambientes colaborativos poderia recomendar melhorias significativas na aprendizagem?

Estas, e outras muitas perguntas sem respostas podem ser entendidas como grandes desafios a serem discutidos em uma visão científica quando estiverem sendo abordadas de maneira empírica as relações existentes da autoavaliação da aprendizagem com a Mineração de Dados Educacionais. As tecnologias alinhadas aos fatores humanos que geram impactos na educação precisam ser expandidas, a partir do momento em que são detectadas novas necessidades para uma sociedade que muda constantemente sua forma de pensar, agir e aprender.

5. Considerações Finais

Este trabalho teve como objetivo apresentar de maneira contextualizada os aspectos fundamentais do fenômeno metacognitivo da autoavaliação, as abordagens da aprendizagem colaborativa em que o mesmo pode ser expandido, além de algumas técnicas da mineração de dados que podem ser implementadas para representar cenários de autoavaliação da aprendizagem através da análise e tratamento dos dados de interação de professores, tutores ou alunos.

Dentre as limitações encontradas pelo autor estão a busca por trabalhos que abordassem o fenômeno da autoavaliação alinhado a mineração de dados educacionais, a apresentação de processos ou métodos que repliquem esta integração de maneira efetiva e a discussão acerca de tecnologias que implementaram soluções com seus respectivos resultados. Em virtude de ser uma temática recente, entende-se que a literatura que aborda esta temática ainda é dispersa, com trabalhos pontuais que abrangem de maneira genérica a autoavaliação da aprendizagem com uso da tecnologia.

Esperam-se como trabalhos futuros a realização de duas pesquisas: uma revisão sistemática literatura que replique todas as evidências que comprovem a necessidade em se investir na pesquisa deste tema, além do desenvolvimento uma solução prática através de um método empírico com um aprofundamento acerca desta problemática tratada no artigo, o levantamento de eventuais hipóteses, uma proposta de solução e a avaliação dos resultados desta integração.

Referências

- BARROS, Daniela V.; **Estilos de uso do espaço virtual: como se aprende e ensina no virtual**. Universidade Federal do Goiás. Revista Inter-ação, edição 34, 2009, p. 51-74.
- BOUD, David; FALCHIKOV, Nancy; **Rethinking Assessment In High Education**. Routledge, 2007
- BRUSILOVSKY, Peter; SOSNOVSKY, Sergey. **Individualized Exercises for Self-Assessment of Programming Knowledge: An Evaluation of QuizPACK**. School of Information Sciences, University of Pittsburgh, 2005.
- CHI, Cheng Man Chi. **Self assessment in the School-based Assessment speaking component**. Dissertation Secondary of Univesity of Hong Kong, 2009.
- DELIALIOGLU, Omer; YILDIRIM, Zahide. **Students' Perceptions on Effective Dimensions of Interactive Learning in a Blended Learning Environment**. International Journal of Educational Technology & Society, vol. 10, 2007, p. 133-146.
- ECONOMIDES, Anastasios; **Adaptive context-aware pervasive and ubiquitous learning**. Computers & Education Volume 57, Issue 4, December 2011, p.2517-2534.

- FREITAS, Luis F. C.; FERREIRA, Salete L.; **Evolução da EAD no Brasil: um comparativo no ensino superior a distância x presencial**. Arquivo da Associação Brasileira de Ensino a Distância (ABED), Rio de Janeiro, 2013.
- GUZMAN, Eduardo; CONEJO, Ricardo; DE-LA-CRUZ, José Luis Pérez; **Improving Student Performance Using Self-Assessment Tests**. Published by the IEEE Computer Society. Issue Nº. 04, vol. 22, 2007, p. 46-52.
- HATZIAPOSTOLOU, Thanos; PARASKAKIS, Iraklis. **Enhancing the Impact of Formative Feedback on Student Learning Through an Online Feedback System**. International Faculty of the University of Sheffield, Thessaloniki, Greece, 2010.
- HEINER, Cecily; HEFFERNAN, Neil; BARNES, Tiffany; **Education Data Mining**. Supplementary Proceedings of the 13th International Conference of Artificial Intelligence in Education, Marina Del Rey, CA, USA, 2007.
- KELVIN, Tan Heng Kiat. **Student Self-Assessment: Assessment, Learning and Empowerment**. Research Publishing, 2012.
- NUNES, V. B.; NOBRE, I.; **O professor tutor em um curso a distância: avaliação, problemas e busca de soluções**. Anais do Congresso do Assoc. Brasileira de EAD, 2013.
- PEREIRA, José; PLEGUEZUELOS, Eulogio; MERI, Alex; MOLINA-ROS, Antoni; MOLINA-TOMAS, Carmen; MASDEU, Carlos. **Effectiveness of using blended learning strategies for teaching and learning human anatomy**. Article in Medical Education, 2007, p. 189-195.
- RAMOS, Jorge L. C.; RODRIGUES, Rodrigo L.; SILVA, João C. S.; GOMES, Alex S. **Analisando Fatores que Afetam o Desempenho de Estudantes Iniciantes em um Curso a Distância**. Anais do Simpósio Brasileiro de Informática na Educação. 2014, p. 99-108.
- RODRIGUES, Rodrigo L.; RAMOS, Jorge L. Cavalcante; SILVA, João Carlos Sedraz; GOMES, Alex Sandro; **A literatura brasileira sobre Mineração de Dados Educacionais**. Anais do III Congresso Brasileiro de Informática na Educação. 2014. p. 621-630.
- SANTANA, Leandro; MACIEL, Alexandre M.A.; RODRIGUES, Rodrigo Lins; **Avaliação dos perfis de Uso no Ambiente Moodle Utilizando Técnicas de Mineração de Dados**. Anais do Simpósio Brasileiro de Informática na Educação. 2014. p. 269-277
- SANTOS, Patricia; PÉREZ-SANAGUSTÍN, M.; HERNÁNDEZ-LEO, D.; BLAT, J.; **QuesTInSitu: from tests to routes for assessment in situ activities**. Dep. of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, 2011.
- SCARTEZINI, R. **O desenvolvimento da identidade profissional docente de professores universitários: um estudo dialógico**. Tese de Doutorado. UnB, Brasília, 2014.
- SILVA, Marjorie Cristina Rocha; NUNES, Maiana Farias Oliveira; SOUZA, Mayra Silva. **Análise da produção científica em painéis dos congressos brasileiros de avaliação psicológica**. Anais do Congresso Brasileiro de Avaliação Psicológica, 2007, p. 239-252
- SOUZA, Bruno M. **Ambiente de aprendizado ubíquo youubi**. Design e avaliação. Tese de Doutorado do Centro de Informática da Universidade Federal de Pernambuco, 2015.

Fala, leitura e escrita em aplicativos móveis: desafios de uma abordagem discursiva

Alline Mayumi Ribeiro Kobayashi¹, Fernanda Freire²

¹Mestranda do Instituto de Estudos da Linguagem – Universidade Estadual de Campinas (UNICAMP)

²Núcleo de Informática Aplicada à Educação – Universidade Estadual de Campinas (UNICAMP)

allinekobayashi@gmail.com, ffreire@unicamp.br

Abstract. *This article reflects on the use of technology in the acquisition of reading and writing, through the analysis of mobile apps developed for this purpose and presents *Língua de Gato*, application for iPad based on theoretical and methodological principles of Discursive Neurolinguistics*

Resumo. *Este artigo apresenta reflexões acerca do uso de tecnologia no processo de aquisição da leitura e da escrita, através da análise de aplicativos móveis voltados para este fim e da apresentação do *Língua de Gato*, aplicativo para iPad desenvolvido a partir de pressupostos teóricos e metodológicos da Neurolinguística Discursiva (ND).*

1. Introdução

Aprender a ler e escrever implica o uso de um sistema que se organiza em função de um repertório próprio da língua - de sons e de letras – que recursivamente origina um conjunto de recursos expressivos (palavras) que podem ser combinados de múltiplas maneiras em função de restrições de, pelo menos, dois tipos: (i) aquelas ditadas pelo próprio sistema, (ii) as que resultam da relação desse sistema com outros sistemas não verbais: processos psíquicos, perceptivos e corporais [FREIRE, 2005].

Assim, a escrita convencional do Português Brasileiro (PB) pressupõe compreender o funcionamento do sistema de escrita alfabética: (i) quais letras compõem o repertório da língua (o alfabeto) e em quais sequências elas podem aparecer (*q* é sempre seguido de *u*, por exemplo); (ii) quais letras se relacionam a um único som da língua (*b* é /b/ e não /d/, por exemplo) e quais delas podem representar mais de um som (*s* pode ser /s/ e /z/, por exemplo); (iii) que existe uma ampla diversidade de grafias usadas em diferentes tipos de materiais impressos ou não (as caligrafias de diferentes pessoas, a letra de imprensa, a letra bastão, *etc.*). Implica ainda aprender a representação gráfica das letras, a direção (da esquerda para a direita) e a orientação (de cima para baixo) mais usuais com que são registradas à mão [FREIRE, 2011].

Todo esse processo convoca um conjunto de conhecimentos que é posto em ação de maneira coordenada, o que exige o trabalho de vários subsistemas cerebrais

responsáveis pela atenção, percepção visual e espacial, gestos e linguagem. À medida que a criança aprende a ler e a escrever de forma significativa e contextualizada, isto é, vivenciando práticas de linguagem da vida em sociedade, pouco a pouco, associações de natureza visual, acústica e cinestésica se integram e os processos iniciais de codificação de som/letra (para escrever) e de decodificação letra/som (para ler) se automatizam [LURIA, 1981]. Em outras palavras, “a experiência com a escrita possibilita ao escrevente que ele não pense mais na letra ou no movimento das mãos para escrever uma palavra. Nesse sentido, os processos de leitura e de escrita sofrem o apagamento da representação geométrica da letra; como na fala, apaga-se a propriocepção da produção dos sons. Lemos e escrevemos sentidos” [BORDIN, 2010, p. 43].

A complexidade desse aprendizado, no entanto, nem sempre é (re)conhecida pelo adulto ou pela escola que muitas vezes consideram “que as crianças, apenas por estarem na escola, irão naturalmente aprender a ler e a escrever” [BORDIN, 2010, p. 56]. Os professores, quase sempre respaldados por um método de alfabetização, oferecem às crianças um único caminho para a leitura/escrita, desconsiderando a variedade que existe entre elas, o que demanda outras possibilidades de *mediação* (material e humana) para este aprendizado [VYGOTSKY, 1984].

Com base nessas considerações, foram analisados alguns aplicativos voltados para a alfabetização disponíveis gratuitamente para iPad e foi desenvolvido o aplicativo *Língua de Gato*. Seu desenvolvimento impõe, em primeiro lugar, o desafio de aliar uma concepção sociocultural de língua (falada, lida e escrita) às possibilidades tecnológicas oferecidas por dispositivos móveis (figuras, fotografias, áudios, *etc.*), considerando as singularidades das crianças que com ele interagem, isto é, seus interesses e as relações que (já) são capazes de estabelecer com o sistema alfabético de escrita. Objetiva-se com o seu uso potencializar a *mediação coadjuvante* (que o aplicativo/dispositivo pode exercer) que pode ser ampliada pela *mediação do outro* (criança e ou adulto) quando participa da brincadeira da criança com o aplicativo, interferindo, potencialmente, na qualidade da relação da criança com a leitura e a escrita.

2. O paradigma da Neurolinguística Discursiva e o uso discursivo de tecnologias

A Neurolinguística Discursiva (ND), iniciada na década de 80 no Instituto de Estudos da Linguagem da Universidade Estadual de Campinas com o trabalho de Coudry [1986], tem como objeto de estudo a *linguagem em funcionamento* (incluindo a fala, a leitura e a escrita) e sua relação com outras atividades cognitivas (atenção, percepção, memória, raciocínio) em meio à vida em sociedade. Assume um posto de observação predominantemente linguístico [COUDRY e FREIRE, 2010] e sua teorização resulta da articulação de autores que consideram a linguagem, o cérebro e o sujeito como construções socioculturais, portanto, heterogêneos e não constituídos *a priori*.

A qualificação discursiva reafirma o modo como a linguagem se apresenta: polissêmica e indeterminada por um lado [FRANCHI, 1977/1992], e determinada ideológica e historicamente, por outro [BAKHTIN, 1929/1999]. Essa condição só se dá a conhecer por meio de práticas de uso social de linguagem, razão pela qual a *interlocução* é um conceito fundamental em seus estudos [COUDRY e FREIRE, 2010].

Para sustentar a natureza complexa do funcionamento das línguas naturais, a ND incorpora um modelo de cérebro igualmente complexo com base em conceitos e postulados encontrados em Vygotsky [1934/1984] e Luria [1981], bem como nos estudos neurofisiológicos de Freud [1891/2013].

A *palavra*, considerada *unidade funcional da linguagem* [FREUD, 1891/2013], é um complexo associativo de elementos auditivos, visuais e cinestésicos. De acordo com Freud, a *representação da palavra* é formada a partir da relação entre a imagem de som (registro da palavra ouvida), a imagem de movimento da fala (gestos articulatórios para se pronunciar tal palavra) e a imagem de leitura (imagem visual da palavra escrita). Esta, por sua vez, relaciona-se à imagem de escrita (movimentos das mãos para se escrever). A partir da imagem do som, todo o esquema da representação de palavra se relaciona a um conjunto de representações de várias naturezas (visuais, táteis, acústicas, etc.) que dão significado à palavra ouvida/lida/escrita. Destaca-se neste modelo a centralidade do elemento sonoro nos processos que envolvem a linguagem.

A respeito da leitura/escrita Coudry [2010], com base na relação entre o “velho” e o “novo” postulada por Freud [1891/2013], argumenta que no início da aquisição da leitura e da escrita, a criança se encontra mais na fala (velho) e na leitura (novo) do que na escrita (novo), sendo que a sua variedade de fala pode aproximá-la ou distanciá-la da leitura, considerando que a fala padrão é mais próxima da escrita convencional. Por isso, ler o que os outros escrevem é tão importante para adquirir o sistema da escrita quanto é aprender a falar com a linguagem dos outros, como Freud afirma, servindo como um elo entre o velho (fala) e o novo (escrita).

Segundo a ND, o eixo fala, leitura e escrita demanda uma relação de concomitância: quando a criança aprende a falar repetindo a fala do adulto estabelece uma forte associação entre o acústico (o que ouve: imagem de som) e o motor (gestos articulatórios: imagem de movimento) que ocorrem concomitantemente na tentativa de aproximar a sua fala da fala do outro e, assim, se automatiza, o que pressupõe estar na relação com o outro. É pelo sentido (repetição/recordação do motor e do acústico e suas possíveis combinações) que a criança “entra” na língua de sua cultura onde funcionam e se articulam as dimensões fonológica, sintática, semântica, pragmática.

De forma análoga, quando a criança ganha autonomia em relação à leitura ela lê sentidos e não mais letras ou sucessão de letras que, até então, eram laboriosamente associadas a sons possíveis e que se associavam, por sua vez, a palavras velhas ou novas. É isso o que se observa na leitura silenciosa, que pressupõe a automatização da relação entre o acústico (imagem de som) e o visual (imagem de leitura), dispensando o apoio da voz (imagem de movimento) ou da mão (imagem de escrita).

Em relação ao uso de tecnologias de informação e comunicação (TICs) a ND concebe que nesse tipo de práticas (digitais) de uso social da linguagem estabelecem-se dois tipos de interação relacionados: a interação entre a criança com o computador/dispositivo mediada pelos recursos verbais e não verbais disponíveis na interface da aplicação utilizada e a interação entre a criança e um outro (adulto ou criança) mediada pela língua [FREIRE, 1999/2006].

O computador/dispositivo, sensível às ações da criança, responde com *feedbacks* que assumem a função de uma *mediação coadjuvante* cuja interpretação pela criança pode ser beneficiada pela mediação do outro [FREIRE, 2012]. A mediação coadjuvante, potencialmente, oferece indícios que ajudam a criança a se manter na interação com a aplicação e a sua qualidade/eficiência depende não só de uma boa usabilidade [NIELSEN, 1993] como também da sua pertinência no contexto da aplicação.

Esse foi um dos desafios do desenvolvimento do *Língua de Gato*: oferecer *feedbacks* que promovem uma interação dinâmica entre a criança e o aplicativo e que provocam uma reflexão da criança sobre a língua, em especial, sobre os elementos visuais e acústicos que compõem a palavra.

3. O paradigma dominante: aplicativos para leitura e escrita

Como parte das práticas de linguagem desenvolvidas no Centro de Convivência de Linguagens do Instituto de Estudos da Linguagem da Universidade Estadual de Campinas junto a crianças e jovem que apresentam dificuldades escolares relativas ao aprendizado da leitura e escrita, realizam-se diversas atividades envolvendo leitura e escrita mediadas pelo uso de computadores. Observa-se que essas crianças têm uma grande motivação para usar a tecnologia e apresentam desenvoltura para acessar a internet, assistir vídeos, jogar, conversar por meio de chats, pesquisar, participar de redes sociais. No entanto, evitam contextos em que é preciso escrever por não saberem como escrever o que desejam. RM, por exemplo, um menino de 8a9m de idade que frequenta o 3º ano de uma escola pública, sabe qual é a função de um site de busca, sabe acessar o *Google* e como funciona, sabe dizer qual palavra-chave deve ser escrita, mas não consegue escrevê-la [FREIRE, 2011].

Diante desse cenário, buscou-se por diversos aplicativos para *tablets* voltados para a aquisição e uso da leitura e da escrita, especialmente aplicativos disponíveis gratuitamente na App Store para iPad, dispositivo que se tinha à disposição na ocasião. Observou-se, então, que a maior parte deles utiliza uma abordagem semelhante à usada em cartilhas de alfabetização, apresentando uma visão da leitura e da escrita que se restringe à codificação/decodificação e memorização. Silva [1998] analisou na década de 90 diferentes softwares que visavam auxiliar no processo de alfabetização voltados para crianças no período pré-escolar e destacou características semelhantes: “o que se viu nos softwares destinados à alfabetização é que se limitam a trazer para a criança uma ‘leitura’ escolar, desvinculada do contexto e da realidade infantil e que se voltam, basicamente, para a decodificação de signos através da memorização” [SILVA, 1998, p.119].

Com base em Cagiliari [2000], pode-se levantar, ao menos, duas questões importantes e relacionadas entre si que são desconsideradas pelos aplicativos analisados: (i) que a criança sabe falar o português, embora a variedade de sua fala possa não corresponder à variedade padrão ou culta e é capaz de refletir sobre a sua própria língua; (ii) que a fala e a escrita diferem entre si, ainda que a escola e esses aplicativos se apoiem em uma fala artificial na tentativa de aproximar a criança da escrita padrão. Observa-se, assim, uma abordagem que pouco considera o usuário final

como um sujeito já inserido em um universo letrado, que é um falante da língua e, por isso, é capaz de refletir sobre ela.

O que se sabe sobre o usuário desses aplicativos? Apenas que são crianças, a julgar pelo design infantil das suas telas. Tal abordagem não abre espaço para que os conhecimentos prévios da criança sobre o universo da leitura e da escrita sejam expostos e explorados, limitando as infinitas possibilidades de criação no mundo das letras a um conjunto limitado de palavras (A de Avião, B de Bola, *etc.*) muitas vezes desinteressante para a criança. Observam-se, assim, atividades “engessadas”, pouco flexíveis, ainda comuns em alguns ambientes escolares, reproduzidas na tela do tablet. Esta abordagem se mostra incondizente não só com a natureza da linguagem, como também com as crianças da atual geração, *nativas digitais* [PRENSKY, 2001], e com a dinamicidade dos vários recursos oferecidos pelos dispositivos móveis (internet, gravação de áudio, vídeo, tecla *touch*, *etc.*).

A este respeito, Silva [op. cit.] afirma que a característica multimídia da tecnologia (movimentos, imagens, sons e cores) dá ao software uma ilusão de novidade e que “o computador poderá ser o ‘novo’ inovador que suscitará discussões importantes para o avanço da compreensão, do desenvolvimento e da aprendizagem das crianças sob vários aspectos, ou poderá ser o velho vestindo uma roupa nova que mascara aspectos já superados no que concerne à alfabetização. Aspectos estes tão arduamente, penosamente, rigorosamente pesquisados através de várias áreas como a linguística, a psicologia, a pedagogia, a psicopedagogia, a história, *etc.*” [SILVA, 1998, p. 101].

Assim, a partir do referencial teórico da ND, surge a questão: como desenvolver softwares voltados para a aquisição e uso da fala, da leitura e da escrita que não sejam apenas o “velho em uma roupa nova”? O desenvolvimento de softwares para a aquisição e uso da linguagem a partir de uma perspectiva discursiva apresenta-se como um desafio a ser debatido, tanto na área de computação, quanto nas áreas relacionadas à linguagem, tais como linguística, pedagogia, psicologia e fonoaudiologia. Do ponto de vista da educação, existe o desafio de inserir a tecnologia na sala de aula de maneira a enriquecer, de fato, a prática escolar e, do ponto de vista computacional, o desafio está em desenvolver softwares para além do padrão cartilhesco e do senso comum a respeito da língua e da linguagem.

Silva [op. cit.], na conclusão do seu trabalho, apresenta alguns possíveis caminhos para a utilização da tecnologia como auxiliar no processo de aquisição de leitura e escrita de maneira realmente inovadora. Um deles é a aproximação do professor dos conhecimentos tecnológicos e a escolha de softwares apropriados para o que se deseja e outro é a formação de equipes multidisciplinares no desenvolvimento de softwares. O que se pretende com este artigo, então, é dar um primeiro passo nesse caminho através do debate interdisciplinar entre a Linguística e a Computação.

4. O Língua de Gato

O Língua de Gato é um aplicativo de criação de jogos de palavras cruzadas para tablets que tem como público alvo crianças em fase de aquisição de leitura e de escrita. As palavras cruzadas são um passatempo clássico em nossa cultura e representam uma prática social de uso de linguagem conhecida de boa parte das crianças. Considerando-

se o fato de os dispositivos móveis serem ricos em possibilidades de interação e permitirem a captura de fotos, vídeos, áudio e acesso à internet, bem como o fato de as crianças gostarem de tecnologia e apresentarem facilidade em seu manuseio, procuramos otimizar esse potencial com vistas a oferecer uma mediação coadjuvante pertinente e eficaz no contexto de leitura e escrita.

Uma criança de 6 anos que vive no meio urbano assiste tv, joga joguinhos no celular dos pais, vai ao shopping. Por que ela gostaria de aprender a ler e a escrever? Por viver em mundo letrado, a escrita faz parte do seu universo: está presente no comercial de tv, no celular, nas revistas, no cardápio do restaurante. Cagliari afirma que “não basta saber escrever, para escrever. É preciso uma motivação para isso” [2000, p.102]. A escrita tem uma função social e é partir desta função, percebida pouco a pouco pela criança com base nas práticas sociais que vivencia e/ou que presencia outros vivenciando, que surge a sua motivação para aprender a ler e a escrever. A criança quer aprender a escrever para escrever seu próprio nome em seu material escolar, ajudar a sua mãe a fazer a lista de compras ou para usar o celular dos pais da mesma forma que eles o fazem. Ela quer ler para compreender o que aparece registrado na tela enquanto joga vídeo game, o que está escrito nas prateleiras do supermercado ou no letreiro do ônibus. Ler e escrever, portanto, tem a ver com o mundo a nossa volta e é nessa premissa que o Língua de Gato se assenta ao possibilitar a criação de palavras cruzadas com elementos presentes no universo da criança.

O aplicativo foi disponibilizado para *download* gratuito na App Store em fevereiro de 2016. Sua interação pressupõe dois papéis: o do usuário responsável por criar o jogo de palavras cruzadas (mediador) inserindo um conjunto de palavras e dicas e o do usuário que vai jogar as palavras cruzadas (criança em fase de aquisição de leitura e escrita). Via de regra, o mediador pode ser o professor ou os pais da criança, mas nada impede que o mediador seja a própria criança que cria para si um jogo, para uma outra criança ou, mesmo, para um adulto. Essa troca de papéis, inclusive, pode ser bastante interessante em termos de exploração da língua e dos recursos do tablet.

A versão atual do aplicativo, v. 1.0.1, apresenta 3 possibilidades de ação (Figura 1): criar um jogo de palavras cruzadas, jogar uma palavra-cruzada gerada aleatoriamente a partir de palavras inseridas pelo mediador e jogar um jogo salvo.



Figura 1. Telas do Língua de Gato, da esquerda para a direita: tela de criação do jogo, tela do jogo e *feedback* do aplicativo quando a criança seleciona uma letra errada.

Cada palavra inserida para a criação de um jogo possui 2 tipos de “dicas”: visuais e/ou auditivas. Por exemplo, ao adicionar a palavra *sorvete* pode-se inserir uma dica visual tirando uma foto de um sorvete ou selecionando uma foto a partir da galeria

de fotos do dispositivo e pode-se também gravar o áudio do mediador dizendo a palavra *sorvete*. A inclusão da dica auditiva tem o propósito de relacionar a fala e a escrita, uma vez que o que se diz não é o exatamente o que se escreve. Tal fato é um complicador para as crianças em fase inicial de aquisição da escrita. Inseridas as palavras e suas respectivas dicas na tela de criação de jogo, o mediador pode salvar este jogo para torná-lo acessível posteriormente ou apenas jogá-lo naquele momento.

A tela de jogo é formada pelas cruzadas e pela imagem de cada dica. Quando uma dica é clicada, a imagem é ampliada e, caso possua uma dica auditiva, o som do áudio é reproduzido. A disposição das palavras na cruzada é gerada aleatoriamente toda vez que a tela do jogo é carregada.

Para completar a cruzada a criança deve arrastar a letras que aparecem distribuídas pelo “tabuleiro” do jogo. Nesta primeira versão do aplicativo todo o conjunto de letras disponível é utilizado para completar o jogo. Ao arrastar uma letra para uma determinada posição da cruzada o dispositivo oferece *feedbacks* visual (escreve em vermelho, como mostra a Figura 1) e auditivo para indicar se a letra foi ou não colocada na posição correta. A interface do aplicativo tem algumas especificidades quando comparado ao de outros jogos de palavras cruzadas, como mostra a Figura 2.

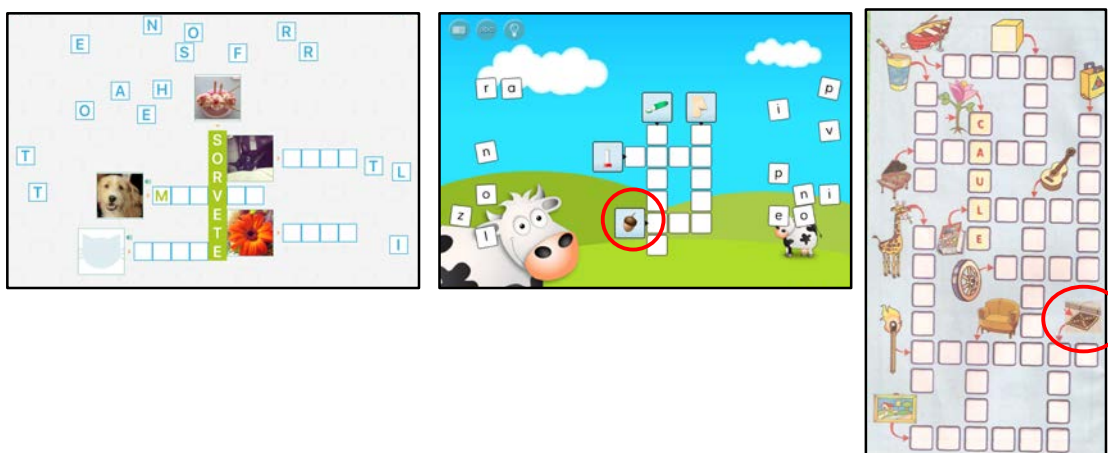


Figura 2. Da esquerda para a direita: aplicativos Língua de Gato e Happi Spells (iPad) e palavras cruzadas em papel (Revista Picolé n.35, Coquetel).

A primeira característica a ser destacada, já mencionada anteriormente, é o fato de as dicas visuais do Língua de Gato serem fotos de elementos do universo da criança, o que confere ao jogo um caráter pessoal, especial, podendo afetar positivamente a motivação da criança. As dicas visuais, por exemplo, podem ser fotografias de seus amigos ou de sua família. Além disso, pelo fato de as dicas visuais serem fotografias e não desenhos, mantém-se as características reais do que se pretende representar, evitando, assim, ambiguidades. Veja-se, por exemplo, na Figura 2, como é difícil perceber visualmente a “noz” do jogo Happi Spells ou o ralo da Revista Coquetel (destacadas com um círculo).

Além disso, pelo fato de as palavras serem customizáveis, pode-se evitar a presença de imagens/palavras pouco familiares à criança que poderiam funcionar como uma dificuldade não desejável para os iniciantes na escrita. A maioria das crianças

brasileiras, por exemplo, pode não identificar o desenho da “noz” como tal, tanto pelo traçado do desenho quanto pelo fato de não conhecer a palavra. Isto não quer dizer que o jogo deva conter apenas palavras já conhecidas pela criança, mas sim que cabe ao mediador determinar quais imagens e palavras são interessantes para determinada criança. Quando se deseja, por exemplo, expandir o vocabulário, pode-se inserir fotografias de objetos que a criança ainda não conhece. Neste caso, está em relação o *velho* e o *novo*: quando a criança já conhece a palavra pela fala, mas não conhece a sua imagem ou, ao contrário, quando a criança conhece a imagem, mas não sabe o seu nome: o velho em cada um dos dois casos pode facilitar o aprendizado do novo.

Outro aspecto a ser destacado na interface do Língua de Gato e também presente no aplicativo Happi Spells é a presença da imagem acústica das palavras, ausente na versão em papel. Conforme detalhado na seção 2, o processo de escrita tem como base a fala e, para escrever, deve existir um vai e vem de relações entre o acústico, o visual e o motor. Ao jogar qualquer um dos jogos da Figura 2 a criança olha a imagem da palavra a ser escrita, identifica-a e diz (nem que seja mentalmente) o que escrever. Quando a dica acústica está presente a criança pode comparar a sua fala com a fala do outro, o que pode gerar uma reflexão importante para os aprendizes sobre diferentes pronúncias, mais ou menos próximas da escrita convencional. Por exemplo, a criança pode falar *pipinu*, mas escreve-se *pepino*.

Uma vez que a criança já sabe o que deverá escrever, no jogo tradicional em papel, ela passa a refletir sobre a relação som/letra e, então, com base na memória que tem das letras do alfabeto, deve selecionar qual letra deve escrever (imagem visual) e, então, como escrevê-la (imagem cinestésica da escrita no papel). Caso sobre ou falte quadradinhos no jogo, a criança perceberá que não fez as seleções adequadas, seja da palavra, seja das letras que a compõem. No Língua de Gato (e também no Happi Spells) este processo torna-se mais simples, uma vez que a representação visual das letras já está dada: ela não precisa resgatar mentalmente qual o formato de determinada letra e como traçá-la, mas deve identificar a letra desejada dentre as letras disponíveis na tela e arrastá-la até a posição desejada.

O *feedback* oferecido pelo aplicativo, caso mova uma letra para uma posição inadequada, difere nos dois aplicativos. No Língua de Gato é possível que a criança mova uma peça para um local inadequado, com o intuito de fazê-la reler/rever o que está tentando escrever. Neste caso, a letra assumirá a cor vermelha, diferente do padrão que é verde, como um indicativo de que há alguma coisa a ser revista. No Happi Spells o movimento de uma letra “errada” não é possível: a criança tenta posicioná-la no jogo, mas ela retorna ao seu lugar de origem na tela. Diferente da abordagem comumente presente no ambiente escolar e nos aplicativos destinados à escrita/leitura, o erro no Língua de Gato não é visto como problema, mas, sim, como uma possibilidade de reflexão sobre as hipóteses possíveis para a escrita de determinada palavra. Para a ND, as tentativas e hesitações da criança que está aprendendo a ler e a escrever, denominadas de atividades epilinguísticas [COUDRY, 1986], são importantes para o aprendizado e, para tal, a criança precisa se colocar na posição de leitora de sua própria produção com o objetivo de refletir sobre ela e corrigi-la.

4.1 Desenvolvimentos futuros

Atualmente, o *Língua de Gato* está em fase de testes com usuários para coletar experiências acerca da interação, interface e jogabilidade. A coleta dessas informações busca não apenas reavaliar tais pontos, mas também construir uma estratégia de jogabilidade, atualmente inexistente, para que ele se torne mais interessante e atraente.

Para possibilitar o trabalho com a linguagem da maneira mais abrangente possível e explorar os recursos tecnológicos disponíveis, algumas funcionalidades serão incluídas. Uma delas é dar à criança um *feedback* sonoro diferenciado da palavra escrita, isto é, a sua leitura automática. É importante que a criança possa comparar a leitura do aplicativo com a imagem de som que ela pretendeu escrever. Encontra-se aqui uma questão técnica a ser estudada: a leitura automática da palavra não poderá ser mecânica a ponto de interferir na prosódia e, com isso, na compreensão de seu sentido.

A respeito da tríade sonoro/visual/motor presente na leitura e escrita, seria desejável a inclusão de atividades que explorem também o aspecto motor da escrita. Serão estudadas possibilidades de desenvolvimento de tais funcionalidades, porém ainda não se sabe se tais funcionalidades serão incluídas dentro do jogo de palavras cruzadas ou em outra modalidade de jogo.

Pretende-se, também, dar ao mediador maior poder de customização sobre o aplicativo a partir de uma tela de configuração, atualmente inexistente, na qual ele poderá escolher entre diferentes tipos de fontes para as letras do jogo e se elas serão maiúsculas ou minúsculas. Desta maneira, o mediador poderá determinar a imagem visual que mais se adequa ao que se deseja trabalhar. Outra possibilidade será a escolha do número de letras do alfabeto disponíveis para se concluir o jogo.

5. Considerações finais

Com base em uma concepção ampla de linguagem proposta pela ND que considera tanto seu aspecto interno, constituinte e construtor do sujeito [FRANCHI, 1977/1992], quanto externo, na relação do sujeito com o outro e com o mundo, inseridos num contexto sócio histórico, buscou-se, por um lado, identificar as abordagens sobre leitura e escrita utilizadas por boa parte dos aplicativos disponíveis para iPad e, por outro, apresentar os desafios (e as razões) envolvidos no desenvolvimento do aplicativo *Língua de Gato*.

A utilização de diferentes aplicativos voltados para a aquisição da leitura e da escrita revelou uma visão de linguagem restrita que desconsidera o fato de a criança ser um falante da língua e seu conhecimento como tal; desconsidera, ainda, o fato de a leitura e a escrita serem práticas sociais cujo aprendizado está diretamente relacionado à interação da criança com o mundo a sua volta.

Quando se considera a linguagem como decorrente da relação do sujeito com o mundo e com o outro, o desenvolvimento de um aplicativo voltado para o processo de aquisição da leitura e escrita precisa considerar o usuário final como um sujeito único, com motivações específicas para entrar no mundo das letras e que trilha um caminho também único até tornar-se um leitor/escrevente autônomo. Desta maneira, o maior desafio no desenvolvimento de um aplicativo para este fim é possibilitar que o sujeito ali se expresse e se veja, ou seja, que o aplicativo reflita a relação singular do sujeito

com a linguagem. Para tal, deve-se compreender que o aplicativo não será uma solução que irá alfabetizar alguém, mas sim mais um *mediador* nesse processo que necessita da interação da criança com o mundo e com o outro.

Referências

- Bakhtin, M. (1929/1999). *Marxismo e Filosofia da Linguagem*, Hucitec, São Paulo.
- Bordin, S. S. B. (2010). *Fala, Leitura e Escrita: encontro entre sujeitos*, *Tese de Doutorado*, Universidade Estadual de Campinas. Campinas.
- Cagliari, L. (2000). *Alfabetização e Linguística*, Editora Scipione, São Paulo.
- Coudry, M. I. H. (1986). *Diário de Narciso: discurso e afasia*. *Tese de Doutorado*, Universidade Estadual de Campinas. Campinas.
- Coudry, M. I. H. (2010). Caminhos da Neurolinguística Discursiva: o velho e novo. In: Coudry, M. I. H.; Freire, F. M. P.; Andrade, M. L.; SILVA, M. A. (orgs.). *Caminhos da Neurolinguística Discursiva: teorização e práticas com a linguagem*, p. 23-48, Mercado de Letras, Campinas.
- Coudry, M. I. H. e Freire, F. M. P. (2010). Pressupostos teórico-clínicos da Neurolinguística Discursiva (ND). In: Coudry, M. I. H.; Freire, F. M. P.; Andrade, M. L.; SILVA, M. A. (orgs.). *Caminhos da Neurolinguística Discursiva: teorização e práticas com a linguagem*, p. 23-48, Mercado de Letras, Campinas.
- Franchi, C. (1977/1992). Linguagem – Atividade Constitutiva. In: *Cadernos de Estudos Lingüísticos*, p. 9-39, v. 22, Campinas.
- Freire, F. M. P. (2005) *Agenda Mágica: linguagem e memória*, *Tese de Doutorado*, Universidade Estadual de Campinas. Campinas.
- Freire, F. M. P. (2006) *Enunciação e Discurso: a linguagem de programação Logo no discurso do afásico*. Mercado de Letras (1999), Campinas.
- Freire, F. M. P. (2011). Práticas digitais informais e leitura/escrita formais. In: *Anais do XVI Congresso Internacional de la Asociación de Lingüística y Filología de la América Latina (ALFAL)*, p. 1-10, Alcalá de Henares.
- Freire, F. M. P. (2012) *Escola e tecnologia: um olhar discursivo sobre essa complexa relação*. In: *Livro de Atas do II Congresso Internacional TIC e Educação*, v. 1, p. 513-533, Lisboa.
- Freud, S. (1891/2013). *Sobre a concepção das afasias: um estudo crítico*. Tradução de Emiliano de Brito Rossi, Editora Autêntica, Belo Horizonte.
- Luria, A. R. (1981). *Fundamentos de Neuropsicologia*. Cultrix, São Paulo.
- Nielsen, J. (1993). *Usability Engineering*. Academic Press, Cambridge.
- Prensky, M. (2001). *Digital Natives, Digital Immigrants*, MCB University Press, Bradford.
- Silva, M. R. G. (1998). *O computador e a alfabetização: estudo das concepções subjacentes nos softwares para a Educação Infantil*, *Dissertação de Mestrado*, Universidade Federal de Santa Catarina. Florianópolis.
- Vygotsky, L. S. (1934/1984). *A Formação Social da Mente*, Martins Fontes, São Paulo.

ENCompIF - Encontro Nacional de Computação dos Institutos Federais

Apresentação

O ENCompIF tem por objetivo reunir professores e alunos de ensino médio dos Institutos Federais e CEFETs de todo o país, que poderão apresentar pesquisas desenvolvidas em suas instituições. Na primeira edição do evento recebemos um grande número de artigos, 88, dos quais 22 foram aceitos. Já na segunda edição do evento recebemos 128 artigos, dos quais 21 foram aceitos. Nesta edição, recebemos 132 submissões, dos quais 20 artigos foram aceitos. Além da apresentação de trabalhos, o evento tem se mostrado como uma excelente oportunidade para a troca de experiências entre os participantes e um estímulo a integração e estabelecimento de parcerias entre professores e alunos e diferentes Institutos Federais e CEFETs.

Carlos Roberto de Oliveira Junior (IFRJ)

Marcelo Rafael Borth (IFPR)

Comitê de Organização

Coordenação Geral

Carlos Roberto de Oliveira Junior (IFRJ)

Coordenação Local

Regio Michelin (IFRS-Restinga)

Coordenação do Comitê de Programa

Marcelo Rafael Borth (IFPR)

Comitê do Programa

Allan Edgard Silva Freitas - IFBA
Amaury Antônio de Castro Junior - UFMS
André Atanasio Maranhão Almeida - IFPB
André Constantino da Silva - IFSP
André Luis Schwerz - UTFPR
Angelo Augusto Frozza - IFC
Ângelo Magno de Jesus - IFMG
Aracele Garcia de Oliveira - IFSULDEMINAS
Carlos Alexandre Silva - IFMG
Carlos Eduardo Paulino - IFMG
César Alberto da Silva - IFSP
Christian Puhlmann Brackmann - IF-Farroupilha
Claudio de Castro Monteiro - IFTO
Daniel C. de Paiva - UTFPR
Daniel Correa Lobato - IFSP
Daniel Gomes Soares - IFC
Daniela dos Santos Santana - IFSP
Denis Rocha de Carvalho - IFMG
Diana Cabral Cavalcanti - IFRS
Edilson da Silva - IF-Goiano
Eduardo Filgueiras Damasceno - UTFPR
Evandro Manara Miletto - IFRS
Fabio de Jesus Lima Gomes - IFPI
Fábio Lorenzi da Silva - IFSUL
Fábio Silveira Vidal - IFTO
Fabíola Gonçalves Coelho Ribeiro - IF-Goiano
Felipe Sampaio Dantas da Silva - IFRN
Fernando Oliveira - IFRJ
Gilvan Martins Durães - IF-Baiano

Glauber da Rocha Balthazar - IFSP
Gleizer Bierhalz Voss - IF-Farroupilha
Guilherme Esmeraldo - IFCE
Hiran Nonato Macedo Ferreira - IFSULDEMINAS
Hugo Alberto Perlin - IFPR
Hylson Vescovi Netto - IFC
Ivo Sócrates Oliveira - IFTO
Joacir Simões Ferreira - IFBA
João Marcelo Ribeiro - IFSULDEMINAS
Jocelma Rios - IFBA
José Almir Freire de Moura Júnior - IFPE
Katusco de Farias Santos - IFPB
Klenilmar Lopes Dias - IFAP
Leandro Marques Queiros - IFPE
Letícia Souza Netto Brandi - IFSP
Lizandro de Souza Oliveira - IFSUL
Lourenço Alves Pereira Júnior - IFSP
Luciana M. Vieira Pöttker - IFPR
Luiz Cláudio Machado dos Santos - IFBA
Manoel Messias Pereira Medeiros - IFPI
Marcelo Damasceno - IFRN
Marcelo José Siqueira Coutinho de Almeida - IFPB
Marcelo Rios Kwecko - IFSUL
Marcílio Ferreira de Souza Júnior - IFAL
Marco Antônio Pereira Araújo - IF Sudeste MG
Maria Terezinha Tamanini Andrade - IFBA
Marlon Cordeiro Domenech - IFC
Mateus Pelloso - IF-Catarinense
Matheus D'Eça Torquato de Melo - IFAL
Nécio de Lima Veras - IFCE
Orlando Pereira Santana Júnior - IFMT
Pablo Matos - IFBA
Patricia Nogueira Hübler - IFRS
Paulo César dos Santos - IFSULDEMINAS
Paulo Sergio Horst - IFPR
Priscylla Maria da Silva Sousa - IFS
Ramon Gustavo Teodoro Marques da Silva - IFSULDEMINAS
Raphael de Aquino Gomes - IFG
Reane Franco Goulart - IFTM
Reginaldo Rubens da Silva - IFC
Renato Novais - IFBA
Ricardo Maroquio Bernardo - IFES
Ricardo Rubens Gomes Nunes Filho - IFAL
Rosana Wagner - IF-Farroupilha
Samuel da Costa Alves Basilio - CEFET-MG
Sandro Renato Dias - CEFET-MG
Silvia de Castro Bertagnolli - IFRS

Simone da Silva Amorim - IFBA
Stefan Oliveira - IFTO
Tereza Kelly Gomes Carneiro - IFBA
Thatiane de Oliveira Rosa - IFTO
Thiago Magela Rodrigues Dias - CEFET-MG
Valério Medeiros Jr - IFRN
Vinicius de Miranda Rios - IFTO
Walkiria Helena Cordenonzi - IFSUL
Wanderson Santiago dos Reis - IFSP

Painéis / Palestras

Curriculum de referência para os cursos tecnológicos da área de informação e comunicação

Igor Steinmacher (Universidade Tecnológica Federal do Paraná (UTFPR-CM) /
Comissão de Educação da SBC)

A Comissão de Educação da Sociedade Brasileira de Computação está se esforçando para construir diretrizes para os cursos da área. Entre esses cursos, estão alguns cursos superiores de tecnologia previstos no Catálogo Nacional de Cursos Superiores de Tecnologia, sob o eixo Informação e Comunicação. Atualmente, não existem diretrizes que possam auxiliar coordenadores e Núcleos Docentes Estruturantes com relação aos eixos de conhecimento e conteúdos que devem ser ministrados nos cursos de tecnologia. Nesta palestra, buscaremos mostrar a necessidade e abrir a discussão sobre possíveis diretrizes que auxiliem a elaboração e manutenção de projetos de curso superiores de tecnologia na área de computação.

Trabalhos aceitos

O uso do @prenda como recurso educativo digital

Walkiria Cordenonzi (Instituto Federal Sul-rio-grandense)

Silvio Normey Gómez (PUCRS)

A construção de objetos de aprendizagem para o ensino de sociologia no ensino médio

Vinicius Hartmann Ferreira (IFRS)

Vanessa Petró (IFRS)

Luis Eich (IFRS)

Murillo Weissheimer (IFRS)

Fernando Zanutto (IFRS)

Alecsander Hennig (IFRS)

Desenvolvimento de protótipo de game para inserção no ensino de ciências

Cleon Pereira Júnior (Instituto Federal Goiano)

Zara Hoffmann (Instituto Federal Goiano)

Thais Aires (Instituto Federal Goiano)

Caroline Castro (Instituto Federal Goiano)

Geovanna Santos (Instituto Federal Goiano)

Feira e-lixo: projeto de ensino multidisciplinar envolvendo re-uso de lixo eletrônico

Ana Elisa Schmidt (Instituto Federal Catarinense)

Mozara Koehler (Instituto Federal Catarinense)

Desenvolvimento do aplicativo XÔ ZIKA para Android: tecnologia educacional aplicada na área de saúde

Cleber F. Serafin (Instituto Federal do Paraná)

Felipe J. R. de Carvalho (Instituto Federal do Paraná)

Marcos Schmitt (Instituto Federal do Paraná)

João V. Zarth (Instituto Federal do Paraná)

HackDay@IFC-Cam: evento de preparação para competições de programação como instrumento de ensino

Ana Elisa Schmidt (Instituto Federal Catarinense)

Mozara Koehler (Instituto Federal Catarinense)

e-Sinais: software tradutor de português sinalizado para sinais em libras

Igor Silva (IFBA)

Bruno Mascarenhas (IFBA)

Saionara Araújo (IFBA)

Daniela Silveira (IFBA)

Pablo Matos (IFBA)

INASH: uma ferramenta de autoria para narrativas interativas digitais

Vinicius Hartmann Ferreira (IFRS)
Luis Eich (IFRS)

Sistema para controle remoto de equipamentos eletroeletrônicos utilizando dispositivos móveis

Fabiano Bhering (CEFET-MG)
Hugo Barbosa Alves (CEFET-MG)
Marina Nunes Silva (CEFET-MG)
Vinicius Ferreira da Silva Bianchi Grilo (CEFET-MG)

Estudo da interação do usuário usando o dispositivo sensível o movimento Kinect e o controle sensível ao movimento WiiMote

Vinicius Silva (IFSP)
Luis Otavio Nascimento (IFSP)
André da Silva (IFSP)

Controle de modelos robóticos por meio de gestos naturais

Paulo Henrique R. de Souza (Instituto Federal de Sergipe)
Welerson Melo (Instituto Federal de Sergipe)
Icaro Santana (Instituto Federal de Sergipe)
Nara Doria (Instituto Federal de Sergipe)
Leila Buarque Couto Matos (Instituto Federal de Sergipe)

Um framework para descoberta de recursos usando computação nas nuvens

Rodrigo Mota (IFG)
Pablo da Silva (IFG)
Clézio de Souza (IFG)
Raphael de Aquino Gomes (IFG)

Conhecimento em redes sociais: um estudo de caso aplicado ao Twitter

Angelo Frozza (IFC)
Miguel Airton Frantz (IFC)
Reginaldo Silva (IFC)

Acessibilidade móvel para alfabetização de deficientes visuais: proposta inicial de um protótipo

Jenifer Melissa de Paula (IFTO)
José Valter Amaral de Freitas (IFTO)
Thatiane Oliveira Rosa (IFTO)

Um protocolo para comunicação entre aeronaves não tripuladas e estações de solo

André Alves dos Reis (IFSP)
Bruno César Vani (IFSP)
Kleber Trevisani (IFSP)

Minimização da intervenção humana para detectar e solucionar anomalias rede de computadores

Alexandre Amaral (IFC)

Ana Paula Malheiro (IFC)

A teoria de redes como subsídio para o estudo da colaboração em projetos de pesquisa

Antônio Raphael Brito Meireles (IFBA)

Ingrid dos Santos Vilas Verde (IFBA)

Maria Andrade (IFBA)

Multi-objective test case selection: local search approaches for the NSGA-II algorithm

Luciano Souza (Federal Institute of Education Science and Technology of the North of Minas Gerais)

O Uso do @prenda como Recurso Educativo Digital

Cordenonzi, W., Normey, S., Cardoso, V. Garay, R.ⁱ, Ortiz, A.M.C.ⁱⁱ

Instituto Federal Sul-rio-grandense – Câmpus Santana do Livramento

Paul Harris 410, Sant’Ana do Livramento – RS – Brasil

{walkiriacordenonzi, silviogomez,vanessacardoso}@ifsul.edu.br

Resumo. *Objetos de aprendizagem (OA), são recursos digitais utilizados para aprender. A realidade do nosso Câmpus são cursos binacionais com alunos brasileiros e uruguaios compartilhando o mesmo espaço educacional. O @prenda é um OA no formato de jogo multi-idioma e interdisciplinar que estimula o aprendizado e a competição, simbolizada por IFCoins. Os resultados observados foram a interação e a competitividade por parte dos alunos buscando figurar na lista dos melhores jogadores.*

1 Introdução.

A conectividade alterou a forma como aprendemos, hoje é possível estudar e interagir com professores e estudantes sem a necessidade de estar presente fisicamente. No contexto deste trabalho objeto de aprendizagem entende-se como “... qualquer recurso digital que pode ser reutilizado para suportar a aprendizagem” Wiley (2002). Eles cobrem diversas modalidades de ensino: presencial, híbrida ou à distância. Ao serem incluídos em OA’s estes são normalmente caracterizados por metadados o que facilita a busca, além de outras ações.

Nos dias de hoje o desafio é o de criar métodos educacionais inovadores que favorecem a disseminação do conhecimento por meio de recursos didáticos adaptados ao contexto da mobilidade. Segundo Barbosa (2007), a aprendizagem móvel em uma perspectiva pedagógica aponta para uma nova dimensão na educação com poder de atender necessidades de aprendizagem imediatas com grande flexibilidade. Somando-se a esta característica o entorno educacional que temos no nosso Câmpus onde metade dos alunos são brasileiros e a outra metade são uruguaios. Este é um projeto piloto de cursos binacionais com dupla validação (o diploma do aluno é validado no Brasil e no Uruguai).

O @prenda reúne os conceitos e tecnologias acima citados. Este pode ser utilizado em dispositivos móveis proporcionando diversas possibilidades de acesso e as suas múltiplas opções de área/subárea e principalmente idioma a ser escolhido faz do @prenda um jogo inovador.

Para um melhor entendimento do trabalho aqui realizado o restante do artigo divide-se da seguinte forma: Seção 2 apresenta um sucinto referencial teórico, a Seção 3 descreve o desenvolvimento e funcionamento do jogo desenvolvido e a Seção 4 apresenta resultados já obtidos. As conclusões e referencias seguem.

2 Referencial Teórico

É consenso entre vários autores sobre a importância e sua contribuição no processo de aprendizagem dos jogos educativos. Segundo Antunes (2002), “no sentido etimológico a palavra jogo expressa um divertimento, brincadeira, passatempo sujeito a regras que devem ser observadas quando se joga”.

Para Borin (2004), o conteúdo deve estar de acordo com o grau de desenvolvimento e, ao mesmo tempo, de resolução possível. Portanto, o jogo não deve ser fácil demais e nem tão difícil, para que os alunos não se desestimulem; sendo uma alternativa para se melhorar o desempenho dos estudantes em alguns conteúdos de difícil aprendizagem (GOMES & FRIEDRICH, 2001). Portanto, o jogo pode ser educativo. Pode ser projetado com o objetivo de estimular uma aprendizagem significativa e a construção do conhecimento e, conseqüentemente a construir conexões.

O @prenda é um OA móvel, aliado ao conceito de gamificação, no sentido que o uso desta técnica permite usar os conceitos intrínsecos aos jogos mesmo em atividades que o aluno não esteja de fato jogando, mas que este perceba que esta atividade pode ter um impacto em determinadas ações ou que construa um conhecimento a partir de um idioma escolhido. A seguinte seção descreve o jogo @prenda de forma detalhada.

3 O Jogo @prenda

O @prenda consiste em um jogo multidisciplinar que desafia o aluno em um conjunto de perguntas e respostas de múltipla escolha. A fim de compreender o nível de aprendizagem do aluno as suas respostas serão armazenadas no seu repositório para acompanhamento do professor e análise dos algoritmos contidos no @prenda.

O papel do professor no jogo é inicialmente cadastrar as perguntas. Para cada pergunta inserida no jogo, o professor deverá selecionar o idioma da questão, a área e subárea do conhecimento ao qual pertence. Por exemplo, a pergunta “É uma instrução DML: ”, ao ser inserida deve-se selecionar o idioma, no caso Português, a área – Banco de Dados e a subárea SQL. Para cada questão cadastrada deve-se armazenar cinco respostas e determinar qual é a correta. Todas as questões possuem um grau de dificuldade (fácil, médio, difícil). As questões podem estar associadas a Dica. Na Figura 01, o balão está mostrando um Dica para a questão, seu tempo restante, a pergunta, suas respostas, e a quantidade de IFCOINS que o jogador já acumulou.

Inicialmente o aluno/jogador deverá fazer o seu *login* no sistema. A Figura 2 apresenta a sequência de funcionamento do jogo. Paralelamente no decorrer do jogo o aluno é frequentemente estimulado a responder as perguntas e para isso ele conta a opção de dicas ajudando-o a solucionar as questões do jogo. Através das perguntas e conforme a assertividade do aluno, o reconhecimento dos alunos será avaliado de acordo com as métricas e parâmetros previamente configurados no jogo.

O reconhecimento e gratificação dos alunos mediante o jogo se dá por meio de moedas virtuais chamada IFCOINS (IFSUL + COINS). Dessa forma o aluno poderá elevar sua pontuação geral no jogo.

Ao escolher as tecnologias a serem adotadas no projeto requisitos foram levados em consideração, tais como, escalabilidade, manutenibilidade, entre outros. Considerou-se o desafio de suportar uma elevada quantidade de requisições ao sistema e seu devido tratamento, e agregar novos programadores com conhecimento nas tecnologias que estão envolvidas no projeto. Desenvolver um projeto com a complexidade do @prenda nos levou adotar uma metodologia de desenvolvimento ágil (Scrum) para obter *feedbacks* rápidos afim de entender se o sistema estava sendo desenvolvido dentro do requerido. Tecnologicamente optou-se por utilizar um SGBDOR (Sistema de Gerenciamento de Banco de Dados Objeto Relacional). Este permitiu garantir a integridade e confiabilidade dos dados.

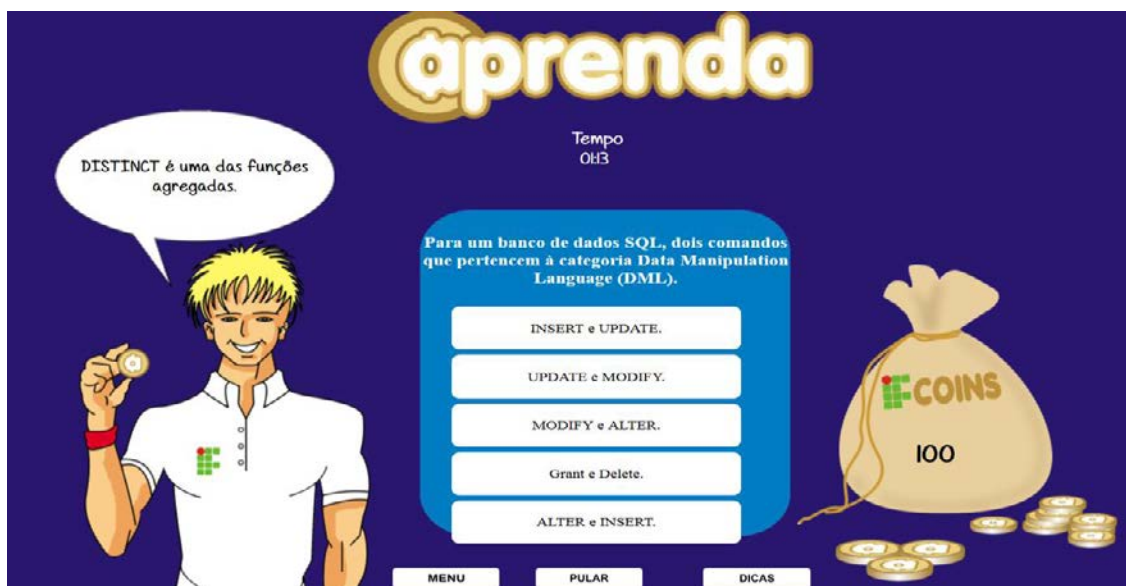


Figura 01– O jogo

Para o desenvolvimento do *back-end* foi estabelecido utilizar a linguagem de programação *Server-Side* PHP5.x juntamente com o Web Server Apache2 ambos sendo uma boa solução para o sistema web.

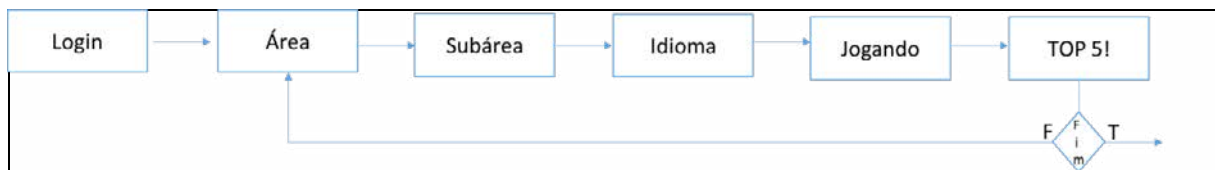


Figura 02 – Sequencia do @prenda

A interface gráfica foi projetada de forma a ser atraente e priorizando aspectos de usabilidade e acessibilidade voltada para Web e principalmente para dispositivos móveis.

3.1 Sobre o gerenciamento do jogo:

Gerenciar o @prenda em nível de dados, seria uma atividade árdua sem o uso de um Sistema de Gerenciamento de Conteúdo (do inglês Content Management System – CMS). Por isso, a administração e manutenção do jogo se dá através de um gerenciador desenvolvido por alunos do curso técnico de Informática para Internet tornando a tarefa eficaz e eficiente. Para garantir a segurança das informações do sistema, o CMS conta com dois diferentes níveis de acessos na área administrativa onde está dividido nos seguintes níveis de acesso: Área Administração e Área do Professor, com módulos e restrições peculiares à cada nível de acesso dos respectivos grupos de usuário. O terceiro módulo é o jogo que está disponível para a comunidade educativa.

4 Resultados e Discussões

A partir do projeto e desenvolvimento do jogo, dentre algumas vantagens com relação aos jogos educativos, e que se encontram implementadas no @prenda, pode-se citar (Grando, 2011): a) fixação de conceitos já aprendidos de uma forma motivadora para o aluno; b) introdução e desenvolvimento de conceitos de difícil compreensão; c) aprender a tomar decisões e saber avaliá-las; d) significação para conceitos aparentemente incompreensíveis; e) propicia o relacionamento das diferentes disciplinas (interdisciplinaridade); f) o jogo requer participação ativa do aluno na construção do seu

próprio conhecimento; g) a utilização dos jogos é um fator de motivação para os alunos; h) o jogo favorece a competição “sadia” e do resgate do prazer em aprender; i) as atividades com jogos podem ser utilizadas para reforçar ou recuperar habilidades de que alunos necessitem. Útil no trabalho com alunos de diferentes níveis.

O jogo está atualmente sendo testado em duas turmas que cursam banco de dados. O conteúdo específico trabalhado foram conhecimentos sobre comandos SQL. Esta utilização resultou em algumas manutenções corretivas e adaptativas e dos itens citados no parágrafo anterior atingiu-se as letras a, b, c, f, g, h (resultado da pesquisa realizada com 20 alunos).

Como resultados qualitativos percebeu-se uma grande motivação por parte dos alunos para responderem o jogo mais rápido e correto para que o jogador figure na lista do TOP 5. Com isso a competitividade foi um fator importante e relevante para o público observado.

Conclusão

A partir da incorporação do jogo no ambiente de aprendizagem pretende-se proporcionar aos alunos a oportunidade de realizar seus estudos independentemente de tempo ou lugar onde se encontram, ou seja, um ensino “globalizado”. O @prenda pode ser uma ferramenta importante na interdisciplinaridade, desde que os professores assim o desejem. A interdisciplinaridade já está sendo projetada para a próxima fase.

Pretende-se ao final do ciclo de vida do @prenda (principalmente após testes exaustivos), disponibilizá-lo em um repositório de OA e/ou em um repositório REA (recurso educativo aberto).

Espera-se a sensibilização do uso de recursos educativos digitais como somadores de materiais didáticos a fim de contribuir com a melhoria do processo de ensino aprendizagem. Não menos importante, fomentar a integração da educação na região de fronteira.

5 Referências

- Antunes, Celso. Jogos para a estimulação das múltiplas inteligências. 11.ed. Rio de Janeiro: Vozes, 2002.
- Barbosa, D. N. F. Um modelo de educação ubíqua orientado à consciência do contexto do aprendiz. UFRGS. Porto Alegre, 2007.
- Borin, Júlia. Jogos e resolução de problemas: uma estratégia para as aulas de matemática. 5ª. ed. São Paulo: CAEM / IME-USP, 2004, 100p.
- Gomes, R. R.; Frieddrich, M. A Contribuição dos jogos didáticos na aprendizagem de conteúdos de Ciências e Biologia. In: EREBIO, 1, Rio de Janeiro, 2001, Anais..., Rio de Janeiro, 2001, p.389-92.
- Grando, R. C. O jogo na educação: aspectos didático-metodológicos do jogo na educação matemática. Unicamp, 2001.
- Wiley, D. Connecting Learning Object s to Instructional Design Theory: a Definition, a Metaphor, and a Taxonomy. In David Wiley (Ed.), 2002.

ⁱ Bolsista de Iniciação Científica com recursos da Pró-Reitoria de Pesquisa – PROPESP - IFSUL

ⁱⁱ Bolsista de Iniciação Científica com recursos da Pró-Reitoria de Pesquisa – PROPESP - IFSUL

A Construção de Objetos de Aprendizagem para o Ensino de Sociologia no Ensino Médio

Vinicius Hartmann Ferreira, Vanessa Petró, Luís Guilherme Eich, Murillo Weissheimer, Fernando Zanutto, Alecsander Hennig

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Campus Feliz
Feliz – RS – Brasil

vinihf@gmail.com, vanessapetro@gmail.com, gui.eich@hotmail.com,
weiss.muri@gmail.com, ferzanuttol999@gmail.com,
hennig.alecsander@gmail.com

***Abstract.** After reinsertion of Sociology's discipline in high school curriculum their curriculum and their practices in the classroom has been discussed in events and research groups. A survey about the practices in the discipline presents the existence of a few interactive features, especially of digital learning objects (LO). This article presents an experience that is being developed for the construction of LOs that began in 2014, whose goal is to help students work out and understand the contents studied. Resources are developed from the demand presented by the students of a school that offer the Technical Course in Computing integrated with high school and are designed and implemented with the collaboration of scholarship students from that course through a research project that integrates the areas of Sociology and Computer Science.*

***Resumo.** Após a reinserção da disciplina de Sociologia no currículo do Ensino Médio, tem se discutido sobre seu currículo e suas práticas em sala de aula em eventos e grupos de pesquisa. Um levantamento realizado sobre as práticas na disciplina aponta para a existência de poucos recursos interativos, sobretudo de Objetos de Aprendizagem digitais (OA). Este artigo apresenta uma experiência que está sendo desenvolvida para a construção de OAs que se iniciou em 2014, cujo objetivo é auxiliar os estudantes a exercitar e fixar os conteúdos estudados. Os recursos são desenvolvidos a partir da demanda apresentada pelos estudantes de uma escola que oferta o Curso Técnico em Informática Integrado ao Ensino Médio e são concebidos e executados com a colaboração de estudantes bolsistas do referido curso por meio de um projeto de pesquisa que integra as áreas de Sociologia e de Informática.*

1. Introdução

Após idas e vindas no currículo do ensino médio, a disciplina de Sociologia voltou a ser reinserida em 2008 como disciplina obrigatória, a partir do momento em que foi sancionada a lei 11.684. O ensino de Sociologia no nível médio enfrenta inúmeros desafios, como a precária formação dos professores, incluindo que de maneira recorrente os titulares da disciplina não têm formação na área de Ciências Sociais; a alta carga horária dos professores da rede pública estadual e os baixos salários (PEREIRA, 2009).

Acrescenta-se a estes desafios os fatos de que a disciplina de Sociologia exige dos estudantes para sua compreensão a capacidade de abstração, uma capacidade que muitos estudantes de ensino médio apresentam deficiência, e a falta de recursos que

possam auxiliar os estudantes com esta dificuldade. Ainda que para a disciplina existam livros didáticos, filmes e documentários, músicas e imagens, poucos são os recursos que permitem ao aluno interagir com o conteúdo e fazer relações entre os conceitos estudados e o seu cotidiano. É neste contexto que recursos digitais interativos, como Objetos de Aprendizagem (OA), podem contribuir para a disciplina de Sociologia (FERREIRA, PETRÓ e EICH, 2015).

Este artigo apresenta um projeto desenvolvido em conjunto entre as disciplinas de Programação e de Sociologia de um curso técnico em informática integrado ao ensino médio. Neste projeto, que se iniciou em 2014, OAs são construídos por alunos do curso a partir das demandas identificadas pela professora da disciplina de Sociologia e dos alunos durante os anos letivos. Este artigo está organizado de forma a apresentar os conceitos que fundamentam a pesquisa na Seção 2, os OAs desenvolvidos e os resultados alcançados na Seção 3 e as conclusões obtidas na Seção 4.

2. Sociologia no Ensino Médio e os Objetos de Aprendizagem

Um OA é qualquer recurso digital que tenha conteúdo educacional e propósito de aprendizagem e que possa ser usado e reutilizado em diferentes contextos (SILVEIRA e CARNEIRO, 2012; WILEY, 2000). Um OA pode ter seu foco em diferentes áreas de conhecimento e ser composto por um ou mais recursos de diferentes formatos, como textos, vídeos, músicas, animações ou jogos.

Em uma busca feita em eventos e periódicos científicos nacionais, tanto da área de Informática na Educação quanto de Sociologia, poucas produções foram encontradas com foco no uso ou produção de recursos digitais interativos para a disciplina de Sociologia. A maioria dos trabalhos relata a utilização de filmes ou documentários e imagens, ou apresenta propostas de metodologias de ensino que utilizem recursos já existentes e que não se caracterizam como interativos (PEREIRA, 2009; LEITE, HUBLER e JUNIOR, 2013).

Os OAs podem se constituir como recursos relevantes para o ensino de Sociologia no nível médio. Entre os conteúdos previstos para esta disciplina estão conceitos teóricos que precisam ser relacionados com o mundo objetivo, além de acontecimentos históricos e pensadores que contribuíram para a formação da Sociologia enquanto ciência, o que também é conteúdo a ser trabalhado em sala de aula.

3. Objetos de Aprendizagem para Sociologia

No início do projeto, em 2014, verificou-se que os alunos apresentavam dificuldades com o conteúdo do 1º ano, relacionado aos pensadores clássicos da Sociologia. Partindo da identificação deste conteúdo, buscou-se definir qual seria o melhor recurso para esta situação em reuniões que contavam com a participação dos professores da disciplina de Programação e de Sociologia e também de um aluno do curso que participava do projeto. Após o refinamento do conteúdo que deveria ser tratado relacionado aos pensadores clássicos, concluiu-se que seria desenvolvido um OA que consistia em uma linha do tempo na qual os alunos deveriam relacionar os pensadores de forma cronológica com suas principais contribuições para a Sociologia (Figura 1).

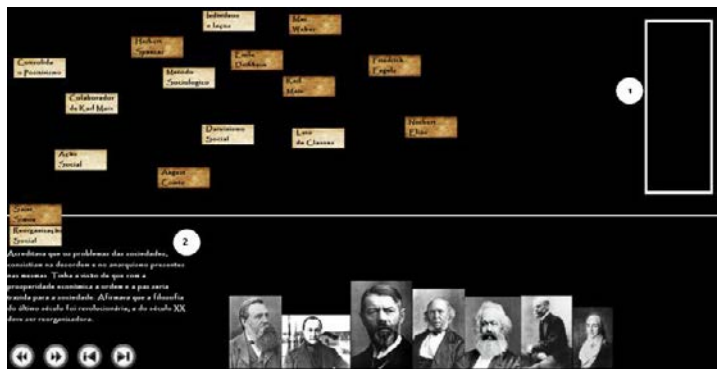


Figura 1. Tela principal do OA Linha do Tempo para Sociologia.

O OA foi desenvolvido de forma que os professores possam inserir informações sobre diferentes pensadores, tornando-o reutilizável, e possui dois módulos, o de atividade, no qual são apresentadas em cor vermelha as relações incorretas e em cor verde as relações corretas, e o de avaliação, no qual é apresentada a pontuação obtida pelos alunos com as relações corretas (Figura 2).

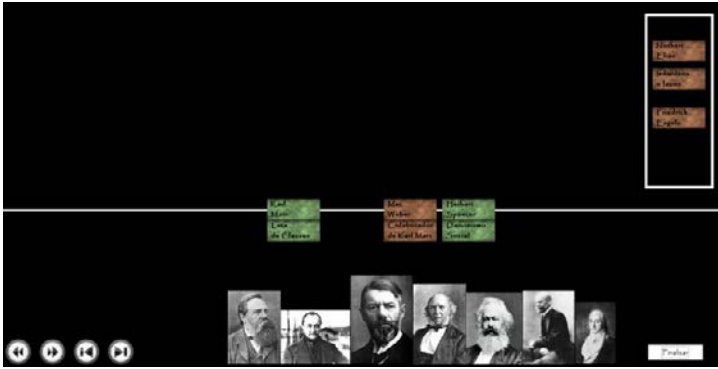


Figura 2. Módulo de atividade do OA Linha do Tempo para Sociologia.

Um experimento foi conduzido com o intuito de avaliar a receptividade dos alunos ao OA. Este experimento foi realizado por 40 alunos de uma turma de 1º ano de ensino médio de um curso técnico em informática integrado. Os alunos utilizaram o recurso na modalidade de avaliação e de atividade durante uma aula de Sociologia. Ao final, responderam um questionário que permitiu identificar, dentre outros aspectos, que o recurso os auxiliou a organizar o pensamento sobre o conteúdo trabalhado (FERREIRA, PETRÓ e EICH, 2015).

No ano de 2015 o projeto iniciou a construção de um novo OA, que tem previsão de conclusão até o final do 1º semestre de 2016. Este novo recurso também está sendo desenvolvido por alunos do curso técnico em informática integrado ao ensino médio e possui fundamentos relacionados à mecânica de jogos. O novo recurso tem por objetivo trabalhar conceitos relacionados às ideologias políticas e à forma como os atores políticos lidam com as mesmas no parlamento.

Neste novo recurso o estudante assume a presidência de um país fictício e deve inicialmente indicar se sua ideologia será voltada para “o bem estar” ou para o neoliberalismo, o que é associado a ideologias de esquerda e de direita no espectro político. A partir desta escolha o estudante deve tomar decisões que afetam indicadores como saúde, educação, economia e segurança. Cada decisão implicará na inclinação de um termômetro que indica a qual das ideologias estão relacionadas suas decisões (Figura 3). Também está em desenvolvimento neste OA um mecanismo que permitirá ao estudante solicitar opinião aos seus conselheiros, que também têm partido e

ideologias próprias. No momento dessa consulta será possível identificar elementos para além das ideologias que exercem influência no posicionamento político, tais como os interesses pragmáticos e privados, as relações pessoais, as trocas de favores, dentre outros.



Figura 3. Tela de administração do país com indicadores da ideologia adotada.

4. Conclusão

Este artigo apresentou os resultados de um projeto interdisciplinar que envolve as disciplinas de Sociologia e de Programação de um curso técnico em informática integrado ao ensino médio. Até o momento foi desenvolvido um OA, intitulado Linha do Tempo para Sociologia e disponível para download em <http://vhflabs.com.br/projetos.html>, que foi utilizado por alunos de uma turma de 1º ano do ensino médio em um experimento que indicou boa aceitação.

Um novo OA está sendo desenvolvido com foco em conteúdos como ideologias políticas. Após sua conclusão pretende-se realizar um experimento com as turmas do 3º ano do ensino médio com o intuito de verificar a recepção deles ao novo recurso e avaliar possíveis contribuições na compreensão dos conteúdos.

Referências

- FERREIRA, V. H.; PETRÓ, V.; EICH, L. G. Um objeto de aprendizagem para a disciplina de sociologia no ensino médio. **RENOTE - Revista Novas Tecnologias na Educação**, Porto Alegre, v. 13, n. 1, p. 1-10, dez. 2015.
- LEITE, C. A. C.; HUBLER, P. N.; JUNIOR, A. C. G. R. A inovação no ensino de sociologia com a utilização de objetos de aprendizagem. In: **CONGRESSO BRASILEIRO DE SOCIOLOGIA**, 16, 2013, Salvador. **Anais...** Salvador: SBS. 2013.
- PEREIRA, T. I. Para além do senso comum: aportes para a construção do conhecimento sociológico na educação básica. In: **CONGRESSO BRASILEIRO DE SOCIOLOGIA**, 14, 2009, Rio de Janeiro. **Anais...** Rio de Janeiro: Sociedade Brasileira de Sociologia, 2009.
- SILVEIRA, M. S.; CARNEIRO, M. L. F. Desconstruindo objetos de aprendizagem: reflexões sobre sua qualidade de uso. In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 23, 2012, Rio de Janeiro. **Anais...** Rio de Janeiro: Sociedade Brasileira de Computação, 2012.
- WILEY, D. A. **Learning object design and sequencing theory**. Provo, Utah: Department Of Instructional Psychology And Technology/Brigham Young University, 2000. 131p. Tese de Doutorado.

Desenvolvimento de Protótipo de Game para Inserção no Ensino de Ciências

Cleon X. Pereira Júnior¹, Zara Hoffmann¹,
Caroline P. Castro¹, Geovanna S. Santos¹, Thaís A. Aires¹

¹Instituto Federal Goiano (IF Goiano)
Rua Rui Barbosa Qd 13 Lt. 1-A Setor Aeroporto – Campos Belos – GO – Brazil

Este trabalho teve apoio financeiro do PIPECTI / IF Goiano para apresentação.

{cleon.junior, zara.hoffmann}@ifgoiano.edu.br

Abstract. *The use of Information and Communication Technologies in the educational environment indicates a possibility of educators and students exercise their own citizens through the use of technology. In this context, the development of diversified didactics actions which combine scientific knowledge to decisions taken on a day-to-day that stimulate significant approaches of understanding scientific knowledge, it becomes a possible alternative, both for the development of software systems and for the Science teaching. Within a development proposal of creativity and initiative, this work aims beyond the interdisciplinary interaction between software development areas with common base classes, it is also going to develop a technological prototype in a game format to be applied in Education Science. We intend apply to freshman students, who had presented a large gap of scientific content, after diagnosis survey.*

Resumo. *O uso das Tecnologias da Informação e Comunicação no meio educacional aponta para uma possibilidade de professores e alunos exercerem sua própria cidadania por meio do uso das tecnologias. Neste contexto, o desenvolvimento de ações didáticas diversificadas que aliem saberes científicos às decisões tomadas no dia-a-dia que estimulem abordagens significativas para compreensão dos conhecimentos científicos, torna-se uma alternativa possível, tanto para o desenvolvimento de sistemas de softwares quanto para o Ensino de Ciências. Dentro de uma proposta de desenvolvimento da criatividade e iniciativa este trabalho tem como objetivo além da interação interdisciplinar entre áreas de desenvolvimento de softwares com disciplinas da base comum, desenvolver um protótipo tecnológico em formato de jogo para ser aplicado no Ensino de Ciências. Pretende-se aplicar o protótipo entre os alunos ingressantes, os quais apresentaram grande defasagem de conteúdos científicos, após levantamento diagnóstico.*

1. Introdução

O termo Tecnologia da Informação e Comunicação (TIC) é usado no contexto educacional para relacionar tanto a educação conectada à internet e seu aspecto social, quanto o uso educacional das tecnologias da informação e comunicação tais como a televisão, o computador e outras mídias digitais. Numa visão determinista, conforme

[Hato Sánchez and Vargas Lema 2013], tais tecnologias apontam para um conjunto de inovações tecnológicas e, seu conjunto de ferramentas, proporcionam uma redefinição radical do comportamento e funcionamento da sociedade contemporânea. Já na visão instrumentalista de [Castels and em Rede 1999], “a tecnologia não determina a sociedade; incorpora-a” visto que há um conjunto de fatores que intervêm nos processos societários, interativos e complexos, como a criatividade e a própria tecnologia. Esta forma de perceber a tecnologia considera-a um instrumento “neutro” utilizada com a finalidade de ter acesso ao conhecimento, também visto como neutro. A essa dualidade de pensamento [Peixoto and dos Santos Araújo 2012] descrevem como racionalidade instrumental: “um movimento que oscila entre a visão da tecnologia como ferramenta [...] e a atribuição à tecnologia o poder de configurar a cultura e a sociedade”, e é esta racionalidade que permeia os projetos pedagógicos de inserção das TICs nos meios educacionais.

Nas relações entre escola e TIC, ao pensar de forma acríica sobre as visões, tanto a determinista quanto a instrumentalista, corre-se o risco de acreditar que: 1 – prover a escola de aparatos tecnológicos, como computadores, tablets, lousa digital, entre outros, propiciará a inclusão digital dos alunos; 2 – ficar contra a inserção das TIC nas escolas, numa posição defensiva, protegerá os alunos e professores da desigualdade facilitada pelas TIC, e 3 – focar na capacitação a priori para uso das TIC, de fora para dentro da escola, pode excluir as possibilidades de apropriação dessas TIC de forma a impedir as mudanças escolares nas suas inadequações. Como deixa transparecer [BUZATO 2006] ao “conceber a tecnologia como ação social coletiva” é possível identificar além das necessidades de infraestrutura e formação, as possibilidades de professores e alunos exercerem sua própria cidadania através do uso das tecnologias.

Para uma prática educativa que atenda tanto os requisitos da Inclusão Digital quanto os Parâmetros Curriculares Nacionais para o Ensino Médio (PCNEM), pautada na interdisciplinaridade, contextualização e desenvolvimento de competências e habilidades, não é possível que o professor se atenha ao livro didático apenas, o qual proporciona um entendimento periférico e fragmentado dos conhecimentos [SEMTEC 1999]. Neste contexto, o Ensino Tecnológico pode contribuir na construção da cidadania, isto é, formar cidadãos capazes de participar criticamente na sociedade, através de uma proposta de ações didáticas inovadoras, com a incorporação das Tecnologias de Informação e Comunicação. Igualmente, entende-se que o Ensino de Ciências ainda é pautado por uma grande quantidade de conceitos teóricos e que a falta de metodologias alternativas voltadas para um aprendizado lúdico e motivador, dificulta o processo de aprendizagem, levando os alunos à memorização em nível superficial. Numa proposta de estratégias motivacionais, dinâmicas e inovadoras, o uso de jogos educativos pode suprir a dificuldade de se ministrarem aulas experimentais devido à falta de laboratórios. Porém, mais que o domínio das tecnologias, desenvolver nos alunos habilidades mentais e pessoais, tais como criatividade, iniciativa, espírito crítico e colaborativo, sociabilidade e capacidade de transformar informações que recebe em conhecimento útil [TIBÚRCIO 1997]. E é nesse sentido que este trabalho propõe o desenvolvimento de uma aplicação no formato de jogo para aprendizagem de conceitos de ciências naturais.

2. Jogos no ensino de ciências

Os jogos, ou elementos de jogos, estão inseridos cada vez mais no meio educacional, e é possível encontrá-los para o ensino das mais diversas disciplinas. Para o ensino de

ciências, [Li and Tsai 2013] apresentam uma revisão dos trabalhos mais relevantes que utilizam jogos ou estratégias de jogos. Na pesquisa é apontada que a partir de 2006 houve um crescimento no uso de jogos para a aprendizagem. Os trabalhos levantados na pesquisa apontam os jogos para trabalhar conceitos, processos científicos, resolução de problemas, afetividade, engajamento ou aprendizagem contexto-social.

Mais específico para o ensino de química, existe uma gama de trabalhos que demonstram o uso de jogos para maior efetividade do ensino. [Boletsis and McCallum 2013] apresentam, através da realidade aumentada, um jogo colaborativo para o ensino da tabela periódica. Já [Capps 2008] apresenta o uso de um jogo de tabuleiro para o reforço dos conceitos básicos de química. Desta forma, é possível ver que, de diferentes formas, os jogos vem crescendo para engajar mais o estudante na aprendizagem, tornando estratégias pedagógicas durante o ensino.

3. O jogo - Os elementos químicos no meio ambiente

O jogo, proposto aqui, tem como enredo a história de um garoto chamado Caio que está a caminho de casa, sendo que o mesmo deve chegar antes do sol se pôr. Durante o caminho, o jogador terá que resolver alguns problemas para permitir que o garoto prossiga. Por exemplo, para atravessar o primeiro obstáculo, o usuário terá que selecionar, dentre os compostos, qual deles representa a água. Respondendo corretamente, o lugar que era um vazio, aparecerá um rio e permitirá que o mesmo atravesse, conforme apresentado na figura 1. Outros desafios como o plantio de árvore para a sobrevivência (Relação do oxigênio com o gás carbônico), combustão do fogo para o aquecimento, etc, são apresentados permitindo que o jogador enxergue os elementos químicos presentes na natureza.

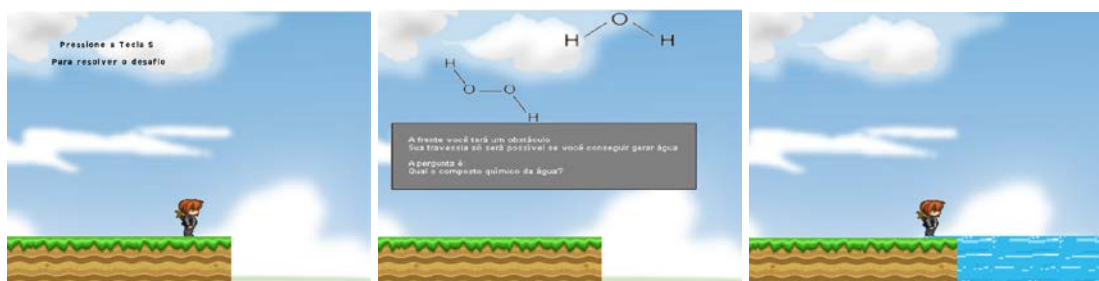


Figura 1. Personagem em frente ao obstáculo; o desafio para atravessá-lo; Desafio cumprido

Caracterizado como uma pesquisa qualitativa, este trabalho vem sendo desenvolvido com a utilização de diversas estratégias de investigação. A princípio foram utilizados dados da avaliação diagnóstica efetuada pelos professores com alunos ingressantes, com a finalidade de detectar suas dificuldades específicas. Esta investigação permitiu determinar o nível de conteúdo científico a ser trabalhado no jogo, sendo a química e a educação ambiental escolhido como foco principal.

O jogo é dividido em fases, onde os acertos permitem que o estudante prossiga e os erros fazem com que o mesmo sofra uma penalidade, perdendo uma porcentagem da vida. Desta forma, acredita-se que o estudante terá uma oportunidade a mais de revisar os conceitos básicos de química e associá-los no meio ambiente entendendo a necessidade dos elementos.

Para o desenvolvimento do jogo foi utilizado o Game Maker Studio ¹ pela facilidade na criação de cenários e possibilidades geradas pelo mesmo. A documentação foi baseada no padrão de jogos, o Game Design. Durante o desenvolvimento foram utilizadas estratégias de engenharia de software, como testes de caixa preta na verificação de erros ou inconsistências.

4. Considerações Finais

Este trabalho apresentou uma estratégia, utilizando um protótipo de game, para o ensino de química e educação ambiental. Espera-se que, desta forma, os estudantes tenham a possibilidade de aprender ou revisar os conteúdos básicos de química de uma forma divertida, contextualizando e entendendo a necessidade dos compostos no meio ambiente.

Como próxima etapa, haverá uma aplicação do protótipo em turmas iniciantes do ensino médio no Instituto federal, procurando assim, enriquecer mais os conhecimentos científicos, onde os mesmos apresentam uma grande defasagem. Espera-se, ainda como continuação deste trabalho, uma melhoria na interface do jogo, enriquecendo com elementos e animações que levem a concentração do jogador. Também espera-se criar novos enredos com objetivos distintos para aprimorar ainda mais o conhecimento dos usuários.

Referências

- Boletsis, C. and McCallum, S. (2013). The table mystery: An augmented reality collaborative game for chemistry education. In *Serious Games Development and Applications*, pages 86–95. Springer.
- BUZATO, M. E. K. (2006). Letramentos digitais e formação de professores. In *III Congresso Ibero-Americano EducaRede*, volume 29.
- Capps, K. (2008). Chemistry taboo: An active learning game for the general chemistry classroom. *Journal of chemical education*, 85(4):518.
- Castels, M. and em Rede, A. S. (1999). A era da informação: economia, sociedade e cultura. *São Paulo: Paz e Terra*, 3.
- Hato Sánchez, M. A. and Vargas Lema, S. T. (2013). El uso de las tics como instrumento pedagógico en el proceso educativo de los estudiantes de bachillerato en químico biólogo del i.t.s. república del ecuador durante el año lectivo 2011-2012.
- Li, M.-C. and Tsai, C.-C. (2013). Game-based learning in science education: a review of relevant research. *Journal of Science Education and Technology*, 22(6):877–898.
- Peixoto, J. and dos Santos Araújo, C. H. (2012). Tecnologia e educação: algumas considerações sobre o discurso pedagógico contemporâneo. *Educação & Sociedade*, 33(118):253–268.
- SEMTEC, M. (1999). Parâmetros curriculares nacionais para o ensino médio. *Brasília: Secretaria de Educação Média e Tecnologia*.
- TIBÚRCIO, C. (1997). Guia de informática e internet para a educação brasileira. *São Paulo*.

¹**Game Maker Studio:** Motor de jogo com suporte à linguagem GML, disponível em <http://yoyogames.com/>

Feira E-Lixo: projeto de ensino multidisciplinar envolvendo re-uso de lixo eletrônico

Ana Elisa Ferreira Schmidt¹, Mozara Dias Koehler¹

¹ Instituto Federal Catarinense – Campus Camboriú (IFC)
Caixa Postal 2016 – 88340-055 – Camboriú – SC – Brasil

{anaelisa,mozara}@ifc-camboriu.edu.br

Abstract. *This paper presents the E-Waste Fair, an educational project that proposes a multidisciplinary science fair event involving high school and technical college students whose projects must use electronic waste components, e-waste. The projects presented at the fair should target teaching-learning contents of high school disciplines or technology. The E-Waste Fair aims to stimulate student's awareness for proper disposal and re-use of electronic components at the same time it encourages creativity, innovative spirit and integration of knowledge from different disciplines.*

Resumo. *Este artigo apresenta o projeto de ensino da Feira E-Lixo que propõe a realização de uma feira científica multidisciplinar com alunos do ensino médio técnico onde os trabalhos apresentados devem envolver componentes de lixo eletrônico, e-lixo. Os trabalhos apresentados na feira devem privilegiar o ensino-aprendizado de conteúdos de disciplinas do ensino médio ou de tecnologia. A Feira E-Lixo tem por finalidade estimular nos estudantes a conscientização para o descarte adequado e re-uso de componentes eletrônicos ao mesmo tempo que incentiva a criatividade, espírito inovador e integração de conhecimentos adquiridos em diferentes disciplinas.*

1. Introdução

A rápida obsolescência dos equipamentos eletrônicos tem gerado um grande problema para o meio ambiente e sociedade: o consumismo gera lixo. Nos últimos dez anos, a população do Brasil aumentou 9,65%, enquanto que, no mesmo período, o volume de lixo cresceu mais do que o dobro disso, 21%. Esta enorme geração de lixo, entretanto, não é acompanhada de um descarte adequado. De acordo com dados da Associação Brasileira de Empresas de Limpeza Pública e Resíduos Especiais [Abrelpe, 2016], só em 2012, dos 64 milhões de toneladas de resíduos produzidos pela população, 24 milhões (37,5%) foram enviados para destinos inadequados [OECD, 2014].

O descarte inadequado de lixo é prejudicial à saúde pública e danoso ao meio ambiente. A fim de enfrentar as consequências sociais, econômicas e ambientais do manejo de resíduos sólidos sem prévio e adequado planejamento técnico, a Lei no 12.305/10 [Brasil, 2010] instituiu a Política Nacional de Resíduos Sólidos (PNRS), regulamentada pela Decreto 7.404/10 [Brasil, 2010]. Esta política propõe a prática de

hábitos de consumo sustentável e contém instrumentos variados para propiciar o incentivo à reciclagem e à reutilização dos resíduos sólidos (reciclagem e reaproveitamento), bem como a destinação ambientalmente adequada dos dejetos [OECD, 2014].

O lixo eletro- eletrônico ou e- lixo é o resíduo material produzido pelo descarte de equipamento eletrônico, entre eles computadores, celulares, eletrodomésticos ou televisões. O descarte inadequado do lixo eletrônico (e-lixo) pode causar um grande impacto ambiental porque tem elementos químicos como arsênio, cobre, alumínio, cádmio, mercúrio, chumbo e berílio que, se entrarem em contato com o solo, podem penetrar até o lençol freático e gerar a contaminação de plantas e animais por meio da água [Mattos, 2008]. Segundo a ONUBr – Nações Unidas no Brasil [ONUBr, 2015] atualmente são gerados aproximadamente 41 milhões toneladas de lixo eletrônico por ano no mundo e entre 60% e 90% são comercializados ilegalmente ou jogados no lixo.

Dentro da proposta apoiar e disseminar a ideia de novo uso para materiais que não serão mais usados em sua forma original, vê-se nas feiras de ciências uma possibilidade conscientização sobre o tema. Assim, surge a proposta da realização de uma feira de ciências multidisciplinar onde o lixo eletrônico deve ser utilizado na elaboração dos trabalhos a serem apresentados. Entende-se a proposta da feira como um projeto de ensino onde a realização do evento amplia o processo ensino-aprendizagem para além da sala de aula, permitindo inclusive que a comunidade externa participe e aprecie os trabalhos desenvolvidos pelos alunos participantes da feira.

Segundo Mezzari (2011) a exposição dos trabalhos em uma feira multidisciplinar à comunidade auxilia na formação do aluno como ser humano integral, com possibilidades de desenvolverem-se nas mais diversas áreas do conhecimento, não se limitando a uma ou outra competência privilegiada. Segundo a lei 11.892 [Brasil, 2008], os Institutos Federais têm por finalidade e característica: realizar e estimular a pesquisa aplicada, a produção cultural, o empreendedorismo, o cooperativismo e o desenvolvimento científico e tecnológico. Esta característica propicia um ambiente muito favorável ao desenvolvimento de projetos com aplicação prática.

2. Projeto de Ensino: Feira E-Lixo

O evento da Feira de Ciências E-lixo teve sua primeira edição realizada no IFC-Camboriú em 2015, dentro do Encontro de Tecnologia da Informação. Os principais objetivos deste projeto de ensino são: promover o desenvolvimento da criatividade e da capacidade inventiva e investigativa dos estudantes; incentivar a atividade científica, através da elaboração e execução de projetos científicos multidisciplinares, auxiliando, assim, na construção e consolidação do conhecimento; estimular o aproveitamento dos resíduos tecnológicos.

Para execução dos projetos apresentados na feira, os grupos devem ser formados por no máximo 3 alunos da instituição, de qualquer curso de nível médio, e um professor orientador (não obrigatório). As propostas devem ter como matéria prima principal o lixo eletrônico e devem se enquadrar nas modalidades: Inovação/Invenção

ou Faça Você Mesmo (Do It Yourself – DIY), onde em cada uma destas modalidades tem-se as categorias: ensino-aprendizagem ou tecnologia.

A modalidade Inovação/Invenção aceita propostas novas ou reformulações de um projeto já existente. No Faça Você Mesmo as propostas podem desenvolver projetos já existentes em tutoriais largamente disponíveis através de sites na internet como Manual do Mundo e Brincando com Robson Sato [Thenório, 2008; Sato, 2015].

Além da elaboração do projeto proposto em si, cada grupo de alunos deve produzir um vídeo-tutorial que explica os fundamentos teóricos e os passos para elaboração do seu projeto. Este vídeo, que deve ser publicado no YouTube pelo grupo, é avaliado por cada avaliador da Feira E-Lixo e considerado na classificação final dos vencedores em cada categoria. Além da prévia avaliação dos vídeos, os trabalhos são expostos para os avaliadores e para a comunidade acadêmica durante o evento da Feira E-Lixo. Os melhores trabalhos de cada modalidade/categoria são premiados com medalhas e brindes, além do recebimento de certificado de participação e exposição dos vídeos produzidos para cada projeto em canal aberto do YouTube.

3. Feira E-Lixo 2015: resultados alcançados

Para a realização do evento foi realizada uma campanha de coleta de e-lixo, que arrecadou 298 kg de lixo eletrônico que, além de serem utilizados nos projetos da feira, foram utilizados em outros projetos do IFC-CAM. O e-lixo que não foi utilizado foi encaminhado à cooperativa de reciclagem.

Durante o evento foram apresentados 07 trabalhos, com 19 alunos envolvidos dos cursos Técnico em Informática e Técnico em Controle Ambiental, cuja matéria prima principal foi o lixo eletrônico, tendo seus vídeos publicados no canal aberto da Feira E-Lixo no YouTube [E-Lixo 2015]. A elaboração dos trabalhos envolveu conhecimentos adquiridos em disciplinas como física, matemática, instalações elétricas, eletrônica, hardware, algoritmos, programação avançada, projeto de interfaces, dentre outras. A figura 1 mostra momentos da feira.



Figura 1. Momentos da Feira de Ciências E-Lixo 2015.

O trabalho “Desenvolvimento de um Aspirador de Pó para Teclado”, teve artigo aceito no 4º Concurso de Trabalhos Técnicos em Informática no Computer on the Beach 2016 [COTB, 2016].

Referências

- Abrelpe. Associação Brasileira de Empresas de Limpeza Pública e Resíduos Especiais. <http://www.abrelpe.org.br/>. Março 2016.
- Brasil. Lei Nº 12.305, de 2 de agosto de 2010. Institui a Política Nacional de Resíduos Sólidos Brasília. http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/lei/112305.htm. Março 2016.
- Brasil. Decreto Nº 7.404, de 23 de dezembro de 2010. Regulamenta a Lei no 12.305, de 2 de agosto de 2010, que institui a Política Nacional de Resíduos Sólidos, cria o Comitê Interministerial da Política Nacional de Resíduos Sólidos e o Comitê Orientador para a Implantação dos Sistemas de Logística Reversa. Brasília. http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2010/Decreto/D7404.htm. Março 2016.
- Brasil. Lei Nº 11.892, de 29 de janeiro de 2008. Institui A Rede Federal de Educação Profissional, Científica e Tecnológica, Cria Os Institutos Federais de Educação, Ciência e Tecnologia, e Dá Outras Providências. Brasília. http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/111892.htm. Março 2016.
- COTB. Computer on the Beach 2016. <http://www.computeronthebeach.com.br/portal/programacao/concurso-de-trabalhos-tecnicos-2016>. Março 2016.
- E-Lixo. Feira de Ciências E-Lixo 2015 do Instituto Federal Catarinense – Campus Camboriú. Vídeos dos trabalhos apresentados. <https://www.youtube.com/channel/UCVxF4xI9O7kcKsndxPdGuqQ>. Março 2016.
- Thenório, Iberê; Fulfaro, Mari. Manual do mundo. 2008. <http://www.manualdomundo.com.br>. Março 2016.
- Mattos, K. M. da C.; Mattos, K. M. da C.; Perales, W. J. S. Os impactos ambientais causados pelo lixo eletrônico e o uso da logística reversa para minimizar os efeitos causados ao meio ambiente. 2008. http://www.abepro.org.br/biblioteca/enegep2008_TN_STP_077_543_11709.pdf. Março 2016.
- Mezzari, S.; Frota, P. R. de O.; Martins, M, da C. Feiras multidisciplinares e o ensino de Ciências. Revista Electrónica de Investigación y Docencia, n.1, p. 107-119, 2011.
- OECD, Associação O Eco. Entenda a Política Nacional de Resíduos Sólidos. Julho, 2014. <http://www.oeco.org.br/dicionario-ambiental/28492-entenda-a-politica-nacional-de-residuos-solidos/>. Março 2016.
- ONUBr– Nações Unidas no Brasil. 2015. <https://nacoesunidas.org/onu-preve-que-mundo-tera-50-milhoes-de-toneladas-de-lixo-eletronico-em-2017/>. Março 2016.
- SATO, Robson. Brincando com Robson Sato. 2015. <http://www.brincandocomrobsonsato.com.br/>. Março 2016.

Desenvolvimento do aplicativo XÔ ZIKA para Android: Tecnologia educacional aplicada na área de saúde

Cleber Fernando Serafin¹, Felipe José Rezende de Carvalho¹, Marcos Fernando Schmitt¹, João Vitor Zarth¹

¹Técnico Integrado em Cooperativismo – IFPR – Capanema – PR – Brasil

{cleber.serafin,felipe.carvalho,marcos.schmitt}@ifpr.edu.br, manoloae2@hotmail.com

Abstract. Contemporaneously Brazil is plagued by a great evil: the proliferation of the mosquito *Aedes aegypti*. Our region has become endemic with growing number of infected people. Lack of awareness and collaboration of the population contributes to the growth of this problem. In order to improve the knowledge of people about this issue, helping to combat the problem, we have developed the free app *Xô Zika* for mobile devices with *Android* operating system. He has issues involving the subject in question in the *quiz* format. In addition, the application has a form so that people can send anonymous complaints of a possible focus. This is sent by the developer team to the competent local authorities.

Resumo. Contemporaneamente o Brasil é assolado por um grande mal: a proliferação do mosquito *Aedes aegypti*. Nossa região tem se tornado endêmica com número crescente de pessoas infectadas. A falta de conscientização e colaboração da população contribui para o crescimento desse problema. Com o objetivo de melhorar o conhecimento das pessoas sobre essa questão, auxiliando no combate ao problema, desenvolvemos o aplicativo gratuito *Xô Zika* para dispositivos móveis com sistema operacional *Android*. Ele conta com questões envolvendo a temática em questão no formato de *quiz*, além disso, dispõe de um formulário para que a população possa enviar denúncias anônimas de um possível foco às autoridades competentes.

1. Introdução

As tecnologias digitais estão cada vez mais presentes nos diversos setores da sociedade, não só no que tange a comunicação, mas também como recurso de ensino e aprendizagem. Segundo pesquisa da Jana (2015), o Brasil conta com 49 milhões de dispositivos móveis, podendo chegar a 72 milhões em 2018, sendo que 90% deles contam com sistema operacional *Android*. O número de usuários desse sistema operacional livre vem crescendo ano a ano, os quais eram de 72,6% em 2012 e 85,1% em 2013 considerando pesquisa da Gartner (2013).

Levando em conta a ampla difusão desses dispositivos e a necessidade de conscientizar a população sobre o grande problema da proliferação do mosquito *Aedes aegypti* que é agente etiológico de várias doenças, nada melhor que juntar as informações necessárias com esse meio de comunicação para alcançar a população. Faz-se necessário então um aplicativo que pudesse ser educativo, interativo e recreativo, com intuito de apresentar os problemas que envolvem esse mosquito.

Segundo Fontes (2002), o processo de aprendizagem que utiliza de recursos digitais e as Novas Tecnologias da Informação (NTI), possui condições inovadoras de produção do conhecimento. Esse contexto institui diferenciadas formas de comunicação

e organização de mensagens entre seus usuários.

Neste trabalho, apresentamos o estado atual de desenvolvimento do nosso aplicativo *Xô Zika*, que conta com questões em um jogo *quiz*, além de uma tela que propicia ao usuário fazer denúncias anônimas de possíveis criadouros do *Aedes aegypti*.

A seção 2, o artigo explana sobre os materiais e métodos utilizados no desenvolvimento do projeto. Já na seção 3, é comentado um pouco sobre a ferramenta *App inventor*. A seção 4 apresenta uma discussão juntamente aos resultados parciais e, por fim, apresenta-se uma conclusão.

2. Material e Métodos

Em um momento de descontração, alguns professores do IFPR Campus Capanema – PR, debatiam sobre possíveis formas de conscientização da população para reduzir a multiplicação do *Aedes aegypti*, problema que assola a população em quase todo Brasil. Com o crescimento da utilização de dispositivos móveis, o desenvolvimento de um aplicativo que, não somente auxiliasse no combate ao mosquito a partir de denúncias de focos, mas também possuísse finalidade educativa à população, pareceu algo promissor a alcançar os objetivos almejados.

Delineada a ideia, foi traçado o planejamento e as metas sobre sua execução. Para isso, foi feita uma chamada interna entre os discentes do primeiro ano do curso Técnico Integrado de Cooperativismo, para que tivessem a oportunidade de participar do trabalho, selecionando-se assim o aluno João Zarth. O mesmo ajudou a equipe com sugestões no projeto e também com seu desenvolvimento, convergindo para a idealização do aplicativo *Xô Zika* a partir do *framework App Inventor*.

O desenvolvimento do aplicativo teve início no começo do mês de março de 2016. Criou-se um roteiro de produção com as etapas e as funções necessárias para o mesmo. Definiu-se que seria um jogo no formato *quiz* e que teria uma tela para denúncia anônima de possíveis focos. O banco de perguntas apresentadas no *quiz* foi formulado com a colaboração de um professor com formação em Farmácia Bioquímica. Em parceria com o aluno, baseados em documentos oficiais da campanha de combate ao *Aedes aegypti* fornecidas pelo Ministério da Saúde (2015), as questões foram formuladas. Essas perguntas estão em constante atualização para informar ao usuário de novidades, inclusive notificando sobre novas áreas endêmicas. Também está em desenvolvimento uma nova funcionalidade para o aplicativo, que permitirá o envio de foto dos possíveis criadouros através da tela de denúncias. O aplicativo está disponível para *download* gratuito na *Play Store* sob o nome de *Xô Zika*.

3. Um pouco sobre o *App inventor*

O *App inventor* é uma plataforma livre de programação com funcionalidades para o desenvolvimento de aplicativos forma trivial, trabalhando com o sistema *blockly*¹, permitindo que a criação seja ágil e eficaz, reduzindo a preocupação com detalhes das linhas de código. Apesar de sua simplicidade de manipulação, é uma ferramenta robusta e dispõe de uma enorme gama de recursos, atendendo às necessidades do projeto.

A tecnologia de blocos do *App Inventor* foi desenvolvida a partir da biblioteca *Open Blocks* para *Java*, que é distribuída pelo *STEP (Scheller Teacher Education Program)* do *MIT (Massachusetts Institute of Technology)*. Uma das vantagens desse

1 Biblioteca para programação visual baseada em blocos.

framework é que pode ser executado de qualquer lugar, ou seja, a plataforma de desenvolvimento está localizada na web, baseando-se nas *cloud computing*². Vale ressaltar que os aplicativos produzidos por ele podem ser *softwares* com arquitetura nativa, como é o caso do *Xô Zika*, que consiste na aplicação elaborada para executar no ambiente do dispositivo móvel.

Um exemplo dessa arquitetura é a utilização do *TinyDB*, um dos bancos de dados que o *framework* utiliza. Ele é descarregado no dispositivo móvel toda vez que a aplicação é instalada ou atualizada. Durante a execução do aplicativo, para a interação com a tela de usuário, as perguntas são buscadas desse banco de dados, não sendo necessário conexão com a internet.

4. Resultados e Discussão

O aplicativo *Xô Zika* é uma iniciativa que visa disseminar informações à solução de problemas relacionados à área social, de saúde pública e educacional, utilizando como meio a tecnologia aplicada nos *smartphones e tablets*. Através de projetos desta natureza, pode-se sensibilizar e provocar uma mudança de comportamento no cidadão a partir de sua interação e investigação, possibilitando que autoridades saibam, quase em tempo real, sobre possíveis criadouros de mosquitos em sua região. O desenvolvimento de aplicativos com propósitos educacionais na área de saúde, especificamente treinamento de inspeção de focos da dengue, mostrou-se uma tarefa inicialmente complexa devido a falta de uma metodologia específica, voltada para o auxílio do Plano de Erradicação do *Aedes aegypti* na área digital. Uma característica particular do presente projeto, é o fato de ele não estar associado a nenhuma outra iniciativa educacional, e portanto de caráter inovador no que se refere a denúncias *on line* de focos de dengue. Espera-se cativar o usuário pelo jogo *quiz*, que está repleto de informações sobre a dengue, que são renovadas e incrementadas periodicamente, permitindo a interação com autoridades municipais responsáveis pelas providências. Observou-se, por outras experiências relatadas na literatura, bem como pelo desenvolvimento deste projeto, que não há uma metodologia apropriada para este tipo de atividade, mas que a Saúde Pública pode se beneficiar muito de ambientes de treinamento, usando a tecnologia de forma simples para melhorar índices de saúde.

Segundo Souza *et all* (2015), responsáveis pela avaliação da ferramenta denominada de “*Avaliação de Mobilidade Precoce*”, aplicativos desenvolvidos com a plataforma *App Inventor* podem ser tão úteis quanto outros aplicativos utilizados na área da saúde para a conscientização da população frente a alguma doença.

As denúncias recebidas pela equipe desenvolvedora, sobre possíveis criadouros do mosquito, são prontamente redirecionadas para o setor de controle de dengue do respectivo município.

Ainda está em etapa de finalização e validação através da escala de REEVES, que utiliza critérios em relação à interface com o usuário (14 critérios) e em relação aos aspectos pedagógicos do aplicativo (10 critérios). Esse método avalia o aplicativo com base na pontuação em escala de conceitos antagônicos, levando em conta critérios relacionados à interface com o usuário (facilidade de navegação, *design* da tela, compatibilidade espacial do conhecimento, apresentação da informação, estética e funcionalidade) e critérios pedagógicos (epistemologia, sequenciamento instrucional, validade experimental, valorização do erro, estruturação, controle do aluno e aprendizado cooperativo). A avaliação está sendo realizada continuamente durante o

2 Computação em nuvens.

desenvolvimento do *aplicativo* pela equipe responsável, a fim de assegurar os objetivos e metas propostos originalmente.

5. Conclusões

Neste trabalho foram apresentadas as etapas do desenvolvimento de um aplicativo *Android* em forma de jogo de perguntas e respostas com a possibilidade de efetuar denúncias anônimas sobre possíveis criadouros do *Aedes aegypti*: o *Xô Zika*. Acredita-se que ele auxiliará estudantes do ensino médio, professores e profissionais da área de Saúde Pública a aumentarem seus conhecimentos sobre esse problema. Cabe ressaltar o papel motivador do número de acertos nas perguntas, trazendo conteúdos significativos e compreensíveis, além de métodos adequados de combate ao mosquito, o que possivelmente auxiliará no processo de ensino e aprendizagem. Destaca-se ainda que o contato do aluno com o desenvolvimento de aplicativo para dispositivos móveis poderá proporcionar-lhe uma futura oportunidade de inserção no mercado de trabalho.



Figura 1: Telas do aplicativo

Referências

- Fontes, M. C. M. (2002) “Aprendizagem de inglês via Internet: descobrindo as potencialidades do meio digital.” São Paulo: PUCSP. Tese (Doutorado em linguística aplicada e estudos da linguagem), Pontifícia Universidade Católica de São Paulo.
- Gartner (2013) <http://olhardigital.uol.com.br/pro/noticia/samsung-detem-mais-da-metade-do-mercado-de-smartphones-no-brasil/40282>, Março.
- Jana (2015) <http://www.tudocelular.com/android/noticias/n61396/Android-Brasil-Mercado.html>, Março.
- Lee, V.; Schneider, H.; Schell, R. (2005) “Aplicações móveis: arquitetura, projeto e desenvolvimento.” Pearson, São Paulo.
- Ministério da Saúde (2015) <http://www.portalsaude.saude.gov.br>, Março.
- Souza, J. F.; Gonçalves F. B.; Queiroz, V. A. R.; Queiroz, R. S. (2015) "Avaliação de um aplicativo para auxílio à tomada de decisão de mobilizar pacientes críticos." Revista Saúde.com <http://www.uesb.br/revista/rsc/v11/v11n1.html>, Maio.

HackDay@IFC-Cam: evento de preparação para competições de programação como instrumento de ensino

Ana Elisa Ferreira Schmidt¹, Mozara Dias Koehler¹

¹ Instituto Federal Catarinense – Campus Camboriú (IFC)
Caixa Postal 2016 – 88340-055 – Camboriú – SC – Brasil

{anaelisa,mozara}@ifc-camboriu.edu.br

***Abstract.** This article presents the HackDay @ IFC-Cam experience report, an event that takes place in a single day and where participants learn to think about, design, develop and present a mobile application, for Android, using AppInventor as programming tool. As one of the HackDay@IFC-Cam's positive results, we mention an app developed by our students that classified among the top 10 Apps in the Maratona de Aplicativos FIAP/Google, 2015 edition.*

***Resumo.** O presente artigo apresenta o relato de experiência da realização do HackDay@IFC-Cam, evento que se realiza em um único dia e onde os participantes aprendem a idear, projetar, desenvolver e apresentar um aplicativo para dispositivo móvel, com sistema operacional Android, utilizando a ferramenta de programação AppInventor. Um dos resultados positivos deste evento foi a classificação do aplicativo desenvolvido por uma equipe do Técnico em Informática entre os 10 melhores na Maratona de Aplicativos FIAP/Google, edição 2015.*

1. Introdução

O HackDay é uma atividade, geralmente, que possui duração de aproximadamente um dia e propõe aos participantes (sozinhos ou organizados em equipes) que realizem um projeto tendo como base um desafio [Lima, 2012]. Pires e Erlich (2015) descrevem o HackDay como a liberdade criativa dentro do ambiente corporativo em uma maratona de programação computacional de apenas um dia, onde os funcionários atuam colaborativamente na criação de projetos livres, apresentando-os na forma de protótipos rápidos de software contendo os elementos essenciais de suas ideias.

Visando incentivar a criatividade, espírito inovador, aprendizado de novos conceitos e consolidação de conhecimentos aprendidos academicamente, em 2014 organizou-se um HackDay com alunos do curso Técnico em Informática do IFC-Cam. Este evento tinha também o propósito de incentivar os alunos a submeterem suas idéias e projetos para a Maratona de Aplicativos FIAP/Google 2014 [FIAP/ Google, 2015].

A realização de um HackDay segue uma sequência de momentos específicos: 1) são definidos os grupos de trabalho; 2) temos a etapa de ideiação e escolha da idéia vencedora a ser desenvolvida pelo grupo; 3) etapa de brainstorming sobre funcionalidades da idéia vencedora. 4) aprendizado da ferramenta de desenvolvimento -

App Inventor; 5) criação do protótipo do software utilizando a ferramenta de desenvolvimento; 6) a apresentação do projeto desenvolvido aos colegas e avaliadores.

A figura 1 mostra os momentos de ideiação, desenvolvimento do protótipo, confraternização e apresentação dos projetos durante o HackDay@IFC-Cam 2014.



Figura 1. Hackday – ideiação, programação, apresentação e confraternização.

2. Maratona de aplicativos

A Maratona de aplicativos FIAP/ Google é uma competição entre estudantes do Ensino Médio em que os participantes deverão criar aplicativos que possam melhorar a vida deles como alunos, sua escola ou a educação no Brasil. Em 2015 foi terceira edição do evento que é organizado pela FIAP em parceria com o Google e a HP [FIAP/ Google, 2015].

Esta maratona é um evento aberto somente aos estudantes do nível médio (ensino médio). Os participantes podem reunir-se em grupos de no mínimo 3 (três) e no máximo 5 (cinco) pessoas, que juntos são denominados simplesmente como “EQUIPE”.

A maratona consiste no desenvolvimento de aplicativos para dispositivos móveis que contenham o sistema operacional Android. A Equipe deve utilizar, obrigatoriamente, a ferramenta AppInventor versão 2 [AppInventor, 2016] para o desenvolvimento do aplicativo e seguir o tema: “Desenvolva um aplicativo que ajude você como aluno, a sua escola ou a educação no Brasil” [FIAP/ Google, 2015].

De acordo com a FIAP/ Google (2015), contribuir com a real inclusão digital dos jovens, transformando-os de meros usuários a desenvolvedores de aplicativos é o principal objetivo. Além disso, a FIAP acredita que compartilhar conhecimento é a chave para alavancar a educação no Brasil e encorajar novos empreendedores a fazer a diferença em nossa sociedade.

3. Ferramenta de desenvolvimento: AppInventor

MIT App Inventor é uma ferramenta de programação visual inovadora permite de introduzir iniciantes na programação e criação de aplicativos. Esta ferramenta transforma a complexidade das linguagens de programação textuais em uma operação de clicar e arrastar de blocos de instruções visuais que em conjunto formam um programa. A interface gráfica simples permite que mesmo um novato sem experiência

em programação tenha a habilidade de criar aplicativos básicos totalmente funcionais dentro de um período de 01 hora ou menos. [MIT, 2016].

O projeto do MIT App Inventor visa democratizar o desenvolvimento de software por capacitar todas as pessoas, especialmente os jovens, para fazer a transição de consumidores de tecnologia para criadores do mesmo.

4. HackDay como instrumento de ensino

Ao realizar um HackDay muitas ideias surgem e muitas dicas dos outros colegas são sugeridas para as propostas apresentadas entre eles, tornando os alunos agentes ativos na melhoria da qualidade da educação. O olhar para o problema com a finalidade de resolvê-lo pode ser um caminho para que o aluno se interesse mais pelo ambiente escolar. Neste contexto, os alunos reconhecem na prática a aplicação dos conhecimentos aprendidos academicamente, compartilham seus saberes e expõem suas dificuldades, levando o ensino-aprendizado para muito além das barreiras da sala de aula.

Um grande número de alunos socializando suas ideias para melhorar o cotidiano de outros alunos é um momento valioso. O HackDay pode proporcionar este espaço de interação e aprendizagem entre grupos com realidades, rotinas e pensamentos distintos.

A utilização do AppInventor no aprendizado de algoritmos, programação, design e projeto de interfaces é descrito com êxito por Gomes(2013) e Duda (2015). É uma ferramenta de programação bastante versátil para ensino, não somente de conteúdos relacionados à informática, mas também de outras áreas do conhecimento.

5. HackDay no IFC-Camboriú: resultados obtidos

A realização do HackDay@IFC-Cam incentivou alunos do Técnico em Informática a aprofundar seus conhecimentos em metodologias de ideiação, ferramentas de gerência de projeto, programação para dispositivos móveis, projeto de interfaces e desenvolvimento de habilidades de apresentação em público e elaboração de artigos. Além disso, criou uma atmosfera de integração interdisciplinar com as demais áreas do conhecimento de onde surgem as mais diferentes demandas por soluções de software através de aplicativos móveis.

No primeiro ano da participação do IFC-Camboriú na Maratona de Aplicativos FIAP/Google, 2014, foram 04 apps enviados por equipes de alunos do Técnico Informática do IFC-Cam, mas não tivemos classificados para o final. No segundo ano, 2015, foram 02 apps enviados e obtivemos o feliz resultado de ter uma equipe classificada para final entre mais de 9600 inscritos e 450 apps do Brasil inteiro. O aplicativo classificado “Formação de Campo para Basquete” [Koehler, 2015] foi desenvolvido pelos alunos Rafael Andrade, Nicolas Oliveira e Gabriel Martins e surgiu da necessidade apontada pela Prof. Alba Garcia, professora de educação física do IFC.

A viagem à São Paulo para apresentação ao vivo do aplicativo dentro da sede da Google Brasil, bem como a palestra oferecida dentro Samsung Ocean [Samsung, 2016] foram momentos muito especiais para os alunos do curso técnico em informática e os transformaram em pequenas celebridades escolares, pois foram várias entrevistas publicadas em jornais locais e regionais, entrevistas ao vivo em programa local de televisão e divulgação no site institucional. A figura 2 mostra alguns destes momentos.

Esta visibilidade advinda de um projeto escolar foi extremamente positiva, pois veio motivar outros alunos a participarem na edição 2016 desta e de outras competições da área da informática, demonstra que buscar alternativas de ensino para melhorar a relação ensino-aprendizagem é extremamente gratificante.



Figura 2- Momentos na Maratona de Aplicativos FIAP/Google 2015.

6. Referências

- Duda, Rodrigo; Silva, Sani de Carvalho Rutz da. Desenvolvimento de aplicativos para android com uso do app inventor: uso de novas tecnologias no processo de ensino-aprendizagem em matemática. Revista Conexão Uepg, Ponta Grossa, v. 11, n. 3, p.310-323, 28 ago. Março 2015.
- FIAP/GOOGLE. Maratona de aplicativos 2015. 2015. <http://www.maratonadeaplicativos.com.br/tutoriais>. Novembro 2015.
- Garcia, Rogério Eduardo; Correia, Ronaldo Celso Messias; Shimabukuro, Milton Hirokazu. Ensino de Lógica de Programação e Estruturas de Dados para Alunos do Ensino Médio. In: Anais do XXVIII Congresso da SBC, Belém do Pará, PA, 2008.
- Gomes, Tancicleide C. S.; Melo, Jeane C. B. de. App Inventor for Android: Uma Nova Possibilidade para o Ensino de Lógica de Programação. Anais dos Workshops do Congresso Brasileiro de Informática na Educação. Porto Alegre: Sociedade Brasileira de Computação, 2013. v. 1, p. 620 - 629.
- Lima Junior, Walter Teixeira. Big Data , Jornalismo Computacional e Data Journalism : estrutura, pensamento e prática profissional na Web de dados. 2012. http://sbpjour.kamotini.kinghost.net/sbpjour/admjor/arquivos/10encontro/comunicacoes_coordenadas/walter_teixeira_lima_junior.pdf. Novembro 2015.
- Koehler, Mozara; Garcia, Alba; Oliveira, Nicolas; Andrade, Rafael. (2015) Aplicativo de Formação de Campo para Auxiliar Treinadores e Professores no Esquema Tático do Jogo. Anais online da VIII Mostra Nacional de Iniciação Científica e Tecnológica Interdisciplinar (MICTI). <http://eventos.ifc.edu.br/micti/anais-2015/>. Março 2016.
- MIT, App Inventor. <http://appinventor.mit.edu/explore/index-2.html>. Março 2016.
- Pires, Cristiana Leal de Lacerda; Erlich, Márcio Barbosa. Prototipagem Urbana: reativando espaços públicos através de maratonas colaborativas de fabricação digital. In: Congresso Internacional Espaços Públicos, 1., Porto Alegre: Edipucrs, 2015.
- Samsung Ocean. <http://www.oceanbrasil.com/>. Março 2016.

e-Sinais: *Software* Tradutor de Português Sinalizado para Sinais em LIBRAS

Igor Q. Silva¹, Bruno S. Mascarenhas², Saionara S. Araújo³,
Daniela G. S. Freitas⁴, Pablo F. Matos⁵

¹Discente Superior em Ciência da Computação, ²Discente Técnico em Eletrônica Integrado, ³Discente Superior em Sistemas de Informação, ⁴Intérprete de LIBRAS, ⁵Docente de Informática

¹Universidade Estadual do Sudoeste da Bahia (UESB)
Estrada Bem Querer – 45.083-900 – Vitória da Conquista – BA – Brasil

²³⁴⁵Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA)
Av. Amazonas, 3150, Zabelê – 45.075-265 – Vitória da Conquista – BA – Brasil

{igorqsilva, brunomask, narabdo, danielasilveira08, pablofmatos}@gmail.com

Abstract. *The e-Signs is an educational software that performs the translation of Signaled Portuguese for signs in LIBRAS and it aims providing to deaf the learning of new words. The idea is that deaf will be independent of a human translator/interpreter in the day-to-day tasks and he may perform their studies or to work with the help of the tool. Moreover, the software may also serve as teaching tool of LIBRAS for listeners, in which it helps the integration of deaf and hearing communities. The e-Signs offers other features such as the addition of new signs inherent of each region and the export of signs to PDF file, which it allows the printing and distribution of signs.*

Resumo. *O e-Sinais é um software educacional que realiza a tradução de Português Sinalizado para sinais em LIBRAS e tem como objetivo proporcionar aos surdos o aprendizado de novas palavras. A ideia é que nas tarefas do dia a dia o surdo seja independente de um tradutor/intérprete humano e consiga realizar seus estudos ou trabalhar com o auxílio da ferramenta. Ademais, o software pode servir também como ferramenta de ensino de LIBRAS para ouvintes, em que auxilia a integração das comunidades surda e ouvinte. O e-Sinais disponibiliza outros recursos como a adição de novos sinais inerentes de cada região e a exportação de sinais para arquivo PDF, que permite a impressão e a distribuição dos sinais.*

1. Introdução

A Língua Brasileira de Sinais (LIBRAS), utilizada pelos surdos, é uma língua natural e, portanto, apresenta os seguintes níveis linguísticos: fonológico, morfológico, sintático e semântico [1]. A LIBRAS é diferenciada da Língua Portuguesa por sua forma de expressão ser visual-espacial e também por possuir estruturas gramaticais diferentes. Neste trabalho, considera-se a expressão Português Sinalizado como sendo um sistema de notação em palavras que tem como característica a apresentação dos verbos no infinitivo e a inexistência de alguns conectivos (conjunções e preposições). Essas diferenças são um dos pontos que dificultam a compreensão, por um surdo, de um texto escrito em Português.

Segundo censo de 2010 pelo IBGE [2], cerca de 9,7 milhões de brasileiros possuem deficiência auditiva, o que representa 5,1% da população total. Na população com deficiência, 61,1% não possuem o fundamental completo, 14,2% possuem o fundamental completo, 17,7%, o médio completo e 6,7%, o superior completo [3]. Segundo a FENEIS [4], quase dois milhões de brasileiros surdos não são alfabetizados em Português escrito, que é sua segunda língua, e se comunicam exclusivamente por meio da LIBRAS. A falta de acessibilidade a textos escritos, além de restringir o acesso à informação para a comunidade surda, é um grande agente desmotivador.

Há alguns *softwares* que lidam com a tradução textual de Português para Português Sinalizado [5, 6, 7]. Porém, o uso desta funcionalidade exige dos surdos um conhecimento de compreensão dos textos em Português, o que não é constatado na prática, já que os surdos necessitam de uma metodologia adequada para compreender o Português escrito [8].

O *software* LIBROL [5], desenvolvido pelo nosso grupo de pesquisa, visa a realizar a tradução de textos em Português para o Português Sinalizado, mais simples de ser compreendido pelo surdo. Após o desenvolvimento do *software* LIBROL, foi realizada uma avaliação da compreensão de textos jornalísticos em Português, em LIBROL e em Português Sinalizado por Estudantes Surdos [9]. Percebeu-se que, com o uso da ferramenta, houve uma melhora na compreensão dos textos escritos por parte dos leitores surdos. Porém, em alguns casos, os surdos não compreendiam o texto, nem em Português Sinalizado, nem em LIBROL, muito menos em Português escrito. Isso aconteceu devido ao desconhecimento das palavras presentes nos textos.

Nesse contexto, surgiu a ideia de criar um *software* complementar ao LIBROL, denominado e-Sinais, *software* tradutor de Português Sinalizado para sinais em LIBRAS, que tem o intuito de possibilitar ao estudante surdo aprender novas palavras independentemente de um tradutor/intérprete humano, o que contribui para aumentar o vocabulário e, por conseguinte, a compreensão dos textos. Ademais, o *software* pode ser usado como ferramenta de ensino de LIBRAS para ouvintes, o que pode propiciar a integração das comunidades surda e ouvinte. O restante deste artigo está estruturado como segue. Na Seção 2, são resumidos os trabalhos correlatos. Na Seção 3, são apresentadas as características do *software* e-Sinais e na Seção 4, são apresentadas as Considerações Finais e os Trabalhos Futuros.

2. Trabalhos Correlatos

Os trabalhos encontrados na literatura têm como objetivo a tradução entre a Língua Portuguesa e a LIBRAS a fim de auxiliar na comunicação entre surdos e ouvintes [6, 10, 11], a tradução textual do Português para o Português Sinalizado [5, 6, 7] ou o ensino-aprendizagem de LIBRAS [5, 12, 13].

Dentre os trabalhos pesquisados, ressalta-se o *Teaching Hand*, o *Luz, Câmera, Libras!* e o *ProDeaf*. O *Teaching Hand* [13] é um *software* que auxilia crianças surdas no aprendizado da gramática da Língua Portuguesa por meio da associação de imagens e sinalizações. O *Luz, Câmera, Libras!* [12] é um *software* direcionado para o aprendizado de sinais por meio do lúdico, no qual os participantes devem adivinhar os sinais elaborados por outro participante. O jogo se baseia em vídeos feitos pelos jogadores. O *ProDeaf* [6] é o *software* mais completo encontrado na literatura. Contém um conjunto de *software* que possibilita a tradução de texto e voz de Português para LIBRAS, com o principal objetivo de permitir a comunicação entre surdos e ouvintes. Ademais, o *ProDeaf* tem soluções para tradução textual com a possibilidade de incluir sinais e consultar o dicionário de palavras. Porém, a tradução é limitada a 140 caracteres, não é possível exportar os sinais para nenhum tipo de arquivo e não tem o foco no ensino-aprendizagem de LIBRAS.

3. Software e-Sinais

O *software* e-Sinais tem como objetivo possibilitar ao surdo interpretar palavras do Português Sinalizado para sinais em LIBRAS, o que viabiliza a aprendizagem de novas palavras com o intuito de obter a independência de um tradutor/intérprete humano nas tarefas do dia a dia, seja na vida profissional ou acadêmica. O e-Sinais tem como entrada de dados palavras em Português Sinalizado e como saída os sinais correspondentes em LIBRAS no formato de imagem (Figura 1). O e-Sinais foi desenvolvido utilizando a Linguagem de Programação Java [14] por ser gratuita e portátil para qualquer sistema operacional. Está disponível sob a licença GNU AGPL e pode ser baixado em <https://sourceforge.net/u/e-sinais/profile/>.

A tradução é realizada em três etapas: (i) divisão do texto em palavra por palavra; (ii) busca destas palavras no banco de dados; (iii) apresentação do sinal retornado pelo banco de dados na tela do *software*. A divisão do texto em várias palavras separadas é a primeira etapa da tradução. Para realizar essa tarefa, foi necessário criar um *array* de variáveis do tipo *string* com as palavras isoladas. O sistema de gerenciamento de Banco de Dados Relacional (BDR) utilizado foi o HSQLDB [15], escolhido por ser desenvolvido em Java, por ser pequeno, por executar na memória principal, sendo assim não necessita de instalação e, por fim, por ser de fácil utilização. A tabela criada no banco de dados possui dois atributos: “*palavra*”, na qual são armazenadas as palavras em Português Sinalizado; e “*sinal*”, na qual são armazenados os endereços das imagens contendo os sinais.



Figura 1. Interface do software e-Sinais.
Fonte. Próprio autor.

Para realizar a busca no banco de dados, foi utilizado o seguinte comando SQL (Linguagem de Consulta Estruturada):

```
SELECT sinal FROM public.tabela WHERE palavra = 'x';
```

Os atributos “*sinal*” e “*palavra*” representam as colunas, a tabela “*public.tabela*” se refere à tabela onde os dados estão armazenados e, 'x' representa a palavra sobre a qual será realizada uma busca no banco de dados. Este comando retornará o endereço em que se encontra o sinal correspondente à palavra 'x'. Para finalizar o processo da tradução, o endereço retornado é utilizado para carregar a imagem e apresentá-la na tela do *software*. Os sinais podem ser armazenados no BDR por imagens estáticas (JPG e PNG) ou por imagens animadas (GIF).

A versão atual do *software* contém 811 sinais no formato de imagem animada GIF, as quais foram obtidas do Glossário de LIBRAS do Incluir Tecnologia¹. Foi necessário converter os arquivos disponíveis em vídeos para o formato GIF e, em seguida, redimensionar as imagens. Os sinais que representavam frases ou palavras compostas e os verbos conjugados foram excluídos. Com relação às palavras que possuíam mais de uma representação em LIBRAS (sinônimos), decidiu-se utilizar somente um sinal e excluir os outros. Ademais, para palavras diferentes que possuíam o mesmo sinal, optou-se por nomear o sinal com um código único e relacioná-lo a todas as palavras correspondentes.

Além do recurso de tradução, o e-Sinais possui mais dois outros importantes recursos: *adição de sinais* e *exportação de sinais para PDF*. O primeiro recurso possibilita ao usuário adicionar um novo sinal ou alterar um sinal cadastrado. Isso é útil para suprir a falta de algum sinal no banco de dados, o que viabiliza o acréscimo de sinais inerentes de cada região. O segundo recurso permite que as palavras traduzidas para sinais sejam exportadas para um arquivo PDF (Formato Portátil de Documento), o que permite tanto que os sinais sejam distribuídos para os que não possuem o *software* quanto sejam impressos e distribuídos em salas de aula e demais ambientes a fim de viabilizar a acessibilidade aos surdos.

Para imagens estáticas no formato JPG e PNG, o programa apenas gera uma imagem do sinal no arquivo PDF. Porém, para imagem no formato GIF, o programa gera uma série de imagens sequenciais que são selecionadas de 4 em 4 *frames*. Optou-se por esta implementação para que o arquivo PDF não fique muito extenso, o que inviabiliza a tradução de sequências longas de palavras. É importante destacar que não há restrição de quantidade de palavras a serem traduzidas. É possível mostrar uma quantidade ilimitada de sinais devido à criação dinâmica de componente, no qual são mostradas as imagens. O recurso de exportação para PDF foi implementado com a biblioteca iTextPDF² que é utilizada para a criação de arquivos em PDF utilizando a linguagem de programação Java.

¹ <https://www.youtube.com/user/incluirtecnologia>

² <http://itextpdf.com/>

4. Considerações Finais e Trabalhos Futuros

O e-Sinais é um *software* educacional que tem o intuito de proporcionar aos surdos o aprendizado de novas palavras, o que permite a independência do tradutor/intérprete humano nas tarefas do dia a dia, seja na vida profissional ou acadêmica. O *software* possui dois relevantes recursos, a saber: (i) adição de sinais e (ii) exportação de sinais para PDF. Estes recursos viabilizarão, respectivamente, ampliar a base de dados com a inclusão de sinais inerentes de cada região e exportar os sinais para um arquivo PDF, o que possibilitará ao surdo estudar com um material digital ou impresso e compartilhar este material com outros surdos. Ademais, o material exportado poderá ser distribuído nas salas de aula e outros ambientes a fim de garantir a acessibilidade aos surdos. É importante ressaltar que o recurso de exportação de sinais para PDF não foi encontrado nos trabalhos correlatos. O e-Sinais também tem como utilidade o ensino de LIBRAS para pessoas ouvintes. Como trabalhos futuros, pretende-se adicionar mais palavras (simples e compostas) e seus correspondentes sinais ao banco de dados, permitir o armazenamento de palavras sinônimas e possibilitar ao usuário a seleção do sinal mais adequado. Também se tem o intuito de realizar um experimento com estudantes surdos, para corroborar a funcionalidade do *software* no ensino-aprendizagem da Língua Portuguesa escrita.

Agradecimentos

Agradecemos o apoio financeiro concedido pelo IFBA por meio do Programa Institucional de Bolsa de Iniciação Científica para o Ensino Médio (PIBIC-EM) – Edital N° 03/2014/PRPGI e ao professor de português do IFBA, Sinval Araújo de Medeiros Júnior, pela revisão do artigo.

Referências

1. Souza, T.A.F. and Monteiro, M.S. (2008) **Libras em Contexto: Curso Básico** - Livro do Professor. WalPrint Gráfica e Editora, Rio de Janeiro.
2. ADAP. (2013) **Deficiência Auditiva Atinge 9,7 milhões de Brasileiros**, <http://goo.gl/l5pZVk>.
3. Oliveira, L.M.B. (2012) **Cartilha do Senso 2010: Pessoas com Deficiência**, <http://goo.gl/AjeimR>.
4. FENEIS. (2016) **Federação Nacional de Educação e Integração dos Surdos**, <http://www.feneis.org.br/>.
5. Carvalho, R.S. *et al.* (2013) LIBROL: Software Tradutor de Português para LIBRAS. In: **ENCompIF**, pages 2098-2101. CSBC.
6. Oliveira, J.P.S. (2012) ProDeaf: Uma Ferramenta Colaborativa para a Produção de Conhecimento em Libras. In: **INES**, pages 1-6, <http://www.prodeaf.net/>.
7. Nunes, M.G.V. *et al.* (2003) **Projeto Pulo**. USP, São Carlos.
8. Guarinello, A.C. *et al.* (2009) Surdez e Letramento: Pesquisa com Surdos Universitários de Curitiba e Florianópolis. **Revista Brasileira de Educação Especial**, 15, 99-120.
9. Silva, I.Q. *et al.* (2014) Avaliação da Compreensão de Textos Jornalísticos em Português, em LIBROL e em LIBRAS por Estudantes Surdos. In: **II ENCompIF**, pages 718-721. CSBC.
10. HandTalk. (2016) **Hand Talk Tradutor para Libras**, <http://www.handtalk.me/>.
11. Januário, G.C., Leite, L.A.F. and Koga, M.L. (2010) **Poli-Libras: Um Tradutor de Português para Libras**. 93 f. TCC (Bacharelado em Eng. da Comp.) – USP, São Paulo.
12. Moura, G. *et al.* (2013) Luz, Câmera, Libras!: How a Mobile Game Can Improve the Learning of Sign Languages. In Marcus, A. (Ed.), **DUXU/HCI**. LNCS, Springer, pages 266-275.
13. Santos, P.V. *et al.* (2015) Teaching Hand: Uma Plataforma Educacional para o Ensino da Língua Portuguesa para Crianças Surdas. In: **IV CBIE - X LACLO**, pages 441-446. CSBC.
14. Deitel, H.M. and Deitel, P.J. (2010) **Java: Como Programar**. Pearson, São Paulo.
15. HyperSQL. (2016) **HSQldb - 100% Java Database**, <http://hsqldb.org/>.

INASH: Uma Ferramenta de Autoria para Narrativas Interativas Digitais

Vinicius Hartmann Ferreira, Luís Guilherme Eich

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Campus
Feliz
Feliz – RS – Brasil

vinihf@gmail.com , gui.eich@hotmail.com

Abstract. *This article presents the INASH tool (Interactive Narratives Shared), which seeks to encourage the practice of reading and writing in students by building interactive narratives. The tool allows the user to create their own narratives, by providing alternative ways for readers, and share them with other users. An experiment with the use of the tool was performed with a technical high school class consists of 24 students. At the end of the experiment the students answered a questionnaire by which it was possible to identify that the tool displays elements that can stimulate students' interest in writing and reading, as the graphical interface construction and its production sharing.*

Resumo. *Este artigo apresenta a ferramenta INASH (Interactive Narratives Shared), que busca estimular a prática da leitura e da escrita em estudantes através da construção de narrativas interativas. A ferramenta permite ao usuário criar suas próprias narrativas, estipulando caminhos alternativos para os leitores escolherem durante a leitura, e compartilhá-las com os demais usuários. Um experimento com o uso da ferramenta foi realizado com uma turma de ensino médio técnico composta por 24 alunos. Ao final do experimento os alunos responderam a um questionário através do qual foi possível identificar que a ferramenta apresenta elementos que podem estimular o interesse dos alunos pela escrita e pela leitura, como a interface gráfica de construção e o compartilhamento de suas produções.*

1. Introdução

A escrita e a leitura são habilidades importantes em todas as áreas de conhecimento. Porém, testes internacionais, como o PISA, indicam que o nível de leitura e escrita dos alunos com 15 anos no Brasil está abaixo da média internacional, alcançando pontuação inferior a países da América do Sul como Chile e Uruguai. Além disso, também foi possível constatar através do PISA que os alunos estão lendo menos por prazer, embora aqueles que indicaram ler por prazer tenham obtido melhor desempenho no teste (OECD, 2013).

Aproximar a escrita e a leitura daquilo que atrai os alunos de ensino médio é uma estratégia relevante para estimulá-los a sua prática (Carbonaro *et al.*, 2008). Neste contexto, narrativas interativas surgem como uma possibilidade de relacionar a produção textual e a leitura aos jogos digitais, um tema que desperta o interesse dos adolescentes. As narrativas interativas caracterizam-se por não apresentar uma condução linear da narração, dando ao leitor a possibilidade de fazer escolhas que irão alterar o seu rumo. É neste ponto que as narrativas interativas apresentam similaridade aos jogos digitais, dando ao leitor a sensação de estar jogando.

Partindo disto, este artigo apresenta uma ferramenta digital para autoria de narrativas interativas digitais, intitulada INASH (Interactive Narratives Shared), desenvolvida por um aluno de curso técnico em informática integrado ao ensino médio. O artigo está estruturado de forma que na seção 2 são apresentados conceitos fundamentais para a compreensão da importância das TICs no apoio a escrita e a leitura, na seção 3 é apresentada a ferramenta desenvolvida, na seção 4 são apresentados a descrição de um experimento e os resultados obtidos e na seção 5 são explicitadas as conclusões obtidas com o experimento.

2. As TICs no Apoio a Escrita e a Leitura

A utilização de novas tecnologias para produção textual, e seu amplo domínio, podem trazer benefícios para um aluno, permitindo que ele consiga compreender como melhorar seu texto. Esta nova forma de desenvolver a produção textual vem alterando a forma como um aluno lê, escreve e compreende um texto. Observa-se isso através da forma como ele agora organiza suas ideias e com a facilidade que encontra para compartilhar suas produções (Cordero *et al.*, 2015).

A partir de trabalhos que envolvem o desenvolvimento ou aplicação de ferramentas digitais como suporte ao letramento ou ao estímulo à escrita (Lysenko e Abrami, 2014; Azeredo e Reategui, 2013; Carbonaro *et al.*, 2008), verifica-se que a utilização de tais recursos nestes processos apresenta resultados positivos, destacando o papel de autor dos alunos. É possível constatar que as ferramentas utilizadas e desenvolvidas variam entre aplicações para Desktop e para a Web e abrangem diferentes contextos, como o ensino ou suporte a prática de escrita em língua nativa ou língua estrangeira.

3. A Ferramenta INASH

A ferramenta INASH foi construída como um sistema para a Web, utilizando tecnologias como HTML, PHP, CSS e Javascript, e possui duas modalidades de interação: (i) escrita; e (ii) leitura. Na modalidade de escrita o usuário pode construir uma narrativa interativa digital utilizando o editor visual (Figura 1) e compartilhá-la com os demais usuários do sistema. Na modalidade de leitura o usuário pode ler as produções dos demais usuários e avaliá-las.

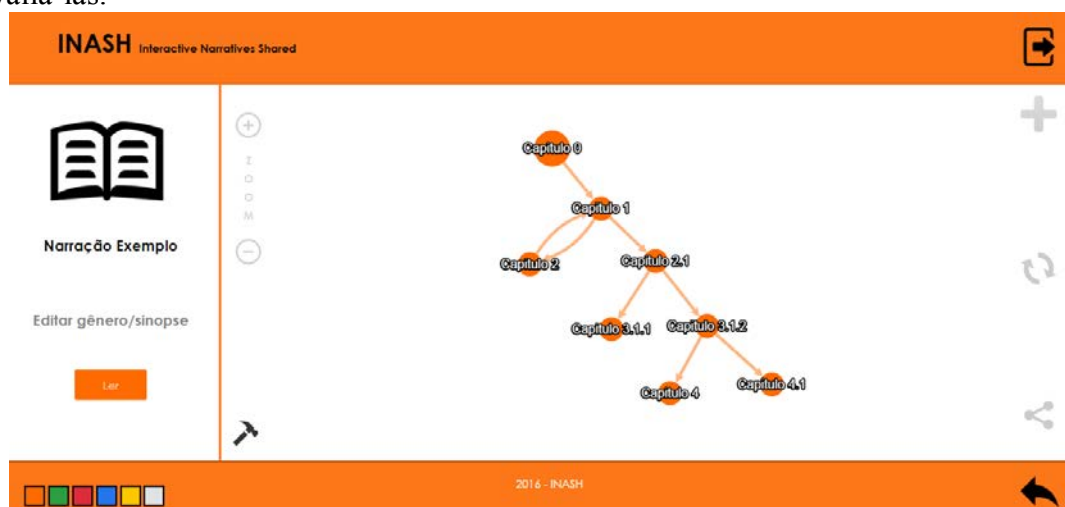


Figura 1. Editor visual da ferramenta INASH.

O editor visual representa a narrativa na forma de um grafo no qual cada vértice é um capítulo e cada aresta é um caminho possível entre os vértices. A Figura 1 apresenta os recursos disponíveis no editor, sendo eles a inclusão de um novo capítulo, um recurso para zoom no grafo, exclusão de capítulos e compartilhamento da produção. Para editar cada um dos capítulos clica-se na aresta desejada e um editor de textos simples é apresentado para o usuário. As relações entre os capítulos são construídas quando o usuário arrasta uma aresta sobre a outra. Desta forma, a lógica sequencial da narrativa é construída e visualizada na forma de um grafo.

A leitura das narrativas produzidas também é realizada por meio da ferramenta Web. Conforme pode ser visto na Figura 2, a ferramenta apresenta um capítulo escrito pelo usuário e ao final da página as opções que o leitor tem para seguir na história. Na parte esquerda da tela de leitura é apresentada ao leitor a estrutura dos capítulos da narrativa e a possibilidade de o leitor avaliá-la, indicando se sua opinião sobre a narrativa é positiva ou negativa.

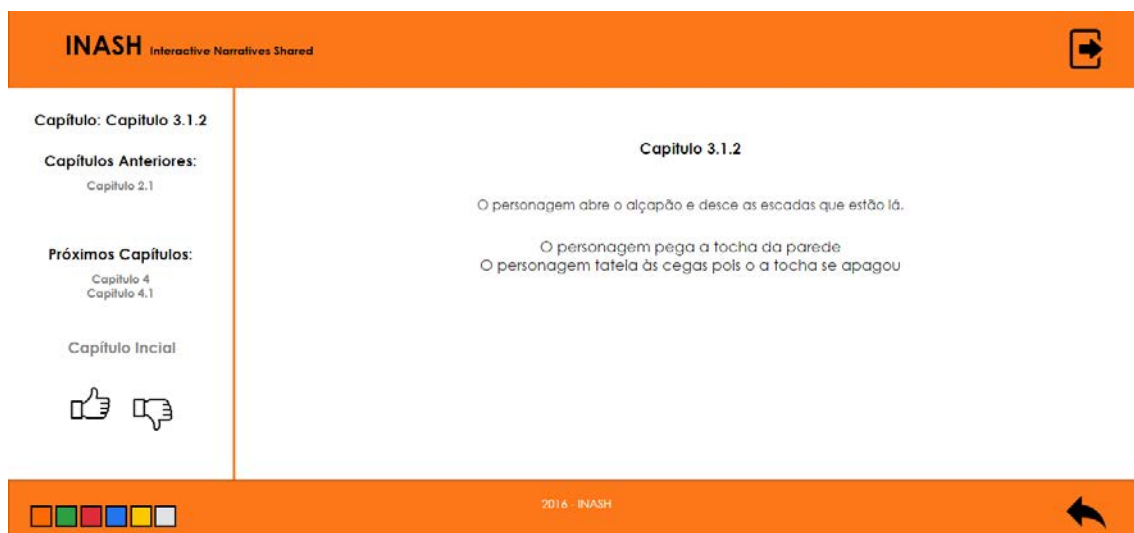


Figura 2. Leitura de uma narrativa na ferramenta INASH.

4. Experimento e Resultados

Com o objetivo de avaliar o funcionamento da ferramenta construída e também o seu potencial, mesmo que de forma inicial, no estímulo a prática da escrita e da leitura, um experimento foi realizado com turma de primeiro ano (24 alunos) de um curso técnico integrado ao ensino médio. Neste experimento, que teve a duração de uma aula (45 minutos), os alunos foram conduzidos na utilização da ferramenta para criar uma pequena narração e compartilhá-la e também ler as produções dos colegas e avaliá-las. Ao final do experimento os alunos responderam a um questionário que buscava identificar a recepção deles a ferramenta, possíveis dificuldades ou sugestões.

Quanto à escrita, por meio do questionário foi possível identificar que 91,7% dos alunos consideraram o editor visual como positivo na construção das narrativas. A maioria dos estudantes (83,3%) também indicou que a produção através do editor visual é mais atrativa do que a realizada por meios tradicionais e 75% dos estudantes indicaram que se sentiram mais motivados para produzir um texto por meio do editor visual.

Quanto à leitura, 91,7% dos estudantes consideraram muito boa a forma como a leitura das narrativas produzidas é realizada na ferramenta. Além disso, apenas um dos estudantes que participou dos testes não leu e não avaliou as produções dos colegas.

Durante a condução do experimento também foi realizada observação do comportamento dos estudantes. Como não havia um tutorial da ferramenta verificou-se que a etapa na qual eles mais precisaram de auxílio para realizar a atividade foi a de criar relações entre os capítulos arrastando-os uns sobre os outros. Além disso, foram identificados textos explicativos na ferramenta que precisavam de maior clareza. Por fim, ainda que o tempo do experimento tenha sido razoavelmente curto foi possível ver que aqueles alunos que já possuem o hábito de escrever narrativas ou ler livros de ficção se interessaram mais pela ferramenta do que os alunos que não possuem estes hábitos.

5. Conclusão

Este artigo apresentou uma ferramenta de autoria para narrativas interativas digitais desenvolvida em um projeto de pesquisa por um aluno de curso técnico em informática integrado ao ensino médio. Além da ferramenta, também foram apresentadas informações sobre a percepção dos estudantes ao usar a ferramenta em um experimento guiado.

Ainda que o intuito do projeto seja estimular a prática da escrita e da leitura em alunos de ensino médio por meio da ferramenta desenvolvida, verificou-se no experimento que este objetivo foi alcançado apenas em alunos que já possuem afinidade com estas atividades. Porém, o fato de aproximar os jogos digitais da produção textual e da leitura e de apresentar um editor visual tende a apresentar melhores resultados quando combinado com práticas pedagógicas utilizadas por mais tempo durante as aulas, conforme relato da professora da turma que acompanhou o experimento.

É válido relatar também que após o experimento e a análise dos resultados obtidos foram realizadas modificações na ferramenta com o intuito de sanar seus problemas técnicos e ela está disponível por meio do endereço <http://vhflabs.com.br/inash>. Além das correções de código também foi elaborado um vídeo tutorial de uso da ferramenta para sanar as dúvidas dos usuários durante o uso da ferramenta.

Referências

- AZEREDO, D. C.; REATEGUI, E. A construção de narrativas digitais como apoio ao processo de letramento. **Novas Tecnologias na Educação**, Porto Alegre, v. 11, n. 1, Jul. 2013.
- CARBONARO, M.; CUTUMISU, M.; DUFF, H.; GILLIS, S.; ONUCZKO, C.; SIEGEL, J.; SCHAEFFER, J.; SCHUMACHER, A.; SZAFRON, D.; WAUGH, K. Interactive story authoring: a viable form of creative expression for the classroom. **Computers and Education**, v. 51, n. 1, p. 687-707, 2008.
- CORDERO, K.; NUSSBAUM, M.; IBASETA, V.; OTAÍZA, M. J.; GLEISNER, S.; GONZALEZ, S.; RODRÍGUEZ-MONTERO, W.; STRASSER, K.; VERDUGO, R.; UGARTE, A.; CHIUMINATTO, P.; CARLAND, C.. Read Create Share (RCS): a new digital tool for interactive reading and writing. **Computers and Education**, v. 82, n. 1, p. 486-496, 2015.
- LYSENKO, L. V.; ABRAMI, P. C. Promoting reading comprehension with the use of technology. **Computers and Education**, v. 75, n. 1, p. 162-172, 2014.
- OECD. PISA 2012 Results in focus: What 15-year-olds know and what they can do what that they know, 2013.

Sistema para Controle Remoto de Equipamentos Eletroeletrônicos utilizando Dispositivos Móveis

Fabiano P. Bhering, Hugo B. Alves, Marina N. Silva, Vinicius F. S. B. Grilo

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
Campus Leopoldina

Rua José Peres, 558 – Centro – Leopoldina - MG – Brasil

fabianobhering@leopoldina.cefetmg.br,
{hugoba18,marinanunessilva}@hotmail.com, ferreira_vinicius@icloud.com

Abstract. *With the growing demand for home automation this paper presents a system for remote control electronic equipment using mobile devices with the Android platform. For the development of the proposal was examined existing technologies and presented differences in relation to cost and practicality to improve comfort and convenience in homes. With the implementation of the system was possible to verify the efficiency and effectiveness of the proposal, proving to be a new product that could benefit many people.*

Resumo. *Com a crescente demanda pela automação residencial este trabalho apresenta um sistema para controle remoto de equipamentos eletroeletrônicos utilizando dispositivos móveis com a plataforma Android. Para o desenvolvimento da proposta foi analisada tecnologias existentes e apresentado diferenciais em relação ao custo e praticidade para a melhoria do conforto e comodidade nas residências. Com a implantação do sistema foi possível constatar a eficiência e eficácia da proposta, demonstrando ser um novo produto que poderá beneficiar muitas pessoas.*

1. Introdução

Com a inclusão tecnológica e a popularização dos aparelhos eletroeletrônicos, a chamada automação residencial vem sendo muito estudada e implementada. É a principal área responsável pela tão buscada facilidade e simplicidade de ações do dia a dia. Um simples dispositivo, como o controle remoto, se tornou altamente presente nas residências, apesar disso apresentam diversos problemas e limitações do ponto de vista dos usuários. Um destes problemas está relacionado à sua área de atuação. Para utilizá-lo é preciso estar próximo (cerca de no máximo 10 metros) do equipamento que se deseja controlar, e livre de barreiras físicas. A incompatibilidade entre os receptores também faz com que seja necessário um controle remoto para cada equipamento (televisão, ar condicionado, etc). Existem no mercado os chamados controles universais, mas suas configurações são específicas, continuam com necessidade de se ter vários controles à disposição. Além disso, são alimentados com pilhas ou baterias que, quando terminam sua vida útil, são altamente prejudiciais ao meio ambiente.

O objetivo deste trabalho é apresentar um sistema de baixo custo para a universalização do controle remoto de equipamentos eletroeletrônicos utilizando os populares dispositivos móveis (smartphone e tablets).

2. Etapas do Desenvolvimento

O desenvolvimento do projeto foi inspirado em diversos trabalhos relacionados e que já são comercializados, sendo os principais: AllShare Control, Griffin Beacon Universal Remote Control, Pell Smart Remote e TouchLight [GooglePlay, 2015]. A partir de comparações e levantamento das limitações apresentadas pelos produtos analisados foi definida uma estrutura para a projeto do sistema (Figura 1) e verificada as possíveis tecnologias para ser apresentada no diferencial do desenvolvimento.



Figura 1. Estrutura de desenvolvimento

O sistema operacional escolhido para o dispositivo móvel foi o Android, por ser uma plataforma open-source e muito popular. Para a recepção, interpretação e execução do comando do usuário, foi constatada a necessidade da confecção de uma central. Para a comunicação com a central foi utilizado um roteador WiFi fazendo o intermédio da comunicação entre o Android e a Central. E a comunicação da Central com o roteador foi de forma cabeada, utilizando o padrão Ethernet com cabos terminais RJ45.

Para alcançar os objetivos foram pesquisadas as tecnologias das áreas de programação e eletrônica, entre elas: Emissor de faixa de luz infravermelho, padrão de emissão de radio frequência adotado nos controles remotos atuais e fundamental para a comunicação com os equipamentos [Lombardi, 2006]; microcontrolador Atmel, principal componente para a programação da central, utilizando o ambiente de desenvolvimento Arduino IDE com bibliotecas escritas na linguagem C; módulo Ethernet, utilizado para estabelecer a conexão da central com o roteador e o Android Studio, ambiente de desenvolvimento para aplicativos móveis, utilizando como principal linguagem de programação o Java com seu kit de desenvolvimento [Lecheta, 2010].

Com o intuito de reunir informações sobre o funcionamento prático do sistema, foram realizados alguns testes relacionados à emissão do sinal infravermelho utilizando a biblioteca IRRemote para Arduino [Paula, 2015]. Com estes testes, foi possível comprovar os diferentes padrões de sinais relacionados à marca e modelo dos eletroeletrônicos. Além disso, foi realizado experimentos de distância e ângulo de incidência dos leds para conseguir abranger todo um cômodo da residência. Sendo assim, foi definido a necessidade da criação de um módulo emissor infravermelho, constituído de sete leds infravermelho dispostos em um ângulo aproximado de 360°, fazendo com que o sinal de luz seja emitido em todas as direções, possibilitando o controle de todo o local. A distância máxima alcançada pelo emissor foi de aproximadamente 7 metros do eletroeletrônico. Para diminuir as ocorrências de interferências físicas foi constatado que o melhor posicionamento do módulo emissor seria no teto do cômodo, formando um ângulo acentuado entre os dois, diminuindo o nível de interferência física e tornando o projeto ainda mais viável.

3. Sistema Proposto

O sistema é composto por uma central eletrônica, módulos emissores de infravermelho e o dispositivo móvel com o aplicativo em Android instalado. A seguir são descritos os componentes do sistema.

3.1. Central Eletrônica

Para a confecção da central de controle foi escolhido o microcontrolador ATmega328P, o mesmo encontrado na placa de prototipagem Arduino UNO. A central foi desmembrada dos leds emissores, permitindo que os módulos emissores fossem distribuídos para cada cômodo e apenas uma Central controlando mais de um cômodo. Assim, a comunicação entre os módulos emissores e a Central é feita de forma cabeada utilizando conectores RJ11. A Figura 2 mostra o esquema da estrutura do sistema.



Figura 2. Esquema do sistema proposto

Para a confecção da central e do módulo emissor foi elaborado o desenho de todo o circuito e transferido para uma placa de fenolite conforme ilustrado na Figura 3.

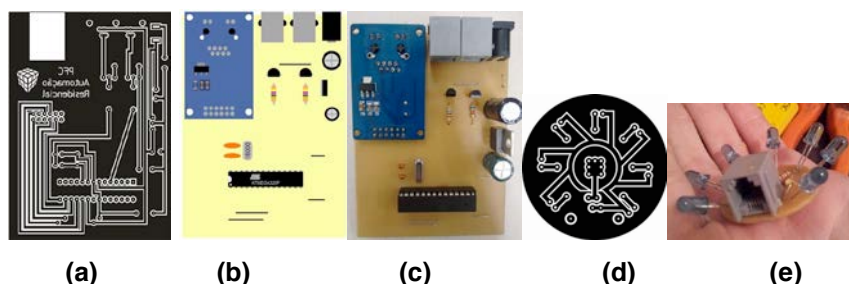


Figura 3. (a) Circuito da Central, (b) Componentes eletrônicos da Central, (c) Central confeccionada, (d) Circuito do módulo emissor e (e) módulo emissor confeccionado

3.2. Aplicativo Android

No desenvolvimento do aplicativo foi utilizada a API de sockets disponibilizada pelo SDK Java [Sierra e Bates, 2010]. Na interface gráfica do usuário tem a opção de criação de cenários da residência, oferecendo diversos cômodos e aparelhos para a escolha do

usuário. A Figura 2 demonstra uma das interfaces para controlar equipamentos similares a uma TV, mas também foram desenvolvidas as interfaces para controlar ar condicionado e iluminação da residência.

4. Experimentos e Resultados

Como demonstrado na Figura 3, a confecção da parte física do sistema atendeu as necessidades, foi desenvolvido um produto compacto e de baixo custo. Com a implementação do sistema foi possível realizar experimentos práticos em alguns cenários reais, onde foi testada a eficiência ao controlar os aparelhos de TV, som, ar condicionado e iluminação em uma residência. Também foi possível garantir sua estabilidade, por utilizar uma placa consistente que diminuiu as interferências e mal contatos.

Os resultados obtidos nos experimentos foram bastante satisfatório. O aplicativo apresentou um interface muito agradável e intuitiva para os usuários que avaliaram o sistema. Foi possível ampliar a área de abrangência para o controle dos equipamentos da residência e diminuiu grande parte das interferências físicas. O sistema também não apresentou nenhum problema em relação ao seu funcionamento e a latência de comunicação com os equipamentos foi semelhante à encontrada em controles remotos tradicionais, menor que 1 segundo. Assim, foi constatada a eficiência e eficácia da proposta.

5. Conclusão

O sistema desenvolvido possibilita a padronização da comunicação remota com os equipamentos eletrônicos domésticos, que fornece a redução do número de controles remotos e, conseqüentemente, lixo eletrônico, que foram algumas das principais motivações para esta proposta. Com o sistema também é possível realizar o chaveamento (ligar e desligar) de equipamentos que não utilizam a tecnologia do infravermelho, como por exemplo as lâmpadas. Assim, foi proposto a implementação de um sistema que poderá ser útil em qualquer tipo de residência, devido ao seu baixo custo e facilidade de implantação.

6. References

- GooglePlay. Aplicativos para Controle Remoto Universal. Disponível em: <<http://play.google.com/>>. Acessado em 21 de março de 2015.
- Lombardi, Rafael R. R. Controle Remoto Infravermelho para Automação. Monografia. FAET - Centro Universitário de Brasília – UNICEUB. Brasília, 2006.
- Lecheta, Ricardo R. Android: Aprenda a criar aplicações para dispositivos móveis com o Android SDK. 2. ed. São Paulo: Novatec, 2010.
- Paula, Luciano I. de. Como funciona: Controle Remoto. Disponível em: <https://www.oficinadanet.com.br/artigo/ciencia/como_funciona_controle_remoto>. Acesso em 21 de março de 2015.
- Sierra, Kathy; Bates, Bert – Use a Cabeça Java – Editora Alta Books, Rio de Janeiro, 2010.

Estudo da Interação do Usuário com o Dispositivo Sensível ao Movimento Kinect e o Controle Sensível ao Movimento WiiMote

Vinicius S. dos S. Silva, Luis Otávio A. Nascimento, André C. da Silva

Instituto Federal de São Paulo – Campus Hortolândia (IFSP)
Av. Thereza Ana Cecon Breda, S/N - Vila São Pedro, Hortolândia - SP, 13183-250

vinicius_simes@hotmail.com, luis.otavio06@gmail.com,
andre.constantino@ifsp.edu.br

Abstract. *New hardware and user interaction software emerged and became popular in the last decade; we can highlight touch-sensitive screens and motion recognition tools. However, the interaction with each one has peculiarities that must be considered in the design time; just changing it could impact on the usability and to conduce to interaction problems. With the motivation to avoid bad solutions to applications that users would use many equipment to interact with, this paper presents an in-progress study about users' interaction with Wii's motion sensitive control and with Kinect motion sensor to investigate the relationship of usability with the size and the distance of the widgets.*

Resumo. *Novos equipamentos e ferramentas de interação de usuário surgiram e se popularizaram na última década, podendo citar telas sensíveis ao toque e ferramentas para reconhecer o movimento. Entretanto a interação com cada equipamento possui peculiaridades que devem ser consideradas, e simplesmente mudar o equipamento pode impactar na usabilidade e ocasiona em problemas de interação. Com a motivação de evitar soluções ruins para as aplicações acessadas por diversos equipamentos, este artigo apresenta um estudo em andamento para investigar a relação da usabilidade com o tamanho e a distância dos elementos da interface quando usado o controle sensível ao movimento do Wii ou o dispositivo sensível ao movimento Kinect.*

1. Introdução

Ao projetar uma interface de usuário, seja para uma aplicação comercial seja um site na Web, o projetista considera quais os equipamentos disponíveis e a sua plataforma. Um dos motivos para considerar o equipamento e plataforma é a consistência da interface a ser projetada com as aplicações usadas pelos usuários. A consistência é um dos princípios de design apontados por pesquisadores da área de IHC e que torna a aplicação mais fácil de ser usada em múltiplos dispositivos. A facilidade de uso, ou seja, a usabilidade, é uma das preocupações ao projetar uma interface de usuário. Nielsen (1993) define usabilidade como a combinação dos elementos facilidade de aprendizado, eficiência, facilidade de lembrar, probabilidade de o usuário cometer poucos erros e a satisfação do usuário.

Dado o crescimento do uso de dispositivos equipados com sensores de movimento, as aplicações podem ser acessadas por esses dispositivos. Torna-se assim necessário estudar problemas recorrentes da mudança de modalidade de interação e identificar soluções com boa usabilidade para múltiplos dispositivos. Este trabalho visa oferecer uma melhor compreensão das mudanças de modalidade, e averiguar as diferenças entre utilizar um periférico sensor de movimento (Kinect) de um controle sensível ao movimento do console Wii (WiiMote) no contexto em estudo, investigando as propriedades de componentes de interface de usuário (como tamanho e distância) quando usuários utilizam os sensores citados, identificando valores mínimos para as propriedades para uma usabilidade adequada conforme sua definição por Nielsen.

Na Seção 2 apresentamos o conceito de modalidade e alguns trabalhos que investigam a mudança de modalidade de interação. Na Seção 3 apresentamos a metodologia utilizada e os resultados parciais deste estudo em andamento. As considerações finais são apresentadas na Seção 4.

2. Fundamentação Teórica

Modalidade é o termo empregado para definir o modo como uma entrada do usuário e uma saída do sistema são expressas. Nigay e Coutaz (1995) definem modalidade como um método de interação que um agente pode usar para alcançar uma meta. Segundo os autores uma modalidade pode ser especificada em termos gerais como “usando fala” ou em termos mais específicos como “usando microfone”. Bernsen (2008) discute que “a modalidade, ou, mais explicitamente, uma modalidade de representação da informação, é uma forma de dispor a informação em algum *medium*. Assim, uma modalidade é definida por seu *medium* e sua particular ‘forma’ de representação”.

Diversas modalidades tornaram-se foco de pesquisa nas últimas décadas, entre elas a fala (reconhecimento e sintetização), a escrita manuscrita (reconhecimento) e a interação com dedos por meio de toques e gestos. Assim, surgiram áreas de pesquisa dentro da Ciência da Computação que visam possibilitar a entrada e a saída de dados por uma modalidade, bem como explorar suas vantagens, como é o caso da Computação Baseada em Caneta e a interação por gestos. Exemplos de trabalhos que estudam a interação com sensores a movimento e gestos são o trabalho de Castellucci e Mackenzie (2008), que estuda o uso de periférico sensível ao movimento na elaboração de textos, o de Fitz-Walter e Tjondronegoro (2008), que estudam a interação com aplicações usando gestos produzidos com os periféricos em questão, e o trabalho de Santhanam (2012), que estuda o uso do periférico por pessoas com paralisia cerebral.

Bernsen (2008) afirma que duas modalidades não são equivalentes, pois as modalidades diferem uma da outra em relação aos pontos fortes e fracos de expressividade e também em sua relação aos sistemas perceptual, cognitivo e emocional do ser humano.

Em nossos estudos anteriores (Silva, Viana e Marques, 2014; Silva et al., 2015), foram estudadas as interações de usuários ao navegar em sites da Web interativos como o *Google Maps* e *Google Street View* com os WiiMote e o Kinect, identificando alguns problemas de interação devido a mudança de modalidade, como a dificuldade em acionar botões pequenos e a movimentação de elemento ao invés de seu acionamento (relacionado ao tempo de resposta e o movimento involuntário da mão para abrir e

fechar ou para pressionar o botão do controle). Algumas questões ficaram em aberto, como: qual o tamanho mínimo adequado para os botões serem acionados com facilidade por meio dos dispositivos estudados? Qual a distância entre os componentes para que a ação iniciada em um componente não acione outro componente devido a movimentação involuntária da mão? São essas perguntas que pretendemos buscar uma resposta.

3. Metodologia

Para a realização deste projeto estamos utilizando um computador para desenvolvimento, um Kinect e um WiiMote. Para mapear os movimentos do usuário captados pelo Kinect em ações do mouse, estamos utilizando a ferramenta Orchestra versão 3.2, enquanto que para mapear os movimentos captados pelo WiiMote em ações do mouse, estamos utilizando a ferramenta TouchMote versão 1.0b13.

Para análise da interação, planejamos a realização de testes com voluntários utilizando uma aplicação desenvolvida com o propósito de coletar dados da interação dos usuários ao interagir com os elementos de interface de usuário mais comuns (botões, menus *drop-down*) e ações do usuário (clique e ação de arrasta e solta). Na Figura 1 é apresentada a interface de usuário da aplicação projetada que emite um log de interação. O propósito é o usuário realizar três baterias de testes na qual cada uma consiste em clicar em nove botões em uma sequência pré-determinada. Cada bateria de testes é repetida com três tamanhos diferentes dos elementos de interface. A ferramenta coleta os erros cometidos pelo usuário (para cálculo da eficácia) e o tempo (para cálculo da eficiência). A opinião subjetiva do usuário é coletada por meio de um questionário, que também é possível obter o perfil do voluntário como, por exemplo, se já utilizou anteriormente o Kinect ou o WiiMote e a sua idade.



Figura 1. Interface da aplicação desenvolvida para investigação dos elementos de interface usando os dispositivos sensíveis ao movimento (a) botões de tamanho mediano e (b) botões de tamanho pequeno.

As dimensões de cada botão são de 60 px de altura por 90 px de largura na bateria 1 (considerados de tamanho mediano), de 45 px de altura por 30 px de largura na bateria 2 (considerados de tamanho pequeno) e de 100 px de altura por 150 px de largura na bateria 3 (considerados de tamanho grande).

A Tabela 1 apresenta dados sobre o tempo e o número de erros de dois voluntários (V1 e V2) utilizando o dispositivo Kinect. Ambos os voluntários possuem 18 anos e possuíam experiência anterior com os dispositivos. Analisando os dados, percebe-se que a eficiência e a eficácia são melhores quanto maior as dimensões dos botões. Entretanto é necessário repetir o estudo com um maior número de voluntários para se realizar uma análise estatística, além de confrontar com a opinião dos usuários.

Tabela 1 – Tempos e erros coletados a partir da interação de dois voluntários utilizando o Kinect.

	Tempo (ms)			Erros		
	Bateria 1	Bateria 2	Bateria 3	Bateria 1	Bateria 2	Bateria 3
V1	18.711	20.013	17.375	0	2	1
V2	19.033	21.328	19.380	1	2	1

4. Considerações Finais

Dado o crescimento do uso de dispositivos equipados com sensores de movimento, as aplicações podem ser acessadas por diversos dispositivos, sendo necessário estudar problemas recorrentes da mudança de modalidade. Este trabalho estuda as diferenças entre utilizar um periférico sensor de movimento (Kinect) de um controle sensível ao movimento (WiiMote) por meio da interação com uma aplicação que coleta o tempo e o número de erros cometidos pelos voluntários e a opinião dos voluntários sobre a interação por meio de um questionário. Percebemos que o tamanho dos elementos interfere no tempo e no número de erros, entretanto um maior número de sessões é necessário para resultados significativos, o que planejamos como trabalhos futuros.

Referências

- Bernsen, N.O. (2008) “Multimodality Theory”, In: *Multimodal User Interfaces: From signal to interaction*, Edited by D. Tzovaras, Berlin, Springer-Verlag Berlin Heidelberg, p. 5-28.
- Castellucci, S. J. and Mackenzie, I. S. (2008) “Unigest: text entry using three degrees of motion”, *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, ACM Press, p. 3549-3554.
- Fitz-Walter, Z., Jones, S. and Tjondronegoro, D. W. (2008) “Detecting gesture force peaks for intuitive interaction”, *Proceedings of the 5th Australasian Conference on Interactive Entertainment*, ACM Press.
- Nielsen, J. (1993) *Usability Engineering*. Morgan Kaufmann.
- Nigay, L. and Coutaz, J. (1995) “A Generic Platform for Addressing the Multimodal Challenge”, *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 13., ACM Press / Addison-Wesley Publishing Co., p. 98-105.
- Santhanan, N. (2012) “Wii remote as a web navigation device for people with cerebral palsy”, *Proceedings of the 14th International ACM SIGACCESS conference on Computers and Accessibility*, ACM Press, p. 303-304.
- Silva, A. C. da; Viana, A. L. C. and Marques, D. (2014) “Identifying Interaction Problems on Internet Navigation Caused by Change of Input Mode: a study case about motion sensor controller, Google Maps and Google Street View”, *Proceedings of International Conference on Interfaces and Human Computer Interaction 2014*, 8, IARIA, p. 271-275.
- Silva, A. C. da, Farias, G. D. R., Gomes, A. S. F. and Silva, F. E. R. (2015) “Identifying Interaction Problems on Web Applications Due to the Change of Input Modality: a study about Kinect, Google Maps and Google Street View”, *Proceedings of International Conference WWW/Internet*, 14., IADIS, p. 192-196.

Controle de Modelos Robóticos por Meio de Gestos Naturais

Paulo Henrique R. de Souza¹, Welerson Augusto L. de Jesus Melo¹, Icaro Meneses F. de Santana², Nara Strappa F. Doria³, Leila Buarque C. de Matos³

¹Discente do Curso Técnico Integrado Informática ao Ensino Médio, ²Discente do Curso Técnico Integrado Eletrônica ao Ensino Médio, ³Docente de Eletrônica
¹²³Instituto Federal de Educação, Ciência, Tecnologia de Sergipe (IFS)
Av. Eng. Gentil Tavares, 1166 - Getúlio Vargas, Aracaju - SE, 49055-260

paulohenriqueifs@live.com, welerson.a.melo@gmail.com, icaro.santana.ifs@gmail.com, narastrappa@gmail.com, leila.buarque.couto.matos@gmail.com

Abstract: *This article describes the creation of new control methods of robotics models through a human-robot interface based on natural gestures, integrating the Arduino and Processing programmable platforms to Kinect and Bluetooth, enabling any person to control remotely the robot using body movements.*

Resumo: *Este artigo descreve a criação de novos métodos de controle de modelos robóticos através de uma interface humano-robô baseada em gestos naturais, integrando a plataforma programável Arduino e Processing ao Bluetooth e Kinect, possibilitando a qualquer pessoa controlar um robô remotamente usando movimentos do corpo.*

1. Introdução

Com a crescente popularização dos modos de interação homem-máquina através da interface natural, foi notável a necessidade de um olhar diferente para essa área de modo que ela possa ser aproveitada ao máximo. O Kinect surge então como grande precursor dessa interface, que segundo Jain et al. (2011), é a forma que permite que os usuários interajam com os computadores da maneira que interagem com o mundo.

Nesse contexto, propõe-se criar novas formas de controle de robôs a partir da interface humano-robô baseada em gestos naturais, integrando a plataforma programável Arduino ao Kinect, possibilitando a qualquer pessoa controlar um robô remotamente usando movimentos comuns do corpo. O restante deste artigo está estruturado como segue: na Seção 2, é apresentado o referencial teórico, enquanto na Seção 3 é resumida a proposta, os materiais utilizados e a implementação. Os resultados obtidos são apresentados na Seção 4 e, por último, na Seção 5 são feitas algumas considerações finais.

2. Referencial Teórico

No mundo da tecnologia há diversos tipos de integração homem-máquina, dentre as quais se podem citar a Interface de Linha de Comando (CLI), a Interface Gráfica (GUI) e a Interface Natural (NUI). Esta última é baseada em comandos realizados pelo corpo humano através de algum objeto ou diretamente pela captação dos gestos. A Interface

Natural Humano-Robô é uma importante função nos sistemas de operação robóticos (Kun Qian, et al., 2013), pois aumenta a interação entre o homem e a máquina.

Pensando na abrangência de uma nova tecnologia com Interface Natural, a Microsoft desenvolveu o dispositivo Kinect. Este dispositivo tem mudado a forma como as pessoas integram-se à tecnologia, e seu impacto vai além da indústria de *games*, com um baixo custo e fácil aquisição. Segundo Zhengyou Zhang (2012) muitas pesquisas vêm sendo realizadas na área da ciência e engenharia, automação (Moreno, 2014) e saúde.

3. Aplicação

A partir do exposto, viu-se a oportunidade de conectar a tecnologia do Kinect à robótica utilizando comunicação sem fio, a fim de criar novos métodos remotos de controle de robôs. Para o projeto, foi desenvolvido um robô móvel com rodas que possuía dois motores DC, um chassi, um Arduino Uno, *Chip* L293D, Módulo Bluetooth e baterias. Além disso, foi criado um braço robótico de ferro com três servomotores simulando os movimentos do ombro, cotovelo e mão humana. No que se refere à Programação Orientada a Objetos, optou-se pelo uso da IDE Processing.

Quatro métodos de controle foram criados para o sistema, sendo um pelo teclado e os outros três pelo Kinect, que detecta automaticamente o corpo do usuário se este estiver devidamente posicionado à frente do dispositivo, tanto em pé como sentado.

No Método 0, o usuário utiliza o teclado para movimentar o robô. O Método 1 capta o ponto mais próximo da tela para mover os motores sem variação de aceleração (Ribeiro et al., 2013). Já o Método 2 utiliza a angulação formada entre os ombros e cotovelos para mover os motores com variação de aceleração (Correa, Marcos Vinicius, 2014). O Método 3, o novo modo criado, simula um volante, através da angulação formada por uma linha imaginária entre as mãos e a distância das mãos ao corpo para mover os motores com variação de aceleração. Os métodos 0 e 2 são utilizados tanto no robô com rodas como no braço robótico.

Enfim, os dados gerados a partir da movimentação do usuário são tratados e transformados de acordo com a programação do método utilizado e enviados em tempo real e via Bluetooth ao Arduino, que lê os dados e escreve as potências de cada motor de acordo com as entradas. A diferença entre o tempo de resposta nos diferentes métodos foi desconsiderada neste trabalho por ser um valor muito pequeno.

3.1 Implementação

Integrando as bibliotecas *open source* OpenNI e Nite2.0, (Melgar e Diez, 2012) foi possível detectar o posicionamento dos membros e retornar as coordenadas no campo 3D dos pés, joelhos, mãos, cotovelos e ombros. Os dados das coordenadas que podem vir com ruídos são tratados com média móvel e ligados para criar a imagem do esqueleto do humano que é sobreposta à sua imagem real. Precisou-se detectar o estado da mão do usuário (aberta ou fechada) e então foi implementado o algoritmo *Flood Fill* que age pelas conexões de pontos em um grafo (Felix e Steven Halim, 2012).

No Método 3, para realizar o movimento de curva com o robô com rodas, obteve-se em relação ao eixo X (horizontal) e ao eixo Y (vertical) a diferença entre as duas mãos e assim foi calculado o coeficiente angular da reta imaginária formada entre

elas definindo para qual lado o robô deve virar e a aceleração centrípeta nas curvas. Já para o braço robótico utilizou-se a programação do Método 2 adaptada para responder o mais próximo possível da posição real do braço do usuário.

4. Resultados

Foram realizados testes de laboratório com alguns voluntários a fim de melhorar a aplicação, avaliar a aceitação e gerar gráficos. A Figura 1 mostra a tela com os dados de potência linear (centro da tela), a proporção entre as potências nos motores (canto esquerdo da tela) e o estado das mãos, necessário para estar no controle do sistema.

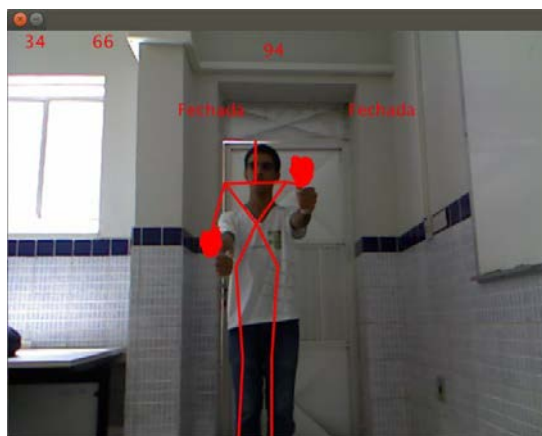


Figura 1. Voluntário utilizando o Sistema no Método 3.

Tomando o Método 0 e o Método 3 no robô com rodas, foram gerados gráficos que mostram a diferença entre os dois métodos numa trajetória que se definiu por variações em uma pista que apresentava curvas para a direita e esquerda. A pista possuía dois metros de comprimento por um metro de largura. Os dados apresentados nos gráficos fazem referência à proporção entre os motores, indicando qual lado está recebendo a maior potência. Ou seja, quanto mais baixo maior a curva para a direita.

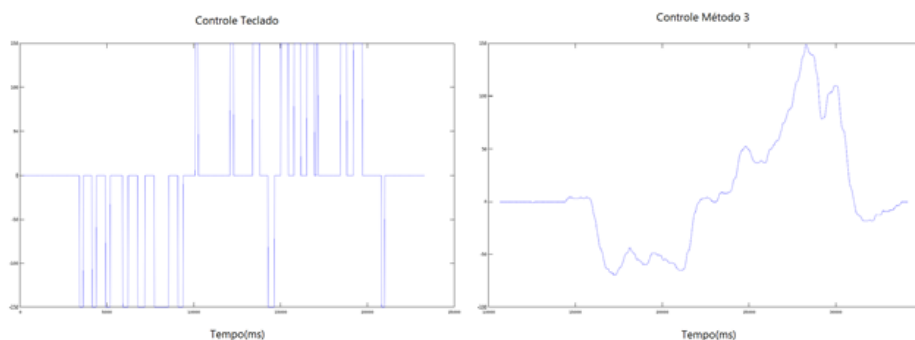


Figura 2. Gráficos dos métodos de controle - Robô com rodas

O gráfico da Figura 2, gerado a partir de amostras, apresenta no eixo horizontal os dados de tempo em milissegundos e no eixo vertical a diferença proporcional de potência entre os motores. Pode-se perceber a maior suavização do movimento com o Kinect. Vemos de modo semelhante nos gráficos da Figura 3 os resultados do Método 0 utilizando o teclado e do Método 2 utilizando o Kinect para controle do braço robótico. A tarefa foi levar o Braço Robótico até o centro de um pneu pequeno, estendê-lo e arremessá-lo para o lado ao fechar a mão. O pneu distava vinte centímetros em relação à base do Braço Robótico e seis centímetros de altura.

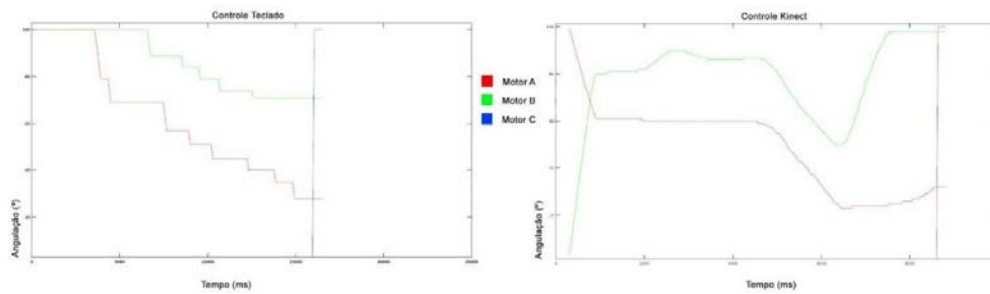


Figura 3. Gráficos dos métodos de controle - Braço Robótico

No gráfico acima, o Motor A representa a articulação do ombro, o Motor B a articulação do cotovelo; o Motor C, a abertura da mão. É possível observar a partir das amostras que o método utilizando o Kinect terminou a tarefa aproximadamente duas vezes mais rápido, além de possuir maior suavização dos movimentos articulares e um controle mais intuitivo. Os resultados podem ser vistos nos vídeos: <https://youtu.be/4iPnecyNJ4> e <https://youtu.be/6AezSTYSE00>.

5. Conclusão e trabalhos futuros

Tendo em vista que a Interface Natural aplicada à robótica tem um grande potencial de desenvolvimento no mundo, o projeto utilizou uma metodologia simples para conectar tecnologias e alcançar o resultado desejado. Foi possível controlar de forma remota um robô com rodas e um braço robótico utilizando o próprio corpo, demonstrando maior eficácia no principal método proposto em relação aos métodos já existentes implementados. Futuramente, visa-se o estudo de técnicas de filtragem e suavização de movimentos para os motores, sendo que a partir dos gráficos pôde-se perceber pequenos erros e oscilações indesejadas devido à baixa qualidade dos dados recebidos do Kinect.

6. Referências

- CORREA, M. V. C. (2014). "Controle de robôs móveis utilizando o Kinect". Escola de Engenharia de São Carlos da Universidade de São Paulo.
- Halim, S., Halim F. (2012). "Competitive Programming 2". Lulu Press.
- Jain, J., Lund, A., Wixon, D. (2011). The future of Natural User Interface. In: CHI '11 Extended Abstracts on Human Factors in Computing Systems, pages 211-214.
- Melgar, E., Diez, C. (2012). "Arduino and Kinect Projects: Design, Blow Their Minds." Technology in Action.
- Moreno, R. J. (2014). "Tracking Humano mediante Kinect para control de robots". Universidad Autónoma de Colombia.
- Niu, J., Quian, R. and Yang, H. (2013). "Developing a Gesture Based Remote Human-Robot Interaction System Using Kinect", In: "International Journal of Smart Home".
- Ribeiro, L. C., Farias, Farias, A.B., Amaral, E.M.A. (2013) "Sistema De Controle De uma Plataforma Robótica Experimental Baseado em Visão Computacional Utilizando Kinect", In: "Mostra Nacional de Robótica".
- Zhang, Z. (2012) "Microsoft Kinect Sensor and Its Effect", In: "IEEE Multimedia", pages 4-10.

Um Framework para Descoberta de Recursos usando Computação nas Nuvens

Rodrigo Sforini Mota¹, Pablo Diego Silva da Silva¹,
Clézio Silva de Souza¹, Raphael de Aquino Gomes^{1,2}

¹Instituto Federal de Goiás - Campus Goiânia
Rua 75, n. 46, Centro. CEP: 74055-110 - Goiânia - GO - Brasil

²Instituto de Informática – Universidade Federal de Goiás
Câmpus Samambaia, Caixa Postal 131, Goiânia - GO, Brasil

primeironome.ultimonome@estudantes.ifg.edu.br, raphael.gomes@ifg.edu.br

Abstract. *The allocation of resources needed to deploy an application is usually performed in an ad hoc way. In this paper we propose a framework to facilitate cloud-based resources specification and discovery. The framework has been partially implemented, and its effectiveness was evaluated.*

Resumo. *A definição de recursos necessários para implantar uma aplicação é comumente executada de maneira ad hoc, o que pode ocasionar perda de qualidade. Neste trabalho propomos um framework para facilitar a especificação e descoberta de recursos em ambientes de computação nas nuvens. O framework foi parcialmente implementado, sendo sua eficácia comprovada.*

1. Introdução

Uma dificuldade comum no desenvolvimento de software é realizar sua implantação. Após a definição da infraestrutura necessária, a implantação comumente é feita mediante a alocação de recursos pré-existentes ou com alocação de capital para a aquisição de novos recursos. Contudo, em diversos cenários nenhuma destas alternativas é viável devido a limitações financeiras. Dessa forma, Computação nas Nuvens constitui uma alternativa promissora. Este modelo permite acesso sob demanda, através de rede, a um conjunto de recursos computacionais (rede, armazenamento, processamento e aplicações) configuráveis, que podem ser rapidamente oferecidos e liberados com mínimo esforço de gerenciamento e interação por seu provedor [Mell e Grance 2009].

Este trabalho apresenta um framework para definição e descoberta de recursos usando computação nas nuvens. É possível encontrar os recursos necessários utilizando múltiplos provedores. O principal objetivo deste trabalho é propor uma alternativa que abstraí detalhes da seleção de recursos e permite analisar a viabilidade de ambientes de nuvem para o provisionamento destes recursos. O framework proposto foi avaliado de forma qualitativa através de um estudo de caso utilizando uma aplicação de uma fábrica de software. O restante do artigo é organizado da seguinte forma: a Seção 2 discute alguns trabalhos relacionados; a Seção 3 apresenta o framework, ao passo que sua avaliação é discutida na Seção 4; na Seção 5 são apresentadas algumas considerações finais.

2. Trabalhos Relacionados

Computação nas Nuvens tem como tecnologia base Arquiteturas Orientadas a Serviços (SOA) para gerenciar os serviços oferecidos e controlar a distribuição, acesso e armazenamento dos dados. Desta forma, os trabalhos mais relacionados a ambientes de nuvem são aqueles que tratam ambientes de SOA. Dentre estes, Wright et al. [Wright et al. 2012] propuseram um modelo baseado em restrições para especificação e seleção de recursos em ambientes de nuvem. Assim como propomos, o modelo é focado na aplicação, de forma que os recursos são especificados usando uma linguagem independente do provedor. Contudo, a linguagem proposta é exclusiva do modelo, dificultando sua utilização de forma ampla. Em [Khajeh-Hosseini et al. 2011] é apresentado o *Cloud Adoption Toolkit* que consiste em um framework organizacional usado para identificar os aspectos relacionados à adoção de ambientes de nuvem com ferramentas para apoiar decisões do cliente. Neste trabalho não é apresentada nenhuma ferramenta que apoie a obtenção de recursos.

3. Framework para Descoberta de Recursos

O modelo proposto para representação dos recursos é um subconjunto do padrão *Open Cloud Computing Interface (OCCI)*¹, que consiste em um conjunto de especificações abertas desenvolvido pelo *Open Grid Forum*², uma comunidade voltada para padronização de computação nas nuvens.

Os tipos de recursos disponíveis no modelo inclui: i) *Resource* que representa um tipo abstrato para as demais classes; ii) *Network* que representa uma entidade de rede (ex. switch). Possui como atributos um identificador VLAN 802.1q; um *token* e o estado corrente da VLAN; iii) *Compute* que representa um recurso genérico de processamento (ex. máquina virtual). Possui como atributos a arquitetura da CPU, o número de núcleos, a frequência, o hostname DNS do recurso, a quantidade de memória RAM, e o estado; e iv) *Storage* que representa recurso que armazenam informação em um dispositivo de armazenamento de dados. Possui como atributo o tamanho do armazenamento em gigabytes.

O módulo responsável pela descoberta de recursos é denominado *Cloud Service Discovery (CSD)*. Sua arquitetura é apresentada na Figura 1. A linguagem utilizada para definição de recursos no CSD é uma implementação do padrão OCCI. Os demais componentes do CSD são descritos a seguir:

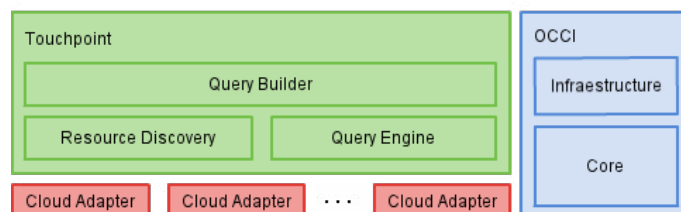


Figura 1. Arquitetura do framework *Cloud Service Discovery*.

O *Touchpoint* constitui o núcleo do módulo. Tem como responsabilidade a agregação dos vários provedores de nuvem, provendo um ponto único de acesso aos mesmos, além de gerar e processar as consultas a serem realizadas. Estas consultas são

¹<http://www.occi-wg.org>

²<http://www.ogf.org>

geradas pelo componente *Query Builder* a partir da definição de recursos da aplicação. A interpretação e processamento das consultas por sua vez são realizados pelo *Query Engine*. O gerenciamento dos provedores de nuvem é feito pelo componente *Service Discovery*. Para um provedor ser considerado nas buscas realizadas ele deve implementar a interface estabelecida pelo *Touchpoint* e se registrar no *Service Discovery*.

O *Cloud Service Provider Adapter* representa a abstração de um provedor de nuvem. Consiste em uma adaptador que deve ser implementada pelo provedor para que ele seja considerado nas consultas realizadas.

Os módulos do framework são executados na estação cliente, sendo a comunicação com o provedor de responsabilidade do adaptador. Além disso, as atividades de descoberta e especificação de recursos são feitas como uma etapa anterior à implantação da aplicação, não afetando na eficiência da mesma.

O framework foi implementado em Java. Para a implementação da descrição dos serviços em OCCI foi realizada uma adaptação da implementação OCCI4Java³. Para permitir a avaliação da abordagem, foi implementado um adaptador para provedor de nuvem. O provedor escolhido foi Amazon⁴, uma vez que este é um dos produtos dominantes na área e já oferece um SDK em Java⁵, permitindo sua fácil integração à implementação realizada. Neste protótipo também não foi considerada a implementação do componente *Query Builder*, de forma que as consultas são construídas manualmente.

4. Avaliação

A avaliação foi realizada através de uma aplicação que visa controlar as atividades pertinentes à realização de análises clínicas de um laboratório. O sistema possui interface Web e foi desenvolvido em PHP, sendo necessário portanto um servidor web e um sistema gerenciador de banco de dados para realizar sua implantação.

Com base nos requisitos descritos, e com a restrição de que a implantação deveria demandar o mínimo de gastos possível pela organização, foi obtida a descrição de recursos apresentada na Figura 2, de acordo com a qual seria necessário um banco de dados com espaço disponível de 10 GB e dois servidores, com arquitetura x86 e configuração mínima de processador com 8 núcleos e memória RAM com 7,5 GB. A duplicação dos servidores se deu visando redundância para atender o requisito de confiabilidade.

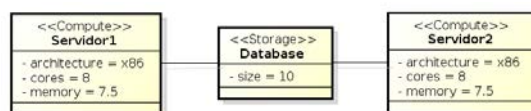


Figura 2. Descrição dos recursos necessários para a aplicação.

Com base nesta descrição de recursos foram geradas as consultas submetidas ao framework. O primeiro passo consiste em definir a consulta utilizando a API implementada pelo componente *Query Engine*. Depois, esta é submetida ao serviço de descobertas. Como resultado é retornado um mapa cujas chaves são identificadores dos provedores

³<https://github.com/occi4java>

⁴<http://aws.amazon.com>

⁵<http://aws.amazon.com/sdkforjava>

cadastrados no sistema; e os valores são recursos que satisfazem a consulta realizada. É importante observar que os recursos retornados tem o mínimo da configuração de hardware solicitada na consulta, sem necessariamente possuir o valor exato.

Realizamos um experimento para comparar a eficácia e eficiência da busca de recursos usando o framework ou usando diretamente a API oferecida pela Amazon.

A mesma consulta foi realizada 10 vezes em ambas as abordagens. Usando diretamente a API Java da Amazon foram encontradas 7248 imagens satisfatórias em um tempo médio de 34,23 segundos, com desvio padrão de 12,06 segundos. Usando o framework foram obtidas 7248 imagens em um tempo médio de 40,69 segundos, com desvio padrão de 10,43. A comparação dos dados permitiu verificar que os resultados eram os mesmos, comprovando a eficácia da proposta. Com relação à eficiência, foi verificada um aumento médio de 6,46 segundos com o uso do framework, o que não constitui um grande problema porque sua utilização é feita de maneira *offline*, sem comprometer a aplicação. Além disso, este resultado pode ter ocorrido em virtude de variações no provedor de serviços, o que pode ser verificado com o alto desvio padrão medido. Os principais resultados da consulta obtidos com o framework são apresentados na Tabela 1.

ID	Tipo de Instância	Cores	RAM (GB)	Custo por hora (\$)
ami-daf005b3	m4.2xlarge	8	32.0	0.54
ami-8ec035e7	m4.2xlarge	8	32.0	0.54
ami-987185f1	m4.2xlarge	8	32.0	0.54
ami-d59d6bbc	m4.2xlarge	8	32.0	0.54
ami-9f4082f6	m4.2xlarge	8	32.0	0.54

Tabela 1. Resultado da busca realizada no experimento (5 melhores instâncias).

5. Considerações Finais

Neste artigo foi apresentado um framework para descoberta de recursos em ambientes de nuvens. O framework pode ser usado para apoiar as atividades de implantação do produto desenvolvido. Ele foi avaliado com uma aplicação de exemplo que demonstrou sua eficácia e eficiência. Como melhorias a serem incluídas no framework propõe-se o suporte a um número maior de tipos de recursos e provedores de nuvem, além de um mecanismo mais aperfeiçoado de seleção dos recursos a serem utilizados.

Agradecimentos

Este trabalho recebe auxílio financeiro das fundações FAPEG (chamadas # 04/2011, 12/2012 e 03/2013) e CNPq (contratos # 249809/2013-3 e 473939/2012-6).

Referências

- Khajeh-Hosseini, A., Greenwood, D., Smith, J. W., e Sommerville, I. (2011). The Cloud Adoption Toolkit: supporting cloud adoption decisions in the enterprise. *Software: Practice and Experience*.
- Mell, P. e Grance, T. (2009). Draft NIST working definition of cloud computing.
- Wright, P., Sun, Y., Harmer, T., Keenan, A., Stewart, A., e Perrott, R. (2012). A constraints-based resource discovery model for multi-provider cloud environments. *Journal of Cloud Computing: Advances, Systems and Applications*, 1(1):6.

Conhecimento em redes sociais: um estudo de caso aplicado ao *Twitter*

Miguel Airton Frantz, Angelo Augusto Frozza, Reginaldo Rubens da Silva

Instituto Federal Catarinense (IFC) – Campus Camboriú
Rua Joaquim Garcia S/N – 88.340-055 – Camboriú – SC

frantz.miguel@gmail.com, {frozza,reginaldo}@ifc-camboriu.edu.br

Resumo. *Este artigo descreve a criação de uma base de conhecimento a partir de dados não estruturados que um usuário produz. O foco consiste no desenvolvimento de uma aplicação que extrai o conhecimento dos textos publicados no Twitter. Como resultado, apresenta-se o conhecimento implícito presente na base de dados do usuário na forma de uma nuvem de tags contendo os termos relevantes que mais aparecerem na base de dados em questão.*

1. Introdução

Os avanços da tecnologia e a facilidade no compartilhamento de informações e conhecimentos, faz com que, diariamente, milhares de pessoas produzam informações novas que circulam pelo mundo. Essas informações compartilhadas criam imensos bancos de dados pessoais que, quando não estão organizadas ou classificadas, acabam tornando difícil a realização de qualquer busca ou até mesmo seu entendimento.

Para Sousa (2012), o ato de classificar acompanha o cotidiano dos seres humanos, muitas vezes de forma imperceptível. Como exemplo, a rede social *Twitter* utiliza um sistema de classificação e indexação opcional das mensagens, que consiste em atribuir uma *tag* (etiqueta) à mensagem para classificá-la e representá-la. Como o *Twitter* não possui comunidades, fóruns ou grupos, as *tags* assumem o papel de realizar a classificação da informação nesta rede social. Esse sistema de classificação e indexação (manual), por muitas vezes acaba não sendo utilizado, o que faz com que as mensagens não possuam nenhuma classificação.

Assim sendo, surge a proposta de um mecanismo de Recuperação da Informação (RI) capaz de extrair as principais informações presentes nos textos publicados por um usuário em sua conta do *Twitter*. Utilizando técnicas de RI, é realizada a busca de palavras que possam expressar o sentido de cada *tweet*, ou seja, um conjunto de palavras chave. A junção destes termos, que podem ser considerados relevantes, ajuda a definir o perfil de determinado usuário e retrata os assuntos que ele mais aborda em sua conta. Para apresentar visualmente as informações coletadas, optou-se por utilizar nuvens de *tags*, nas quais as palavras são dispostas aleatoriamente, variando seu tamanho de acordo com o número de vezes que a mesma aparece no conjunto de texto analisado.

O restante deste artigo está organizado da seguinte maneira: na seção 2 são apresentados os materiais e métodos; a seção 3 apresenta os principais resultados obtidos; a seção 4 apresenta as considerações finais.

2. Materiais e Métodos

A aplicação proposta, denominada *TweetKnowledge*¹, tem sua arquitetura estruturada em camadas, de forma que cada camada procura ser, tanto possível, independente das demais. Optou-se pelo desenvolvimento em camadas para que, em projetos futuros, seja possível reaproveitar facilmente os componentes de *software* produzidos:

- a) *Camada de autenticação*: é responsável pelo processo de identificação do usuário. Para isso, utiliza-se o *login* social do *Twitter*, que autoriza o acesso aos dados cadastrais do usuário. Feita a autenticação, a aplicação recebe algumas informações sobre o usuário, que são utilizadas para realizar o controle de quem está submetendo uma base de dados, além de permitir a publicação da nuvem de *tags* no *Twitter*;
- b) *Camada de upload dos arquivos*: é responsável por armazenar (*upload*) no banco de dados da aplicação os arquivos de *tweets* do usuário, os quais estão em formato JSON. O *upload* do histórico de *tweets* do usuário dá um ganho de performance considerável em comparação com a opção de buscar os dados de forma *on-line*;
- c) *Camada de manipulação dos dados e indexação das palavras*: após o *upload* de um arquivo é feita a indexação do texto. Esta camada teve como base a infraestrutura de *Full Text Searching* (FTS) fornecida pelo próprio *PostgreSQL* (através do módulo *TSearch 2*). O FTS consiste em uma técnica de pesquisa e recuperação de informações de texto armazenado no banco de dados. Foi necessário compreender como funciona a indexação e os dicionários de palavras, uma vez que os mesmos são utilizados em diversas etapas de filtragem e limpeza dos dados;
- d) *Camada de apresentação e geração da nuvem de tags*: responsável pela geração da nuvem de *tags* através do *framework D3-Cloud* (<https://github.com/jasondavies/d3-cloud>). Também permite compartilhar a nuvem de *tags* do usuário.

A dificuldade em classificar e encontrar informações não é algo presenciado somente no *Twitter*. Entretanto, escolheu-se utilizá-lo por se tratar de uma das três maiores redes sociais ativas, com aproximadamente 500 milhões de usuários (MASHABLE, 2012). Redes sociais permitem que usuários criem conteúdo e vêm se tornando chave em pesquisas relacionadas ao tratamento de grandes quantidades de dados, além de constituírem um ambiente propício para extração de conhecimento (BENEVENUTO, ALMEIDA e SILVA, 2012). Com o intuito de facilitar a interpretação das informações obtidas pela aplicação, optou-se por apresentar o resultado final no formato de uma nuvem de *tags*. Além disso, este formato permitirá fazer a ligação dos termos com outras bases de dados, de uma maneira muito natural para o usuário final.

3. Resultados e Discussão

O *TweetKnowledge* consiste em uma aplicação prática para demonstrar a resolução de problemas associados com Recuperação da Informação (RI) sobre uma base de dados produzida por um usuário que compartilha notícias e informações por meio de sua conta do *Twitter*. A RI sobre textos, em geral, consiste em diminuir o número palavras de um texto, sempre tomando o cuidado para que este não perca o seu sentido inicial. Assim, os textos são reduzidos até que reste o menor número possível de termos que expressem sentido. Esses termos podem ser entendidos como as palavras chave do texto em questão e são utilizados na aplicação desenvolvida neste projeto.

¹ <http://www.tweetknowledge.com>

Para realizar esta tarefa, o *PostgreSQL* trabalha com a indexação de textos e a análise dos documentos em *tokens*, que podem ser entendidos como fragmentos brutos do texto do documento. Também utiliza o termo “lexema” para referir-se às palavras consideradas úteis para indexação e busca. Estes lexemas são obtidos após o tratamento dos *tokens* mencionados anteriormente. Dicionários são utilizados para realizar esta etapa. No *PostgreSQL* há vários dicionários padrões, mas também pode-se personalizar dicionários para necessidades específicas. O uso de dicionários permite realizar um controle refinado sobre como os termos (*tokens*) são normalizados, por exemplo: definir *stop words* - conjunto de palavras comuns que não possuem um sentido sozinhas, como artigos, preposições etc.; mapear sinônimos para uma única palavra - utilizando *ispell* (dicionário com palavras do idioma escolhido); mapear frases para uma única palavra - utilizando *thesaurus* (dicionário de sinônimos); mapear diferentes variações de uma palavra para uma forma canônica - utilizando regras de *Snowball stemmer* (redução das palavras por meio de algoritmos específicos).

Neste trabalho, as etapas descritas anteriormente resultam em um conjunto de lexemas presentes em cada *tweet*. Para obter a frequência que cada termo aparece, basta realizar uma contagem por meio de uma consulta SQL. Em função de ter-se adotado o paradigma da codificação por camadas (Figura 1), no caso da adaptação da aplicação para usar outras fontes de dados, é necessário reescrever apenas a camada de entrada de dados.

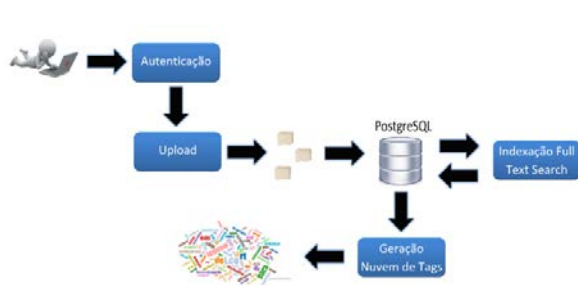


Figura 1 – Arquitetura da aplicação *TweetKnowledge*.



Figura 2 – Nuvem de *tags* gerada.

Percebe-se a importância desta pesquisa para a comunidade em geral, em especial, para aquelas pessoas que produzem algum tipo de conhecimento. Com isso, além de servir como mecanismo de indexação, o sistema permite que o usuário dê maior visibilidade à sua produção, por meio da nuvem de *tags* que pode ser compartilhada nas redes sociais.

Os termos apresentados na nuvem de *tags* (Figura 2) são as palavras consideradas relevantes após a manipulação e indexação dos dados na terceira camada. A quarta camada recebe como entrada uma lista de palavras e a frequência que cada uma aparece na base de dados e se encarrega de distribuí-las, determinando maior tamanho de fonte para palavras que tem maior número de repetições, além da variação das cores que ocorre de forma aleatória, com o intuito de facilitar a distinção entre os termos.

4. Considerações finais

Tendo em vista a importância da RI em redes sociais e, também, a relevância da classificação do conhecimento utilizando *tags* e palavras chave que representem uma mensagem ou texto, o presente artigo apresenta um mecanismo para extração de

conhecimento em cima das mensagens (*tweets*) de um usuário e permite a representação das informações obtidas de maneira que o usuário possa entender com facilidade.

Para tanto, é disponibilizada uma ferramenta que realiza a busca de palavras chave presentes em todo o histórico de publicações no *Twitter* de um usuário. Para obter os termos relevantes é realizado um tratamento dos textos anexados, utilizando dicionários de palavras, removendo as palavras que não possuem sentido e realizando diversas etapas de filtragem e limpeza dos dados. Após o tratamento dos textos, tem-se como resultado uma lista de termos relevantes, acompanhados do respectivo número de ocorrências. Com o objetivo de facilitar o entendimento e a assimilação da informação obtida, os termos relevantes são apresentados por meio de uma nuvem de *tags*. Um protótipo funcional do *TweetKnowledge* pode ser acessado através do link <http://www.tweetknowledge.com>.

Atualmente, a aplicação usa como entrada apenas os textos de *tweets*, porém, trabalhos futuros preveem o uso de outras fontes de conteúdo geradas pelos usuários, como outras redes sociais, documentos de textos (PDF, DOC etc.) e textos de *e-mails*. Além disso, pretende-se efetuar diversas melhorias, como utilizar *Linked Data* para adicionar conhecimento à lista de termos, sendo que cada termo da nuvem de *tags* direcionará para os textos (*tweets*) de sua origem e também para fontes externas.

Até o presente momento, foi encontrado apenas um trabalho que se propõem a fazer algo semelhante, chamado *TweetCloud* (<http://tweetcloud.icodeforlove.com>). Entretanto, segundo análise realizada, o mesmo apresenta diversos problemas, como lentidão por buscar os textos dos *tweets on-line* (o que também restringe a quantidade de texto a ser indexada), falhas na filtragem dos termos, utilização de um *layout* de nuvem simples, entre outros.

Referências bibliográficas

- BENEVENUTO, F.; ALMEIDA, J.; SILVA, A. Coleta e análise de grandes bases de dados de redes sociais online. In: JORNADA DE ATUALIZAÇÕES EM INFORMÁTICA (JAI). Cap. 2. **Anais do XXXII Congresso da Sociedade Brasileira de Computação (CSBC)**. Curitiba: SBC, 2012.
- FELIX, V. **Faça o download do histórico do Twitter**. In: <<http://blogs.estadao.com.br/link/faca-o-download-do-historico-do-twitter/>>. São Paulo: Estadão, 18 jan. 2013.
- MASHABLE. **Will You Be Twitter's 500 Millionth User?** 2012. Disponível em: <<http://mashable.com/2012/02/22/twitters-500-million-user/>>. Acesso em: 25 jul. 2014.
- POSTGRESQL. Chapter 12. **Full Text Search**. Disponível em: <<http://www.postgresql.org/docs/9.3/static/textsearch.html>>. Acesso em: 20 mar. 2016.
- SOUSA, A. M. de. **Organização em sistema caótico: uso das tags para classificação da informação pelos usuários da rede social Twitter**. 106 f. 2012. Mestrado (Mestrado em Ciência da Informação) – UFRJ, Rio de Janeiro (RJ).

Acessibilidade Móvel para Alfabetização de Deficientes Visuais: Proposta Inicial de um Protótipo

Jenifer Melissa de Paula¹, José Valter Amaral de Freitas¹, Thatiane de Oliveira Rosa¹

¹ Curso de Gestão da Tecnologia da Informação - Campus Paraíso do Tocantins do Instituto Federal de Educação, Ciência e Tecnologia do Tocantins (IFTO) Distrito Agroindustrial, BR 153, KM 480 – Caixa Postal 151 – 77.600-000 – Paraíso do Tocantins – TO – Brasil

{jenifermelissadep, josewalteraf}@gmail.com, thatiane@ifto.edu.br

Abstract. *According to the 2010 census, 18.8% of the population has visual impairment, and such portion requires special attention in several respects, especially with regard to the educational process. Another relevant fact is that today information technology is used as a support for execution of several processes, including the teaching/learning, which extends the possibilities of interaction. a trend towards adoption of mobile technologies being perceived. Therefore, this work aims to identify desirable requirements for a mobile application touch screen, aimed assist in literacy visually impaired.*

Resumo. *De acordo com o censo de 2010, 18,8% da população brasileira possui deficiência visual, sendo que tal parcela necessita de atenção especial em diversos aspectos, principalmente no que tange ao processo educativo. Outro fator relevante é que hoje a tecnologia da informação é utilizada como suporte para execução de diversos processos, dentre eles o de ensino/aprendizagem no qual amplia as possibilidades de interação. Sendo percebida uma tendência à adoção de tecnologias móveis. Diante disso, esse trabalho possui como objetivo identificar requisitos desejáveis a uma aplicação móvel touch screen, que vise auxiliar a alfabetização de deficientes visuais.*

1. Introdução

A aplicação de novas práticas didáticas aliadas a recursos tecnológicos diversos vêm ampliando as possibilidades de ensino em diversas áreas. Desse modo, essas práticas têm sido muito úteis para a educação, pois os conteúdos podem ser abordados de várias formas tais como imagens, vídeos, hipertextos, animações, simulações, jogos educacionais, páginas web, entre outras. Esses métodos servem como incentivo e ferramenta para práticas pedagógicas mais criativas, inovadoras e eficazes, proporcionando novas maneiras de pensar e de manifestar ideias, tanto sobre educação, quanto sobre comunicação, tecnologia e interação com a realidade (BEHAR; TORREZAN, 2009).

Em harmonia com esse contexto, surgiram os “nativos digitais” ou “geração multitarefa”, que segundo Tori (2010) são crianças, adolescentes e jovens que nasceram na era da internet, que estão conectados ao mundo e que sem sair de casa, conseguem

estudar, fazer pesquisas na internet, conversar com os amigos e assistir a programas televisivos simultaneamente. Eles conseguem absorver e descartar grande quantidade de informações e exercer atividades paralelas, no entanto precisam ser frequentemente motivados e recompensados. A partir das considerações desse mesmo autor, entende-se que uma criança, adolescente ou jovem com deficiência visual atende pelas mesmas características.

Diante disso, ao considerar que 18,8% da população brasileira possui deficiência visual (IBGE, 2010), e que tal parcela necessita de uma atenção especial no processo educativo, visto que possuem algumas limitações, mostra-se importante realizar um estudo que aborde o viés educação de deficientes visuais apoiada por tecnologia da informação, por meio de dispositivos móveis.

Nesse contexto, com relação à tecnologia, existem aplicações de acessibilidade em diversos ambientes digitais, como páginas web, desktops e dispositivos móveis. Entretanto, de acordo com a pesquisa bibliográfica realizada até o momento neste trabalho, existem aplicações móveis que auxiliam na alfabetização, entretanto as mesmas não são desenvolvidas para atender às necessidades apresentadas pelos deficientes visuais. Dentre as aplicações móveis existentes para auxiliar na alfabetização, porém, não voltadas aos deficientes visuais, pode-se citar a m-EQUIV, que tem a finalidade de apoiar o ensino de leitura para crianças (SILVA; PINHEIRO, 2014). Logo, a área de aplicações móveis para deficientes visuais ainda é pouco explorada, possuindo assim recursos, informações e conhecimento limitados.

Diante do contexto apresentado, a área de aplicações móveis voltadas à acessibilidade visual é de extrema importância para a sociedade, não apenas pelo valor tecnológico, mas também pelos valores sociais e culturais. Portanto, a partir desse cenário, vê-se a necessidade de realizar uma pesquisa científica que tenha como objetivo identificar os requisitos desejáveis para desenvolver uma aplicação móvel *touch screen*, que vise auxiliar a alfabetização de deficientes visuais.

2. Metodologia

A pesquisa a ser realizada neste trabalho pode ser classificada como exploratória e qualitativa, pois visa uma investigação por meio de estudos bibliográficos, entrevistas e análises de ambientes reais, para que a partir da interpretação das informações coletadas possam ser identificados os requisitos desejáveis à aplicação pretendida. Conforme ilustra a Figura 1, a pesquisa é composta por três fases: Pesquisa Bibliográfica, Estudo de Campo e Prototipagem.

A primeira fase consiste na realização de uma pesquisa bibliográfica simples direcionada à investigação de conceitos relacionados à acessibilidade, deficiência visual e suas leis, metodologias de alfabetização de crianças com deficiência visual, e utilização da TI no processo educacional.

A fase 2 consiste em um estudo de campo, o qual tem como objetivo identificar as principais dificuldades no processo de alfabetização de crianças com deficiência visual, e a partir disto ser possível projetar soluções tecnológicas. O estudo de campo será realizado com base em entrevistas com envolvidos no processo de alfabetização,

com familiares, professores e estudantes. Além de observação e acompanhamento do cotidiano do processo de alfabetização de crianças com deficiência visual.

Por fim, a fase 3 tem o objetivo de desenvolver um protótipo de alta fidelidade de uma aplicação móvel para a alfabetização de deficientes visuais, elaborada com base nos conceitos estudados durante a fase 1, e nas informações obtidas na fase 2.

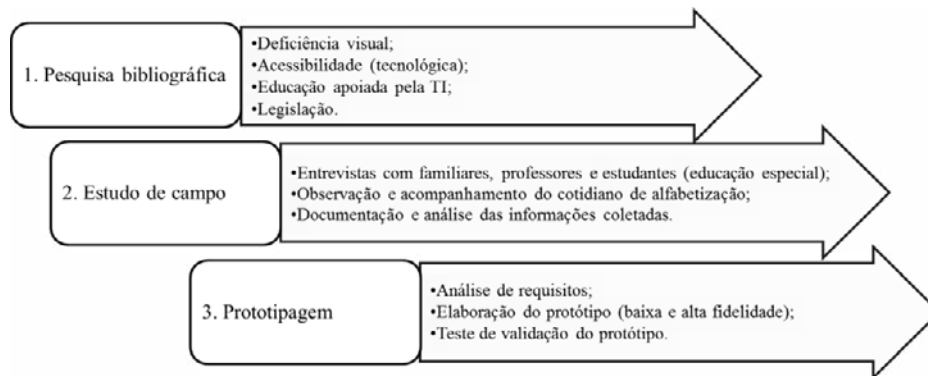


Figura 1. Fases da metodologia.

Até o momento, através da pesquisa bibliográfica foi desenvolvida uma primeira versão do protótipo, levando em consideração as normas e requisitos de desenvolvimento de software acessível identificados na pesquisa, e baseando-se também em softwares e jogos que possuem características semelhantes e favoráveis ao projeto, porém, que possuem outros objetivos. Vale ressaltar que a fase 2 está em plena execução.

3. O Protótipo

O objetivo dessa primeira versão do protótipo é servir de base para o estudo de campo (segunda fase), sendo apresentado aos profissionais da educação especial para avaliação e sugestões, juntamente com um questionário sobre as metodologias de ensino especial, suas dificuldades e necessidades. Após análise e avaliação das sugestões e dos resultados do questionário, além dos requisitos já levantados, será desenvolvida uma segunda versão do protótipo, obedecendo todas as considerações feitas até o momento (terceira fase).

O protótipo será uma espécie de jogo interativo, onde a criança deverá aprender as letras do alfabeto, fazendo associação de cada letra a um objeto ou coisa, através do som produzido por ele, com a letra “c”, por exemplo, seria um “carro“, que produz um som único “Vrum vrum”. O jogo simulará um tabuleiro, onde cada parada representa uma letra. A criança deve ouvir o som e fazer associação da letra com o objeto que a representa.

A comunicação entre a criança e o jogo será feita através de comandos de voz e de toques na tela, onde o jogo “conversa” com a criança e a faz perguntas. Para iniciar a interação, a criança precisará apenas manter o dedo pressionado na tela para responder cada uma das perguntas apresentadas. A arquitetura de funcionamento proposto é ilustrada na Figura 2.



Figura 2. Arquitetura de funcionamento.

Os requisitos identificados até o momento para o jogo foram os seguintes: Entrar no Jogo, Salvar Jogo, Sair do Jogo, Iniciar Fase, Reiniciar Fase, Avançar Fase, Encerrar Fase, Escutar Desafio, Responder Desafio, Configurar Personagem, Apresentar Orientação e Acionar Ajudar. Tais requisitos são ilustrados através do diagrama de casos de uso da Figura 3.

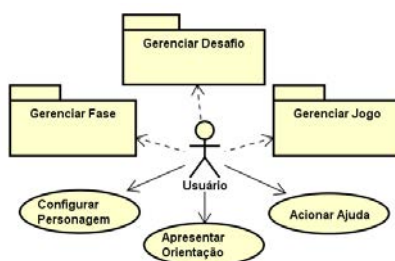


Figura 3. Diagrama de Caso de Uso.

Conforme citado anteriormente, os estudos ainda não foram finalizados, sendo este apenas o início dos resultados alcançados até o momento. A partir dos quais buscase fomentar discussões acerca do uso das tecnologias móveis no processo de alfabetização de crianças com deficiência visual.

Referências

BEHAR, Patrícia Alejandra; TORREZAN, Cristina A. W. **Parâmetros para a construção de materiais educacionais digitais do ponto de vista do design pedagógico**. In: BEHAR, Patrícia Alejandra (Org). Modelos Pedagógicos em Educação a Distância. Porto Alegre: Artmed, 2009. P.33-65.

IBGE – Instituto Brasileiro de Geografia e Estatísticas. **Censo 2010**. Disponível: <www.ibge.gov.br/home/presidencia/noticias/imprensa/ppts/00000009352506122012255229285110.pdf>. Acesso em 22 fev. 2015.

SILVA, Roberto A. D., PINHEIRO, Edson Pinheiro. **Proposta de Aplicativo em Dispositivos Móveis para Ensino de Leitura Baseado no Controle por Unidades Mínimas**. Universidade Federal do ABC (UFABC) Santo André – SP – Brasil, 2014. Disponível em: <www.tise.cl/volumen10/TISE2014/tise2014_submission_140.pdf> Acesso em 20 mai. 2015.

TORI, Romero. **Educação sem distância**. As tecnologias interativas na redução de distâncias em ensino e aprendizagem. São Paulo: Editora SENAC São Paulo, 2010.

Um Protocolo para Comunicação entre Aeronaves não Tripuladas e Estações de Solo

André Alves Dos Reis, Bruno César Vani, Kleber Manrique Trevisani

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) – Campus de Presidente Epitácio – Rua José Ramos Júnior, 27-50 – Jardim Tropical, Presidente Epitácio - São Paulo.

andrereis_live@hotmail.com, brunovani@ifsp.edu.br, kleber@ifsp.edu.br

Abstract. *This paper presents the development of a communication protocol that coordinate communication between unmanned aircraft and ground stations. This protocol defines structures of the messages involved in this context, which allow controlling aircraft with different characteristics.*

Resumo. *Este trabalho apresenta o desenvolvimento de um protocolo de comunicação para coordenar a comunicação entre aeronaves não tripuladas e estações de solo. O protocolo de comunicação desenvolvido define estruturas de mensagens envolvidas neste contexto, que possibilitam o controle de aeronaves com diferentes características.*

1. Introdução

Os Veículos Aéreos Não Tripulados (VANTs) são aeronaves sem tripulação que podem ser remotamente pilotadas ou autônomas, cuja utilização não seja meramente recreativa (STEINER, 2012; ANAC, 2012). Seu principal objetivo é realizar missões aéreas para reduzir custos e/ou riscos em comparação com operações realizadas por aeronaves tripuladas (NERIS, 2001). Como não necessitam de tripulação, os VANTs podem realizar missões que seriam impossíveis ou mais custosas para aeronaves tripuladas, como por exemplo, monitorar locais de difícil acesso.

Segundo Amaral, Nascimento Jr. e Adabo (2009) a operação ou monitoramento de um VANT requer a integração de três segmentos: segmento aéreo, segmento de comunicações e segmento de solo. O segmento aéreo refere-se a uma aeronave, incluindo toda sua estrutura física e os sistemas embarcados, podendo variar seu tipo de acordo com sua necessidade de aplicação (PASTOR; LOPEZ; ROYO, 2007). O segmento de solo é composto pela estação de controle de solo (*Ground Control Station*, GCS), que permite a visualização dos dados de voo pelo piloto e o envio de comandos e dados de planejamento dos voos autônomos para a aeronave (AUSTIN, 2010). O segmento de comunicações interliga os segmentos aéreo e de solo através de um enlace de dados em que podem ser transmitidos a telemetria (dados de voo enviados da aeronave para estação de solo), os comandos (enviados da estação de solo para a aeronave) e sinais de vídeo (da aeronave para a estação de solo). Este segmento engloba os transceptores de radiofrequência e os protocolos de comunicação utilizados.

Este trabalho apresenta o desenvolvimento de um protocolo de comunicação que permite a comunicação entre uma GCS e um VANT. O protocolo desenvolvido foi definido em nível da camada de enlace de dados em relação ao modelo de referência ISO/OSI (TANENBAUM; WETHERALL, 2010). Foram definidas regras de

delimitação de quadros, checagem de erros e as regras do protocolo para transmissão de telemetria e de telecomandos. O cenário de operação de um VANT pode incluir um canal de transmissão de sinais de vídeo, contudo, este tipo de transmissão não faz parte do escopo deste trabalho.

2. Descrição do Protocolo

O protocolo desenvolvido prevê a transmissão de dados entre a aeronave e o GCS. Dois grupos de dados são definidos: **dados de telecomando** e **dados de telemetria**.

O grupo de **dados de telecomando** possui quatro tipos de quadros: Quadro de Requisição de Configuração, Quadro de Configuração, Quadro Habilita Controle e Quadro de Telecomando. O Quadro de Requisição de Configuração é enviado da GCS para a aeronave, por solicitação do operador, para requisitar um Quadro de Configuração. O Quadro de Configuração, por sua vez, é enviado da aeronave para a GCS e descreve todos os controles da aeronave, incluindo suas características, que poderão ser manipulados pela estação de solo. Este quadro é importante, pois permite que uma GCS possa manipular aeronaves de diferentes tipos e configurações.

O *Quadro Habilita Controle* é enviado, inicialmente, da GCS para a aeronave, para solicitar uma permissão de controle, quando o operador deseja assumir o controle manual da aeronave. É importante evidenciar que não foi contemplada no desenvolvimento deste trabalho um quadro específico para realizar uma solicitação de voo autônomo. Contudo, dentro do *Quadro Habilita Controle* foi definido um atributo que define se o controle será realizado manualmente, pelo operador da GCS, ou autônomo, conforme plano de voo definido. Por fim, os *Quadros de Telecomando*, enviados da GCS para a aeronave, são montados de acordo com as configurações definidas no *Quadro de Configuração*. Dessa forma, o protocolo possibilita o envio dos comandos seguindo o formato especificado pela aeronave.

O grupo de **dados de telemetria** possui apenas um tipo de quadro: o *Quadro de Telemetria*. Esse quadro contém dados, transmitidos de forma contínua e repetitiva para a GCS, que permitem ao operador controlar a aeronave remotamente de forma segura ou monitorar um voo autônomo. Tais dados dependem da leitura contínua de dispositivos embarcados na aeronave, tais como giroscópios, barômetros, sensores de pressão e receptores posicionamento GNSS.

A técnica utilizada para delimitar os quadros é denominada *byte stuffing*. Tal técnica consiste em inserir bytes no início e fim de cada quadro para que fique explícito ao receptor onde inicia e onde termina cada quadro, permitindo assim contornar possíveis problemas de sincronização dos quadros em caso de problemas na transmissão. Para a detecção de erros, foi utilizado o algoritmo CRC-16 (*Cyclic Redundance Code* de 16 bits) (TANENBAUM; WETHERALL, 2010). Embora o protocolo de comunicação possua cinco tipos de quadros diferentes, todos seguem uma mesma estrutura: seu cabeçalho e rodapé permanecem idênticos para todos os quadros. Devido a limitação de espaço, detalhes sobre o formato dos quadros serão omitidos.

2.1. Lógica do Protocolo

Antes e durante um voo, é necessário que a aeronave e a GCS realizem trocas de quadros definidos pelo protocolo de comunicação desenvolvido, em uma sequência lógica definida. A Figura 1 demonstra essa sequência lógica.

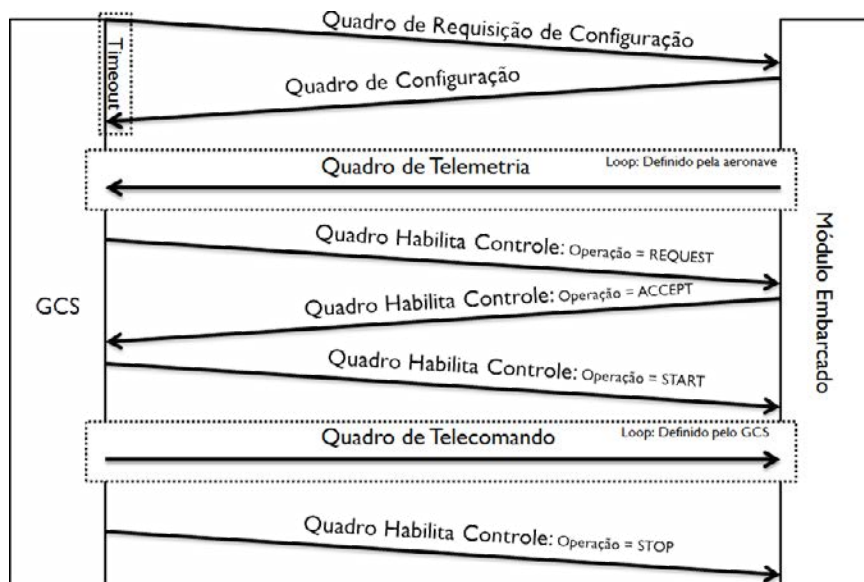


Figura 1. Troca de Quadros.

Inicialmente, a GCS envia uma solicitação de configuração inicial utilizando um *Quadro de Requisição de Configuração* e espera que a aeronave envie um *Quadro de Configuração*. Um *timeout* é definido na GCS para solicitar uma possível retransmissão em caso de falha no estabelecimento da comunicação (serviço confiável). Após o recebimento do *Quadro de Configuração*, a GCS conhecerá as configurações da aeronave e passa a receber os *Quadros de Telemetria*.

Quando solicitado pelo operador, a GCS transmite o *Quadro Habilita Controle* com o atributo Operação contendo o valor *REQUEST*. A aeronave responde com o mesmo tipo de quadro com o campo Operação contendo o valor *ACCEPT*. Por fim, a GCS encaminha outro quadro com o valor do campo Operação alterado para *START*. Este fluxo indica que a GCS solicita à aeronave a concessão dos comandos; a aeronave aceita o pedido e o GCS conclui o processo indicando que está iniciando o controle. A partir desse momento os *Quadros de Telecomando* podem então ser encaminhados para a aeronave.

Destaca-se que há um serviço confiável para os comandos com necessidade de confirmação, que podem ser úteis para acionamento, por exemplo, de câmeras, paraquedas e gaiolas. Já os comandos relacionados ao controle da aeronave não necessitam ser confirmados, pois o mais importante é a fluência dos comandos. A utilização de um serviço confiável apenas quando necessário implica diretamente em uma melhor utilização da largura de banda disponível. Eventualmente, para que o controle seja reestabelecido para um transmissor de rádio controle, por exemplo, o operador solicitará que o GCS desabilite o comando da aeronave, logo o GCS transmitirá um *Quadro Habilita Controle* com o valor *STOP* no atributo Operação.

3. Considerações Finais

A proposta realizada por este trabalho contempla o desenvolvimento de um protocolo para controlar a comunicação entre uma GCS e uma aeronave. Tal protocolo permite o controle de aeronaves com características variadas e demonstrou ser capaz de ser utilizado para controlar um VANT a partir de uma GCS. Nos testes realizados o protocolo se comportou da maneira esperada, porém é necessária uma certa cautela sobre a capacidade de processamento e limites de memória nos sistemas embarcados onde o protocolo é implementado. Devido a flexibilidade do protocolo permitir o controle de aeronaves com diferentes configurações, alguns quadros não possuem tamanho definido (tamanho dinâmico), isso implica em mais processamento e alocação memória.

O contexto deste trabalho envolveu características que não foram possíveis de serem implementadas, mas não foram significativamente relevantes para a fase atual de desenvolvimento. Uma característica desejável e não implementada é a possibilidade da aeronave informar a GCS sobre quais serão os dados de telemetria transmitidos, permitindo que a mesma represente somente os dados enviados.

Algumas melhorias também podem ser implementadas em projetos futuros, como por exemplo, a implementação de algoritmos de criptografia de dados para evitar quebras de confidencialidade e integridade. Também se sugere a investigação de métodos de autenticação, como por exemplo, certificação digital, para que a GCS possa autenticar-se de forma segura antes de controlar a aeronave.

4. Referências

- ANAC - AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **Instrução Suplementar: IS 21-002A**. 2012.
- AMARAL, G. F.; NASCIMENTO JR, C. L.; ADABO, G. J. Predição e Melhoria da Confiabilidade do Sistema de Controle de Voo de um Veículo Aéreo Não-Tripulado. In: **2009 Brazilian Symposium on Aerospace Eng. & Applications / 3o CTA-DLR Workshop on Data Analysis & Flight Control**, 2009, São José do Campos. Anais... São José do Campos, 2009.
- AUSTIN, R. **Unmanned Aircraft Systems: UAVs Design, Development and Deployment**. 1. ed. Reino Unido: John Wiley & Sons Ltd. 2010.
- NERIS, L. O. **Um Piloto Automático para as Aeronaves do Projeto ARARA**. 2001. 113 f. Dissertação (Mestrado) – Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo, São Carlos.
- PASTOR, E; LOPEZ, J; ROYO, P. A Hardware/Software Architecture for UAV Payload and Mission Control. In: **Digital Avionics Systems Conference, 2006 IEEE/AIAA**, 25, 2006, Portland. Anais... Portland: IEEE, 2006. p. 1-8.
- STEINER, E. M. **Gerenciamento de Configuração de uma Linha de Produtos de Software de Veículos Aéreos não Tripulados**. 2012. 121 f. Dissertação (Mestrado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo.
- TANENBAUM, A. S.; WETHERALL, D. J. **Computer Networks**. 5. ed. Pearson Prentice Hall, 2010.

Minimização da Intervenção Humana para Detectar e Solucionar Anomalias Rede de Computadores

Alexandre Amaral¹, Ana Paula Malheiro²

¹Instituto Federal de Santa Catarina (IFSC)
CEP 89254-430 – Jaraguá do Sul – SC – Brasil

²Instituto Federal Catarinense (IFC)
Caixa Postal 2016 – Camboriú – SC – Brasil

alexandre.amaral@ifsc.edu.br, anapaula@ifc-camboriu.edu.br

Abstract. *This paper proposes a system to detect and apply corrective actions when anomalous events occur in the network. Management goals are defined through the metrics and the self-healing property of the autonomic computing is used, empowering the system to perform corrective actions without human intervention. NEMES (Network Metric Specification) a domain-specific language was developed to build the metrics. The system uses IP flows to reduce the volume of data to be processed, allowing its use in large-scale networks. Tests performed in a real environment have shown the effectiveness and potential of the proposed system to assist in the network management.*

Resumo. *Neste trabalho é proposto um sistema para detectar e aplicar ações corretivas na ocorrência de eventos anômalos na rede. Os objetivos da gerência são definidos através de métricas e a propriedade de autorreparo da computação autônoma é utilizada para que o sistema execute ações corretivas sem a intervenção humana. NEMES (Network Metric Specification), uma linguagem de domínio específico foi desenvolvida para a escrita das métricas. O sistema utiliza fluxos IP que reduz o volume de dados a serem processados, permitindo sua utilização em redes de grande escala. Testes realizados em um ambiente real demonstraram a eficácia e o potencial do sistema proposto para auxiliar no gerenciamento de rede.*

1. Introdução

Esforços têm sido realizados pela comunidade científica a fim de desenvolver mecanismos que auxiliem no monitoramento e na segurança da rede. Algumas soluções do tipo NIDS (*Network-based Intrusion Detection System*) têm sido propostas. Porém, [Bhuyan *et al.* 2014] apresentaram recentemente diversos fatores que limitam a aplicação desses mecanismos para funcionamento em tempo real em redes de grande escala. A inspeção de todos os pacotes transmitidos pela rede requer um alto processamento e espaço de armazenamento. Os alarmes emitidos na ocorrência de um evento anômalo requer uma análise manual do administrador de rede para que o problema seja reparado. A aplicação dos NIDS em um escopo mais amplo de gerenciamento não tem sido endereçada. Isto inclui o monitoramento de serviços e usuários, identificação de pontos de gargalo na rede, além de não fornecerem um meio para definição e monitoramento de KPIs (*Key Performance Indicator*) específicos, como *uptime/downtime* de um recurso de rede.

Nesse trabalho é apresentado um sistema para detectar eventos anômalos na rede e aplicar ações corretivas, com funcionamento em tempo real. Os objetivos da gerência são definidos através de métricas, e o sistema aplica a propriedade de autorreparo da computação autônoma [Magalhães e Silva 2013], a fim de que as ações corretivas e/ou preventivas sejam executadas sem a intervenção humana. NEMES (*Network Metric Specification*), uma linguagem de domínio específico foi desenvolvida para a escrita das métricas. Diferentemente das propostas que utilizam como fonte de dados os pacotes IP, o sistema proposto utiliza os fluxos IP. Esta escolha visa reduzir o volume de dados a serem processados e a aplicação do sistema em redes de grande escala. Testes realizados em um ambiente real demonstraram a eficácia e o potencial desse sistema para auxiliar no gerenciamento de rede.

2. Sistema Proposto

A Figura 1 apresenta os principais componentes do sistema. Os fluxos IP coletados dos dispositivos ativos na rede são processados pelo *FE* (*Feature extractor*). O *FE* obtém as propriedades dos fluxos (e.g., endereço IP de origem, número de pacotes) enviando os para o componente *ME* (*Metric processor*). Utilizando as métricas lidas da *KB* (*Knowledge base*), as propriedades dos fluxos são analisadas pelo *ME*, com o objetivo de identificar os fluxos anômalos ou indesejáveis. As métricas contêm as condições das propriedades dos fluxos considerados anômalos e as ações que devem ser executadas caso eles ocorram na rede. Quando um fluxo é considerado anômalo a ação especificada na métrica é executada pelo componente *AP* (*Action processor*), e o resultado de sua execução é armazenado na *KB*.

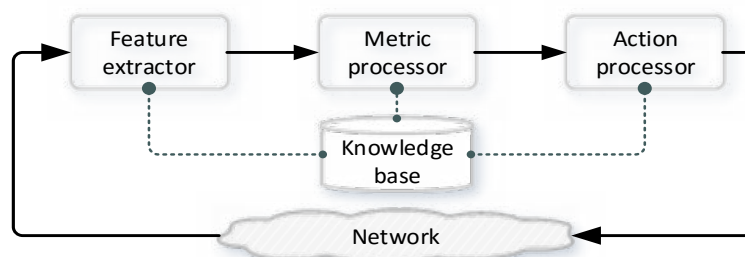


Figura 1. Principais componentes do sistema.

Neste trabalho, uma métrica é uma condição definida pelo administrador que caso ela ocorra na rede uma ação é disparada. A Figura 2 apresenta uma métrica para detecção de um ataque *Ping of Death*. Este ataque consiste no envio de fluxos ICMP com pacotes maiores do que 65535 bytes. Tais características são definidas na *condition* da métrica. Na ocorrência desse ataque, a métrica especifica duas ações a serem realizadas: *execute* e *notify*. O método *execute* aplica uma ação corretiva e/ou preventiva. No exemplo, um *script* será executado com o propósito de bloquear o ataque. Simultaneamente, uma notificação será enviada para o administrador de rede através do método *notify*. O meio de comunicação utilizado no exemplo é o e-mail. Entretanto, o sistema está sendo desenvolvido como uma interface que permita a utilização de outros meios, tais como SMS (*Short Message Service*) e *log*.

```

{
  "metric": {
    "id": "030511",
    "status": "active",
    "description": "Detecção de ataques Ping of Death",
    "condition": {
      "and": [
        {"gt": {"sizepackets": 65535}},
        {"eq": {"proto": "ICMP"}}
      ]
    },
    "action": {
      "methods": [
        {
          "name": "execute",
          "process": "/action/scripts/blockPingOfDeathAttack.sh"
        },
        {
          "name": "notify",
          "mean": "email",
          "msg": "Ping of Death detectado. Fonte %srcip% e o(s) destino(s) %dstip%."
        }
      ]
    }
  }
}

```

Figura 2. Exemplo de uma métrica para detectar um ataque *Ping of Death*.

Uma linguagem de domínio específico, denominada NEMES (*Network Metric Specification*) foi desenvolvida para a criação das métricas. NEMES utiliza a notação JSON como mostra a Figura 2. A escolha do JSON dentre outras linguagens (e.g., XML), se por algumas razões. O JSON é mais leve, mais fácil para leitura e escrita por humanos e por estar sendo amplamente utilizado nos últimos anos [Yang 2012]. Antes de serem disponibilizadas na base de conhecimento, as métricas escritas pelo administrador de rede são processadas pelo compilador NEMES, desenvolvido com o propósito de realizar a análise léxica, análise sintática e análise semântica.

2.1. Aplicação do autorreparo

Quando a condição de uma métrica é atendida o componente *AP (Action Processor)* é acionado. Através deste componente, o sistema proposto implementa o método *execute* com o objetivo de aplicar a propriedade de autorreparo da computação autônoma. No contexto do gerenciamento de redes, uma gama de ações pode ser definida pelo administrador para fins de autorreparo. O bloqueio de acesso e a finalização de uma sessão de usuário pelo abuso dos recursos de um servidor são procedimentos que podem ser especificados, a fim de que os serviços prestados pela rede continuem operando sem danos.

3. Estudo de Caso

Testes em um ambiente real de rede foram realizados com o propósito de validar o sistema proposto. Para este fim, um *testbed* foi montado, como ilustra a Figura 3. Fluxos IP são coletados do gateway/firewall através do protocolo NetFlow v9 a cada cinco minutos. O sistema foi desenvolvido com a linguagem Java. Para a base de conhecimento foi utilizada o banco de dados Neo4J [Neo4J 2016].

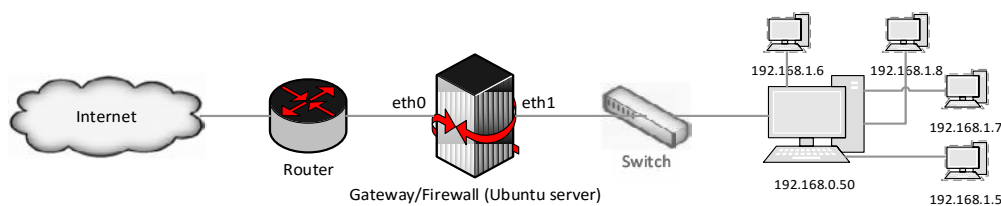


Figura 3. Testbed para realização dos testes.

Nos experimentos, as métricas foram criadas para a detecção de *Land attack*, *Smurf attack*, detecção de acesso não permitido a diversos sites proibidos pela instituição, incluindo redes sociais e de conteúdo adulto. Para geração dos ataques *Land attack* e *Smurf attack* foi utilizada a ferramenta Hayane [Hyenae 2016] lançados a partir das máquinas virtuais hospedadas na máquina 192.168.0.50. Para aplicação das ações de autorreparo foi utilizado o firewall Iptables instalado na máquina Gateway/Firewall.

Os testes realizados demonstraram a eficácia do sistema para detectar e aplicar as correções especificadas nas métricas. As ações testadas correspondem ao método *execute* e *notify*. Em todos os testes foram utilizados *scripts* para bloquear os fluxos indesejados via firewall. O método *notify* foi testado utilizando o e-mail e o armazenamento textual do resultado da aplicação da ação em arquivos de *log*.

4. Conclusão

Nesse trabalho foi apresentado um sistema para detecção de atividades anômalas na rede. Através de uma linguagem de domínio específica denominada NEMES, métricas podem ser escritas para que o sistema identifique atividades que vão além dos ataques e ações maliciosas endereçadas pelas soluções atuais. A utilização da propriedade de autorreparo da computação autonômica permite que o sistema monitore, detecte as ações especificadas pelas métricas e execute ações para solucioná-las. Os testes demonstraram a capacidade do sistema de tomar ações sem a intervenção humana, auxiliando o administrador na árdua tarefa do gerenciamento de rede. Como trabalhos futuros, desejamos explorar outras propriedades da computação autonômica, realizar mais testes e utilizar o sistema de forma distribuída.

5 Referências

- Bhuyan, M.H. Bhattacharyya, D.K. e Kalita, J.K. (2014) “Network Anomaly Detection: Methods, Systems and Tools”. *Communications Surveys & Tutorials*, IEEE , vol.16, no.1, p. 303-336, 2014.
- Hyenae. Disponível em: <http://sourceforge.net/projects/hyenae/>
- Magalhães J. P. e Silva L. M., "Self-healing Performance Anomalies in Web-based Applications," *Network Computing and Applications (NCA)*, 2013 12th IEEE International Symposium on, Cambridge, MA, pp. 81-88, 2013.
- Neo4j. “The World's Leading Graph Database”. Disponível em: <http://www.neo4j.org/>
- Yang Y. “Impact data-exchange based on XML”, *Computer Science & Education (ICCSE)*, 7th International Conference. p.1147-1149, 2012.

A teoria de redes como subsídio para o estudo da colaboração em projetos de pesquisa

Antônio Raphael Brito Meireles¹, Ingrid dos Santos Vilas Verde¹, Maria Teresinha Tamanini Andrade¹

¹Instituto Federal de Educação Ciência e Tecnologia da Bahia – IFBA
Via Universitária, s/n - Pintaguinha, Simões Filho - BA, 43700-000 – Brazil

arb.meireles@outlook.com, vilas.ingrid@gmail.com, tamanini@ifba.edu.br

Abstract. *In this paper we study the collaborative networks of researchers who participate in research projects in an academic community. Methods and indices Network Theory were used for the analysis. The results suggest networks topologically characterized as small world. This means that it is an enabling environment for collaboration, since it favors the articulation among researchers.*

Resumo. *Neste artigo estudamos as redes de colaboração formadas por pesquisadores que participam de projetos de pesquisa em uma comunidade acadêmica. Métodos e índices da Teoria de Redes foram utilizados para a análise. Os resultados sugerem redes caracterizadas topologicamente como mundo pequeno (small-world), gerando um ambiente propício à colaboração, pois favorece a articulação entre os pesquisadores.*

1. Introdução

A colaboração científica é um tema que vem sendo estudado mais intensamente nos últimos anos, mas não é recente [Andrade et al. 2015, Andrade et al. 2014]. Segundo [Sonnenwald 2006] colaboração é a interação que facilita não só a realização de tarefas, mas, também, o compartilhamento do significado desta tarefa, relacionada a um objetivo maior compartilhado entre dois ou mais cientistas.

O estudo de redes tem suas origens na teoria dos grafos, um ramo da matemática. Assim, uma rede é um grafo constituído de um conjunto de elementos chamados vértices ou nós, que são ligados por outro conjunto de elementos chamados de arestas que fazem conexões com os vértices.

Uma rede social é um conjunto de pessoas ou grupos de pessoas com algum padrão de contato ou interação entre elas [Wasserman e Faust, 1997]. Segundo [Barabási 2003], uma rede complexa é um grafo que apresenta uma estrutura ou propriedades topológicas não triviais, composto por um conjunto de vértices (nós) que são interligados por meio de arestas.

Com os índices da teoria das redes é possível caracterizar uma rede como aleatória, livre de escala ou *small world*, que são os modelos mais difundidos. Uma rede apresenta o comportamento *small world* se seu coeficiente de aglomeração médio for

muito maior que o coeficiente de aglomeração de uma rede aleatória e se seu caminho mínimo médio for comparável com o caminho mínimo médio da rede aleatória correspondente.

Para discutir e caracterizar as relações de colaboração foi utilizada três medidas de centralidade: centralidade de grau, centralidade de proximidade e centralidade de intermediação.

2. Objetivo

O objetivo desta pesquisa é estudar as redes de colaboração formadas por pesquisadores que participam dos projetos de pesquisa em uma comunidade acadêmica enquanto parâmetro para compreender como ocorre a colaboração nestas comunidades, identificar o modelo de rede associado e as implicações desse modelo nas comunidades às que pertencem os pesquisadores.

3. Metodologia

Analisamos um Programa de Pós-Graduação na área disciplinar da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). A escolha deste PPG deve-se a presença de pesquisas de mestrado e doutorado e ao fato de possuir conceito 5 (cinco) pela CAPES. Os cadernos de indicadores de Projeto de Pesquisa são preparados com base nos dados que os Programas de Pós-Graduação informam anualmente.

Os cadernos foram obtidos no formato PDF e neles foram coletados os dados referentes aos projetos de pesquisa (título, participantes do projeto, período, financiamento e situação).

Estudamos os cadernos de indicadores de quatro triênios (2001/2003, 2004/2006, 2007/2009 e 2010/2012); e o período completo (2001-2012). Cabe ressaltar que foi considerado para análise o período das avaliações trienais e relatórios da coleta CAPES a partir do início das atividades do Comitê Internacional Interdisciplinar da CAPES. Foram criados arquivos em formato TXT para cada triênio e outro para o período completo com os nomes dos pesquisadores dos respectivos projetos.

Utilizamos o software Ucinet para gerar redes no formato Pajek; a seguir foram calculados, com o auxílio dos softwares Ucinet e Pajek, os índices e parâmetros da Teoria de Redes.

4. Discussão dos Resultados

Nesta seção apresentamos alguns resultados referentes às redes de Projetos de Pesquisa por triênios e do período total, bem como os índices e os seus parâmetros da Teoria de Redes.

Na Figura 1 apresentamos a rede de Projetos de Pesquisa. Os vértices são os pesquisadores e as arestas são as colaborações dos participantes nos projetos de pesquisa.

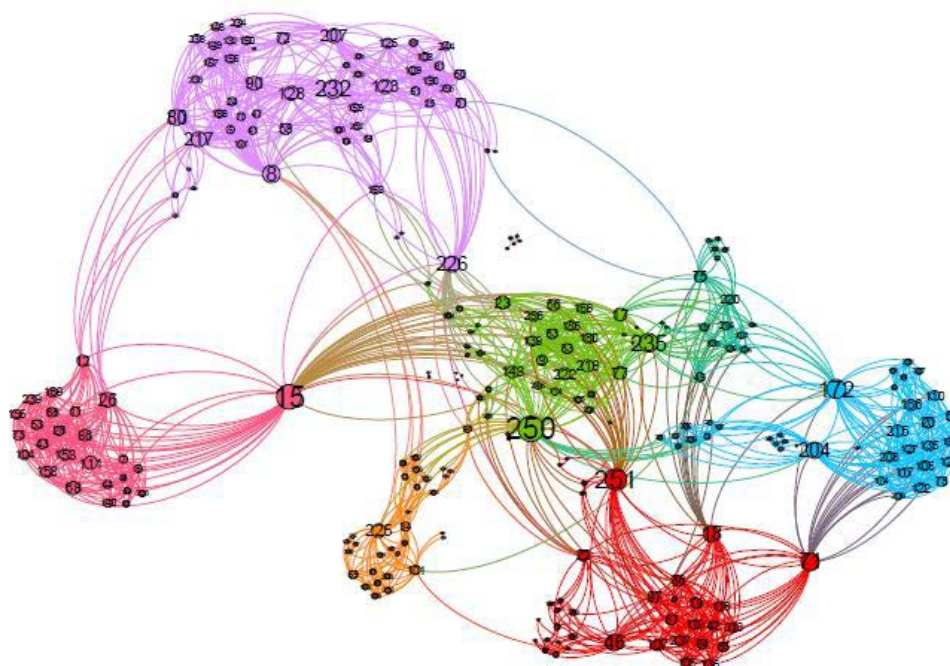


Figura 1. Rede de Projetos de Pesquisa

Na Tabela 1 apresentamos os índices e parâmetros da Teoria de Redes para Projetos de Pesquisa. O C_r é o coeficiente de aglomeração da rede aleatória e o L_r o caminho mínimo médio da rede aleatória.

Tabela 1. Índices de Projetos de Pesquisa

ÍNDICES	TRIÊNIO 1	TRIÊNIO 2	TRIÊNIO 3	TRIÊNIO 4	PERÍODO TOTAL
Vértices	13	109	135	135	256
Componentes	11	2	3	8	6
Densidade	0.0256	0.117	0.060	0.081	0.051
Diâmetro	2	7	9	7	6
Coefficiente de aglomeração C	0.000	0.876	0.883	0.873	0.833
C_r	0.000	0.129	0.066	0.088	0.051
Caminho mínimo médio L	1.250	2.060	2.570	2.276	2.476
L_r	1.333	2.603	4.446	3.000	3.065
Grau Médio	0.307	12.642	8.059	10.874	13.031

As redes estudadas são caracterizadas topologicamente como redes mundo pequeno (*small-world*). Isso significa que é um ambiente propício à colaboração, pois favorece a articulação entre os pesquisadores. O coeficiente de aglomeração alto ($0 < C < 1$) retrata uma situação onde existe uma forte interlocução entre os pesquisadores; o caminho mínimo baixo sugere que o grupo de pesquisadores é ágil em relação ao acesso e contato entre os pesquisadores.

Nos índices obtidos para as centralidades de grau (CG), de proximidade (CP) e de intermediação (CI) observamos pesquisadores que se destacam. Os docentes P18, P142, P251 e P253 (Figura 1) são os pesquisadores com maior destaque nas redes de colaboração, visto que aparecem em todos os índices de centralidades. Observa-se que a presença desses integrantes é fundamental para o desenvolvimento da colaboração e difusão das informações na rede. Estes pesquisadores que se destacam nos índices analisados são vértices importantes, podendo interagir rápido com os demais e exercendo controle nesta, formando um núcleo de pesquisadores que fomentam o desenvolvimento e o fortalecimento das redes do programa interna e externamente.

5. Considerações Finais

Houve um aumento significativo do número de pesquisadores ao longo dos triênios, logo, ao analisar triênios mais recentes é possível observar maior quantidade de trabalhos sendo desenvolvidos. Com isso é possível constatar que alguns fatores como, maior investimento financeiro e/ou demanda externa ou amadurecimento da comunidade científica ocasionaram maior contingente de pesquisadores atuando e desenvolvendo Projetos de Pesquisa.

As redes obtidas nos dois primeiros triênios possuem características muito comuns. Uma das principais é o fato de terem poucos vértices quando comparadas as redes maiores, como a rede dos triênios 3 e 4. Isso pode ser interpretado como uma fragilidade momentânea devido ao programa estar iniciando e as redes de colaboração ainda estarem por se formar. A metodologia de análise de redes é uma ferramenta que facilita a identificação da função dos participantes nas redes de projetos de pesquisa.

Referências

- Andrade, M. T. T., Rosa, M. G., Fadigas, I. S., Pereira, H. B. B. (2015). “O papel dos docentes em programas de pós graduação: uma abordagem baseada em redes”, In CSBC-BraSNAM, p. 1-6.
- Andrade, M. T. T., Braga, P. F., Carneiro, T. K. G., Ribeiro, N. M., Moret, M. A., Pereira, H. B. B. (2014). “Contextualized Analysis of Social Networks: Collaboration in Scientific Communities”, *Social Networking*, v. 3, p. 71-79.
- Barabási, A.L. (2003). “Linked: How everything is connected to everything else and what it means for business, science and everyday life”. Plume.
- Rosa, M. G., Fadigas, I. S., Andrade, M. T. T., Pereira, H. B. B. (2014). “Clique Approach for Networks: Applications for Coauthorship Networks”. *Social Networking*, v. 3, p. 80-85.
- Sonnenwald, D. H. (2006). “Scientific Collaboration”, In: *Annual Review of Information Science and Technology*, Medford, NJ: Information Today, ed. B. Cronin, v. 41, p. 643–681
- Wasserman, S., Faust, K. (1997). “Social network analysis: methods and applications”. Cambridge: Cambridge University Press.

Multi-Objective Test Case Selection: Local Search Approaches for the NSGA-II algorithm

Luciano Soares de Souza¹

¹Instituto Federal do Norte de Minas Gerais (IFNMG) - *Câmpus* Pirapora
R. Dr. Humberto Malard, 1355, 39270-000, Pirapora - MG - Brasil

luciano.souza@infnmg.edu.br

Abstract. *The software testing process can be very expensive and it is important to find ways in order to reduce its costs. Test case selection techniques can be used in order to reduce the amount of tests to execute and this way reducing the costs. Search algorithms are very promising approach to deal with the test case selection problem. This work proposes new hybrid algorithms for multi-objective test case selection by adding local search mechanisms into the NSGA-II algorithm. The results showed that some of the mechanisms were capable of improve the NSGA-II algorithm.*

Resumo. *O processo de testes de software pode ser bastante caro, portanto é importante se encontrar formas de reduzir os custos desse processo. Técnicas de seleção de casos de teste podem ser usadas de forma a reduzir a quantidade de testes a executar e dessa forma reduzir os custos. A utilização de algoritmos de busca são uma maneira promissora de resolver o problema de seleção de casos de teste. Esse trabalho propõe novos algoritmos híbridos de busca para seleção multiobjetivo de casos de teste através da inserção de mecanismos de busca local no algoritmo NSGA-II. Os resultados mostraram que alguns dos mecanismos foram capazes de introduzir melhoras no algoritmo NSGA-II.*

1. Introduction

Software testing is an expensive and time consuming process, which may reach about 40% of total costs involved in software development [Ramler and Wolfmaier 2006]. As such, automation emerges as the key solution for improving the efficiency and effectiveness of the testing process, as well as to reduce its costs.

Fortunately, it is possible to identify, in test suites, redundant test cases concerning a piece of code. Thus, we can envision ways to reduce the suites in order to fit the available resources without seriously compromising the coverage of the adequacy criterion, and thus the quality of the testing process.

The task of reducing a test suite based on a *selection criterion* is known as *Test Case selection*. Given an input test suite, test case selection aims to find a relevant subset regarding the adopted test adequacy criterion, such that the test cases that do not improve the reduced suite coverage can be eliminated. Clearly, the selection criterion relies upon the coverage of the adopted adequacy criterion.

Test case selection is not easy or trivial since there may be a large number of combinations to consider when searching for an adequate subset. A very promising approach to deal with this problem relies upon the use of search optimization techniques

(see [Harman 2011]), which is the focus of this research. Here, the aim is to search for a subset of test cases which optimizes a given objective function (i.e., the given selection criterion).

Within this work we addressed the test case selection problem using two selection criteria: (1) branch coverage, and (2) execution cost (time). In this light, the TC selection problem here was treated as a multi-objective optimization problem.

The focus of this work was to investigate whether local search mechanisms were capable of improve the results of the Non-Dominated Sorting Genetic Algorithm (NSGA-II) [Deb et al. 2000]. For that, we combined three local search strategies with the NSGA-II algorithm in order to create three new hybrid algorithms: (1) NSGA-II-FSBE, (2) NSGA-II-1opt and (3) NSGA-II-AG.

2. Problem Formulation

In this work, the solution is defined as a binary vector representing a candidate subset of TCs to be applied in the software testing process. Let $T = \{T_1, \dots, T_n\}$ be a test suite with n test cases. A solution is defined as $\mathbf{t} = (t_1, \dots, t_n)$, in which $t_j \in \{0, 1\}$ indicates the presence (1) or absence (0) of the test case T_j within the subset of selected TCs.

As said, two objective functions were adopted: branch coverage and execution cost. The branch coverage (function to be maximized) consists of the ratio (in percentage) between the amount of code branches covered by a solution \mathbf{t} and the amount of branches covered by T . In turn, the execution cost (function to be minimized) represents the amount of time required to execute the selected suite. Finally, the proposed algorithms are used to deliver a good Pareto frontier regarding the objective functions.

3. Local Search Approaches

Generally speaking, local search algorithms choose, at each step, the locally best node (which yields the best objective evaluation). The local search algorithms used in this work are the Forward Selection (FS) [Webb 2002], the Backward Elimination (BE) [Webb 2002], the 1-opt [Papadimitriou and Steiglitz 1998] and the Additional Greedy (AG) algorithm [Elbaum et al. 2000].

3.1. NSGA-II-FB

In order to create the NSGA-II-FB algorithm, we introduced the local search procedure as the last step of the main loop as follows:

1. Select 10% of the non-dominated solutions by using a Roulette Wheel;
2. For each selected solution \mathbf{t} randomly select one objective to improve
3. *IF* the selected objective is to be maximized, *THEN* use the FS local search algorithm; *ELSE* (if it is to be minimized) use the BE local search algorithm.

The FS algorithm takes the solution \mathbf{t} and iterates as follows: for each test case t_j not yet present in the current solution \mathbf{t} (i.e., for each $t_j = 0$) a new candidate solution \mathbf{t}' is produced by setting $t_j \leftarrow 1$. For each candidate solution, the previously chose objective function is computed. The candidate solution which yields the highest objective value and is not dominated is then adopted as the new current solution in the search process.

The algorithm stops (1) when all candidate solutions found at an the current iteration are dominated, or (2) when all test cases have been already added to the current solution. Contrarily, the BE algorithm takes the solution \mathbf{t} and iterates as follows: for each TC present in the current solution (i.e., for each $t_j = 1$) a candidate solution \mathbf{t}' is produced by setting $t_j \leftarrow 0$. The objective function is computed and the candidate solution which yields the lowest objective value and is not dominated is considered as the current solution for the next iteration.

It is important to highlight that all non-dominated solutions found during the local search process are stored within the non dominated solutions of the NSGA-II algorithm.

3.2. NSGA-II-1opt

The NSGA-II-1opt algorithms was created by adding the 1-opt algorithm as local search in the following way:

1. Select 10% of the non-dominated solutions stored in the EA by using a Roulette Wheel;
2. For each solution \mathbf{t} do:
 - (a) For each test case $t_j \in \mathbf{t}$ do:
 - i. Flip the bit of t_j (i.e if $t_j = 0$ then $t_j \leftarrow 1$, or the contrary) in order to create the neighborhood solution \mathbf{t}' ;
 - ii. Evaluate \mathbf{t}' according to the objectives and store if it is a non-dominated solution;

3.3. NSGA-II-AG

In order to create the NSGA-II-AG algorithm we adapted the idea of the aforementioned traditional Additional Greedy algorithm in the following way:

1. Select 10% of the non-dominated solutions stored in the EA by using a Roulette Wheel;
2. For each solution \mathbf{t} do:
 - (a) REPEAT
 - i. For each test case $t_j \in \mathbf{t}$ where $t_j = 0$ do:
 - A. Creates \mathbf{t}' by making $t_j = 1$;
 - B. Computes the additional coverage per cost of \mathbf{t}' ;
 - ii. Choose the \mathbf{t}' with the best additional coverage per cost value and update the current solution $\mathbf{t} = \mathbf{t}'$
 - (b) UNTIL it is not possible find a \mathbf{t}' that enhances \mathbf{t}
 - (c) Store each non-dominated solution found during the previous steps

4. Experiments and Results

The experiments were performed using 4 programs (*flex*, *grep*, *sed* and *space*) from the SIR (Software-artifact Infrastructure Repository). Each algorithm (NSGA-II, NSGA-II-FB, NSGA-II-1opt and NSGA-II-AG) was executed 100 times with a total of 50000 functions evaluations and they used the parameters' values proposed in [Deb et al. 2000]. Furthermore, three well known multi-objective metrics (Hypervolume - HV, Generational Distance - GD and Inverted Generational Distance - IGD) were used in order to evaluate the algorithms.

Table 1. Mean and standard deviation values for each algorithm and each metric.

	HV	GD	IGD		HV	GD	IGD		HV	GD	IGD		HV	GD	IGD				
flex	NSGA-II	0.793 (0.012)	0.018 (0.004)	0.021 (0.001)	grep	NSGA-II	0.707 (0.013)	0.025 (0.005)	0.025 (0.001)	sed	NSGA-II	0.776 (0.021)	0.030 (0.008)	0.019 (0.001)	space	NSGA-II	0.852 (0.017)	0.003 (0.001)	0.016 (0.001)
	NSGA-II-FB	0.909 (0.064)	0.018 (0.011)	0.009 (0.006)		NSGA-II-FB	0.769 (0.072)	0.033 (0.013)	0.018 (0.007)		NSGA-II-FB	0.940 (0.085)	0.008 (0.009)	0.004 (0.006)		NSGA-II-FB	0.976 (0.001)	0.001 (0.001)	0.001 (0.001)
	NSGA-II-1opt	0.763 (0.011)	0.024 (0.003)	0.022 (0.001)		NSGA-II-1opt	0.672 (0.012)	0.037 (0.005)	0.029 (0.001)		NSGA-II-1opt	0.788 (0.021)	0.027 (0.005)	0.016 (0.001)		NSGA-II-1opt	0.920 (0.009)	0.004 (0.001)	0.006 (0.001)
	NSGA-II-AG	0.697 (0.12)	0.027 (0.009)	0.024 (0.001)		NSGA-II-AG	0.592 (0.011)	0.054 (0.010)	0.038 (0.001)		NSGA-II-AG	0.670 (0.020)	0.069 (0.010)	0.024 (0.001)		NSGA-II-AG	0.746 (0.024)	0.015 (0.004)	0.019 (0.001)

Table 1 shows the mean and standard deviation values for each algorithm and for each metric. The values were compared using the Mann-Whitney U statistical test and the best value for each metric is highlighted.

According to the results from the Table 1, it is possible to see that the NSGA-II-FB outperformed the others in almost all cases. Hence, the FB local search mechanism indeed was able to improve the results of the NSGA-II algorithm in the studied cases. In turn, the 1-opt mechanism was able to improve the results only in some cases. Finally, the AG mechanism was not able to improve the results of the NSGA-II algorithm. In fact, the results of the NSGA-II-AG algorithm were worse than the NSGA-II results.

5. Conclusions

This paper proposes the use of three local search mechanisms in order to create new hybrid algorithms. These algorithms were used in the context of multi-objective test case selection. Experiments, using programs from the SIR repository, were performed in order to evaluate whether the proposed hybrid algorithms were capable of improving the results of the NSGA-II algorithm.

The results showed that the NSGA-II-FB algorithm was able to improve the results of the NSGA-II algorithm in almost all observed cases. Hence it indicates that the FB local search mechanism indeed improved the NSGA-II algorithm. Furthermore, the NSGA-II-FB algorithm was considered the best. Further studies will expand these experiments in order to allow the comparison of the algorithms in more diverse environments.

References

- Deb, K., Agrawal, S., Pratap, A., and Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In *Parallel Problem Solving from Nature PPSN VI*, volume 1917 of *Lecture Notes in Computer Science*, pages 849–858. Springer Berlin Heidelberg.
- Elbaum, S., Malishevsky, A. G., and Rothermel, G. (2000). Prioritizing test cases for regression testing. *SIGSOFT Softw. Eng. Notes*, 25(5):102–112.
- Harman, M. (2011). Making the case for morto: Multi objective regression test optimization. In *Fourth International IEEE Conference on Software Testing, Verification and Validation*, pages 111–114. IEEE Computer Society.
- Papadimitriou, C. and Steiglitz, K. (1998). *Combinatorial optimization: algorithms and complexity*. Dover books on mathematics. Dover Publications.
- Ramler, R. and Wolfmaier, K. (2006). Economic perspectives in test automation - balancing automated and manual testing with opportunity cost. In *Workshop on Automation of Software Test, ICSE 2006*.
- Webb, A. R. (2002). *Statistical Pattern Recognition, 2nd Edition*. John Wiley & Sons.

I ETC – I Encontro de Teoria da Computação

Apresentação

Este ETC é a primeira edição de um evento voltado para a grande área de Teoria da Computação, proposto e organizado por membros da Comissão Especial em Algoritmos, Combinatória e Otimização (CE-ACO), cujo propósito maior é integrar e aproximar mais efetivamente os pesquisadores da área e a área à Sociedade Brasileira de Computação (SBC), bem como promover a sua divulgação entre a comunidade de pesquisadores, profissionais e estudantes que participam dos congressos da SBC. Com o entendimento claro desse propósito, os pesquisadores e estudantes da área responderam à chamada de trabalhos com verdadeiro entusiasmo: foram 40 trabalhos submetidos, de qualidade inegável, levando a organização a estender o período de realização do evento em mais um dia, de forma a trazer para apresentação uma amostra da qualidade da pesquisa na área no Brasil. Como resultado, foram aceitos 34 trabalhos, dos quais 28 serão apresentados ao longo dos dias 04 e 05 de julho de 2016, intercalados por palestras de pesquisadores renomados da área. Os demais trabalhos serão apresentados como pôsteres. Todos os trabalhos submetidos tiveram entre 2 e 3 pareceres emitidos por especialistas, mobilizando 62 pesquisadores que emitiram 115 pareceres. Os autores estão em 21 instituições distintas, no Brasil e no exterior.

Adicionalmente, o I ETC integra também o I Desafio de Algoritmos, Combinatória e Otimização (I DACO), com o fito de despertar talentos e entusiasmo entre estudantes de mestrado e graduação em computação para essas áreas.

Gostaríamos de expressar nossos agradecimentos ao Prof. Jayme Szwarcfiter, que generosamente nos orientou ativamente na organização do I ETC, com a maior presteza, sempre que solicitado, aceitando também fazer a abertura do evento, aos palestrantes Profa. Celina de Figueiredo, Profa. Luciana Buriol, Prof. Luis Lamb e Prof. Luiz Satoru Ochi, que aceitaram prontamente o nosso convite de abrilhantar o evento, ao membros do Comitê Científico, Comitê do DACO e pareceristas externos pelo seu comprometimento e excelente trabalho, e à organização do CSBC pelo apoio imensurável à realização deste evento, que esperamos ser um marco e o primeiro de uma sequência feliz.

Cláudia Linhares Sales (DC, UFC)
Rosiane de Freitas (IComp, UFAM)

Comitê de Organização

Coordenação Geral

Cláudia Linhares Sales (DC, UFC)
Rosiane de Freitas (IComp, UFAM)

Coordenação Local

Alfio Martini (PUC-RS)

Comitê Científico

Ana Teresa Martins (DC, UFC)
Cláudia Linhares Sales (DC, UFC)
Claudson Bornstein (DCC - IM, UFRJ)
Edson Cáceres (FACOM, UFMS)
Fábio Protti (IC, UFF)
Flavio Keidi Miyazawa (IC, UNICAMP)
Jayme Szwarefiter (COPPE, IM e NCE, UFRJ)
Luciana Buriol (INF, UFRGS)
Luiz Satoru Ochi (IC, UFF)
Mario Benevides (IM, UFRJ)
Rosiane de Freitas (IComp, UFAM)
Vinicius Gusmão Pereira de Sá (DCC - IM, UFRJ)
Vinicius Santos (DCC, UFMG)

Comitê do DACO

Claudson Bornstein (DCC - IM, UFRJ)
Rosiane de Freitas (IComp, UFAM)
Vinicius Gusmão Pereira de Sá (DCC - IM, UFRJ)
Vinicius Santos (DCC, UFMG)

Palestras

Lógica em Computação, IA e Cognição

Luis C. Lamb (Universidade Federal do Rio Grande do Sul)

Desde o início do século XX, estudos em teoria da recursão e compatibilidade levaram a resultados científicos que tiveram destacada relevância em Ciência da Computação. O impacto da lógica foi tão significativo que a mesma foi considerada por pesquisadores eminentes como o "cálculo da Ciência da Computação". Posteriormente, o desenvolvimento de áreas como a Inteligência Artificial contribuiu para o desenvolvimento da lógica computacional. Nesta palestra, apresentaremos uma breve análise da contribuição da lógica à Ciência da Computação, com destaque aos resultados que contribuíram diretamente no desenvolvimento dessa ciência e que tiveram impacto e aplicação em diversas tecnologias.

Intatibilidade e Otimização: uma Homenagem ao David Johnson

Celina de Figueiredo (Universidade Federal do Rio de Janeiro)

Luciana Buriol (Universidade Federal do Rio Grande do Sul)

Faremos uma homenagem a David Johnson (1945–2016), destacando as suas contribuições para a análise teórica e experimental de algoritmos. Ao longo da sua brilhante carreira de 40 anos no Bell Las Research, foi chefe do departamento de Fundamentos Matemáticos de Computação e do departamento de Algoritmos e Otimização. David Johnson liderou na ACM a área de Algoritmos e Teoria da Computação, através da criação da conferência ACM-SIAM SODA e do grupo de interesse ACM SIGACT. O seu livro "Computers and Intractability: A Guide to the Theory of NP-Completeness" e a sua série "An Ongoing Guide on NP-completeness" constituem os fundamentos para o desenvolvimento da teoria que identifica os problemas difíceis. Ele criou e liderou nos últimos 25 anos as DIMACS Implementation Challenges para computação experimental buscando o rigor científico na avaliação empírica de algoritmos.

Problemas de Edição de Arestas em Grafos

Luiz Satoru Ochi (Universidade Federal Fluminense)

O Problema de Clusterização por Edição de Arestas (PCEA) consiste em transformar um grafo G em um "cluster graph" (união disjunta de subgrafos completos), através do menor número de operações de edição de arestas. Cada operação de aresta consiste em uma adição de uma nova aresta ou em uma remoção de uma aresta existente. O Problema de Biclusterização por Edição de Arestas (PBEA) é um problema de biclusterização similar ao PCEA, dado como entrada um grafo bipartido G , seu objetivo é transformar G em uma união disjunta de grafos bipartidos completos (bicliques) através do menor número de operações de edição de aresta. Problemas de Edição de Arestas em Grafos apresentam diversas aplicações, dentre elas: processamento de imagem, projeto de redes multicast, análise de dados biológicos e formação de células

de manufatura. Nesta palestra será apresentado o desenvolvimento de novos resultados teóricos, meta-heurísticas, métodos exatos e algoritmos híbridos para os Problemas de Edição de Aresta em Grafos. Este é um trabalho em conjunto com Anand Subramanian (UFPB), Fábio Protti (UFF), Gilberto Sousa (UFF e UFPB), Ivan Martins (UFF), Lucas Bastos (FINEP), Lucídio Cabral (UFPB), Luidi Simonetti (UFRJ), Rian Pinheiro (UFRPE) e Teobaldo Bulhões Júnior (UFF).

Trabalhos aceitos

Minimal Obstructions of (2,1) - cographs with External Restrictions

Raquel Bravo (Universidade Federal Fluminense - Brasil)

Fabio Protti (Universidade Federal Fluminense - Brasil)

Loana Nogueira (Universidade Federal Fluminense - Brasil)

Cláutenis Viana (Instituto Federal do Piauí - Brasil)

On Linial's Conjecture for Split Digraphs

Maycon Sambinelli (Universidade Estadual de Campinas - Brasil)

Orlando Lee (Universidade Estadual de Campinas - Brasil)

Cândida Silva (Universidade de São Carlos at Sorocaba - Brasil)

On Colored Edge Cuts in Graphs

Luerbio Faria (Universidade do Estado do Rio de Janeiro - Brasil)

Sulamita Klein (Universidade Federal do Rio de Janeiro - Brasil)

Ignasi Sau (CNRS, LIRMM, Montpellier - France)

Uéverton Souza (Universidade Federal Fluminense - Brasil)

Rubens Sucupira (Universidade do Estado do Rio de Janeiro - Brasil)

Frustração de Arestas em (3,6) - Fullerenes

Diego Nicodemos (Universidade Federal do Rio de Janeiro - Brasil)

Sulamita Klein (Universidade Federal do Rio de Janeiro - Brasil)

Luerbio Faria (Universidade do Estado do Rio de Janeiro - Brasil)

Descriptive Complexity of Probabilistic Complexity Classes through Second Order Generalized Quantifiers

Thiago Rocha (Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Brasil),

Ana Teresa Martins (Universidade Federal do Ceará - Brasil)

A Gramática de Grafos como uma Alternativa para o Desenvolvimento do Pensamento Computacional na Educação Básica

Braz Junior (Universidade Federal de Pelotas - Brasil)

Simone Cavalheiro (Universidade Federal de Pelotas - Brasil)

Luciana Foss (Universidade Federal de Pelotas - Brasil)

Sobre Coloração Total dos Grafos r-partidos Completos

Raphael Martins (Universidade do Estado do Rio de Janeiro - Brasil)

Diana Sasaki (Université Paris Dauphine – France)

Amostragem Gaussiana na Criptografia Baseada em Reticulados

Jheyne N. Ortiz (Universidade Estadual de Campinas - Brasil)

Ricardo Dahab (Universidade Estadual de Campinas - Brasil)

Diego Aranha (Universidade Estadual de Campinas – Brasil)

Achieving CCA1-security in Homomorphic Encryption

Eduardo Moraes (Universidade Estadual de Campinas - Brasil)
 Diego Aranha (Universidade Estadual de Campinas - Brasil)
 Ricardo Dahab (Universidade Estadual de Campinas - Brasil)

Análise de um Algoritmo Aproximativo para o Problema de Escalonamento de Tarefas com Restrições de Precedência em Máquinas Paralelas Idênticas

Elton Lever (Universidade Federal do Amazonas – Brasil)
 Omar Vilca (Universidade Federal do Amazonas - Brasil)
 Rosiane de Freitas (Universidade Federal do Amazonas – Brasil)

Some Families of 0-rotatable Graceful Caterpillars

Atílio Luiz (Universidade Estadual de Campinas - Brasil)
 C. N. Campos (Universidade Estadual de Campinas - Brasil)
 Bruce Richter (University of Waterloo – Canada)

Computação sobre Dados Cifrados em GPGPUs

Pedro Alves (Universidade Estadual de Campinas - Brasil)
 Diego Aranha (Universidade Estadual de Campinas - Brasil)

Algoritmos Branch-and-Price para o Problema de Empacotamento em Recipientes com Restrições de Classe

Yulle Glebbyo Felipe Borges (Universidade Estadual de Campinas - Brasil)
 Flávio Miyazawa (Universidade Estadual de Campinas - Brasil)
 Rafael Schouery (Universidade Estadual de Campinas - Brasil)
 Eduardo Xavier (Universidade Estadual de Campinas – Brasil)

Uma Aproximação para o Problema de Alocação de Terminais

Lehilton Pedrosa (Universidade Estadual de Campinas - Brasil)
 Vinicius dos Santos (Universidade Federal de Minas Gerais - Brasil)
 Rafael Schouery (Universidade Estadual de Campinas – Brasil)

UKP5: Solving the Unbounded Knapsack Problem

Henrique Becker (Universidade Federal do Rio Grande do Sul - Brasil)
 Luciana Buriol (Universidade Federal do Rio Grande do Sul - Brasil)

Uma Abordagem baseada nas Preferências dos Docentes para o Problema de Programação de Horários em Universidades

Osmar Leandro Dantas da Silva (Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Brasil)
 Diego Rocha Lima (Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Brasil)

On a Leasing Variant of the Online Connected Facility Location Problem

Murilo de Lima (Universidade Estadual de Campinas - Brasil)
 Mario San Felice (Universidade de São Paulo - Brasil)
 Orlando Lee (Universidade Estadual de Campinas – Brasil)

Sobre a Minimização de Transdutores Sequenciais

Rodrigo de Souza (Universidade Federal Rural de Pernambuco – Brasil)

Quantificando o Vazamento de Informações sobre Estratégias

Mário Alvim (Universidade Federal de Minas Gerais - Brasil)

Piotr Mardziel (University of Maryland, College Park - USA)

Michael Hicks (University of Maryland, College Park – USA)

Compartilhamento de Custos de Empacotamento

Flávio Miyazawa (Universidade Estadual de Campinas - Brasil)

Rafael Schouery (Universidade Estadual de Campinas – Brasil)

Algoritmos Aleatorizados com Oráculo para MCSP: Aplicações para o Problema do Resíduo Quadrático e do Logaritmo Discreto

Nicollas Sdroievski (Universidade Tecnológica Federal do Paraná - Brasil)

Murilo da Silva (Universidade Tecnológica Federal do Paraná – Brasil)

Um Leilão à Prova de Estratégia para o Compartilhamento de Viagens Dinâmico com Múltiplos Passageiros

Leonardo Schwarzstein (Universidade Estadual de Campinas - Brasil)

Rafael Schouery (Universidade Estadual de Campinas - Brasil)

Flávio Miyazawa (Universidade Estadual de Campinas - Brasil)

Caracterizações de Convexidades Geométricas de Grafos

Rafael Teixeira de Araujo (Universidade Federal do Ceará - Brasil)

Rudini Sampaio (Universidade Federal de Ceará – Brasil)

Results on Circular-Arc Bigraphs

Fabricio Kolberg (Universidade Federal do Paraná - Brasil)

Marina Groshaus (Universidad de Buenos Aires - Argentina)

André Guedes (Universidade Federal do Paraná - Brasil)

Renato Carmo (Universidade Federal do Paraná – Brasil)

O Problema da Atribuição Dupla de Custo Mínimo

Vinicius dos Santos (Universidade Federal de Minas Gerais - Brasil)

Sebastian Alberto Urrutia (Universidade Federal de Minas Gerais – Brasil)

The Geodetic Carathéodory Number

Eduardo Silva Lira (Universidade Federal de Goiás - Brasil)

Diane Castonguay (Universidade Federal de Goiás - Brasil)

Erika Coelho (Universidade Federal de Goiás - Brasil)

Hebert da Silva (Universidade Federal de Goiás – Brasil)

Padrão de Desenho para a Família de Grafos Gêmeos

Cassio Ferreira Merlo (Universidade Federal do Espírito Santo - Brasil)

Mauro Pinheiro (Universidade Federal do Espírito Santo - Brasil)

Marcia Paiva (Universidade Federal do Espírito Santo – Brasil)

An Efficient Algorithm for the Closest String Problem

Omar Latorre (Universidade Federal do Amazonas - Brasil)

Rosiane de Freitas (Universidade Federal do Amazonas - Brasil)

Pôsteres**Stochastic Scenario Generation: an Empirical Approach**

Alan Delgado de Oliveira (Universidade Federal do Rio Grande do Sul - Brasil)

Tiago Filomena (Universidade Federal do Rio Grande do Sul – Brasil)

Análise e Comparação dos Algoritmos de Dijkstra e A-Estrela na Descoberta de Caminhos Mínimos em Mapas de Grade

Marcel Rios (Universidade Federal do Amazonas - Brasil)

Francisco S.S Neto (Universidade Federal do Amazonas - Brasil)

José Francisco Netto (Universidade Federal do Amazonas – Brasil)

Um modelo RCPSP para Gestão Ágil Scrum

Osmar Leandro Dantas da Silva (Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Brasil)

Emanuel Dantas Filho (Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Brasil)

Diego Rocha Lima (Instituto Federal de Educação, Ciência e Tecnologia do Ceará - Brasil)

KM-Finder: Uma Ferramenta para Detecção de Motivos

Lucas Vilhagra (Universidade Federal de Mato Grosso do Sul - Brasil)

Luciana Montera (Federal University of Mato Grosso do Sul – Brasil)

Taina Alencar (FIOCRUZ- Brasil)

Estruturas de Dados Probabilísticas para Representação de Conjuntos

Juan Lopes (Universidade do Estado do Rio de Janeiro - Brasil)

Paulo Pinto (Universidade do Estado do Rio de Janeiro - Brasil)

Fabiano Oliveira (Universidade do Estado do Rio de Janeiro - Brasil)

Análise Empírica do Algoritmo Shellsort

Paulo Pinto (Universidade do Estado do Rio de Janeiro - Brasil)

Fabiano Oliveira (Universidade do Estado do Rio de Janeiro - Brasil)

Raquel Souza (Universidade do Estado do Rio de Janeiro - Brasil)

Obstruções Minimais de Cografos-(2,1) com restrição externa

Raquel S. F. Bravo¹, Loana T. Nogueira¹, Fábio Protti¹, Clautenis Vianna²

¹Instituto de Computação – Universidade Federal Fluminense (UFF)

²Instituto Federal do Piauí (IFPI) – Teresina, PI – Brazil

{raquel, fabio, loana}@ic.uff.br, clau.nivica@gmail.com

Abstract. We characterize cographs whose set of vertices can be partitioned into k independent sets and ℓ clique, named (k, ℓ) -graphs and consider all possible restrictions between the parts (external restrictions), that is, for any two distinct parts i and j can be completely adjacent, completely non-adjacent, or with no restrictions. We determine all the possible minimal obstruction cographs-(2, 1) with external restrictions.

Palavras Chave: Cografos, Partição em grafos, M -partição, M -obstrução, Cografos- (k, ℓ) .

Resumo. Caracterizamos os cografos cujo conjunto de vértices pode ser particionado em k conjuntos independentes e ℓ cliques, denominados grafos- (k, ℓ) e consideramos todas as possíveis restrições entre as partes (restrições externas), isto é, quaisquer duas partes distintas i e j podem ser completamente adjacentes, completamente não adjacentes ou sem restrições. Determinamos todas as possíveis obstruções minimais dos cografos-(2, 1) com restrições externas.

Keywords: Cographs, Graph Partition, M -partition, M -obstruction, Cographs- (k, ℓ) .

1. Introdução

Os problemas de partição de grafos consistem em particionar o conjunto V dos vértices de um grafo em subconjuntos V_1, V_2, \dots, V_k onde $V_1 \cup V_2 \cup \dots \cup V_k = V$ e $V_i \cap V_j = \emptyset$, $i \neq j$, $1 \leq i \leq k$ e $1 \leq j \leq k$, exigindo-se, porém, algumas propriedades sobre estes subconjuntos de vértices. Estas propriedades podem ser *internas*, como por exemplo exigir que os vértices de cada subconjunto V_i sejam completamente adjacentes (isto é, V_i é uma clique) ou completamente não-adjacentes (isto é, V_i é um conjunto independente), ou *externas*, onde as exigências são feitas sobre os pares (V_i, V_j) , isto é, todos os vértices de V_i e os vértices de V_j podem ser completamente adjacentes entre si ou completamente não-adjacentes entre si. Brandstädt propôs uma generalização dos grafos split, que definiu uma nova classe de grafos, a classe dos grafos- (k, ℓ) , como sendo aquela formada pelos grafos cujo conjunto de vértices pode ser particionado em k conjuntos independentes e ℓ cliques. Reconhecer grafos- (k, ℓ) para $k \geq 3$ ou $\ell \geq 3$ é NP -Completo [Brandstädt (1996)]. Como o reconhecimento desta classe é NP -completo, para $k \geq 3$ ou $\ell \geq 3$, alguns autores estudaram o problema quando restrito à subclasses de grafos. Hell, Klein, Nogueira e Protti (2004) apresentaram uma caracterização e um algoritmo de reconhecimento com complexidade $O(n(n+m))$ para os grafos cordais- (k, ℓ) . Outra forma conveniente de representar o particionamento em grafos é através da M -partição. O problema da M -partição foi definido por Feder *et. al* (1999), que generaliza o problema de grafos- (k, ℓ) , consistindo em particionar o conjunto de vértices de um dado grafo em m partes A_1, A_2, \dots, A_m , com certas restrições para cada A_i , $i = 1, \dots, m$. Os vértices

em A_i devem ser completamente adjacentes ou completamente não-adjacentes (restrições internas) ou para cada par A_i, A_j , os vértices em A_i e A_j são completamente adjacentes ou não-adjacentes entre si (restrições externas).

Com a finalidade de expressar essas restrições de forma mais detalhada, Feder *et. al.* (1999) definiram uma matriz simétrica M de ordem $m \times m$ em que cada elemento $M_{i,i}$ representa a restrição interna no conjunto A_i , e cada elemento fora da diagonal principal $M_{i,j}$, $i \neq j$, representa uma restrição externa entre os subconjuntos A_i e A_j . Mais especificamente, seja uma matriz simétrica, onde as entradas $M_{i,j}$ podem ser igual a 0, 1 ou *. Assim, temos que A_i é um conjunto independente, se $M_{i,i} = 0$, uma clique, se $M_{i,i} = 1$, ou sem nenhuma restrição interna, se $M_{i,i} = *$. Por outro lado, as entradas fora da diagonal principal, A_i e A_j são completamente não-adjacentes, se $M_{i,j} = 0$, completamente adjacentes, se $M_{i,j} = 1$, ou sem restrição externa, se $M_{i,j} = *$. Uma M -obstrução H é minimal quando para qualquer vértice v de H , $H - v$ é M -particionável.

Neste trabalho, consideramos a classe dos cografos, classe esta livre de P_4 induzido. Essa classe está contida na classe dos grafos perfeitos, o que a torna uma classe interessante para estudo. Mais especificamente, caracterizamos os cografos M -particionáveis em termos de obstruções e determinamos todas as M -obstruções minimais dos cografos quando M é uma matriz de ordem igual a três.

2. Preliminares

Dado um grafo simples $G = (V, E)$, denotamos por \bar{G} o complemento de G . Para $V' \subseteq V$, denotamos por $G[V']$ o subgrafo de G induzido por V' . Uma *clique* (*conjunto independente*) é um subconjunto de vértices que induz um subgrafo completo (sem arestas), não necessariamente maximal e denotada por K_p (I_p) uma clique (conjunto independente) de p vértices. G é um *grafo*-(k, ℓ) se V pode ser particionado em k conjuntos independentes e ℓ cliques. Dados dois grafos $G_1 = (V_1, E_1)$ e $G_2 = (V_2, E_2)$, o grafo $G_1 \cup G_2$ (chamado de *união* de G_1 e G_2) é um grafo com o conjunto de vértices $V_1 \cup V_2$ e conjunto de arestas $E_1 \cup E_2$, e o grafo $G_1 + G_2$ (chamado de *junção* de G_1 e G_2) é o grafo com o conjunto de vértices $V_1 \cup V_2$ e o conjunto de arestas $E_1 \cup E_2 \cup \{(x, y) \mid x \in V_1, y \in V_2\}$. Um *caminho* num grafo G é uma sequência de vértices $P = v_1, v_2, \dots, v_k$, onde os v_i 's são vértices (dois a dois distintos), e $(v_i, v_{i+1}) \in E(G)$, $1 \leq i \leq k - 1$. Uma *corda* em P é uma aresta que liga dois vértices não-consecutivos de P . Um *caminho induzido* é um caminho sem cordas, e denotado por P_k o caminho induzido com k vértices. Dizemos que um grafo é *livre de P_k* quando não contém P_k como subgrafo induzido. Um *cografo* G é um grafo livre de P_4 , isto é, G não contém P_4 como subgrafo induzido.

3. Resultado Principal

Nesta seção consideramos o problema da M -partição para matrizes quadradas M de ordem $m = 3$ e apresentamos uma caracterização dos cografos M -particionáveis em termos de M -obstruções, quando $k = 2$ e $\ell = 1$. Usaremos a notação $(k, \ell) \{a_1^2, a_1^3, \dots, a_1^{k+\ell}, \dots, a_i^{i+1}, \dots, a_i^{k+\ell}\}$, $i = 1, \dots, k + \ell - 1$ para representar uma partição em k conjuntos independentes X_1, X_2, \dots, X_k e ℓ cliques $X_{k+1}, \dots, X_{k+\ell}$, onde cada a_i^j representa a restrição externa entre os conjuntos X_i e X_j . Observe que um grafo- $(k, \ell) \{a_1^2, a_1^3, \dots, a_1^{k+\ell}, \dots, a_i^{i+1}, \dots, a_i^{k+\ell}\}$ representa um grafo que admite uma M -partição onde M é uma matriz com k elementos nulos na diagonal principal, ℓ 1's na diagonal principal, com qualquer elemento fora da diagonal principal podendo pertencer ao conjunto $\{0, 1, *\}$.

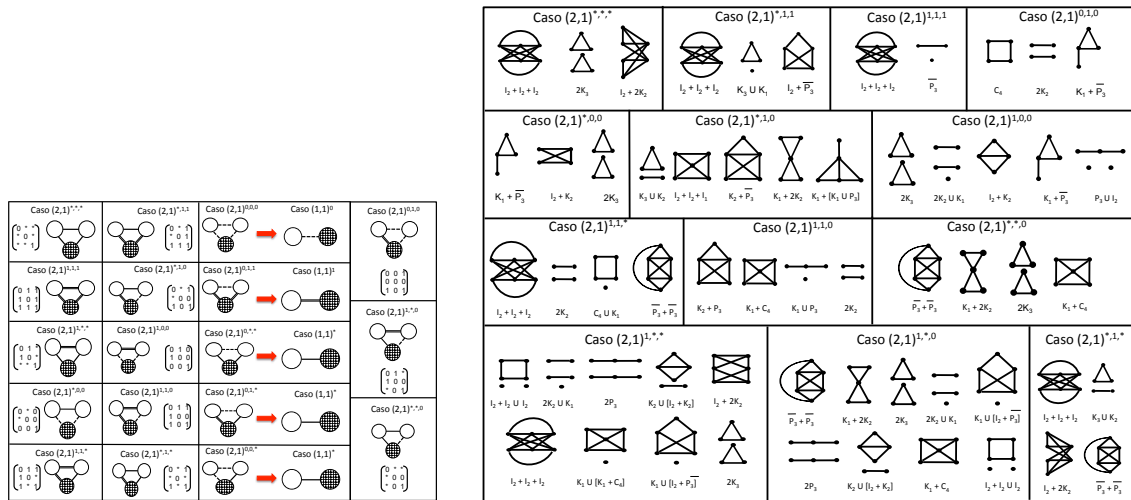


Figure 1. Obstruções minimais de todos os Casos (2, 1) que não são redutíveis

3.1. Casos (2,1)

Na Figura 1 são listadas todas as matrizes de ordem 3×3 referentes aos Casos (2, 1) que serão considerados nesta subseção. Observe que o caso de se verificar se um grafo G é $(2, 1)^{0,0,0}$ equivale a verificar se o grafo é $(1, 1)^0$. Da mesma forma, o caso de se verificar se um grafo é $(2, 1)^{0,1,1}$ equivale a se verificar se o grafo é $(1, 1)^1$. Note também, que os casos de se verificar se um grafo é $(2, 1)^{0,*,*}$ ou $(2, 1)^{0,1,*}$ ou $(2, 1)^{0,0,*}$ são equivalentes a verificar se o grafo é $(1, 1)^*$. As obstruções minimais de cada caso (2,1) se encontra explícito na Figura 1.

Teorema 1 *Seja G um cografo. G é um grafo $(2, 1)^{*,1,0}$ se e somente se G não contém nenhum dos grafos da Figura 2 como subgrafo induzido.*

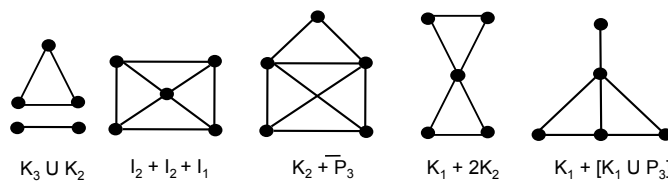


Figure 2. Obstruções minimais do caso $(2, 1)^{*,1,0}$

Prova: (\Rightarrow) É fácil ver que se G contém algum dos subgrafos acima, como subgrafo induzido, então G não é um cografo $(2, 1)^{*,1,0}$. (\Leftarrow) Seja G um cografo minimal que não é $(2, 1)^{*,1,0}$, ou seja, para todo vértice $v \in V(G)$, $G - v$ é um grafo $(2, 1)^{*,1,0}$. Suponha, por contradição que G não contém nenhum dos subgrafos da Figura 2 como subgrafo induzido. Assim, consideraremos dois casos: (i) G é desconexo. $G = G_1 \cup G_2 \cup \dots \cup G_k$, $k \geq 2$, onde cada G_i é conexo. Se G_i for trivial, para algum $i = 1, 2, \dots, k$, então G é $(2, 1)^{*,1,0}$, pois por minimalidade, $G - G_i$ é $(2, 1)^{*,1,0}$. Contradição. Dessa forma, $|G_i| \geq 2$, para todo $i = 1, \dots, k$. Temos que G não é $(2, 0)^*$, caso contrário G seria $(2, 1)^{*,1,0}$, uma contradição. Logo, G contém K_3 , e desta forma, G_i contém K_3 . Como cada G_i , $|G_i| \geq 2$ e $k \geq 2$, então G contém $K_3 \cup K_2$. Absurdo. (ii) \overline{G} é desconexo.

Pela propriedade da decomposição modular, $G = G_1 + G_2 + \dots + G_k$, com $k \geq 2$ e cada G_i é trivial ou desconexo. Se G_i for trivial, para todo $i = 1, 2, \dots, k$, então G é $(2, 1)^{*,1,0}$, uma contradição. Desta forma, existe pelo menos um G_i que possui pelo menos um par de vértices não adjacentes, $i = 1, 2, \dots, k$. Assim, analisaremos os dois subcasos ($k \geq 3, k = 2$): (ii.1) $k \geq 3$: se existem G_i e G_j , ambos contendo I_2 , então G contém $I_2 + I_2 + I_1$. Absurdo. Se apenas G_i contém I_2 , isto quer dizer que $G - G_i$ não contém I_2 e por sua vez é uma clique. Logo, se G_i for $(1, 1)^1$, então G é $(2, 1)^{*,1,0}$, uma contradição. Temos que G_i contém C_4 ou \overline{P}_3 , e desta forma, se G_i contém C_4 , então G contém $C_4 + K_1$, absurdo. E se G_i contém \overline{P}_3 , então G contém $\overline{P}_3 + K_2$, absurdo. (ii.2) $k = 2$: consideraremos mais dois subcasos (ii.2.1) G_1 contém I_2 e G_2 é trivial. Se G_1 for $(1, 1)^1$, então G é $(2, 1)^{*,1,0}$. Absurdo! Logo, G_1 contém C_4 ou \overline{P}_3 . Considerando que G_1 contenha C_4 , então G contém $I_2 + I_2 + I_1$. E, se G_1 contém \overline{P}_3 , então, analisaremos as seguintes situações: se G_1 é $(1, 1)^0$, então G é $(2, 1)^{*,1,0}$. Contradição. Logo, G_1 contém P_3 ou $2K_2$. Se G_1 contém P_3 , então como G_1 é desconexo temos que G_1 contém $P_3 \cup K_1$, logo G contém $K_1 + (K_1 \cup P_3)$. (ii.2.2) Ambos G_1 e G_2 contém I_2 . Se G é $(2, 0)^*$ então G é $(2, 1)^{*,1,0}$. Contradição. Logo, G contém K_3 . Já que G contém K_3 , então pelo menos um G_i , digamos G_1 , contém uma aresta. Se G_2 também contiver K_2 , então G contém $\overline{P}_3 + K_2$, já que G_1 contém \overline{P}_3 , pois é desconexo, absurdo. No entanto, se apenas G_1 contém K_2 e G_2 é tal que $V(G_2)$ é um conjunto independente, então: se G_1 for $(1, 1)^0$ então G é $(2, 1)^{*,1,0}$. Contradição. Se G_1 não for $(1, 1)^0$ então G_1 contém $2K_2$ ou P_3 . Dessa forma, considerando que G_1 contenha $2K_2$, então G contém $K_1 + 2K_2$, absurdo! E, se G_1 contém P_3 , então G contém $K_1 + (P_3 \cup K_1)$, já que G_1 é desconexo, absurdo. Concluimos que o cografo G é $(2, 1)^{*,1,0}$, se e somente se, não contém nenhum dos grafos $K_3 \cup K_2, I_2 + I_2 + I_1, K_2 + \overline{P}_3, K_1 + 2K_2$ e $K_1 + [K_1 \cup P_3]$ como subgrafo induzido.

Por falta de espaço, omitiremos as provas dos outros casos.

4. Conclusão

Neste trabalho, usamos fortemente a estrutura dos cografos para caracterizarmos por obstruções minimais todos os cografos $-(2, 1)$ com e sem restrição externa.

References

- [1] **Brandsstädt, A.** Partitions of graphs into one or two independent sets and cliques. *Discrete Mathematics* 152 (1996) 47 – 54.
- [2] **Brandsstädt, A.** The complexity of some problems related to graph 3-colorability. *Discrete Applied Mathematics* 89 (1998) 59 – 73.
- [3] **Bravo, R. S. F., Klein, S., Nogueira, L. T., and Protti, Fábio** Characterization and recognition of P_4 -sparse graphs partitionable into independent sets and cliques. *Discrete Applied Mathematics* 159 (2011) 165 – 173.
- [4] **Corneil, D. G., Lerchs, H., and Burlingham, L. S.** Complement reducible graphs. *Discrete Applied Mathematics* 3 (1981) 163 – 174.
- [5] **Feder, T., Hell, P., Klein, S., and Motwani, R.** List partitions. *SIAM Journal on Discrete Mathematics* 16 (2003) 449 – 478.
- [6] **Feder, T., and Hell, P.** Matrix partitions of perfect graphs. Special Issue of *Discrete Mathematics*, 306 (2006) 2450 – 2460
- [7] **Feder, T., Hell, P., and Hochstättler, W.** Generalized Colouring (Matrix Partitions) of Cographs. *Trends in Mathematics* 2006 149 – 167
- [8] **Hell, P., Klein, S., Nogueira, L. T., and Protti, F.** Partitioning chordal graphs into independent sets and cliques. *Discrete Applied Mathematics* 141 (2004) 185 – 194.

On Linial's Conjecture for Split Digraphs

Maycon Sambinelli^{1*}, Cândida Nunes da Silva², Orlando Lee^{1†}

¹Institute of Computing – University of Campinas (Unicamp)
13083-852 – Campinas – SP – Brazil

²Department of Computing – Federal University of São Carlos (Ufscar)
18052-780 – Sorocaba – SP – Brazil

{msambinelli, lee}@ic.unicamp.br, candida@ufscar.br

Abstract. *In this paper we show that Linial's Conjecture holds for two classes of split digraphs, namely the spider digraphs and the k -loose digraphs.*

1. Introduction

The digraphs considered in this text do not contain loops or parallel arcs and by path we mean directed path. Let D be a digraph. We denote by $V(P)$ the set of vertices of a path P . The **size** of a path P , denoted by $|P|$, is $|V(P)|$ ¹. We denote by $\lambda(D)$ the size of the longest path in D and by $\alpha(D)$ the size of a maximum stable set. A *path partition* \mathcal{P} of D is a set of vertex-disjoint paths of D that cover $V(D)$. We say that \mathcal{P} is an **optimal** path partition if there is no path partition \mathcal{P}' of D such that $|\mathcal{P}'| < |\mathcal{P}|$. We denote by $\pi(D)$ the size of an optimal path partition of a digraph D .

Dilworth [Dilworth 1950] showed that for every transitive acyclic digraph D we have $\pi(D) = \alpha(D)$. Note that this equality is not valid for any digraph; for example, if D is a directed cycle with 5 vertices, then $\pi(D) = 1$ and $\alpha(D) = 2$. However, Gallai and Milgram [Gallai and Milgram 1960] have shown that $\pi(D) \leq \alpha(D)$ for every digraph D .

Greene and Kleitman [Greene and Kleitman 1976] proved a generalization of Dilworth's Theorem described next. Let k be a positive integer. The **k -norm** of a path partition \mathcal{P} , denoted by $|\mathcal{P}|_k$, is defined as $|\mathcal{P}|_k = \sum_{P \in \mathcal{P}} \min\{|P|, k\}$. We say that \mathcal{P} is a **k -optimal path partition** if there is no path partition \mathcal{P}' such that $|\mathcal{P}'|_k < |\mathcal{P}|_k$. We denote by $\pi_k(D)$ the k -norm of a k -optimal path partition of D . A **k -partial coloring** \mathcal{C}^k is a set of k disjoint stable sets called **color classes** (empty color classes are allowed). The **weight** of a k -partial coloring \mathcal{C}^k , denoted by $\|\mathcal{C}^k\|$, is defined as $\|\mathcal{C}^k\| = \sum_{C \in \mathcal{C}^k} |C|$. We say that \mathcal{C}^k is an **optimal k -partial coloring** if there is no k -partial coloring \mathcal{B}^k such that $\|\mathcal{B}^k\| > \|\mathcal{C}^k\|$. We denote by $\alpha_k(D)$ the weight of an optimal k -partial coloring of D . Given these definitions, what Greene and Kleitman [Greene and Kleitman 1976] showed was that for every transitive acyclic digraph D , we have $\pi_k(D) = \alpha_k(D)$. Note that $\pi(D) = \pi_1(D)$ and $\alpha(D) = \alpha_1(D)$. Thus, Dilworth's Theorem is a particular case of Greene-Kleitman's Theorem in which $k = 1$.

As Gallai-Milgram's Theorem extends Dilworth's Theorem, it is a natural question whether Greene-Kleitman's Theorem can be extended to digraphs in general. More precisely, is it true that for every digraph D we have that $\pi_k(D) \leq \alpha_k(D)$? Linial [Linial 1981] conjectured that the answer for this question is positive.

*Supported by National Counsel of Technological and Scientific Development - CNPq (grant 141216/2016-6).

†Supported by National Counsel of Technological and Scientific Development - CNPq (grants 311373/2015-1 and 477692/2012-5).

¹Usually $|P|$ denotes the length of a path (number of arcs), but here it denotes the number of vertices.

Linial's Conjecture [Linial 1981]. *Let D be a digraph and k be a positive integer. Then, $\pi_k(D) \leq \alpha_k(D)$.*

Linial's Conjecture remains open, but we know it holds for acyclic digraphs [Saks 1979], bipartite digraphs [Berge 1982], digraphs which contain a Hamiltonian path [Berge 1982], $k = 1$ [Linial 1978], $k = 2$ [Berger and Hartman 2008] and $k \geq \lambda(D) - 3$ [Herskovics 2013]. In this paper we give partial results on Linial's Conjecture for split digraphs.

2. Split digraphs

Let D be a digraph and let $X \subseteq V(D)$. We denote by $D[X]$ the subdigraph of D induced by X . A digraph D is a **split digraph** if there is a partition of $V(D)$ into two sets X and Y , such that $D[X]$ is a tournament and $D[Y]$ is a stable set. We shall use the notation $D[X, Y]$ to indicate that D is a split digraph with such partition $\{X, Y\}$.

In this section we shall prove an approximation to Linial's Conjecture for split digraphs, i. e., that $\pi_k(D) \leq \alpha_k(D) + 1$ for every split digraph D , as stated in Theorem 1. For that, we need Rédei's Theorem and Lemmas 1 and 2 below.

Rédei's Theorem [Rédei 1934]. *Every tournament contains a Hamiltonian path.*

Lemma 1. *Let $D[X, Y]$ be a split digraph. Then, $\pi_k(D) \leq |Y| + \min\{|X|, k\}$.*

Proof. By Rédei's Theorem, the tournament $D[X]$ contains a path P such that $V(P) = X$. Let $\mathcal{P} = \{P\} \cup \{(y) : y \in Y\}$. Clearly, \mathcal{P} is a path partition of D for which $|\mathcal{P}|_k = \min\{|X|, k\} + |Y|$. Therefore, $\pi_k(D) \leq |\mathcal{P}|_k = \min\{|X|, k\} + |Y|$. ■

Lemma 2. *Let $D[X, Y]$ be a split digraph. Then, $\alpha_k(D) \geq |Y| + \min\{|X|, k - 1\}$. Moreover, when $|X| < k$, we have that $\alpha_k(D) = |V(D)|$.*

Proof. First, suppose that $|X| \leq k - 1$. Let $\mathcal{C}^k = \{Y\} \cup \{\{x\} : x \in X\}$. Note that \mathcal{C}^k is a k -partial coloring of D with $\|\mathcal{C}^k\| = |V(D)|$. Therefore, $\alpha_k(D) = \|\mathcal{C}^k\| = |Y| + |X| = |Y| + \min\{|X|, k - 1\}$ and the result follows. We may assume that $|X| \geq k$. Let $\mathcal{C}^k = \{Y\} \cup \{\{x\} : x \in S\}$, where $S \subseteq X$ such that $|S| = k - 1$. Clearly, \mathcal{C}^k is a k -partial coloring for which $\|\mathcal{C}^k\| = |Y| + k - 1$. Therefore, $\alpha_k(D) \geq \|\mathcal{C}^k\| = |Y| + k - 1 = |Y| + \min\{|X|, k - 1\}$. ■

Theorem 1. *Let $D[X, Y]$ be a split digraph. Then, $\pi_k(D) \leq \alpha_k(D) + 1$.*

Proof. The result follows immediately from Lemmas 1 and 2. ■

In Section 2.1 we introduce k -loose digraphs and show that Linial's Conjecture holds for them and in Section 2.2 we show that it holds for spider digraphs [Hoàng 1985].

2.1. k -loose digraphs

A split digraph $D[X, Y]$ is **k -loose** if either $|X| < k$ or there is a $S \subseteq X$ such that $|S| = k$ and no vertex $y \in Y$ is adjacent to every vertex in S . A split digraph $D[X, Y]$ that is not k -loose is called **k -tight**. We show in this section that Linial's Conjecture holds for every k -loose digraph (Theorem 2) and for split digraphs such that $|X| \leq k$ (Theorem 3). For that, we need Lemmas 3 and 4 below.

Lemma 3. *Let $D[X, Y]$ be a split digraph. Then, D is k -loose if and only if $\alpha_k(D) \geq |Y| + \min\{|X|, k\}$.*

Proof. Consider that D is k -loose. If $|X| < k$, then by Lemma 2, $\alpha_k(D) = |V(D)| = |Y| + |X| \geq |Y| + \min\{|X|, k\}$. We may thus assume that $|X| \geq k$ and there is $S \subseteq X$ such that $|S| = k$ and no vertex $y \in Y$ is adjacent to every vertex in S . Assume $S = \{x_1, x_2, \dots, x_k\}$ and let $\mathcal{C}_0^k = \{C_1, C_2, \dots, C_k\}$ be a k -partial coloring where $C_i = \{x_i\}$ for $i = 1, 2, \dots, k$. For each $y \in Y$ choose some vertex x_i not adjacent to y (which exists by definition) and add y in color class C_i . The k -partial coloring \mathcal{C}^k thus obtained has weight $|Y| + k = |Y| + \min\{|X|, k\}$ as expected.

Conversely, consider that $\alpha_k(D) \geq |Y| + \min\{|X|, k\}$. If $|X| < k$, then D is k -loose by definition. So, we may assume that $|X| \geq k$ and, whence, $\alpha_k(D) \geq |Y| + k$. We conclude that \mathcal{C}^k must have exactly k vertices of X , besides all $|Y|$ vertices from Y . Let $S = \{x : x \in C_i \cap X \text{ for } i = 1, 2, \dots, k\}$. Since all vertices of Y belong to \mathcal{C}^k , then there is no vertex in Y which is adjacent to every vertex of S . Therefore, D is k -loose. ■

Theorem 2. *Let $D[X, Y]$ be a k -loose split digraph. Then, $\pi_k(D) \leq \alpha_k(D)$.*

Proof. By Lemma 3, $\alpha_k(D) \geq |Y| + \min\{|X|, k\}$. On the other hand, by Lemma 1 $\pi_k(D) \leq |Y| + \min\{|X|, k\}$ and the result follows. ■

Lemma 4. *Let $D[X, Y]$ be a split digraph such that $\lambda(D) > |X|$. Then, $\pi_k(D) \leq \alpha_k(D)$.*

Proof. If $\alpha_k(D) = |V(D)|$, then the result follows trivially. Thus, we may assume that $\alpha_k(D) < |V(D)|$. By Lemma 2 we have that $|X| \geq k$ and also that $\alpha_k(D) \geq |Y| + \min\{|X|, k - 1\} = |Y| + k - 1$. Since $\lambda(D) > |X|$, there exists a path P in D such that $|P| = |X| + 1$. Let $\mathcal{P} = \{P\} \cup \{(v) : v \notin V(P)\}$. Clearly, \mathcal{P} is a path partition of D and $|\mathcal{P}|_k = |Y| + k - 1$. Therefore, $\pi_k(D) \leq |\mathcal{P}|_k \leq \alpha_k(D)$. ■

Theorem 3. *Let $D[X, Y]$ be a split digraph such that $|X| \leq k$. Then, $\pi_k(D) \leq \alpha_k(D)$.*

Proof. If D is k -loose, then the result follows by Theorem 2. So, we may assume that D is not k -loose. Hence, $|X| = k$ and there exists a vertex $y \in Y$ which is adjacent to every vertex of X . Therefore, $D[X \cup \{y\}]$ is a tournament and by Rédei's Theorem it has a Hamiltonian path P such that $|P| = |X| + 1$. As P is a path in D as well, we conclude that $\lambda(D) \geq |X| + 1$ and the result follows by Lemma 4. ■

2.2. Spider digraphs

We denote by $\mathcal{N}(v)$ the set of vertices that are adjacent to $v \in V(D)$ (regardless the direction of the arcs). A split digraph $D[X, Y]$ is **spider** [Hoàng 1985] if (i) $|X| = |Y| \geq 2$; and (ii) there exists a bijective function $f : X \rightarrow Y$ such that either $\mathcal{N}(x) = \{f(x)\}$ for all $x \in X$ (in this case, we say that D is a **thin** spider) or $\mathcal{N}(x) = Y - f(x)$ for all $x \in X$ (in this case, we say that D is a **thick** spider). Note that thin spider digraphs are k -loose, but thick spider digraphs are k -tight, as long as $|X| > k$. The following theorem shows that Linial's Conjecture holds for spider digraphs.

Theorem 4. *Let $D[X, Y]$ be a spider digraph. Then, $\pi_k(D) \leq \alpha_k(D)$.*

Proof. Let $\ell = |X| = |Y|$. If $\ell \leq k$, then the result follows by Theorem 3. Thus, we may assume that $|X| > k$. Clearly, $\pi_k(D) \leq |V(D)|$ and we deduce that $\alpha_k < |V(D)|$. If D is a thin spider digraph, whence k -loose, the result follows by Theorem 2. Therefore, we may assume that D is a thick spider graph. Since $D[X]$ is a tournament, by Rédei's Theorem, there exists a path P such that $V(P) = X$. Let $P = (x_1, x_2, \dots, x_\ell)$. Since D is a thick spider digraph, there exists one single vertex $y_i \in Y$ that is not adjacent to x_i , for $i = 1, \dots, \ell$. Note that if $\lambda(D) > |X|$, then the result follows by Lemma 4. So we may assume that $\lambda(D) \leq |X|$.

Let Px_i denote the subpath (x_1, x_2, \dots, x_i) and let x_iP denote the subpath $(x_i, x_{i+1}, \dots, x_\ell)$. We denote by $W \circ Q$ the concatenation of two paths W and Q .

Claim 1: *If $x_i \in X$, $y_j \in Y$ and $i < j$, then $(x_i, y_j) \in A(D)$.*

We prove this claim by induction on i . If $i = 1$, assume by contradiction that $(y_j, x_1) \in A(D)$; then $P' = (y_j, x_1) \circ P$ is a path in D such that $|P'| = |X| + 1$, a contradiction. Hence, $(x_1, y_j) \in A(D)$. Consider now $i > 1$. Recall that y_j is adjacent to every vertex in $X - \{x_j\}$. Thus, y_j is adjacent to every vertex of $V(Px_i)$. By induction hypothesis, we have $(x_{i-1}, y_j) \in A(D)$. Suppose by contradiction that $(y_j, x_i) \in A(D)$. Then, there is a path $P' = Px_{i-1} \circ (x_{i-1}, y_j, x_i) \circ x_iP$ such that $|P'| = |X| + 1$, a contradiction. Therefore, $(x_i, y_j) \in A(D)$. This completes the proof of Claim 1.

Claim 2: *If $x_i \in X$, $y_j \in Y$ and $j < i$, then $(y_j, x_i) \in A(D)$.*

We omit the proof of Claim 2, as it is analogous to that of Claim 1.

We claim that both $P_0 = (x_1, y_2, x_3, y_4, \dots)$ and $P_1 = (y_1, x_2, y_3, x_4, \dots)$ are paths in D . By Claim 1 we have that $(x_i, y_{i+1}) \in A(D)$ for $i = 1, 3, \dots$, and by Claim 2 we have that $(y_j, x_{j+1}) \in A(D)$ for $j = 2, 4, \dots$. Hence P_0 is a path in D . The proof is analogous for P_1 . Clearly, $\mathcal{P} = \{P_0, P_1\}$ is a path partition of D . Moreover, $|P_0| = |P_1| = \ell$ and $|\mathcal{P}|_k = 2 \min\{\ell, k\} = 2k$. Since $|X| > k$, we have that $\min\{\ell, k\} = k \leq |X| - 1 = |Y| - 1$. Thus, $|\mathcal{P}|_k = 2k \leq k + |Y| - 1$. On the other hand, by Lemma 2, $\alpha_k(D) \geq |Y| + \min\{|X|, k - 1\} = |Y| + k - 1$. Therefore, $\pi_k(D) \leq |\mathcal{P}|_k = 2k \leq |Y| + k - 1 \leq \alpha_k(D)$. ■

3. Conclusion

We showed that Linial's Conjecture holds for k -loose digraphs and for some subclasses of k -tight digraphs, namely those with $|X| = k$ and the thick spider digraphs. It is easy to see that for k -tight digraphs, $\alpha_k(D) = |Y| + k - 1$. Therefore, it is clear that any approach to prove Linial's Conjecture for k -tight digraphs must involve finding a path partition with k -norm less than or equal to $|Y| + k - 1$. We are currently working on this idea.

References

- Berge, C. (1982). k -optimal partitions of a directed graph. *European Journal of Combinatorics*, 3(2):97–101.
- Berger, E. and Hartman, I. B.-A. (2008). Proof of Berge's strong path partition conjecture for $k = 2$. *European Journal of Combinatorics*, 29(1):179–192.
- Dilworth, R. P. (1950). A decomposition theorem for partially ordered sets. *Annals of Mathematics*, 51(1):161–166.
- Gallai, T. and Milgram, A. N. (1960). Verallgemeinerung eines graphentheoretischen satzes von rédei. *Acta Sci Math*, 21:181–186.
- Greene, C. and Kleitman, D. J. (1976). The structure of Sperner k -families. *Journal of Combinatorial Theory, Series A*, 20(1):41–68.
- Herskovics, D. (2013). Proof of Berge's path partition conjecture for $k \geq \lambda - 3$. Technical Report TR-2013-08, Egerházy Research Group, Budapest.
- Hoàng, C. T. (1985). *Perfect Graphs*. PhD thesis, School of Computer Science, McGill University Montreal.
- Linial, N. (1978). Covering digraphs by paths. *Discrete Mathematics*, 23(3):257–272.
- Linial, N. (1981). Extending the Greene-Kleitman theorem to directed graphs. *Journal of Combinatorial Theory, Series A*, 30(3):331–334.
- Rédei, L. (1934). Ein kombinatorischer satz. *Acta Litt. Szeged*, 7(39-43):97.
- Saks, M. (1979). A short proof of the existence of k -saturated partitions of partially ordered sets. *Advances in Mathematics*, 33(3):207–211.

On Colored Edge Cuts in Graphs

Luerbio Faria², Sulamita Klein¹, Ignasi Sau³, Uéverton S. Souza⁴, Rubens Sucupira^{1,2}

¹Universidade Federal do Rio de Janeiro, PESC/COPPE, Brazil

²Universidade do Estado do Rio de Janeiro, Brazil

³CNRS, LIRMM, Montpellier, France

⁴Universidade Federal Fluminense, Brazil

{luerbio, sula}@cos.ufrj.br, {ignasi.sau, rasucupira}@gmail.com,

uevertonssouza@yahoo.com.br

Abstract. *In this work we present some results on the classical and parameterized complexity of finding cuts in edge-colored graphs. In general, we are interested in problems of finding cuts $\{A, B\}$ which minimize or maximize the number of colors occurring in the edges with exactly one endpoint in A .*

1. Introduction

A *cut* in a graph G is a partition of $V(G)$ into two disjoint non-empty sets A and B . The *cut-set* of a cut $\{A, B\}$, denoted by $[A, B]$, is defined as the set of edges of G with exactly one endpoint in A . If s and t are two distinct vertices of a graph G , an (s, t) -*cut* of G is a cut $\{A, B\}$ of G such that $s \in A$ and $t \in B$.

Motivation. MINIMUM CUT and MAXIMUM CUT are two of most popular graph problems and they have several applications. An important application of graph edge cuts is *image segmentation*, in the field of *computer vision*. Image segmentation can be defined as the task of distinguishing objects from background in images, more precisely, it is the process of partitioning a digital image into multiple segments (sets of pixels, also known as superpixels). Classical methods to obtain image segmentations are based on finding edge cuts in uncolored-edge graphs [Shi and Malik 1997, Felzenszwalb and Huttenlocher 2004]. However, as images can be easily seen as graphs, where pixels' colors can be interpreted as edge's colors, the study of combinatorial problems related to cuts in colored-edge graphs seems to be useful to construct new techniques for image segmentation and for other applications as well.

Let $c : E(G) \rightarrow \mathbb{N}$ be a (not necessarily proper) coloring function defined on the edges of a graph G and let $\{A, B\}$ be a cut in G . We denote by $c(A, B)$ the set of colors that appear in the edges of $[A, B]$, that is,

$$c(A, B) = \{i \in \mathbb{N} : \text{there exists } e \in [A, B] \text{ with } c(e) = i\}.$$

Let $\text{im}(c)$ be the image of the coloring function $c : E(G) \rightarrow \mathbb{N}$.

In this paper we consider the following combinatorial problems related to cuts in edge-colored graphs:

MINIMUM COLORED CUT**Input:** A connected graph G and a coloring function $c : E(G) \rightarrow \mathbb{N}$.**Output:** A cut $\{A, B\}$ of G that minimizes $|c(A, B)|$.**MINIMUM COLORED (s, t) -CUT****Input:** A connected graph G , two distinct vertices $s, t \in V(G)$, and a coloring function $c : E(G) \rightarrow \mathbb{N}$.**Output:** An (s, t) -cut $\{A, B\}$ of G that minimizes $|c(A, B)|$.**MAXIMUM COLORED CUT****Input:** A graph G and a coloring function $c : E(G) \rightarrow \mathbb{N}$.**Output:** A cut $\{A, B\}$ of G that maximizes $|c(A, B)|$.**COLORFUL CUT****Input:** A graph G and a coloring function $c : E(G) \rightarrow \mathbb{N}$.**Output:** A cut $\{A, B\}$ of G such that $c(A, B) = \text{im}(c)$, if it exists.

As we can observe, COLORFUL CUT is a special case of MAXIMUM COLORED CUT, and MINIMUM COLORED CUT are easily Turing-reducible to MINIMUM COLORED (s, t) -CUT.

In this article we focus on the computational complexity of these problems, with special emphasis on their parameterized complexity for several choices of the parameters. For an introduction to the field of Parameterized Complexity, see [Flum and Grohe 2006, Niedermeier 2006, Downey and Fellows 2013, Cygan et al. 2015]. We use standard graph-theoretic notation [Diestel 2010]. Throughout the article, we denote by n the number of vertices of the graph.

2. Minimum colored cut and colored (s, t) -cuts

To the best of our knowledge, the MINIMUM COLORED CUT and MINIMUM COLORED (s, t) -CUT problems were first introduced in [Coudert et al. 2007], using different terminology.

Note that if the coloring function c is injective, that is, if all edges get different colors, then the MINIMUM COLORED CUT and MINIMUM COLORED (s, t) -CUT problems correspond exactly to the MINIMUM CUT and MINIMUM (s, t) -CUT problems, respectively, hence they can be both solved in polynomial time by a classical MAXIMUM FLOW algorithm [Diestel 2010].

Theorem 1.

- MINIMUM COLORED (s, t) -CUT *cannot be approximated within a factor of $(1 - \epsilon) \ln(c)$ for any constant $\epsilon > 0$ unless $P = NP$, even if the input graph is bipartite planar graph or complete graph;*
- MINIMUM COLORED (s, t) -CUT *is $W[2]$ -hard parameterized by the cost of the solution, even if the input graph is bipartite planar.*

Proof. Reduction from SET COVER which is known to be $W[2]$ -hard when parameterized by the size of the solution [Flum and Grohe 2006] and $(1 - \epsilon) \ln(c)$ inapproximable (for any constant $\epsilon > 0$ unless $P = NP$) [Dinur and Steurer 2014]. \square

Theorem 2. MINIMUM COLORED (s, t) -CUT *on planar bipartite graphs G remains NP-complete even when:*

- either, each color of G occurs at most three times and each (s, t) -path in G has length two;
- or, each color of G occurs at most twice and each (s, t) -path in G has length at most three.

Proof. Reduction from VERTEX COVER on cubic graphs [Blin et al. 2014]. \square

Theorem 3. MINIMUM COLORED (s, t) -CUT can be solved in polynomial time when each color appears in at most two (s, t) -paths of G .

Proof. We can show this using the polynomial algorithm for MONOTONE WEIGHTED SAT where each variable occurs at most twice [Porschen and Speckenmeyer 2007]. \square

Corollary 4. MINIMUM COLORED (s, t) -CUT can be solved in polynomial time when each color occurs at most twice and each (s, t) -path has length two.

Theorem 5. MINIMUM COLORED CUT and MINIMUM COLORED (s, t) -CUT are FPT when parameterized by the number of colors.

Note that MINIMUM COLORED (s, t) -CUT can be solved in polynomial time when $c = O(\log n)$, and can be solved in pseudo-polynomial time when G has an uncolored (s, t) -cut of size $O(\log n)$. The following result complements Theorem 5 above.

Lemma 6. MINIMUM COLORED (s, t) -CUT does not admit polynomial kernels when parameterized by the number of colors, unless $\text{NP} \subseteq \text{coNP/poly}$.

Proof. The proof uses OR-composition. \square

Lemma 7. MINIMUM COLORED (s, t) -CUT are FPT when parameterized by the number of (s, t) -paths and the size of the solution.

Lemma 8. MINIMUM COLORED (s, t) -CUT does not admit polynomial kernels when parameterized by the number of (s, t) -paths and the size of the solution, unless $\text{NP} \subseteq \text{coNP/poly}$.

Lemma 9. MINIMUM COLORED CUT can be solved in polynomial time when G has an uncolored (s, t) -cut of constant size, and it can be solved in $O(n^5)$ time on planar graphs.

Theorem 10. MINIMUM COLORED CUT for directed graphs is NP-hard.

We denote by *span* of a color c_i the number of connected components in the graph induced by the set of edges colored with c_i .

Lemma 11. Given an edge-colored graph (G, c) , let c_2 denote the number of colors with span at least two. The MINIMUM COLORED CUT problem can be solved in time $3^{c_2} \cdot n^{O(1)}$.

Lemma 12. Given an edge-colored graph (G, c) and a positive integer p , let c_p denote the number of colors with span at least p . The MINIMUM COLORED CUT problem can be solved by a randomized algorithm in time $3^{c_p} \cdot n^{O(1)}$, where the degree of the polynomial depends on p .

3. Maximum colored cut and colorful cut

In this section we list our results on COLORFUL CUT and MAXIMUM COLORED CUT.

Theorem 13. MAX COLORED CUT remains NP-hard even when restricted to complete graphs.

Theorem 14. MAX COLORED CUT admits a polynomial 2-approximation algorithm.

Theorem 15. MAX COLORED CUT can be solved in polynomial time on graphs G colored with a constant number of colors.

Theorem 16. COLORFUL CUT is NP-complete even when each color class induces a clique.

Theorem 17. COLORFUL CUT can be solved in polynomial time on planar graphs if each color class induces a clique.

Theorem 18. COLORFUL CUT is NP-complete even on planar graphs where each color occurs at most twice and each vertex has degree at most 4.

Corollary 19. COLORFUL CUT is NP-complete, even when restricted to planar graphs with odd cycle transversal number at most 1.

Theorem 20. COLORFUL CUT is NP-complete, even when restricted to graphs with vertex cover number at most 4.

Theorem 21. COLORFUL CUT is NP-complete, even when restricted to planar graphs with feedback vertex set number at most 2.

Theorem 22. MAXIMUM COLORED CUT admits a cubic kernel when parameterized by the number of colors.

Theorem 23. MAXIMUM COLORED CUT admits a cubic kernel when parameterized by the cost of the solution.

References

- Blin, G., Bonizzoni, P., Dondi, R., Rizzi, R., and Sikora, F. (2014). Complexity insights of the minimum duplication problem. *Theoretical Computer Science*, 530:66 – 79.
- Coudert, D., Datta, P., Perennes, S., Rivano, H., and Voge, M. (2007). Shared risk resource group complexity and approximability issues. *Parallel Process. Lett.*, 17(2):169–184.
- Cygan, M., Fomin, F. V., Kowalik, L., Lokshtanov, D., Marx, D., Pilipczuk, M., Pilipczuk, M., and Saurabh, S. (2015). *Parameterized Algorithms*. Springer.
- Diestel, R. (2010). *Graph Theory*, volume 173. Springer-Verlag, 4th edition.
- Dinur, I. and Steurer, D. (2014). Analytical approach to parallel repetition. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing, STOC '14*, pages 624–633, New York, NY, USA. ACM.
- Downey, R. G. and Fellows, M. R. (2013). *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer.
- Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181.
- Flum, J. and Grohe, M. (2006). *Parameterized Complexity Theory*. Texts in Theoretical Computer Science. Springer.
- Niedermeier, R. (2006). *Invitation to Fixed-Parameter Algorithms*, volume 31. Oxford University Press.
- Porschen, S. and Speckenmeyer, E. (2007). Algorithms for variable-weighted 2-sat and dual problems. In *Theory and Applications of Satisfiability Testing–SAT 2007*, pages 173–186. Springer.
- Shi, J. and Malik, J. (1997). Normalized cuts and image segmentation. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 731–737. IEEE.

Frustração de Arestas em $(3, 6)$ -Fullerenes

Diego S. Nicodemos¹, Sulamita Klein², Luerbio Faria³

¹Colégio Pedro II / COPPE – Sistemas, Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

²COPPE–Sistemas, Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

³Departamento de Matemática – Universidade do Estado do Rio de Janeiro (UERJ)
Rio de Janeiro – RJ – Brasil

nicodemos@cos.ufrj.br, sula@cos.ufrj.br, luerbio@cos.ufrj.br

Resumo. Grafos $(3, 6)$ -fullerenes são grafos planares, 3-conexos, cúbicos cujas faces têm tamanho 3 ou 6. Determinar o menor número de arestas a serem deletadas de um grafo de modo a obter um subgrafo gerador bipartido é conhecido na literatura [Došlić and Vukičević 2007] como o Problema de Frustração de Arestas. Neste trabalho, abordamos o Problema da Frustração de Arestas em grafos $(3, 6)$ -fullerenes. Mostramos que todo grafo $(3, 6)$ -fullerene com n vértices torna-se bipartido após a retirada de no máximo $\sqrt{\frac{4}{3}n}$ arestas.

Abstract. A $(3, 6)$ -fullerene graph is a cubic bridgeless plane graph with all faces of size 3 or 6. Determining the smallest number of edges that have to be deleted from the graph to obtain a bipartite spanning subgraph is known in the literature [Došlić and Vukičević 2007] as the Bipartite Edge Frustration Problem. In this paper, we investigate the Bipartite Edge Frustration Problem in $(3, 6)$ -fullerene graphs. We show that every graph $(3, 6)$ -fullerene on n vertices becomes bipartite after deleting at most $\sqrt{\frac{4}{3}n}$ edges.

1. Introdução

Segundo Došlić e Vukičević [Došlić and Vukičević 2007] uma aresta $e \in E$ é dita *frustrada* com respeito à bipartição (V_1, V_2) de V se ambas as extremidades de e pertencem a mesma classe da bipartição. O menor número de arestas a serem deletadas de um grafo de modo a obter um subgrafo gerador bipartido é conhecido na literatura como o Problema da Frustração de Arestas e é representado pelo parâmetro τ_{odd} .

Um grafo $(3, 6)$ -fullerene é um grafo planar, cúbico, 3-conexo cujas faces têm tamanho 3 ou 6. Pela Relação de Euler sabe-se que todo grafo $(3, 6)$ -fullerene possui exatamente 4 faces triangulares (faces de tamanho 3). O resultado central deste trabalho fornece uma cota superior para o Problema da Frustração de Arestas em grafos $(3, 6)$ -fullerenes e estende o resultado similar obtido para grafos fullerenes.

Teorema 1.1 *Se G é um grafo $(3, 6)$ -fullerene com n vértices, então $\tau_{odd}(G) \leq \sqrt{\frac{4}{3}n}$.*

O restante do artigo está dividido da seguinte maneira. Na seção 2, discutimos conceitos específicos para a abordagem do Problema da Frustração de Arestas e introduzimos a ideia de remendos e fossos. Na seção 3, provamos o Teorema 1.1.

2. Preliminares

As definições usadas neste trabalho são definições padrões. Para maiores detalhes o leitor é convidado a consultar [Bondy and Murty 2008].

O *grafo dual* de um $(3, 6)$ -fullerene é uma *triangulação planar* sem loops ou arestas múltiplas e todos os seus vértices têm graus 3 ou 6. No dual de um $(3, 6)$ -fullerene os vértices de grau 3 são também chamados de *vértices defeituosos*.

Por toda esta seção, G é uma triangulação planar com todos os vértices de grau menor que 6. Sejam um grafo G e um conjunto de vértices $T \subseteq V(G)$ tal que $|T|$ é par. Uma *T -junção* de G é um subconjunto $J \subseteq E(G)$ tal que T é o conjunto dos vértices de grau ímpar em $G[J]$. Observe que se T é o conjunto dos vértices de grau ímpar de G e J é uma T -junção de G então $|T|$ é par (pois o número de vértices de grau ímpar em um grafo é sempre par) e cada vértice de $G - J$ tem grau par (pois retiramos de G uma quantidade ímpar de arestas referentes aos vértices de grau ímpar e uma quantidade par de arestas referentes aos vértices de grau par). O tamanho da menor T -junção de G é denotado por $\tau(G, T)$.

Seja $\delta_G(X)$ o conjunto de arestas em um grafo G com exatamente um vértice em $X \subseteq V(G)$. Um conjunto C de arestas de G é um *corte de aresta* de G se $C = \delta_G(X)$, para algum $X \subseteq V(G)$. Sejam $X \subseteq V(G)$ e T o conjunto dos vértices de grau ímpar de G . Um *T -corte* é um corte de aresta $\delta(X)$ tal que $|T \cap X|$ é ímpar.

Um *empacotamento* de T -cortes de G é uma coleção disjunta $\delta(\mathcal{F}) = \{\delta(X) \mid X \in \mathcal{F}\}$ de T -cortes de G . Se T é o conjunto dos vértices de grau ímpar de G , então denotaremos por $\nu(G, T)$ o tamanho do maior empacotamento de T -cortes de G . Uma *inclusão sábia minimal* é um conjunto dentre uma coleção de conjuntos que não contém qualquer outro conjunto da coleção. Dado um empacotamento de T -cortes, um T -corte $\delta_G(X)$ é uma inclusão sábia minimal quando $\delta_G(X)$ não contém qualquer T -corte do empacotamento de T -cortes. Uma família \mathcal{F} é dita *laminar* se para cada par $X, Y \in \mathcal{F}$, tem-se $X \subseteq Y$, $Y \subseteq X$, ou $X \cap Y = \emptyset$.

Seja G^* o dual de um $(3, 6)$ -fullerene. O grafo G^* não é bipartido, pois suas faces são todas triangulares. O grafo G^{*l} obtido subdividindo as arestas de G^* é bipartido, pois todas as suas faces têm tamanho 6. Considere o grafo $G^{*\Delta}$ obtido, a partir de G^{*l} , adicionando três novas arestas dentro de cada face de G^{*l} , incidentes a cada um dos 3 vértices de grau 2. Chamamos $G^{*\Delta}$ de *refinamento* de G^* . Todos os vértices em $V(G^{*\Delta}) - V(G^*)$ têm grau 6 em $G^{*\Delta}$, portanto se D é o conjunto dos vértices defeituosos de G^* , então D é também o conjunto dos vértices defeituosos de $G^{*\Delta}$.

O Lema 2.1 foi provado por Klein, Faria e Stehlik [Klein et al. 2012].

Lema 2.1 *Para toda triangulação planar G e todo subconjunto $T \subseteq V(G)$ tal que $|T|$ é par, $\tau(G, T) = \frac{1}{2}\nu(G^\Delta, T)$. Além disso, existe um empacotamento de T -cortes em G^Δ que é laminar, ótimo formado por inclusões sábias minimais.*

Sejam $X \subset V(G)$ e $G[X]$ um subgrafo, 2-conexo, de G tal que todas as faces de $G[X]$, exceto a face exterior, são triângulos. Um *fosso* de largura k em G ao redor do remendo $G[X]$ é um subconjunto de $E(G)$ definido como:

$$\delta_G^k(X) = \bigcup_{i=0}^{k-1} \delta_G(N^i[X]).$$

Em particular, $\delta_G^1(X) = \delta_G(X)$. Se $\sum_{v \in X} (6 - d(v)) = d$, então $\delta_G^k(X)$ é um d -fosso de largura k , como mostra a Figura 1.

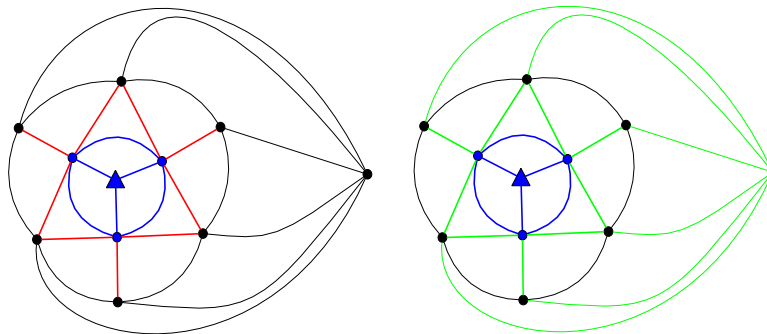


Figura 1. Nas duas figuras os subgrafos azuis representam remendos $G[X]$. As arestas de vermelho indicam um fosso de largura 1 e as arestas de verde um fosso de largura 2. Ambos são 3-fossos.

Se G^* é o dual de um $(3, 6)$ -fullerene, então os possíveis tipos de fossos de G^* são os 3-fossos, os 6-fossos, os 9-fossos e os 12-fossos.

Para todo fosso $\delta_G^k(X)$ corresponde um conjunto $|\delta_G^k(X)|$ de faces triangulares. Dizemos que as faces incidentes a pelo menos uma aresta de $\delta_G^k(X)$ são *geradas* pelo fosso $\delta_G^k(X)$. Se G é um $(3, 6)$ -fullerene, então o número de arestas em um 3-fosso de G^* é facilmente determinado.

Lema 2.2 *Sejam G um grafo $(3, 6)$ -fullerene, G^* o dual de G e D o conjunto dos vértices defeituosos de G^* . Se $d_{G^*}(u) = 3$, e nenhuma aresta de $\delta^{k-1}(u)$ é incidente a vértices do conjunto $D - \{u\}$, então $|\delta_{G^*}^k(u)| = 3k^2$.*

Demonstração. Observe que em cada camada do 3-fosso de G^* a quantidade de arestas é dada pela sequência $(3, 3 + 6, 3 + 6 \cdot 2, 3 + 6 \cdot 3, \dots, 3 + 6 \cdot (k - 1))$. Portanto, $|\delta^k(u)| = \sum_{i=0}^{k-1} |\delta(N^i[u])| = 3k + 6(1 + 2 + 3 + \dots + k - 1) = 3k^2$. \square

3. Resultados Centrais

O problema da frustração de arestas para os grafos $(3, 6)$ -fullerene será resolvido através de sua versão dual, ou seja, determinaremos o menor número de arestas a serem deletadas de um grafo G de modo que o grafo remanescente não possua vértices de grau ímpar. Este parâmetro é denotado por $\tau(G)$. Já discutimos que quando G é um grafo $(3, 6)$ -fullerene o seu correspondente dual G^* é uma triangulação planar cujos vértices têm graus 3 ou 6 e existe pelo Lema 2.1, um empacotamento de T -cortes em $G^{*\Delta}$ que é laminar, ótimo, consistindo apenas de inclusões sábias minimais. Chamamos um tal empacotamento de um *empacotamento de fossos de T -cortes* que, neste caso, é composto apenas de 3-fossos de G^* .

Teorema 3.1 *Seja G^* o grafo dual de um $(3, 6)$ -fullerene. Se f é o número de faces de G^* e T é o conjunto dos vértices de grau ímpar de G^* , então $\tau(G^*, T) \leq \sqrt{\frac{4f}{3}}$.*

Demonstração. Seja $G^{*\Delta}$ o refinamento de G^* . Assim $G^{*\Delta}$ é uma triangulação planar com $4f$ faces e todos os vértices de graus 3 ou 6. Pelo Lema 2.1, existe um empacotamento de fossos $\delta_{G^{*\Delta}}(\mathcal{F})$. Seja m_3 o número de arestas em um 3-moat de $\delta_{G^{*\Delta}}(\mathcal{F})$. Definimos o vetor de incidência $\vec{r} \in \mathbb{R}^4$ da seguinte maneira: para cada $u \in T$, seja r_u a largura do 3-fosso centrado em u .

Definimos o *produto interno* $\langle \cdot, \cdot \rangle$ em \mathbb{R}^4 por $\langle \vec{a}, \vec{b} \rangle = \sum_{u \in T} a_u b_u$. Também definimos a *norma* $\| \cdot \|$ por $\| \vec{a} \| = \langle \vec{a}, \vec{a} \rangle$.

Pela otimalidade de $\delta_{G^{*\Delta}}(\mathcal{F})$,

$$\tau(G^*, T) = \frac{1}{2} \nu(G^{*\Delta}, T) = \frac{1}{2} \langle \vec{r}, \vec{1} \rangle, \text{ sendo } \vec{1} = (1, 1, 1, 1). \quad (1)$$

Suponha que $\delta_{G^{*\Delta}}^r(u)$ é um 3-fosso de $\delta_{G^{*\Delta}}(\mathcal{F})$, para algum $u \in P$. Lembre-se que pelo Lema 2.2, $|\delta_{G^{*\Delta}}^r(u)| = 3r_u^2$, logo somando sobre todos os 3-fossos,

$$m_3 = 3 \sum_{u \in T} r_u^2 = 3 \|\vec{r}\|^2. \quad (2)$$

O grafo $G^{*\Delta}$ tem $4f$ triângulos, e os 3-fossos geram m_3 triângulos de $G^{*\Delta}$. Estes triângulos são mutuamente disjuntos.

Usando (2), temos: $4f \geq m_3 \geq 3\|\vec{r}\|^2$.

Consequentemente temos,

$$\sqrt{\frac{4f}{3}} \geq \|\vec{r}\|. \quad (3)$$

Portanto, por (1), (3) e pela desigualdade de Cauchy-Schwarz, segue que:

$$\tau(G^*, T) = \frac{1}{2} \langle \vec{r}, \vec{1} \rangle \leq \frac{1}{2} \|\vec{r}\| \|\vec{1}\| = \|\vec{r}\|.$$

Concluimos que $\tau(G^*, T) \leq \sqrt{\frac{4f}{3}}$. □

Aplicando o Teorema 3.1 ao grafo dual, obtemos a prova do Teorema 1.1.

Prova do Teorema 1.1. Seja G um grafo planar, cúbico, 3-conexo com n vértices e com todas as faces de tamanho 3 ou 6. O grafo dual de G é uma triangulação planar com n faces e todos os vértices de graus 3 ou 6. Sejam T o conjunto dos vértices de grau ímpar de G^* , J^* uma T -junção mínima de G^* , e J o conjunto de arestas de G que corresponde à J^* . Como $G^* - J^*$ não tem vértices de grau ímpar, $G - J = (G^* - J^*)^*$ é bipartido, e pelo Teorema 3.1, $|J| = |J^*| \leq \sqrt{\frac{4n}{3}}$. □

4. References

Referências

- [Bondy and Murty 2008] Bondy, J. A. and Murty, U. S. R. (2008). *Graph theory*. Macmillan/Elsevier, Canada.
- [Došlić and Vukičević 2007] Došlić, T. and Vukičević, D. (2007). Computing the bipartite edge frustration of fullerene graphs. *Discrete Applied Mathematics*, 155(10):1294–1301.
- [Klein et al. 2012] Klein, S., Faria, L., and Stehlík, M. (2012). Odd cycle transversals and independent sets in fullerene graphs. *SIAM Journal of Discrete Mathematics*, 48(3):1458–469.

Descriptive Complexity of Probabilistic Complexity Classes through Second Order Generalized Quantifiers

Thiago Alves Rocha^{1,2}, Ana Teresa Martins²

¹Departamento de Computação
Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Maracanaú – CE – Brazil

²Departamento de Computação – Universidade Federal do Ceará (UFC)
Caixa Postal 6.021 – 60.455-970 – Fortaleza – CE – Brazil

thiago.alves@ifce.edu.br, ana@lia.ufc.br

Abstract. *Descriptive Complexity deals with the relationship between logical definability and computational complexity on finite structures. As an example in the case of probabilistic complexity classes, we have that BPP is equivalent to the class of problems definable by a randomised inflationary fixed-point logic with counting BPIFP(C). In this paper, we show that we can define logics with generalized second order quantifiers equivalent to probabilistic complexity classes. This quantifiers are used to simulate the behavior of probabilistic Turing machines.*

Resumo. *Complexidade Descritiva lida com a relação entre definibilidade lógica e complexidade computacional em estruturas finitas. Como exemplo no caso de classes de complexidade probabilísticas, temos que BPP é equivalente à classe de problemas definíveis por uma versão randômica da lógica de ponto-fixa inflacionário com contagem BPIFP(C). Neste artigo, nós mostramos que podemos definir lógicas com quantificadores generalizados de segunda ordem equivalentes à classes de complexidade probabilísticas. Estes quantificadores são usados para simular o comportamento de máquinas de Turing probabilísticas.*

1. Introduction

Descriptive Complexity [Grädel et al. 2005] characterize the complexity of a problem based on a logical language needed to express it rather than physical measures such as time and space. We say that a logic \mathcal{L} captures a complexity class \mathcal{C} if \mathcal{L} express all and only the problems of this class. In [Eickmeyer 2011], probabilistic logics $BPFO$ and $BPIFP(C)$ are defined and they capture probabilistic complexity classes $BPAC^0$ and BPP , respectively.

Our contribution is to define logics with generalized second order quantifiers and without randomisation and show that they are strong enough to capture probabilistic complexity class. The approach is to use second order quantifiers to simulate the behavior of randomised algorithms for problems in these classes. We demonstrate this by showing that our logics with generalized second order quantifiers are equivalent to randomised logics defined in the framework of [Eickmeyer 2011].

2. Probabilistic Complexity Classes

We assume basic knowledge of computational complexity classes as P and NP , Turing Machines and languages. In order to investigate probabilistic algorithms, we may formally introduce probabilistic Turing machines as in [Arora and Barak 2009].

Definition 1 (Probabilistic Turing Machine). *A probabilistic Turing machine (PTM) is a Turing machine with two transition functions δ_0, δ_1 . To execute a PTM M on an input x , we choose in each step with probability $1/2$ which transition function to use. For a function $T : N \rightarrow N$, we say that M runs in $T(n)$ -time if for any input x , M halts on x within $T(|x|)$ steps regardless of the random choices it makes.*

Bellow, we have the class BPP that aims to capture efficient probabilistic computation.

Definition 2 (BPP). *The class BPP (Bounded Error Probability) contains all languages L for which there is a probabilistic Turing machine M polynomially bounded in time with the following property: for all inputs w , if $w \in L$, then $\Pr[M(x) = 1] \geq 2/3$ and if $w \notin L$, then $\Pr[M(x) = 0] \geq 2/3$.*

Note that the PTM in the previous definition satisfies a very strong property: For every input, it either accepts it with probability at least $2/3$ or rejects it with probability at least $2/3$. Since a deterministic TM is a special case of a PTM (where both transition functions are equal), we have that $P \subseteq BPP$ while it is not known whether $BPP \subseteq NP$.

3. Randomised Logics

For a τ -structure \mathcal{A} , we denote the class of all $(\tau \cup \rho)$ -expansions of \mathcal{A} by $\chi(\mathcal{A}, \rho)$. We can view $\chi(\mathcal{A}, \rho)$ as a probability space with the uniform distribution. A structure $\mathcal{B} \in \chi(\mathcal{A}, \rho)$ can be seen as a $(\tau \cup \rho)$ -structure such that for all k -ary $R \in \rho$ and all tuples (a_1, \dots, a_k) in the domain of \mathcal{A} we can decide whether $(a_1, \dots, a_k) \in R$ with probability $\frac{1}{2}$.

Definition 3. *Let \mathcal{L} be a logic and $0 \leq \alpha \leq \beta \leq 1$. A formula $\phi \in \mathcal{L}[\tau \cup \rho]$ has a $(\alpha, \beta]$ -gap if for all τ -structures \mathcal{A}*

$$\Pr_{\mathcal{B} \in \chi(\mathcal{A}, \rho)}(\mathcal{B} \models \phi) \leq \alpha \text{ or } \Pr_{\mathcal{B} \in \chi(\mathcal{A}, \rho)}(\mathcal{B} \models \phi) > \beta.$$

Definition 4. *Let \mathcal{L} be a logic and $0 \leq \alpha \leq \beta \leq 1$. The logic $P_{(\alpha, \beta]}\mathcal{L}$ is defined as follows: for each vocabulary τ ,*

$$P_{(\alpha, \beta]}\mathcal{L}[\tau] = \bigcup_{\rho} \{\phi \in \mathcal{L}[\tau \cup \rho] \mid \phi \text{ has a } (\alpha, \beta]\text{-gap}\},$$

where the union ranges over all vocabularies ρ disjoint from τ . Let $\phi \in P_{(\alpha, \beta]}\mathcal{L}[\tau]$, the semantics is defined bellow:

$$\mathcal{A} \models \phi \text{ if and only if } \Pr_{\mathcal{B} \in \chi(\mathcal{A}, \rho)}(\mathcal{B} \models \phi) > \beta.$$

To finish, let $BP\mathcal{L} = P_{(1/3, 2/3]}\mathcal{L}$ and define the randomised logics $BPFO$ and $BPFP(C)$ using Definition 4 such that FO stands for First-Order Logic and $IFP(C)$ stands for Inflationary Fixed-Point with Counting [Eickmeyer 2011].

4. Descriptive Complexity

The Descriptive Complexity [Grädel et al. 2005] deals with the relationship between logical definability and computational complexity on finite structures. While the computational complexity is interested with the cost of computational resources, as time and

space, to decide if structures in a class have a certain property, the Descriptive Complexity is concerned with the logical expression of this property.

We are concerned with the results which states that a logic \mathcal{L} captures a complexity class \mathcal{C} on class of structures \mathcal{D} (notation $\mathcal{L} = \mathcal{C}$). It means that the \mathcal{L} -definable properties of structures in \mathcal{D} are precisely those that are decidable in \mathcal{C} . The definition of $\mathcal{L} = \mathcal{C}$ is introduced by the following two definitions [Grädel et al. 2005].

Definition 5 ($\mathcal{L} \subseteq \mathcal{C}$). *Let \mathcal{L} be a logic, \mathcal{D} a domain of finite structures, and \mathcal{C} a computational complexity class. The logic \mathcal{L} is in \mathcal{C} on \mathcal{D} if, for every fixed vocabulary τ and fixed sentence $\phi \in \mathcal{L}(\tau)$, the complexity of evaluating ϕ on $\mathcal{D}(\tau)$ is a problem in \mathcal{C} .*

Definition 6 ($\mathcal{C} \subseteq \mathcal{L}$). *Let \mathcal{L} be a logic, \mathcal{D} a domain of finite structures, and \mathcal{C} a computational complexity class. \mathcal{C} is in \mathcal{L} on \mathcal{D} if, for every model class $\mathcal{K} \subseteq \mathcal{D}(\tau)$ such that the membership problem is in \mathcal{C} , there is a sentence $\phi \in \mathcal{L}(\tau)$ such that $\mathcal{K} = \{\mathcal{A} \in \mathcal{D}(\tau) \mid \mathcal{A} \models \phi\}$.*

Below, we state and briefly explain some results of Descriptive Complexity that will be used in this work. The next result is the seminal theorem of the Descriptive Complexity area.

Theorem 1 (Fagin’s Theorem). $SO\exists = NP$.

Using this result, we can define any NP problem in a formula of $SO\exists$. For instance, the formula

$$\phi_{3color} = \exists R \exists B \exists G \forall x ((R(x) \vee G(x) \vee B(x)) \wedge \forall y (E(x, y) \rightarrow (\neg(R(x) \wedge R(y)) \wedge \neg(G(x) \wedge G(y)) \wedge (B(x) \wedge B(y)))))$$

define the class of graphs that are 3-colorable, i.e., a graph G satisfies ϕ_{3color} if and only if G is 3-colorable. Three colorability of graphs is an NP -complete problem. For the case of probabilistic complexity classes, [Eickmeyer 2011] has the following results:

Theorem 2 ([Eickmeyer 2011]). $BPFO = BPAC^0$.

where AC^0 is a complexity class of problems that are recognized by a family of Boolean circuits $(C_n)_{n \geq 1}$ such that each circuit C_n has n inputs, one output and the total number of gates is polynomially bounded by n . Besides that, there is a $d > 1$ such that all circuits C_n have depth at most d and we can test whether a circuit $C \in (C_n)_{n \geq 1}$ in $dlogtime$. $BPAC^0$ is defined similarly as BPP .

Theorem 3 ([Eickmeyer 2011]). $BPIFP(C) = BPP$.

5. Logics with Generalized Quantifiers

Generalized quantifiers are generalizations of standard quantifiers \exists and \forall . A general way of defining generalized quantifiers that quantify over second order variables is introduced in [Andersson 2002]. For example, we can define a new quantifier $Q_{>\frac{1}{2}}^k$:

$$Q_{>\frac{1}{2}}^k = \{\langle A, P \rangle \mid P \subseteq \mathcal{P}(A^k) \text{ and } |P| > 2^{|A|^k - 1}\}.$$

Now we can define a logic $\mathcal{L}(Q_{>\frac{1}{2}}^k)$ adding formulas of the form $Q_{>\frac{1}{2}}^k X^k \varphi(X)$ for $\varphi(X) \in \mathcal{L}$. For instance, we can define $FO(Q_{>\frac{1}{2}}^k)$ adding the following case in the semantics of FO :

$$\mathcal{A} \models Q_{>\frac{1}{2}}^k X^k \varphi(X) \text{ if and only if } |\{R \in \mathcal{P}(A^k) \mid \mathcal{A} \models \varphi(R)\}| > 2^{|A|^k - 1}.$$

Now we show that we can define logics with generalized second order quantifiers equivalent to randomized logics $P_{\alpha,\beta}\mathcal{L}$. Let \mathcal{L} be a logic. First, we define $G\mathcal{L}((\mathcal{Q}_{\leq\alpha}^k)_{k\geq 1}, (\mathcal{Q}_{>\beta}^k)_{k\geq 1})$ as

$$G\mathcal{L}((\mathcal{Q}_{\leq\alpha}^k)_{k\geq 1}, (\mathcal{Q}_{>\beta}^k)_{k\geq 1}) = \{ \mathcal{Q}_{>\beta}^k X\varphi(X) \mid k \geq 1, \varphi(X) \in \mathcal{L} \text{ and for all structures } \mathcal{A}, \text{ we have } \mathcal{A} \models \mathcal{Q}_{>\beta}^k X\varphi(X) \text{ or } \mathcal{A} \models \mathcal{Q}_{\leq\alpha}^k X\varphi(X) \}.$$

Now we can show the equivalence:

Theorem 4. *Let $0 \leq \alpha \leq \beta \leq 1$. $G\mathcal{L}((\mathcal{Q}_{\leq\alpha}^k)_{k\geq 1}, (\mathcal{Q}_{>\beta}^k)_{k\geq 1})$ is equivalent to $P_{(\alpha,\beta)}\mathcal{L}$.*

Proof. Let τ be a vocabulary, $\mathcal{Q}_{>\beta}^k X\varphi(X) \in G\mathcal{L}((\mathcal{Q}_{\leq\alpha}^k)_{k\geq 1}, (\mathcal{Q}_{>\beta}^k)_{k\geq 1})$ and \mathcal{A} a τ -structure. We have that $\varphi(X) \in \mathcal{L}[\tau \cup \{X\}]$, thus

$$P_{\mathcal{B} \in \mathcal{X}(\mathcal{A}, \rho)}(\mathcal{B} \models \varphi(X)) = \frac{|\{\mathcal{B} \in \mathcal{X}(\mathcal{A}, \rho) \mid \mathcal{B} \models \varphi(X)\}|}{|\{\mathcal{B} \mid \mathcal{B} \in \mathcal{X}(\mathcal{A}, \rho)\}|} = \frac{|\{X \in \mathcal{P}(A^k) \mid (\mathcal{A}, X) \models \varphi(X)\}|}{2^{|\mathcal{A}|^k}}.$$

$$\mathcal{A} \models \mathcal{Q}_{>\beta}^k X\varphi(X) \text{ iff } |\{R \in \mathcal{P}(A^k) \mid (\mathcal{A}, R) \models \varphi(R)\}| > \beta \times 2^{|\mathcal{A}|^k} \text{ iff } \frac{|\{R \in \mathcal{P}(A^k) \mid (\mathcal{A}, R) \models \varphi(R)\}|}{2^{|\mathcal{A}|^k}} > \beta \text{ iff } P_{\mathcal{B} \in \mathcal{X}(\mathcal{A}, \rho)}(\mathcal{B} \models \varphi(X)) > \beta.$$

$$\mathcal{A} \models \mathcal{Q}_{\leq\alpha}^k X\varphi(X) \text{ iff } |\{R \in \mathcal{P}(A^k) \mid (\mathcal{A}, R) \models \varphi(R)\}| \leq \alpha \times 2^{|\mathcal{A}|^k} \text{ iff } \frac{|\{R \in \mathcal{P}(A^k) \mid (\mathcal{A}, R) \models \varphi(R)\}|}{2^{|\mathcal{A}|^k}} \leq \alpha \text{ iff } P_{\mathcal{B} \in \mathcal{X}(\mathcal{A}, \rho)}(\mathcal{B} \models \varphi(X)) \leq \alpha.$$

□

6. Capturing Results and Future Work

From Theorem 4, we can obtain the following results:

Theorem 5. $GFO((\mathcal{Q}_{\leq 1/3}^k)_{k\geq 1}, (\mathcal{Q}_{> 2/3}^k)_{k\geq 1}) = BPAC^0$.

Theorem 6. $GIFP(C)((\mathcal{Q}_{\leq 1/3}^k)_{k\geq 1}, (\mathcal{Q}_{> 2/3}^k)_{k\geq 1}) = BPP$.

For future work, we want to define logics with second order generalized quantifiers to capture probabilistic exponential time classes.

Acknowledgement

We are grateful to Francicleber Martins Ferreira for his valuable contributions to this paper.

References

- Andersson, A. (2002). On Second-order Generalized Quantifiers and Finite Structures. *Ann. Pure Appl. Logic*, 115(1-3):1–32.
- Arora, S. and Barak, B. (2009). *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition.
- Eickmeyer, K. (2011). *Randomness in Complexity Theory and Logics*. PhD thesis, Humboldt University of Berlin. <http://d-nb.info/1015169163>.
- Grädel, E., Kolaitis, P. G., Libkin, L., Marx, M., Spencer, J., Vardi, M. Y., Venema, Y., and Weinstein, S. (2005). *Finite Model Theory and Its Applications (Texts in Theoretical Computer Science. An EATCS Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

A Gramática de Grafos como uma alternativa para o desenvolvimento do Pensamento Computacional na Educação Básica*

Braz A. S. Junior¹, Simone A. C. Cavalheiro¹, Luciana Foss¹

¹Laboratório de Fundamentos da Computação – Universidade Federal de Pelotas (UFPEL)
Caixa Postal 354 – 96.010-610 – Pelotas – RS – Brasil

{badsjunior, simone.costa, lfoss}@inf.ufpel.br

Abstract. *This paper proposes the creation of a game based on Graph Grammars. The goal of this work is offering a playful option to develop Computational Thinking without using a computer. Particularly, this approach includes the concepts of algorithms, simulation, abstraction, parallelism and analysis.*

Resumo. *Este artigo propõe a criação de um jogo fundamentado em Gramática de Grafos. O objetivo desta proposta é oferecer uma opção lúdica para o desenvolvimento de habilidades do Pensamento Computacional sem o uso do computador. Particularmente, esta abordagem engloba os conceitos de algoritmos, simulação, abstração, paralelismo e análise.*

1. Introdução

Em 2006, Jeannette Wing [Wing 2006] expôs que as habilidades que os profissionais da ciência da computação desenvolvem não são úteis somente para sua área, mas são habilidades de propósito geral, fundamentais para todos em um mundo moderno. O conjunto destas habilidades representa o conceito de Pensamento Computacional (PC). Este conceito foi consolidado por profissionais da área da educação, a *International Society for Technology in Education* (ISTE) e a *Computer Science Teachers Association* (CSTA), através da criação de uma definição operacional para o termo Pensamento Computacional [CSTA and ISTE 2011], que exploramos na seção 2.

Uma das formas de se desenvolver o PC que tem ganhado popularidade recentemente é através do ensino de programação. Em 2013, houve uma campanha global de desmitificação da programação, a *Hour of Code*, com um vídeo¹ viral incentivando as pessoas a aprenderem a programar, que contou com personalidades do mundo da tecnologia como Mark Zuckerberg, do Facebook, e Bill Gates, da Microsoft. As escolhas comuns de abordagens para a introdução à solução de problemas na educação básica e, conseqüente desenvolvimento do PC, são: programação em blocos, linguagens imperativas, funcionais, entre outras.

Este trabalho visa explorar a Gramática de Grafos (GG) como uma abordagem alternativa para se desenvolver o PC na Educação Básica. A inovação desta proposta está na adoção de uma linguagem que permite uma forma rigorosa e eficaz de modelagem, assim, viabilizando a introdução da verificação formal já no primeiro contato com

*Este trabalho é parcialmente financiado pelo CNPq - Processo 447463/2014-4.

¹What Most Schools Don't Teach - <https://www.youtube.com/watch?v=nKIu9yen5nc>

a especificação de soluções de problemas. Além disso, GG possui descrição visual, e consegue expressar de forma natural construções complexas, como as de distribuição, concorrência e paralelismo [Ehrig et al. 1997], que são pouco exploradas nas abordagens comumente utilizadas no ensino de programação na Educação Básica.

A proposta deste trabalho é a especificação e criação de um jogo educacional fundamentado em GG. O escopo de aplicação é a educação básica e, para se tornar atrativo aos estudantes, prioriza-se a existência de componentes lúdicos. Além disso, a abordagem se caracteriza como uma atividade desplugada, já que não necessita de computador. Como relatado em [Bell et al. 2015], a ciência da computação é frequentemente ensinada priorizando programação, mas nem todo aluno acha isto muito motivante, na verdade, o computador muitas vezes se torna uma distração. Além disso, é preciso considerar que muitos ambientes de aprendizado não possuem a infraestrutura e o suporte necessário para oferecer aulas regulares com a utilização de computadores e rede conectada à internet.

Na seção 2, são apresentados a definição operacional e os conceitos básicos do PC. Na seção 3, é descrita uma breve introdução à GG. Na seção 4, são levantados alguns dos conceitos e habilidades do PC que podem ser exploradas com GG. A seção 5 trata da descrição do jogo e a seção 6 traz algumas considerações finais.

2. Pensamento Computacional

Segundo sua definição operacional [CSTA and ISTE 2011], PC é um processo de resolução de problemas que inclui (mas não é limitado a) as seguintes características: formular problemas de uma maneira que nos permita usar um computador ou outras ferramentas para ajudar a resolvê-los; organizar e analisar dados logicamente; representar dados através de abstrações, como modelos e simulações; automatizar soluções através do pensamento algorítmico (uma série de passos ordenados); identificar, analisar e implementar possíveis soluções com o objetivo de alcançar a mais eficiente e efetiva combinação de passos e recursos; e generalizar e transferir este processo de solução de problemas para uma ampla variedade de problemas.

Essas habilidades são sustentadas e aprimoradas por um número de disposições ou atitudes que são dimensões essenciais do PC. Essas disposições ou atitudes incluem: confiança em lidar com complexidade; persistência em trabalhar com problemas difíceis; tolerância à ambiguidade; a habilidade de lidar com problemas em aberto; e a habilidade de se comunicar e trabalhar em grupo para alcançar uma solução ou objetivo comum.

A abstração e a análise constituem pilares do PC. A abstração é a capacidade de extrair apenas as características importantes de um problema para chegar à sua solução. Problemas com elevado grau de complexidade tornam-se difíceis de serem resolvidos, caso não haja uma abstração adequada. A análise é o estudo dos resultados gerados pela especificação. Se eles não saírem como o esperado, há a possibilidade de não ter sido escolhido o nível de abstração adequado ou até mesmo de não ter sido planejada a solução corretamente.

No *toolkit* [CSTA and ISTE 2011] elaborado pela ISTE e CSTA, destaca-se um quadro de progressão com nove conceitos da computação fundamentais para o desenvolvimento do PC: coleta de dados, análise de dados, representação de dados, decomposição de problemas, algoritmos e procedimentos, automação, simulação e paralelismo.

3. Gramática de Grafos

Em termos gerais, as GG [Ehrig et al. 1997] são uma generalização de gramáticas de Chomsky, substituindo *strings* por grafos. Uma GG é composta por três elementos:

Grafo tipo: o grafo tipo estabelece os tipos de vértices e arcos permitidos no sistema. Ou seja, um sistema especificado em GG é restrito ao uso dos tipos de vértices e arcos contidos em seu grafo tipo.

Grafo inicial: o grafo inicial descreve a configuração inicial do sistema.

Regras: regras são morfismos entre grafos que definem as possíveis mudanças do sistema, sendo compostas por: um grafo esquerdo (L); um grafo direito (R); e um homomorfismo de grafos de L em R. O lado esquerdo determina os componentes que devem estar presentes no grafo que descreve a configuração corrente do sistema (grafo estado) para que a regra seja aplicada. Enquanto o lado direito especifica as modificações que devem ocorrer após a aplicação da regra.

A semântica de uma GG é descrita por aplicações de regras a partir do grafo inicial. As regras são aplicadas a um grafo estado quando existe um *match* (um morfismo total entre grafos), o qual identifica que todos os elementos do lado esquerdo da regra estão presentes no grafo estado. A especificação em GG permite a simulação e a verificação formal do sistema descrito, o que possibilita investigar a validade de propriedades.

4. Gramática de Grafos e o Pensamento Computacional

A adoção e promoção de GG como fundamento para a criação de um jogo educacional se justifica por sua relação com conceitos fundamentais do PC, destacados a seguir:

Abstração: a representação e especificação da solução de um problema em GG se dá em um nível de abstração mais alto do que as linguagens de programação usuais.

Análise: o uso de uma linguagem formal na especificação, permite a verificação precisa e análise de propriedades. Bem como a visualização e organização dos dados é facilitada e direta através do uso de uma linguagem visual.

Simulação: na especificação de uma GG, a aplicação de regras a partir do estado inicial simula os comportamentos possíveis do sistema descrito.

Paralelismo: GG permite explorar conceitos usualmente complexos como concorrência e distribuição de forma natural. Devido ao seu estilo declarativo (através do uso de regras), várias regras podem ser executadas ao mesmo tempo, desde que não deletem/modifiquem os mesmos elementos do grafo que representa o estado.

5. O Jogo

O jogo proposto possui um tabuleiro, cartas e personagens infantis, compondo uma GG a ser manipulada pelos jogadores. Logo, ao jogar, os alunos deverão obedecer às definições formais de especificação, manipulação e transformação de grafos em uma GG, tendo assim um primeiro contato com estes conceitos, mas de forma lúdica.

Para englobar diferentes níveis de conhecimento acerca da GG, o jogo pode ser executado nos seguintes modos: o modo simples, em que os jogadores recebem um objetivo e regras prontas, devendo aplicar as regras de uma GG para chegar ao seu objetivo; o modo intermediário, em que eles recebem um objetivo, então devem definir e aplicar as regras para alcançá-lo; e o modo avançado em que eles validam propriedades do sistema ao aplicar regras. As diferentes formas de jogá-lo permitem também com que ele

seja utilizado individualmente ou em grupo, podendo ser adaptado para atender a variadas situações, bastando ao educador julgar qual é mais conveniente.

Para ilustrar o funcionamento do jogo, a Figura 1, apresenta uma configuração do jogo no modo simples. Os vértices da GG são representados por árvores, pelo macaco (personagem do jogador) e pelo cacho de bananas (objetivo do jogo), enquanto as arestas são representadas por cipós. Neste exemplo, o jogador recebe: o objetivo (chegar à árvore que possui o cacho de bananas); uma única regra, a regra de movimentação do macaco (à direita na Figura 1); e o grafo inicial (à esquerda na Figura 1).

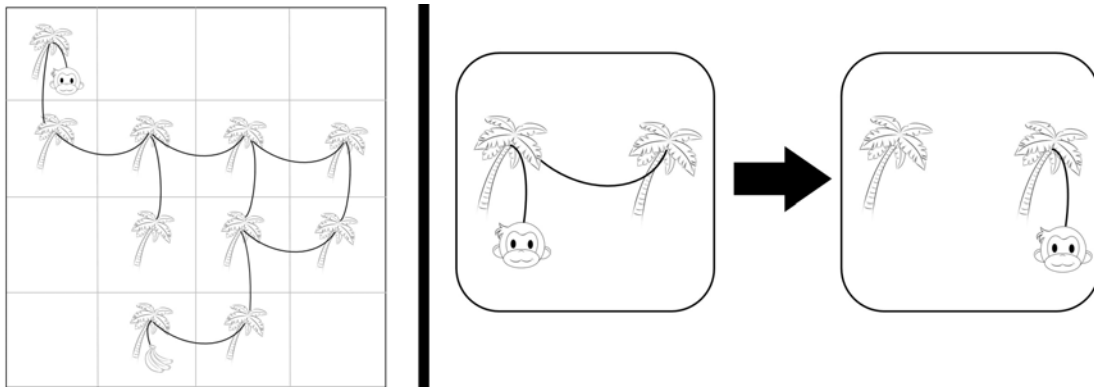


Figura 1. Grafo Inicial (à esquerda) e Regra (à direita)

Partindo do grafo inicial recebido, o jogador deve aplicar, iterativamente, a regra de movimentação do macaco até que o objetivo seja alcançado, ou não haja mais possibilidade de aplicações da regra. O jogador deverá observar que para mover o macaco de uma árvore à outra, é necessário existir um cipó entre elas e este cipó deve ser retirado do jogo após a movimentação do macaco (como descrito pela regra). Portanto, o jogador deve estar atento às aplicações que podem excluir cipós essenciais para o caminho até seu objetivo (isolando-o em uma árvore sem bananas) e evitá-las.

6. Conclusão

Neste artigo é proposta uma abordagem fundamentada em GG como uma alternativa para o desenvolvimento do PC na educação básica. Para isto, é apresentada a proposta de um jogo educacional capaz de simular uma GG. O produto é um jogo versátil, que pode ser jogado de diferentes maneiras, para explorar conceitos do PC de acordo com a faixa etária dos estudantes. E ainda, utiliza-se o conceito de atividade desplugada, para reduzir eventuais dificuldades em sua implementação, como a dependência da infraestrutura.

Referências

- Bell, T., Witten, H., and Fellows, M. (2015). *CS Unplugged*.
- CSTA and ISTE (2011). *Computational Thinking Leadership Toolkit*.
- Ehrig, H. et al. (1997). *Handbook of Graph Grammars and Computing by Graph Transformation, Volume 1*.
- Wing, J. (2006). Computational thinking. In *Communications of the ACM*, volume 49, pages 33–35. ACM.

Sobre coloração total dos grafos r -partidos completos

Raphael Martins¹, Diana Sasaki¹

¹Instituto de Matemática e Estatística da UERJ, Rio de Janeiro, RJ, Brasil

raphaelbgmartins@hotmail.com, diana.sasaki@ime.uerj.br

Abstract. *In this work, we investigate the total coloring and the equitable total coloring problems in the family of complete r -partite graphs. We prove that every complete bipartite graph $K_{n,n}$ cannot be colored with a $(\Delta + 1)$ -total coloring and we prove that every complete bipartite graph $K_{n,m}$ ($n > m \geq 1$) has a $(\Delta + 1)$ -equitable total coloring. Furthermore, we present a property on equitable total colorings of balanced complete r -partite graphs K_n^r .*

Resumo. *Neste trabalho, investigamos os problemas de coloração total e coloração total equilibrada na família dos grafos r -partidos completos. Provamos que todo grafo bipartido completo $K_{n,n}$ não possui uma $(\Delta + 1)$ -coloração total e provamos que todo grafo bipartido completo $K_{n,m}$ ($n > m \geq 1$) possui uma $(\Delta + 1)$ -coloração total equilibrada. Além disso, apresentamos uma propriedade de coloração total equilibrada dos grafos r -partidos completos balanceados K_n^r .*

1. Introdução

Coloração de grafos é um problema desafiador que modela diversas situações reais onde as adjacências representam conflitos. Um grafo $G = (V, A)$ consiste em um conjunto de vértices V conectados por um conjunto de arestas A . Uma k -coloração total de G é uma atribuição de k cores aos elementos (vértices e arestas) de G de forma que elementos adjacentes possuam cores diferentes. O número cromático total χ'' de G é o menor k tal que o grafo possui uma k -coloração total. A Conjectura da Coloração Total afirma que $\Delta + 1 \leq \chi'' \leq \Delta + 2$, onde Δ é o grau máximo do grafo [Behzad 1965, Vizing 1968]. Grafos com $\chi'' = \Delta + 1$ são chamados de *Tipo 1*, caso contrário, são chamados de *Tipo 2*. Uma coloração total é *equilibrada* quando a diferença entre as quantidades de vezes em que a cor que mais aparece e a cor que menos aparece é de no máximo um. Similarmente, o número cromático de coloração total equilibrada χ_e'' de G é o menor k tal que o grafo possui uma k -coloração total equilibrada. Assim como na coloração total, foi conjecturado que o número cromático de coloração total equilibrada é no máximo $\Delta + 2$ [Wang 2002]. Dizemos que a cor i está *representada* em um vértice v , se o vértice v ou suas arestas incidentes possuem a cor i .

Um grafo G é *bipartido* se V pode ser particionado em conjuntos Y e U tal que toda aresta de G possui um extremo em Y e outro extremo em U . Um grafo é *bipartido completo* quando todo vértice de Y é adjacente a todos os vértices de U . Denotamos um grafo bipartido completo por $K_{n,m}$, onde $|Y| = n$ e $|U| = m$; $Y = \{y_1, y_2, y_3, \dots, y_n\}$ e $U = \{u_1, u_2, u_3, \dots, u_m\}$. Um grafo é dito *r -partido completo* quando este pode ser particionado em r partes de modo que vértices de mesma parte não sejam adjacentes e todo vértice de cada parte seja adjacente a todos os vértices das outras partes. Quando

um r -partido completo possui n vértices em cada parte, denotamos estes grafos por K_n^r e dizemos que este é um grafo r -partido completo balanceado.

Neste trabalho, verificamos que todo grafo $K_{n,n}$ é Tipo 2, o que já sabíamos em [Behzad et al. 1967]; apresentamos uma outra prova de que todo grafo $K_{n,m}$, com $n \neq m$, possui uma $(\Delta + 1)$ -coloração total equilibrada, resultado originalmente citado em [Fu 1994]. Além disso, apresentamos uma propriedade de coloração total equilibrada para os grafos K_n^r .

2. Grafos bipartidos completos

Teorema 1. *Todo grafo $K_{n,n}$ é Tipo 2. Além disso, todo $K_{n,n}$ admite $(\Delta + 2)$ -coloração total equilibrada.*

Demonstração. Verificamos facilmente que o grafo $K_{1,1}$ é Tipo 2. Considere $n \geq 2$. Inicialmente provamos que não existe $(\Delta + 1)$ -coloração total do grafo $K_{n,n}$ utilizando uma mesma cor para dois ou mais vértices de mesma parte. De fato, suponha que exista uma $(\Delta + 1)$ -coloração total do grafo $K_{n,n}$ tal que os vértices y_1 e y_2 possuam mesma cor 1, sem perda de generalidade. Como o grafo é Δ -regular, todo vértice precisa ter a cor 1 representada. No vértice u_1 a única forma de se representar a cor 1, é atribuindo esta cor à aresta u_1y_3 ou u_1y_4 ou \dots ou u_1y_n . Em qualquer escolha, a cor 1 estará representada em mais um vértice de Y . Após, a cor 1 estará representada em três vértices de Y e em um vértice de U . Se continuarmos representando cor 1 nos vértices de U , assim que faltarem dois vértices de U para terem a cor 1 representada, todos os vértices de Y estarão com a cor 1 representada. Assim, o único modo de representar a cor 1 nestes dois vértices é atribuindo a cor aos próprios vértices, o que não é possível pois são adjacentes à y_1 e y_2 . A Figura 1 (esquerda) apresenta uma ilustração deste caso.

Agora, note que não existe $(\Delta + 1)$ -coloração total do $K_{n,n}$ utilizando n cores nos n vértices de uma mesma parte. De fato, suponha que os n vértices de Y possuam cores diferentes de 1 até n (ou seja, Δ cores). Isto implica que os vértices de U precisam possuir a cor $n + 1$, pois são adjacentes a todos os vértices de Y , o que não é possível, pois atribuindo mesma cor a todos os vértices de U , cairemos no caso anterior. A Figura 1 (direita) apresenta o grafo $K_{4,4}$ com alguns vértices coloridos. É imediato ver que a única cor disponível para os vértices ainda não coloridos é a cor 5. Ao colorir todos os vértices de uma parte com a cor 5, pelo caso anterior, temos que esta coloração não poderá ser estendida para uma $(\Delta + 1)$ -coloração total do $K_{4,4}$.

Pelo conhecido Teorema de König, existe uma Δ -coloração das arestas do $K_{n,n}$. Assim, colorindo os vértices de Y com cor $\Delta + 1$ e de U com cor $\Delta + 2$, temos que todo $K_{n,n}$ é Tipo 2. Mais especificamente, todo $K_{n,n}$ possui uma $(\Delta + 2)$ -coloração total equilibrada. \square

Teorema 2. *Todo $K_{n,m}$ ($n > m \geq 1$) possui uma $(\Delta + 1)$ -coloração total equilibrada.*

Demonstração. Começamos apresentando uma $(\Delta + 1)$ -coloração total equilibrada para o grafo $K_{n,m}$. Notemos que $\Delta = n$. Atribuímos a cor 1 a todos os m vértices de U . Assim, esta cor não poderá ser mais utilizada nos outros elementos. Atribuímos as cores 2, 3, \dots , n e $n + 1$, às arestas $u_1y_1, u_1y_2, \dots, u_1y_n$, respectivamente. Em seguida, as arestas $u_2y_1, u_2y_2, \dots, u_2y_n$, recebem as cores 3, 4, \dots , $n + 1$ e 2, resp. Continuando

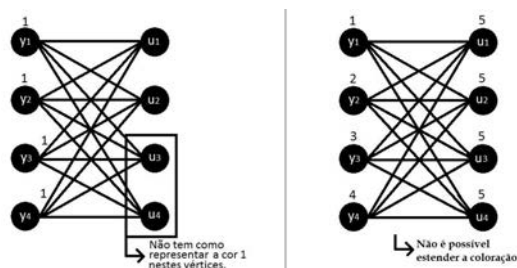


Figura 1. Ilustração da prova do Teorema 1.

a coloração de forma análoga, nas arestas incidentes ao vértice u_m a ordem de cores será $m + 1, m + 2, \dots, \Delta + 1, 2, 3, \dots, m$.

Com a atribuição feita acima, vejamos que não há conflito de cores nas arestas incidentes a cada vértice de Y . De fato, às m arestas incidentes ao vértice y_i , ou seja, $u_1y_i, u_2y_i, \dots, u_my_i, i = 1, \dots, n$, atribuímos m cores seguindo a ordem seguinte: $i + 1, i + 2, \dots, n, n + 1, 2, 3, \dots, i - 1$. Note que, em cada vértice de Y não haverá repetição de cores nas arestas, pois para isto ocorrer precisaríamos atribuir pelo menos $n + 1$ cores às arestas, o que não é possível.

Resta atribuir cores aos vértices de Y . Observe que restam $n + 1 - m - 1 = n - m$ cores disponíveis para cada vértice de Y , pois foram utilizadas m cores para as arestas incidentes a eles e uma cor para os vértices de U . Atribuímos a cor $n + 1$ ao vértice y_1 . Observe que esta cor não foi utilizada em nenhuma aresta incidente à ele, pois foram utilizadas as primeiras m cores da listagem acima para estas arestas. Para $i \geq 2$, atribuímos a cor i ao vértice y_i . De fato, esta cor não é utilizada na listagem de cores das arestas incidentes ao vértice.

Vamos provar que esta coloração é equilibrada. Atribuímos a cor 1 para todos os m vértices de U . Todas as outras cores $2, 3, \dots, \Delta + 1$ foram representadas uma única vez em cada vértice de U , aparecendo também m vezes cada uma. Por fim, cada vértice de Y recebe cores diferentes entre si dentre $2, 3, \dots, \Delta + 1$. Assim, todas essas cores aparecem um total de $m + 1$ vezes e a cor 1 aparece m vezes. Dessa forma, concluímos que todo grafo $K_{n,m}$ possui uma $(\Delta + 1)$ -coloração total equilibrada. Uma 6-coloração total equilibrada do grafo $K_{5,4}$ é apresentada na Figura 2. \square

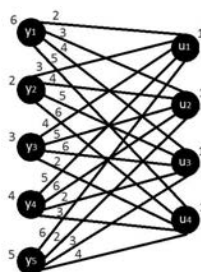


Figura 2. A 6-coloração total equilibrada obtida pelo Teorema 2 do grafo $K_{5,4}$.

3. Grafos r -partidos completos

A seguir, apresentamos uma propriedade de coloração total equilibrada nos grafos r -partidos completos balanceados que nos ajudará na determinação de colorações totais

equilibradas para estes grafos.

Propriedade 1. *Em toda $(\Delta + 1)$ -coloração total equilibrada do grafo K_n^r , $r \geq 3$ e n par, o número de vértices coloridos com mesma cor é exatamente 2.*

Demonstração. Como n é par, o número de vértices do grafo K_n^r também é par. Suponha que um número ímpar de vértices possuem mesma cor i em uma $(\Delta + 1)$ -coloração total equilibrada. Neste caso, restará um número ímpar de vértices sem cor i representada. Como a única forma de se representar a cor i é através das arestas, sempre restará exatamente um vértice sem a cor i representada, pois atribuindo uma cor a uma aresta, estamos representando esta cor em dois vértices.

Afirmamos que sempre teremos cores que serão atribuídas apenas às arestas. De fato, suponha que foram coloridos pelo menos dois vértices de cada cor, totalizando pelo menos $\frac{rn}{2}$ cores utilizadas em vértices. Temos que $\Delta = (r - 1)n = rn - n$, que é sempre maior que $\frac{rn}{2}$, para $r \geq 3$. Assim, como o número total de cores é $\Delta + 1 = (r - 1)n + 1 = rn - n + 1$, este valor também será sempre maior que $\frac{rn}{2}$. Isto implica que sobram cores para serem utilizadas somente nas arestas. Cada classe de cor que aparece somente nas arestas tem um total de $\frac{rn}{2}$ elementos.

Suponha agora que temos um número par $P \geq 2$ de vértices coloridos com mesma cor i . O número de arestas coloridos com a cor i é igual a $\frac{(r-1)n}{2} + \frac{n-P}{2}$, pois precisamos ter a cor i representada (através de arestas) em todos os n vértices de cada uma das outras $r - 1$ partes e nos demais $n - P$ vértices da parte que possui os P vértices coloridos com cor i . Assim, o total de elementos (vértices e arestas) coloridos com mesma cor i é igual a $P + \frac{(r-1)n}{2} + \frac{n-P}{2} = \frac{2P+(r-1)n+n-P}{2} = \frac{P}{2} + \frac{rn}{2}$.

Observe que o único caso em que teremos a diferença permitida de 1 unidade é quando $P = 2$, já que temos classes de cores utilizando $\frac{rn}{2}$ elementos (apresentado acima). Para todos os valores de P maiores do que 2, a coloração não é equilibrada. \square

4. Conclusões

Neste trabalho, investigamos a coloração total dos grafos bipartidos e r -partidos completos balanceados. Como trabalhos futuros, continuaremos a coloração total na classe dos grafos r -partidos completos balanceados.

Referências

- Behzad, M. (1965). *Graphs and their chromatic numbers*. PhD thesis, Michigan State University.
- Behzad, M., Chartrand, G., and Cooper Jr, J. K. (1967). The colour numbers of complete graphs. *J. Lond. Math. Soc.*, 42:226–228.
- Fu, H. L. (1994). Some results on equalized total coloring. *Congr. Numer.*, 102:111–119.
- Vizing, V. G. (1968). Some unsolved problems in graph theory. *Russian Math. Surveys*, 23(6):125–141.
- Wang, W. F. (2002). Equitable total coloring of graphs with maximum degree 3. *Graphs Combin.*, 18:677–685.

Amostragem Gaussiana na Criptografia Baseada em Reticulados

Jheyne N. Ortiz¹, Diego F. Aranha¹, Ricardo Dahab¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas – SP – Brasil

Abstract. *In Lattice-Based Cryptography, some cryptosystems require sampling lattice points and integers over a distribution that, conventionally, follows a Gaussian function. This work presents a constant-time implementation of the Knuth-Yao method for Gaussian sampling over integers. For Gaussian sampling over lattices, the results refer to the sampling over NTRU lattices. Our experiments target an Intel Ivybridge desktop and all implementations are described using C++ language with support of the NTL library of Victor Shoup.*

Resumo. *Na Criptografia Baseada em Reticulados, diversos esquemas requerem a amostragem de vetores de um reticulado e de inteiros seguindo uma distribuição que, convencionalmente, é uma função Gaussiana. Este trabalho apresenta uma implementação com tempo de execução constante para o método Knuth-Yao apropriado à amostragem Gaussiana sobre os inteiros. Para a amostragem de vetores de um reticulado, os resultados se referem à amostragem sobre reticulados NTRU. Os experimentos têm como alvo um computador pessoal Intel Ivybridge e as implementações são descritas em linguagem C++ com aporte da biblioteca NTL de Victor Shoup.*

1. Introdução

Esquemas sobre reticulados como, por exemplo, Encriptação Baseada em Identidades, funções de resumo seguidas de assinatura e Encriptação Baseada em Atributos requerem a amostragem de vetores de um reticulado. A amostragem sobre os inteiros é um passo em métodos de amostragem de pontos de um reticulado além de ser necessária no esquema de encriptação [Lyubashevsky et al. 2012] e no esquema de assinatura [Ducas et al. 2013] baseados no problema Ring-LWE. Usualmente, ambas amostragens sobre reticulados e sobre os inteiros seguem uma distribuição Gaussiana.

No esquema de encriptação baseado no problema Ring-LWE, a amostragem de ruídos conforme a Gaussiana ocorre na fase de encriptação que comumente é implementada em um dispositivo vulnerável a ataques por canais laterais. Similarmente, como demonstrado por [Bruinderink et al. 2016], o esquema de assinatura BLISS é suscetível a ataques por canais laterais bem como o amostrador Gaussiano utilizado na fase de assinatura. Por isso, é importante que implementações de tais algoritmos sejam não somente eficientes, como também incorporem técnicas de proteção contra ataques de canais laterais. Neste trabalho, somente ataques por canais de tempo são tratados.

Objetivos. Neste trabalho, os objetivos consistem na implementação de um algoritmo específico para amostragem Gaussiana sobre reticulados NTRU bem como a implementação em software de um algoritmo para amostragem sobre os inteiros tendo em vista eficiência e resistência contra ataques por canais de tempo.

2. Implementação de Amostragem de Gaussianas Discretas

A função Gaussiana determina a probabilidade de amostragem dos elementos de um determinado domínio. Para o conjunto dos inteiros, a distribuição Gaussiana, parametrizada pelo centro $c \in \mathbb{R}$ e um fator $\sigma \in \mathbb{R}^+$, é uma função que determina a probabilidade de um valor inteiro x ser amostrado em relação aos demais, conforme a Equação 1.

$$\mathcal{D}_{\mathbb{Z},\sigma,c}(x) = \frac{\rho_{\sigma,c}(x)}{\sum_{y=-\infty}^{\infty} \rho_{\sigma,c}(y)}, \text{ com } \rho_{\sigma,c}(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (1)$$

No contexto deste trabalho, devido à estrutura isométrica da base curta que gera o reticulado NTRU, a ortogonalização de Gram-Schmidt da base do reticulado é obtida através da chamada ao procedimento BLOCK-GSO proposto por [Lyubashevsky and Prest 2015]. O método híbrido de [Ducas and Prest 2015] para amostragem Gaussiana de pontos de um reticulado, denotado HYBRID-SAMPLER, é otimizado para amostragem de reticulados sobre o anel de inteiros de um corpo ciclotômico e consiste em uma adaptação do método de Klein [Gentry et al. 2008] sobre anéis. Em uma etapa intermediária, o método HYBRID-SAMPLER requer inteiros amostrados conforme uma distribuição Gaussiana, tarefa desempenhada pelo algoritmo Knuth-Yao [Knuth and Yao 1976].

O algoritmo Knuth-Yao utiliza o conceito de árvore DDG (*Discrete Distribution Generating tree*) para gerar uma amostra, que representa o rótulo de um nó desta árvore. A árvore DDG é descrita pela matriz de probabilidades $\mathbf{P} \in \mathbb{Z}_2^{2t\sigma \times \lambda}$, que consiste na expansão binária das probabilidades dos elementos no intervalo $[c-t\sigma, c+t\sigma] \cap \mathbb{Z}$. Sendo um método finito, as probabilidades são calculadas com λ bits de precisão e podem ser armazenadas preliminarmente para uso na amostragem.

Grosso modo, o método Knuth-Yao consiste em percorrer a matriz de probabilidades computando a distância entre o nó sendo visitado e o nó mais à direita na árvore DDG. Para que a implementação tenha tempo de execução independente da entrada, a execução do algoritmo é extrapolada para o pior caso sem que haja sobre-escrita de amostras. Neste caso, uma solução é a combinação de uma somatória com operações aritméticas que substituem desvios condicionais inerentes ao algoritmo. As operações binárias testam se a distância, que determina que uma amostra foi encontrada, é igual a -1 , utilizando o Código 1 e definem o valor lógico da variável de controle `enable` como verdadeiro em caso afirmativo.

Código 1. Teste de igualdade em tempo constante.

```
unsigned isEqual(int x, int y) {
    unsigned equal = (x-y);
    return (1 ^ ((equal | -equal) >> 31) & 1);
}
```

Uma variável lógica, `hit`, guarda a informação de que uma amostra já foi aceita. Com esses dois valores inteiros, um seletor determina se o valor inválido `invalidSample`, que se encontra fora do intervalo de amostragem $[c-t\sigma, c+t\sigma] \cap \mathbb{Z}$, ou o valor do índice da coluna `col`, que é o rótulo da amostra, será somado à variável de saída denotada por S . O seletor é uma função com tempo de execução constante para atribuição condicional que é determinada pela variável `bit` conforme o Código 2.

Código 2. Seletor para atribuição condicional.

```

int Select(int a, int b, unsigned bit) {
    unsigned mask = -bit;
    return (mask & (a ^ b)) ^ a;
}

```

No término do algoritmo, o valor de S é da forma $(k \cdot \text{invalidSample} + c)$, com $k \in \mathbb{Z}$. Assim sendo, a redução módulo invalidSample remove o erro e a operação $S = S - (t + \sigma) + c$ obtém o valor da amostra corrigido.

3. Resultados Experimentais

A Figura 1 apresenta os tempos de execução em segundos para a amostragem de coeficientes inteiros para um polinômio com diferentes dimensões utilizando o método Knuth-Yao. Na versão em tempo constante do método Knuth-Yao, denotada por Knuth-YaoC, a execução prossegue é extrapolada para o pior caso e, então, toda a estrutura que representa a matriz de probabilidades é visitada. Portanto, esse é o principal fator na perda de desempenho da versão Knuth-YaoC, cuja redução no tempo de execução pode chegar em até sete vezes em comparação com a implementação direta do método Knuth-Yao quando o polinômio tem 8192 coeficientes. Todos os resultados desta seção foram obtidos de uma máquina Intel Ivybridge com processador Intel®Core™i5-3570 @ 3.40 GHz e 8 GB de memória física.

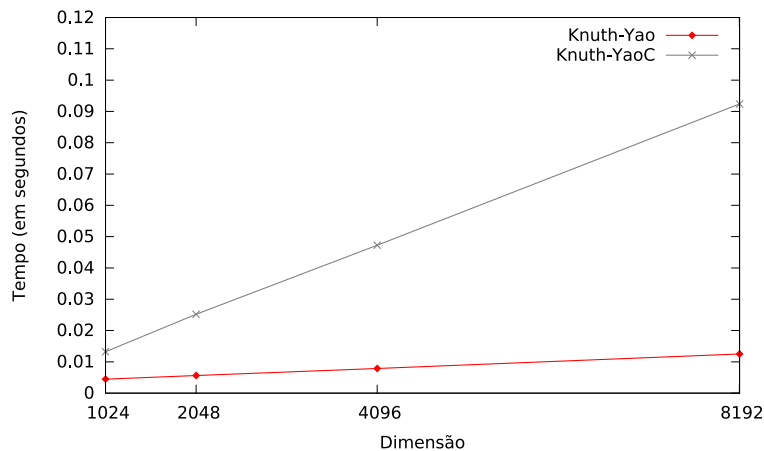


Figura 1. Comparação entre as implementações para o Knuth-Yao com $\sigma = 3, 19$ e $\lambda = 107$.

A Tabela 1 descreve os tempos de execução em segundos para todas as etapas da amostragem utilizando a implementação do método híbrido de [Ducas and Prest 2015]. Os tempos de execução para o experimento sobre reticulados NTRU são a média de cem tempos de execução para os algoritmos KEYGEN, BLOCK-GSO, OFFLINE-KLEIN e OFFLINE-PEIKERT. Para a amostragem propriamente dita, os tempos são a média de dez execuções.

O algoritmo KEYGEN é responsável pela geração da base do reticulado NTRU $\mathbf{B} \in \mathbb{Z}^{2N \times 2N}$. Os algoritmos OFFLINE-KLEIN e OFFLINE-PEIKERT denotam as fases de pré-computação associadas aos algoritmos RING-KLEIN, que é instanciado como o método híbrido, e RING-PEIKERT, que fornece amostras do anel de inteiros \mathcal{R} ao algoritmo RING-KLEIN, respectivamente.

Algoritmo	Dimensão (N)			
	128	256	512	1024
KEYGEN	0, 10	0, 44	2, 20	11, 02
BLOCK-GSO	0, 11	0, 45	2, 37	7, 43
OFFLINE-KLEIN	3, 60	17, 15	97, 46	505, 39
OFFLINE-PEIKERT	3, 19	5, 60	20, 94	82, 37
HYBRID-SAMPLER	15, 58	72, 38	373, 23	2042, 71

Tabela 1. Tempos de execução, em segundos, para o método híbrido com variação no valor de N .

Em suma, os tempos de execução para o método híbrido [Ducas and Prest 2015] representam a ineficiência de algoritmos para amostragem Gaussiana sobre reticulados. Por outro lado, as fases de pré-computação são invocadas com baixa frequência uma vez que a base do reticulado usualmente é a chave privada mestre do sistema criptográfico. Os resultados na Figura 1 ilustram a dificuldade em desenvolver implementações eficientes e seguras contra ataques por canais laterais para métodos para amostragem Gaussiana sobre os inteiros. A necessidade de extrapolar a execução do método Knuth-Yao para o pior caso visando uma implementação com tempo de execução constante reflete o balanço entre segurança e eficiência na implementação de métodos para amostragem Gaussiana discreta.

Referências

- Bruinderink, L. G., Hülsing, A., Lange, T., and Yarom, Y. (2016). Flush, Gauss, and Reload – A Cache Attack on the BLISS Lattice-Based Signature Scheme. Cryptology ePrint Archive, Report 2016/300. <http://eprint.iacr.org/>.
- Ducas, L., Durmus, A., Lepoint, T., and Lyubashevsky, V. (2013). *Advances in Cryptology – CRYPTO 2013: 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, chapter Lattice Signatures and Bimodal Gaussians, pages 40–56. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Ducas, L. and Prest, T. (2015). A Hybrid Gaussian Sampler for Lattices over Rings. Cryptology ePrint Archive, Report 2015/660. <http://eprint.iacr.org/>.
- Gentry, C., Peikert, C., and Vaikuntanathan, V. (2008). Trapdoors for Hard Lattices and New Cryptographic Constructions. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing, STOC '08*, pages 197–206, New York, NY, USA. ACM.
- Knuth, D. and Yao, A. (1976). *Algorithms and Complexity: New Directions and Recent Results*, chapter The Complexity of Nonuniform Random Number Generation. Academic Press, New York, j. f. traub edition.
- Lyubashevsky, V., Peikert, C., and Regev, O. (2012). On Ideal Lattices and Learning with Errors Over Rings. Cryptology ePrint Archive, Report 2012/230. <http://eprint.iacr.org/>.
- Lyubashevsky, V. and Prest, T. (2015). Quadratic Time, Linear Space Algorithms for Gram-Schmidt Orthogonalization and Gaussian Sampling in Structured Lattices. Cryptology ePrint Archive, Report 2015/257. <http://eprint.iacr.org/>.

Achieving CCA1-security in homomorphic encryption

Eduardo Morais¹, Diego F. Aranha¹, Ricardo Dahab¹

¹Institute of Computing – University of Campinas (Unicamp)

{dfaranha, rdahab}@ic.unicamp.br, emorais@lasca.ic.unicamp.br

Abstract. *This paper proposes the combination of homomorphic encryption and verifiable computation to avoid key recovery attacks and achieve CCA1-secure constructions of Somewhat Homomorphic Encryption (SHE) schemes described in the literature. We also provide concrete parameters, based on the best-attack analysis, concluding that the AGCD [van Dijk et al. 2010] family of SHE schemes may be the best implementation choice under certain circumstances.*

Resumo. *Este artigo propõe a combinação de encriptação homomórfica e computação verificável para evitar ataques de recuperação de chaves e obter segurança CCA1 em construções de esquemas parcialmente homomórficos descritos na literatura. Além disso, são propostos parâmetros concretos, baseados na análise do melhor ataque, concluindo que a família AGCD [van Dijk et al. 2010] de esquemas SHE pode ser consideradas a melhor escolha em determinadas circunstâncias.*

1. Introduction

Homomorphic encryption has been a topic of great interest for the Cryptology community over the last few years, since Craig Gentry’s breakthrough in 2009 [Gentry 2009]. Although Fully Homomorphic Encryption (FHE) is yet to become practical, SHE schemes can be used to construct practical applications. An important drawback, however, is the fact that all but one of the SHE schemes described in the literature are susceptible to *key recovery attacks*, a concrete threat in many scenarios. In this work, we investigate how verifiable computation can be combined with homomorphic encryption in order to avoid key recovery attacks. Indeed, we show that it is possible to achieve CCA1-security for homomorphically evaluating quadratic multivariate polynomials.

1.1. Homomorphic encryption

Informally, homomorphic encryption provides the possibility of having a pair of encryption and decryption functions, ENC, DEC, that allows the computation of a function f on an encrypted text c , such that $\text{DEC}(f(c)) = f(m)$, where $c = \text{ENC}(m)$. That is, allowing functions to be computed on encrypted texts without the need for decrypting them first. The following two definitions formalize this notion.

Definition 1.1. *Correctness.* A scheme $\mathcal{E}(\text{KEYGEN}, \text{DEC}, \text{ENC}, \text{EVAL})$ is correct if, for a determined circuit \mathbf{C} and every key pair (sk, pk) , where sk is the private key and pk is the public key generated by KEYGEN , any message tuple $\bar{m} = \langle m_1, \dots, m_t \rangle$ and corresponding ciphertexts $\bar{c} = \langle c_1, \dots, c_t \rangle$, that is, $c_i = \text{ENC}_{pk}(m_i)$ for $1 \leq i \leq t$, then we have that

$$\text{DEC}_{sk}(\text{EVAL}_{pk}(\mathbf{C}, \bar{c})) = \mathbf{C}(\bar{m}).$$

Furthermore, algorithms KEYGEN , DEC , ENC and EVAL must have polynomial complexity.

Definition 1.2. *Fully Homomorphic Encryption.* A scheme \mathcal{E} is correct for a class $\mathbf{S}_{\mathbf{C}}$ of circuits, if it is correct for each $\mathbf{C} \in \mathbf{S}_{\mathbf{C}}$. Moreover, \mathcal{E} is denominated fully homomorphic if it is correct for every algebraic circuit, or, equivalently, if it is correct for every Boolean circuit.

SHE schemes corresponds to homomorphic encryption schemes that are correct for circuits whose multiplicative depth is limited by a certain upper bound, denoted by ℓ .

1.2. Security model

We say that a cryptosystem is secure against *chosen ciphertext attacks* (CCA2) if there is no polynomial time adversary that can win the following game with non-negligible probability.

Setup. The challenger obtains $(sk, pk) = \text{KEYGEN}(\lambda)$, where λ is a security parameter, and sends pk to the adversary \mathcal{A} .

Queries. \mathcal{A} sends ciphertexts to the challenger, before or after the challenge, who returns the corresponding plaintexts.

Challenge. The adversary randomly generates two plaintexts $m_0, m_1 \in \mathcal{M}$ and sends to the challenger, who then randomly chooses a bit $b \in \{0, 1\}$ and computes the ciphertext $c = \text{ENC}_{pk}(m_b)$. The challenger sends c to \mathcal{A} .

Answer. \mathcal{A} sends a bit b' to the challenger and wins the game if $b' = b$.

If we allow queries only before the challenge, we say that the cryptosystem is secure against CCA1 adversaries (lunchtime attacks). Queries can be interpreted as accesses to a *decryption oracle*. If, instead, we only allow access to an *encryption oracle*, namely the adversary can choose any message to be encrypted under the same key pair, then we say that the cryptosystem is secure against *chosen plaintext attacks* (CPA). Clearly, resistance to CCA2 attacks are the most desirable.

In homomorphic encryption, it is impossible to achieve CCA2 security, because the adversary can simply add to the encrypted message some encryption of zero, which can be obtained by querying the encryption oracle, and send it back to the decryption oracle. Many FHE schemes have as public value an encryption of the private key bits, which can be sent to the decryption oracle before the challenge, making such schemes insecure against CCA1 adversaries. Indeed, a *key recovery* attack is stronger than a CCA1 attack;

also, Loftus et al [Loftus et al. 2011] showed that Gentry's construction over ideal lattices is vulnerable to key recovery attacks and presented the only somewhat homomorphic encryption scheme that is known to be CCA1-secure.

In 2015, Dahab, Galbraith and Morais showed that the NTRU-based family of SHE schemes is vulnerable to key recovery attacks [Dahab et al. 2015]. Hence, except for Loftus et al's [Loftus et al. 2011] scheme, no other known SHE proposal achieves CCA1 security.

1.3. Homomorphic verifiable computation

Although homomorphic encryption is a very flexible cryptographic primitive, when applied to the cloud computing scenario, it lacks an important property: the ability to verify if a given homomorphic computation corresponds to what the client desired. A verifiable computation scheme could solve this problem, provided two requirements are met. First, the cloud must not spend much more time to perform the verifiable computation when compared to the non-verifiable solution. Second, the client must be able to verify the result faster than the time it takes to perform the entire computation by himself. There are proposals [Gennaro et al. 2010, Chung et al. 2010] that use homomorphic encryption to construct a verifiable scheme, because it is possible to offer input and output privacy, since both are encrypted. However, the underlying security model does not allow verification queries. Recently, Fiore, Gennaro and Pastro [Fiore et al. 2014] proposed a new construction that does allow verification queries, improving on the security model. They showed how to solve practical problems, such as computing quadratic multivariate polynomials over encrypted data, which can be used to homomorphically compute statistical functions. We remark that this application requires only one level of multiplications, which is an important characteristic to be considered in order to calculate the parameters of the underlying SHE scheme.

Definition 1.3. A verifiable computation scheme \mathcal{VC} is defined by the algorithms (KEYGEN, PROBGEN, COMPUTE, VERIFY), as follows:

Key generation. Algorithm KEYGEN($1^\lambda, f$) generates secret key sk and evaluation key esk .

Problem generation. Using secret key sk , algorithm PROBGEN receives as input ciphertext c_i and computes the corresponding authentication tag σ_i , such that $\sigma_i = \text{AUTH}_{vk}(c_i, (\cdot, i))$.

Verification. Given the secret sk , tag σ and ciphertext c , we have that $\text{VERIFY}_{sk}(\sigma, c)$ returns 1 if $c = f([c_i])$ and $\sigma = \text{AUTH}_{vk}(c, (\cdot, i))$. Otherwise, it returns 0.

Evaluation. Given $\sigma_1, \dots, \sigma_t$ and the description of function f , algorithm $\text{COMPUTE}_{esk}([\sigma_i], \Delta, f)$ returns the authentication tag σ that corresponds to the ciphertext $c = \text{EVAL}_{edk}([c_i], f)$ obtained by running the EVAL algorithm from the underlying homomorphic encryption scheme. We say that the \mathcal{VC} scheme is correct if $\text{VERIFY}_{sk}(\sigma, c)$ outputs 1.

2. Main contributions

Here we list the main contributions of this work:

- the combination of verifiable computation and homomorphic encryption in order to avoid key recovery attacks and achieve CCA1-security. This is relevant because almost every SHE scheme in the literature is vulnerable to CCA1 attacks, which are a feasible scenario in cloud computing;
- the comparison of AGCD-based [van Dijk et al. 2010] and BGV-based [Brakerski et al. 2011] SHE schemes for a specific scenario, where we can homomorphically evaluate quadratic multivariate polynomials. In general, BGV is considered the best choice for the implementation of somewhat homomorphic encryption; in this particular case, however, the AGCD scheme, which is simpler, also offers better performance in practice;
- the proposal of concrete parameters for a specific multiplicative depth to be secure against the attacks described in the literature, based on the utilization of some variation of the LLL algorithm.

References

- Brakerski, Z., Gentry, C., and Vaikuntanathan, V. (2011). Fully homomorphic encryption without bootstrapping. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:111.
- Chung, K., Kalai, Y., and Vadhan, S. (2010). Improved delegation of computation using fully homomorphic encryption. In Rabin, T., editor, *Advances in Cryptology - CRYPTO 2010*, volume 6223 of *Lecture Notes in Computer Science*, pages 483–501. Springer Berlin Heidelberg.
- Dahab, R., Galbraith, S., and Morais, E. (2015). Adaptive key recovery attacks on NTRU-based somewhat homomorphic encryption schemes. In Lehmann, A. and Wolf, S., editors, *Information Theoretic Security*, volume 9063 of *Lecture Notes in Computer Science*, pages 283–296. Springer International Publishing.
- Fiore, D., Gennaro, R., and Pastro, V. (2014). Efficiently verifiable computation on encrypted data. *Cryptology ePrint Archive*, Report 2014/202. <http://eprint.iacr.org/>.
- Gennaro, R., Gentry, C., and Parno, B. (2010). Non-interactive verifiable computing: Outsourcing computation to untrusted workers. In Rabin, T., editor, *Advances in Cryptology - CRYPTO 2010*, volume 6223 of *Lecture Notes in Computer Science*, pages 465–482. Springer Berlin Heidelberg.
- Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 169–178, New York, NY, USA. ACM.
- Loftus, J., May, A., Smart, N. P., and Vercauteren, F. (2011). On CCA-secure somewhat homomorphic encryption. In *In Selected Areas in Cryptography*, pages 55–72.
- van Dijk, M., Gentry, C., Halevi, S., and Vaikuntanathan, V. (2010). Fully homomorphic encryption over the integers. In *Proceedings of the 29th Annual International Conference on Theory and Applications of Cryptographic Techniques, EUROCRYPT'10*, pages 24–43, Berlin, Heidelberg. Springer-Verlag.

Análise de um algoritmo aproximativo para o problema de escalonamento de tarefas com restrições de precedência em máquinas paralelas idênticas

Elton Lever¹, Omar Vilca¹, Rosiane de Freitas¹

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Manaus, AM – Brasil

{elton,omarlatorre,rosiane}@icompu.ufam.edu.br

Abstract. *This paper presents an unit time precedence constrained scheduling, considering a variable number of identical parallel machines, which is NP-Hard. For this problem, the best approximation ratio was provided by Coffman and Graham's algorithm (1972), $2 - \frac{2}{m}$, with $m \geq 3$ parallel machines, which has long been the best ratio until recently an approximation algorithm that guarantees schedules which are at most $2 - \frac{7}{3m+1}$ factor as long as the optimal, was proposed by Gangal and Ranade (2008), providing a small improvement but with interesting theoretical concepts, whose analysis is presented.*

Resumo. *Neste artigo é apresentado o problema de escalonamento de tarefas com tempos de execução unitários e restrições de precedência, considerando um número variável de máquinas paralelas idênticas, que é NP-difícil. Para este problema, a melhor razão de aproximação era fornecida pelo algoritmo de Coffman e Graham (1972), de fator de aproximação $2 - \frac{2}{m}$, para $m \geq 3$ máquinas paralelas idênticas, e que perdurou por um longo tempo até que recentemente um algoritmo de fator de aproximação $2 - \frac{7}{3m+1}$ foi proposto por Gangal e Ranade (2008), proporcionando uma melhoria pequena mas significativa e com conceitos teóricos interessantes, cuja análise é apresentada.*

1. Introdução

O escalonamento de tarefas com tempos de execução unitários e restrições de precedência, considerando máquinas (ou processadores) paralelas idênticas, consiste em um importante problema de otimização combinatória com aspectos de complexidade computacional de grande interesse. Deste modo, dadas $J = \{J_1, J_2, \dots, J_j, \dots, J_n\}$ tarefas de tempos unitários de processamento, $p_j = 1$, a serem escalonadas em $|P|$ máquinas paralelas idênticas, onde as tarefas $|P| = n$ possuem relações de precedência entre si (*prec*), deseja-se escalonar as n tarefas nas m máquinas de modo a minimizar o *makespan*, ou seja, minimizar o tempo de completude (C_j) da última tarefa a ser escalonada. Usando a notação clássica de 3-campos [Graham et al. 1979], tem-se $P_m|prec, p_j = 1|C_{max}$, ou somente $P|prec, p_j = 1|C_{max}$, para o caso onde o número de máquinas for parte da entrada, ou seja, se tiver um número variável de máquinas, provado ser NP-difícil [Ullman 1975]. Se o número de máquinas for fixo, com $|P| = k$, então, o problema ainda está em aberto e corresponde ao oitavo problema da lista de 12 problemas em aberto publicada no livro *Computers and Intractability* [Garey and Johnson 1979], da qual apenas este problema de escalonamento e o problema de isomorfismo em grafos, permanecem em aberto.

O problema de interesse deste trabalho é o de escalonamento em um número variável de máquinas, para o qual existem diversos algoritmos aproximativos, sendo por muito tempo o melhor deles o proposto por Coffman e Graham, de razão de aproximação $2 - \frac{2}{m}$, para $m \geq 3$ máquinas paralelas idênticas [Coffman and Graham 1972]. Recentemente, um algoritmo de fator de aproximação $2 - \frac{7}{3m+1}$ foi proposto [Gangal and Ranade. 2008], proporcionando uma melhoria pequena mas significativa e com conceitos teóricos interessantes. Este algoritmo é analisado no restante deste artigo.

2. Algoritmo aproximativo para $P|prec, p_j = 1|C_{max}$

Nesta seção serão brevemente apresentados os principais algoritmos aproximativos para $P|prec, p_j = 1|C_{max}$. O algoritmo pioneiro foi proposto em 1961 [Hu 1961], e mostrou que resolve o problema $P|prec, p_j = 1|C_{max}$ em tempo polinomial quando as restrições de precedência das tarefas são árvores enraizadas (*in-tree*, *out-tree*, *in-forest* e *out-forest*). Para tais casos, a complexidade do algoritmo é da ordem de $O(|V|\log|V|)$ [Papadimitriou and Yannakakis 1979], onde $|V|$ é o número de vértices, assim o algoritmo oferece um escalonamento ótimo quando $|P| = 2$, porém, para $|P| \geq 3$ o algoritmo dista $2 - \frac{1}{|P|-1}$ do ótimo [Feng 1975]. Em 1972, o algoritmo de Hu foi refinado, ficando conhecido como algoritmo CG [Coffman and Graham 1972], cuja complexidade ficou $O(|V| + |E|)$. Lam & Sethi mostraram que a razão de aproximação do algoritmo de CG quando $|P| \geq 3$ é de $2 - \frac{2}{|P|}$, onde $|P|$ é o número de processadores. Em 1994, foi corrigido um erro na análise de Lam & Sethi e deram com melhoria no limite inferior gerado [Braschi and Trystram 1994]. Trinta e cinco anos mais tarde, em 2008, um algoritmo com melhor aproximação do ótimo que todos os anteriores foi proposto [Gangal and Ranade. 2008] e será apresentado seguir.

2.1. Algoritmo aproximativo de Gangal & Ranade(GR)

O algoritmo de melhor razão de aproximação para o problema $P|prec, p_j = 1|C_{max}$ foi proposto em 2008 [Gangal and Ranade. 2008], onde foi estabelecida uma razão de aproximação de $2 - \frac{7}{3|P|+1}$, onde $|P|$ é a quantidade de processadores para o caso de $|P| > 3$, portanto, melhor que os seus antecessores.

Os passos detalhados do algoritmo são apresentados a seguir (Algoritmo 1). Cada vértice folha será calculado pelo caminho do comprimento mais longo no grafo. N é a quantidade de sucessores de v no caminho mais longo até suas folhas. L é o número de arcos no caminho de v até u , $|P|$ são os processadores e D o maior prazo do caminho. As colunas representam o tamanho do tempo no escalonamento. A equação (1) é utilizada para calcular os prazos $D(v)$ para cada vértice v no sentido inverso da ordem de classificação na ordenação topológica, começando pelas folhas das árvores em que $N(v, D, L)$ é o maior conjunto de vértices (v, u) e, será calculado o menor prazo, tendo um caminho de comprimento pelo menos $L + 1$, conectando v até u com $D(v) \leq D$, onde D é o prazo do caminho mais longo no grafo direcionado. O alcance máximo de valores inteiros que $L, D, |N(v, D, L)|$ pode tomar é $|V|$, contudo, a complexidade de tempo do algoritmo é $|V|^4$ para calcular todos prazos.

$$D(v) = D - L - \lceil \frac{N(v, D, L)}{P} \rceil \quad (1)$$

Algoritmo 1: Algoritmo - (Gangal & Ranade)

Entrada: Grafo $G(V, E)$, P : Conjunto de processadores.

inicio

1 - Inclua um vértice a com arestas $\forall v \in V(G)$

para cada vértice $v \in V(G)$ **hacer**

se v é vértice sem sucessores **então**

 Seja $D(v) = \Delta$, onde Δ é um número qualquer, $D(v)$ = comprimento do caminho mais longo em G ;

senão

 Calcule os prazos $D(v)$ para os outros vértices de acordo com a Equação (1);

fim se

fin para cada

2 - Ordene os vértices em ordem não decrescente de acordo com seu prazo, realizando a redução de transitividade ;

3 - Escalone a no tempo 1 do processador 1;

para cada vértice $v \in V(G)$ **hacer**

enquanto Reorganiza predecessores (v) **faça**

 1. Seja t' uma coluna com maior prazo, tal que contém alguns predecessores de v ;

 2. Seja A o conjunto de vértices na coluna de $t' - 1$ e coluna t' ;

se todos os vértices em A tiver o mesmo prazo, e A tem no máximo $|P|$ predecessores de v , e coluna t' tem menos de $|P|$ vértices **então**

 - Mova os predecessores de v para a posição $t' - 1$;

 - Mova outros vértices da coluna $t' - 1$ para a coluna t' se for necessário;

fim se

fim enqto

 Seja t o menor tempo em que v pode ser escalonado;

 Escalone v na coluna t livre do primeiro processador disponível.

fin para cada

fin

Primeiramente, um vértice artificial a é criado e conectado com todos os outros vértices, e se estima prazos $D(v)$ para todos os vértices v , que define o comprimento do caminho mais longo no grafo (G). Os prazos dos outros vértices são, então, calculados utilizando-se a equação (1). Depois de calculado os prazos, o algoritmo realiza uma redução de transitividade no grafo. Tal processo, com a reorganização dos predecessores, onde os vértices que estão na coluna $t' - 1$, com t' tendo algum prazo, não afeta as restrições de precedências, pois as tarefas estão prontas para serem escalonadas na quantidade de processadores disponíveis naquele momento. Se todos os antecessores de v se moverem para a coluna de $t' - 1$, em seguida, v é escalonado na coluna t' . O benefício deste rearranjo é exemplificado na Figura 1, onde é demonstrado que a reorganização dos predecessores no primeiro bloco não é possível, no entanto, no segundo bloco da instância do exemplo percebe-se que o tempo é quebrado, e isto ocorre devido a reorganização das tarefas seguintes. A tarefa j que iniciava no tempo 7 foi para o tempo 6. Por consequência, a tarefa g que estava iniciando no tempo 6 agora inicia no tempo 7. Com isto, as tarefas sucessoras de j passam a ser $\{n,p,o\}$ e, conseqüentemente, z diminui um período de tempo, para o início da execução. As tarefas l e n podem iniciar no tempo 7, pois, a quantidade de tarefas é menor que de processadores. Como a tarefa i é uma folha e tem baixa prioridade, inicia no tempo 8. A tarefa p vai para o tempo 7 e, conseqüentemente, z vai para o tempo 8 (terminando assim o escalonamento). Pode-se perceber claramente

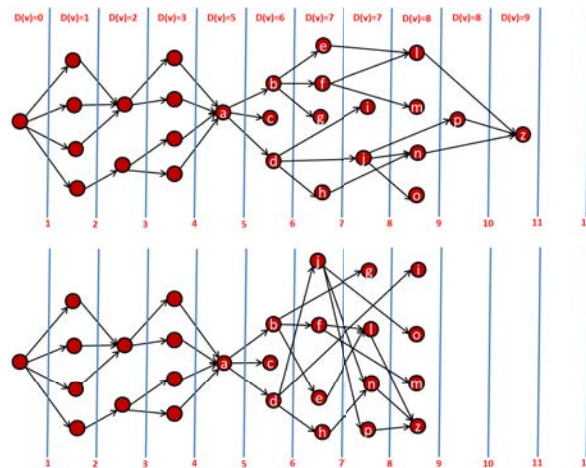


Figura 1. Alg. GR, com reorganização dos predecessores para $|P| = 4$.

que após tal reorganização, o tempo diminui de 11 para 9, mas 9 é o comprimento do caminho mais longo no grafo e, portanto, é um limite inferior. Assim a Figura 1 mostra o escalonamento ótimo para este grafo.

3. Considerações Finais

Neste trabalho foi abordado o algoritmo de melhor razão de aproximação para $P|prec, p_j = 1|C_{max}$, sendo de fator $2 - \frac{7}{3|P|+1}$, considerando $|P| > 3$. É ótimo para 2 máquinas e, para 3, mantém o fator de $2 - \frac{2}{m}$, dado pelo algoritmo de CG.

Referências

- Braschi, B. and Trystram, D. (1994). A new insight into the coffman-graham algorithm. *SIAM J. Comput.*, 23:662–669.
- Coffman, E. and Graham, R. (1972). *Optimal Scheduling of coffman-graham algorithm for a new order class*. Acta informatica, 1:200-213.
- Feng, T., editor (1975). *Parallel Processing, Proceedings of the Sagamore Computer Conference, August 20-23, 1974*, volume 24 of *Lecture Notes in Computer Science*. Springer.
- Gangal, D. and Ranade., A. (2008). Precedence constrained scheduling in $(2 - \frac{7}{3p+1})$ para $p \geq 4$ optimal. *Elsevier. Journal of Computer and System Sciences*, 74 (2008) 1139-1146., Vol:1:1–13.
- Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.
- Graham, R. L., Lawler, E. L., Lenstra, J. K., and Kan, A. H. G. R. (1979). Optimization and approximation in deterministic sequencing and scheduling: a survey. *Annals of Discrete Mathematics*.
- Hu, T. C. (1961). Parallel sequencing and assembly line problems. *IBM Research Center, Yorktown, New York*, 1(1):8.
- Papadimitriou, C. and Yannakakis, M. (1979). *Scheduling interval-ordered tasks*. *SIAM J. Comput.*, 8(3): 405-409.
- Ullman, J. (1975). Np-complete scheduling problems. *Journal of Computer and System Sciences*, 10(3):384 – 393.

Some families of 0-rotatable graceful caterpillars

Atílio G. Luiz¹, C. N. Campos¹, R. Bruce Richter²

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas – São Paulo – Brasil

²Department of Combinatorics and Optimization – University of Waterloo
Waterloo – Ontario – Canada

gomes.atilio@gmail.com, campos@ic.unicamp.br, brichter@uwaterloo.ca

Abstract. A graceful labelling of a tree T is an injective function $f: V(T) \rightarrow \{0, 1, \dots, |E(T)|\}$ such that $\{|f(u) - f(v)|: uv \in E(T)\} = \{1, 2, \dots, |E(T)|\}$. A tree T is said to be 0-rotatable if, for any $v \in V(T)$, there exists a graceful labelling f of T such that $f(v) = 0$. In this work, it is proved that the following families of caterpillars are 0-rotatable: caterpillars with perfect matching; caterpillars obtained by identifying a central vertex of a path P_n with a vertex of K_2 ; caterpillars obtained by identifying one leaf of the star $K_{1,s-1}$ to a leaf of P_n , with $n \geq 4$ and $s \geq \lceil \frac{n-1}{2} \rceil$; caterpillars with diameter five or six; and some families of caterpillars with diameter at least seven. This result reinforces the conjecture that all caterpillars with diameter at least five are 0-rotatable.

1. Introduction

Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and edge set $E(G)$. A graceful labelling of G is an injection $f: V(G) \rightarrow \{0, 1, \dots, |E(G)|\}$ such that $\{|f(u) - f(v)|: uv \in E(G)\} = \{1, 2, \dots, |E(G)|\}$. We say that G is *graceful* if it has a graceful labelling.

In 1967, Rosa introduced four types of labellings of graphs, including graceful labellings, and posed the *Graceful Tree Conjecture* which states that all trees are graceful [Rosa 1967]. Rosa proved that the Graceful Tree Conjecture is a strengthened version of the well-known *Ringel-Kotzig Conjecture* which states that the complete graph K_{2m+1} has a cyclic decomposition into subgraphs isomorphic to a given tree T with m edges. The Graceful Tree Conjecture is a very important open problem in Graph Theory, with more than a thousand papers about it [Gallian 2015].

As soon as one starts investigating graceful labellings of trees, it becomes clear the importance of knowing how to construct graceful labellings with the label 0 appearing in a given vertex. The importance of label 0 in a graceful labelling of a tree T is due to the fact that it is easy to grow T by adding k new leaves to the 0-labelled vertex and expand the graceful labelling by assigning labels $|E(T)| + 1, \dots, |E(T)| + k$ to these new leaves. A tree T is 0-rotatable if, for any $v \in V(T)$, there exists a graceful labelling f of T such that $f(v) = 0$.

The importance of 0-rotatability of trees was first noted by Rosa in his seminal paper [Rosa 1967], in which the author stated, without proof, that all paths are 0-rotatable. Ten years later, the author published a proof of this result [Rosa 1977]. Meanwhile, in 1969, some examples of non 0-rotatable trees were discovered [Gallian 2015]. As an

example, the smallest non 0-rotatable tree is the tree obtained by identifying one leaf of the star $K_{1,3}$ to a leaf of P_3 . Posteriorly, Chung and Hwang investigated the 0-rotatability of a product of trees called Δ -construction and proved that if two trees T_1 and T_2 are 0-rotatable, then their product $T_1\Delta T_2$ is also 0-rotatable [Chung and Hwang 1981]. In 2004, Bussel [Bussel 2004] showed that all trees with diameter at most three are 0-rotatable. The author also showed that there exist non 0-rotatable trees with diameter four. In fact, he completely characterized the diameter-four non 0-rotatable trees using the following result.

Theorem 1 ([Bussel 2004]). *Let T be a tree of diameter four such that its center v has degree two. Let v_1, v_2 be the vertices adjacent to v and m_1, m_2 be the number of leaves adjacent to v_1, v_2 , respectively. Assume $m_1 \geq m_2$. The tree T has a graceful labelling f with $f(v) = 0$ if and only if there exist integers x and r such that $m_1 = (m_2 + 2 - x)(r - 1) - x$, with: (i) x, r not both odd; (ii) $2 \leq r \leq |E(T)|/2$; and (iii) $0 \leq x \leq \min\{r - 1, m_2\}$. \square*

Let \mathcal{D} denote the class of diameter-four trees whose center has degree two and that do not satisfy the conditions of Theorem 1. Let \mathcal{D}' be the class of trees built by identifying a leaf of an arbitrary path P_n , $n \geq 1$, with the center of a tree in \mathcal{D} . Bussel proved that, given a tree T with diameter four, T is 0-rotatable if and only if $T \notin \mathcal{D}'$. Additionally, he showed that all trees with at most 14 vertices and that are not 0-rotatable belong to the class \mathcal{D}' . Thus, based on these results, the author posed the following conjecture.

Conjecture 2 ([Bussel 2004]). *The class \mathcal{D}' contains all non 0-rotatable trees.*

From the time it was first studied, 50 years ago, 0-rotatability of trees has been considered a possible way to approach the Graceful Tree Conjecture, and also a challenging problem by itself. In particular, a family of trees for which the 0-rotatability property is not known is the family of caterpillars, defined as follows. A tree T is a *caterpillar* if either T is a path or the subgraph obtained by deleting all its leaves (the *base* of T) is a path.

In fact, note that, if Conjecture 2 is true, then it implies that every caterpillar with diameter at least five is 0-rotatable. Considering these observations, in this work, we investigate Conjecture 2 restricted to caterpillars and prove that the following families of caterpillars are 0-rotatable: (i) caterpillars with perfect matching; (ii) caterpillars obtained by identifying a central vertex of P_n with a vertex of K_2 ; (iii) caterpillars obtained by identifying one leaf of $K_{1,s-1}$ to one leaf of P_n , with $n \geq 4$ and $s \geq \lceil \frac{n-1}{2} \rceil$; (iv) caterpillars with diameter five or six; and (v) some families of caterpillars with diameter at least seven. These results reinforce Conjecture 2.

2. Preliminaries

A *matching* of a graph G is a set of pairwise nonadjacent edges of G . Let M be a matching of a graph G . A vertex $v \in V(G)$ is *saturated* by M if v is incident with an edge of M . A *perfect matching* of G is a matching that saturates all the vertices of G . Let T be a tree with a perfect matching M . The *contree* of T is the tree T' obtained from T by contracting all the edges of M .

Broersma and Hoede [Broersma and Hoede 1999] introduced the concept of strongly graceful labellings of trees defined as follows. Let T be a tree with a perfect

matching M . A labelling f of T is *strongly graceful* if f is a graceful labelling and if $f(u) + f(v) = |E(T)|$ for every edge $uv \in M$. The authors proved that the Graceful Tree Conjecture is true if and only if every tree with a perfect matching has a strongly graceful labelling. They also studied the label 0 in strong graceful labellings, as presented in the next lemma. This result is important for the proof of Theorem 5.

Lemma 3 ([Broersma and Hoede 1999]). *Let T be a tree with a perfect matching M and $uv \in M$, $u, v \in V(T)$. Let T' be the contree of T and let $x \in V(T')$ be the vertex corresponding to the edge uv . If T' has a graceful labelling f' , with $f'(x) = 0$, then T has two strongly graceful labellings f_1 and f_2 , such that: (i) $f_1(u) = 0$ and $f_1(v) = |E(T)|$; (ii) $f_2(u) = |E(T)|$ and $f_2(v) = 0$. \square*

Given a graceful labelling f of a tree T , the *complementary labelling* of f is the labelling \bar{f} defined by $\bar{f}(v) = |E(T)| - f(v)$ for each $v \in V(T)$. Note that the complementary labelling is also a graceful labelling since: (i) $f(v)$ is an injection from $V(T)$ to $\{0, \dots, |E(T)|\}$; and (ii) for each $uv \in E(T)$, $|\bar{f}(u) - \bar{f}(v)| = |(m - f(u)) - (m - f(v))| = |f(v) - f(u)|$.

A technique used in our proofs is the method of transfers, defined as follows. Let u, v, u_1 be distinct vertices of a tree T , such that u_1 is adjacent to u . We call *transfer*, the operation of deleting the edge u_1u from T and adding the edge u_1v . After the transfer operation, we say that u_1 has been *moved* from u to v . The following lemma determines when a transfer performed over a graceful tree generates another graceful tree.

Lemma 4 ([Hrnčiar and Haviar 2001]). *Let f be a graceful labelling of a tree T and let $u, v \in V(T)$ be two distinct vertices. If u is adjacent to leaves $u_1, u_2 \in V(T)$, such that $u_1 \neq v$, $u_2 \neq v$ and $f(u_1) + f(u_2) = f(u) + f(v)$, then the tree T' obtained by moving u_1, u_2 from u to v is also graceful. \square*

3. Results

In this section, we state our main results. In particular, Theorems 7 and 8, and the second family stated in Theorem 6 show that, for each integer $d \geq 5$, there exist 0-rotatable caterpillars with diameter d and an arbitrarily large number of vertices. These results reinforce the conjecture that all caterpillars with diameter at least five are 0-rotatable.

Theorem 5. *Every caterpillar with a perfect matching is 0-rotatable.*

Proof. Let T be a caterpillar with a perfect matching M and let $uv \in M$, $u, v \in V(T)$. Let T' be the contree of T and let $x \in V(T')$ be the vertex corresponding to the edge uv . Since T has a perfect matching, we have that T' is a path. Rosa proved that every path is 0-rotatable [Rosa 1977]. Therefore, T' is 0-rotatable. Hence, T has a graceful labelling f' such that $f'(x) = 0$. By Lemma 3, T has two strongly graceful labellings f_1 and f_2 such that: $f_1(u) = 0$ and $f_1(v) = |E(T)|$; $f_2(u) = |E(T)|$ and $f_2(v) = 0$. Therefore, there exist strongly graceful labellings of T which assign the label 0 to vertex u or v . Since uv is arbitrary, we conclude that T is 0-rotatable. \square

Theorem 6. *The following families of caterpillars are 0-rotatable: (i) caterpillars obtained by identifying a vertex of K_2 with a central vertex of P_n ; (ii) caterpillars obtained by identifying one leaf of the star $K_{1,s-1}$ to a leaf of P_n , with $n \geq 4$ and $s \geq \lceil \frac{n-1}{2} \rceil$.*

Outline of the proof. Let T be a caterpillar as defined in the hypothesis and let $v \in V(T)$ be an arbitrary vertex. First, we specify an edge $wz \in E(T)$, $w, z \in V(T)$, and remove

wz from T , thus obtaining two vertex-disjoint subgraphs $H_1 \subset T$ and $H_2 \subset T$ such that $w, v \in V(H_1)$ and $z \in V(H_2)$. Consider a bipartition $\{V_1, V_2\}$ of $V(H_1)$ such that $v \in V_1$ and define $k = |V_1|$. Thus, we construct injective labellings f_1, f_2 for H_1, H_2 , respectively, where $f_1: V(H_1) \rightarrow \{0, \dots, k-1\} \cup \{k + |E(H_2)| + 1, \dots, |E(T)|\}$, $f_2: V(H_2) \rightarrow \{k, k+1, \dots, k+|E(H_2)|\}$, and such that: (i) $f_1(v) = 0$; (ii) the edge labels induced by f_2 are $1, 2, \dots, |E(H_2)|$; (iii) the edge labels induced by f_1 are $|E(H_2)| + 2, \dots, |E(T)|$; and (iv) $f_1(w)$ and $f_2(z)$ are such that $|f_1(w) - f_2(z)| = |E(H_2)| + 1$. Finally, we define a labelling f of T as follows: for $u \in V(T)$, $f(u) = f_1(u)$ if $u \in H_1$; and $f(u) = f_2(u)$ if $u \in H_2$. Therefore, f is a graceful labelling of T with $f(v) = 0$ and, since v is an arbitrary vertex, we obtain that T is 0-rotatable. \square

Theorem 7. *If T is a caterpillar with diameter five or six, then T is 0-rotatable.*

Outline of the proof. Let T be a caterpillar with diameter five or six. For each vertex $v \in V(T)$ in the base of T , we construct a graceful labelling f of T that assigns label 0 to v and assigns label $|E(T)|$ to any leaf $u \in V(T)$ adjacent to v . Consequently, given any of these graceful labellings f , one can use its complementary labelling \bar{f} in order to obtain $\bar{f}(u) = 0$ and $\bar{f}(v) = |E(T)|$. Since \bar{f} is also a graceful labelling and f was constructed considering an arbitrary vertex v of the base of T , we obtain that T is 0-rotatable. \square

Theorem 8. *If T has odd diameter at least seven and each vertex of its base is adjacent to a positive even number of leaves, then T is 0-rotatable. Additionally, If T has even diameter at least eight and each vertex of its base is adjacent to an even number of at least 4 leaves, then T is 0-rotatable.*

Outline of the proof. The technique used in this proof is similar to the technique used in the proof of Theorem 7. \square

4. Acknowledgments

This work was funded by FAPESP, grants 2014/16987-1, 2014/16861-8, 2015/03372-1, and NSERC grant 41705-2014 057082.

References

- Broersma, A. J. and Hoede, C. (1999). Another equivalent of the graceful tree conjecture. *Ars Combinatoria*, 51:183–192.
- Bussel, F. V. (2004). 0-Centred and 0-ubiquitously graceful trees. *Discrete Mathematics*, 277:193–218.
- Chung, F. R. K. and Hwang, F. K. (1981). Rotatable graceful graphs. *Ars Combinatoria*, 11:239–250.
- Gallian, J. A. (2015). A dynamic survey of graph labeling. *The Electronic Journal of Combinatorics*, DS6, 1–389.
- Hrnčiar, P. and Haviar, A. (2001). All trees of diameter five are graceful. *Discrete Mathematics*, 233:133–150.
- Rosa, A. (1967). On certain valuations of the vertices of a graph. *Theory of Graphs (Internat. Sympos., Rome, 1966) Gordon and Breach, New York; Dunod, Paris*, 349–355.
- Rosa, A. (1977). Labeling snakes. *Ars Combinatoria*, 3:67–73.

Computação sobre dados cifrados em *GPGPUs*

Pedro Geraldo M. R. Alves , Diego F. Aranha

Instituto de Computação – Universidade Estadual de Campinas (Unicamp)
Cidade Universitária Zeferino Vaz – CEP 13083-970 – Campinas – SP – Brazil

{pedro.alves,dfaranha}@ic.unicamp.br

Abstract. *Under the dominant cloud computing paradigm, employing encryption for data storage and transport may not be enough. Security guarantees should also be extended to data processing. Homomorphic encryption schemes are natural candidates for computation over encrypted data since they are able to satisfy the requirements imposed by the cloud environment. This work presents CUYASHE as a GPGPU implementation of the leveled fully homomorphic scheme YASHE. It employs CUDA, the Chinese Remainder Theorem and the Fast Fourier Transform to obtain significant performance improvements. In particular, there was a speedup between 6 and 35 times for homomorphic multiplication.*

Resumo. *No contexto da computação na nuvem, a aplicação de métodos criptográficos exclusivamente no armazenamento e transporte dos dados não é suficiente, uma vez que precisam ser revelados ao serviço para ocorrer processamento. Esquemas de cifração homomórfica são candidatos naturais para computação sobre dados cifrados, o que os torna capazes de satisfazer esse novo requisito de segurança. Este trabalho apresenta a CUYASHE, uma implementação em GPGPUs do criptosistema homomórfico YASHE. A CUYASHE emprega CUDA, o Teorema Chinês do Resto e a Transformada Rápida de Fourier para obter ganho de desempenho sobre o estado da arte. Em especial, destaca-se uma redução de 6 até 35 vezes no tempo de execução da operação de multiplicação.*

1. Introdução

A computação em nuvem tem sido responsável por uma profunda mudança na comunidade de processamento distribuído. A possibilidade de terceirizar a instalação, manutenção e a escalabilidade de servidores, somada a preços competitivos, faz com que esses serviços se tornem altamente atraentes.

Diversos esquemas criptográficos são utilizados como padrão para o armazenamento e transferência de dados. Contudo, no caso de serviços na nuvem existe a possibilidade de se lidar com um adversário que não apenas tenha acesso aos dados como também ao *hardware* que realiza seu processamento. Desse modo, é necessário que se estabeleça uma trajetória completamente sigilosa para os dados durante o transporte, armazenamento e processamento para que seja preservada a privacidade.

A criptografia homomórfica se mostra promissora para satisfazer esse novo requisito de segurança. Os esquemas dessa classe permitem que o processamento seja feito sobre criptogramas mesmo sem o conhecimento das chaves de cifração ou decifração. Assim, não há razão para que dados sejam revelados no momento do processamento.

O objetivo deste trabalho consistiu em obter ganho de desempenho no estado da arte do criptosistema homomórfico em nível YASHE [Bos et al. 2013]. Para isso, foram aplicadas técnicas de computação paralela em *GPGPUs* por meio da arquitetura CUDA; utilizou-se

o Teorema Chinês do Resto para simplificar a manipulação de inteiros grandes; e a Transformada Rápida de Fourier foi aplicada para reduzir a complexidade computacional das operações de multiplicação polinomial utilizadas por este criptossistema.

2. YASHE - *Yet Another Somewhat Homomorphic Encryption*

A classe de criptossistemas homomórficos é caracterizada por esquemas que permitem a aplicação de operações de adição ou multiplicação sobre seus criptogramas. O resultado dessas operações gera um novo criptograma que, quando decifrado, se apresenta de forma equivalente ao que seria esperado em uma operação sobre texto claro.

YASHE é um criptossistema completamente homomórfico em nível baseado em reticulados ideais [Bos et al. 2013]. Sua classificação significa que suporta uma quantidade limitada de operações de adição e multiplicação. Além disso, é um esquema probabilístico e tem sua segurança baseada nos problemas *RLWE* e *DSPR*. Ainda, o YASHE atinge o padrão de segurança *IND-CPA*, o que significa que seus criptogramas são resistentes a um adversário que seja capaz de cifrar texto claro.

3. Metodologia

Como forma de reduzir o tempo de execução das operações do YASHE, este trabalho se focou em paralelizar a implementação das operações polinomiais nas quais o esquema se baseia. Isso foi feito por meio da plataforma CUDA, escolhida principalmente devido a seu excepcional desempenho na aplicação de *GPGPUs* na resolução de problemas baseados em dados.

O Teorema Chinês do Resto, ou *CRT*, foi utilizado como ferramenta para simplificar a manipulação dos coeficientes polinomiais [Ding et al. 1996]. Sua função é mapear um polinômio com coeficientes inteiros arbitrariamente grandes em diversos polinômios com coeficientes inteiros arbitrariamente pequenos chamados polinômios residuais, ou simplesmente resíduos. Essa substituição permite que os coeficientes sejam manipulados utilizando uma aritmética mais simples e suportada nativamente pela arquitetura.

As operações do YASHE dependem de adições e multiplicações polinomiais. Enquanto aquelas tem complexidade linear e são trivialmente paralelizáveis, estas são consideravelmente mais complexas e requerem uma estratégia para a redução de sua complexidade computacional. Dessa forma, este trabalho estudou a aplicação da Transformada Rápida de Fourier, ou *FFT*, para a implementação dessa operação [Cooley and Tukey 1965]. Em seu domínio, a multiplicação polinomial é aplicada como uma simples multiplicação coeficiente-a-coeficiente. Assim, sua complexidade torna-se $\Theta(n \log n)$ com o custo da aplicação da transformada.

4. Implementação

A biblioteca *CUYASHE* foi desenvolvida durante a execução deste trabalho. A implementação consolida a metodologia descrita e permite a avaliação dos ganhos com a sua aplicação. Seu código está disponível ao público sob uma licença GNU GPLv3 [Alves and Aranha 2016]. Apesar do foco na execução em *GPUs*, algumas operações sobre inteiros grandes são realizadas pela *CPU*, como por exemplo a formatação e a cópia dos operandos entre as memórias. Essas operações são realizadas por meio da biblioteca *NTL* [Shoup 2003].

A aplicação do *CRT* nos operandos da *CUYASHE* é feita completamente na *GPU*. Dessa forma, uma vez que os resíduos ainda não foram calculados, o suporte a aritmética de inteiros grandes torna-se necessário. A biblioteca *RELIC* [Aranha and Gouvêa 2016] foi utilizada como base e as rotinas requeridas (adição, subtração, multiplicação, divisão e resto modular) foram adaptadas para CUDA.

A implementação da *FFT* foi realizada sobre a biblioteca *CUFFT* [NVIDIA 2015]. Para isso, precisou-se aplicar operações de conversão dos coeficientes entre o conjunto dos inteiros e o corpo dos complexos. Além disso, aritmética de ponto-flutuante utilizada por esta biblioteca gerou erros de precisão que precisaram ser mitigados para o funcionamento correto do algoritmo.

Para maximizar os ganhos com o uso das estratégias descritas, a *CUYASHE* trabalha com resíduos dentro do domínio da transformada. Ou seja, aplica-se o *CRT* seguido da *FFT* em cada um dos polinômios residuais. Assim, as operações polinomiais de adição e multiplicação tem aplicação coeficiente-a-coeficiente, o que implica em complexidade linear. Além disso, os resíduos são mantidos exclusivamente na memória da *GPU*, dispensando os custos envolvidos com a cópia de dados entre as memórias.

5. Resultados

Com o intuito de avaliar a qualidade da implementação e das estratégias propostas, as operações da *CUYASHE* foram comparadas com três trabalhos no estado da arte: [Bos et al. 2013], *BLLN*; [Lepoint and Naehrig 2014], *LN*; [Dowlin et al. 2015], *SEAL*. Foram utilizados os parâmetros de configuração do *YASHE* propostos por *BLLN* [Bos et al. 2013] e *LN* [Lepoint and Naehrig 2014] como padrão para as comparações de tempo de execução ¹. Os autores argumentam que esses parâmetros oferecem segurança de 80 *bits*.

Para as implementações disponíveis à comunidade, *LN* e *SEAL*, realizou-se medições dos tempos de execução das principais operações do *YASHE*: cifração, decifração, adição e multiplicação homomórfica. Os resultados apresentados são os tempos médios calculados a partir de 100 execuções isoladas de cada operação. A máquina utilizada possui uma *CPU Intel Xeon E5-2630 @ 2,6GHz* com uma *GPU NVIDIA GeForce GTX TITAN Black @ 0,98 GHz*. Além disso, utiliza-se a versão 7.5 do kit de desenvolvimento *CUDA* e as versões 4.8.4 dos compiladores *gcc* e *g++*. Por fim, empregou-se a biblioteca *NTL 9.1.0* compilada sobre a *GMP 6.0.0*. A comparação com a *BLLN* é feita por meio dos resultados oferecidos pelos autores, uma vez que seu código não foi disponibilizado.

Tabela 1. Comparação entre a *CUYASHE* e as implementações *LN*, *SEAL* e *BLLN*. Os valores para as três primeiras foram medidos na mesma máquina, enquanto a última teve seus tempos fornecidos pelos autores. Os tempos para multiplicação homomórfica incluem o custo de relinearização.

Operação	<i>CUYASHE</i> (ms)	<i>LN</i> (ms)	<i>SEAL</i> (ms)	<i>BLLN</i> (ms)
Cifração	1,85	15,4	34,93	27
Decifração	2,01	13,71	34,1	5
Adição Homomórfica	0,02	0,59	0,18	0,02
Multiplicação Homomórfica	5,49	31,07	194,94	31

Como pode ser visto na Tabela 1, as operações de cifração, decifração e multiplicação homomórfica da *CUYASHE* apresentaram tempos 8, 7 e 9 vezes menores do que o trabalho *LN* e 15, 6 e 2,5 vezes menores do que o *BLLN*, respectivamente. Sobre a *SEAL* os ganhos foram ainda mais expressivos, atingindo 19, 17 e 35 vezes, respectivamente.

A principal motivação para a utilização de um criptossistema homomórfico é poder operar sobre criptogramas. Dessa forma, apesar de ganhos na cifração e decifração serem importantes, operações homomórficas são o alvo principal para otimizações. Assim, os expressivos ganhos

¹ $R = \mathbb{Z}[X] / (x^{4096} + 1)$, $q = 2^{127} - 1$, $w = 2^{32}$, $t = 2^{10}$.

de velocidade apresentados sobre o estado da arte tem grande importância do ponto de vista de viabilizar aplicações práticas.

Por fim, este documento apresenta aperfeiçoamentos na implementação em relação a resultados preliminares [Alves and Aranha 2015]. A revisão na máquina de estados da CUYASHE permitiu a implementação da aritmética com operandos não apenas mapeados em resíduos pelo CRT como também interpolados para o domínio da FFT. Assim, a complexidade computacional dessas operações se tornou linear e evitou-se o custo de aplicação dessa transformada. Além disso, toda a aritmética passou a ser computada na GPU. Restando à CPU apenas o gerenciamento das operações.

6. Conclusão

Este trabalho investigou estratégias para a implementação do criptossistema YASHE em GPGPUs por meio da plataforma CUDA. Com os resultados obtidos a biblioteca CUYASHE foi desenvolvida, otimizada e disponibilizada à comunidade [Alves and Aranha 2016].

Aplicou-se o CRT para simplificar a manipulação de inteiros grandes na GPU e a FFT para a redução da complexidade de operações de multiplicação polinomial, nas quais o YASHE é altamente dependente.

A comparação com outros trabalhos da literatura demonstrou uma redução expressiva nos tempos de execução desse criptossistema, o que sugere que a metodologia proposta foi adequada ao contexto. Como trabalho futuro, espera-se a aplicação da CUYASHE em uma implementação de um protocolo que preserve privacidade.

Referências

- [Alves and Aranha 2015] Alves, P. and Aranha, D. (2015). cuYASHE: Computação sobre dados cifrados em GPGPUs. In *XV Simpósio Brasileiro de Segurança da Informação e Sistemas Computacionais (SBSeg 2015)*, pages 55–60.
- [Alves and Aranha 2016] Alves, P. and Aranha, D. (2016). cuYASHE. <https://github.com/cuyashe-library/cuyashe>. Acessado pela última vez: 19/05/2016.
- [Aranha and Gouvêa 2016] Aranha, D. F. and Gouvêa, C. P. L. (2016). RELIC is an Efficient Library for Cryptography. <https://github.com/relic-toolkit/relic>.
- [Bos et al. 2013] Bos, J., Lauter, K., Loftus, J., and Naehrig, M. (2013). Improved Security for a Ring-Based Fully Homomorphic Encryption Scheme. Springer Berlin Heidelberg.
- [Cooley and Tukey 1965] Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301.
- [Ding et al. 1996] Ding, C., Pei, D., and Salomaa, A. (1996). *Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography*. World Scientific Publishing Co., Inc.
- [Dowlin et al. 2015] Dowlin, N., Gilad-Bachrach, R., Laine, K., Lauter, K., Naehrig, M., and Wernsing, J. (2015). Manual for Using Homomorphic Encryption for Bioinformatics.
- [Lepoint and Naehrig 2014] Lepoint, T. and Naehrig, M. (2014). A Comparison of the Homomorphic Encryption Schemes FV and YASHE. Springer International Publishing.
- [NVIDIA 2015] NVIDIA (2015). CUDA Toolkit Documentation. <http://docs.nvidia.com/cuda/cufft/>. Acessado pela última vez: 12/08/2015.
- [Shoup 2003] Shoup, V. (2003). NTL: A library for doing number theory. <http://www.shoup.net/ntl>. Acessado pela última vez: 05/03/2016.

Algoritmos Branch-and-Price para o Problema de Empacotamento em Recipientes com Restrições de Classe*

Yulle Glebbyo¹, Flávio K. Miyazawa¹, Rafael C. S. Schouery¹, Eduardo C. Xavier¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251 – 13083-852 – Campinas – SP – Brazil

glebbyo@gmail.com, {fkm,schouery,eduardo}@ic.unicamp.br

Abstract. *In the Class Constrained Binpacking Problem, a set of items of various sizes and classes must be packed in bins, each with a capacity B and a limit C on the amount of items from different classes used, in a way that minimizes the number of bins used. We present Branch-and-Price algorithms for this problem, as well as algorithms for the Class Constrained Knapsack Problem, which appears as a pricing subproblem. In the experiments, the best algorithm proposed solved 93% of the instances in less than 30 seconds.*

Resumo. *No problema de Empacotamento em Recipientes Com Restrições de Classe, um conjunto de itens de variados tamanhos e classes deve ser empacotado em recipientes, cada um com capacidade B e limite C na quantidade de classes diferentes acomodadas, minimizando o número de recipientes usados. Apresentamos algoritmos de Branch-and-Price para este problema, assim como algoritmos para o Problema da Mochila Com Restrições de Classe, que aparece como subproblema de pricing. Nos experimentos executados, o melhor algoritmo proposto resolveu 93% das instâncias em menos de 30 segundos.*

1. Introdução

Problemas de empacotamento são frequentemente estudados na ciência da computação devido sua alta aplicabilidade na indústria. O problema de empacotamento em recipientes clássico foi abordado de diversas formas, incluindo algumas técnicas de algoritmos exatos. Uma variante do problema clássico, onde itens possuem classes, também foi estudada, porém os resultados obtidos se focaram em algoritmos de aproximação (e.g. [Shachnai and Tamir 2001], [Xavier and Miyazawa 2008] e [Epstein et al. 2010]). Neste trabalho, propomos atacar o Problema de Empacotamento em Recipientes com Restrições de Classe de maneira exata utilizando técnicas que se provaram eficazes para o problema clássico, como o *Branch-and-Price* [Vance et al. 1994], que utiliza do Problema da Mochila com Restrições de Classe como subproblema.

2. Problema de Empacotamento em Recipientes com Restrições de Classe

Dado um conjunto de recipientes de capacidade B , um limite C na quantidade de classes, e um conjunto de itens também $I = [n]$, onde cada item $i \in I$ possui um tamanho s_i e

*Este trabalho foi parcialmente financiado pelos processos nº 2013/21744-8 e 2014/25892-4, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e processos nº 311499/2014-7, 306358/2014-0 e 479070/2012-1, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq.)

uma classe $c_i \in [Q]$ sendo Q o número de classes diferentes na instância, o objetivo do CCBP é empacotar todos os itens utilizando a quantidade mínima de recipientes. Neste artigo, apresentamos algoritmos de *Branch-and-Price* para encontrar soluções exatas para o CCBP. Considerando P o conjunto de todos os padrões viáveis (subconjuntos de itens que podem ser empacotados em um mesmo recipiente) possíveis, e λ^p , uma variável binária que indica se um padrão p é usado na solução, podemos descrever o CCBP como o problema de minimizar a soma de λ^p , para todo $p \in P$, de forma que cada item esteja pelo menos um dos padrões utilizados na solução [Gilmore and Gomory 1961].

Em nossos algoritmos utilizamos a regra de ramificação simples [Vance 1998] para o CCBP. Ela consiste em escolher uma coluna p que possua valor fracionário após a solução da relaxação do problema mestre, e forçar duas situações opostas: de um lado, a coluna p deve ser sempre incluída na solução ($\lambda^p = 1$); e do lado oposto, a coluna p não poderá fazer parte da solução ($\lambda^p = 0$), ou seja, será um padrão proibido. Como o número de padrões viáveis é exponencial, escolhemos um conjunto inicial de padrões e geramos novos padrões durante a execução do algoritmo conforme necessário. Este processo é chamado de *pricing*, e para o CCBP, recai no Problema da Mochila com Restrições de Classe (CCKP) com padrões proibidos onde, dados um conjunto de itens, onde um item i tem um valor v_i , classe c_i , e tamanho s_i , e uma lista de padrões proibidos, o objetivo é selecionar um subconjunto de valor máximo deste itens satisfazendo as restrições de capacidade e do limite de itens de classes diferentes na mochila, de forma que este subconjunto não forme uma solução presente na lista de padrões proibidos, dada como entrada.

Em nossos algoritmos, implementamos dois métodos de exploração da árvore de busca. O primeiro trata-se do *Best-Bound*, no qual cada nó é armazenado e indexado pelo seu valor de relaxação, e os nós com menor valor de relaxação são explorados primeiro, obedecendo uma fila de prioridades. No segundo método, *Diving*, escolhemos um lado para ser sempre explorado primeiro até que uma solução inteira seja encontrada. Uma vez encontrada a solução inteira, o nó com menor relaxação linear é escolhido para o início do próximo mergulho. Para a formulação proposta, escolhemos explorar primeiro o lado da sub-árvores no qual uma coluna é fixada na solução. Desta forma, postergamos a exploração de nós onde o *pricing* requer evitar padrões proibidos.

3. Problema da Mochila Com Restrições de Classe

Como mencionado na seção anterior, o problema de *pricing* que devemos resolver recai no CCKP com padrões proibidos. Porém, é interessante considerar também algoritmos para a versão sem padrões proibidos, já que por vezes não temos tais padrões (como na raiz da árvore de busca) e mesmo que tivermos, podemos executar um tal algoritmo com a esperança que o mesmo encontre um padrão não proibido e, caso o mesmo encontre um padrão proibido, podemos utilizar outro algoritmo como *fallback*. Nesta seção discutimos os algoritmos implementados para resolver este problemas. Além dos algoritmos discutidos nesta seção, consideramos também uma formulação natural do programa linear inteiro para este problema (com padrões proibidos), que foi resolvida utilizando o Gurobi.

DPI Quando não temos padrões proibidos, considere $U(b, c, q)$ como o valor máximo de uma solução de tamanho até b , usando até c classes de 1 até q . Considere também que para cada classe $i \in [Q]$, temos que n_i representa a quantidade de itens da classe i . Para encontrar $U(b, c, q)$, precisamos primeiro computar, para cada classe $i \in [Q]$, o

valor máximo de uma solução de tamanho s , utilizando apenas itens da classe i . Considere $U'_i(s, j)$ como o valor máximo de uma solução de tamanho até s , utilizando apenas os primeiros j itens da classe i . Este valor é calculado com a recorrência $U'_i(s, j) = \max\{U'_i(s, j-1), U'_i(s-s_j, j-1) + v_j\}$. Com isso, podemos calcular o valor máximo de uma solução de tamanho até s , utilizando apenas itens da classe i como $U'_i(s, n_i)$, para cada classe em $[Q]$. A partir destes valores, podemos finalmente computar $U(b, c, q)$ como

$$U(b, c, q) = \max \left\{ \begin{array}{l} U(b, c, q-1) \\ \max_{1 \leq b' \leq b} \{U(b-b', c-1, q-1) + U'_q(b', n_q)\} \end{array} \right\},$$

sendo que o valor de uma solução ótima é dado por $U(B, C, Q)$.

DPII Novamente, considerando que não temos padrões proibidos, considere agora os itens ordenados de acordo com as classes as quais eles pertencem. Isto nos garante que todos os itens pertencentes a uma mesma classe estarão em sequência. Considere $V(i, b, c, u)$ como o valor máximo de uma solução utilizando itens de 1 até i , tamanho máximo da mochila b , utilizando até c classes diferentes, e tendo u como a variável indicadora que mostra se a classe do item i é utilizada ou não na solução. Considere também, l_i como o índice do último item pertencente a classe anterior a de i . Com isso, podemos computar $V(i, b, c, u)$ de forma que

$$V(i, b, c, 0) = \max \{V(l_i, b, c, 0), V(l_i, b, c, 1)\}$$

$$V(i, b, c, 1) = \begin{cases} \max \{V(i-1, b, c, 1), V(i-1, b-s_i, c, 1) + v_i\} & \text{caso } c_i = c_{i-1} \\ \max \left\{ \begin{array}{l} V(i-1, b, c-1, 1), \quad V(i-1, b, c-1, 0) \\ V(i-1, b-s_i, c-1, 1) + v_i, \quad V(i-1, b-s_i, c-1, 0) + v_i \end{array} \right\} & \text{caso } c_i \neq c_{i-1} \end{cases}$$

sendo que o valor de uma solução ótima é dado por $\max\{V(n, B, C, 0), V(n, B, C, 1)\}$.

Branch-and-Bound: Implementamos um algoritmo de *Branch-and-Bound* no qual exploramos os itens em ordem não crescente de eficiência, definida por v_i/s_i . Calculamos o limitante utilizando um algoritmo guloso. Durante o processo de atualização do valor incumbente, utilizamos uma estrutura de *hash* para checar se o padrão obtido nesta ramificação esta na lista de padrões proibidos e, em caso positivo, esta solução não é utilizada, e o algoritmo continua em busca de uma solução não proibida.

4. Experimentos e Resultados

Realizamos experimentos em 320 instâncias com diferentes valores de Q , C , n , B e intervalo de tamanho de cada item. Registramos o tempo necessário até a obtenção da solução ótima com um limite de 1800 segundos para cada instância. Utilizamos o Gurobi para resolver os limitantes e relaxações de programação linear inteira. Nestes experimentos, consideramos diferentes combinações de algoritmos para resolver o *pricing* assim como técnicas de exploração da árvore. Os algoritmos de Branch & Bound e o Programa Linear Inteiro (resolvido pelo *Gurobi*) são os únicos algoritmos que resolvem o problema de maneira definitiva, uma vez que os algoritmos de Programação Dinâmica podem retornar padrões proibidos. Além disso, o tempo de execução da DPI é $\Theta(BCQ + nB)$, enquanto que o tempo de execução da DPII é $\Theta(nBC)$ e, assim, dependendo da instância uma pode ser melhor do que a outra.

A Figura 1, apresenta a fração das instâncias resolvidas em função do tempo, considerando as combinações dos algoritmos de *pricing* básicos, *Branch-and-Bound* (B&B)

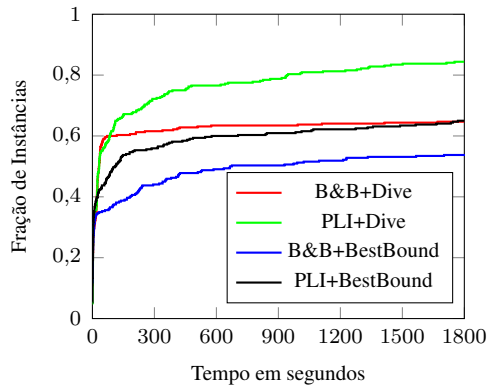


Figura 1. Básicos

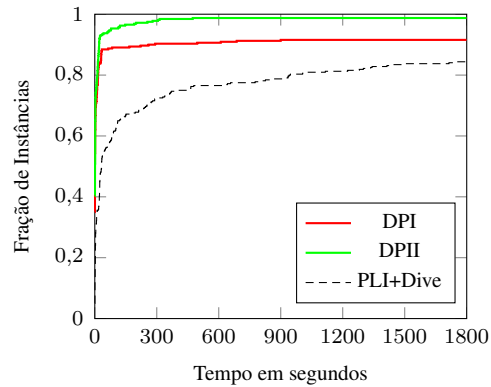


Figura 2. DPI / DPPII

e Programação Linear Inteira (PLI), e os métodos de exploração *Diving* e *Best-Bound*. Podemos observar que o método *Diving* apresenta resultados melhores independente do algoritmo de *pricing* utilizado, e embora o algoritmo de B&B resolva até 63% das instâncias mais rápido do que sua contraparte em PLI, a diferença no tempo de solução para estas instâncias não é grande. Por outro lado, a combinação PLI e *Diving* consegue resolver cerca de 85% das instâncias antes do limite de tempo, enquanto B&B e *Diving* resolve em torno de 64%. A Figura 2, apresenta a melhor solução do Figura 1 e as incrementa com o uso dos algoritmos programação dinâmica para acelerar a geração de colunas. Tais algoritmos sempre são utilizados, mesmo na presença de padrões proibidos. Porém, caso a solução encontrada seja de fato um padrão proibido, nós resolvemos o problema utilizando o PLI. Podemos observar que o acréscimo das DPs melhora consideravelmente o tempo de solução das instâncias. Para a melhor combinação, *Diving* + PLI acrescentado de DPPII, temos 93% das instâncias resolvidas em menos de 30 segundos, e totalizando em torno de 99% até o limite de tempo.

Estes resultados nos dão confiança para continuar atacando este problema com esta técnica, uma vez que não existem outros estudos experimentais publicados sobre este problema para comparação direta. Estamos realizando experimentos adicionais envolvendo outros algoritmos para resolução do *pricing*, assim como outro modelo de ramificação, sob um conjunto de instâncias ainda maior.

Referências

- Epstein, L., Imreh, C., and Levin, A. (2010). Class constrained bin packing revisited. *Theoretical Computer Science*, 411(34):3073–3089.
- Gilmore, P. C. and Gomory, R. E. (1961). A linear programming approach to the cutting-stock problem. *Operations Research*, 9(6):849–859.
- Shachnai, H. and Tamir, T. (2001). Polynomial time approximation schemes for class-constrained packing problems. *Journal of Scheduling*, 4(6):313–338.
- Vance, P. H. (1998). Branch-and-price algorithms for the one-dimensional cutting stock problem. *Computational Optimization and Applications*, 9(3):211–228.
- Vance, P. H., Barnhart, C., Johnson, E. L., and Nemhauser, G. L. (1994). Solving binary cutting stock problems by column generation and branch-and-bound. *Computational Optimization and Applications*, 3(2):111–130.
- Xavier, E. C. and Miyazawa, F. K. (2008). The class constrained bin packing problem with applications to video-on-demand. *Theoretical Computer Science*, 393(1):240–259.

Uma Aproximação para o Problema de Alocação de Terminais

Lehilton L.C. Pedrosa¹, Vinícius F. dos Santos², Rafael C.S. Schouery^{1†}

¹Instituto de Computação – Universidade Estadual de Campinas
Campinas, São Paulo

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais

{lehilton,schouery}@ic.unicamp.br, viniuciussantos@dcc.ufmg.br

Abstract. *In the center p -hub problem, we are given a metric space (V, d) and a set $D \subseteq V^2$, that represents flow demands. The objective is to select a subset of V of size p , which are called hubs, and assign each demand uv to a terminal w , minimizing the maximum distance of traveling from origin u , hub w , and destination v . Although hub location problems comprise a very active area, there are few works on approximation. For the center p -hub problem, we notice that there is a lower bound of 2, and obtain the first approximation, that achieves factor 3.*

Resumo. *No problema de centros dos p -terminais são dados um espaço métrico (V, d) e um conjunto $D \subseteq V^2$ que representa demandas de fluxo. O objetivo é selecionar um subconjunto de V de tamanho p , denominado terminais, e associar cada demanda $uv \in D$ a um terminal w , de forma a minimizar o maior custo de percurso entre uma origem u , o terminal associado w e o destino v . Embora os problemas de alocação de terminais componham uma área bastante ativa nos últimos anos, ainda há poucos trabalhos de aproximação. Para o problema de centros dos p -terminais, nós verificamos um limitante inferior de 2 para o fator e obtemos um primeiro algoritmo de aproximação, com fator 3.*

1. Introdução

Problemas de alocação de terminais aparecem em diversas aplicações, quando há a necessidade de satisfazer demandas de fluxo entre pares dos mais variados objetos, como no deslocamento de pessoas, de transporte de *commodities*, ou de transmissão de informação. Nessas situações, ao invés de conectar os objetos (os chamados clientes) diretamente, muitas vezes é desejável a criação de terminais (os chamados *hubs*), com os quais a conexão é mais barata e eficiente. O exemplo clássico é o das malhas aeroportuárias, quando um voo entre duas cidades contém uma parada em algum aeroporto de transferência: o problema é como escolher e alocar uma quantidade limitada de terminais, minimizando os custos de conexão. Em outro exemplo, o objetivo é alocar um conjunto de *hubs* em uma rede de computadores, de forma a minimizar a maior latência entre dois nós.

Diversas variantes são consideradas, dependendo se o objetivo é minimizar o custo total de conexão, ou o maior custo de conexão; se instalar um terminal tem um custo, ou há uma quantidade limitada de terminais; se a demanda por uma unidade de fluxo passa

[†]Pesquisa parcialmente apoiada pelo processo nº 2013/21744-8, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

por um ou mais terminais, etc. Neste trabalho, estamos interessados em uma variante, que denominamos de *problema de centros dos p -terminais*, em que há p terminais disponíveis, cada demanda por fluxo passa por apenas um terminal e o objetivo é minimizar a maior distância entre um nó de origem, o terminal associado e o nó destino de uma demanda.

Embora estes problemas sejam estudados na teoria clássica de localização desde o início do século 20, os problemas de localização e alocação de terminais são estudados mais formalmente apenas a partir da década de oitenta, particularmente com o trabalho de [O’Kelly 1986]. Ao contrário dos problemas clássicos de localização, para os quais há diversos algoritmos de aproximação, os problemas de alocação de terminais foram estudados principalmente por meio de heurísticas e métodos exatos [Farahani et al. 2013].

Poucos algoritmos de aproximação foram propostos para problemas de alocação de terminais. Para uma variante chamada problema de centros dos p -terminais em estrela (*star p -hub center problem*) [Yaman and Elloumi 2012], quando o subgrafo induzido pelos terminais é uma estrela e cada cliente só pode se conectar a um único terminal folha, [Liang 2013] mostrou que o problema é inaproximável por um fator melhor que 1,25 e obteve uma 3,5-aproximação. Quando o objetivo é minimizar a soma dos custos de conexão, uma variante comumente considerada é o problema de terminais e raios de alocação simples (*single allocation hub-and-spoke problem*), quando o fluxo de uma demanda pode percorrer dois ou mais hubs, os p terminais são completamente interconectados e há um fator de desconto α , com $0 \leq \alpha \leq 1$, para conexão entre terminais. [Sohn and Park 1997] mostraram que o problema é NP-difícil mesmo para $p = 3$. [Iwasa et al. 2009] obtiveram uma aproximação com fator 3 e uma aproximação aleatorizada com fator 2.

Neste trabalho, nós damos mais um passo na direção de algoritmos de aproximação para problemas de terminais e apresentamos uma 3-aproximação para o problema de centros dos p -terminais. A principal técnica é um préprocessamento da instância utilizada no chamados *problemas de gargalo*, como o clássico problema dos k -centros [Hochbaum and Shmoys 1986]. Nós observamos que o problema não tem uma aproximação melhor que 2, a não ser que $P = NP$.

2. Preliminares

Definição No *problema de centros dos p -terminais*, uma instância é dada por um conjunto V , de clientes, uma função de custo de conexão associada $d : V \times V \rightarrow \mathbb{R}^+$, um conjunto de pares $D \subseteq V \times V$, que representam demandas por fluxo entre pares de clientes, e um inteiro positivo p . Uma solução é um subconjunto H de V com tamanho p , que representa terminais a serem instalados, e uma atribuição $\phi : D \rightarrow H$, que associa cada demanda a um terminal. O objetivo é encontrar uma solução que minimize

$$\max_{uv \in D} d(u, \phi(uv)) + d(\phi(uv), v).$$

Nós assumimos que a função d é métrica, isso é, é simétrica e satisfaz a desigualdade triangular $d(u, v) \leq d(u, w) + d(w, v)$ para todos $u, v, w \in V$. Observe que para descrever uma solução para o problema, basta obter um conjunto de terminais a serem instalados, já que uma demanda uv sempre será atribuída a um terminal instalado w cujo valor $d(u, w) + d(w, v)$ é mínimo.

Podemos reduzir o k -centros, que não admite fator de aproximação melhor que 2 [Hochbaum and Shmoys 1986]. Dada uma instância (V, k) do k -centros, criamos uma instância do nosso problema com demanda uu para cada $u \in V$ e $p = k$. Obtemos:

Teorema 1. *Suponha que exista uma ρ -aproximação para o Problema de Centros dos p -Terminais com $\rho < 2$, então $P = NP$.*

3. Uma 3-aproximação

3.1. Algoritmo

Nós podemos interpretar o problema dos p -terminais como um problema de gargalo. Nesses problemas, o valor da solução ótima é dado pelo peso de uma aresta do grafo subjacente, então podemos enumerar todos os valores possíveis e utilizar um préprocessamento para reduzir o problema para o caso particular em que métrica é dada por um grafo com pesos unitários. Embora o valor ótimo no problema dos p -terminais não seja necessariamente a distância entre dois clientes de V , ainda podemos utilizar essa estratégia. Nosso algoritmo é dividido em duas partes: *préprocessamento*, que reduz uma instância para uma instância de um problema auxiliar; e *clustering*, que obtém uma partição dos clientes de tal forma que cada parte tenha um centro e exista um terminal instalado “próximo” ao centro. Cada parte está descrita a seguir. O processo está resumido no Algoritmo 1.

Préprocessamento Dado $\tau \geq 0$, construa o grafo bipartido G_τ entre demandas D e candidatos a terminais V , tal que existe aresta entre $uv \in D$ e $w \in V$ se, e somente se, $d(u, w) + d(w, v) \leq \tau$. O objetivo do subproblema é: ou obter um certificado de que não há um conjunto de p terminais em G_τ cuja vizinhança cubra todas as demandas, ou obter uma solução para o problema original cujo valor é no máximo 3τ . Para $\tau = \text{OPT}$, onde OPT é o valor ótimo, o conjunto dos p terminais em uma solução ótima cobre todas as demandas, assim podemos testar todos os possíveis valores de $\tau \leq \text{OPT}$ e obter uma solução de custo no máximo $3\tau \leq 3\text{OPT}$. Seja $\tau_0 = \max_{uv \in D} d(u, v)$. Trivialmente temos $\text{OPT} \geq \tau_0$, já que o fluxo de cada demanda uv em uma solução ótima é pelo menos $d(u, v)$. Portanto, precisamos testar apenas valores $\tau \geq \tau_0$. Note que para $\tau \geq \tau_0$, cada demanda uv tem pelo menos um vizinho em G_τ .

Clustering Seja $N_{G_\tau}(v)$ o conjunto de vizinhos de v em G_τ . Dizemos que duas demandas uv e $u'v'$ são vizinhas se elas compartilharem um vizinho em G_τ , isto é, se $N_{G_\tau}(uv) \cap N_{G_\tau}(u'v') \neq \emptyset$. O conjunto de clientes é particionado de maneira gulosa, considerando as vizinhanças. Em cada iteração, uma demanda ainda não particionada é selecionada como “centro” e uma nova parte é criada com o centro e todas as demandas vizinhas do centro. O algoritmo termina quando todas as demandas estiverem particionadas. Em seguida, para cada parte com centro uv , um terminal é escolhido entre as adjacências de uv em G_τ . Em particular, note que $u \in N_{G_\tau}(uv)$.

3.2. Análise

Primeiro, obtemos o seguinte lema auxiliar, cuja demonstração segue diretamente da construção do algoritmo.

Lema 2. *Para cada τ , existe um conjunto $F_\tau \subseteq D$ tal que:*

1. *para todo $uv, u'v' \in F_\tau$, com $uv \neq u'v'$, $N_{G_\tau}(uv) \cap N_{G_\tau}(u'v') = \emptyset$;*
2. *para todo $u'v' \in D$, existe $uv \in F_\tau$ tal que $N_{G_\tau}(uv) \cap N_{G_\tau}(u'v') \neq \emptyset$ e $u \in H_\tau$.*

Agora podemos demonstrar a aproximação.

Teorema 3. *O Algoritmo 1 é uma 3-aproximação para o problema de centros dos p -terminais.*

Algoritmo 1: Aproximação para o problema de centros dos p -terminais

Entrada: V, D, p, d

```

1  $\tau_0 \leftarrow \max_{uv \in D} d(u, v)$ 
2  $A \leftarrow \{d(u, w) + d(w, v) : uv \in D, w \in V, d(u, w) + d(w, v) \geq \tau_0\}$ 
3 para  $\tau \in A$  em ordem crescente faça
4   Crie o grafo bipartido  $G_\tau$ 
5    $D' \leftarrow D$ 
6    $H_\tau \leftarrow \emptyset$ 
7   enquanto  $D' \neq \emptyset$  faça
8     Seja  $uv \in D'$ 
9      $H_\tau \leftarrow H_\tau \cup \{u\}$ 
10     $D' \leftarrow D' \setminus \{u'v' \in D' : N_{G_\tau}(uv) \cap N_{G_\tau}(u'v') \neq \emptyset\}$ 
11    se  $|H_\tau| \leq p$  então
12      devolva  $H_\tau$ 

```

4. Extensões e trabalhos futuros

Ao contrário dos k -centros, para o qual o fator de aproximação corresponde ao limitante de aproximação de 2, para o problema dos p -terminais o fator é 3 e o limitante inferior é 2; uma questão natural é, portanto, se ambos problemas podem ser aproximados pelo mesmo fator, ou se o problema dos p -terminais é mais difícil de aproximar. Em trabalhos futuros pretendemos considerar generalizações do problema, como a variante em que mais de um terminal pode ser percorrido, quando há incentivo no percurso entre terminais.

Referências

- [Farahani et al. 2013] Farahani, R. Z., Hekmatfar, M., Arabani, A. B., and Nikbakhsh, E. (2013). Hub location problems: A review of models, classification, solution techniques, and applications. *Computers & Industrial Engineering*, 64(4):1096–1109.
- [Hochbaum and Shmoys 1986] Hochbaum, D. S. and Shmoys, D. B. (1986). A Unified Approach to Approximation Algorithms for Bottleneck Problems. *Journal of the ACM*, 33(3):533–550.
- [Iwasa et al. 2009] Iwasa, M., Saito, H., and Matsui, T. (2009). Approximation algorithms for the single allocation problem in hub-and-spoke networks and related metric labeling problems. *Discrete Applied Mathematics*, 157(9):2078–2088.
- [Liang 2013] Liang, H. (2013). The hardness and approximation of the star-hub center problem. *Operations Research Letters*, 41(2):138–141.
- [O’Kelly 1986] O’Kelly, M. E. (1986). The Location of Interacting Hub Facilities. *Transportation Science*, 20(2):92–106.
- [Sohn and Park 1997] Sohn, J. and Park, S. (1997). A linear program for the two-hub location problem. *European Journal of Operational Research*, 100(3):617–622.
- [Yaman and Elloumi 2012] Yaman, H. and Elloumi, S. (2012). Star p -hub center problem and star p -hub median problem with bounded path lengths. *Computers & Operations Research*, 39(11):2725–2732.

UKP5: Solving the Unbounded Knapsack Problem

Henrique Becker , Luciana S. Buriol

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{hbecker,buriol}@inf.ufrgs.br

Abstract. *In this extended abstract we present UKP5, an algorithm for solving the unbounded knapsack problem. UKP5 is based on dynamic programming, but implemented in a non traditional way: instead of looking backward for stored values of subproblems, it stores incremental lower bounds forward. UKP5 is considerably simpler than EDUK2, the state-of-the-art algorithm for solving the problem. We run UKP5 and EDUK2 on a benchmark of 4540 hard instances proposed by the authors of EDUK2. The results reveal that UKP5 outperforms EDUK2, being 47 times faster on the average.*

Resumo. *Nesse resumo estendido nós apresentamos o UKP5, um algoritmo para solucionar o unbounded knapsack problem (problema da mochila com repetições). O UKP5 é baseado em programação dinâmica, mas implementado de uma forma não-tradicional: ao invés de olhar para trás para usar soluções de subproblemas previamente computados, ele armazena limites inferiores a frente. O UKP5 é consideravelmente mais simples que o EDUK2, o algoritmo do estado da arte para solucionar o problema. Nós executamos o UKP5 e o EDUK2 em uma bateria de testes contendo 4540 instâncias consideradas difíceis pelos autores do EDUK2. Os resultados mostram que o UKP5 é, em média, 47 vezes mais rápido que o EDUK2.*

1. Introduction

The unbounded knapsack problem (UKP) is a simpler variation of the well-known bounded knapsack problem (BKP). UKP allows the allocation of an unbounded quantity of each item type. The UKP is NP-Hard, and thus has no known polynomial-time algorithm for solving it. However, it can be solved by a pseudo-polynomial dynamic programming algorithm.

Two techniques are often used for solving UKP: dynamic programming (DP) [Andonov et al. 2000], [Garfinkel and Nemhauser 1972, p. 214], and branch and bound (B&B) [Martello and Toth 1990]. The state-of-the-art solver for the UKP, introduced by [Poirriez et al. 2009], is a hybrid solver that combines DP and B&B. The solver's name is PYAsUKP, and it is an implementation of the EDUK2 algorithm.

An UKP instance is composed by a capacity c , and a list of n items. Each item can be referenced by its index in the item list $i \in \{1 \dots n\}$. Each item i has a weight value w_i , and a profit value p_i . A solution is an item multiset, i.e, a set that allows multiple copies of the same element. The sum of the items weight, or profit, of a solution s is denoted by w_s , or p_s . A valid solution s has $w_s \leq c$. An optimal solution s^* is a valid solution

with the greatest profit among all valid solutions. The UKP objective is to find an optimal solution for the given UKP instance. The mathematical formulation of UKP is:

$$\text{maximize } \sum_{i=1}^n p_i x_i \quad (1) \quad \text{subject to } \sum_{i=1}^n w_i x_i \leq c \quad (2) \quad x_i \in \mathbb{N}_0 \quad (3)$$

The quantities of each item i in an optimal solution are denoted by x_i , and are restricted to the non-negative integers, as (3) indicates. The efficiency of an item i is the ratio $\frac{p_i}{w_i}$. We use w_{min} and w_{max} to denote the smallest item weight, and the biggest item weight, respectively.

2. UKP5: The Proposed Algorithm

UKP5 is inspired by the DP algorithm described by Garfinkel and Nemhauser (we will reference it as G&N) [Garfinkel and Nemhauser 1972, p. 221]. The name “UKP5” is due to five improvements applied over that algorithm: **Symmetry pruning**: symmetric solutions are pruned in a more efficient fashion than in G&N; **Sparsity**: not every position of the optimal solutions value array has to be computed; **Dominated solutions pruning**: we never generate some solutions if they are worse than solutions already generated (bigger weight and smaller profit); **Time/memory tradeoff**: the test $w_i \leq y$ from G&N was removed in cost of more $O(w_{max})$ memory; **Periodicity**: the periodicity check suggested in [Garfinkel and Nemhauser 1972] (but not implemented there) was adapted and implemented.

The g is a sparse array where we store solutions profit. If $g[y] > 0$ then there exists a non-empty solution s with $w_s = y$ and $p_s = g[y]$. The d array stores the index of the last item used on a solution. If $g[y] > 0 \wedge d[y] = i$ then the solution s with $w_s = y$ and $p_s = g[y]$ has at least one copy of item i . Our first loop (lines 4 to 9) stores all solutions comprised of a single item in the arrays g and d . After this setup, we simply iterate g and, when we find a stored solution, we create new solutions combining the current solution with single items. We only prune symmetric or dominated solutions, all other valid solutions are generated. Consequently, one of those solutions is guaranteed to be an optimal solution, and opt will end with the optimal solution profit value.

With the intent of making easier to the reader to understand the core idea of the UKP5 algorithm, the pseudocode presented at Algorithm 1 was stripped of many small optimizations. Some of them are: all the items are sorted by non-increasing efficiency; the y^* periodicity bound is computed as in [Garfinkel and Nemhauser 1972, p. 223], and used to reduce the c value; an UKP5-specific periodicity check is used, we don’t describe it here because of the page limit. The solution assemble phase also isn’t described here, but it’s similar to the one used by the DP method described in [Garfinkel and Nemhauser 1972, p. 221, Steps 6-8].

3. Computational Results, Analysis and Conclusions

The computer used on the experiments was an ASUS R552JK-CN159H (Intel Core i7-4700HQ Processor, 6M Cache, 3.40 GHz). The operating system used was Linux 4.3.3-2 x86_64. The number of instances of each different class are: Subset-Sum (400); Strong Correlation (240); Postponed Periodicity (800); No Collective Dominance (2000);

Algorithm 1 UKP5 – Computation of opt

```

1: procedure UKP5( $n, c, w, p, w_{min}, w_{max}$ )
2:    $g \leftarrow$  array of  $c + w_{max}$  positions each one initialized with 0
3:    $d \leftarrow$  array of  $c + w_{max}$  positions each one initialized with  $n$ 
4:   for  $i \leftarrow 1, n$  do ▷ Stores one-item solutions
5:     if  $g[w_i] < p_i$  then
6:        $g[w_i] \leftarrow p_i$ 
7:        $d[w_i] \leftarrow i$ 
8:     end if
9:   end for
10:   $opt \leftarrow 0$ 
11:  for  $y \leftarrow w_{min}, c$  do ▷ Can end early because of periodicity check
12:    if  $g[y] \leq opt$  then ▷ Handles sparsity and pruning of dominated solutions
13:      continue ▷ Ends current iteration and begins the next
14:    end if
15:     $opt \leftarrow g[y]$ 
16:    for  $i = 1, d[y]$  do ▷ Creates new solutions (never symmetric)
17:      if  $g[y + w_i] < g[y] + p_i$  then
18:         $g[y + w_i] \leftarrow g[y] + p_i$ 
19:         $d[y + w_i] \leftarrow i$ 
20:      end if
21:    end for
22:  end for
23:  return  $opt$ 
24: end procedure

```

SAW (1100). These datasets aim to reproduce the ones described in [Poirriez et al. 2009] (see the paper for more information). The same tool was used to generate the datasets (PYAsUKP). The capacity of the instances vary between 10^6 and 10^8 , and the number of items vary between 10^3 and 10^5 . The sources can be found at <https://github.com/henriquebecker91/masters/tree/v0.1>¹.

As we can see in Figure 3, for many instances PYAsUKP is faster than UKP5. However, PYAsUKP can also be one or more orders of magnitude slower than UKP5 on some instances. Our tests have shown when PYAsUKP is faster than UKP5 this is often because of its B&B phase, that solves some instances almost instantly. When B&B fails to solve the instance in a short time, PYAsUKP fallbacks to a DP algorithm that seems to be many times slower than UKP5. It's important to note that both algorithms have a pseudo-polynomial worst-case ($O(c \times n)$). Integrating a B&B phase to UKP5 seems promising, and will be the theme for future works.

¹The UKP5 implementation is at **codes/cpp/** and two versions of PYAsUKP are at **codes/ocaml/**. The *pyasukp_site.tgz* is the version used to generate the instances, and was also available at <http://download.gna.org/pyasukp/pyasukpsrc.html>. A more stable version was provided by the authors. This version is in *pyasukp_mail.tgz* and it was used to solve the instances the results presented in Figure 3. The *create_*_instances.sh* scripts inside **codes/sh/** were used to generate the instance datasets.

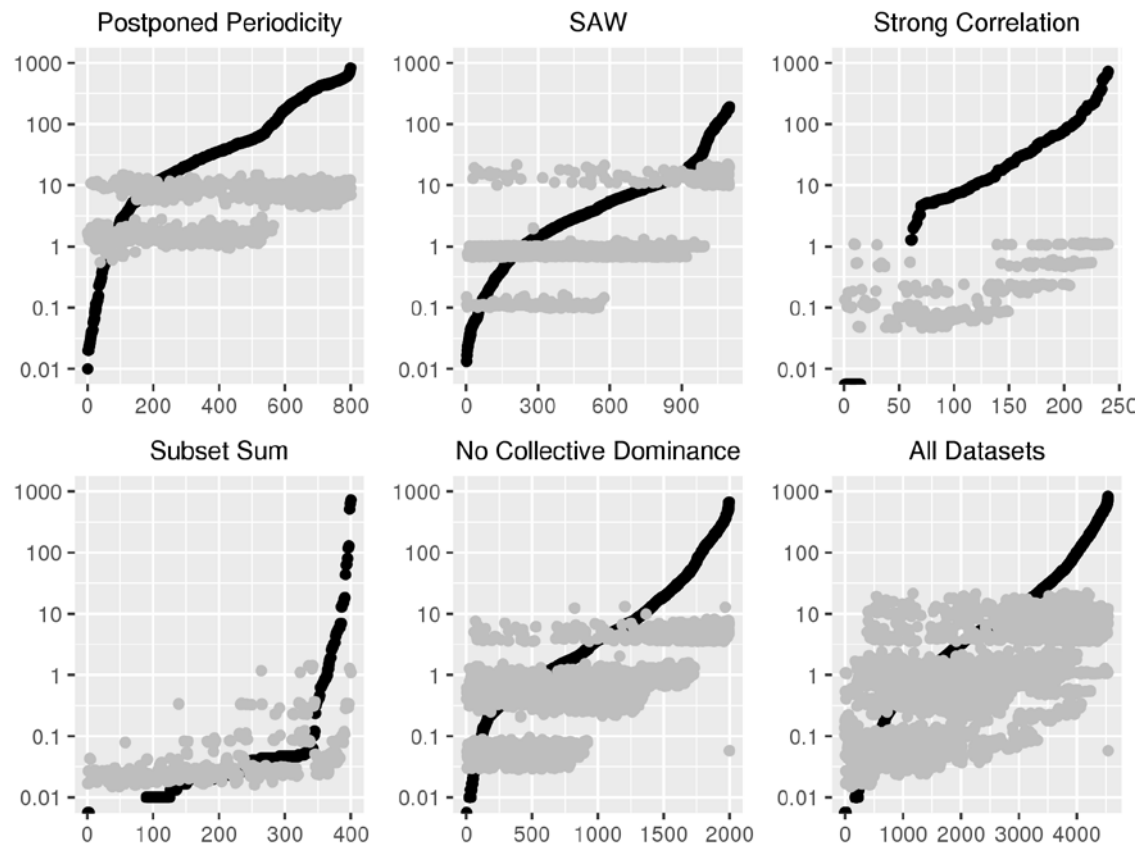


Figure 1. The times used by UKP5 and PYAsUKP for each instance of each class. The black dots represent PYAsUKP times. The gray dots represent UKP5 times. The y axis is the time used to solve an UKP instance, in seconds. The x axis is the instance index when the instances are sorted by the time PYAsUKP took to solve it. Note that the y axis is in logarithmic scale.

Acknowledgments We are thankful to the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support.

References

- Andonov, R., Poirriez, V., and Rajopadhye, S. (2000). Unbounded knapsack problem: Dynamic programming revisited. *European Journal of Operational Research*, 123(2):394–407.
- Garfinkel, R. S. and Nemhauser, G. L. (1972). *Integer programming*, volume 4. Wiley New York.
- Martello, S. and Toth, P. (1990). An exact algorithm for large unbounded knapsack problems. *Operations research letters*, 9(1):15–20.
- Poirriez, V., Yanev, N., and Andonov, R. (2009). A hybrid algorithm for the unbounded knapsack problem. *Discrete Optimization*, 6(1):110–124.

Uma Abordagem baseada nas Preferências dos Docentes para o Problema de Programação de Horários em Universidades

Osmar Leandro Dantas da Silva¹, Diego Rocha Lima¹

¹Instituto Federal do Ceará campus Crato (IFCE)
Caixa Postal 63115-500 – Crato – CE – Brasil

osmarleandrods@gmail.com, diego.rocha@ifce.edu.br

Abstract. *The university timetabling problem is usually solved by the responsible manager on planning with the teaching staff of the institution. The preferences of teachers may not be relevant in certain respects, which does not guarantee they are always enforced. Thus, this work will approach the issue of allocation of time at universities, with the proposal of an accurate model. The optimal solutions of the tests were found in medium 38 minutes for the instance with 263.040 decision variables, which makes it feasible to use the model to generate the six-month timetable.*

Resumo. *O problema de programação de horários em universidades é resolvido normalmente pelo gestor responsável através de planejamento com a equipe pedagógica da instituição. As preferências dos docentes podem não ser relevantes em certos aspectos, o que não garante que sempre sejam obedecidas. Dessa forma, neste trabalho será abordado o problema de alocação de horários em universidades, com a proposta de um modelo exato. As soluções ótimas dos testes foram encontradas no período médio de 38 minutos para a instância com 263.040 variáveis de decisão, o que torna viável a utilização do modelo para gerar a grade horária semestral.*

1. Introdução

A alocação de horários em universidades consiste na programação de docentes, disciplinas e seus respectivos horários semestrais, respeitando certas restrições. Algumas das restrições são quantidade de salas de aula e laboratórios, quantidade de docentes, dias letivos e disciplinas. Os recursos da instituição são alocados de forma a alcançar um objetivo pré-definido (BUCCO, 2014).

O responsável pela programação de horários geralmente realiza suas atividades manualmente, ou assistido por computador, alocando um recurso humano para cada aula em um horário. Atender as restrições do corpo docente e da coordenação institucional geralmente dificulta a obtenção de um horário sólido, uma vez que estas alterações podem influenciar em outros horários de forma encadeada.

Segundo Schaerf (1999), a instância deste problema geralmente possui milhares de variáveis, sendo que estas variáveis devem respeitar as restrições da própria instituição. Assim, o planejamento pode envolver dias, o que nem sempre torna viável uma programação de horários de forma satisfatória, o que justifica a busca por uma solução automatizada.

Este trabalho tem como objetivo apresentar um modelo matemático que, considerando as preferências dos docentes por disciplinas, dias e horários, realize a alocação de horários de um curso superior.

2. Trabalhos Relacionados

O problema de escalonamento de horários é recorrente e bastante estudado. Diversas soluções para cenários distintos já foram apresentados e consolidados na literatura. Dessa forma, esta revisão tem como base trabalhos publicados entre 2005 e 2015, que tratam algum tipo de preferência acadêmica.

A solução desenvolvida por Hime (2015) consiste em um modelo matemático inteiro, em conjunto com uma interface web desenvolvida para interação com o usuário. A instância utilizada por ele foi construída com base nos cursos de Pós-graduação do Departamento de Estatística e Informática da UFRPE, que envolve 58 disciplinas e 30 professores. Neste trabalho, o autor permitiu que o corpo docente indicasse suas preferências via e-mail.

A solução proposta por Bucco (2014) contribuiu para a literatura através de uma revisão bibliográfica bastante completa de diversos trabalhos que envolvem o problema. Além disso, seu trabalho propõe um modelo linear que se aproxima bastante da realidade. O modelo do autor foi implementado utilizando o Microsoft Access 2007, para carregamento de dados, Microsoft Visual C++ 2010 Express, para implementação do modelo matemático, e o solver IBM ILOG CPLEX, para solução da instância.

Já o modelo de Costa et al. (2012) foi desenvolvido em programação inteira buscando maximizar as preferências dos professores do departamento de Engenharia de Produção da UFPB, onde os docentes indicariam seu benefício para cada disciplina do curso, semelhante ao que acontece no trabalho de Hime (2015). Com relação às restrições, os autores definem carga horária para os docentes.

Com relação ao trabalho de Souza Filho e Gomes (2009), o problema é retratado no Departamento de Tecnologias e Linguagens do Instituto Multidisciplinar da UFRRJ. Neste trabalho, os autores apresentam um modelo matemático que foi testado utilizando instâncias com 36 disciplinas, 18 professores e 8 salas por dia, em dois horários noturnos, resultando em 3.240 variáveis inteiras e 167 restrições.

3. Modelagem do Problema

O desenvolvimento do modelo teve como base os trabalhos de Bucco (2014), Hime (2015) e Souza Filho e Gomes (2009). Os dados utilizados no modelo são definidos matematicamente através de conjuntos e matrizes, e serão explanados a seguir.

$p \in P$	conjunto de docentes.
$d \in D$	conjunto de dias letivos, exemplo: segunda-feira, sábado.
$s \in S$	conjunto de salas de aula, exemplo: sala 1, laboratório 1.
$t \in T$	conjunto de disciplinas semestrais, exemplo: Cálculo Diferencial I.
$h \in H$	conjunto de horários, exemplo: 07:00 às 09:20.
CH_t	conjunto de carga horária de cada t-ésima disciplina.
$TP_{p,t}$	conjunto de preferências do p-ésimo docente para a t-ésima disciplina.
$DP_{p,d,h}$	preferências do p-ésimo docente para o d-ésimo dia no h-ésimo horário.

O conjunto CH_t indica a quantidade de aulas da t-ésima disciplina durante a semana, no caso deverá ser uma ou duas vezes, enquanto que os conjuntos $TP_{p,t}$ e $DP_{p,d,h}$

representam as preferências por disciplinas e por dias/horários, respectivamente. Para as restrições, considere o modelo abaixo.

$$\text{maximize} \quad \sum_{p \in P, d \in D, s \in S, h \in H, t \in T} x_{pdhst} \times TP_{pt} \times DP_{pdh} \quad (01)$$

$$\sum_{t \in T, s \in S} x_{pdhst} \leq 1 \quad \forall p \in P, d \in D, h \in H \quad (02)$$

$$\sum_{p \in P, t \in T} x_{pdhst} \leq 1 \quad \forall d \in D, h \in H, s \in S \quad (03)$$

$$\sum_{s \in S} y_{st} \leq 1 \quad \forall t \in T \quad (04)$$

$$\sum_{p \in P} z_{pt} \leq 1 \quad \forall t \in T \quad (05)$$

$$\sum_{h \in H} x_{pdhst} \leq 1 \quad \forall p \in P, d \in D, s \in S, t \in T \quad (06)$$

$$x_{pdhst} \leq y_{st} \quad \forall p \in P, d \in D, h \in H, s \in S, t \in T \quad (07)$$

$$\sum_{d \in D, h \in H, s \in S} x_{pdhst} \leq z_{pt} \times ch_t \quad \forall p \in P, t \in T \quad (08)$$

$$x_{pdhst} \in \{0,1\} \quad (09)$$

$$y_{st} \in \{0,1\} \quad (10)$$

$$z_{st} \in \{0,1\} \quad (11)$$

A função objetivo em (01) tem como foco encontrar a solução ótima considerando o valor das preferências dos docentes. Dessa forma, espera-se que a alocação seja realizada buscando maximizar este o valor resultante. Na restrição (02) é definido que cada professor em um horário e um dia deve estar alocado apenas a uma disciplina em uma sala. Já na restrição (03), cada sala de aula deve estar alocada para apenas uma aula (professor/disciplina) em um dia e um horário. Concomitantemente, a restrição (04) indica cada disciplina deve ter suas aulas estar alocadas em apenas uma sala. A restrição (05) define a variável z_{pt} , onde cada disciplina deve ter suas aulas alocadas para, no máximo, um docente.

A restrição (06) mostra que a quantidade de aulas (professor/disciplina) por dia não deve ultrapassar um, permitindo que o modelo não realize a alocação consecutiva de aulas para uma mesma disciplina. Em seguida, a restrição (07) define a alocação de apenas uma sala para cada aula no mesmo instante, limitando a quantidade de aulas consecutivas à quantidade de salas. A restrição seguinte (08) tem como objetivo alocar todas as aulas de uma disciplina para um único docente, além de realizar o controle da carga horária da disciplina.

Por fim, as restrições (09), (10) e (11) representam o domínio das variáveis de decisão. Considere que x_{pdhst} deve assumir o valor 1 quando indicar a alocação do p -ésimo professor à t -ésima disciplina, no h -ésimo horário do d -ésimo dia na s -ésima sala. Com isso, a variável deverá assumir o valor 0 caso contrário. Por conseguinte, as variáveis y_{st} e z_{pt} indicam respectivamente a alocação da s -ésima sala para a t -ésima disciplina e a alocação do p -ésimo docente à t -ésima disciplina.

4. Estudo de Caso e Resultados

Para realizar os testes, foram utilizadas duas instâncias de tamanhos diferentes, sendo que uma delas contém os dados reais das preferências coletados com os docentes através da

tecnologia Formulários Google. Já a máquina utilizada continha um processador Intel Core i7 3612QM (3,1Ghz), 8GB de memória RAM DDR3 e sistema operacional Linux Ubuntu 15.10 de 64 bits e AMPL for *Students* com o Gurobi 6.5. Antes de iniciar os testes com a instância da pesquisa em (P1), foi utilizada uma outra instância (P2) com a finalidade de testar a capacidade do modelo.

Tabela 1. Resultados dos testes por instância.

Instância	Docentes	Disciplinas	Salas	Total de Variáveis	Tempo Resolução
P1	15	43	08	45.623	6,59 segundos
P2	25	80	13	263.040	39,3 minutos

O estudo de caso consistiu na coleta das preferências dos docentes do curso de Sistemas de Informação do Instituto Federal do Ceará campus Crato, por meio de uma pesquisa on-line realizada com Formulários Google. Como mostrado anteriormente, participaram 15 docentes, que informaram suas preferências sobre as 43 disciplinas, os cinco dias da semana e dois horários noturnos.

5. Conclusão

Pode-se concluir a partir do estudo de caso que a qualidade da alocação depende sumariamente da disponibilidade do corpo docente para as disciplinas do curso superior. No estudo, todas as disciplinas foram alocadas, já que existiam horários suficientes para escalonar todo o curso. Concomitantemente, uma quantidade insuficiente de docentes ou de salas influencia negativamente no resultado. Finalmente, conclui-se que o modelo pode substituir o método tradicional de programação de horário, permitindo tempo de resolução abaixo de 38 minutos. Esse tempo é aceitável uma vez que apresenta ganho de tempo comparando-se com o método tradicional. Como trabalhos futuros pretende-se criar uma aplicação web e disponibilizar para os outros cursos da instituição.

Referências

- Bucco, G. B. (2014) “Construção de um modelo de programação linear para o University Timetabling Problem”. 156 f. Dissertação (Mestrado em Administração) - Universidade Federal do Rio Grande do Sul. Porto Alegre, jun. 2014.
- Costa, L. C. A.; Acioli, L. F.; Subramanian, A. (2012) “Um modelo de programação inteira para o problema de alocação de professores a turmas do departamento de engenharia de produção de uma instituição de ensino superior”. XLIV Simpósio Brasileiro de Pesquisa Operacional, Rio de Janeiro.
- Hime, R. (2015) “Uma aplicação da Programação Inteira no School Timetabling Problem”. Universidade Federal Rural de Pernambuco.
- Schaerf, A. (1999) A survey of automated timetabling. In *Artificial intelligence review*, v. 13, n. 2, p. 87-127.
- Souza Filho, E. M.; Gomes, C. R. (2009) “Programação do quadro de horários de disciplinas de uma universidade via programação inteira”. In XLI Simpósio Brasileiro de Pesquisa Operacional, Porto Seguro, 357-368.
- Neukirchen, F. V. P. et al. (2014) “Um estudo de caso sobre a geração de quadros de horários no departamento de Ciência da Computação da UFRGS”. XLIV Simpósio Brasileiro de Pesquisa Operacional, Salvador – BA, p. 3272–3279.

On a Leasing Variant of the Online Connected Facility Location Problem

Murilo Santos de Lima^{*1}, Mário César San Felice^{†2}, Orlando Lee^{‡1}

¹Institute of Computing – Campinas State University (Unicamp)
Av. Albert Einstein, 1251 – Cidade Universitária – 13083-852 – Campinas – SP – Brazil

²Institute of Mathematics and Statistics – University of São Paulo (USP)
Rua do Matão, 1010 – 05508-090 – São Paulo – SP – Brazil

{ra137105, felice, lee}@ic.unicamp.br

Abstract. *In the leasing optimization model, resources are leased for K different time periods, instead of being acquired for unlimited duration. The goal is to use these temporary resources to maintain a dynamic infrastructure that serves n requests while minimizing the total cost. We propose and study a leasing variant of the online connected facility location problem, which we call the **online connected facility leasing** problem. In this problem each client that arrives must be connected to a temporary facility, which in turn must be connected to a root facility using permanent edges. We present an algorithm that is $O(K \cdot \lg n)$ -competitive if the scaling factor is $M = 1$.*

1. Introduction

In traditional network design problems, the goal is to acquire resources corresponding to nodes or connections in order to build a minimum-cost network infrastructure that serves client requests. For example, in the **facility location** (FL) problem, we wish to determine where to install servers to satisfy connection requests by a set of clients.

However, in many scenarios, it is not plausible for a small company to install its own servers, and requests may be concentrated in certain time intervals. Therefore, it may be cheaper to **lease** a server in the cloud (from a service like Amazon AWS, Google App Engine or Microsoft Azure) for specific time intervals. This scenario has recently motivated the proposal of a class of problems known as **leasing optimization** [Meyerson 2005, Anthony and Gupta 2007, Nagarajan and Williamson 2013, Abshoff et al. 2015]. In those problems, instead of acquiring resources that last for unlimited duration, each resource may be leased for different time periods (e.g., a day, a week, a month). In this context, it is cheaper to lease a resource for a longer period than for smaller periods totalizing an equivalent time; this represents an economics of scale. This model may be applied to **offline** problems, which are traditional optimization problems in which the whole input is known in advance and there is no causality restriction for a solution, and to **online** problems, in which at every instant a request arrives and must be served without knowledge of future requests, and previous decisions cannot be revoked.

*Supported by FAPESP Proc. 2014/18781-1, and CNPq Proc. 142161/2014-4.

†Supported by CAPES PNPd scholarship 1522390.

‡Supported by CNPq Proc. 311373/2015-1, and CNPq Proc. 477692/2012-5.

In the **connected facility location** (CFL) problem, we are given a graph with edge costs, a subset F of the vertices (potential facilities), an opening cost for each facility, a subset D of the vertices (clients) and a constant M . The objective is to select a subset of the facilities to open and a subset of the edges that connects the open facilities, so to minimize the cost of the open facilities, plus M times the cost of the edges that connect the facilities, plus the sum of the distances from each client to its closest open facility.

In this paper we propose a leasing variant of this problem, which we call the **connected facility leasing** (CFLE) problem. In this version, we are given a set F of potential facilities, and clients arriving in different times that we wish to connect to open facilities. However, in the leasing version we no longer buy a facility for an unlimited time, but instead we can lease each facility for one of K different periods of time. Suppose those periods of time have lengths $\delta_1, \dots, \delta_K$, and leasing a facility $f \in F$ for a lease of type k costs $\gamma_f^k \in \mathbb{R}_+$. If we lease a facility f with leasing type k at instant \hat{t} , then we say that facility lease (f, k, \hat{t}) is **active** during interval $[\hat{t}, \hat{t} + \delta_k)$. At every instant $t \in \{0, \dots, T - 1\}$ we receive a set D_t of clients which we wish to connect to active facilities. We are given a distance $d(f, j)$ between each potential facility f and each client j . We also have to connect the facilities to a root facility r ; however the edges that connect facilities to the root cost M times the cost of an edge connecting a client to a facility, just like a backbone cable costs more than a common cable for a network service provider. The objective is to minimize the cost of leasing the facilities, plus the sum of the distances between each client and the facility to which it is connected, plus M times the cost of the edges connecting the facilities to the root. Note that this problem has a natural temporal requirement, even in an offline setting. Note, also, that we have leasing just on facilities: this models, for example, the problem of a network service provider, who has to install cables between routers to serve clients, but has routers with different limited lifetimes. Our focus will be in online algorithms, in which clients are given in an online manner, facility leases and core edges cannot be removed, and a new facility lease cannot be used to serve an old client request.

Related Work. [Meyerson 2005] proposed the **parking permit** (PP) problem, which has become the central problem for the leasing optimization model. In the same paper, Meyerson studied the online leasing version for the **Steiner forest** (SF) problem.

[Anthony and Gupta 2007] presented approximation algorithms for offline leasing versions of several NP-hard network design problems. For the **facility leasing** (FLE) problem, they obtained an $O(K)$ -approximation, where K is the number of lease types.

[Nagarajan and Williamson 2013] improved the result for the offline version of problem FLE, obtaining a 3-approximation. For the online version, the **online facility leasing** (OFLE) problem the authors showed an algorithm with competitive factor $O(K \lg n)$, where n is the number of client requests.

[Abshoff et al. 2015] improved the result on problem OFLE, presenting an online $O(\delta_K \lg \delta_K)$ -competitive algorithm, where δ_K is the longest lease duration.

Problem CFL is a generalization of the Steiner tree (ST) problem, so it is NP-hard, and the best known offline algorithm for problem CFL is a 3.19-approximation [Grandoni and Rothvoß 2011]. The **online connected facility location** (OCFL) problem, similarly, has lower bound $\Omega(\lg n)$ for its competi-

tive factor due to the lower bound on the **online Steiner tree (OST)** problem [Imase and Waxman 1991]. There are known $O(\lg n)$ -competitive online algorithms for problem OCFL [Umboh 2015, San Felice et al. 2016].

Our Contribution. We propose a first leasing variant of problem CFL, and give an $O(K \lg n)$ -competitive online algorithm for the case when the scaling factor is $M = 1$.

Text Organization. In Section 2 we give a formal definition of the problem and present our algorithm. In Section 3 we present some conclusions and future research directions.

2. Online Simple Connected Facility Leasing

We give the following definition of the problem. To simplify our notation, let $\mathcal{D} := \{(j, t) : j \in D_t, t \in \{0, \dots, T-1\}\}$ be the set of all client requests; we denote a client request as a pair (j, t) since the same client can request a connection at different instants.

Problem $\text{CFLE}(G, d, r, F, K, \gamma, \delta, M, D_0, \dots, D_{T-1})$: *The input consists of a graph $G = (V_G, E_G)$, a metric distance function $d : V_G \times V_G \mapsto \mathbb{R}_+$ on the vertices of G , a root vertex $r \in V_G$, a set $F \subseteq V_G$ of potential facilities, an integer $K > 0$ that represents the number of lease types, a cost $\gamma_f^k \in \mathbb{R}_+$ for leasing facility $f \in F$ with lease type $k \in [K]$, a function $\delta : [K] \mapsto \mathbb{N}$ that maps each lease type to a time length, a constant $M \in \mathbb{R}_+$, and a sequence $D_0, \dots, D_{T-1} \subseteq V_G$ of client sets. The goal is to find a set $X \subseteq F \times [K] \times \{0, \dots, T-1\}$ of leased facilities, a function $a : \mathcal{D} \mapsto X$ that maps each client request (j, t) to a facility leasing $(f, k, \hat{t}) \in X$ such that $t \in [\hat{t}, \hat{t} + \delta_k)$, and a set $\mathcal{T} \subseteq E_G$ of edges that connects the leased facilities in X to r , so to minimize $\sum_{(f,k,\hat{t}) \in X} \gamma_f^k + \sum_{(j,t) \in \mathcal{D}} d(a(j,t), j) + M \cdot \sum_{e \in \mathcal{T}} d(e)$.*

We suppose that $\delta_1 \leq \delta_2 \leq \dots \leq \delta_K$, and that $\gamma_f^k / \delta_k \leq \gamma_f^\ell / \delta_\ell$ for every $f \in F$ and $k > \ell$.

Online Algorithm. We propose the following online algorithm, based on the algorithm by [San Felice et al. 2016] for problem OCFL. Our algorithm uses the primal-dual algorithm by [Nagarajan and Williamson 2013] for problem OFLE, which we denote by NWFLLE and has offline parameters $(G, d, F, K, \gamma, \delta)$ and an online sequence of client sets. Algorithm NWFLLE assumes that facility leases occur only in specific moments; i.e., $X \subseteq \{(f, k, c \cdot \delta_k) : f \in F, k \in [K], c \in \mathbb{Z}_+\}$. This assumption increases the competitive factor of any algorithm for problem CFLE at most by a constant ratio.

Our algorithm maintains a set \mathcal{D} of client requests, and a set $\mathcal{D}^\mathcal{T}$ of clients that are connected to the core network. At each client request, the algorithm forwards the request to algorithm NWFLLE and updates the virtual solution. Then, the client is added to the core network with probability $1/M$. In that case, the algorithm adds to the core network the shortest path between the client and the tree, opens the facility indicated by the virtual solution and adds another core edge between the client and the new facility. If the client is not added to the core network, then it is served by the closest active facility lease. Note that we ensure that the root facility is always active at zero cost.

We have a proof that this algorithm is $O(K \cdot \lg n)$ -competitive for the special case when $M = 1$, where n is the total number of client requests.¹

¹The proof is available at <http://www.ime.usp.br/~mslima/appendix-etc2016.pdf>.

Algorithm OCFLE-ALG($G, d, r, F, K, \gamma, \delta, M$)

```

01  set  $\delta_r^K \leftarrow 0, X \leftarrow \emptyset, \mathcal{D} \leftarrow \emptyset, \mathcal{D}^T \leftarrow \emptyset, \mathcal{T} \leftarrow (\{r\}, \emptyset)$ 
02  initialize NWFLE with  $(G, d, F, K, \gamma, \delta)$ ; let  $(X', a')$  be the virtual solution
03  upon receipt of  $D_t$  do
04    if  $t \bmod \delta_K = 0$  then  $X \leftarrow X \cup \{(r, K, t)\}$ ; send  $(r, t)$  to NWFLE
05    for each  $j \in D_t$  do
06      send  $(j, t)$  to NWFLE and update the virtual solution  $(X', a')$ 
07      include  $j$  in  $\mathcal{D}^T$  with probability  $1/M$ 
08      if  $j \in \mathcal{D}^T$  then
09         $\mathcal{T} \leftarrow \mathcal{T} \cup \text{path}(j, V(T)), (f, k, \hat{t}) \leftarrow a'(j, t)$ 
10        if  $(f, k, \hat{t})$  is not opened then
11           $X \leftarrow X \cup \{(f, k, \hat{t})\}, \mathcal{T} \leftarrow \mathcal{T} \cup \{(f, j)\}$ 
12        choose  $(f, k, \hat{t}) \in X$  active in  $t$  and closest to  $j$ 
13         $\mathcal{D} \leftarrow \mathcal{D} \cup \{(j, t)\}, a(j, t) \leftarrow (f, k, \hat{t})$ 
14  return  $(X, a, \mathcal{T})$ 

```

3. Conclusion and Future Work

We propose a leasing variant of the online connected facility location problem, and give an algorithm which is $O(K \cdot \lg n)$ -competitive if $M = 1$. We wish to prove the same competitive factor for arbitrary M , maybe by modifying the algorithm so to take into account core edges when deciding to lease a facility; currently we have partial results.

References

- Abshoff, S., Kling, P., Markarian, C., auf der Heide, F. M., and Pietrzyk, P. (2015). Towards the price of leasing online. *Journal of Combinatorial Optimization*, pages 1–20.
- Anthony, B. M. and Gupta, A. (2007). Infrastructure leasing problems. In *Integer Programming and Combinatorial Optimization*, pages 424–438. Springer.
- Grandoni, F. and Rothvoß, T. (2011). Approximation algorithms for single and multi-commodity connected facility location. In Günlük, O. and Woeginger, G. J., editors, *Integer Programming and Combinatorial Optimization*, volume 6655 of *Lecture Notes in Computer Science*, pages 248–260. Springer Berlin Heidelberg.
- Imase, M. and Waxman, B. M. (1991). Dynamic Steiner tree problem. *SIAM Journal on Discrete Mathematics*, 4(3):369–384.
- Meyerson, A. (2005). The parking permit problem. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 274–282.
- Nagarajan, C. and Williamson, D. P. (2013). Offline and online facility leasing. *Discrete Optimization*, 10(4):361–370.
- San Felice, M. C., Williamson, D. P., and Lee, O. (2016). A randomized $O(\log n)$ -competitive algorithm for the online connected facility location. *Algorithmica*, pages 1–19.
- Umboh, S. (2015). Online network design algorithms via hierarchical decompositions. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1373–1387.

Sobre a minimização de transdutores sequenciais

Rodrigo de Souza^{1*}

¹DEINFO – Universidade Federal Rural de Pernambuco (UFRPE)
Dois Irmãos – 52.171-900 – Recife – PE – Brazil

rodrigo.npmsouza@ufrpe.br

Abstract. We study the minimization of sequential transducers with outputs in structures which we call monoids with gcd (“greatest common divisor”). We prove the existence of a minimal sequential transducer for a sequential function $\Sigma^* \rightarrow M$, where M is a cancellative monoid such that every non-empty subset has a single gcd. This result includes the minimization for outputs in a free monoid or in the additive monoid of the non-negative real numbers, among others. Next, we characterize the sequential functions $\Sigma^* \rightarrow M$, where M is a cancellative monoid with gcd, by using the right congruence of a function.

Resumo. Estudamos a minimização de transdutores sequenciais utilizando para as saídas uma família de monóides que chamamos de monóides com mdc. Provamos a existência de um transdutor minimal para funções sequenciais $\Sigma^* \rightarrow M$, onde M é um monóide cancelativo com mdc único. Esse resultado inclui diversos monóides de interesse, como os monóides livres, e o monóide aditivo dos números reais não-negativos. Também apresentamos uma caracterização das funções sequenciais $\Sigma^* \rightarrow M$, onde M é um monóide cancelativo com mdc, utilizando a congruência à direita de uma função.

1. Introdução

Os autômatos finitos são tão essenciais na Ciência da Computação quanto os espaços topológicos na Análise e os grafos na Combinatória, por representarem a noção de *computação* em sua forma mais límpida. Em consequência, a disciplina hoje denominada Teoria dos Autômatos firmou-se como um dos pilares da Ciência da Computação.

Em sua definição mais simples, *transdutores*¹ são uma extensão do modelo de autômato no qual as transições são rotuladas por um par de palavras, uma dita *de entrada* e outra *de saída*. Ao invés de reconhecer linguagens, os transdutores realizam *relações* entre linguagens, chamadas *transduções* em algumas referências (especialmente americanas), ou ainda *relações racionais* (especialmente pelos autores filiados à escola francesa). A definição formal repousa no produto cartesiano de dois monóides livres Σ^* e Γ^* , $\Sigma^* \times \Gamma^*$, um monóide cuja operação é a aplicação das concatenações de Σ^* e Γ^* nas coordenadas respectivas. De forma mais geral, um monóide qualquer, como os números reais com a operação de soma, pode ser utilizado no lugar do monóide de saída Γ^* .

*Este trabalho é apoiado pelo projeto *Problemas estruturais em modelos formais de Computação*, Edital MCTI/CNPQ/Universal 14/2014 (Processo 459957/2014-7).

¹Por limitação de espaço, não apresentamos neste resumo definições e notações básicas sobre autômatos e transdutores. Esse material pode ser consultado no livro de Jacques Sakarovitch [Sakarovitch 2009].

Versões restritas de transdutores, as chamadas máquinas de Mealy e Moore, estão entre os primeiros tipos de autômatos da literatura, tendo sido propostos na década de 1960. Desde então, os transdutores se firmaram como uma das pedras angulares da Teoria dos Autômatos, citando aqui o compêndio clássico de Samuel Eilenberg [Eilenberg 1974]. Transdutores também tem muitas aplicações práticas. Por exemplo, algoritmos para determinização, minimização e composição de transdutores são extensivamente utilizados em sistemas de tratamento de linguagem natural [Benesty et al. 2008]. Um aspecto importante dessas ferramentas é sua sustentação em propriedades profundas de transdutores cujo estudo remonta à década de 1970 [Choffrut 1979] mas que foram desde então revisitadas através de novas abordagens [Béal et al. 2003, Choffrut 2003]. Um panorama atualizado da extensão da teoria dos transdutores pode ser consultado em [Sakarovitch 2009].

As relações racionais que são funções parciais – ou *funções racionais* – tiveram uma posição predominante nesse contexto devido a propriedades notáveis dessa família (decidibilidade, não-ambiguidade), e em particular devido ao interesse teórico e prático envolvido no conceito de determinismo (na entrada). Contrariamente ao caso clássico dos autômatos sobre um monóide livre, nem todo transdutor funcional (= realiza uma função) pode ser determinizado; a caracterização das funções racionais que admitem uma leitura determinística, ou *sequencial*, em nossa terminologia, é um resultado profundo, obtido por Choffrut em 1979 [Choffrut 1979]. Aquelas que podem ser realizadas por um *transdutor sequencial*² – as *funções sequenciais* – são importantes em aplicações práticas.

O presente resumo aborda o conceito de *minimização de transdutores sequenciais*, que estende os resultados clássicos sobre autômatos determinísticos. O estudo de um transdutor minimal que realiza uma função sequencial aparece implicitamente em trabalhos da década de 1970 de Choffrut [Choffrut 1979], e explicitamente em trabalhos posteriores de Reutenauer [Reutenauer 1990] e Choffrut [Choffrut 2003]. Nesses dois últimos, são consideradas funções com saídas em um monóide livre. Ao mesmo tempo, um conceito de transdutor minimal para funções sequenciais com saídas em outros tipos de monóides, como o monóide dos números reais com a operação de adição, também foi estudado por Mohri e colaboradores, em conexão com aplicações em tratamento de linguagem natural [Mohri 2000]. A abordagem ao assunto não é uniforme na literatura: de um lado, constituiu-se uma noção de congruência sintática para funções, semelhante à congruência de Nerode para linguagens, que permite caracterizar algebricamente o transdutor minimal para uma função *com saídas em um monóide livre*. Como ocorre com os autômatos, essa máquina minimal tem o menor número de estados dentre todas aquelas equivalentes, o que pode ser observado através de *morfismos entre transdutores*. Esse tratamento, utilizado por Eilenberg no estudo da minimização de autômatos, aparece superficialmente no trabalho de Reutenauer, e explicitamente no trabalho de Choffrut de 2003. Nessa abordagem, o primeiro passo é definir o transdutor sequencial minimal associado a uma função, de maneira puramente algébrica. Em seguida, utilizando morfismos de transdutores sequenciais, demonstra-se a minimalidade desse transdutor. No trabalho de Mohri, a abordagem é algorítmica: o objetivo é descrever a construção de um transdutor minimal, tanto para emissões em um monóide livre, quanto em um monóide numérico.

Nossa intenção é apresentar um formalismo permitindo descrever o conceito de

²Preferimos o termo *sequencial* a *subsequencial*, usado originalmente. Uma discussão sobre a terminologia (e algoritmos para decidir a sequencialidade) é feita em [Lombardy and Sakarovitch 2006].

autômato minimal de forma unificada, considerando em uma única linguagem diversos monóides já abordados na literatura. Para tanto, definimos uma família de monóides, que chamamos de *monóides com mdc*, que abstrai as propriedades essenciais do monóide de saída para a definição de um transdutor minimal. Nas próximas seções, esboçamos nossos principais resultados, e fazemos referência à nossa dissertação de mestrado para uma discussão mais extensa sobre o assunto e as demonstrações completas [de Souza 2004].

2. Transdutores sequenciais minimais

Seja M um monóide. Dados elementos a e b em M , escrevemos $a|b$ se existir um elemento b' em M tal que $b = ab'$. Essa relação de “divisibilidade” é transitiva, reflexiva, mas não necessariamente anti-simétrica (considere o caso de um grupo). Dado um subconjunto não-vazio S de M , dizemos que um elemento $m \in M$ é um *máximo divisor comum*, ou *mdc*, de S , se $m|a$, para todo $a \in S$, e para cada $n \in M$, se $n|a$, para todo $a \in S$, então $n|m$. Dizemos que M é um monóide com *mdc* se todo subconjunto não-vazio de M tiver pelo menos um *mdc*. Interessamos o caso de o *mdc* de todo subconjunto não-vazio de M ser único. Dizemos nesse caso que M é um *monóide com mdc único*. Exemplos dessa estrutura são: monóides livres (o *mdc* é a operação de maior prefixo comum de uma linguagem); monóides livres comutativos; o monóide (\mathbb{Z}, \cdot) dos inteiros positivos, com a operação de produto (com o *mdc* usual); o monóide $(\mathbb{R}_+, +)$ dos números reais não-negativos, com a operação de adição (o *mdc* é o ínfimo de um conjunto). Intuitivamente, o *mdc* abstrai a extração do maior prefixo comum de uma linguagem, operação essencial na definição do transdutor minimal com emissões em um monóide livre.

O escopo de nossa definição de transdutor minimal é o das funções sequenciais $f : \Sigma^* \rightarrow M$, onde Σ é um alfabeto, e M é um monóide cancelativo com *mdc* único. A teoria aqui desenvolvida é então uma extensão do trabalho de Choffrut de 2003, no qual o monóide de saída é sempre um monóide livre [Choffrut 2003]. Ademais, Choffrut define o transdutor minimal utilizando a congruência sintática de uma função. Fazemos uma construção direta do transdutor minimal a partir de f , seguindo as ideias desenvolvidas nas seções III.3 e XII.5 do livro de Eilenberg [Eilenberg 1974].

Em linhas gerais, a construção do transdutor sequencial minimal para f , que denotamos \mathcal{T}_f , parte do conjunto Q^0 de todas as funções parciais não-vazias $g : \Sigma^* \rightarrow M$ tais que $\text{mdc } g = 1$ (onde 1 é a unidade do monóide; para o monóide livre, é a palavra vazia). Esse conjunto é análogo ao conjunto potência de Σ^* na definição do autômato minimal para uma linguagem sobre Σ (mas note a restrição imposta sobre o *mdc*). De forma puramente algébrica, definimos então uma ação \cdot de Σ^* sobre Q^0 , e uma emissão $*$ de Σ^* sobre Q^0 em M (que correspondem às regras de transição e emissão do transdutor). As restrições de \cdot e $*$ a uma parte especial de Q^0 constituem o transdutor \mathcal{T}_f . A propriedade mais relevante dessa construção, que garante a minimalidade de \mathcal{T}_f , é:

Teorema 2.1 *Seja $f : \Sigma^* \rightarrow M$ uma função sequencial, onde M é um monóide cancelativo com mdc único. Para todo transdutor sequencial bi-acessível \mathcal{T} que realiza f , existe um morfismo próprio $\varphi : \mathcal{T} \rightarrow \mathcal{T}_f$.*

Como $\varphi : \mathcal{T} \rightarrow \mathcal{T}_f$ é uma função sobrejetora, $n_f \leq n$, onde n_f e n são os números de estados de \mathcal{T}_f e \mathcal{T} , respectivamente. Essa propriedade não apenas mostra que \mathcal{T}_f é finito, se f for sequencial, mas também que \mathcal{T}_f tem o menor número de estados dentre todos os transdutores que realizam f . Ademais, é fácil ver que, se \mathcal{T} tiver o mesmo número de estados de \mathcal{T}_f e for prefixo, então \mathcal{T} e \mathcal{T}_f são isomorfos.

3. Congruência de uma função e redução

Nosso estudo sobre o transdutor minimal com emissões em um monóide cancelativo com mdc único completa-se com os seguintes resultados: definimos uma congruência \sim_f de uma função entre monóides, que estende o conceito de congruência à direita, ou congruência de Nerode, de uma linguagem; descrevemos um conceito de equivalência \equiv de estados para transdutores sequenciais, que estende o conceito correspondente para autômatos finitos. Com essa ideia, é possível construir um transdutor minimal a partir de uma partição do conjunto de estados de um dado transdutor sequencial. Em se tratando do monóide livre, o conceito da congruência à direita de uma função apareceu no trabalho fundamental de Schützenberger sobre funções sequenciais [Schützenberger 1977]; os conceitos de estados equivalentes e transdutor quociente estão em [Reutenauer 1990]. Os resultados relevantes nesse contexto são:

Teorema 3.1 *Sejam Σ um alfabeto, e M um monóide cancelativo com mdc único. Uma função $f : \Sigma^* \rightarrow M$ é seqüencial se, e somente se, o quociente Σ^* / \sim_f é finito.*

Teorema 3.2 *Seja \mathcal{T} um transdutor sequencial bi-acessível e prefixo que realiza uma função sequencial $f : \Sigma^* \rightarrow M$. Então, $\mathcal{T} / \equiv \cong \mathcal{T}_f$.*

Referências

- Béal, M. P., Carton, O., Prieur, C., and Sakarovitch, J. (2003). Squaring transducers: an efficient procedure for deciding functionality and sequentiality. *Theoretical Computer Science*, 292(1):45–63.
- Benesty, J., Sondhi, M., and Huang, Y., editors (2008). *Springer Handbook of Speech Processing*. Springer-Verlag Berlin Heidelberg.
- Choffrut, C. (1979). A generalization of Ginsburg and Rose’s characterization of gsm mappings. In *ICALP’79*, volume 71 of *Lecture Notes in Computer Science*, pages 88–103. Springer-Verlag.
- Choffrut, C. (2003). Minimizing subsequential transducers: A survey. *Theoretical Computer Science*, 292(1):131–143.
- de Souza, R. (2004). Propriedades de algumas classes de relações racionais. Master’s thesis, Instituto de Matemática e Estatística, Universidade de São Paulo.
- Eilenberg, S. (1974). *Automata, Languages, and Machines*. Academic Press.
- Lombardy, S. and Sakarovitch, J. (2006). Sequential? *Theoretical Computer Science*, 356(1-2):224–244.
- Mohri, M. (2000). Minimization algorithms for sequential transducers. *Theoretical Computer Science*, 234(1-2):177–201.
- Reutenauer, C. (1990). Subsequential functions: Characterizations, minimization, examples. In *International Meeting of Young Computer Scientists*, volume 464 of *Lecture Notes in Computer Science*, pages 62–79.
- Sakarovitch, J. (2009). *Elements of Automata Theory*. Cambridge University Press.
- Schützenberger, M. P. (1977). Sur une variante des fonctions sequentielles. *Theoretical Computer Science*, 4(1):47–57.

Quantificando Vazamento de Informação sobre Estratégias

Mário S. Alvim¹, Piotr Mardziel², Michael Hicks²

¹ Universidade Federal de Minas Gerais
msalvim@dcc.ufmg.br

²University of Maryland, College Park
{piotrm,mwh}@cs.umd.edu

Abstract. *The field of quantitative information flow concerns the rigorous mathematical assessment of the amount of secret information leaked by computational systems. We report first steps towards a formal model for strategic leakage. We generalize the representation of prior adversarial knowledge from a distribution on secrets to a distribution on strategies for generating secrets, which we call an environment. Applying information-theoretic techniques to environments allows us to disentangle information leakage about a secret from leakage about how users generate secrets, i.e., their strategy.*

Resumo. *Este resumo reporta primeiros passos na formalização do conceito de vazamento de estratégias. Nós generalizamos a representação do conhecimento prévio de um adversário de uma distribuição de probabilidade sobre os segredos para uma distribuição sobre estratégias que geram segredos, chamando esta nova distribuição de ambiente. Aplicando técnicas de teoria da informação a ambientes, conseguimos distinguir o vazamento de informação sobre segredos do vazamento de informação sobre a maneira como segredos são gerados, i.e., sobre a estratégia.*

1. Introdução

O campo de *fluxo de informação quantitativo* (QIF, do inglês *quantitative information flow*) diz respeito à mensuração da informação sigilosa vazada por sistemas computacionais. Baseando-se no rigor matemático da teoria da informação, dois princípios fundamentais de QIF são: (i) um segredo é considerado “seguro” na medida em a distribuição de probabilidade sobre os valores sigilosos tem alta entropia; e (ii) o vazamento de informação de um sistema é uma medida de quanto o comportamento observável do sistema, durante o processamento de um valor sigiloso, degrada a entropia desse segredo. Estes princípios têm sido usados para criar modelos de QIF cada vez mais sofisticados para representar sistemas e medir vazamentos. No entanto, pouca atenção tem sido dedicada a entender as fontes que injetam a entropia inicial na distribuição sobre os segredos.

Muitas vezes, o processo de geração de segredos não é completamente aleatório. Por exemplo, usuários podem ter uma “estratégia” para escolher senhas que seja parcialmente previsível (por exemplo, usando sua data de nascimento), e um adversário pode usar isso para prever senhas futuras. Estudos têm mostrado que forçar usuários a alterar frequentemente suas senhas pode não ser tão benéfico quanto se acreditava [Zhang et al. 2010, Chiasson and van Oorschot 2015]. Além disso, um segredo que varia no tempo e sofre repetidas observações, em alguns casos, pode ser aprendido *mais rapidamente* se for alterado (e observado) mais frequentemente [Mardziel et al. 2014].

Neste resumo investigamos em que medida o conhecimento sobre *estratégias* ajuda o adversário a inferir o segredo. Desenvolvemos um modelo que não considera apenas o espaço de segredos possíveis, mas também o espaço de possíveis estratégias que os geram. Nosso modelo nos permite considerar adversários capazes de aprender a estratégia e quantificar a vantagem que eles obtêm, desacoplando a medida do valor de se aprender a estratégia da medida do valor de se aprender o segredo. No resto deste resumo, apresentamos o nosso modelo inicial e caracterizações de vazamento de estratégias.

2. Preliminares

Faremos uma breve revisão de conceitos fundamentais de *fluxo de informação quantitativo* (QIF). Um *adversário* tem apenas informação parcial sobre o valor de um *segredo*, modelada por uma distribuição de probabilidade chamada de *prior*. Denotamos por \mathcal{X} o conjunto de segredos possíveis e por $\mathbb{D}\mathcal{X}$ o conjunto de distribuições de probabilidade sobre \mathcal{X} . Normalmente usamos p_X para denotar uma *prior*. Uma medida de informação é uma função $\mathbb{V}:\mathbb{D}\mathcal{X}\rightarrow\mathbb{R}$ com imagem real. Uma medida de informação pode medir *vulnerabilidade* - quanto maior o valor, menos seguro é o segredo - ou *incerteza/entropia* - quanto maior o valor, mais seguro é o segredo. Definições de medidas de informação na literatura variam de acordo com a interpretação operacional da medida. Recentemente, o arcabouço de *g-vulnerability* [Alvim et al. 2012] foi proposto para capturar vários modelos de adversário. Neste artigo vamos utilizar o termo “vulnerabilidade” para medidas em geral, embora as nossas reivindicações também se aplicam a incerteza medidas. Uma *hiper-distribuição* [McIver et al. 2014] (ou *hiper*) é uma distribuição de probabilidade sobre distribuições. Uma hiper sobre o conjunto \mathcal{X} é do tipo $\mathbb{D}^2\mathcal{X}$, significando $\mathbb{D}(\mathbb{D}\mathcal{X})$. Usamos Δ para denotar uma hiper, $[\Delta]$ para seu *suporte* (o conjunto de valores com probabilidade não nula), e $[p_X]$ para hiper puntual associando probabilidade 1 para p_X .

3. Estratégias e sua vulnerabilidade

Esta seção apresenta um modelo que explicita que segredos são gerados pela aplicação de uma estratégia particular, e apresenta generalizações de vulnerabilidade correspondentes.

3.1. Vulnerabilidade contextual do segredo

Uma *estratégia* é uma regra probabilística para a geração de segredos, representada por uma distribuição de probabilidade sobre o conjunto \mathcal{X} de segredos. O conjunto \mathcal{S} de todas as estratégias é, portanto, $\mathbb{D}\mathcal{X}$. Modelamos o *ambiente* em que o segredo é gerado como uma distribuição p_S sobre o conjunto \mathcal{S} de estratégias. O conjunto $\mathbb{D}\mathcal{S}$ de todos os ambientes é o conjunto das distribuições de probabilidade sobre estratégias, ou seja, o conjunto $\mathbb{D}^2\mathcal{X}$ de hipers sobre \mathcal{X} . O ambiente representa o espaço de onde estratégias são amostrados. Assumimos que a *prior* p_X sobre segredos é *consistente* com o ambiente p_S , isto é, que a *prior* é obtida como o comportamento esperado do ambiente na falta qualquer conhecimento mais refinado sobre estratégia utilizada: $p_X = \mathbb{E} p_S$. Caso haja uma única estratégia p_X para gerar segredos, o ambiente p_S é a hiper puntual $[p_X]$.

Enquanto em modelos tradicionais de QIF o adversário só sabe a *prior* p_X , consideramos aqui também adversários que conhecem o ambiente p_S . A vulnerabilidade do segredo para este tipo adversário mais poderoso é definida a seguir.

Definition 1. Dada uma medida de informação \mathbb{V} , a *vulnerabilidade contextual (do segredo)* é uma função $\mathbb{V}_C:\mathbb{D}^2\mathcal{X}\rightarrow\mathbb{R}$ definida como $\mathbb{V}_C(p_S) \stackrel{\text{def}}{=} \mathbb{E}_{p_S} \mathbb{V}$ representando a expectativa da vulnerabilidade do segredo quando o ambiente é conhecido pelo adversário.

Nosso primeiro resultado é que um adversário que conhece o ambiente nunca está em desvantagem em relação a um adversário que só conhece a *prior* sobre os segredos.

Proposition 1. *Se \mathbb{V} é uma g -vulnerability, então $\mathbb{V}_C(p_S) \geq \mathbb{V}(p_X)$.*

Além disso, no caso de o ambiente p_S ser uma hiper puntual $[p_X]$, a vulnerabilidade contextual $\mathbb{V}_C(p_S)$ se reduz a $\mathbb{V}(p_X)$.

Proposition 2. *Se $p_S = [p_X]$, então $\mathbb{V}_C(p_S) = \mathbb{V}(p_X)$.*

3.2. Segurança real e segurança por agregação

A modelagem do conhecimento do adversário como apenas uma *prior* p_X ignora como o adversário pode explorar o conhecimento do ambiente p_S para inferir segredos. Demonstramos este fato com um exemplo em que segredos distribuídos de acordo com uma mesma *prior* apresentam vulnerabilidades contextuais drasticamente diferentes.

Sejam $\mathcal{X} = \{x_1, x_2\}$ um conjunto de valores secretos binários, e $\mathcal{S} = \{s_1, s_2, s_3\}$, um conjunto de possíveis estratégias, onde $s_1 = [1, 0]$ sempre gera o valor x_1 , $s_2 = [0, 1]$ sempre gera o valor x_2 e $s_3 = [1/2, 1/2]$ gera cada valor com igual probabilidade. Considere dois ambientes $p_S^1 = [1/2, 1/2, 0]$, em que as estratégias s_1 e s_2 pode ser adotadas com igual probabilidade,

e $p_S^2 = [0, 0, 1]$, em que a estratégia s_3 é sempre adotada. Retratamos tais estratégias e ambientes na Tabela 1. Colunas listam estratégias; o primeiro agrupamento de linhas contém a definição da estratégia (isto é, a probabilidade de cada segredo ser escolhido), o agrupamento seguinte de linhas contém a definição de cada ambiente, um por linha, fornecendo a probabilidade de cada estratégia. Ambos os ambientes produzem a mesma *prior* $p_X = \mathbb{E} p_S^1 = \mathbb{E} p_S^2 = [1/2, 1/2]$, assim um adversário que não pode observar a estratégia utilizada obterá a mesma vulnerabilidade prévia em ambos os ambientes. Por exemplo, para a vulnerabilidade de Bayes $\mathbb{V}^{\text{Bayes}}(p_X) \stackrel{\text{def}}{=} \max_{x \in \mathcal{X}} p_X(x)$, o adversário obterá uma vulnerabilidade de $\mathbb{V}^{\text{Bayes}}(p_X) = 1/2$. No entanto, um adversário que possa aprender a estratégia utilizada obterá diferentes vulnerabilidades do segredo em cada ambiente. No ambiente p_S^1 a vulnerabilidade contextual é $\mathbb{V}_C^{\text{Bayes}}(p_S^1) = 1/2 \cdot p(s_1) + 1/2 \cdot p(s_2) = 1/2 \cdot 1 + 1/2 \cdot 1 = 1$, enquanto no ambiente p_S^2 a vulnerabilidade contextual é $\mathbb{V}_C^{\text{Bayes}}(p_S^2) = 1 \cdot p(s_3) = 1 \cdot 1/2 = 1/2$. Note que no ambiente p_S^2 , o valor para a vulnerabilidade contextual e vulnerabilidade prévia é o mesmo ($\mathbb{V}_C^{\text{Bayes}}(p_S^2) = \mathbb{V}^{\text{Bayes}}(p_X) = 1/2$), assim um adversário que aprende a estratégia utilizada não terá vantagem sobre um adversário que só conhece a *prior*. Por outro lado, no ambiente p_S^1 , a vulnerabilidade contextual excede a vulnerabilidade prévia ($\mathbb{V}_C^{\text{Bayes}}(p_S^1) = 1 > 1/2 = \mathbb{V}^{\text{Bayes}}(p_X)$), e um adversário que aprende a estratégia utilizada obterá sucesso duas vezes mais frequentemente que um adversário que só conhece a *prior*.

	s_1	s_2	s_3
x_1	1	0	1/2
x_2	0	1	1/2
p_S^1	1/2e1/2e0		
p_S^2	0	0	1

Tabela 1.

Podemos obter dois *insights* a partir deste exemplo. Em primeiro lugar, *segurança por agregação* ocorre quando a vulnerabilidade contextual excede a vulnerabilidade prévia: $\mathbb{V}_C(p_S) \gg \mathbb{V}(p_X)$. Neste caso, o segredo está protegido pela falta de conhecimento por parte do adversário sobre a estratégia utilizada, e se o adversário aprender a estratégia a vulnerabilidade do segredo pode aumentar significativamente. Um exemplo de segurança por agregação é um cenário em que todos usuários escolhem senhas com estratégias determinísticas, mas o adversário não sabe qual usuário está gerando a a senha. Se houver um grande número de usuários, e se são as suas estratégias são variada o suficiente, as senhas podem ser considerados “seguras” apenas no sentido de que o adversário não pode usar o conhecimento sobre o ambiente para identificar a estratégia utilizada.

Por outro lado, *segurança real* ocorre quando vulnerabilidades contextual e prévia têm valores semelhantes: $\mathbb{V}_C(p_S) \approx \mathbb{V}(p_X)$. Neste caso, o segredo está protegido pela imprevisibilidade (ou incerteza) no âmbito das estratégias que geram o segredo, e mesmo que a estratégia se torne conhecida pelo adversário, a vulnerabilidade do segredo não irá aumentar significativamente. Um exemplo de segurança real é um sistema de banco em que senhas são escolhidas automaticamente para cada usuário de modo uniforme. Mesmo que o algoritmo seja conhecido do adversário, a vulnerabilidade do segredo não aumenta.

3.3. Vulnerabilidade da estratégia

Consideramos agora como o conhecimento de um ambiente afeta o conhecimento do adversário sobre a estratégia a ser utilizada, definindo uma medida $\mathbb{V}_S(p_S): \mathbb{DS} \rightarrow \mathbb{R}$ da *vulnerabilidade da estratégia*. A ideia central é que $\mathbb{V}_S(p_S)$ deve refletir a “semelhança” entre as estratégias no suporte de p_S . Do ponto de vista do adversário, duas estratégias devem ser consideradas “semelhantes” se levarem a “semelhantes” vulnerabilidades do segredo, conforme medido de acordo com uma $\mathbb{V}: \mathbb{DX} \rightarrow \mathbb{R}$ de interesse. Assim, a vulnerabilidade de um ambiente deve ser mais elevada quando o seu suporte consistir em estratégias semelhantes, e mais baixa quando seu suporte consistir em estratégias muito diferentes. Uma estratégia é, *para fins práticos*, conhecida pelo adversário quando $\mathbb{V}(p_X) \approx \mathbb{V}_C(p_S)$, ou, equivalente, quando $\mathbb{V}(\mathbb{E} p_S) \approx \mathbb{E}_{p_S} \mathbb{V}$.

Definition 2. Se \mathbb{V} é uma medida de vulnerabilidade, então a *vulnerabilidade da estratégia* é a relação $\mathbb{V}_S(p_S) \stackrel{\text{def}}{=} \mathbb{V}(\mathbb{E} p_S) / \mathbb{V}_C(p_S)$. A razão é invertida para medidas de incerteza.

Note que temos sempre $0 \leq \mathbb{V}_S(p_S) \leq 1$ (pela Proposição 1), e $\mathbb{V}_S(p_S)$ é máxima quando $\mathbb{V}(\mathbb{E} p_S) = \mathbb{V}_C(p_S)$. Além disso, note que a definição é consistente com a de decomposição vulnerabilidade prévia no produto de vulnerabilidade estratégica e vulnerabilidade contextual: $\mathbb{V}(p_X) = \mathbb{V}_S(p_S) \cdot \mathbb{V}_C(p_S)$. Este é o desacoplamento que buscávamos.

4. Conclusão

Atualmente estamos trabalhando na caracterização de níveis intermediários de conhecimento do adversário, em que seu conhecimento é mais refinado do que uma *prior*, mas não tão refinado quanto o conhecimento completo do ambiente.

Referências

- [Alvim et al. 2012] Alvim, M. S., Chatzikokolakis, K., Palamidessi, C., and Smith, G. (2012). Measuring information leakage using generalized gain functions. In *Proceedings of the IEEE Computer Security Foundations Symposium (CSF)*.
- [Chiasson and van Oorschot 2015] Chiasson, S. and van Oorschot, P. C. (2015). Quantifying the security advantage of password expiration policies. *Journal of Designs, Codes, and Cryptography*, 77(2-3):401–408.
- [Mardziel et al. 2014] Mardziel, P., Alvim, M. S., Hicks, M., and Clarkson, M. (2014). Quantifying information flow for dynamic secrets. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*.
- [McIver et al. 2014] McIver, A., Meinicke, L., and Morgan, C. (2014). Compositional closure for Bayes risk in probabilistic noninterference. In *Proc. ICALP'10*.
- [Zhang et al. 2010] Zhang, Y., Monrose, F., and Reiter, M. K. (2010). The security of modern password expiration: an algorithmic framework and empirical analysis. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS)*.

Compartilhamento de Custos de Empacotamento*

Flávio K. Miyazawa¹, Rafael C. S. Schouery¹

¹Instituto de Computação - Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, CEP 13083-852, Campinas/SP - Brasil

{fkm, schouery}@ic.unicamp.br

Abstract. *In the bin packing problem, one desires to pack a set of items in bins while respecting the bins' capacities with the objective of minimizing the number of bins used. In this paper, we consider the problem of sharing the cost of packing between the participant agents.*

Resumo. *No Problema do Empacotamento, queremos empacotar um conjunto de itens em recipientes de forma a respeitar a capacidade dos recipientes e a minimizar o número de recipientes usados. Neste artigo, abordamos o problema de compartilhar o custo do empacotamento entre os agentes participantes.*

1. Introdução

No *Problema do Empacotamento*, queremos empacotar um conjunto I de n itens (onde a_i é o tamanho do item i) em *recipientes* de capacidade 1 de forma a respeitar a capacidade dos recipientes e a minimizar o número de recipientes usados. Aplicações do problema incluem armazenamento de cargas, escalonamento de tarefas computacionais, etc. Em muitos desses cenários, é natural considerar a participação de diversos agentes colaborando para empacotar seus itens, já que, de tal forma, é possível utilizar uma quantidade menor de recipientes do que se os agentes empacotassem seus itens individualmente.

Neste artigo consideramos o Problema do Empacotamento do ponto de vista da Teoria dos Jogos Cooperativos, considerando que cada item é controlado por um jogador diferente. De forma geral, considere que um conjunto \mathcal{A} de jogadores desejam cooperar para realizar uma tarefa e que, se um conjunto $S \subseteq \mathcal{A}$ (uma *coalizão*) de jogadores cooperarem, então eles conseguem realizar a tarefa com um custo $c(S)$ (no nosso caso, o número de recipientes necessários para empacotar os itens do conjunto S). Assim, desejamos compartilhar o custo $c(\mathcal{A})$ entre os jogadores definindo um preço p_i para cada jogador $i \in \mathcal{A}$. Dizemos que a precificação $p: \mathcal{A} \rightarrow \mathbb{R}$ pertence ao *núcleo* se ela satisfaz as seguintes propriedades: $\sum_{i \in \mathcal{A}} p_i = c(\mathcal{A})$ (*orçamento balanceado*) e $\sum_{i \in S} p_i \leq c(S)$ para todo $S \subseteq \mathcal{A}$ (*propriedade do núcleo*). Isto é, os preços pagos pelos jogadores é igual ao custo de realizar a tarefa e, para qualquer coalizão S de jogadores, não é vantajoso para os jogadores em S empacotarem seus itens eles mesmos ao invés de realizar a tarefa com a *grande coalizão* \mathcal{A} dos jogadores.

Porém, para muitos jogos e, em particular, para muitas instâncias do Problema do Empacotamento, o núcleo é vazio. Assim, duas relaxações são consideradas na literatura. No *núcleo ε -taxado*, exigimos o orçamento balanceado mas relaxamos a propriedade do núcleo, exigindo apenas que $\sum_{i \in S} p_i \leq (1 + \varepsilon)c(S)$ para todo $S \subseteq \mathcal{A}$. Isto é,

*Parcialmente financiado pelo processo nº 2013/21744-8, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e processo nº 311499/2014-7, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq.)

consideramos que uma coalizão S teria uma taxação adicional de $\varepsilon c(S)$ no custo de realizar a tarefa apenas entre eles. Já no *núcleo γ -aproximado*, exigimos a propriedade do núcleo, mas relaxamos a exigência do orçamento balanceado, exigindo apenas que $\gamma c(\mathcal{A}) \leq \sum_{i \in \mathcal{A}} p_i \leq c(\mathcal{A})$, isto é, coletamos pagamentos dos jogadores de forma a cobrir pelo menos uma fração γ do custo de realizar a tarefa. Neste artigo estamos interessados no conceito de núcleo γ -aproximado para o Problema do Empacotamento.

Trabalhos Relacionados [Faigle and Kern 1993] consideraram uma versão do Problema da Mochila Múltipla (porém, o denominaram como Problema do Empacotamento) e mostraram que, se todo item cabe em qualquer um dos recipientes, então o núcleo $1/2$ -taxado é não-vazio. O resultado mais recente para tal problema é de [Qiu and Kern 2016] que consideram o caso onde todas as mochilas têm a mesma capacidade e mostraram que o núcleo $1/4$ -taxado é não-vazio para toda instância do problema. Apesar da semelhança com o nosso trabalho, tais resultados consideram um problema e um conceito de relaxação de núcleo diferente dos considerados nesse artigo.

2. Resultados para o núcleo γ -aproximado

Teorema 2.1. Para todo $0 < \varepsilon \leq 1/2$, existe uma instância do Problema do Empacotamento com núcleo $(1/2 + \varepsilon)$ -aproximado vazio.

Demonstração. Seja $k > 1/(2\varepsilon)$ um inteiro e considere a instância onde temos $k + 1$ itens de tamanho $1/k$. Note que $c(I) = 2$ e que, para toda coalizão $S \subsetneq I$, $c(S) = 1$. Portanto, para toda coalizão S de k itens, temos que $\sum_{i \in S} p_i \leq 1$. Somando sobre toda possível escolha de S , temos que $\sum_{i \in I} \binom{k}{k-1} p_i \leq \binom{k+1}{k}$. Portanto, $\sum_{i \in I} p_i \leq \frac{k+1}{k} = \left(\frac{1}{2} + \frac{1}{2k}\right) c(I) < \left(\frac{1}{2} + \varepsilon\right) c(I)$, de onde segue que nenhuma precificação pertence ao núcleo $(1/2 + \varepsilon)$ -aproximado. \square

Seja \mathcal{A} um conjunto de jogadores e $c : 2^{\mathcal{A}} \rightarrow \mathbb{R}$ uma função de custo, um *esquema de compartilhamento de custos* é uma função $\xi : \mathcal{A} \times 2^{\mathcal{A}} \rightarrow \mathbb{R}$ tal que, para todo $S \subseteq \mathcal{A}$ e todo $i \notin S$, $\xi(i, S) = 0$. Dizemos que ξ é *γ -orçamento balanceado* se, para todo conjunto $S \subseteq \mathcal{A}$, temos que $\gamma c(S) \leq \sum_{i \in S} \xi(i, S) \leq c(S)$ e que ξ é *monotônico cruzado* se, para todo $S, T \subseteq \mathcal{A}$ e $i \in S$, $\xi(i, S) \geq \xi(i, S \cup T)$.

A existência de um esquema de compartilhamento de custos monotônico cruzado γ -orçamento balanceado ξ implica na existência de uma precificação no núcleo γ -aproximado (basta considerar a precificação dada por $\xi(\cdot, \mathcal{A})$).

Além disso, no cenário onde um leiloeiro decide quais jogadores terão seus itens empacotados e o preço p_i a ser pago por cada jogador i de acordo com os lances submetido pelos jogadores, sendo que cada jogador i obtém um valor v_i por ter o seu item empacotado, obtendo utilidade $v_i q_i - p_i$ onde q_i é uma variável indicadora se i é escolhido ou não, um resultado de [Moulin 1999] mostra que a existência de tal ξ implica na existência de um mecanismo à prova de estratégia de grupos γ -orçamento balanceado. Isto é, se o conjunto vencedor é S , então $\gamma c(S) \leq \sum_{i \in S} p_i \leq c(S)$ e nenhuma coalizão, ao relatar lances diferentes de seus reais valores dados por v , é tal que pelo menos um jogador obtém uma utilidade maior do que a obtida ao relatar o seu real valor e os outros jogadores obtém pelo menos a mesma utilidade do que a obtida ao relatar o seu real valor.

Teorema 2.2. Existe um esquema de compartilhamento de custos monotônico cruzado 1/2-orçamento balanceado para o Problema do Empacotamento.

Demonstração. Considere o esquema de compartilhamento de custos ξ onde, para um conjunto $S \subseteq I$ e $i \in S$, temos que se $\sum_{i \in S} a_i \leq 1$, então $\xi(i, S) = a_i / (\sum_{j \in S} a_j)$ e, caso contrário, $\xi(i, S) = a_i$. Como é fácil ver que ξ é monotônico cruzado, basta então provar que ξ é 1/2-orçamento balanceado.

Para um conjunto S , se $c(S) = 1$, então $\sum_{i \in S} \xi(i, S) = 1$. Se $c(S) \geq 2$, considere uma solução ótima do Problema do Empacotamento com instância S e seja δ o peso mínimo de um recipiente em tal solução. Se $\delta \geq 1/2$, então $\sum_{i \in S} \xi(i, S) \geq c(S)/2$. Caso contrário, então todo recipiente exceto o de peso mínimo tem peso maior do que $1 - \delta$ e, portanto, $\sum_{i \in S} \xi(i, S) = \sum_{i \in S} a_i \geq (1 - \delta)(c(S) - 1) + \delta \geq \frac{1}{2}(c(S) - 1) + \frac{1}{2} = c(S)/2$. Por outro lado, é claro que $\sum_{i \in S} \xi(i, S) \leq c(S)$, já que as somas dos pesos dos itens de uma instância do Problema do Empacotamento é um limitante inferior para o custo de uma solução ótima. Assim, ξ é 1/2-orçamento balanceado. \square

3. Resultados para o núcleo γ -aproximado assintótico

No projeto de algoritmos de aproximação para o Problema do Empacotamento é usual considerarmos o conceito de *razão de aproximação assintótica*, onde dizemos um algoritmo A tem razão de aproximação assintótica α quando $\lim_{OPT(I) \rightarrow \infty} A(I)/OPT(I) \leq \alpha$, onde, para uma instância I , $A(I)$ é o valor da solução encontrada pelo algoritmo A e $OPT(I)$ é o valor de uma solução ótima.

Inspirados nesta definição, introduzimos o conceito de *núcleo assintoticamente γ -aproximado*. Uma precificação p está no núcleo assintoticamente γ -aproximado se, além de satisfazer a propriedade de núcleo, temos que $\lim_{c(I) \rightarrow \infty} \sum_{i \in A} p_i / c(I) \geq \gamma$. No caso do Problema do Empacotamento, o preço pago pelos jogadores tende a γ vezes o número de recipientes usados em uma solução ótima quando tal número tende ao infinito. De forma semelhante, definimos um esquema de compartilhamento de custos γ -assintoticamente orçamento balanceado.

Note agora que, no caso do Problema do Empacotamento, se \mathcal{P} é a família de conjuntos de itens que podem ser empacotado em um único recipiente, a propriedade do núcleo pode ser substituída pela restrição que $\sum_{i \in P} p_i \leq 1$ para todo $P \in \mathcal{P}$. Assim, podemos considerar o seguinte programação linear (DGG) para encontrar uma precificação com a propriedade de núcleo que maximize a soma dos preços: $z^* = \max\{\sum_{i \in I} p_i : \sum_{i \in P} p_i \leq 1, \forall P \subseteq \mathcal{P} \wedge p_i \geq 0, \forall i \in I\}$.

De fato, (DGG) é o programa linear dual da relaxação linear do modelo de programação inteira de [Gilmore and Gomory 1961] para o Problema do Empacotamento. Como mostrado por [Karmarkar and Karp 1982], $c(I) \leq \lceil z^* \rceil + O(\log^2 c(I))$ e, apesar do problema da separação associado ser NP-difícil, podemos, para todo δ constante, encontrar uma solução p de (DGG) em tempo polinomial tal que $\sum_{i \in I} p_i \geq z^* - \delta$. Assim, existe uma precificação no núcleo assintoticamente 1-aproximado que pode ser encontrada em tempo polinomial (para δ constante). Além disso, caso a conjectura de [Scheithauer and Terno 1997] que afirma que $c(I) \leq \lceil z^* \rceil + 1$ for verdadeira, então para todo δ seria possível computar uma precificação em tempo polinomial que deixa de pagar por no máximo $2 + \delta$ recipientes em relação a uma solução ótima.

Porém, não é claro como utilizar tal programa linear para obter um esquema de precificação monotônico cruzado que mantenha estas garantias em relação ao orçamento. Por exemplo, se temos 6 itens com tamanhos 0.1, 0.6, 0.6, 0.3, 0.1 e 0.4, então a única solução ótima para o subconjunto 0.1, 0.6 e 0.6 é tal que o primeiro item tem preço 0. Porém, para o conjunto completo de itens, toda solução onde o preço do primeiro item é 0 não é ótima.

Teorema 3.1. Existe um esquema de precificação monotônico cruzado 0.591-assintoticamente orçamento balanceado para o Problema do Empacotamento.

Demonstração. Seja $M \geq 12$ um inteiro e I uma instância do Problema do Empacotamento. Considere o algoritmo HARMONIC_M [Lee and Lee 1985] que empacota os itens de forma que apenas itens do mesmo tipo fiquem num mesmo recipiente, onde um item i tem tipo k para $1 \leq k < M$ se $1/(k+1) < a_i \leq 1/k$ e tipo M se $a_i \leq 1/M$. Assim, dizemos que um recipiente é do tipo k se os itens empacotados nele são do tipo k e dizemos que um recipiente do tipo k está cheio para $1 \leq k < M$ se ele tem k itens e para $k = M$ se o espaço restante do recipiente é menor do que $1/M$. A solução encontrada pelo algoritmo HARMONIC_M com entrada S utiliza $\text{SOL}(S)$ recipientes, onde $\text{SOL}(S) \leq \sum_{j=1}^M m_j(S) + M - 1$ e $m_j(S)$, para $1 \leq j \leq M$, é o número de recipientes cheios do tipo j . Considere agora a seguinte função peso $g: I \rightarrow \mathbb{R}_+$ onde, para um item i de tipo k com $1 \leq k < M$, $g_i = 1/k$ e para um item i de tipo M , $g_i = Ma_i/(M-1)$. Note que todo recipiente cheio tem peso pelo menos 1. Segue que $\sum_{j=1}^M m_j(S) \leq \sum_{i \in S} g_i$.

Definimos o preço de um item i como $p_i = g_i/\alpha$, onde $\alpha = 1.692$. Como provado [Lee and Lee 1985], para qualquer conjunto T que pode ser empacotado em um recipiente temos que $\sum_{i \in T} g_i \leq \alpha$ e, portanto, p satisfaz a propriedade do núcleo. Assim, $c(S) \leq \text{SOL}(S) \leq \sum_{j=1}^M m_j(S) + M - 1 \leq \sum_{i \in S} g_i + M - 1 = \alpha \sum_{i \in S} p_i + M - 1$, de onde o resultado segue. \square

Referências

- Faigle, U. and Kern, W. (1993). On some approximately balanced combinatorial cooperative games. *Zeitschrift für Operations Research*, 38(2):141–152.
- Gilmore, P. C. and Gomory, R. E. (1961). A linear programming approach to the cutting-stock problem. *Operations Research*, 9(6):849–859.
- Karmarkar, N. and Karp, R. M. (1982). An efficient approximation scheme for the one-dimensional bin-packing problem. In *23rd Annual Symposium on Foundations of Computer Science*, pages 312–320.
- Lee, C. C. and Lee, D. T. (1985). A simple on-line bin-packing algorithm. *Journal of the ACM*, 32(3):562–572.
- Moulin, H. (1999). Incremental cost sharing: Characterization by coalition strategy-proofness. *Social Choice and Welfare*, 16(2):279–320.
- Qiu, X. and Kern, W. (2016). Approximate core allocations and integrality gap for the bin packing game. *Theoretical Computer Science*, 627:26 – 35.
- Scheithauer, G. and Terno, J. (1997). Theoretical investigations on the modified integer round-up property for the one-dimensional cutting stock problem. *Operations Research Letters*, 20(2):93–100.

Algoritmos aleatorizados com oráculo para MCSP: aplicações para o problema do resíduo quadrático e do logaritmo discreto

Nicollas M. Sdroievski Murilo V. G. da Silva

Departamento Acadêmico de Informática
Universidade Tecnológica Federal do Paraná (UTFPR) – Curitiba
Av. Sete de Setembro, 3165 – 80230-901 – Curitiba-PR, Brasil

Abstract. *In this paper we study the relation between three problems that are believed to be NP-intermediate. More precisely, we explicitly present algorithms with access to an oracle for the minimum circuit size problem, which run in randomized polynomial time. The first algorithm solves the quadratic residue problem and the second algorithm solves the discrete logarithm problem.*

Resumo. *Neste artigo estudamos a relação entre três problemas que são candidatos a serem NP-intermediários. Mais precisamente, apresentamos explicitamente dois algoritmos polinomiais aleatorizados que fazem acesso a um oráculo para o problema de minimização de circuitos. O primeiro algoritmo resolve o problema do resíduo quadrático e o segundo o problema do logaritmo discreto.*

1. Introdução

Richard Ladner mostrou em 1975 que se $P \neq NP$, então existem problemas, conhecidos como problemas NP-intermediários, que estão em $NP \setminus P$ e que não são NP-completos [Ladner 1975]. Embora existam alguns candidatos a pertencerem a tal classe, mostrar diretamente que um problema é NP-intermediário é extremamente difícil, pois tal resultado implicaria que $P \neq NP$. Por conta disso, maior parte da pesquisa nesta área gira em torno criação de conexões entre problemas candidatos a serem NP-intermediários e em esclarecer como estes problemas se relacionam com outras classes de complexidade [Arora and Barak 2009]. Em dois trabalhos nesta linha [Allender et al. 2006, Allender and Das 2014] foi mostrado que um oráculo para o problema de minimização de circuitos (MCSP) tem o poder de resolver em tempo polinomial aleatorizado problemas como fatoração e isomorfismo de grafos, e até mesmo qualquer problema da classe SZK, que conjectura-se estar estritamente contida em $NP \setminus P$.

Neste artigo mostramos explicitamente dois algoritmos polinomiais aleatorizados com oráculo para MCSP, um para o problema do resíduo quadrático (RQUAD) e outro para o problema do logaritmo discreto (LOGD). Os três problemas (MCSP, RQUAD e LOGD) são candidatos a serem NP-intermediários. O fato de que os algoritmos apresentados neste artigo devam existir já era conhecido, mas de maneira indireta, sendo consequência de reduções. Ambos os problemas estão em RP^{MCSP} pois possuem provas de conhecimento zero do tipo “v-bit” [Allender and Das 2014], e no caso específico de RQUAD sabemos também que $RQUAD \in ZPP^{MCSP}$, pois este é redutível ao problema da fatoração, que por sua vez está em ZPP^{MCSP} [Allender et al. 2006]. Os dois algoritmos que apresentamos aqui são bastante simples e diretos.

Na seção 2 definimos os problemas MCSP, RQUAD e LOGD e na seção 3 descrevemos brevemente por que um oráculo para o problema MCSP é bastante poderoso. Na seção

4 apresentamos os dois algoritmos, que são a contribuição deste trabalho. Por questões de espaço, vamos omitir definições usuais de classes de complexidade, oráculos, provas iterativas, provas de conhecimento zero e teoria dos grupos. As definições e notações usadas aqui são as mesmas usadas em [Arora and Barak 2009].

2. Problemas candidatos à classe NP-intermediária

Pouco se sabe sobre a complexidade de MCSP¹, e embora acredite-se que o problema não esteja em **P**, uma prova de NP-completude utilizando reduções independentes à oráculos (como é o caso de todas as reduções atuais para MCSP) causaria um colapso ao segundo nível na hierarquia polinomial [Hirahara and Watanabe 2015], algo que conjectura-se improvável. Ambos RQUAD e LOGD também são considerados intratáveis, porém possuem algoritmos sub-exponenciais, o que refutaria a hipótese de tempo exponencial no caso de serem provados NP-completos. Definimos agora os três problemas:

MCSP: dados (x, k) , onde x é uma string que representa a tabela verdade de uma função booleana f_x e $k \in \mathbb{N}^*$, responder se f_x possui circuito de tamanho menor ou igual a k .

RQUAD: dados (z, n) , onde $z \in \mathbb{Z}_n^*$, responder se existe $x \in \mathbb{Z}_n^*$ tal que $z = x^2 \pmod{n}$. Note que o conjunto dos resíduos quadráticos módulo um inteiro n , chamado de RQ_n é um grupo com multiplicação. Dessa forma, se $x, y \in RQ_n$, então $xy \in RQ_n$. Além disso, se $y \in RQ_n$ e r é um elemento aleatório de RQ_n , então yr também é um elemento aleatório de RQ_n . Por último, se $y \in RQ_n$ e $x \notin RQ_n$, então $xy \notin RQ_n$.

LOGD: dados (z, b, n) onde $z, b \in \mathbb{Z}_n$ e n é primo, encontrar um inteiro $x \in \mathbb{Z}_n$ tal que $z = b^x \pmod{n}$. Note que x pode não existir caso b não seja um gerador de \mathbb{Z}_n . Denotando por $H \subseteq \mathbb{Z}_n$ o subgrupo gerado por b , observe que se $x, y \in H$, então $xy \in H$. Além disso, se $y \in H$ e r é um elemento aleatório de H , então yr também é um elemento aleatório de H . Por último, se $y \in H$ e $x \notin H$, então $xy \notin H$.

3. O poder do problema de Minimização de Circuitos

Começamos esta seção com a definição de complexidade de Kolmogorov limitada em tempo, um conceito que tem conexões interessantes com MCSP. Seja U uma Máquina de Turing e x uma string, definimos:

$$\text{KT}_U(x) = \min\{|d|+t; \forall b \in \{0, 1, *\} \forall i \leq |x|+1 : U^d(i, b) \text{ aceita em } t \text{ passos sse } x_i = b\}$$

Ou seja, a máquina U , com acesso a descrição d , deve reconhecer corretamente os símbolos de x , dado seu índice. Como esse mecanismo não determina o tamanho de x , assumimos que para a posição $i = |x| + 1$ o símbolo correto é $*$. Para simplificar, podemos fixar uma Máquina de Turing Universal U qualquer e definir $\text{KT}(x) = \text{KT}_U(x)$, pois a medida KT pode sofrer um *overhead* no máximo logarítmico caso a máquina U seja substituída por alguma outra [Allender et al. 2006].

Considerando que a string x representa a tabela verdade de uma função booleana f_x , temos que o tamanho do menor circuito que computa f_x é aproximadamente o valor de $\text{KT}(x)$. Sendo s o tamanho desse circuito e $|x| = m$, temos [Allender and Das 2014]:

$$\left(\frac{s}{\log m}\right)^{\frac{1}{4}} \leq \text{KT}(x) \leq \mathcal{O}(s^2(\log s + \log \log m))$$

¹A sigla MCSP vem do inglês *minimum circuit size problem*.

Levando em consideração que strings x geradas por geradores pseudoaleatórios possuem $KT(x)$ pequeno, uma máquina com oráculo para MCSP pode tomar como entrada a string x e aceitar se e somente se x possui circuitos maiores que, por exemplo, $\sqrt{|x|}$. Isso é interessante pois garantimos que essa máquina aceita a maior parte das strings (pois quase todas as funções booleanas precisam de circuitos grandes para serem computadas [Arora and Barak 2009]), porém não aceita nenhuma string x com $KT(x)$ pequeno, o que torna essa máquina um ótimo teste para distinguir dentre a distribuição uniforme e uma distribuição gerada por um gerador pseudoaleatório [Allender and Das 2014].

Devido à forte e clássica conexão entre geradores pseudoaleatórios e funções irreversíveis [Håstad et al. 1999], temos então o seguinte teorema:

Teorema 3.1. [Allender et al. 2006] *Seja L uma linguagem de densidade polinomial² tal que para algum $\epsilon > 0$ e para todo $x \in L$, $KT(x) \geq |x|^\epsilon$. Seja $f(y, x)$ uma função computável em tempo polinomial no tamanho de x . Então existe um algoritmo aleatorizado polinomial N^L e um polinômio q tal que:*

$$\Pr_{x \in \{0,1\}^n, s} [f(y, N^L(y, f(y, x), s)) = f(y, x)] \geq \frac{1}{q(n)}$$

A string s acima representa as escolhas aleatórias de N^L . Pelas características já apontadas de MCSP, podemos utilizar um oráculo para esse problema no lugar da linguagem L do Teorema 3.1. Esse poder de inverter funções polinomiais em uma fração polinomial das entradas permite à uma máquina com oráculo para MCSP ser utilizada para obter algoritmos aleatorizados para vários problemas candidatos à NP-intermediários, como fatoração [Allender et al. 2006], isomorfismo de grafos e distância estatística, um problema que é SZK-completo [Allender and Das 2014].

O seguinte teorema é de especial interesse, pois conecta provas de conhecimento zero perfeitas do tipo “v-bit” (ambos RQUAD e LOGD admitem tais provas [Goldwasser et al. 1989, Malka 2008]) com o problema MCSP.

Teorema 3.2. [Allender and Das 2014] *Todas as linguagens pertencentes à classe PZK com protocolos do tipo “v-bit” também pertencem à $\mathbf{RP}^{\text{MCSP}}$*

4. Algoritmos aleatorizados com oráculo para minimização de circuitos

Nessa seção se encontra a principal contribuição desse trabalho, algoritmos aleatorizados para RQUAD e LOGD, que se utilizam de um oráculo para MCSP. Pelo Teorema 3.2 já era conhecido que ambos pertencem a $\mathbf{RP}^{\text{MCSP}}$, porém a prova do referido teorema não apresenta diretamente os algoritmos.

Para o problema RQUAD, definimos primeiramente a função polinomial $f : \mathbb{Z}_n^* \times \mathbb{Z}_n^* \rightarrow \mathbb{Z}_n^*$, que em conjunto com a máquina N^{MCSP} nos permitirá encontrar um certificado para a instância de RQUAD. Definimos $f(u, v) = f_u(v) = uv^2 \pmod{n}$.

Pelo teorema 3.1, então sabemos que existe um algoritmo aleatorizado polinomial com oráculo N^{MCSP} e um polinômio q tal que:

$$\Pr_{v \in \mathbb{Z}_n^*, s} [f_u(N^{\text{MCSP}}(u, f_u(v), s)) = f_u(v)] \geq \frac{1}{q(|v|)}$$

² L possui pelo menos $2^n/n^k$ strings para cada tamanho n e para algum k

Isso é, dado um resultado $f_u(v) = y$, o algoritmo N^{MCSP} consegue encontrar um outro número $t \in \mathbb{Z}_n^*$ tal que $f_u(v) = f_u(t)$ com probabilidade igual ou maior que $1/q(|v|)$.

Ideia principal do algoritmo: computar $y = f_z(r)$ (onde r é um elemento aleatório de \mathbb{Z}_n^*) porém executar o algoritmo N^{MCSP} com o argumento r^2 fixo para inverter o resultado, e não z . Esse passo é possível pois se $z \in RQ_N$, então o elemento y aparece na imagem de f_{r^2} , e o valor x tal que $f_{r^2}(x) = y$ é justamente o certificado para z .

Algoritmo para RQUAD: Dados como entrada (z, n) tal que $z \in \mathbb{Z}_n^*$, o algoritmo proposto executa independentemente os seguintes passos $2q(|n|)$ vezes:

(1) Sorteie aleatoriamente um número $r \in \mathbb{Z}_n^*$ e uma string aleatória s . (2) Compute $y = f_z(r) = zr^2 \pmod n$. (3) Execute $N^{\text{MCSP}}(r^2, y, s)$ e obtenha o número $x \in \mathbb{Z}_n^*$. (4) Reporte sucesso se $z = x^2 \pmod n$.

O algoritmo aceita se pelo menos umas das tentativas obtém sucesso. Observe que se $z \notin RQ_n$, então não haverá nenhum sucesso, pois y não aparece na imagem de f_{r^2} já que é o produto de um resíduo e um não resíduo. Por outro lado, se $z \in RQ_n$ temos probabilidade $1/q(|n|)$ de encontrar x tal que $z = x^2 \pmod n$, e executar os passos apresentados $2q(|n|)$ vezes garante que a probabilidade de que haja pelo menos um sucesso seja maior ou igual a $3/4$. Dessa forma, o algoritmo apresentado é da classe $\mathbf{RP}^{\text{MCSP}}$.

Podemos utilizar a mesma ideia para encontrar um algoritmo para o problema do LOGD, tendo como entrada (z, b, n) , primeiro testamos se n é primo. Definimos a função $f : \mathbb{Z}_n \times \mathbb{Z}_n \rightarrow \mathbb{Z}_n$ como $f(u, v) = f_u(v) = ub^v \pmod n$. Basta então computar $y = f_z(r)$ para r aleatório e utilizar a máquina N^{MCSP} com a entrada (b^r, y, s) para tentar inverter a função com o parâmetro fixo b^r . O algoritmo então é completamente análogo, também é executado $2q(|n|)$ vezes e apresenta x caso esse seja encontrado, caso contrário indica que z não está na órbita de b . Argumentos semelhantes aos utilizados na ideia do algoritmo para RQUAD podem ser apresentados para mostrar que esse algoritmo é da classe $\mathbf{RP}^{\text{MCSP}}$.

Referências

- Allender, E., Buhrman, H., Koucký, M., van Melkebeek, D., and Ronneburger, D. (2006). Power from random strings. *SIAM Journal on Computing*, 35(6):1467–1493.
- Allender, E. and Das, B. (2014). Zero knowledge and circuit minimization. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:68.
- Arora, S. and Barak, B. (2009). *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition.
- Goldwasser, S., Micali, S., and Rackoff, C. (1989). The knowledge complexity of interactive proof systems. *SIAM J. Comput.*, 18(1):186–208.
- Hirahara, S. and Watanabe, O. (2015). Limits of minimum circuit size problem as oracle. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:198.
- Håstad, J., Impagliazzo, R., Levin, L. A., and Luby, M. (1999). A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396.
- Ladner, R. E. (1975). On the structure of polynomial time reducibility. *J. ACM*, 22(1):155–171.
- Malka, L. (2008). *A Study of Perfect Zero-knowledge Proofs*. PhD thesis, Victoria, B.C., Canada, Canada. AAINR47335.

Um Leilão à Prova de Estratégia para o Compartilhamento de Viagens Dinâmico com Múltiplos Passageiros*

Leonardo Y. Schwarzstein¹, Flávio K. Miyazawa¹, Rafael C. S. Schouery¹

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251 – 13083-852 – Campinas – SP – Brazil

leoyvens@gmail.com, {fkm, schouery}@ic.unicamp.br

Abstract. *Dynamic ridesharing systems arrange rides between drivers and passengers on short notice. In this paper, we propose an auction that selects passengers to be served and define prices they should be charged so that prices are fair for drivers and passengers while providing profit for the company.*

Resumo. *Sistemas de compartilhamento de viagens dinâmico designam passageiros a motoristas em curto prazo. Neste artigo, propomos um leilão que seleciona passageiros a serem servidos e define preços que estes devem pagar, de maneira justa para motoristas e passageiros, provendo lucro para a empresa.*

1. Introdução

No contexto da mobilidade urbana, o *compartilhamento de viagens* ocorre quando uma viagem em um veículo pessoal é compartilhada por pessoas com trajetos similares, no intuito de dividir custos. Isto motiva o problema do compartilhamento de viagens dinâmico, que, em curto prazo, designa passageiros a motoristas e determina os preços a serem cobrados. Leilões são estudados como um mecanismo para solucionar este problema.

Duas propriedades se destacam neste contexto. Um leilão é *à prova de estratégia* se o passageiro obtém o maior ganho possível ao reportar como lance o seu real valor para a viagem. Ou seja, o passageiro não consegue manipular o leilão individualmente para alterar o resultado do leilão a seu favor. Um leilão tem um *orçamento balanceado* quando o preço pago ao motorista sempre cobre os custos para transportar os passageiros.

Trabalhos relacionados. Existem dois trabalhos sobre leilões para compartilhamento de viagens dinâmico. O primeiro [Kamar and Horvitz 2009] propõe um sistema para múltiplos motoristas baseado no leilão Vickrey-Clark-Groves (VCG) [Vickrey 1961] que, por usar heurísticas que garantem melhor desempenho computacional, deixa de garantir a propriedade de ser *à prova de estratégia*. O segundo [Kleiner et al. 2011] considera um leilão *à prova de estratégia* (uma adaptação do leilão de segundo preço [Vickrey 1961]) para um único motorista e um único passageiro. Um levantamento [Furuhata et al. 2013] cita desafios nestes trabalhos, que são garantir a propriedade de ser *à prova de estratégia* e permitir múltiplos passageiros por viagem. De fato, o leilão que propomos atende estas duas propriedades. A formulação é baseada no Dial-A-Ride-Problem (DARP), como modelado em [Cordeau 2003].

*Parcialmente financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 2013/21744-8 e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo 311499/2014-7.

Resultados. Neste artigo, propomos um leilão, chamado *leilão de segundo excedente mínimo*, para o compartilhamento de viagens dinâmico com motorista único e múltiplos passageiros servidos. Apesar de aplicações reais envolverem múltiplos motoristas, estudar o caso do motorista único é interessante como um subproblema do caso geral, além de trazer novas ideias para atacar o problema mais geral. Provamos que o leilão é à prova de estratégia e tem um orçamento balanceado. Estas propriedades se mantêm mesmo que a solução computacional não seja exata, uma característica importante para poder resolver o problema na prática. O leilão pode receber mais que o necessário para pagar o custo do motorista sendo que o excedente é o lucro da empresa. Como o leilão é genérico em certos critérios, propomos uma variação que busca maximizar o lucro da empresa. Os experimentos indicam a viabilidade do leilão proposto.

2. Modelo e Notação

No modelo que consideramos, temos um único motorista e um conjunto de passageiros. Nos referimos a um passageiro qualquer como passageiro i . Uma entrada do problema é representada por um grafo completo onde os vértices são os pontos de partida e de chegada do motorista e os pontos de coleta e entrega dos passageiros. Os pontos de coleta possuem uma janela de tempo onde a coleta é válida. Cada aresta possui um custo e um tempo para percorrê-la. Tanto o motorista como os passageiros possuem um tempo máximo de viagem. O veículo tem uma capacidade máxima no número de pessoas que pode transportar. Estas são as restrições de roteamento, inspiradas no modelo DARP [Cordeau 2003] e buscam tornar a viagem conveniente para todos.

O maior *valor* que o passageiro i pode pagar para ser servido pelo motorista é dado por v_i . Como esta informação não é conhecida pelo sistema, cada passageiro fornece um número não-negativo b_i como lance que representa o seu valor (porém, ele pode reportar $b_i \neq v_i$ se assim o desejar). O leilão deve determinar o conjunto de passageiros a ser servido pelo motorista, denotado por P^* , e o preço pago por cada passageiro servido, denotado p_i . Se $i \notin P^*$ então o passageiro i não é servido e $p_i = 0$. Definimos a *utilidade* de um passageiro i como $u_i = v_i - p_i$ se $i \in P^*$ e $u_i = 0$ se $i \notin P^*$. Um leilão é *à prova de estratégia* se, para todo passageiro i , nenhum valor de b_i resulta em uma utilidade maior para i do que a utilidade obtida fazendo $b_i = v_i$. Isto é, o passageiro i não obtém vantagens mentindo.

O custo do motorista para servir um conjunto P de passageiros é denotado por $c_m(P)$. Tal custo não depende dos lances. Definimos que um leilão tem um orçamento balanceado se $\sum_{i \in P^*} p_i \geq c_m(P^*)$. Ou seja, o preço total cobrado dos passageiros é suficiente para pagar o custo para servi-los.

3. Leilão de Segundo Excedente Mínimo

Esta seção detalha o leilão de segundo excedente mínimo e suas propriedades. Primeiro, determinam-se os conjuntos que participam do leilão, depois, qual será servido pelo motorista e o preço a ser pago por cada passageiro. Paga-se ao motorista o valor $c_m(P^*)$, e o restante do pagamento fica para a empresa, que gerencia o sistema. Participam do leilão apenas conjuntos para os quais existe uma rota que respeita as restrições de roteamento e serve apenas os passageiros do conjunto. Caso haja mais de uma rota viável, é escolhida a mais curta. Denotamos a família de conjuntos participantes do leilão por \mathcal{B} . Se um passageiro i estiver simultaneamente em dois conjuntos participantes, o leilão poderia não ser

à prova de estratégia. Para lidar com essa situação, tomaremos \mathcal{B} como uma partição do conjunto de passageiros, restringindo quais conjuntos participam do leilão. Como existem várias maneiras de se construir \mathcal{B} , apresentamos um possível critério na Seção 4.

3.1. Determinando o Conjunto Servido e a Precificação

Nesta seção descrevemos a escolha do conjunto \mathcal{B} a ser servido e o preço para cada passageiro servido. Parece natural servir um conjunto de passageiros com lances altos. Porém isto pode resultar em um benefício baixo para a empresa e para o motorista caso o custo de servir este conjunto também seja alto. Por isso, consideramos o excedente do lance em relação ao custo ao escolher os passageiros servidos. A seguir, formalizamos as definições de custo do passageiro e excedente do passageiro.

Dado $P \in \mathcal{B}$, o custo para cada passageiro $i \in P$, denotado por c_i , é escolhido de forma que $\sum_{i \in P} c_i = c_m(P)$. Ou seja, o custo do motorista para servir o conjunto é dividido entre os passageiros. Esta divisão pode ser feita através de qualquer critério que não dependa dos lances como, por exemplo, uma divisão uniforme. O excedente de i , denotado s_i , é definido como $s_i = b_i - c_i$.

Garantimos o orçamento balanceado, impondo que $p_i \geq c_i$ para todo $i \in P^*$. Assim o preço para um passageiro servido será o custo c_i mais uma parcela de lucro para a empresa. Esta parcela será igual para todos os passageiros servidos. Definimos o *excedente mínimo* de um conjunto P , denotado por $s_{\min}(P)$, como $s_{\min}(P) = \min\{s_i \mid i \in P\}$. Com isso, escolhemos $P^* \in \mathcal{B}$ tal que $s_{\min}(P^*) = \max\{s_{\min}(P) \mid P \in \mathcal{B}\}$, ou seja, o conjunto servido é aquele com o maior excedente mínimo, sendo que se este valor for negativo, então ninguém é servido.

Após escolher P^* , determinamos p_i para $i \in P^*$. Aqui nos inspiramos no leilão de segundo preço, e definimos o *segundo excedente mínimo*, denotado por s_2 , como $s_2 = \max\{\max\{s_{\min}(P) \mid P \in \mathcal{B}, P \neq P^*\}, 0\}$. Ou seja, s_2 é o segundo maior excedente mínimo, ou 0 caso este seja negativo. Por fim, definimos $p_i = c_i + s_2$ para todo $i \in P^*$, de modo que p_i não depende de b_i . A parcela c_i será paga ao motorista e a parcela s_2 será paga para a empresa.

3.2. Propriedades do Leilão de Segundo Excedente Mínimo

Teorema 3.1. O leilão de segundo excedente mínimo tem o orçamento balanceado.

Demonstração. Como $s_2 \geq 0$, temos $\sum_{i \in P^*} (c_i + s_2) = c_m(P^*) + |P^*|s_2$ e da definição de p_i para $i \in P^*$ conclui-se $\sum_{i \in P^*} p_i \geq c_m(P^*)$. \square

Teorema 3.2. O leilão de segundo excedente mínimo é a prova de estratégia.

Demonstração. Considere um passageiro i . Se o passageiro i dá o lance $b_i = x$, então seja u_x sua utilidade, $p_x = c_i + s_2^x$ o seu preço caso seja servido, P_x^* o conjunto servido e s_2^x o segundo excedente mínimo. Note que $u_{v_i} \geq 0$, já que um passageiro nunca paga mais do que o seu lance no leilão de segundo excedente mínimo. Se $i \notin P_x^*$ então $u_x = 0$ e $u_{v_i} \geq u_x$. Se $i \in P_x^*$ e $i \in P_{v_i}^*$ então $u_x = u_{v_i}$ pois p_i não depende do lance de i . Se $i \in P_x^*$ e $i \notin P_{v_i}^*$ então quando $b_i = v_i$ o excedente de i é o mínimo do seu conjunto logo $v_i - c_i \leq s_{\min}(P_{v_i}^*)$. Quando $b_i = x$ o excedente de i supera $s_{\min}(P_{v_i}^*)$ fazendo $s_{\min}(P_{v_i}^*) = s_2^x$. Substituindo obtemos $v_i - c_i \leq s_2^x$, logo $v_i \leq p_i^x$ e $u_x \leq 0 \leq u_{v_i}$. \square

4. Experimentos Computacionais

Geramos entradas em um mapa real de aproximadamente 100 quilômetros quadrados do centro da cidade de São Paulo, distribuindo os vértices de maneira uniformemente aleatória. Por simplicidade, dividimos o custo igualmente entre os passageiros, ou seja, $c_i = c_m(P)/|P|$.

A operação de maior custo computacional é construir \mathcal{B} . O critério pode ser qualquer desde que não considere os lances, caso contrário poderia violar a propriedade de ser à prova de estratégia. Nos experimentos utilizamos um algoritmo que minimiza, de maneira gulosa, o custo por passageiro. A cada iteração do algoritmo o conjunto com menor custo por passageiro é adicionado, até não ser possível adicionar mais conjuntos. Implementamos este algoritmo de maneira exata utilizando programação linear inteira e o resolvidor Gurobi. Limitamos o número máximo de passageiros servidos a 4. Em um computador Intel Core i7 2,7 GHz, a implementação obteve solução para uma entrada de 20 passageiros em 10 segundos e para uma entrada de 100 passageiros em 10 minutos. Isto demonstra que o problema é tratável em tempo prático, porém para entradas maiores é necessário implementar heurísticas para atender aos requisitos de um sistema dinâmico.

5. Conclusão

O leilão de segundo excedente mínimo combina as propriedades desejáveis de ser à prova de estratégia e ter um orçamento balanceado, respondendo a questões em aberto na literatura. O arcabouço teórico é robusto pois permite liberdade nos critérios de divisão do custo e de escolha dos conjuntos participantes. O leilão é permissivo em relação a implementação pois o uso de heurísticas no passo mais computacionalmente custoso (construir \mathcal{B}) não prejudica as propriedades teóricas. Isto é uma vantagem sobre o leilão VCG que requer implementações exatas. Os experimentos estão em fase inicial e pretendemos fazer experimentos comparativos com o leilão VCG, bem como implementar e analisar a performance de heurísticas. As principais limitações do leilão são a restrição a um único motorista e a restrição de que \mathcal{B} deve particionar o conjunto de passageiros. O desafio futuro mais interessante é generalizar o leilão para múltiplos motoristas.

Referências

- Cordeau, J.-f. (2003). A Branch-and-Cut Algorithm for the Dial-a-Ride Problem. *Operations Research*, 54:573–586.
- Furuhata, M., Dessouky, M., Ordonez, F., Brunet, M.-E., Wang, X., and Koenig, S. (2013). Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57:28–46.
- Kamar, E. and Horvitz, E. (2009). Collaboration and Shared Plans in the Open World: Studies of Ridesharing. In *Proceedings of the 21th International Joint Conference on Artificial Intelligence*, pages 187–194.
- Kleiner, A., Nebel, B., and Ziparo, V. A. (2011). A Mechanism for Dynamic Ride Sharing Based on Parallel Auctions. In *Proceedings of the 22th International Joint Conference on Artificial Intelligence*, pages 266–272.
- Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, 16(1):8–37.

Caracterizações de convexidades geométricas de grafos

Rafael Teixeira de Araújo¹, Rudini Menezes Sampaio¹

¹Departamento de Computação – Universidade Federal do Ceará (UFC)
60455760 - Fortaleza, CE - Brasil

rafaelteixeira@lia.ufc.br, rudini@lia.ufc.br

Abstract. A graph convexity is geometric if every convex set is the convex hull of its extreme vertices. Several papers have characterized classes of graphs for which a given convexity is geometric, such as the monophonic convexity in chordal graphs, the geodesic convexity in ptolemaic graphs [Farber and Jamison 1986] and the m_3 -convexity in weak-bipolarized graphs [Dragan et al. 1999]. In this paper, we obtain characterizations for the convexities P_3 , P_3^* , P_4^+ and triangle-path. We also obtain results in the opposite direction: is there a graph convexity which is geometric exactly in a given graph class?

Resumo. Uma convexidade de grafos é geométrica se todo conjunto convexo é o fecho convexo de seus vértices extremos. Vários trabalhos têm caracterizado as classes de grafos para os quais uma dada convexidade é geométrica, como a convexidade monofônica em grafos cordais, a convexidade geodésica em grafos ptolemaicos [Farber and Jamison 1986] e a convexidade m_3 em grafos bipolarizados fracos [Dragan et al. 1999]. Nesse artigo, obtemos caracterizações para as convexidades P_3 , P_3^* , P_4^+ e triangle-path. Também obtemos resultados no sentido oposto: dada uma classe de grafos, existe alguma convexidade que só é geométrica nessa classe?

1. Introdução

Uma convexidade de grafos é um par ordenado (G, \mathcal{C}) , onde G é um grafo simples e \mathcal{C} é uma coleção de subconjuntos de $V(G)$ (chamados *conjuntos convexos*) tal que $\emptyset, V(G) \in \mathcal{C}$ e a interseção de dois conjuntos convexos é convexo. Dado um grafo G e uma coleção de caminhos \mathcal{P} em G , a convexidade de caminhos \mathcal{P} é tal que um conjunto $S \subseteq V(G)$ é convexo se e só se todo caminho em \mathcal{P} com extremidades em S está contido em S .

Entre as convexidades de caminho mais conhecidas, destacam-se a convexidade P_3 (caminhos com 3 vértices) [Campos et al. 2015], convexidade *geodésica* (caminhos mínimos) [Araújo et al. 2013], convexidade *monofônica* (caminhos induzidos) [Dourado et al. 2010], convexidade m_3 (caminhos induzidos com mais de 3 vértices) [Dragan et al. 1999] e convexidade *triangle-path* (T-caminhos: caminhos cujas cordas formam triângulos no caminho) [Changat and Mathew 1999].

Dada uma convexidade (G, \mathcal{C}) , o *fecho convexo* $\mathcal{H}(S)$ de um conjunto $S \subseteq V(G)$ é o menor conjunto convexo que contém S . Dado um conjunto convexo $S \subseteq V(G)$, dizemos que $v \in S$ é um *extremo* de S se $S \setminus \{v\}$ também é convexo. Uma convexidade (G, \mathcal{C})

é *geométrica* se satisfaz a propriedade de Minkowski-Krein-Milman: todo conjunto convexo é fecho convexo de seus extremos. Sabe-se que essa propriedade é equivalente a propriedade *Antiexchange*: para todo $S \subseteq V(G)$ e quaisquer vértices distintos $x, y \notin S$, temos que $x \in \mathcal{H}(S \cup \{y\})$ implica $y \notin \mathcal{H}(S \cup \{x\})$.

Uma questão bastante estudada é caracterizar a classe de grafos para os quais uma dada convexidade é geométrica.

2. Resultados conhecidos

O seguinte teorema obtém uma caracterização para a convexidade monofônica. Um grafo é *cordal* se não possui *holes* (ciclo induzido de tamanho maior que três).

Teorema 2.1 ([Farber and Jamison 1986]). *A convexidade monofônica de um grafo G é geométrica se e só se G é cordal.*

O seguinte teorema obtém uma caracterização para a convexidade geodésica. Um *3-fan* é um grafo com 5 vértices tal que um vértice é universal e os demais induzem um P_4 . Um grafo é *ptolemaico* se para quaisquer 4 vértices a, b, x, y de uma mesma componente conexa:

$$d(a, b) \cdot d(x, y) \leq d(a, x) \cdot d(b, y) + d(b, x) \cdot d(a, y).$$

Teorema 2.2 ([Farber and Jamison 1986]). *Seja G um grafo. Então as seguintes propriedades são equivalentes: (a) a convexidade geodésica de G é geométrica; (b) G é um grafo ptolemaico; (c) G é cordal e não tem 3-fan induzido; (d) G é cordal e todo caminho induzido é mínimo.*

O teorema abaixo obtém uma caracterização para a convexidade m_3 . Um grafo é *bipolarizado-fraco* se é livre de HHDA (livre de *hole*, *house*, *dominó* e o grafo A).

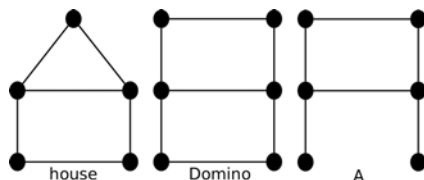


Figura 1. Grafos *house*, *dominó*, A

Teorema 2.3. [Dragan et al. 1999] *A convexidade m_3 de um grafo G é geométrica se e só se G é um grafo bipolarizado fraco.*

3. Resultados novos

Seguindo a sequência desses resultados, nós obtivemos em [Araújo and Sampaio 2013] alguns resultados novos: uma caracterização para as convexidades *triangle-path*, P_3 , P_3^* e P_4^+ .

Dado um grafo G , um conjunto S é P_3^* -convexo, se para todo P_3 induzido abc com $a, c \in S$, então $b \in S$. Similarmente, definimos um conjunto S como sendo P_4^+ -convexo, se para todo P_4 induzido $abcd$ com $a, c, d \in S$, então $b \in S$.

Teorema 3.1. *Seja G um grafo. A convexidade triangle-path de G é geométrica se e só se G é uma floresta. A convexidade P_3 de G é geométrica se e só se G é uma floresta de estrelas. A convexidade P_3^* de G é geométrica se e só se G é um cografo cordal. A convexidade P_4^+ de G é geométrica se e só se G é um cografo.*

Devido a restrição de espaço, faremos apenas as provas para as convexidades P_3 e *triangle-path*.

Prova para Convexidade P_3 . Suponha que a convexidade P_3 é geométrica em G . Se G possui um $K_3 \{x, y, z\}$, temos que $S = \{z\}$ é um conjunto convexo, $x \in \mathcal{H}(S \cup \{y\})$ e $y \in \mathcal{H}(S \cup \{x\})$, o que contradiz a propriedade *Antiexchange*. Logo G é livre de triângulos. Se G possui um C_4 induzido $\{a, b, c, d\}$ com arestas ab, bc, cd, da , temos que $S = \{c, d\}$ é um conjunto convexo pois G é livre de triângulos e portanto não existe vértice adjacente a c e d . Portanto, $a \in \mathcal{H}(S \cup \{b\})$ e $b \in \mathcal{H}(S \cup \{a\})$, o que contradiz a propriedade *Antiexchange*. Logo G é livre de C_4 . Se G possui um C_5 induzido $\{a, b, c, d, e\}$ com arestas ab, bc, cd, de, ea , temos que $S = \{c, d, e\}$ é um conjunto convexo (pois G não possui K_3 nem C_4 induzido). Portanto, $a \in \mathcal{H}(S \cup \{b\})$ e $b \in \mathcal{H}(S \cup \{a\})$, o que contradiz a propriedade *Antiexchange*. Logo G é livre de C_5 induzido.

Se G possui um P_4 induzido $wxyz$, então $S = \{w, z\}$ é P_3 -convexo. Isso porque se existisse um vértice u adjacente a w e z , teríamos que ux e uy não seriam arestas (pois G é livre de triângulos) e portanto teríamos o C_5 induzido $\{w, x, y, z, u\}$, uma contradição. Portanto, $x \in \mathcal{H}(S \cup \{y\})$ e $y \in \mathcal{H}(S \cup \{x\})$, o que contradiz a propriedade *Antiexchange*. Logo G é livre de P_4 induzido. Ou seja, G é um cografo livre de K_3 e C_4 .

Sabe-se que todo cografo conexo G é a junção de dois cografos G_1 e G_2 (incluindo todas as arestas entre G_1 e G_2). Se G_1 possui uma aresta ab então teríamos um $K_3 \{abz\}$ com qualquer vértice z de G_2 , uma contradição. Logo, G_1 e G_2 não possuem arestas. Se G_1 e G_2 possuem mais de 2 vértices cada, então teríamos um C_4 induzido $\{u_1, v_1, u_2, v_2\}$ com quaisquer vértices $u_1, v_1 \in V(G_1)$ e $u_2, v_2 \in V(G_2)$, uma contradição. Portanto, G_1 ou G_2 possui apenas um vértice. Portanto G é uma estrela $K_{1,p}$. Se G é desconexo, então G é uma floresta de estrelas.

Se G é uma estrela com pelo menos 3 vértices e centro c , então c é o único vértice com grau maior que um e portanto todos os P_3 induzidos contêm c . Isso implica que a propriedade *Antiexchange* é satisfeita e portanto a convexidade P_3 é geométrica. \square

Prova para convexidade triangle-path. Suponha que a convexidade *triangle-path* é geométrica em G . Se G possui um K_3 induzido $\{x, y, z\}$, temos que $S = \{z\}$ é um conjunto convexo, $x \in \mathcal{H}(S \cup \{y\})$ (pois há o T-caminho yxz entre y e z passando por x com uma única corda yz que forma um triângulo) e $y \in \mathcal{H}(S \cup \{x\})$ (pelo mesmo motivo anterior), o que contradiz a propriedade *Antiexchange*. Logo G é livre de triângulos.

Se G possui um ciclo induzido v_1, v_2, \dots, v_k para $k \geq 4$, temos que $S = \{v_2, v_3\}$ é um conjunto T-convexo, pois não existem T-caminhos entre v_2 e v_3 visto que G é livre de triângulos. Note que $v_1 \in \mathcal{H}(S \cup \{v_4\})$ e $v_4 \in \mathcal{H}(S \cup \{v_1\})$, o que contradiz a propriedade *Antiexchange*. Logo G é livre de ciclos induzidos. Como todo ciclo contém um ciclo induzido, temos que G é livre de ciclos. Ou seja G é uma floresta.

Por outro lado, suponha que G é uma floresta. Logo G não contém triângulos. Com isso, todo T-caminho de G é um caminho induzido, e vice-versa. Ou seja, todo conjunto T-convexo de G é monofonicamente convexo, e vice-versa. Isso quer dizer que a convexidade *triangle-path* e a convexidade monofônica são idênticas em G . Como G é uma floresta, então G é cordal, o que implica que a convexidade monofônica é geométrica em G e portanto a convexidade *triangle-path* também é geométrica em G . \square

Também obtivemos resultados no sentido inverso: dada uma classe de grafos, existe uma convexidade que é geométrica apenas nos grafos dessa classe?

Seja H um grafo com pelo menos dois vértices. Dado um grafo G , dizemos que um conjunto $S \subseteq V(G)$ é H -free convexo se para todo $S' \subseteq S$, $|S'| = |V(H)| - 1$ temos que, se $S' \cup \{x\}$ induz um grafo H , então $x \in S$. Dada uma família \mathcal{F} de grafos com pelo menos dois vértices, dizemos que um conjunto $S \subseteq V(G)$ é \mathcal{F} -free convexo se é H -free convexo para todo $H \in \mathcal{F}$.

Teorema 3.2. *A convexidade \mathcal{F} -free é geométrica em G se e só se G é livre de \mathcal{F} .*

Demonstração. Se G não possui nenhum subgrafo induzido $H \in \mathcal{F}$, então por definição todo subconjunto $S \subseteq V(G)$ é H -free convexo. Portanto, a propriedade *Antiexchange* é sempre satisfeita e a convexidade é geométrica.

Suponha que G possui um subgrafo induzido em \mathcal{F} . Seja $H \in \mathcal{F}$ um subgrafo induzido de G com número mínimo de vértices. Sejam S um conjunto de vértices de G que induz H , x, y dois vértices distintos de S e $S' = S \setminus \{x, y\}$. Note que S' é \mathcal{F} -free convexo, pela minimalidade de H . Note ainda que $y \in \mathcal{H}(S' \cup \{x\})$ e que $x \in \mathcal{H}(S' \cup \{y\})$, o que contradiz a propriedade *Antiexchange*. Logo, a convexidade não é geométrica. \square

Corolário 3.1. *Seja $\mathcal{F} = \{C_{2k+1} : k = 1, 2, 3, \dots\}$ a família infinita dos ciclos ímpares. A convexidade \mathcal{F} -free é geométrica em G se e só se G é bipartido.*

Corolário 3.2. *Seja $\mathcal{F} = \{TK_5, TK_{3,3}\}$, onde T representa topological minor, a família infinita dos grafos que podem ser obtidos a partir do K_5 ou do $K_{3,3}$ através de subdivisão de arestas. A convexidade \mathcal{F} -free é geométrica em G se e só se G é planar.*

Referências

- Araújo, R. T., Sampaio, R., and Szwarcfiter, J. (2013). The convexity of induced paths of order three. *Electronic Notes in Discrete Mathematics*, 44(0):109 – 114.
- Araújo, R. T. and Sampaio, R. (2013). Convexidades de caminhos e convexidades geométricas. *Dissertação de Mestrado (MDCC-UFC), Fortaleza, Brazil*.
- Campos, V., Sampaio, R. M., Silva, A., and Szwarcfiter, J. L. (2015). Graphs with few P_4 's under the convexity of paths of order three. *Discrete Applied Mathematics*, 192:28 – 39. 11th Cologne/Twente Workshop on Graphs and Combinatorial Optimization (CTW 2012).
- Changat, M. and Mathew, J. (1999). On triangle path convexity in graphs. *Discrete Mathematics*, 206(1–3):91 – 95.
- Dourado, M. C., Protti, F., and Szwarcfiter, J. L. (2010). Complexity results related to monophonic convexity. *Discrete Applied Mathematics*, 158(12):1268 – 1274.
- Dragan, F. F., Nicolai, F., and Brandstädt, A. (1999). Convexity and HHD-free graphs. *SIAM Journal on Discrete Mathematics*, 12(1):119–135 (electronic).
- Farber, M. and Jamison, R. (1986). Convexity in graphs and hypergraphs. *SIAM Journal on Algebraic Discrete Methods*, 7(3):433–444.

Results on Circular-Arc Bigraphs

(Extended Abstract)

Fabricio Schiavon Kolberg^{1,*}, Marina Groshaus^{2,†},
André Luiz Pires Guedes¹, Renato Carmo¹

¹Universidade Federal do Paraná, Curitiba, Brazil

²Universidad de Buenos Aires / CONICET, Buenos Aires, Argentina.

Abstract. We present a series of results related to the structural properties of the bipartite graph class known as circular-arc bigraphs. We also propose the definition of a Helly circular-arc bigraph subclass, based on a concept known as bipartite-Helly, along with a few results related to its structural properties.

1. Introduction

A *circular-arc graph* is a graph that admits an intersection model of arcs on a circle. Arising as a generalization of *interval graphs*, the class has been extensively studied by many researchers since 1964 [Lin and Szwarcfiter 2009], yielding a plethora of relevant results. Trotter and Moore [Trotter and Moore 1976] presented infinite sets of minimally non circular-arc graphs. Francis et al. [Francis et al. 2014] presented a characterization of circular-arc graphs, along with a certifying recognition algorithm. Furthermore, Hell and Huang [Hell and Huang 2004] characterized *two-clique* circular arc graphs as the complements of *interval bigraphs*. Soulignac, in his thesis [Soulignac 2010], characterized different circular-arc subclasses, including the *Helly* subclass, based on the concept of *Helly families* [Helly 1923]. A linear-time algorithm for recognition of the class was presented by McConell [McConnell 2003], based on the so-called *circular-arc matrices*.

The concept of *circular-arc bigraphs* arises as a bipartite variation of the circular-arc concept, similar to interval bigraphs with relation to interval graphs. We treat circular-arc bigraph representations as *bi-circular-arc models* $\mathcal{B} = (C, \mathbb{I}, \mathbb{E})$, where C is a circle, and \mathbb{I}, \mathbb{E} are families of arcs over C . It is then said that \mathcal{B} represents a bipartite graph $G = (U, V, E)$ if there is a one-to-one correspondence between U and \mathbb{I} , and one between V and \mathbb{E} , such that, for every $u \in U$ and every $v \in V$, $\{u, v\} \in E$ if and only if the arc corresponding to u in \mathbb{I} and the one corresponding v in \mathbb{E} intersect.

Being a rather recent subject, it has yet to be as extensively studied as the other classes cited. Basu et al. [Basu et al. 2013] made a series of characterizations of circular-arc bigraphs, as well as the *proper* and *unit* subclasses, based on the so-called *biadjacency matrices*. Das et al. [Das and Chakraborty 2015] found a forbidden structure characterization of proper circular-arc bigraphs and interval bigraphs.

In this paper, we present a handful of results about the structural properties of circular-arc bigraphs, as well as the Helly subclass, which we introduce. Our study's main focus were the structural properties of circular-arc bigraphs, based on their graph structures instead of matrix structures. A handful of potentially useful results were found, which we present in the next section.

*Partially supported by Coordenação de Aperfeiçoamento Pessoal de Nível Superior, CAPES

†Partially supported by ANPCyT PICT-2013-2205 and CONICET

2. Results on Circular-Arc Bigraphs

To represent circular arcs on a circle C of length n , we represent each point as a real number $0 \leq r < n$, which denotes a clockwise offset through the circle from a fixed zero point in C . It is then possible to denote a circular arc by a pair (r_1, r_2) of real numbers, where the arc is traced clockwise from point r_1 to point r_2 , respectively called the counter-clockwise and clockwise endpoints of the arc. Furthermore, in the enunciation of the following results, every index sum is circular, which means that in an enumerated set $\{a_1, \dots, a_n\}$, we have $a_{n+1} = a_1$ and $a_{1-1} = a_n$.

Lemma 1. *Let $B = (U, V, E)$ be a bipartite graph, with $|V| = n$ and $|U| = m$. If there is an ordering $S = (v_1, \dots, v_n)$ of V such that, for every $u \in U$, the neighborhood of u is circularly consecutive with relation to S , then B is a circular-arc bigraph.*

Proof. Start with a circle C of length $n + 1$. For each $v_i \in V$, draw an arc $(i, i + 1) \in \mathbb{I}$ to represent it. Then, for each $u \in U$ whose neighborhood ranges from v_i to v_j , we draw an arc $(i + \frac{1}{2}, j + \frac{1}{2}) \in \mathbb{E}$ to represent it. $(C, \mathbb{I}, \mathbb{E})$ corresponds to B . \square

Lemma 1 implies that graphs such as *generalized crowns* and *bipartite permutation graphs* are all circular-arc bigraphs.

Definition 1. *Let $a > 0$ and $0 \leq b < a$. The generalized crown graph S_b^a is a bipartite graph with vertex set $\{u_0, \dots, u_{a-1}\} \cup \{v_0, \dots, v_{a-1}\}$, such that the neighborhood of v_i is $U - \{u_i, \dots, u_{i+b}\}$.*

Definition 2. *Let $G = (V, E)$ be a graph. Then G is a permutation graph if there exist two permutations P_1, P_2 of its vertex set such that, if the index of v in P_1 is less than that of w , then $\{v, w\} \in E$ if and only if the index of v in P_2 is greater than that of w . A bipartite permutation graph is simply a permutation graph that is bipartite.*

Let \mathbb{A} be a family of arcs on a circle C . An arc B of C is said to *minimally intersect* $A \in \mathbb{A}$ if it intersects A without intersecting another arc $A' \in \mathbb{A}$ such that $A' \subset A$. Lemma 2 allows us to describe a type of forbidden structure for the class.

Lemma 2. *Let $0 < m < n$, and let \mathbb{A} be a family of arcs on a circle C , with $|\mathbb{A}| = n$, such that no two arcs have coinciding endpoints. There are at most n subfamilies $\mathbb{A}' \subset \mathbb{A}$ such that $|\mathbb{A}'| = m$ for which it is possible to draw an arc on C that intersects every arc of \mathbb{A}' without intersecting any of $\mathbb{A} - \mathbb{A}'$.*

Proof. Let B be an arc on C ($B \notin \mathbb{A}$) that intersects exactly m arcs of \mathbb{A} . Let \mathbb{A}' be the family of arcs intersected by B , and let $S = (A_1, \dots, A_m)$ be the order in which the counter-clockwise endpoint of each arc in \mathbb{A}' is first encountered by trailing the circle clockwise from the clockwise endpoint of B . Let A_i be the first arc in the order S that is minimally intersected by B . We shall call A_i the *first minimally intersected arc* of B . Notice that \mathbb{A}' contains all arcs that contain A_i . Let $k > 0$ be the number of such arcs. In particular, notice that every arc in $\{A_1, \dots, A_{i-1}\}$ contains A_i , otherwise, either they would be minimally intersected themselves, or they'd contain another arc that is minimally intersected and comes before A_i in S . The other arcs that \mathbb{A}' contains will be exactly the first $m - k$ arcs that are not contained in (nor contain) A_i and whose counter-clockwise endpoints come immediately after the counter-clockwise endpoint of A_i . Therefore, if A_i is the first minimally intersected arc of a new arc B that intersects m arcs, \mathbb{A}' will always be the same. That implies that for each arc $A \in \mathbb{A}$, there is at

most one subfamily of m arcs which can be intersected by a new arc with A being the first minimally intersected arc (there might also be none, for instance, if an arc is contained in more than $m - 1$ arcs). Therefore, there are at most n subfamilies of m arcs that can be exclusively intersected by a new arc. \square

Corollary 3. *Let $B = (U, V, E)$ be a bipartite graph, such that $|U| = n$, $|V| = n + 1$, and there is a k such that $|N(v)| = k$ for all $v \in V$. If the neighborhoods of the vertices of V are all pairwise distinct, then B is not a circular-arc bigraph.*

Figure 1 contains a handful of forbidden subgraphs we found for the class.

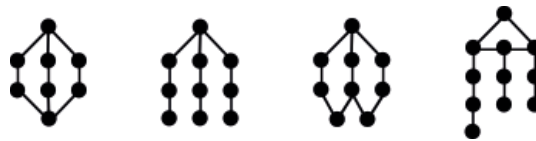


Figure 1. List of forbidden subgraphs we found for the circular-arc bigraph class.

2.1. Studies on the Helly Subclass

The concept of Helly families [Helly 1923] lends itself to the definition of many extensively studied graph classes [Szwarcfiter 1997, Groshaus and Szwarcfiter 2007, Groshaus and Szwarcfiter 2008]. An adaptation of the definition of Helly families, known as *bipartite-Helly*, was formulated by Groshaus and Szwarcfiter [Groshaus and Szwarcfiter 2010], and allows us to define Helly-like properties between pairs of families. For the results in this section, we first introduce *interval bigraphs* in Definition 3, and then *Helly circular-arc bigraphs* and *Helly interval bigraphs* in Definition 4, loosely based on the bipartite-Helly concept.

Definition 3. *A bipartite graph $B = (U, V, E)$ is an interval bigraph if it admits a bi-interval representation. A bi-interval representation of B is a mapping between $U \cup V$ and a family of intervals on the real number line, such that $u \in U$ and $v \in V$ are neighbors if and only if their corresponding intervals intersect.*

For the following definition, we consider a *biclique* of a given graph to be a maximal bipartite-complete induced subgraph.

Definition 4. *A bipartite graph B is a Helly circular-arc bigraph if it admits a Helly bi-circular-arc model. A bi-circular-arc model $(C, \mathbb{I}, \mathbb{E})$ is Helly if, for any biclique in the graph it represents, there is a point $p \in C$ such that all arcs corresponding to vertices of the biclique contain p . A bipartite graph B is a Helly interval bigraph if it admits a Helly bi-interval representation. A bi-interval representation is Helly if, for any biclique of the graph it represents, the intervals corresponding to the vertices of the biclique all contain a common point p in the real number line.*

It is easy to verify that both classes are hereditary over induced subgraphs, and that Helly interval bigraphs are a proper subclass of Helly circular-arc bigraphs. Lemma 4 presents a sufficient condition for a graph to be a Helly circular-arc bigraph. We denote by (C_{2k}, S_ℓ) the graph obtained by adding ℓ isolated vertices to a C_{2k} .

Lemma 4. *Let G be a bipartite graph. If G is $K_{1,3}$ -free and (C_{2k}, S_ℓ) -free (with $k > 2$), then G is a Helly circular-arc bigraph.*

Proof. If the graph contains a C_{2n} , for $n > 2$, and doesn't contain any induced $K_{1,3}$ nor any isolated vertices, then its only component is the C_{2n} , which can easily be verified to be a Helly circular-arc bigraph. Now suppose the graph does not contain a C_{2n} , for $n > 2$. Then, since the graph is $K_{1,3}$ -free, every single one of its components is either a path or a $K_{2,2}$. It is easy to verify that both paths and the $K_{2,2}$ are Helly interval bigraphs, meaning that a graph that has those graphs as connected components also is. \square

Lemma 4 is a simple sufficient condition for the Helly subclass, serving to prune the search space for forbidden structures. It is known to not be necessary, since $K_{1,3}$ by itself is a Helly circular-arc bigraph.

References

- Basu, A., Das, S., Ghosh, S., and Sen, M. (2013). Circular-arc bigraphs and its subclasses. *Journal of Graph Theory*, 73(4):361–376.
- Das, A. K. and Chakraborty, R. (2015). *Algorithms and Discrete Applied Mathematics: First International Conference, CALDAM 2015, Kanpur, India, February 8-10, 2015. Proceedings*, chapter New Characterizations of Proper Interval Bigraphs and Proper Circular Arc Bigraphs, pages 117–125. Springer International Publishing, Cham.
- Francis, M. C., Hell, P., and Stacho, J. (2014). Forbidden structure characterization of circular-arc graphs and a certifying recognition algorithm. *CoRR*, abs/1408.2639.
- Groshaus, M. and Szwarcfiter, J. (2007). Biclique-helly graphs. *Graphs and Combinatorics*, 23:633 – 645.
- Groshaus, M. and Szwarcfiter, J. (2008). On hereditary helly classes of graphs. *Discrete Mathematics & Theoretical Computer Science*, 10(1).
- Groshaus, M. and Szwarcfiter, J. L. (2010). Biclique graphs and biclique matrices. *Journal of Graph Theory*, 63(1):1–16.
- Hell, P. and Huang, J. (2004). Interval bigraphs and circular arc graphs. *J. Graph Theory*, 46:313–327.
- Helly, E. (1923). Über Mengen konvexer Körper mit gemeinschaftlichen Punkten. *Jahresber. Dtsch. Math.-Ver.*, 32:175–176.
- Lin, M. C. and Szwarcfiter, J. L. (2009). Characterizations and recognition of circular-arc graphs and subclasses: A survey. *Discrete Mathematics*, 309(18):5618–5635. *Combinatorics 2006, A Meeting in Celebration of Pavol Hell's 60th Birthday (May 1-5, 2006)*.
- McConnell, R. M. (2003). Linear-time recognition of circular-arc graphs. *Algorithmica*, 37(2):93–147.
- Souignac, F. J. (2010). *Sobre Grafos Arco-Circulares Propios y Helly*. PhD thesis, Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Computacion.
- Szwarcfiter, J. L. (1997). Recognizing clique-Helly graphs. *Ars Comb.*, 45:29–32.
- Trotter, Jr., W. T. and Moore, Jr., J. I. (1976). Characterization problems for graphs, partially ordered sets, lattices, and families of sets. *Discrete Mathematics*, 16(4):361–381.

O Problema da Atribuição Dupla de Custo Mínimo

Vinícius F. dos Santos¹, Sebastián Urrutia¹

¹ Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) Belo Horizonte, MG, Brazil

{vinciussantos, surrutia}@dcc.ufmg.br

Abstract. *In this work we introduce the Minimum Cost Double Assignment Problem, in which the input is a pair X, Y of n -dimensional arrays of costs and one has to find a pair (i, j) for which a given function $f(i, j)$ is true, minimizing $x_i + y_j$. We present exact polynomial time algorithms for its solution and discuss the best scenarios for each algorithm.*

Resumo. *Neste trabalho introduzimos o Problema da Atribuição Dupla de Custo Mínimo, no qual a entrada é um par X, Y de vetores n -dimensionais de custos e é necessário encontrar um par (i, j) para o qual uma função $f(i, j)$ dada é verdadeira, minimizando $x_i + y_j$. Apresentamos algoritmos polinomiais para a solução do problema e discutimos os melhores cenários para os algoritmos apresentados.*

1. Introdução

Considere duas tarefas a serem executadas. Ambas as tarefas precisam de determinados recursos (tempo, espaço, ferramentas, matéria-prima, mão-de-obra, etc) para a sua execução. Cada uma das tarefas pode ser executada em n modos distintos e cada um destes modos tem um custo associado e utiliza quantidades diferentes de recursos. Como recursos são compartilhados pelas duas tarefas, alguns pares de modos (um de cada tarefa) são incompatíveis.

No Problema da Atribuição Dupla de Custo Mínimo, é necessário atribuir um modo de execução a cada tarefa de forma que os modos selecionados sejam compatíveis e a soma dos custos dos modos selecionados seja tão pequena quanto possível.

Como exemplo prático, considere o Problema de Alocação de Cais e Guindaste [Meisel and Bierwirth 2009] no qual duas embarcações podem precisar utilizar as instalações de um porto para entregar um número de contêineres e receber outros. Cada embarcação demanda espaço, tempo, mão-de-obra e equipamentos específicos para a execução das tarefas. A quantidade e a qualidade dos recursos designados a esta tarefa (o modo de execução) determina o custo financeiro que deve ser pago à administração do porto e estes recursos devem ser compartilhados pelas duas embarcações. Neste caso, o dono das embarcações gostaria de resolver o Problema da Atribuição Dupla de Custo Mínimo.

Neste trabalho apresentaremos diferentes soluções de tempo polinomial para este problema. Iniciaremos com uma formalização do mesmo e, em seguida, apresentaremos diferentes estratégias de solução. Todas as estratégias apresentadas são originais. Finalmente, concluiremos com direções de trabalhos futuros.

2. Formalização

Denotaremos por $[n]$ o conjunto $\{1, \dots, n\}$. Sejam X e Y vetores inteiros ordenados de dimensão n e $f : [n] \times [n] \rightarrow \{0, 1\}$ uma função booleana, que recebe índices dos vetores como parâmetros. Neste trabalho assumiremos que f pode ser computada em tempo constante. Se $f(i, j) = 1$, dizemos que i e j são compatíveis.

O Problema da Atribuição Dupla de Custo Mínimo consiste em determinar um par de índices compatíveis de menor soma, isto é, encontrar um par (i, j) tal que $f(i, j) = 1$ e $x_i + y_j$ seja tão pequeno quanto possível. Se tal par não existe, então o problema não possui solução.

Em determinados casos, estaremos interessados em saber qual a posição k do primeiro par compatível quando todos os pares estão ordenados pela soma de seus elementos correspondentes nos vetores. Definimos $k = n^2$ se não existe par compatível. Por simplicidade e restrições de espaço, vamos assumir que não existem duas somas iguais. O caso mais geral pode ser tratado sem grandes mudanças nos algoritmos.

3. Problemas relacionados

Alguns problemas similares já foram considerados na literatura. Dados vetores X e Y , a soma cartesiana $X + Y$ é a matriz cujo elemento na linha i coluna j tem valor $x_i + y_j$. Três problemas foram bastante estudados para matrizes $X + Y$: ordenação, busca e seleção.

No problema de ordenação, a entrada são os vetores X e Y e deseja-se ordenar totalmente os elementos da matriz $X + Y$. Note que há n^2 elementos a serem ordenados, assim, um limite inferior natural é $\Omega(n^2)$. Em [Fredman 1976] é fornecida uma prova não construtiva de que os elementos de $X + Y$ podem ser ordenados utilizando-se $O(n^2)$ comparações, mas apenas em [Lambert 1992] é apresentado o primeiro algoritmo que resolve o problema neste número de comparações, embora o tempo total do algoritmo seja $O(n^2 \log n)$.

O problema de busca consiste em, dado um valor z , encontrar índices i e j tais que $a_i + b_j = z$. Em [Cosnard et al. 1990] é demonstrado que este problema pode ser resolvido em tempo $\Theta(n)$.

Finalmente o problema de seleção tem como entrada os vetores X e Y e um inteiro k , $1 \leq k \leq n^2$ e consiste em determinar o índices i e j tais que $x_i + y_j$ seja o k -ésimo elemento de ordenação total dos elementos de $X + Y$. Algoritmos para este problema foram apresentados resolvendo-o em tempo $O(n)$ [Frederickson and Johnson 1982, Mirzaian and Arjomandi 1985].

4. Algoritmos

Algoritmo trivial (AT). A ideia mais ingênua consiste em avaliar $f(i, j)$ para todos os n^2 pares de índices válidos e retornar o par que minimiza a soma dos valores correspondentes nos vetores. Este algoritmo claramente leva tempo $O(n^2)$. Uma otimização simples seria evitar testes de pares dominados por um par compatível já encontrado. Isto é, se (i, j) é uma solução válida já encontrada, não é necessário testar nenhum par $(i + s, j + t)$, para s e t positivos.

Algoritmo com fila de prioridades (AFP). Outra estratégia consiste em avaliar a função f em ordem crescente de $x_i + y_j$. Assim, tão logo um par válido seja encontrado, o

algoritmo encerraria a busca, uma vez que esta seria a solução ótima. Isto pode ser feito através de uma fila de prioridades. Inicialmente, apenas o par $(1, 1)$ é colocado na fila de prioridades e, a cada iteração, um par (i, j) é removido da fila e os pares $(i+1, j)$ e $(i, j+1)$ são adicionados à fila, se ainda não tiverem sido adicionados anteriormente e se os índices forem no máximo n . Alguns detalhes de implementação devem ser observados, mas o algoritmo pode ser implementado em tempo $O(k \times \log(k))$ utilizando heaps binomiais.

Note que nenhum dos algoritmos apresentados domina o outro. Para valores pequenos de k ($k = O(n)$), AFP é melhor que AT. Por outro lado, no caso em que $k = O(n^2)$, AT é melhor que AFP. Motivados por esta não dominância, na próxima seção apresentaremos algoritmos cuja complexidade depende de k , como em AFP, mas que domina AT, igualando sua complexidade no pior caso, porém melhor no caso geral.

5. Algoritmos de complexidade $O(n \times \sqrt{k})$

A ideia dos dois algoritmos apresentados a seguir é avaliar na i -ésima iteração todos os pares com soma menor ou igual que a do par (i, i) . Dado que o a menor posição possível do par (i, i) na ordenação de todos os pares é i^2 os algoritmos garantidamente avaliam o par na posição k no pior caso na iteração \sqrt{k} .

Os algoritmos possuem algumas diferenças fundamentais. O Algoritmo 1 é conceitualmente mais simples e sempre avalia $O(n)$ elementos por iteração, e o número de iterações é limitado por \sqrt{k} . Já o Algoritmo 2 em algumas instâncias pode avaliar apenas $\Theta(k)$ posições, porém ao custo de espaço adicional da ordem de $\Theta(n)$. Ainda assim, embora o Algoritmo 2 possa ser executado em tempo $O(n \times \sqrt{k})$ em algumas instâncias, é possível construir instâncias nas quais o algoritmo também leva tempo $\Theta(n \times \sqrt{k})$.

6. Trabalhos futuros

O principal problema que deixamos em aberto é determinar se é possível resolver o problema em tempo linear em k e n , como ocorre em alguns problemas relacionados. Além disso, como trabalho futuro, podemos mencionar a comparação, analítica e experimental, em instâncias reais e artificiais, dos algoritmos propostos.

Referências

- Cosnard, M., Duprat, J., and Ferreira, A. G. (1990). The complexity of searching in $X + Y$ and other multisets. *Information Processing Letters*, 34(2):103–109.
- Frederickson, G. and Johnson, D. (1982). The complexity of selection and ranking in $X + Y$ and matrices with sorted columns. *J. of Comput. and Syst. Sci.*, 24(2):197–208.
- Fredman, M. L. (1976). How good is the information theory bound in sorting? *Theoretical Computer Science*, 1(4):355–361.
- Lambert, J.-L. (1992). Sorting the sums $(x_i + y_j)$ in $O(n^2)$ comparisons. *Theoretical Computer Science*, 103(1):137–141.
- Meisel, F. and Bierwirth, C. (2009). Heuristics for the integration of crane productivity in the berth allocation problem. *Transportation Research Part E: Logistics and Transportation Review*, 45(1):196 – 209.
- Mirzaian, A. and Arjomandi, E. (1985). Selection in $X + Y$ and matrices with sorted rows and columns. *Information processing letters*, 20(1):13–17.

Algoritmo 1: Avalia uma linha e uma coluna por iteração.

```

início
  se  $f(1, 1) = 1$  então
    | retorna (1,1)
   $i \leftarrow 2, best \leftarrow +\infty$ 
  enquanto  $i \leq n$  e  $best > x_{i-1} + y_{i-1}$  faça
    | para  $j \leftarrow i$  até  $n$  faça
      | se  $f(i-1, j) = 1$  e  $x_{i-1} + y_j < best$  então
        | |  $best \leftarrow x_{i-1} + y_j$ 
        | |  $bestpair \leftarrow (i-1, j)$ 
      | se  $f(j, i-1) = 1$  e  $x_j + y_{i-1} < best$  então
        | |  $best \leftarrow x_j + y_{i-1}$ 
        | |  $bestpair \leftarrow (j, i-1)$ 
      | se  $f(i, i) = 1$  e  $x_i + y_i < best$  então
        | |  $best \leftarrow x_i + y_i$ 
        | |  $bestpair \leftarrow (i, i)$ 
    |  $i \leftarrow i + 1$ 
  se  $best < +\infty$  então
    | retorna  $bestpair$ 
  senão
    | retorna “não existe solução”

```

Algoritmo 2: Avalia apenas elementos menores que (i, i)

```

início
   $i \leftarrow 1, best \leftarrow +\infty$ 
  para  $j \leftarrow 1$  to  $n$  faça
    |  $contourline[j] \leftarrow 1$ 
  enquanto  $i \leq n$  e  $best = +\infty$  faça
    | para  $j \leftarrow 1$  to  $n$  faça
      | enquanto  $contourline[j] < n$  e  $x_j + y_{contourline[j]} \leq x_i + y_i$  faça
        | se  $f(j, contourline[j]) = 1$  e  $x_j + y_{contourline[j]} < best$  então
          | |  $best \leftarrow x_j + y_{contourline[j]}$ 
          | |  $bestpair \leftarrow (j, contourline[j])$ 
          | |  $countline[j] \leftarrow contourline[j] + 1$ 
        |  $i \leftarrow i + 1$ 
  se  $best < +\infty$  então
    | retorna  $bestpair$ 
  senão
    | retorna “não existe solução”

```

The Geodesic Carathéodory Number

Eduardo S. Lira¹, Diane Castonguay¹, Erika M. M. Coelho¹, Hebert Coelho¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia – 74.690-900 – Goiânia – GO – Brazil

{eduardolira,diane,erikamorais,hebert}@inf.ufg.br

Abstract. From Carathéodory's theorem arises the definition of the Carathéodory number for graphs. This number is well-known for monophonic and Triangle-path convexities. It is limited for some classes of graphs on P_3 and geodesic convexities but is known to be unlimited only on P_3 -convexity. In this paper, we prove that the Carathéodory number is unlimited on geodesic convexity.

Resumo. Do teorema de Carathéodory surge a definição do número de Carathéodory para grafos. Este número é bem conhecido nas convexidades monofônica e de caminho de triângulos. Ele é limitado para algumas classes de grafos nas convexidades P_3 e geodésica, mas apenas na convexidade P_3 sabe-se que ele é ilimitado. Neste artigo, nós provamos que o número de Carathéodory é ilimitado na convexidade geodésica.

1. Introduction

In 1911, Constantin Carathéodory published a theorem stating that every u in the convex hull of a subset $S \subseteq R^d$, is also in the convex hull of a subset F of S of order at most $d + 1$ [Carathéodory 1911]. The ideas provenient from this theorem, brought to light an invariant in graph theory, called Carathéodory number, widely studied since then. We will focus on the study of this number on geodesic convexity.

In this article, we used the following definition for the Carathéodory number of a graph. A subset S of vertices is a *Carathéodory set* if there exists a vertex in the convex hull of S which does not belong to the convex hull of the $S \setminus \{s\}$ for every $s \in S$. The *Carathéodory number* $c(G)$ of a graph G is the maximum cardinality of a Carathéodory set [E. M. M. Coelho and Szwarcfiter 2013].

The Carathéodory number is well-known for some convexities. On monophonic convexity, $c(G) = 1$ for complete graphs, and 2 for other graphs [Duchet 1988]. On Triangle-path convexity $c(G) = 2$ [Changat and Mathew 1999]. On P_3 -convexity, $c(G) \leq 3$ for multipartite tournaments [D.B. Parker and Wolf 2008], however, it is unlimited considering general graphs [E. M. M. Coelho and Szwarcfiter 2013]. Considering the geodesic convexity, [M.C. Dourado 2013] showed that it is NP-complete to determine whether $c(G)$ is at least a given k , even when restricted to bipartite graphs, and that $c(G)$ is at most 3 for split graphs.

This article brings a new result on this matter: the Carathéodory number is unlimited on geodesic convexity.

2. Preliminaries

Let G be a finite graph given by a set of vertices $V(G)$ and a set of edges $E(G)$. A set \mathcal{C} of subsets of $V(G)$ is a *convexity* if $\emptyset, V(G) \in \mathcal{C}$ and \mathcal{C} is closed under intersections. The elements of \mathcal{C} are called *convex sets*. Given a set $S \subset V$, the *convex hull* of S is the smallest convex set $H(S) \in \mathcal{C}$, such that $S \subseteq H(S)$.

Several convexities are defined by a set \mathcal{P} of paths in a graph. In this scenario, a subset $\mathcal{C} \in V(G)$ is convex when for all $x, y \in \mathcal{C}$, if $p \in \mathcal{P}$ is a path between x and y , then all vertices in p are also in \mathcal{C} .

The *geodesic convexity* is obtained when \mathcal{P} is the set of all geodesics, i.e., all the shortest paths. A *chord* of a path p in a graph G is any edge joining a pair of nonadjacent vertices of p . The *monophonic convexity* is defined using the set of all chordless paths, i.e., induced paths. The *triangle-path convexity* arise from the set of all paths in which every chord is a *short chord*, i.e., chords joining vertices at distance 2 apart. And the P_3 -convexity, also known as *two-path convexity*, is obtained from the set of all paths of length 2.

The *Carathéodory number* of a graph G , written as $c(G)$, is the smallest integer k such that for every subset S of $V(G)$ and every element s in $H(S)$, there is a subset F of S with $|F| \leq k$ and $s \in H(F)$.

A subset $S \subset V(G)$ is called *Carathéodory set* when $\partial H(S) = H(S) \setminus \bigcup_{s \in S} H(S \setminus \{s\})$ is nonempty. We can also define the Carathéodory number as the maximum cardinality of a Carathéodory set [E. M. M. Coelho and Szwarcfiter 2013].

3. Geodesic Carathéodory Number is Unlimited

We now construct inductively a family of graphs G_i , with $i \geq 1$ that have an unlimited Carathéodory number on geodesic convexity:

- G_1 is the graph with $V(G_1) = \{v_1\}$, and $E(G_1) = \emptyset$;
- G_2 is the graph with $V(G_2) = V(G_1) \cup \{v_2, v_3\}$, and $E(G_2) = E(G_1) \cup \{v_1v_3, v_2v_3\}$;
- G_3 is the graph with $V(G_3) = V(G_2) \cup \{v_4, v_5, v_6\}$, and $E(G_3) = E(G_2) \cup \{v_1v_4, v_2v_4, v_3v_4, v_3v_5, v_4v_6, v_5v_6\}$;
- G_{i+1} , for $i \geq 3$, is the graph with $V(G_{i+1}) = V(G_i) \cup \{v_{n-2}, v_{n-1}, v_n\}$, and $E(G_{i+1}) = E(G_i) \cup \{v_{n-4}v_{n-2}, v_{n-3}v_{n-2}, v_{n-2}v_{n-1}, v_{n-1}v_n, v_4v_n\}$ where $n = 3i$.

Theorem 1. *Every G_i , constructed as above, has a Carathéodory set of cardinality i .*

Proof. For $i = 1$, $S_1 = \{v_1\}$ is a Carathéodory set of G_1 . For $i = 2$, $S_2 = \{v_1, v_2\}$ and $v_3 \in \partial H(S_2)$. We will now show by induction that $S_i = \{v_1, v_2\} \cup \{v_{3j} \mid 2 \leq j \leq i\}$ is a Carathéodory set of G_i , with cardinality i , for $i \geq 3$. For $i = 3$, note that $v_5 \in \partial H(S_3)$. For $i = 4$, we have that $v_7, v_8 \in \partial H(S_4)$ and for $i = 5$, one can see that $\partial H(S_5) = \{v_{10}, v_{11}\}$. This proves that S_i is a Carathéodory set of G_i , with cardinality i , for $1 \leq i \leq 5$.

Take $i \geq 5$ and assume (as inductive hypothesis) that the longest geodesic of G_i is at most 4, and S_i is a Carathéodory set of cardinality i , with $\{v_{n-5}, v_{n-4}\} = \partial H(S_i)$, where $n = 3i$.

Take $S_{i+1} = S_i \cup \{v_n\}$, and note that all vertices in S_{i+1} are connected to v_4 , and that any other vertex in $V(G_{i+1})$ is connected to at least one vertex in S_{i+1} , by construction of G_{i+1} . Therefore, the longest geodesic of G_{i+1} is at most 4. There are three edges connecting the vertices in $V(G_i)$ to those new vertices added to build G_{i+1} , namely $v_{n-4}v_{n-2}$, $v_{n-3}v_{n-2}$, and v_4v_n . Choose any two vertices x, y in $V(G_i)$, other than v_4 (note that the geodesic distance of any vertex to v_4 is at most 2). If you take a path between x and y that uses both edges $v_{n-4}v_{n-2}$ and $v_{n-3}v_{n-2}$, it is bigger than their geodesic in at least one unit, since $v_{n-4}v_{n-3} \in E(G_i)$. If you take a path between x and y that uses $v_{n-4}v_{n-2}$ or $v_{n-3}v_{n-2}$, and v_4v_n , this path is bigger than 4. So, by inductive hypothesis, in G_{i+1} no shorter path was introduced between two vertices of G_i , implying that $H(S_i)$ in G_{i+1} is exactly $V(G_i)$.

All the geodesics between v_n and another vertex in S_{i+1} have size 2, and between v_n and another vertex not in S_{i+1} have size at most 3. Take any vertex in $V(G_i)$ that is not in S_{i+1} , other than v_{n-4} . If a path between such vertex and v_n uses either edge $v_{n-4}v_{n-2}$ or $v_{n-3}v_{n-2}$, it is bigger than 3, so it is not a geodesic. The geodesics between v_{n-4} and v_n are $v_{n-4}v_{n-2}v_{n-1}v_n$ and $v_{n-4}v_{n-3}v_4v_n$. This means the only minimum path between a vertex in $V(G_i)$ and v_n that uses the vertices v_{n-2} and v_{n-1} is the one between v_{n-4} and v_n . Without loss of generality, using the inductive hypothesis, we can conclude that $\partial H(S_{i+1}) = \{v_{n-2}, v_{n-1}\}$ (meaning S_{i+1} is a Carathéodory set), and $|S_{i+1}| = i + 1$, completing the proof. □

Corollary 1. *The Carathéodory number on geodesic convexity is unlimited.*

Proof. By Theorem 1, for all integer $i \geq 1$, there is a Carathéodory set of cardinality i in G_i . We conclude that $c(G_i) \geq i$ and thus yields the results. □

References

- Carathéodory, C. (1911). Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen. In *Rend. Circ. Mat. Palermo*, pages 193 – 217.
- Changat, M. and Mathew, J. (1999). On triangle path convexity in graphs. In *Discrete Math*, pages 91–95.
- D.B. Parker, R. W. and Wolf, M. (2008). On two-path convexity in multipartite tournaments. In *European J. Combin.*, pages 641–651.
- Duchet, P. (1988). Convex sets in graphs ii. minimal path convexity. In *J. Combin Theory, Ser. B*, pages 307–316.
- E. M. M. Coelho, M.C. Dourado, D. R. and Szwarcfiter, J. (2013). The Carathéodory number of the p_3 convexity of chordal graphs. In *The Seventh European Conference on Combinatorics, Graph Theory and Applications*, pages 209–214.
- M.C. Dourado, D. Rautenbach, e. a. (2013). On the Carathéodory number of interval and graph convexities. In *Theoretical Computer Science*, pages 127–135.

Padrão de Desenho para a Família de Grafos Gêmeos

Cassio Ferreira Merlo¹, Mauro Pinheiro Rodrigues¹, Marcia Helena Moreira Paiva²

¹Laboratório de Projetos em Design (ProDesignUfes)
Universidade Federal do Espírito Santo (UFES) – Vitória – ES – Brasil

²Laboratório de Telecomunicações (LabTel)
Universidade Federal do Espírito Santo (UFES) – Vitória – ES – Brasil

cassiofmerlo@gmail.com, {mauro.pinheiro,marcia.paiva}@ufes.br

Abstract. *This paper is about the critical thinking involved in the development of a graphic pattern for the family of twin graphs, considering as its main objective the characterization of the family, and the ease of design. This work is based on the search for developing a visual pattern for this purpose, starting from empty until the final form.*

Resumo. *Este artigo trata da reflexão crítica envolvida no desenvolvimento de um padrão gráfico para a família de grafos gêmeos, considerando como seu objetivo principal a caracterização da família e a facilidade de concepção. Este trabalho é baseado na busca pelo desenvolvimento de um padrão visual para tal propósito, partindo do vazio até a forma final.*

1. Introdução

Um grafo $G = (V, A)$ é uma estrutura consistindo em um conjunto de vértices V e um conjunto de arestas A , onde os vértices, que também podem ser chamados de nós, são pontos distribuídos no espaço, e as arestas são linhas representando relações entre pares de vértices. Em outras palavras, grafos são conjuntos tais que sua representação geométrica se dá pela relação entre vértices e arestas, de forma que as arestas estejam conectadas em função de relações entre os vértices.

Neste trabalho buscou-se, a princípio, identificar e estudar as características de grafos, particularmente grafos gêmeos, estudando-se a viabilidade de uma representação padrão para esta família de grafos, que possui diversas características interessantes para problemas de roteamento com proteção e, em particular, para o projeto de redes de telecomunicações [PAIVA et al. 2013]. Redes de telecomunicações devem atender certos requisitos como confiabilidade, desempenho e custo. A partir da modelagem via grafos, é possível definir as características topológicas de uma rede de telecomunicações, traduzidas em parâmetros provenientes da teoria de grafos, tais como diâmetro, grau máximo e conectividade. Em geral usa-se ferramentas computacionais tais como yEd Graph Editor¹ ou NetworkX² para desenho automático de grafos. Entretanto, nem sempre o desenho do grafo obtido por estas ferramentas permite a identificação imediata das características topológicas da rede de telecomunicações associada. Nesse contexto, este trabalho busca desenvolver uma representação padrão para os grafos gêmeos, a exemplo da representação padrão de árvores proposta em [HARARY 1969].

¹www.yworks.com/products/yed

²networkx.github.io

Grafos gêmeos são grafos 2-geodesicamente conexos minimais com respeito ao número de arestas [FARLEY e PROSKUROWSKI 1997], ou seja, são grafos que possuem pelo menos duas geodésicas disjuntas (i.e., dois *menores* caminhos disjuntos) para cada par de vértices não-adjacentes. Todos os grafos gêmeos são bipartidos e planares. Com exceção dos grafos gêmeos, os únicos grafos 2-geodesicamente conexos minimais são o C_3 (ciclo de ordem 3) e o Q_3 (cubo de ordem 8) [FARLEY e PROSKUROWSKI 1997].

A partir de reflexões sobre essas características, foi possível entender que essas propriedades deveriam ser preservadas na concepção de um modelo visual. Ao mesmo tempo foi considerado que a facilidade de identificação e leitura do grafo deveriam ser uma característica, auxiliando assim o usuário no reconhecimento do mesmo.

2. Primeiros esboços

Utilizando-se do processo de resolução de problemas proposto em [LOBACK 2001], decidiu-se iniciar tal desenvolvimento por meio de esboços manuais, seguido de estudos e desenhos com o programa yEd Graph Editor. A partir dessa decisão, realizou-se um levantamento gráfico de grafos gêmeos existentes, que apresentassem características semelhantes às encontradas anteriormente, mas com desenhos diferentes. Foram encontrados alguns poucos desenhos publicados no artigo [PAIVA et al. 2013], um dos quais é reproduzido na Figura 1. Esses primeiros desenhos serviram como início para os primeiros estudos e esboços de um modelo adequado de visualização dos grafos.

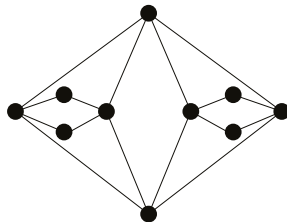


Figura 1. Desenho inicial encontrado em [PAIVA et al. 2013].

Simultaneamente aos desenhos iniciais, foi observado como o grafo foi percebido por um grupo de usuários, principalmente como sua visualização espacial e fisionomia são agradáveis na percepção e apesar de apresentar parcialmente uma boa aparência, o principal fator negativo observado foi a complexidade para a construção do grafo a partir do vazio, tornando a tarefa complicada e demorada. Concluiu-se então, que a adoção desse aspecto visual, tornar-se-ia mais trabalhosa na disseminação e construção de um modelo útil, necessitando então incluir como característica a simplicidade e rapidez na concepção estrutural do grafo.

3. Desenvolvimento

Para demonstração da construção do novo padrão visual, utilizamos como base o primeiro grafo encontrado no artigo [PAIVA et al. 2013]. Primeiramente, precisamos observar e analisar o grafo quanto a quantidade de vértices e o grau de cada um, ou seja, a quantidade de ligações de cada um. Na Figura 2 observamos que o grafo é composto por 10 vértices, sendo 4 vértices com 2 ligações e 6 vértices com 4 ligações. A partir dessa informação é possível construir o grafo no novo padrão.

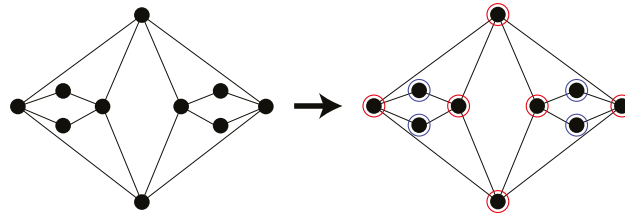


Figura 2. Primeira análise, observando-se os graus dos vértices.

Para a construção do novo modelo, primeiramente divide-se o número de vértices por dois; no caso do grafo que estamos usando, obtemos o valor cinco. Em seguida, desenha-se um quadrado com 4 vértices. Os dois vértices superiores levam os dois primeiros números sequenciais (1 e 2, no caso), e os vértices inferiores levaram os dois primeiros números sequenciais a partir da metade do número de vértices (6 e 7, no caso). Ao mesmo tempo, determinou-se um padrão onde os números ímpares são expostos no lado esquerdo e os pares do lado direito. O resultado desse processo é ilustrado na Figura 3.

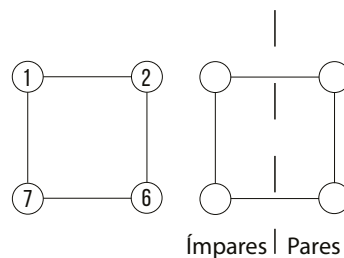


Figura 3. Passo-a-passo inicial.

Dessa forma, é necessário somente organizar os números em seqüências. Na parte superior deve estar a primeira metade dos vértices (até o vértice 5), na metade inferior os demais. Após isso, inicialmente é necessário ligar os vértices com lado de menor quantidade de vértices, ao outro lado superior, por exemplo ligar o vértice 2 ao vértice 3, em seguida ligar ao lado oposto do mesmo, vértice 4, sequencialmente repetiremos o primeiro passo, ligaremos o vértice 4 ao vértice 5. Esses passos são ilustrados na Figura 4. Posteriormente, deve-se fazer o mesmo procedimento com todos os lados do grafo. A Figura 5 ilustra o resultado final obtido por esse procedimento, que corresponde ao novo padrão gráfico proposto a partir do grafo da Figura 1.

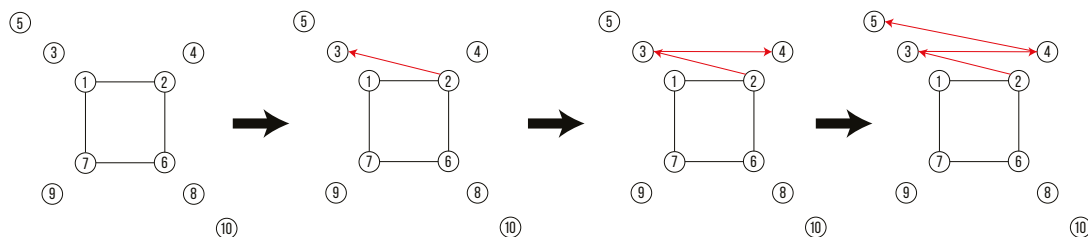


Figura 4. Passo-a-passo final.

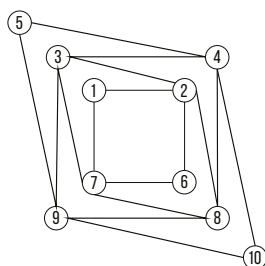


Figura 5. Padrão final de desenho do grafo gêmeo mostrado na Figura 1.

4. Resultados

Com o desenvolvimento da pesquisa foi possível observar o processo de construção de um padrão para grafos gêmeos, preservando suas características e sendo adequado às propostas a serem desenvolvidas. Além disso, a pesquisa levou à criação de um modelo final de padrão gráfico a ser explorado pelos usuários. Tal produto servirá como insumo para novos estudos e pesquisas, bem como ampliará o acervo gráfico de grafos gêmeos, se tornando patrimônio público imaterial da comunidade científica.

Como continuação deste trabalho, estamos elaborando um algoritmo para o desenho automático de grafos gêmeos, a partir do método aqui apresentado. Simultaneamente com o insumo produzido, a pesquisa apresenta um processo metodológico que pode ser utilizado em outros casos semelhantes, sendo uma contribuição para futuras pesquisas na área de desenvolvimento de padrões para desenho de grafos.

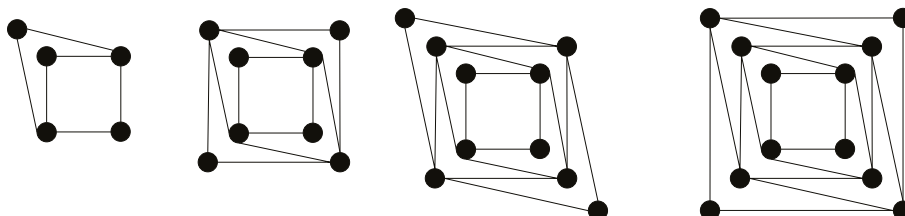


Figura 6. Ilustrações do novo padrão obtido para o desenho de grafos gêmeos.

Agradecimentos

Este trabalho é parcialmente financiado pelo CNPq (processo 462477/2014-2) e pela CAPES (processo PNP20130772).

Referências

- FARLEY, A. e PROSKUROWSKI, A. (1997). Minimum self-repairing graphs. *Graphs and Combinatorics*, 13:345–351.
- HARARY, F. (1969). *Graph Theory*. Addison-Wesley Series in Mathematics. Addison-Wesley Publishing Company.
- LOBACK, B. (2001). *Design industrial: bases para a configuração dos produtos industriais*. Ed. Blucher, São Paulo.
- PAIVA, M., CAPOROSSI, G., e SEGATTO, M. (2013). Twin graphs for OTN physical topology design. *Les Cahiers du GERAD*, 48:1–12.

An efficient algorithm for the Closest String Problem

Omar Vilca¹, Rosiane de Freitas¹

¹Institute of Computation – Federal University of Amazonas (UFAM)
Manaus, AM – Brasil

{omarlatorre, rosiane}@icomp.ufam.edu.br

Abstract. *The closest string problem that arises in computational molecular biology and coding theory is to find a string that minimizes the maximum Hamming distance from a given set of strings, the CSP is NP-hard problem. This article proposes an efficient algorithm for this problem with three strings. The key idea is to apply normalization for the CSP instance. This enables us to decompose the problem in five different cases corresponding to each position of the strings. Furthermore, an optimal solution can be easily obtained in linear time. A formal proof of the algorithm will be presented, also numerical experiments will show the effectiveness of the proposed algorithm.*

1. Introduction

String selection problems are among the most important topics facing researchers in computational biology. Combinatorial optimization is a possible approach to solving selection sequences problems. There are some previously exact algorithms for CSP with 3-sequences [Gramm et al. 2001, Liu et al. 2011], in both papers presented algorithms in linear time. Meneses obtains optimal solutions via integer programming [Meneses et al. 2004, Meneses et al. 2005], In [Vilca 2013] solves this problem by cutting planes algorithm.

The CSP is defined as follows: Given a finite set $S = \{s^1, s^2, \dots, s^n\}$ of strings with alphabet Σ , each string with length m , find a center string t of length m minimizing d , such that for every string $s^i \in S$, $d_H(t, s^i) \leq d$, we mean the Hamming distance between t and s^i . This is a NP-hard problem according to Frances and Litman [Frances and Litman 1997]. In the following, we show a formal proof of correctness from an efficient linear time algorithm for CSP instances with 3-sequences, computational experiments are reported, and finally, concluding remarks are presented.

2. Efficient linear time algorithm for 3-sequences

The 3-CSP-A algorithm is designed based on the isomorphic instance [Gramm et al. 2001]. An exact algorithm for 3-CSP with alphabet $\Sigma = 2$ is found in [Liu et al. 2011].

Definition 1 (Normalized instance) *Let S an instance, that is, $S = \{s^1, \dots, s^n\}$, where $|s^i| = m$, with $1 \leq i \leq n$ and $1 \leq j \leq m$. Let $M_{n \times m}$ a matrix of characters from S each*

column is a position of the strings in S , so we have $M[c_1, c_2, \dots, c_m]$

$$\phi(c'_j)_{1 \leq i \leq n, 1 \leq j \leq m} = \begin{cases} \lambda_1 & \text{if } \sigma_1 \in c_j : \max_{\sigma_i} |c'_j|^i = \sigma_1 | \\ \lambda_2 & \text{else if } \sigma_2 \in c_j : \max_{\sigma_i} |c'_j|^i \setminus \sigma_1 = \sigma_2 | \\ \dots & \\ \lambda_k & \text{else if } \sigma_k \in c_j : \max_{\sigma_i} |c'_j|^i \setminus \{\sigma_1, \dots, \sigma_{k-1}\} = \sigma_k | \end{cases}$$

Let $M'[c'_1, \dots, c'_m]$, and S' an instance based on characters from M' , as a result S' is called normalized instance from S .

Theorem 1 Let a CSP instance with 3-sequences, which denotes 3-CSP, that is, $S = \{s^i \in \Sigma^m, i = 1, 2, 3\}$ with alphabet $|\Sigma| > 2$, so the 3-CSP-A algorithm always finds an exact solution to 3-CSP.

Proof 1 Let $S = \{s^1, s^2, s^3\}$ a 3-CSP instance, let ϕ a bijective function that transforms S into S' , called normalized instance that is equivalent to S for ϕ according to [Gramm et al. 2001]. Let $M_{3 \times m}$ a matrix of characters from S each column is a position of the strings in S , so we have $M[c_1, c_2, \dots, c_m]$ with m length of strings in S

$$\phi(c_j)_{1 \leq i \leq 3, 1 \leq j \leq m} = \begin{cases} a & \text{if } \sigma_1 \in c_j : \max_{\sigma_i} |c_j|^i = \sigma_1 | \\ b & \text{else if } \sigma_2 \in c_j : \max_{\sigma_i} |c_j|^i \setminus \sigma_1 = \sigma_2 | \\ c & \text{else if } \sigma_3 \in c_j : \max_{\sigma_i} |c_j|^i \setminus \{\sigma_1, \sigma_2\} = \sigma_3 | \end{cases}$$

We assume wlog five different cases in c_i , that is, $v_1 = aaa, v_2 = baa, v_3 = aba, v_4 = aab$ and $v_5 = abc$. In order to get an optimal solution, divide $M'[c_1, c_2, \dots, c_m]$ in blocks of 3-length and 2-length, consider all possible combinations of $\{v_2, \dots, v_5\}$, then drop v_1 since it is a trivial case.

In the case of 3-length blocks with repetition columns we have, $d_H(v_i v_j v_k, baa) = 2$ and without repetitions $d_H(v_i v_j v_k, aaa) = 1$ for $2 \leq i, j, k \leq 4$ now inserting v_5 in our analysis with column repetition $d_H(\{v_5 v_i v_j, v_i v_5 v_j, v_i v_j v_5\}, aaa) = 2$, and finally $d_H(v_5 v_5 v_5, abc) = 2$.

In the case of 2-length blocks we have, $d_H(v_i v_j, aa) = 1$ and $d_H(v_i v_i, ab) = 1$ with $2 \leq i, j \leq 4$ after add v_5 , $d_H(v_2 v_5, aa) = 1, d_H(v_3 v_5, ab) = 1, d_H(v_4 v_5, ac) = 1$, and finally $d_H(v_5 v_5, aa) = 2$.

So in truth we are interested in the cases when the Hamming distance is equals to 1. Let l_j a number that accounts for each time the case v_j is repeated where $j \in \{1, 2, \dots, 5\}$, let t' a string of size m , an optimal solution of S' , in order to minimize the maximum Hamming distance $d_H(t', S')$, we verified various reductions for each case v_j presented. We are checking each kind of reductions for 2-length and 3-length blocks.

Assume wlog $l_2 \leq l_3 \leq l_4$, then $l_3 = l_3 - l_2, l_4 = l_4 - (l_2 + l_3)$, after these calculations, one of them is greater or equal to zero. let $\rho_{iab} = \lfloor \frac{1}{2} l_i \rfloor$ for $2 \leq i \leq 4$

- If $l_5 \bmod 3 > 0$ and $\{l_2, l_3, l_4\} \bmod 2 > 0$ then $\{\rho_{5ac}, \rho_{ab}, \rho_{aa}\} = 1$
- If $l_5 \bmod 2 > 0$ then $\rho_{5ab} = 1$

Let t a string that represents an optimal solution of 3-CSP-A, with $1 \leq i \leq m$, we

have:

$$\begin{array}{ll}
 s_i^1 = s_i^2 & t_i = s_i^1 = s_i^2 \\
 s_i^1 = s_i^2 & s_i^2 \neq s_i^3 & \rho_{2ab} > 0 & t_i = s_i^3 \\
 s_i^1 \neq s_i^2 & s_i^1 = s_i^3 & \rho_{3ab} > 0 & t_i = s_i^2 \\
 s_i^1 \neq s_i^2 & s_i^1 \neq s_i^3 & s_i^2 = s_i^3 & \rho_{4ab} > 0 & t_i = s_i^1 \\
 s_i^1 \neq s_i^2 & s_i^2 \neq s_i^3 & s_i^2 \neq s_i^3 & \rho_{5ab} > 0 & t_i = s_i^2 \\
 s_i^1 \neq s_i^2 & s_i^2 \neq s_i^3 & s_i^2 \neq s_i^3 & \rho_{5ac} > 0 & t_i = s_i^3 \\
 s_i^1 \neq s_i^2 & s_i^2 \neq s_i^3 & s_i^2 \neq s_i^3 & \rho_{ab} > 0 & t_i = s_i^2 \\
 s_i^1 \neq s_i^2 & s_i^2 \neq s_i^3 & s_i^2 \neq s_i^3 & \rho_{aa} > 0 & t_i = s_i^1
 \end{array}$$

Thus the theorem holds.

Example 1 Let S_1 a 3-CSP instance, and let S_2 its normalized instance, we have:

$$S_1 = \begin{cases} AGTATTGGTG \\ CCCTTTGAGA \\ TAGTGGGTCT \end{cases} \quad S_2 = \begin{cases} aaabaaaaaa \\ bbbaaaabbb \\ cccabbaccc \end{cases}$$

After using the 3-CSP-A algorithm we have an optimal solution for the normalized instance S_2 , that is, $t = caccaaaaa$, with Hamming distance $d_H(t, S_2) = 4$.

Algorithm 1: Linear time algorithm 3-CSP-A, for 3-sequences.

```

Procedure Algorithm 3-CSP-A ( $n, m, S$ )
 $n$  : number of strings
 $m$  : size of strings
 $S$  : a 3-CSP instance, that is,  $S = \{s^1, s^2, s^3\}$ 
Input : Normalized Instance  $S = \{s^1, s^2, s^3\}$ 
Output: optimal solution  $t \in \Sigma^m : d_H(t, S) \leq d$ 
// Let  $v_i$ : number of times that {aaa, aab, aba, baa, abc} appears in the  $j$ th column with  $1 \leq i \leq 5$  and  $1 \leq j \leq m$ 
if  $|S| = 3$  then
     $smallest \leftarrow \text{getSmallest}(v_2, v_3, v_4)$ ;
    // subtract the lowest value for cases: {aab, aba, baa}
     $v_2 \leftarrow v_2 - smallest$ ;
     $v_3 \leftarrow v_3 - smallest$ ;
     $v_4 \leftarrow v_4 - smallest$ ;
    for  $i=1$  to  $m$  do
        if  $s_i^1 = s_i^2$  then
             $t_i \leftarrow s_i^1$ ;
        end if
        else if  $s_i^2 = s_i^3$  then
             $t_i \leftarrow s_i^2$ ;
        else if  $s_i^1 = s_i^3$  then
             $t_i \leftarrow s_i^1$ ;
            // when all the characters are different, case: abc
            else if  $v_4 > 0$  then
                 $t_i \leftarrow s_i^1$ ;  $v_4 \leftarrow v_4 - 1$ ;
                else if  $v_3 > 0$  then
                     $t_i \leftarrow s_i^2$ ;  $v_3 \leftarrow v_3 - 1$ ;
                else if  $v_2 > 0$  then
                     $t_i \leftarrow s_i^3$ ;  $v_2 \leftarrow v_2 - 1$ ;
                else
                     $t_i \leftarrow s_i^j$ ; //where  $j \in \{1, 2, 3\}$  one of them each time
                end if
            end if
        end if
    end for
end if

```

Instance m	2 Characters			4 Characters			20 Characters		
	seed	Val	time(s)	seed	Val	time(s)	seed	Val	time(s)
100	542319	27	< 1	791034	48	< 1	691425	63	< 1
200	7121	52	< 1	52151	91	< 1	351554	124	< 1
300	64874	77	< 1	68724	132	< 1	98121	183	< 1
400	6487	112	< 1	193185	180	< 1	246754	244	< 1
500	94115	124	< 1	15364	215	< 1	658745	311	< 1
600	5419	153	< 1	5419	260	< 1	154525	373	< 1
700	43212	180	< 1	524514	306	< 1	487754	437	< 1
800	2454	215	< 1	55364	354	< 1	754812	494	< 1
900	645387	234	< 1	6487	389	< 1	722451	557	< 1
1000	94315	260	< 1	153364	444	< 1	34567	616	< 1
2000	264554	508	< 1	6487	881	1	65743	1219	1
3000	68174	765	2	2454	1305	2	4432	1857	1
4000	4212	1003	3	53214	1760	3	543	2484	3
5000	722312	1278	5	934145	2188	6	344	3099	5

Table 1. Linear time algorithm 3-CSP-A, for 3-sequences.

3. Computational Experiments and Results

Some computational experiments was done involving forty two 3-CSP instances, using the instance generator described in the literature [Meneses et al. 2004]. These instances consider alphabets with two, four and twenty characters, which obtained optimal solutions using the 3-CSP-A algorithm presented in the Section 2. The header on Table 2 has the following meanings: the first column indicates the tested instance, indicating the parameters (m) size of string; the columns (2,4, and 20 Characters) indicates the (seed) number to generate a random instance, (Val) values of optimal solutions to 3-sequences for binary, DNA, and protein types as well as their execution times in seconds.

4. Final Remarks

We proposed an exact algorithm for the special case of CSP with 3-sequences and alphabet size $|\Sigma| > 2$, given the corresponding theoretical analysis.

References

- Frances, M. and Litman, A. (1997). On covering problems of codes. *Theory of Computing Systems*, 30(2):113–119.
- Gramm, J., Niedermeier, R., and Rossmanith, P. (2001). Exact solutions for closest string and related problems. In *Proceedings of the 12th International Symposium on Algorithms and Computation*, ISAAC '01, pages 441–453.
- Liu, X., Liu, S., Hao, Z., and Mauch, H. (2011). Exact algorithm and heuristic for the closest string problem. *Computers & Operations Research*, 38(11):1513–1520.
- Meneses, C., Lu, Z., Oliveira, C., and Pardalos, P. (2004). Optimal solutions for the closest string problem via integer programming. *INFORMS Journal on Computing*, 16(4):419–429.
- Meneses, C., Oliveira, C., and Pardalos, P. (2005). Optimization techniques for string selection and comparison problems in genomics. *Engineering in Medicine and Biology Magazine, IEEE*, 24(3):81–87.
- Vilca, O. L. (2013). Métodos para problemas de seleção de cadeias de caracteres. Master's thesis, Universidade Federal do ABC, Santo André, São Paulo.

Stochastic scenario generation: An empirical approach

A. D. Oliveira¹, T. P. Filomena¹

¹Business School, Federal University of Rio Grande do Sul (UFRGS)

alan.delgado@ufrgs.br, tpfilomena@ea.ufrgs.br

Abstract. *We briefly discuss the differences among several methods to generate a scenario tree for stochastic optimization. First, the Monte Carlo Random sampling is presented, followed by the Fitting of the First Two Moments sampling, and lastly the Michaud sampling. Literature results are reviewed, taking into account distinctive features of each kind of methodology. According to the literature results, it is fundamental to consider the problem's unique characteristics to make the more appropriate choice on sampling method.*

1. Introduction

Stochastic optimization (SP) is used to model environments in a more realistic manner when uncertainty is inherent. Many situations in the real world are characterized by random events, which often are viewed as random variables. Random variables are combined into subsets, for that we denote by ξ . So that, we associate a collection of random events with a *probability space*, which contains a collection of all possible events and its probability. From these concepts, we formalize a generic stochastic optimization problem, based on [Dempster et al. 2011], as denoted by equation (1)

$$\max_{x \in X} \int_F f(x, \xi) dF(\xi), \quad (1)$$

where f is the value function defined in terms of both uncertainty and decision spaces, x is the decision variable defined over the feasible set $X \subset \mathbb{R}$, and ξ is a I -dimensional random realization of random variable ξ that is defined by the cumulative distribution function $F : \mathbb{R}^I \rightarrow [0, 1]$. Here, dF represents the probability measure on the probability space F of the underlying multivariate stochastic process.

The literature has discussed the way of finding the best approximation of the continuous distribution; efforts concentrating, for example, on having state-space distribution moments matched [Høyland et al. 2003, Høyland and Wallace 2001], minimizing Wassertein probability metrics [Heitsch and Romisch 2005, Romisch 2003], Latin hypercube sampling [McKay et al. 1979], or Michaud sampling [de Oliveira et al. 2016]. The main goal of this research is to design a discrete-space approximation model which approximates a continuous-space stochastic process. Moreover, empirical tests have been made in order to find the most suitable sampling option, [Dempster et al. 2011, Homem-de-Mello and Bayraksan 2014]. According to [Pflug 2001], continuous problems become easier to be solved if we reduce them to discrete-state multiperiod optimization problems. This logic structure, which contains a pertaining history process, may be seen as a tree. As the scenario generation methodology becomes a key part of process of stochastic optimization, we turn our attention to distinct manners of generating scenarios. We consider the Monte Carlo Random sampling detailed in Section 2, the Fitting of

First Two Moments sampling presented at Section 3, and the Michaud sampling discussed in Section 4. Following [de Oliveira et al. 2016], we concentrate our efforts on generating scenarios for Asset Liability Management (ALM).

2. ALM with the Monte Carlo Random sampling

The traditional way to generate a scenario tree for ALM is through Monte Carlo sampling, where uniformly distributed pseudo-random numbers are appropriately transformed into the target distribution, [Dempster et al. 2011]. Hence, we generate W_1, \dots, W_I random vectors from the standard normal distribution. As noted by [Homem-de-Mello and Bayraksan 2014], in this case, the vectors W_1, \dots, W_I are mutually independent. It is the detail that characterizes this sampling method. This method is more detailed in [Glasserman 2003].

According to [Kouwenberg 2001], though this approach may be quite intuitive, when the states of scenario tree are sampled randomly, the mean and covariance matrix will not be correctly specified in most nodes of the tree. As this information is an input for the optimization model, the optimizer chooses an investment strategy from erratic or miss specified parameters. An obvious way to deal with this problem is to increase the number of nodes in the randomly sample event tree. However, the stochastic program might become computationally intractable due to the exponential growth rate of the tree.

3. ALM with the Fitting the First Two Moments sampling

In this approach, proposed by [Høyland and Wallace 2001], we construct an event tree that fits the mean and the covariance matrix of the underlying distribution. Besides that, [Høyland et al. 2003] provides a source code to this method. The first step is to generate the random vectors from the standard normal distribution, in the same way as described in Section 2. After that, with the I random vectors, we transform them to exhibit a given correlation by pre-multiplying the vectors with lower triangular matrix L of covariance matrix Σ ,

$$W_j^\top = LW_j, \Sigma = LL^\top, j = 1, \dots, I, \quad (2)$$

where L can be obtained by applying Cholesky decomposition. In another words, as mentioned by [Kouwenberg 2001], we specify that the average of the disturbances should be zero, and they should have a covariance matrix equal to Σ . Therefore, we may denote this matching as (3) and (4),

$$\frac{1}{S} \sum_{s=1}^S W_{js} = 0 \quad \forall j \in 1, \dots, I, \quad (3)$$

$$\frac{1}{S-1} \sum_{s=1}^S W_{js} W_{is} = \Sigma_{ij} \quad \forall j, i \in 1, \dots, I, \quad (4)$$

As mentioned by [Kouwenberg 2001, Dempster et al. 2011, Löhndorf 2016], it is possible to argue that this sampling outperform other methods as Monte Carlo Random sampling, Wasserstein Distance sampling, or even Latin Hypercube sampling.

4. ALM with the Michaud Sampling

Another approach is to simulate the generation of scenarios for multiple trees, and to take the solution of an optimization problem of smaller size for each tree. This technique has some similarities with the resampled efficient frontier method proposed by [Michaud and Michaud 2008] for the construction of portfolio of risky securities and it is applied by [de Oliveira et al. 2016]. This sampling has four steps. First, we define the number of trees to be solved for each parametrization, see, e.g., [Michaud and Michaud 2008]. The value of this realizations must be sufficiently large so that the portfolio allocations are stable and small enough, ensuring that the approach does not become computationally prohibitive. Once the number of instances is defined, the second step is to generate the scenarios for each tree in accordance with its model. In Step 3, we solve to optimality the optimization problem corresponding to each tree. In Step 4, we evaluate the results based on the optimal solutions of the trees.

The solutions provided by the resampling avoid corner allocations, i.e. extreme weight allocations, which may be seen in the classical mean-variance portfolio selection model. Furthermore, it decrease the estimation bias, [Fletcher and Hillier 2001]. However, [Scherer 2002] draws attention to when long and short positions are allowed. In this case, the resampled efficient frontier does not present notorious differences from the classical efficient frontier.

5. Final Considerations

In the previous sections, we addressed some literature results to identify and assess strengths and weaknesses of different scenario generation tree methods for stochastic programming. The main objective of these sampling methods is to consider many scenarios while preserving computational tractability. We noticed that main drawback of Monte Carlo Random sampling is the need to assign a large number of scenarios to achieve better results. If we consider, for example, a model with I independent random variables, each with only two possible alternatives; the total number of scenarios is thus 2^I for only one stage, and so even for moderate values of I it becomes impractical to take all possible outcomes into account. In case of the Fitting of the First Two Moments, although it has achieved good performance [Kouwenberg 2001, Löhndorf 2016], [Dempster et al. 2011] outline some weakness when it is applied to the risk control problems given underestimation of the problem's volatility. If we consider the Michaud sampling, the literature have not yet found a consensus on the pros and cons of the results delivered by resampling [Markowitz and Usmen 2003, Ulf and Raimond 2006, Becker et al. 2015]. Furthermore, this method still lacks a theoretical foundation. Therefore, as we can see, due to vulnerabilities of each sampling method, the choice of the appropriate sampling method depends on the problem to be modeled and its specific characteristics.

Referências

- Becker, F., Gurtler, M., and Hibbeln, M. (2015). Markowitz versus Michaud: portfolio optimization strategies reconsidered. *The European Journal of Finance*, 21(4):269–291.
- de Oliveira, A., Filomena, T., Perlin, M., Lejeune, M., and Macedo, G. (2016). A multistage stochastic programming asset-liability management model: an application to the brazilian pension fund industry. *Optimization and Engineering*, forthcoming.

- Dempster, M., Medova, E., and Yong, Y. (2011). Comparison of sampling methods for dynamic stochastic programming. In *Stochastic Optimization Methods in Finance and Energy*, chapter 16, pages 389–425. Springer Science+Business Media, New York.
- Fletcher, J. and Hillier, J. (2001). An examination of resampled portfolio efficiency. *Financial Analysts Journal*, 57(5):66–74.
- Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*, volume 53. Springer-Verlag New York, New York, 1 edition.
- Heitsch, H. and Romisch, W. (2005). Generation of multivariate scenario trees to model stochasticity in power management. In *IEEE St. Petersburg Power Tech*.
- Homem-de-Mello, T. and Bayraksan, G. (2014). Monte carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56–85.
- Høyland, K., Kaut, M., and Wallace, S. (2003). A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2):169–185.
- Høyland, K. and Wallace, S. (2001). Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307.
- Kouwenberg, R. (2001). Scenario generation and stochastic programming models for asset liability management. *European Journal of Operational Research*, 134(2):279 – 292.
- Löhndorf, N. (2016). An empirical analysis of scenario generation methods for stochastic optimization. WU Vienna University of Economics and Business.
- Markowitz, H. and Usmen, N. (2003). Resampled frontiers versus diffuse bayes: An experiment. *Journal of Investment Management*, 1(4):9–25.
- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.
- Michaud, R. and Michaud, R. (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press, 2 edition.
- Pflug, G. (2001). Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89(2):251–271.
- Romisch, W. (2003). Stability of stochastic programming problems. In Ruszczyński, A. and Shapiro, A., editors, *Stochastic Programming, Volume 10 of Handbooks in Operations Research and Management Science*, chapter 8, pages 483–554. Elsevier, Amsterdam.
- Scherer, B. (2002). Portfolio resampling: Review and critique. *Financial Analysts Journal*, 58(6):98–109.
- Ulf, H. and Raimond, M. (2006). Portfolio choice and estimation risk. a comparison of bayesian to heuristic approaches. *Astin Bulletin*, 36(1):135–160.

Análise e Comparação dos Algoritmos de Dijkstra e A-Estrela na Descoberta de Caminhos Mínimos em Mapas de Grade

Marcel L. Rios^{1,2}, Francisco S. S. Neto², José F. Magalhães Netto²

¹Instituto Federal de Rondônia - Campus Porto Velho Calama
(IFRO) - Porto Velho, RO - Brasil

²Programa de Pós Graduação em Informática - Universidade Federal do Amazonas
(UFAM) - Manaus, AM - Brasil

marcel.rios@ifro.edu.br, laertesneto@gmail.com, jfmnetto@gmail.com

Abstract. *This paper presents a comparison between two algorithms for finding shortest paths in grid maps. First of them is Dijkstra, a well known greedy algorithm for finding the shortest path between vertex in a given graph. And the second one is A* (A-star), an algorithm that uses heuristics to predict their behavior, also for going through a graph and finding the shortest path between vertex. For each algorithm a stop condition has been added, showing their pseudocode, analyzing their complexity and showing promising results of A-Star compared to Dijkstra.*

Resumo. *Este trabalho apresenta uma comparação entre dois algoritmos de descoberta de caminhos mínimos em mapas de grade. O primeiro deles é o Dijkstra, um algoritmo guloso muito conhecido por encontrar o caminho mais curto entre vértices em um dado grafo. O segundo é o A* (A-Estrela), um algoritmo que utiliza heurística para prever seu comportamento, também percorrendo um grafo e encontrando o menor caminho entre vértices. Para cada algoritmo foi adicionada uma condição de parada, demonstrando os respectivos pseudocódigos, analisando a complexidade dos mesmos e apresentando resultados promissores do A-Estrela em comparação ao Dijkstra.*

1. Introdução

Os algoritmos de caminho mínimo, por definição, buscam encontrar, a partir de uma posição inicial, o caminho de custo mínimo até uma posição final (alvo). Segundo [Djojo and Karyono 2013] o problema do caminho mínimo é um dos problemas de otimização mais frequentemente estudado e adaptado em várias aplicações.

Segundo [Russell and Norvig 2013] os algoritmos de descoberta de caminhos mínimos podem ser classificados de duas formas: Busca Não Informada, quando o algoritmo não faz uso de heurísticas para descobrir o menor caminho entre origem e destino. E Busca Informada, quando o algoritmo faz uso de heurísticas para estimar o caminho de custo mínimo.

Nessa perspectiva, o presente trabalho busca analisar o comportamento e a complexidade de dois importantes algoritmos: Dijkstra (Busca Não-Informada) e A-Estrela (Busca Informada), ambos aplicados em mapas de grade, cujas arestas possuem custos uniformes. Os resultados demonstraram a eficiência do A-Estrela em face do Dijkstra, evidenciando a importância da heurística adotada.

2. Abordagem e Métodos

O algoritmo de Dijkstra, segundo [Cormen et al. 2009], resolve o problema de caminhos mais curtos de origem única em um grafo orientado ponderado $G = (V, E)$ para o caso no qual todos os pesos de arestas são não negativos.

Para a implementação do algoritmo de Dijkstra em um mapa de grades, conforme demonstra a Figura 1, foi necessário acrescentar uma condição de parada (linha 14), onde o algoritmo, a cada vértice examinado, verifica se chegou ao nó de destino.

```

1 Program Dijkstra(G,w,s,a) ;
2 Begin
3     d[s] = 0;
4     For cada v ∈ V[G]-{s} do
5         d[v] = ∞;
6         π[v] = NIL;
7     Q = V[G];
8     While Q ≠ ∅ do
9         u = ExtraiMin(Q);
10        For cada v ∈ Adj[u] do
11            if d[v] < d[u]+w(u,v) then
12                d[v] = d[u]+w(u,v);
13                π[v] = u;
14            if v == a then
15                return constroiCam(π[], v);
16        return failure;
17 End.
```

Figura 1. Algoritmo de Dijkstra com Condição de Parada.

Além disso, caso o nó verificado seja o nó de destino, então o algoritmo chama uma função de construção do caminho mínimo (linha 15) entre os dois vértices. É possível observar que esta função, conforme demonstra a Figura 2, pode levar o tempo de $O(V)$, considerando sempre o pior caso.

```

1 Function constroiCam(π[],v) ;
2 Begin
3     caminhoFinal = v;
4     While v ∈ π[value] do
5         v = π[v];
6     caminhoFinal = caminhoFinal U {v};
7     return caminhoFinal;
8 End.
```

Figura 2. Algoritmo de Construção do Caminho Mínimo.

Levando em consideração que, tradicionalmente, o algoritmo de Dijkstra, segundo [Cormen et al. 2009], possui o custo de tempo $O(E \log V)$, com a inserção da função supracitada, o Dijkstra recebeu $O(\max(V + E \log V))$, mas permanecendo em $O(E \log V)$.

O algoritmo A* (A-Estrela), segundo [Russell and Norvig 2013], é a forma de solução mais amplamente conhecida da busca de melhor escolha, ela avalia os nós através da combinação de $g(n)$, o custo para alcançar o nó visitado, e $h(n)$, o custo para ir do nó visitado ao nó de destino.

Segundo [Djojo and Karyono 2013] o algoritmo A-Estrela realiza a busca de caminhos mínimos utilizando funções heurísticas, ou seja, a seleção dos nós é baseado na

distância a partir do nó de início mais a distância aproximada até ao destino. Essa estimativa de aproximação pode ser representada pela função $f(n) = g(n) + h(n)$.

Segundo [Eraghi et al. 2014] cada célula do mapa de grades tem uma coordenada que é mostrada em (X,Y). Essa coordenada é utilizada por funções heurísticas que realizam a soma das distâncias ortogonais entre dois pontos, sendo X a posição do nó na coluna e Y a posição do nó na linha.

Para a implementação do algoritmo A-Estrela é necessário utilizar uma lista aberta (para os nós não examinados) e uma lista fechada (para os nós já examinados), além da função heurística: $f(n) = g(n) + h(n)$ existente para estimar a distância até o destino, conforme demonstra a Figura 3.

```

1 Program A-Estrela(G,w,inicio,alvo);
2 Begin
3     ls_fechada = 0;
4     ls_aberta = inicio;
5     g[inicio] = 0;
6     f[inicio] = g[inicio] + h(inicio, alvo);
7     While ls_aberta ≠ 0 do
8         u = ExtraiMin(ls_aberta);
9         ls_fechada = u;
10        For cada v ∈ Adj[u] do
11            if v ∈ ls_fechada then
12                else
13                    geraG = g[u]+w(u,v);
14                    if v ∉ ls_aberta or geraG < g[v] then
15                        π[v] = u;
16                        g[v] = geraG;
17                        f[v] = g[v] + h(v, alvo);
18                        if v == goal then
19                            return constroiCam(π[],v);
20                        if v ∉ ls_aberta then
21                            ls_aberta = v;
22        return failure;
23 End.
```

Figura 3. Algoritmo A-Estrela.

O algoritmo A-Estrela também faz uso da função de construção do caminho mínimo (Figura 2). Além disso, o A-Estrela deve utilizar uma heurística admissível. Tendo em vista a utilização de mapas de grade, a heurística adotada foi a Distância de Manhattan, com movimentos na horizontal e na vertical, conforme demonstra a Figura 4.

```

1 Function h(v,alvo) ;
2 Begin
3     int D = 1;
4     int dx = abs(v[x]-alvo[x]);
5     int dy = abs(v[y]-alvo[y]);
6     return (D.(dx+dy));
7 End.
```

Figura 4. Algoritmo da Heurística Distância de Manhattan.

Apesar da função heurística ser $O(1)$, o algoritmo A-Estrela também possui a função de construção do caminho mínimo $O(V)$, que adicionado a sua complexidade de tempo ficaria $O(max(V + VlogV))$, não aumentando a complexidade inicial $O(VlogV)$.

3. Resultados e Conclusões

Os experimentos evidenciaram o comportamento dos algoritmos de Dijkstra e A-Estrela durante a expansão dos nós. Os resultados em um mapa de grade (7x7), onde o nó de partida (verde) e o nó de destino (vermelho) estão nos extremos, são mostrados na Figura 5.

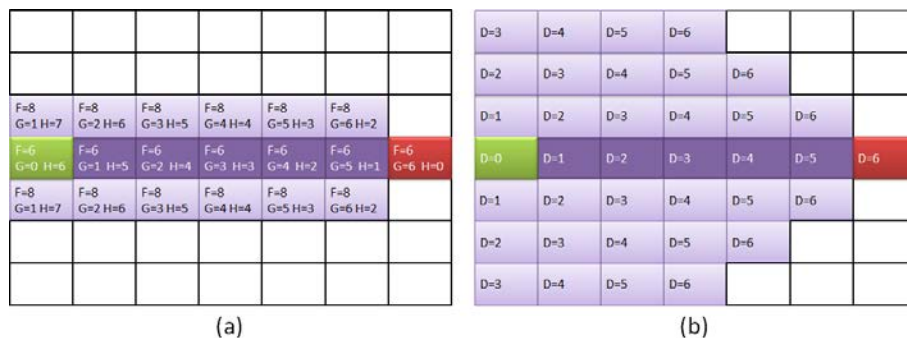


Figura 5. Comportamento dos Algoritmos A-Estrela (a) e Dijkstra (b).

A quantidade de operações realizadas em cada algoritmo diverge significativamente de acordo com o tamanho dos mapas de grade, conforme demonstra a Figura 6.

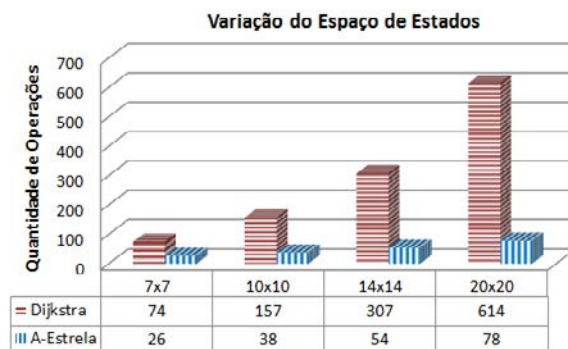


Figura 6. Gráfico Comparativo das Operações em Distintos Mapas de Grade.

Dessa forma, apesar dos algoritmos estudados possuírem, no pior caso, a mesma complexidade de tempo, é possível constatar que o A-Estrela, utilizando um heurística admissível, resolve o problema caminho mínimo de forma mais eficiente que o Dijkstra.

Referências

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition.

Djojo, M. and Karyono, K. (2013). Computational load analysis of dijkstra, a*, and floyd-warshall algorithms in mesh network. In *Robotics, Biomimetics, and Intelligent Computational Systems (ROBIONETICS), 2013 IEEE International Conference on*, pages 104–108.

Eraghi, N., Lopez-Colino, F., de Castro, A., and Garrido, J. (2014). Path length comparison in grid maps of planning algorithms: Hctnav, a*; and dijkstra. In *Design of Circuits and Integrated Circuits (DCIS), 2014 Conference on*, pages 1–6.

Russell, S. J. and Norvig, P. (2013). *Artificial Intelligence: A Modern Approach*. Pearson Education, 2 edition.

Um modelo RCPSP para Gestão Ágil Scrum

Osmar Leandro Dantas da Silva¹, Emanuel Dantas Filho², Diego Rocha Lima¹

¹ Instituto Federal do Ceará campus Crato (IFCE)
Caixa Postal 63115-500 – Crato – CE – Brasil

² Instituto Federal da Paraíba campus Monteiro (IFPB)
Caixa Postal 58500-000 – Monteiro – PB – Brasil

osmarleandros@gmail.com, emanuel.filho@ifpb.edu.br,
diego.rocha@ifce.edu.br

Abstract. *The RCPSP is part of combinatorial problems of class scheduling, which seeks to optimize limited human resources in a project. The purpose of this research is to evaluate the adoption of a mathematical model for the RCPSP in agile management of a team of software development. The analyzed scenario consists to use the model in the allocation of teams using the Scrum methodology. As result, simulations were realized on the distribution activities of Sprints resources, and the proposed model indicated allocations with a gain of up to 16.60 % in the project duration.*

Resumo. *O RCPSP faz parte da classe de problemas combinatórios de escalonamento, onde busca-se otimizar os recursos humanos limitados em um projeto. O propósito dessa pesquisa é avaliar a adoção de um modelo matemático para o RCPSP na gestão ágil de uma equipe de desenvolvimento de software. O cenário analisado consiste em utilizar o modelo na alocação de equipes que utilizam a metodologia Scrum. Como resultados, foram realizadas simulações na distribuição de atividades dos recursos em Sprints, e o modelo proposto indicou alocações com ganho de até 16,60 % no tempo de duração de projetos.*

1. Introdução

Para Versionone (2008), o uso de metodologias ágeis tem sido crescente em equipes de desenvolvimento de software, apresentando resultados positivos em termos de prazo, custo e qualidade. Tais abordagens visam à desburocratização das atividades (SCHWABER, 2013) e empregam princípios como ciclos iterativos e entregas frequentes de software (BECK, 2001).

No universo da Pesquisa Operacional (PO), o problema de alocação de recursos humanos em projetos (RCPSP) é conhecido como uma variação do problema do escalonamento. Para resolver problemas que envolvem PO, pode-se utilizar algoritmos exatos ou heurísticos.

O objetivo deste trabalho consiste em avaliar o uso de um modelo matemático para realizar a otimização de alocações na gestão de um Scrum Team. Para avaliar os resultados, foi comparado as alocações simuladas pelo modelo com alocações reais em equipes de projetos de uma fábrica de software.

2. Trabalhos Relacionados

Em Kolisch (2006), o problema RCPSP é tratado para a minimização de dias do projeto (makespan), e maximização do valor presente líquido (VPL). Em Rodrigues (2010) temos uma pesquisa para otimização de tempo e custos na alocação de recursos em projetos usando o algoritmo MBA (Minimum Bounding Algorithm).

Na literatura, existem estudos de diversas meta-heurísticas e heurísticas que são usadas para resolver o problema de alocação de recursos em projetos. Pode-se citar as alocações baseadas na Busca Tabu, Simulated Annealing e Algoritmos Genéticos, respectivamente dos trabalhos de Baar (1999), Bouleimen (2003) e Ranjbar (2008).

Em projetos de desenvolvimento de software, existem restrições específicas que precisam ser tratadas no problema RCPSP. Algumas dessas restrições, como a alocação de recursos fracionados e/ou recursos heterogêneos, são mencionadas no trabalho de Filho e Gomes (2014).

3. Metodologia

Inicialmente foi realizado um estudo sistemático sobre Pesquisa Operacional, Programação Linear Inteira, e trabalhos relacionados do problema RCPSP. Em seguida foi adaptado o modelo matemático proposto por Filho e Gomes (2014) com adição de restrições específicas para alocações de equipes de desenvolvimento Scrum. Por fim, para subsidiar os resultados desta pesquisa, foram instanciados cenários utilizando o modelo e comparado com as alocações reais de uma equipe.

Com relação aos objetivos e procedimentos, a pesquisa é explicativa e experimental. O objeto de estudo corresponde à otimização de tempo na alocação de recursos humanos em equipes de desenvolvimento que utilizam a metodologia Scrum. Para avaliar os resultados do modelo matemático proposto, foram instanciados no software LINGO algumas Sprints presentes no backlog de uma fábrica de software.

4. Resultados

O modelo RCPSP desta pesquisa é uma adaptação ao modelo proposto por Filho e Gomes (2014) para satisfazer o cenário de equipes Scrum. Foram adicionados ao modelo as restrições dispostas nas equações 4.1 e 4.2.

$$F(a) = \left(\sum_{r,a \in R_A} M_{ra} + \sum_{a,b \in C_{critico}} S_{ab} \right) - 1, \quad \forall a \in A \quad (eq\ 4.1)$$

$$S(a,p) = \left(\sum_{p \in C_{normal}} F_p \right) + 1, \quad \forall (a,p) \in C_{critico} \quad (eq\ 4.2)$$

Na equação 4.1, para cada atividade (a), seu instante de término (F) deverá ser a soma de seus instantes de execução (M) com o momento de seu início (S), subtraindo-se o instante atual. Essa restrição é necessária para que o modelo refaça o caminho crítico das atividades do projeto.

Seguindo essa abordagem, a equação 4.2 mostra que para cada atividade (a), seu instante de início (S) é o próximo instante de tempo após o término (F) da sua atividade precedente (p). Portanto, deve-se definir previamente o início das atividades e suas precedências, que são representados pelo conjunto S(a,p).

Para avaliar os resultados do modelo, foi realizada uma análise comparativa da alocação real de uma equipe Scrum e a alocação proposta na simulação utilizando o modelo matemático RCPSP. Para tanto, foi selecionado um projeto em execução de uma fábrica de software que utiliza a metodologia Scrum. Esse projeto serviu de subsídios na avaliação dos resultados da pesquisa.

As estimativas das atividades e a escolha dos recursos são tarefas realizadas de forma manual pela equipe e conduzida pelo Scrum Master, caracterizando uma alocação real. Este mesmo cenário de recursos e atividades foi instanciado no modelo matemático. Para as simulações de alocações foi utilizado a versão de testes do solver Lingo 15.0, gentilmente cedida pelo LINDO Systems para a pesquisa.

A primeira instância analisada corresponde a uma Sprint 01 com duração estimada de 120 horas (3 semanas). Nessa Sprint foram alocadas trinta e três atividades para os seis recursos humanos disponíveis no projeto. A alocação real foi proposta pelo Scrum Team durante uma Sprint Planning. O mesmo backlog foi instanciado no modelo RCPSP e os recursos humanos cadastrados.

Utilizando o solver Lingo, o modelo gerou uma alocação para as mesmas atividades e propôs uma alocação com apenas 103 horas de duração. Importante frisar que todas as 33 atividades foram contempladas na alocação simulada pelo modelo.

Seguindo a mesma abordagem foram instanciadas as próximas seis Sprints do projeto. O gráfico com os resultados da análise comparativa, entre a alocação real e o a alocação simulada pode ser observada na Figura 1.

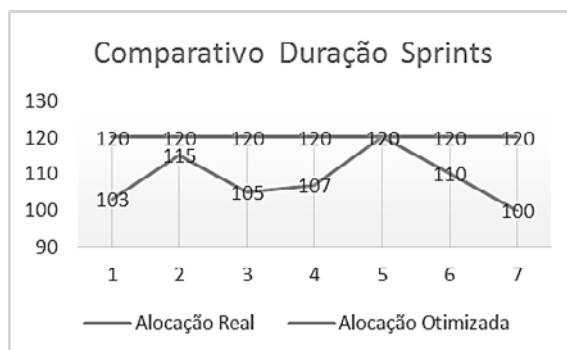


Figure 1. Comparativo de duração das Sprints.

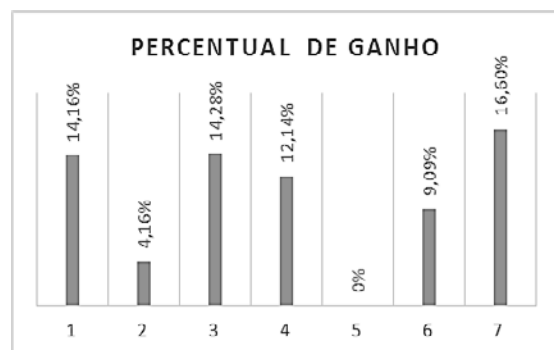


Figure 2. Percentual de ganho com o modelo.

Como visto no gráfico da Figura 1, em todas as Sprints analisadas o tempo de duração proposto pelo modelo RCPSP foi menor ou equivalente ao tempo de duração na alocação real. A Sprint 05 teve o tempo de planejamento igual ao tempo proposto na simulação do modelo matemático. Já na Sprint 07 foi obtido o melhor resultado. Nesta, o modelo propôs uma alocação com 100 horas, a maior discrepância com a alocação real dos cenários analisados.

Na Figura 2 temos a ilustração dos percentuais de ganho para cada Sprint. Esses resultados mostram a vulnerabilidade de executar alocações de forma manual, que é o processo utilizado atualmente pela equipe em questão. O ganho em cada Sprint é uma relação percentual da quantidade de horas do planejamento real, com a quantidade de horas proposta pela simulação do mesmo cenário com o modelo RCPSP.

Ao otimizar o uso dos recursos, além de eliminar situações de atividades paralelas, o modelo proposto busca balancear a carga horária dos recursos. Outra situação tratada consiste em alocar para cada atividade o recurso com mais afinidade disponível.

5. Conclusão

Como observado nos resultados, pode-se concluir que um modelo para automatizar e otimizar a alocação de equipes na gestão de projetos ágeis é de suma importância para área de TI. Vale observar que o problema investigado trata-se de um problema NP-hard, e que a alocação realizada de forma manual normalmente apresenta altas taxas de erros.

O modelo RCPSP busca otimizar o uso dos recursos, normalmente alocações realizadas sem auxílio de uma ferramenta tornam o processo susceptíveis a falhas. Visto a grande quantidade de variáveis a se analisar, realizar alocações é uma tarefa custosa e passível de erros constantes na gestão de projetos que utilizam a metodologia ágil Scrum.

Os resultados obtidos nesta pesquisa foram satisfatórios e cumpriram as expectativas de apresentar alocações com diminuição de tempo na distribuição das atividades para um Scrum Team. Propõe-se como trabalho futuro avaliar a aderência do modelo em outras equipes de desenvolvimento de software que utilizem a metodologia Scrum.

6. Agradecimentos

Nossos agradecimentos ao Instituto Federal do Ceará pelo financiamento da pesquisa e para LINDO Systems que disponibilizou a licença do solver Lingo para essa pesquisa.

Referências

- Baar, T., Brucker, P., & Knust, S. (1999) “Tabu search algorithms and lower bounds for the resource-constrained project scheduling problem”, Springer US, pages 1-18.
- Beck, K. et al. (2001) “Manifesto for Agile Software Development”, <http://www.agilemanifesto.org>, Fevereiro 2015.
- Bouleimen, K., Lecocq, H. (2003) “A new efficient simulated annealing algorithm for the resource – constrained project scheduling problem and its multiple modes version”, *European Journal of Operational Research*, 149(2), 268–281.
- Filho, E. D., and Gomes, M. J. N. (2015) Modelos para Alocação de Recursos Humanos de Diferentes Perfis em Projetos de TI, In *Gestão e Projetos: GeP*, 6(1), 63-78.
- Kolisch, R., and Hartmann, S. (2006) “Experimental investigation of heuristics for resource-constrained project scheduling: An update”, *European journal of operational research*, 174(1), 23-37.
- Ranjbar, M., Kianfar, F., and Shadrokh, S. (2008) “Solving the resource availability cost problem in project scheduling by path relinking and genetic algorithm”. *Applied Mathematics and Computation*, 196(2), 879-888.
- Rodrigues, S. B., and Yamashita, D. S. (2010) An exact algorithm for minimizing resource availability costs in project scheduling. In *European Journal of Operational Research*, 206(3), 562-568.
- Schwaber, K.; Sutherland, J. (2013) “Guia do Scrum: Um guia definitivo para o Scrum: As regras do jogo”. [SI].
- Versionone. (2008) The State of Agile Development Survey Results. http://www.versionone.com/pdf/3rdAnnualStateOfAgile_FullDataReport.pdf, Janeiro 2015.

KM-Finder: Uma Ferramenta para Detecção de Motivos

Luciana Montera¹, Lucas Akayama Vilhagra¹, Tainá Raiol²

¹Faculdade de Computação – Universidade Federal de Mato Grosso do Sul (UFMS)
Caixa Postal 549. CEP 79070-900. Campo Grande - MS - Brasil

²Instituto Leônidas e Maria Deane – Fundação Oswaldo Cruz
Manaus – AM – Brasil

montera@facom.ufms.br, akayamalucas1@gmail.com, tainaraiol@amazonia.fiocruz.br

Abstract. *The identification of patterns within specific regions of the HPV genome may contribute to the understanding of the viral pathogenesis. In this work, a tool for searching nucleotide patterns (motifs) in one or a group of DNA sequences was developed, which continues to be a computational challenge.*

Resumo. *A identificação de padrões em regiões específicas nos genomas de HPV pode contribuir para o entendimento da patogênese viral. Neste trabalho, foi desenvolvida uma ferramenta para busca por padrões de nucleotídeos (motivos) em uma ou grupos de sequências de DNA, tarefa que continua a ser um desafio computacional.*

1. Introdução

Os papilomavírus humanos (HPV), fatores etiológicos do câncer do colo do útero, são classificados em cutâneos ou mucosotrópicos. O genoma viral codifica proteínas estruturais necessárias à replicação e proteínas codificadas por oncogenes, que estão diretamente associadas com o desenvolvimento de câncer. O genoma viral possui cerca de 7.900 pares de bases (pb), sendo dividido em 8 genes codificadores de proteínas (L1, L2, E1, E2, E4, E5, E6 e E7) e duas regiões não codificadoras: NCR (*Noncoding Region*) e LCR, que se localiza entre L1 e E6, e possui 883 pb. A expressão dos oncogenes virais é controlada pela LCR (*Long Control Region*), que apresenta diversos sítios de ligação para fatores transcricionais celulares e virais [Bernard 2013]. Acredita-se que o *enhancer* epitélio específico, localizado na LCR, contribui para o tropismo celular, propriedade importante na patogênese viral.

O objetivo deste trabalho é a identificação de padrões de nucleotídeos nas LCRs de HPV de um mesmo grupo e, possivelmente, inter-grupos que podem estar relacionados ao tropismo viral. Para isso, foi desenvolvida uma ferramenta de busca e outras duas ferramentas *online* foram utilizadas para comparação. Estão disponíveis 26 genótipos de genomas de HPV cutâneo e 13 de HPV mucosotrópico.

2. Fundamentação Teórica

Motivos são pequenas sequências que se repetem ao longo de uma molécula de DNA (ou entre moléculas distintas de DNA) as quais presume-se que tenham alguma função biológica [D'haeseleer 2006]. [Rajasekaran 2005] define três versões para o problema da busca por motivos em sequências biológicas:

Definição 1 *Planted(l,d)-Motif Search Problem* - Dadas n sequências de mesmo tamanho e dois inteiros l e d , o problema consiste em encontrar o motivo de maior score de tamanho l presente em todas as sequências n . Variações do motivo também são retornadas desde que difiram do motivo por uma distância de Hamming [Hamming 1950] não superior a d .

Definição 2 *Edited Motif Search Problem* - Dadas n sequências e três inteiros l , d e q , são buscados todos os padrões das sequências de entrada de tamanho l que ocorram em pelo menos q das n sequências. Um padrão U é considerado uma ocorrência de outro padrão V se a distância de edição [Levenshtein 1966] entre eles é no máximo d .

Definição 3 *Simple Motif Search Problem* - Dadas n sequências e um inteiro l , padrões definidos como uma string de símbolos juntamente com o caracter "?" (curinga) são buscados nas sequências. Um padrão não começa ou termina com "?". O objetivo é identificar todos os padrões de tamanho máximo l , com número de curingas podendo variar de 0 a $\lfloor \frac{l}{2} \rfloor$, bem como o número de ocorrências destes padrões.

Na literatura existem diversos algoritmos de busca por motivos, dentre eles [Pevzner and Sze 2000], [Buhler and Tompa 2001], [Adebiyi and Kaufmann 2002] e [Dinh et al. 2012].

3. Ferramentas e Resultados

Ferramentas *online*, tais como SMILE [Marsan and Sagot 2000] e MEME [Bailey and Elkan 1994] realizam a busca por motivos em sequências biológicas. Neste trabalho, SMILE¹ e MEME², bem como uma ferramenta própria³ denominada KM-Finder (*K-mer and Motif Finder*) foram utilizadas para buscar por padrões nas LCRs de HPV.

SMILE implementa um algoritmo combinatório baseado na construção de uma árvore de sufixo generalizada enquanto MEME [Bailey et al. 2009] é uma suíte de ferramentas que além de permitir realizar a busca por motivos permite a comparação de novos motivos encontrados com motivos já conhecidos e predição de função biológica.

A implementação proposta neste trabalho, KM-Finder, inicialmente determina grupos de *substrings* de tamanho k (*k-mers*), comparando cada *k-mer* e adicionando ao grupo os que possuem distância de Hamming de no máximo m com relação ao *k-mer* representante. Os grupos que não tiverem ocorrências em no mínimo q e no máximo Q sequências do conjunto em estudo, são descartados. Para cada um destes grupos, são construídos *motif-profiles*, que são matrizes de frequências de cada nucleotídeo presente nos *k-mers* deste grupo. A concatenação dos nucleotídeos mais frequentes resulta em um candidato a motivo [Jones and Pevzner 2004]. A qualidade desse candidato é calculada fazendo-se a média aritmética dos valores dos nucleotídeos mais frequentes e, somente os motivos que tiverem qualidade superior a p ($0 < p \leq 1$) serão mostrados na saída. Para o *motif-profile* mostrado na Figura 1(b) a qualidade é 0.81. A Figura 1 mostra como o motivo GCGACCGA foi determinado dado a busca no conjunto de LCR de mucosa com $k = 8$, $q = 13$, $Q = 13$, $m = 2$ e $p = 0.8$. Na Figura 1(a) são mostrados 5 dos

¹disponível em mobyli.pasteur.fr/cgi-bin/portal.py#forms::smile

²disponível em meme-suite.org/tools/meme

³disponível em pintado.facom.ufms.br/hpv

23 elementos do grupo de k -mers utilizados para construir o *motif-profile* apresentado na Figura 1(b).

GCGACGA	Posição 1	Posição 2	Posição 3	Posição 4	Posição 5	Posição 6	Posição 7	Posição 8
GGGACCGA	0.04	0	0.39	0.87	0.04	0	0	0.43
GCAACCGT	0.04	0.83	0.04	0.13	0.87	1	0	0.26
GCGACCGC	0.91	0.17	0.57	0	0.09	0	1	0.13
GCAACCGA	0	0	0	0	0	0	0	0.17
GCAACCGG								

(a) Grupo de k -mers

(b) *motif-profile*

Figura 1. Grupo de k -mers e respectiva *motif-profile* que resulta no motivo GC-GACCGA.

A fim de comparar as três ferramentas citadas, as buscas foram realizadas com parâmetros iguais ou similares. Os tamanhos de motivos buscados variaram entre 6 e 19. Para o grupo de HPV cutâneo na busca por motivos de tamanho 6, KM-Finder e SMILE encontraram a sequência AATAAA, enquanto MEME, além de encontrar essa sequência, encontrou também TGCCAA. Buscando por motivos de tamanho 19, KM-Finder encontrou a sequência AGCGACCGATTTCGGTACC, enquanto MEME encontrou GATTGTTGCCAACAATCAT e SMILE retornou um resultado inesperado com 74.325 motivos. Para o grupo de HPV mucosotrópico, buscando motivos de tamanho 12, SMILE, KM-Finder e MEME encontraram a sequência ACCGATTTCGGT. SMILE encontrou ainda outras 8 sequências e MEME encontrou ACCGAAAACGGT e AACCGAAATCGG.

As divergências nos resultados vem do fato das ferramentas implementarem algoritmos diferentes, sendo assim, para uma mesma entrada, as saídas podem ser diferentes. Pode-se notar que quanto menos conservados são os motivos, mais os resultados apresentados pelas ferramentas divergem.

O tempo de resposta do MEME foi o melhor entre as ferramentas testadas e a apresentação dos resultados, por meio da construção de logos e outros gráficos, facilita a análise. SMILE responde rapidamente buscando motivos com tamanho até 16, porém, para tamanhos superiores, demora várias horas. Além da diferença no algoritmo implementado, um importante diferencial do KM-Finder é a possibilidade de busca simultânea em dois grupos distintos, pois, no caso do HPV, deseja-se observar possíveis semelhanças e diferenças entre cada grupo.

A Figura 2 apresenta o Diagrama de Venn resultante da busca por motivos para $k = 9$, $m = 2$ e $p = 0.8$. Foram encontrados 603 motivos exclusivos de mucosa e 17 exclusivos de cutâneo enquanto que 9 motivos estão presentes em ambos os grupos. Padrões que caracterizem cada grupo de vírus podem ser utilizados como método de diagnóstico para identificar a origem de determinado tumor, além de contribuir para o conhecimento sobre a patologia da doença.

4. Conclusão e Trabalhos Futuros

O prognóstico de uma neoplasia depende do tecido afetado que, no caso de tumores causados pela infecção por HPV, pode ter origem no epitélio ou mucosa. Neste trabalho, foi

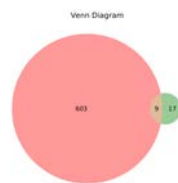


Figura 2. Diagrama de Venn resultante da busca por motivos em sequências de HPV de mucosa (rosa) e cutâneo (verde).

possível identificar padrões únicos que diferem entre os grupos de HPV, indicando potenciais marcadores para diagnóstico. Este estudo, ainda, gerou dados preliminares para a futura identificação de potenciais motivos envolvidos no tropismo viral. Para tal, existem ferramentas, como TFBind⁴, cujas análises podem ser incorporadas na ferramenta proposta. É possível ainda melhorar o algoritmo pela implementação de métodos mais sofisticados para mensurar a qualidade dos motivos a fim de tentar reduzir ainda mais falsos positivos e implementar outras representações gráficas para auxiliar no entendimento dos resultados.

Referências

- Adebiyi, E. F. and Kaufmann, M. (2002). Extracting common motifs under the levenshtein measure: theory and experimentation. pages 140–156. Proc. Workshop on Algorithms for Bioinformatics (WABI).
- Bailey, T. L., Boden, M., Buske, F., Frith, M., vGrant, C., and et al. (2009). Meme suite: tools for motif discovery and searching. pages 202–208. Nucleic Acids Research.
- Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. pages 28–36. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology.
- Bernard, H.-U. (2013). Regulatory elements in the viral genome. *Virology*, 445(1):197–204.
- Buhler, J. and Tompa, M. (2001). Finding motifs using random projections. pages 269–278. Proc. Fifth Annual International Conference on Computational Molecular Biology (RECOMB).
- D’haeseleer, P. (2006). What are dna sequence motifs? pages 423–425. Nature Biotechnology.
- Dinh, H., Rajasekaran, S., and Davila, J. (2012). qpms7: A fast algorithm for finding (l,d)-motifs in dna and protein sequences. In *PLoS ONE*.
- Hamming, R. W. (1950). Error detecting and error correcting codes. The Bell System Technical Journal.
- Jones, N. C. and Pevzner, P. A. (2004). *An Introduction to Bioinformatics Algorithms*. The MIT Press.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics-Doklady.
- Marsan, L. and Sagot, M.-F. (2000). Algorithms for extracting structured motifs using a suffix tree with an application to promoter and regulatory site consensus identification. pages 345–362. Journal of Computational Biology.
- Pevzner, P. and Sze, S. H. (2000). Combinatorial approaches to finding subtle signals in dna sequences. pages 269–278. Proc. Eighth International Conference on Intelligent Systems for Molecular Biology.
- Rajasekaran, S. (2005). *Algorithms for Motif Search in Handbook of Computational Molecular Biology*. Chapman and Hall/CRC. chapter 37.

⁴disponível em tfbind.hgc.jp

Estruturas de Dados Probabilísticas para Representação de Conjuntos

Juan P. A. Lopes¹, Paulo E. D. Pinto^{1*}, Fabiano de S. Oliveira^{1*}

¹IME/DICC, Universidade do Estado do Rio de Janeiro (UERJ)
Av. São Francisco Xavier, 524, Maracanã – 20550-013 – Rio de Janeiro – RJ

me@juanlopes.net, pauloedp@ime.uerj.br, fabiano.oliveira@ime.uerj.br

Abstract. *This paper describes the most important aspects of the probabilistic data structures Bloom Filter, Count-Min Sketch, MinHash, and HyperLogLog, that allow, through the use of hash functions, to estimate some properties of sets and multisets without needing to keep them in memory. All data structures described in this paper are constructed using easily parallelizable online algorithms, making them especially important in the big data scenario where an approximate answer is acceptable.*

Resumo. *Este artigo descreve os aspectos mais importantes das estruturas de dados probabilísticas Filtro de Bloom, Count-Min Sketch, MinHash e HyperLogLog, que permitem, através do uso de funções de hash, estimar algumas propriedades sobre conjuntos e multiconjuntos sem a necessidade de mantê-los em memória. Todas as estruturas apresentadas neste artigo são construídas através de algoritmos online facilmente paralelizáveis, o que as torna especialmente importantes no cenário de grandes volumes de dados onde uma resposta aproximada é aceitável.*

1. Introdução

A emergência de sistemas que lidam com grandes volumes de dados exigiu que fossem elaboradas novas técnicas para processá-los. Muitas vezes, armazená-los para posterior processamento pode ser inviável. Este cenário levou à popularização de uma nova classe de aplicações que operam sobre fluxos de dados [Babcock et al. 2002]. Dado o volume e a natureza virtualmente infinita desses fluxos, é desejável que os algoritmos e estruturas de dados que os processem sejam capazes de computar resultados de forma contínua e em apenas uma passagem, utilizando o mínimo possível de recursos. Para muitos problemas, não são conhecidos algoritmos determinísticos que satisfaçam estas características. As técnicas probabilísticas podem endereçar estas dificuldades ao custo de fornecerem uma resposta correta com certa probabilidade [Gibbons and Matias 1999].

Nas próximas seções serão abordadas quatro estruturas de dados probabilísticas que, utilizando o *hash* dos elementos de conjuntos e multiconjuntos, são capazes de representar operações sobre eles.

*Projeto parcialmente financiado por FAPERJ.

2. Filtro de Bloom

Um *filtro de Bloom* representa um conjunto e permite verificar a pertinência de elementos com possibilidade de falsos positivos [Bloom 1970]. É uma representação bastante compacta: são necessários menos de 10 bits por elemento para uma probabilidade de 1% de falsos positivos [Bonomi et al. 2006]. A estrutura representa um conjunto S de cardinalidade n utilizando um vetor B de m bits e k funções de hash $h_i : S \rightarrow [1..m]$, com $i \in [1..k]$, cujas funções de dispersão possuem distribuição uniforme. Inicialmente, $B[i] = 0$ para todo $i \in [1..m]$. Para inserir um elemento x , é preciso atribuir $B[h_i(x)] \leftarrow 1$ para todo $i \in [1..k]$. Para verificar a pertinência de um elemento x , a estrutura responde afirmativamente precisamente quando $B[h_i(x)] \leftarrow 1$ para todo $i \in [1..k]$. Naturalmente, a condição é necessária mas não suficiente para garantir a pertinência, e portanto falsos positivos podem existir. Quanto mais bits forem usados no filtro de Bloom, menor a probabilidade de um falso positivo.

Em [Bonomi et al. 2006], mostra-se que a probabilidade de falso positivo é minimizada quando $k = q \ln 2$, onde $q = m/n$ (o número de bits alocados por elemento), e dada por

$$\Pr[\text{FALSOPOSITIVO}] \approx ((1/2)^{\ln 2})^q.$$

Note que, com apenas 10 bits por elemento, é possível obter uma probabilidade de falsos positivos de aproximadamente 0.8%.

3. Count-Min Sketch

Count-Min Sketch [Cormode and Muthukrishnan 2005] permite estimar os elementos de um vetor A sem guardá-los todos em memória. O vetor $A[1..n]$ é definido incrementalmente através de operações de atualização na forma de pares ordenados (i, c) , representando um acréscimo de c unidades na i -ésima posição do vetor (i.e., $A[i] \leftarrow A[i] + c$). Além disso, permite estimar o produto escalar de dois vetores representados por tal estrutura. A estrutura é composta por uma matriz $M[1..k, 1..m]$, atualizada de forma análoga a um filtro de Bloom, porém cada função de hash mapeia uma linha distinta. A atualização de um par ordenado (i, c) se dá através de $M[j, h_j(i)] \leftarrow M[j, h_j(i)] + c$ para todo $j \in [1..k]$.

A estimativa $\widehat{A}[i]$ é definida por $\min_{j=1}^k M[j, h_j(i)]$. É possível mostrar que para uma matriz com $m = \lceil e/\epsilon \rceil$ e $k = \lceil \ln(1/\delta) \rceil$, $A[i] \leq \widehat{A}[i] \leq A[i] + \epsilon \|A\|_1$ com probabilidade $1 - \delta$, onde $\|A\|_1 = \sum_{i=1}^n A[i]$. Também é possível estimar o produto escalar $A \cdot B$ de dois vetores, estimado por $\widehat{A \cdot B} = \min_{j=1}^k (\sum_{r=1}^m M_A[j, r] \cdot M_B[j, r])$. Mostra-se que $A \cdot B \leq \widehat{A \cdot B} \leq A \cdot B + \epsilon \|A\|_1 \|B\|_1$, com probabilidade $1 - \delta$.

4. MinHash

MinHash [Broder 1997] permite estimar a semelhança entre conjuntos através da aproximação do coeficiente $J(A, B)$ de similaridade de *Jaccard*, definido para dois conjuntos A e B , como [Real and Vargas 1996]

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Dada uma função de *hash* h , definimos $h_{\min}(A)$ como o menor valor de h aplicado aos elementos de A , isto é, $h_{\min}(A) = \min_{x \in A} h(x)$. Definimos a assinatura *MinHash* de A associado às funções de *hash* h_1, \dots, h_k como sendo a tupla $(h_{1,\min}(A), \dots, h_{k,\min}(A))$. É possível mostrar que $\Pr[h_{\min}(A) = h_{\min}(B)] = J(A, B)$. Assim, a comparação de cada posição nas assinaturas *MinHash* dos conjuntos A e B (que resulta em 1 se as respectivas posições são iguais e 0 caso contrário) resulta em k variáveis de Bernoulli que podem ser utilizadas como um estimador de $J(A, B)$. O erro do estimador pode ser calculado aplicando o limite de Chernoff onde, para haver erro menor que θ , com confiança de $1 - \delta$, é preciso escolher k tal que

$$k \geq \frac{2 + \theta}{\theta^2} \times \ln(2/\delta).$$

Como ilustração, para $k = 512$, o erro deve ser menor que 12.37%, com 95% de confiança.

5. HyperLogLog

HyperLogLog [Flajolet and Martin 1985] permite estimar o número de elementos distintos em um fluxo de dados, utilizando memória sublinear. Mais especificamente, é possível estimar a cardinalidade de elementos distintos em um conjunto com vários bilhões de elementos, com 2% de erro padrão, utilizando menos de 1.5KB de memória. A estrutura baseia-se na observação do padrão de bits do *hash* dos elementos do conjunto. Note que a probabilidade do *hash* iniciar com um certo número de bits zero é dada por

$$\begin{aligned} \Pr[h(x) = 1\dots] &= 2^{-1} \\ \Pr[h(x) = 01\dots] &= 2^{-2} \\ \Pr[h(x) = 001\dots] &= 2^{-3} \\ &\vdots \\ \Pr[h(x) = 0^{p-1}1\dots] &= 2^{-p} \end{aligned}$$

O algoritmo consiste em particionar o fluxo em m subfluxos disjuntos e, para cada um, observar o maior prefixo $0^{p-1}1$, indicativo de que a cardinalidade naquele fluxo é, com alta probabilidade, da ordem de 2^p . Quanto maior o valor de m , mais precisa se torna a estimativa. De fato, em [Flajolet and Martin 1985], mostra-se que o erro relativo padrão do estimador é igual a $1.04/\sqrt{m}$. Para $m = 2048$ (aproximadamente 1.3KB, usando 5 bits para registrar o maior prefixo de cada fluxo), o erro padrão esperado é de 2.3%.

6. Análises Práticas e Conclusão

Para verificar os limites teóricos previstos, implementamos e testamos as estruturas apresentadas utilizando todas as obras de Shakespeare como conjunto de dados. Para cada estrutura foram utilizados parâmetros compatíveis com as operações. A Figura 1 mostra os resultados encontrados em contraste com o erro previsto pela teoria.

Este resultado ajuda a mostrar a utilidade destas estruturas. Há grande valor teórico e prático em poder estimar propriedades de conjuntos sem mantê-los em memória. Por exemplo, filtros de Bloom são usados desde a década de 70 para economizar operações de E/S; *HyperLogLog* tem aplicação direta na contagem de eventos únicos em aplicações web (usuários únicos, geolocalização, etc.); existem aplicações para *MinHash*

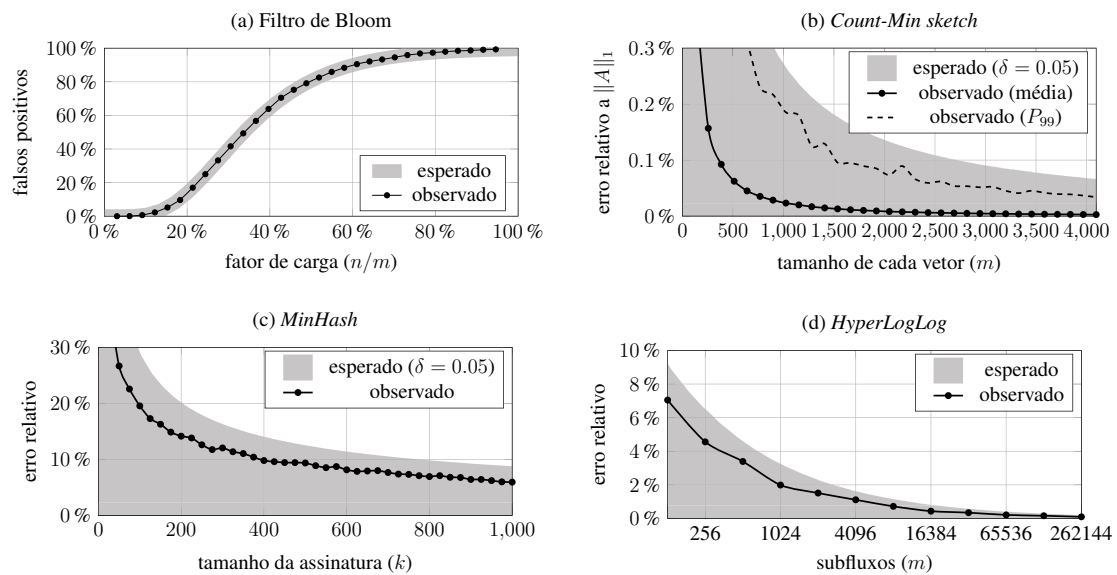


Figura 1. Erro observado ao variar parâmetros das estruturas

e *Count-Min sketch* para computar propriedades grafos muito grandes, que não caberiam em memória para serem utilizados com algoritmos determinísticos. Embora estes algoritmos sejam bastante recentes, percebe-se que há grande potencial para desenvolvimento de novas técnicas que permitam estimar a resposta para problemas clássicos com uso mais eficiente de recursos.

Referências

- Babcock, B., Babu, S., Datar, M., Motwani, R., and Widom, J. (2002). Models and issues in data stream systems. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 1–16. ACM.
- Bloom, B. H. (1970). Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426.
- Bonomi, F., Mitzenmacher, M., Panigrahy, R., Singh, S., and Varghese, G. (2006). An improved construction for counting bloom filters. In *Algorithms-ESA 2006*, pages 684–695. Springer.
- Broder, A. Z. (1997). On the resemblance and containment of documents. In *Compression and Complexity of Sequences 1997. Proceedings*, pages 21–29. IEEE.
- Cormode, G. and Muthukrishnan, S. (2005). An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms*, 55(1):58–75.
- Flajolet, P. and Martin, G. N. (1985). Probabilistic counting algorithms for data base applications. *Journal of computer and system sciences*, 31(2):182–209.
- Gibbons, P. B. and Matias, Y. (1999). Synopsis data structures for massive data sets. In *Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms*, pages 909–910. Society for Industrial and Applied Mathematics.
- Real, R. and Vargas, J. M. (1996). The probabilistic basis of jaccard’s index of similarity. *Systematic biology*, pages 380–385.

Análise Empírica do Algoritmo Shellsort

Raquel M. de Souza^{1*}, Fabiano de S. Oliveira^{1†}, Paulo Eustáquio D. Pinto^{1†}

¹Universidade do Estado do Rio de Janeiro, Brasil

raquelmarcolino25@gmail.com, {fabiano.oliveira,pauloedp}@ime.uerj.br

Abstract. *This work aims at the study of the time complexity of the Shellsort algorithm from an empirical perspective. The theoretical performance of this algorithm depends on a sequence of integers used. Several classical sequences are studied in the literature, for most of which the time complexity is known. However, for some sequences, the exact time complexity is still open. Our approach is to study such classical sequences empirically to confirm the known time complexities and to gather evidences for those still unknown.*

Resumo. *O objetivo deste trabalho é estudar a complexidade de tempo do algoritmo de ordenação Shellsort de um ponto de vista empírico. O desempenho teórico deste algoritmo depende de uma sequência de inteiros usada. Diversas sequências clássicas são estudadas na literatura, para a maioria das quais a complexidade de tempo é conhecida. No entanto, para algumas, ainda não se conhece a complexidade de tempo justa. Nossa abordagem foi a de estudar tais sequências clássicas empiricamente para ratificar as complexidades de tempo conhecidas e juntar evidências sobre aquelas desconhecidas.*

1. Introdução

O algoritmo *Shellsort* é um algoritmo de ordenação por comparação proveniente de uma generalização do algoritmo de ordenação por inserção (*InsertionSort*). A ordenação por *Shellsort* constitui-se de execuções sequenciais de *InsertionSort* sobre subconjuntos tomados dos n elementos do vetor de entrada, sendo que o último subconjunto é aquele de todos os elementos (ou seja, é uma execução padrão do *InsertionSort*). Esta estratégia tem potencial de obter melhor desempenho quando (i) as ordenações anteriores à última atuam sobre subconjuntos de elementos de tamanhos controlados, de modo que o custo do pior caso para estas execuções iniciais tenham peso reduzido no custo total; e (ii) os elementos escolhidos para serem pré-ordenados são tais que evitam o pior caso quadrático do último *InsertionSort*, de modo que a diminuição do custo na última execução compense as ordenações parciais anteriores. A motivação de Shell [Shell 1959] ao introduzir esta ideia em 1959 foi a de tentar fazer com que o *InsertionSort*, um algoritmo de ordenação local (isto é, que faz uso de memória auxiliar constante, que não é o caso, por exemplo, do *MergeSort*) com complexidade de tempo de pior caso $\Theta(n^2)$, fosse tão eficiente quanto possível para um algoritmo de ordenação por comparação (como é o *MergeSort*, com tempo ótimo de $\Theta(n \log n)$).

Mais especificamente, o *Shellsort* escolhe tais subconjuntos de elementos que serão pré-ordenados através da noção de *sequência de passos*. Se $S = p_1, p_2, p_3, \dots$, com

*Projeto financiado por CAPES.

†Projeto parcialmente financiado por FAPERJ.

$p_1 = 1$, é a sequência de passos a ser usada num algoritmo de *Shellsort* e p_k é o maior passo menor que n , então o Shellsort é descrito pelo Algoritmo 1.

Entrada: $V[1..n]$ de inteiros

Resultado: V ordenado

para $j \leftarrow k$ **até** 1 **passo** -1 **faça**

para $i \leftarrow 1$ **até** p_j **faça**

 InsertionSort($V, \{i, i + p_j, i + 2p_j, \dots, i + tp_j\}$), com $t = \lfloor \frac{n-i}{p_j} \rfloor$

Algoritmo 1: Pseudo-código do Shellsort. A chamada InsertionSort(V, S) ordena os elementos de V nas posições contidas em S por *InsertionSort*.

Desde sua introdução, o *Shellsort* foi objeto de investigação por diversos autores (ver Tabela 1). Note que, para algumas sequências, a complexidade de tempo resultante foi mostrada ser $o(n^2)$, como por exemplo, aquelas estudadas em [Frank and Lazarus 1960, Hibbard 1963, Pratt 1972] de tempo $O(n^{3/2})$ e em [Incerpi and Sedgewick 1983] de tempo $O(n^{1+\sqrt{8 \cdot \ln(5/2)/\ln(n)}})$. Note também que algumas destas sequências possuem sua complexidade de pior caso exata desconhecida.

Neste trabalho, é feita uma análise empírica da complexidade de tempo de várias das sequências da Tabela 1 com objetivo inicial de confirmar aquelas complexidades já conhecidas. Com este passo, validamos o processo de aferição empírica da complexidade. Em seguida, passamos a usar o mesmo método para conjecturar as complexidades de tempo desconhecidas. Em especial, analisamos a complexidade de caso médio das diversas sequências, desconhecidas por análise quase na totalidade.

2. Medida empírica da complexidade de tempo

A ferramenta EMA (*EMpirical analysis of Algorithms*) foi desenvolvida para estimar a complexidade de tempo e espaço de algoritmos [Oliveira 2016]. Mais especificamente, a proposta do EMA é fornecer a complexidade que mais se ajusta aos resultados de experimentos planejados e executados pelo próprio EMA. Os experimentos estão sujeitos a entradas do usuário (como tempo/espaço máximo permitido de execução) que influenciam na qualidade da estimativa. A entrada do EMA consiste de: (i) um programa \mathcal{A} que será analisado; (ii) a lista $\mathcal{V} = v_1, v_2, \dots, v_r$ de variáveis das quais a complexidade de tempo/espaço de \mathcal{A} dependa (para o caso do Shellsort, $\mathcal{V} = n$); e (iii) um programa \mathcal{B} que tem por entrada uma valoração V_1, \dots, V_r das respectivas variáveis v_1, \dots, v_r e produz uma entrada \mathcal{I} para \mathcal{A} na qual $v_i = V_i$ para todo $1 \leq i \leq r$. O processamento do EMA segue as seguintes etapas: (i) (*calibração*) para todo $1 \leq i \leq r$, determina valores V_i^{\max} de modo que o tempo de execução de \mathcal{A} fique limitado a certa constante e que o espaço de memória utilizado não ultrapasse a memória física livre do sistema para toda entrada \mathcal{I} na qual $v_i \leq V_i^{\max}$; (ii) (*simulações*) define-se um subconjunto $S_i \subseteq \{1, \dots, V_i^{\max}\}$ para todo $1 \leq i \leq r$ com valores de v_i que serão analisados: para cada $(V_1, \dots, V_r) \in S_1 \times \dots \times S_r$, \mathcal{B} é usado para gerar uma entrada \mathcal{I} de \mathcal{A} na qual $v_i = V_i$ para $1 \leq i \leq r$, e, em seguida, \mathcal{A} é executado tendo \mathcal{I} por entrada, armazenando ao final o tempo de execução e o espaço máximo de memória ocupado. Portanto, forma-se uma base de dados com $\prod_{i=1}^r |S_i|$ execuções; e (iii) (*análise*) por meio de várias regressões não-lineares, estima-se uma forma funcional para a complexidade de tempo/espaço.

Para o caso de uma variável, se $T(n)$ é a função que representa o recurso (tempo ou espaço) requerido por \mathcal{A} sob entrada na qual $v_1 = n$, o EMA determinará os parâmetros que minimizam a soma dos quadrados dos resíduos da equação

$$T(n) = a_0 n^{a_1} a_2 n^{a_3} \log n^{a_4} \log^{a_5} n, \text{ com } a_0, a_2 > 0.$$

Tabela 1. Sequências de passos estudadas na literatura para o Shellsort.

Termo geral [Autor Ano]	Sequência	Pior caso
$\lfloor \frac{n}{2^k} \rfloor, k \geq 1$ [Shell 1959]	$\lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{4} \rfloor, \dots, 1$	$\Theta(n^2)$ [Frank and Lazarus 1960]
$2^k - 1, k \geq 1$ [Hibbard 1963]	1, 3, 7, 15, 31, 63, ...	$\Theta(n^{3/2})$ [Pratt 1972]
$2^p 3^q, p \in \mathbb{N} \text{ e } q \in \mathbb{N}$ [Pratt 1972]	1, 2, 3, 4, 6, 8, ...	$\Theta(n \log^2 n)$ [Pratt 1972]
$a_k = 3 \cdot a_{k-1} + 1, k \geq 2 \text{ e } a_1 = 1$ [Knuth 1973]	1, 4, 13, 40, 121, ...	$\Theta(n^{3/2})$ [Pratt 1972]
$\prod_{0 < k < r, k \neq ((r^2+r)/2)-q} a_k$, onde $r = \lfloor \sqrt{2q} + \sqrt{2q} \rfloor$ e $a_k = \min \left(m \in \mathbb{N} : m \geq \left(\frac{5}{2}\right)^{k+1}, \forall p : 0 \leq p < k \Rightarrow \text{mdc}(a_p, m) = 1 \right)$ [Incerpi and Sedgewick 1983]	1, 3, 7, 21, 48, ...	$O(n^{1+\sqrt{8 \ln(5/2)/\ln n}})$ [Incerpi and Sedgewick 1983]
$4^k + 3 \cdot 2^{k-1} + 1, k \geq 2$ [Sedgewick 1986]	1, 8, 23, 77, 281, ...	$O(n^{4/3})$ [Sedgewick 1986]
$a_k = \max \left(\lfloor \frac{5a_{k-1}}{11} \rfloor, 1 \right), k \geq 2 \text{ e } a_1 = n$ [Gonnet and Baeza-Yates 1991]	$\lfloor \frac{5n}{11} \rfloor, \lfloor \frac{5}{11} \lfloor \frac{5n}{11} \rfloor \rfloor, \dots, 1$	em aberto
$\lfloor \frac{9^k - 4^k}{5 \cdot 4^{k-1}} \rfloor, k \geq 1$ [Tokuda 1992]	1, 4, 9, 20, 46, ...	em aberto
Sequência obtida empiricamente [Ciura 2001]	1, 4, 10, 23, 57, 132, 301, 701, 1750	em aberto

3. Resultados Obtidos e Conclusão

A Tabela 2 sumariza as complexidades de tempo obtidas. Na fase das simulações, o EMA utilizou a faixa $2 \times 10^6 \leq n \leq 25 \times 10^6$, amostrando 100 pontos neste intervalo ($|S_1| = 100$). Na análise de caso médio, foram feitos 30 sorteios com distribuição uniforme de vetores para cada valor de n amostrado, tomando-se a média dos tempos obtidos. Os experimentos foram conduzidos em um computador com processador Intel Core(TM) i7 - 3770 CPU @ 3.40GHz de 8 núcleos e 16GB de memória RAM. Como conclusões, notamos que as complexidades empíricas concordam com aquelas analíticas existentes, sugerindo as complexidades empíricas como candidatas a analíticas quando estas são desconhecidas. Além disso, notamos que o *Shellsort* atinge na prática o objetivo original de ser ótimo (ou próximo de) para várias sequências, ao menos no seu caso médio.

Este trabalho foi desenvolvido como projeto de graduação, que se transformou recentemente em estudo de mestrado. Visamos caracterizar o pior caso do *Shellsort* de maneira algorítmica, de forma que as complexidades empíricas obtidas neste caso sejam

referentes ao pior caso, e não ao caso médio. O objetivo é acumular evidências para o preenchimento da lacuna de conhecimento do tempo de pior caso do *Shellsort* para algumas sequências. Os resultados parciais nesta linha de investigação parecem promissores.

Tabela 2. Complexidades analisadas.

Sequência [Autor Ano]	Existente	Obtida
Pior Caso Analítico		Empírico
Shell, 1959	$\Theta(n^2)$ [Frank and Lazarus 1960]	$\Theta(n^2)$
Pratt, 1971	$\Theta(n \lg^2 n)$ [Pratt 1972]	$\Theta(n \lg^2 n)$
Caso Médio Analítico		Empírico
Pratt, 1971	$\Theta(n \lg^2 n)$ [Pratt 1972]	$\Theta(n \lg^2 n)$
Caso Médio Empírico		Empírico
Shell, 1959	$\Theta(n^{1,226})$ [Shell 1959]	$\Theta(n \lg^3 n)$
Hibbard, 1963	$\Theta(n^{1,26})$ ou $\Theta(n \lg^2 n)$ [Knuth 1973]	$\Theta(n^{1,203})$
Knuth, 1973	$\Theta(n^{1,25})$ ou $\Theta(n \lg^2 n)$ [Weiss 1991]	$\Theta(n \lg^3 n)$
Incerpi e Sedgewick, 1985	em aberto	$\Theta(n \lg n)$
Sedgewick, 1986	$\Theta(n^{7/6})$ [Weiss 1991]	$\Theta(n \lg^2 n)$
demais autores (a partir de 1991)	em aberto	$\Theta(n \lg n)$

Referências

- Ciura, M. (2001). Best increments for the average case of shellsort. In *Fundamentals of Computation Theory*, pages 106–117. Springer.
- Frank, R. M. and Lazarus, R. B. (1960). A high-speed sorting procedure. *Communications of the ACM*, 3(1):20–22.
- Gonnet, G. H. and Baeza-Yates, R. A. (1991). Handbook of algorithms and data structures: in pascal and c.
- Hibbard, T. N. (1963). An empirical study of minimal storage sorting. *Communications of the ACM*, 6(5):206–213.
- Incerpi, J. and Sedgewick, R. (1983). Improved upper bounds on shellsort. In *Foundations of Computer Science, 1983., 24th Annual Symposium on*, pages 48–55. IEEE.
- Knuth, D. E. (1973). The art of computer programming 3: Sorting and searching.
- Oliveira, F. S. (2016). EMA - WebPage. <http://fabianooliveira.ime.uerj.br/ema>. [última vez acessado: 07 de Junho de 2016].
- Pratt, V. R. (1972). Shellsort and sorting networks. Technical report, DTIC Document.
- Sedgewick, R. (1986). A new upper bound for shellsort. *Journal of Algorithms*, 7(2):159–173.
- Shell, D. L. (1959). A high-speed sorting procedure. *Communications of the ACM*, 2(7):30–32.
- Tokuda, N. (1992). An improved Shellsort. In *12th World Computer Congress on Algorithms, Software, Architecture-Information Processing*, pages 449–457. N.-Holland.
- Weiss, M. A. (1991). Short note empirical study of the expected running time of shellsort. *The Computer Journal*, 34(1):88–91.

Failcon Brazil

Apresentação

Você não é um empreendedor de verdade até que você tenha que lidar com algum fracasso. Lidar com a instabilidade pessoal e financeira que um fracasso proporciona pode ser fatal para o seu negócio. A FailCon Brazil é a edição brasileira da conferência mundial que surgiu em São Francisco em 2009 com o objetivo de proporcionar a troca de experiências entre os empreendedores no âmbito das falhas. Não se trata de “glamourizar” o erro, mas sim de assimilar as lições, tirando o máximo de aprendizado possível. Venha você também ouvir e compartilhar os seus fracassos e surpreenda-se com o resultado ao final do evento!

Rafael Matone Chanin (PUCRS)

Flávio Steffens de Castro (Bicharia)

Comitê de Organização

Coordenação Geral

Rafael Matone Chanin (PUCRS)
Flávio Steffens de Castro (Bicharia)

Painéis / Palestras

Nem toda solução tem um problema

Rafael Matone Chanin (Failcon Brazil)

Flávio Steffens de Castro (Failcon Brazil)

Nesta palestra, os idealizadores da Failcon Brazil contam a história de duas startups que desenvolveram em meados de 2010. As histórias são recheadas de fatos surpreendentes e, principalmente, de muito aprendizado.

Cupcake: lições de uma empresa de jogos

João Vítor de Souza (Cupcake Entertainment)

Serão compartilhadas lições aprendidas na criação de uma empresa de jogos, desde como montar o time certo até achar o modelo de negócio certo.

Quando a dor nos fortalece

Eduardo Meira Peres (DBServer)

Vamos compartilhar momentos de dificuldades enfrentados ao longo dos 23 anos da DBServer e que geraram aprendizagens a partir das quais a empresa encontrou ressignificados que sustentaram seu fortalecimento e posicionamento de mercado.

Onde foi que eu (não) errei?

André Echeverria Flores (CStorm)

Nesta palestra, compartilharei minha história de vida, que é cheia de acertos e erros. Mostrarei como errar (e aprender) é fundamental para que se alcance os nossos objetivos.

Erros em série no empreendedorismo digital

Rafael Zatti (EntregueEmCasa.com)

Como fracassar em 5 startups (de canais de conteúdo à fintechs) pode deixar lições e força de vontade para fazer tudo certo da próxima vez. Mostrarei como a tecnologia pode ajudar a reduzir o tamanho dos seus erros.

Seja feliz falhando

Eduardo Arruda (Doc.Space e professor da PUCRS)

Um relato das mais bem sucedidas falhas que alguém pode cometer e como ser feliz depois delas.

1º Workshop do testbed FIBRE: usando testbeds de Internet do Futuro para aprendizado de Redes de Computadores

Apresentação

O *testbed* FIBRE (*Future Internet Brazilian environment for Experimentation*) (<https://fibre.org.br>) é uma infraestrutura de pesquisa e experimentação em redes de computadores. A infraestrutura atual é composta por uma federação de 11 *testbeds* locais, também conhecidos como “ilhas de experimentação”, localizado em diferentes organizações de ensino e pesquisa no Brasil. Ainda em 2016 está prevista sua ampliação para mais 4 instituições.

Um de seus maiores desafios é oferecer uma visão unificada de recursos computacionais geridos por diferentes organizações e sua integração com outros *testbeds* internacionais. Essa federação é facilitada através do Centro de Operações do testbed (NOC), operado pela Rede Nacional de Ensino e Pesquisa (RNP), que atua como um ponto de interconexão entre as instituições provedoras de recursos para o testbed.

A construção do testbed, realizada no âmbito de um projeto financiado pela primeira chamada coordenada Brasil-Europa no domínio das TIC, foi concluída em outubro de 2014. No entanto, a partir de 2015, as instituições participantes assumiram a operação da infraestrutura legada do FIBRE, passando a oferecer o testbed como um serviço.

Além de ser uma plataforma para experimentação, testbeds de Internet do Futuro podem ser muito úteis no ensino de Redes de Computadores, tanto em ambientes de aprendizado convencionais como no contexto de cursos online. O acesso a ambientes de experimentação de larga escala oferece aos alunos uma experiência hands-on que tipicamente não é possível oferecer com os recursos locais de uma instituição de ensino.

O objetivo do 1º Workshop do testbed FIBRE é apresentar sua infraestrutura para experimentação e encorajar professores a fazerem uso do testbed em sala de aula.

O público-alvo são professores e alunos de disciplinas de Redes de Computadores e Sistemas Distribuídos.

O Workshop foi ministrado por Tiago Salmito (RNP), Gustavo Dias (RNP), Daniel Marques (RNP).

Leandro Neumann Ciuffo
Coordenador Geral

Tiago Coelho Ferreto
Coordenador Local

JAI - Jornada de Atualização em Informática

Apresentação

A Jornada de Atualização em Informática (JAI), promovida pela Sociedade Brasileira de Computação, é um dos mais importantes eventos acadêmicos de atualização científica e tecnológica da comunidade de Computação do Brasil e compreende trabalhos de pesquisadores sêniores da nossa comunidade, oferecendo uma oportunidade única para acadêmicos e profissionais atualizarem-se em temas diversos, interagindo com líderes das mais diversas áreas de pesquisa no Brasil.

Neste ano de 2016, temos a 35ª edição da Jornada de Atualização de Informática. A JAI é tradicionalmente oferecida no escopo do Congresso da Sociedade Brasileira de Computação (CSBC). O CSBC ocorre na cidade Porto Alegre, no Estado do Rio Grande do Sul, de 4 a 7 de julho, na PUCRS.

Primeiramente, gostaríamos de agradecer imensamente aos autores que submeteram as propostas para este ano de 2016 e aos membros do Comitê Científico, os quais – autores e avaliadores – contribuíram para a seleção de um conjunto de textos que tratam de temas avançados e de relevância técnico-científica, com amplas possibilidades de contribuição para soluções de alto impacto social. Agradecemos também a atenção e o apoio da organização geral do CSBC 2016 e da Diretoria da SBC que propiciaram todas as condições para a realização desta edição da JAI.

É importante destacar que a JAI favorece a participação de pesquisadores renomados, nacional e internacionalmente, com a colaboração de talentos emergentes, propiciando um ambiente rico para a evolução e disseminação do conhecimento no âmbito do CSBC. Gostaríamos inclusive de sugerir fortemente a oferta desses cursos das JAIs em outros eventos científicos, oportunizando a jovens e profissionais nas diversas regiões do Brasil, e eventualmente na América Latina, o contato com esses excelentes autores e pesquisadores. Salientamos que, por decisão do Conselho da SBC, os textos da JAI 2016 passarão a ser de acesso aberto a partir de 6 meses de sua edição.

Nesta edição da JAI, são apresentados cinco textos, sendo dois convidados da 34ª edição. Os cinco textos abordam um amplo espectro de temas relevantes para a computação e suas aplicações e contribuem certamente para soluções complexas e multidisciplinares relacionadas aos Grandes Desafios da Computação.

Os textos cobrem desde aspectos básicos de teoria de computação, passando por aspectos de comunicação usuário computador a aspectos de integração de soluções computacionais na perspectiva de sistemas de sistemas.

O texto “Teoria da Computação: uma Introdução à Complexidade e à Lógica Computacional” apresenta uma introdução às duas subáreas fundamentais da teoria da computação: complexidade e lógica computacional. Segundo os autores, essas duas áreas contribuem significativamente para o desenvolvimento da Ciência da Computação, não somente do ponto de vista fundamental, mas também sob uma perspectiva de aplicações.

Por outro lado, os textos “Cidades Inteligentes: Tecnologias, Aplicações, Iniciativas e Desafios” e “Computação Urbana: Técnicas para o Estudo de Sociedades com Redes de Sensoriamento Participativo” apresentam um cenário decorrente do crescimento da população urbana e como os seus efeitos negativos poderiam ser mitigados pela Tecnologia da Informação e Comunicação (TIC), no sentido de tornar as cidades mais inteligentes e poder ajudar a melhorar os serviços urbanos, as questões relacionadas com a dinâmica de cidades e o comportamento social urbano, consequentemente, aprimorando a qualidade de vida de seus cidadãos.

Os outros dois textos trazem contribuições significativas que abrem horizontes de desenvolvimento de soluções e aplicações que certamente impactarão a dinâmica das relações sociais e econômicas. O texto “Agentes Conversacionais Incorporados: Desenvolvimento e Aplicações” apresenta um panorama sobre Agentes Inteligentes Conversacionais que visam a facilitar a interação com o usuário a ponto de que o interlocutor tenha a ilusão de estar conversando com outro ser humano. O texto “Interfaces Cérebro-Computador” traz uma perspectiva de indivíduos conectarem-se a dispositivos externos usando sinais cerebrais.

O conjunto de textos apresentado nesta edição evidencia a relevância de investimentos para o avanço da computação, e de TICs de uma forma geral, para a concepção e implementação de sistemas (sistemas de sistemas) computacionais que promovam a sustentabilidade, a inclusão e o acesso a uma ampla gama de cidadãos almejando uma sociedade mais justa e equilibrada.

Desejamos a todos os participantes do CSBC uma excelente Jornada de Atualização de Informática e reiteramos nossos imensos agradecimentos à comunidade. Esperamos que a compilação desses ricos textos contribua para motivar soluções elegantes e comprometidas com a evolução da riqueza social. Inovar e empreender é preciso!

Comitê de Organização

Coordenação Geral

José Carlos Maldonado (ICMC-USP)
José Viterbo Filho (UFF)
Marcio Eduardo Delamaro (ICMC-USP)

Coordenação Local

Sabrina Marczak (PUCRS)

Comitê Científico

Adenilso Simão ICMC-USP
Alessandro Garcia PUC-RIO
Bernadette Loscio UFPE
Daltro Nunes UFRGS
Fatima Nunes EACH-USP
Guilherme Horta Travassos COPPE/UFRJ
Marco Aurélio Gerosa IME-USP
Mariza Bigonha UFMG
Mirella Moro UFMG
Nelson Fonseca UNICAMP
Raimundo Barreto UFAM
Rodolfo Azevedo UNICAMP
Routo Terada IME-USP
Sandra C.P.F. Fabbri - UFSCar
Thais Vasconcelos Batista UFRN

Minicursos

JAI 1 - Cidades Inteligentes: Tecnologias, Aplicações, Iniciativas e Desafios

Ministrantes: Fabio Kon e Eduardo Zambom

Com o crescimento da população urbana, problemas de infraestrutura e de acesso limitado a recursos em diversas cidades ao redor mundo afetam negativamente a vida de bilhões de pessoas. Tornar as cidades mais inteligentes pode ajudar a melhorar os serviços urbanos aumentando a qualidade de vida de seus cidadãos. A Tecnologia da Informação e Comunicação (TIC) são meios fundamentais para esse objetivo. Uma plataforma de software pode ser usada para facilitar enormemente a criação e integração de aplicações robustas para cidades inteligentes. Entretanto, ainda existem desafios técnicos e científicos significativos que necessitam ser enfrentados antes que essas plataformas possam ser amplamente utilizadas. Este capítulo apresenta o estado da arte e o estado da prática em iniciativas e ambientes de cidades inteligentes. Para isso, analisamos onze iniciativas internacionais de cidades inteligentes e onze projetos científicos para a criação de plataformas de software para cidades inteligentes. Descrevemos as tecnologias mais utilizadas por esses projetos e seus requisitos funcionais e não-funcionais. Finalmente, apresentamos os desafios de pesquisa ainda em aberto e discutimos a nossa visão para a área no futuro.

JAI 2- Computação Urbana: Técnicas para o Estudo de Sociedades com Redes de Sensoriamento Participativo

Ministrantes: Thiago H. Silva e Antonio A. F. Loureiro

A computação urbana (urban computing) é um tema recente de pesquisa que visa obter e analisar dados urbanos de diversas fontes, como as tradicionais redes de sensores sem fio (RSSFs) e as emergentes redes de sensoriamento participativo (RSP), com o objetivo de entender e tratar questões enfrentadas pelas cidades. As RSPs são particularmente interessantes nesse caso, pois contam com a colaboração dos usuários no sensoriamento urbano, permitindo a observação das ações das pessoas em larga escala em tempo (quase) real durante longos períodos de tempo. Dados de RSPs aumentam o nosso conhecimento sobre diferentes aspectos de nossas vidas em cenários urbanos o que pode ser bastante útil no desenvolvimento de aplicações mais sofisticadas para diversos segmentos, principalmente os relacionados com o entendimento de sociedades urbanas. O objetivo deste minicurso é discutir o conceito de computação urbana e de sensoriamento urbano com redes de sensoriamento participativo. Visamos mostrar a relevância da computação urbana e motivar a construção de novas aplicações que sirvam para tratar questões relacionadas com a dinâmica de cidades e do comportamento social urbano. Além disso, este minicurso discutirá como trabalhar com RSPs, ao analisar as suas propriedades e a sua utilidade no desenvolvimento de novas aplicações na área de computação urbana.

JAI 3- Teoria da Computação: uma Introdução à Complexidade e à Lógica Computacional

Ministrantes: Celina de Figueiredo e Luis Lamb

A teoria da computação é uma área de pesquisa abstrata e matemática, e é motivada pelos desafios da prática da computação. O objetivo da teoria da computação é explorar e entender a natureza da computação, de modo a oferecer metodologias eficientes e corretas. Este curso apresenta uma introdução à duas subáreas fundamentais da teoria da computação: complexidade e lógica computacional. Estas duas áreas têm contribuído significativamente para o desenvolvimento da Ciência da Computação, não somente do ponto de vista fundamental, mas também sob uma perspectiva de aplicações.

JAI 4- Agentes Inteligentes Conversacionais: Conceitos Básicos e Desenvolvimento

Ministrantes: Patrícia Tedesco e Flavia Barros

Este curso tem por objetivo apresentar um panorama sobre Agentes Inteligentes Conversacionais, incluindo um histórico da área, conceitos básicos, desenvolvimento e aplicações. Tais agentes, também denominados de Chatterbots, são projetados para interpretar a pergunta/intervenção do usuário e oferecer respostas diretas, buscando manter a ilusão de que o interlocutor está conversando com outro ser humano. Atualmente, encontramos esses agentes nas mais diversas aplicações, tais como ensino a distância, jogos sérios, em SACs de lojas online, e mesmo em sistemas voltados a pessoas com deficiência física. Daremos maior ênfase aos Agentes Conversacionais Incorporados, que possuem representação gráfica, movimentos faciais e corporais, e possibilitam diálogos via voz, o que torna a interação mais natural e realista.

JAI 5- Interfaces Cérebro-Computador

Ministrante: Joao Luis Garcia Rosa

Interface cérebro-computador (BCI) é uma forma de comunicação que permite que indivíduos incapazes de realizar movimentos se conectem a dispositivos externos usando sinais cerebrais. BCIs não-invasivas capturam mudanças no fluxo sanguíneo ou flutuações de campos elétricos e magnéticos causados pela atividade de grandes populações de neurônios. A eletroencefalografia (EEG) mede a atividade elétrica do cérebro em diferentes locais da cabeça, geralmente usando eletrodos colocados no couro cabeludo. Com a remoção adequada de artefatos, processamento de sinal e aprendizagem de máquina, o EEG humano carrega informações suficientes sobre a intenção de planejamento e execução. Os objetivos do minicurso são mostrar como o entendimento da atividade do cérebro pode fornecer uma maneira para permitir a comunicação sem movimentos musculares.

SBCUP 2016 - VIII Simpósio Brasileiro de Computação Ubíqua e Pervasiva

Apresentação

É com grande satisfação que damos as boas vindas aos participantes do VIII Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2016). Como já é tradicional, o evento ocorrerá de forma integrada ao Congresso da Sociedade Brasileira de Computação.

A Computação Ubíqua e Pervasiva, em seus vários desdobramentos e aplicações, é considerada por muitos como o novo paradigma da Computação para o século XXI, o qual permitirá o acoplamento do mundo físico ao mundo da informação e fornecerá uma abundância de serviços e aplicações, permitindo que usuários, máquinas, dados, aplicações e objetos do espaço físico interajam uns com os outros de forma transparente. A Computação Ubíqua e Pervasiva é considerada um dos grandes desafios da pesquisa em Computação pela National Science Foundation (NSF) e está também presente no relatório Grandes Desafios da Pesquisa em Computação no Brasil 2006-2016, publicado pela Sociedade Brasileira de Computação (SBC).

O SBCUP, enquanto evento satélite do CSBC, vem se consolidando como um fórum dedicado à apresentação e discussão de resultados de pesquisa na área de Computação Ubíqua e Pervasiva em âmbito nacional. Nesta oitava edição o evento contou com 72 submissões oriundas de instituições de pesquisa localizadas em diferentes regiões do país, sendo aceitos ao todo 31 trabalhos completos, que serão apresentados em Sessões Técnicas ou Sessões de Pôsteres.

Queremos apresentar nosso agradecimento aos integrantes do Comitê de Programa e aos Revisores Convidados. Sabemos do elevado esforço que está associado à realização de avaliações com critério, seriedade e zelo.

Registramos, também, o apoio constante do Comitê Organizador do CSBC 2016 nas diversas atividades inerentes ao SBCUP 2016. Nosso sincero agradecimento pela rapidez e dedicação como foram tratadas as diferentes demandas de planejamento.

Por fim, nosso agradecimento aos autores por sua inestimável contribuição, seja pela submissão de artigos qualificados, seja pela participação apresentando e discutindo trabalhos. Desejamos a todos, alegres e produtivos dias na bela capital dos gaúchos.

Carlos André Guimarães Ferraz (UFPE)
Raquel Aparecida de Freitas Mini (PUC Minas)
Coordenação do Comitê de Programa

César Marcon (PUCRS)
Coordenação Local

Adenauer Corrêa Yamin (UCPEL/UFPEL)
Coordenação Geral

Comitê de Organização

Steering Committee

Adenauer Yamin - Universidades Católica e Federal de Pelotas
André Aquino - Universidade Federal de Alagoas
Cristiano Costa - Universidade do Vale do Rio dos Sinos
Fabio Costa - Universidade Federal de Goiás
Hyggo Almeida - Universidade Federal de Campina Grande
Iara Augustin - Universidade Federal de Santa Maria

Coordenação Geral

Adenauer Yamin - Universidades Católica e Federal de Pelotas

Coordenação Local

César Marcon - Pontifícia Universidade Católica do Rio Grande do Sul

Coordenação do Comitê de Programa

Carlos Ferraz – Universidade Federal de Pernambuco
Raquel Mini - Pontifícia Universidade Católica de Minas Gerais

Comitê de Programa

Adenauer Yamin – Universidades Católica e Federal de Pelotas
Ana Marilza Pernas - Universidade Federal de Pelotas
André Aquino - Universidade Federal de Alagoas
Angelo Perkusich - Universidade Federal de Campina Grande
Antonio Alfredo Ferreira Loureiro - Universidade Federal de Minas Gerais
Atslands Rocha - Universidade Federal do Ceará
Bruno Silvestre - Universidade Federal de Goiás
Carlos Ferraz - Universidade Federal de Pernambuco
Carlos Kamienski - Universidade Federal do ABC
Carlos Mauricio Figueiredo - Universidade do Estado do Amazonas
Claudio Geyer - Universidade Federal do Rio Grande do Sul
Cristiano Costa - Universidade do Vale do Rio dos Sinos
Cristiano Silva - Universidade Federal de São João del Rei
Daniel Fernandes Macedo - Universidade Federal de Minas Gerais
Danielo G. Gomes - Universidade Federal do Ceará
Eduardo Cerqueira - Universidade Federal do Pará

Eduardo Nakamura - Universidade Federal do Amazonas
Fabio Costa - Universidade Federal de Goiás
Flavia Delicato - Universidade Federal do Rio de Janeiro
Francisco Silva - Universidade Federal do Maranhão
Frederico Lopes - Universidade Federal do Rio Grande do Norte
Giovani Rubert Librelotto - Universidade Federal de Santa Maria
Heitor Ramos - Universidade Federal de Alagoas
Horácio Oliveira - Universidade Federal do Amazonas
Humberto Marques - Pontifícia Universidade Católica de Minas Gerais
Hyggo Almeida - Universidade Federal de Campina Grande
Iara Augustin - Universidade Federal de Santa Maria
Iwens Sene Jr - Universidade Federal de Goiás
Jó Ueyama - Universidade de São Paulo
João Carlos Lima - Universidade Federal de Santa Maria
Joel Rodrigues - Instituto de Telecomunicações, University of Beira interior
Jorge Barbosa - Universidade do Vale do Rio dos Sinos
Jorge Sá Silva - Universidade de Coimbra
Kiev Gama - Universidade Federal de Pernambuco
Leandro de Sales - Universidade Federal de Alagoas
Leandro Silva - Universidade Federal de Alagoas
Leandro Villas - Universidade Estadual de Campinas
Lisandro Zambenedetti Granville - Universidade Federal do Rio Grande do Sul
Luiz Henrique Correia - Universidade Federal de Lavras
Mario Dantas - Universidade Federal de Santa Catarina
Michele Nogueira - Universidade Federal do Paraná
Orlando Loques - Universidade Federal Fluminense
Patricia Dockhorn Costa - Universidade Federal do Espírito Santo
Patrícia Tedesco - Universidade Federal de Pernambuco
Raquel Mini - Pontifícia Universidade Católica de Minas Gerais
Ricardo Neisse - IPSC - Joint Research Center
Rodrigo Righi - Universidade do Vale do Rio dos Sinos
Roseclea Medina - Universidade Federal de Santa Maria
Rossana Andrade - Universidade Federal do Ceará
Vaninha Vieira - Universidade Federal da Bahia
Wanderley Lopes de Souza - Universidade Federal de São Carlos

Revisores Convidados

Alisson Alves - Universidade Federal de Minas Gerais
Amirton Chagas - Universidade Federal de Pernambuco
Carlos Quadros - Universidade Federal do Pará
Celso Brennand - Universidade Estadual de Campinas
Denis Rosário - Universidade Federal do Pará
Diego Oliveira Rodrigues - Centro Universitário Luterano de Palmas
Guillermo Ponce - Universidad Católica San Pablo
Igor Dos Santos - Universidade Federal do Rio de Janeiro

Ivo Calado - Instituto Federal de Alagoas
Jaidilson Jó da Silva - Universidade Federal de Campina Grande
Jesus Talavera Portocarrero - Universidade Federal do Rio de Janeiro
Jorge Fonseca - Universidade Federal de Pernambuco
José Viterbo - Universidade Federal Fluminense
Junio Lima - Instituto Federal Goiano
Kyller Gorgônio - Universidade Federal de Campina Grande
Leandro Freitas - Instituto Federal de Goiás
Marcelo Alves - Universidade Federal do Rio de Janeiro
Pedro Libório - Universidade Estadual de Campinas
Raphael Gomes - Instituto Federal de Goiás
Raquel S. Cabral - Universidade Federal de Minas Gerais
Sergio T. Carvalho - Universidade Federal de Goiás
Thiago Borges de Oliveira - Universidade Federal de Goiás
Thiago de Sales - Universidade Federal de Alagoas

Palestra

UbiComp & Interdisciplinaridade: Ubiquidade, Zen-Budismo e Computação

Prof. Jorge Luis Victória Barbosa

Programa de Pós-Graduação em Computação Aplicada (PIPCA)

Universidade do Vale do Rio dos Sinos (UNISINOS)

Homepage: <http://professor.unisinos.br/barbosa>

Lattes: <http://lattes.cnpq.br/6754464380129137>

Inicialmente, a palestra aborda dois focos de interdisciplinaridade da UbiComp. O primeiro é inerente à evolução da computação e contempla a essência do paradigma introduzido por Mark Weiser [1], o qual atualmente estimula uma abordagem holística de disciplinas emergentes, tais como a Computação em Nuvem, o BigData e a Internet das Coisas. O segundo foco consiste na aplicação da UbiComp em diferentes disciplinas [2], tais como, Saúde (u-health) [3], Acessibilidade (u-accessibility) [4], Educação (u-learning) [5], Comércio (u-commerce) [6] e Entretenimento (u-games) [7]. Nessa etapa são apresentados projetos do palestrante que estão concretizando ambos os focos de interdisciplinaridade. Logo após, a palestra evolui para a discussão de tendências de pesquisa relacionadas principalmente à exploração multi-temporal [8] do conceito de Contexto introduzido por Dey, Abowd e Salber [9], focando em estudos relacionados com Histórico de Contextos (passado) [10], Gerenciamento de Perfis (presente) [11] e Previsão de Contextos (futuro) [12]. Novamente, o palestrante apresenta seus trabalhos nesses temas, destacando publicações realizadas em periódicos internacionais qualificados. O final da palestra converge para a discussão de ubiquidade, tempo, contextos e computação, extrapolando para uma especulação de como a UbiComp em um estágio avançado impactaria a percepção espaço-temporal do usuário, criando uma condição mental semelhante à descrita em filosofias orientais como o Zen-Budismo.

Referências:

[1] M. Weiser. The Computer for the 21st Century. *Scientific American*, v.265, p.94–104, 1991.

Available: <http://dx.doi.org/10.1145/329124.329126>

[2] J. L. V. Barbosa. Ubiquitous Computing: Applications and Research Opportunities (Invited Talk). VI IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, Índia, p.1–8, 2015.

Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7435625>

[3] H. D. Vianna, J. L. V. Barbosa. A Model for Ubiquitous Care of Noncommunicable Diseases. *IEEE Journal of Biomedical and Health Informatics*, v.18, n.5, p.1597–1606, 2014.

Available: <http://dx.doi.org/10.1109/JBHI.2013.2292860>

- [4] J. E. R. Tavares, J. L. V. Barbosa, I. G. Cardoso, C. A. Costa, A. C. Yamin, R. A. Real. Hefestos: an intelligent system applied to ubiquitous accessibility. *Universal Access in the Information Society*, p.1–19, 2015.
Available: <http://dx.doi.org/10.1007/s10209-015-0423-2>
- [5] J. L. V. Barbosa, D. N. F. Barbosa, J. M. Oliveira, S. A. J. Rabello. A Decentralized Infrastructure for Ubiquitous Learning Environments. *Journal of Universal Computer Science*, v.20, n.2, p.1649–1669, 2014.
Available: <http://dx.doi.org/10.3217/jucs-020-12-1649>
- [6] J. L. V. Barbosa; C. J. Martins, L. K. Franco, D. N. F. Barbosa. TrailTrade: A model for trail-aware commerce support. *Computers in Industry*, v.80, p.43–53, 2016.
Available: <http://dx.doi.org/10.1016/j.compind.2016.04.006>
- [7] W. Segatto, E. Herzer, C. L. Mazzotti, J. R. Bittencourt, J. L. V. Barbosa. moBIO Threat: a Mobile Game based on the Integration of Wireless Technologies. *Computers in Entertainment*, v.6, n.3, p.1–14, 2008.
Available: <http://dx.doi.org/10.1145/1394021.1394032>
- [8] J. H. Rosa, J. L. V. Barbosa, M. R. Kich, L. K. Brito. A Multi-Temporal Context-aware System for Competences Management. *International Journal of Artificial Intelligence in Education*, v. 25, n.4, p.455–492, 2015.
Available: <http://dx.doi.org/10.1007/s40593-015-0047-y>
- [9] A. K. Dey, G. D. Abowd, D. Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware application. *Human-Computer Interaction*, v.16, n.2, pp. 97–166, 2001.
Available: http://dx.doi.org/10.1207/S15327051HCI16234_02
- [10] J. M. Silva, J. H. Rosa, J. L. V. Barbosa, D. N. F., Barbosa, L. A. M. Palazzo. Content Distribution in Trail-aware Environments. *Journal of the Brazilian Computer Society*, v.16, n.3, pp.163–176, 2010.
Available: <http://link.springer.com/article/10.1007/s13173-010-0015-1>
- [11] A. Wagner, J. L. V. Barbosa, D. N. F. Barbosa. A Model for Profile Management Applied to Ubiquitous Learning Environments. *Expert Systems with Applications*, v.41, n.4, p.2023–2034, 2014.
Available: <http://dx.doi.org/10.1016/j.eswa.2013.08.098>
- [12] J. H. Rosa, J. L. V. Barbosa, G. O. Barcelos. ORACON: An Adaptive Model For Context Prediction. *Expert Systems with Applications*, v.45, n.1, p.56–70, 2016.
Available: <http://dx.doi.org/10.1016/j.eswa.2015.09.016>

Trabalhos aceitos

Abordagem Atividade-Intensidade para o Reconhecimento de Atividades

Eduardo Gomes (UFF)
Natália Melgaço Pereira Torres (UFF)
Izabela Mocaiber (UFF)
Danilo Artigas (UFF)
Alessandro Copetti (UFF)

Avaliação de Desempenho na Comunicação em Redes Acústicas Aquáticas Utilizando Modems Reais de Baixo Custo

Thales T. Almeida (UFV)
Lucas Bragança (UFV)
Rafael Antônio (UFMG)
Francisco Lima (UFMG)
Marcos A. M. Vieira (UFMG)
Jose G. R. Júnior (CEFET-MG)
Luiz F. M. Vieira (UFMG)
José A. M. Nacif (UFV)

Avaliando Algoritmos para um Elemento Sensor Autônomo

Alexandra da Silva Pereira (UFV e UFMG)
Thais R. M. B. Silva (UFV)
Fabrício A. Silva (UFV)
Antonio A. F. Loureiro (UFMG)
Linnyer B. Ruiz (UEM)

BBAware - A Context-Aware Mobile and Wearable Architecture for Monitoring Beta-Blocked Cardiac Patients

Alberto Vianna Dias da Silva (UFBA)
Vaninha Vieira (UFBA e FPC-UFBA)

Black Box: Uma Aplicação de Gerenciamento e Entrega Tolerante a Atrasos e Desconexões de Vídeos Para Veículos

Charles T. B. Garrocho (UFOP)
Maurício J. da Silva (UFOP)
Ricardo A. R. Oliveira (UFOP)

BlinGui: uma Solução Vestível de Apoio a Pessoas Portadoras de Deficiência Visual na Detecção de Obstáculos Estáticos em Ambientes Internos

Elidiane P. dos Santos (UFBA)
Flávia S. Nascimento (UFBA e IFBA)
Vaninha Vieira (UFBA e FPC-UFBA)

Como Inferir o Meio de Transporte das Pessoas em Cenários Urbanos a Partir de Smartphones

Carlos Alvaro de M. S. Quintella (UNIRIO)

Carlos Alberto V. Campos (UNIRIO)

Leila C. V. Andrade (UNIRIO)

Controle de Consumo de Água Baseado em Reconhecimento Facial

Pedro E. Santos (UI)

Roubert Edgar S. Silva (UI)

Dalmy F. Carvalho Jr (UI)

Ilo A. S. Rivero (UI)

Raquel A. F. Mini (PUC Minas)

: Uma Abordagem Utilizando

Teoria dos Jogos

Thiago Montenegro Uchoa (UFPE)

Heitor S. Ramos (UFAL)

(UFPE)

Explorando Modelos Contextuais Híbridos: uma Abordagem Composicional

Roger Machado (UFPE)

Ricardo Almeida (UFPE)

Diórgenes da Rosa (UFPE)

João Lopes (UFRGS)

Ana Pernas (UFPE)

Adenauer Yamin (UFPE)

From a Smart House to a Connected City: Connecting Devices Services Everywhere

João Paulo Cardoso de Lima (UFSC)

Leandro Buss Becker (UFSC)

Frank Siqueira (UFSC)

Analucia Schiaffino Morales (UFSC)

Gustavo Medeiros de Araujo (UFSC)

Internet das Coisas Aplicada nas Redes Elétricas Inteligentes Brasileiras

Efrem E. O. Lousada (IFNMG)

Raquel A. F. Mini (PUC Minas)

Lotus@Runtime: Uma Ferramenta para Monitoramento e Verificação em Tempo de Execução para Sistemas Autoadaptativos

Davi Monteiro Barbosa (UECE)

Paulo Henrique Mendes Maia (UECE)

Evilásio Costa Júnior (UECE)

(Meta)Modelagem de Espaços Inteligentes Pessoais e Espaços Inteligentes Fixos para Aplicações Ubíquas

Marcos Alves Vieira (IF Goiano)
Sergio T. Carvalho (UFG)

Modelagem Genética para Estender o Tempo de Vida de Redes de Sensores Sem Fio com Sorvedouro Móvel

Sérgio de Oliveira (UFSJ)
Pedro Mitsuo Shiroma (UFSJ)
Marconi de Arruda Pereira (UFSJ)
Rone Ilídio da Silva (UFSJ)
Cristiano Maciel da Silva (UFSJ)

Planejamento Regional Adaptativo em Sistemas Self-Adaptive de Larga Escala

Eudes S. Andrade (IFBA)
Gabriel F. Sousa (IFBA)
Sandro S. Andrade (IFBA)

Projetando a Infraestrutura de Comunicação em Redes Veiculares para a Distribuição de Conteúdos Demandando Níveis Distintos de Desempenho

Cristiano M. Silva (UFSJ)
Cristiano G. Pitangui (UFSJ)
Alex V. Bastos (UFSJ)
Rone I. Silva (UFSJ)
. Silva (UFV)

Rede de Sensores para Controle Inteligente de Ambientes

Artejose Revoredo da Silva (UFRN)
João C. Xavier-Júnior (UFRN)
Ivanovitch Medeiros Dantas da Silva (UFRN)

Resolvendo Conflitos em Aplicações de Distribuição de Conteúdo em Redes Veiculares

Ronan D. Mendonça (UFMG e UFV)
Thais R. M. Braga Silva (UFV)
Fabrício A. Silva (UFV)
Linnyer B. Ruiz (UFMG e UEM)

Sistema Embarcado para o Controle de Acesso em Áreas Restritas de Prédios Inteligentes

Hyuri S. Maciel (UFAL)
David F. Silva (UFAL)
Clóvis G. M. do Nascimento (UFAL)
Andre L.L Aquino (UFAL)

Smart Station: Um Sistema Pervasivo de Notificação em Paradas de Ônibus para Pessoas com Deficiência Visual

Lucas Agostini (UFPel)

Alex Braga (UFPel)

Julio Machado (UFPel)

Alysson Nogueira (UFPel)

Tatiana Tavares (UFPel)

Adenauer Yamin (UFPel)

The classification of algorithms for Privacy Management in Ubiquitous Environments

Valderi R. Q. Leithardt (UFRGS) (DEI, UC-PT)

Carlos O. Rolim (UFRGS)

Anubis G. M. Rossetto (UFRGS)

Guilherme A. Borges (UFRGS)
(DEI, UC-PT)

Claudio F. R. Geyer (UFRGS)

Um Mecanismo Baseado em Sensibilidade à Situação para Otimizar a Transmissão de Dados de Aplicações de Sensoriamento Urbano

Carlos O. Rolim (UFRGS)

Anubis G. Rossetto (UFRGS)

Valderi R. Q. Leithardt (UFRGS)

Guilherme A. Borges (UFRGS)

Tatiana F. M. dos Santos (UFSC)

Adriano M. Souza (UFSC)

Claudio F. R. Geyer (UFRGS)

Um Modelo de Monitoramento de Pacientes Utilizando a Ciência de Situação

Lucas Closs (UNISINOS)

(UNISINOS)

Rodrigo da Rosa Righi (UNISINOS)

Um Modelo Proativo de Ações de Times de Resposta Rápida Baseado em Análise Preditiva

Fabio de Oliveira Dias (UNISINOS e IFSul)

Cristiano André da Costa (UNISINOS)

Rodrigo da Rosa Righi (UNISINOS)

Um Modelo Ubíquo para Retriagem de Pacientes Baseado na Internet das Coisas

Guilherme Wunsch (UNISINOS)

Cristiano André da Costa (UNISINOS)

Rodrigo da Rosa Righi (UNISINOS)

Um Serviço de Representação Ontológica de Contexto Baseada no Padrão de Projeto Estímulo-Sensor-Observação

Ernesto Fonseca Veiga (UFG)

Renato de Freitas Bulcão Neto (UFG)

Uma Arquitetura Hierárquica Multinível para Ciência de Situação em Segurança da Informação

Ricardo Borges Almeida (UFPEL)

Roger da Silva Machado (UFPEL)

Diógenes Y. L. da Rosa (UFPEL)

Lucas Medeiros Donato (DMU-UK)

Adenauer Yamin (UFPEL)

Ana Marilza Pernas (UFPEL)

Uma Arquitetura para IoT Direcionada à Ciência do Contexto Baseada em Eventos Distribuídos

Rodrigo Souza (UFRGS e IFSUL)

João Lopes (UFRGS e IFSUL)

Anderson Cardozo (UCPEL)

Tainã Carvalho (UFPEL)

Patrícia Davet (UFPEL)

Alexandre Wolf (UNISINOS)

Adenauer Yamin (UCPEL e UFPEL)

Jorge Barbosa (UNISINOS)

Cláudio Geyer (UFRGS)

Uma Proposta de Instrumentação Remota Explorando a Ubiquidade da Infraestrutura da IoT

Lucas M. S. Xavier (UCPEL e UFPEL)

Huberto Kaiser Filho (UFPEL)

Patrícia T. Davet (UFPEL)

Adenauer C. Yamin (UCPEL e UFPEL)

Uma Proposta de Rede Social Espontânea para Interação Ubíqua em Eventos

Marcelo da Silva dos Santos (UNISINOS)

Cristiano André da Costa (UNISINOS)

Jorge Luis Victória Barbosa (UNISINOS)

Abordagem Atividade-Intensidade para o Reconhecimento de Atividades

Eduardo Gomes¹, Natália Melgaço Pereira Torres¹, Izabela Mocaiber², Danilo Artigas¹, Alessandro Copetti¹

¹Instituto de Ciência e Tecnologia – Universidade Federal Fluminense (UFF)
28.895-532 – Rio das Ostras – RJ – Brasil

²Instituto de Humanidades e Saúde – Universidade Federal Fluminense (UFF)
28.895-532 – Rio das Ostras – RJ – Brasil

eduardosg2@gmail.com, natalia.melgaco@gmail.com, mocaiber@vm.uff.br,
daniloartigas@puro.uff.br, alessandro.copetti@gmail.com

Abstract. *In pervasive healthcare monitoring, a critical application in support to the patient is activity recognition. Despite the great number of studies on this topic, a parameter that has received less attention is intensity recognition. In the present work, we propose the coupling between activity and intensity, namely, Activity-Intensity, in accelerometer data to better describe daily activities of patients. Additionally, we investigated the feasibility of Fuzzy Logic in activity recognition. Results showed the viability of the classification and the good accuracy of Activity-Intensity recognition.*

Resumo. *No monitoramento em saúde pervasivo, uma aplicação essencial no acompanhamento do dia a dia da pessoa é o reconhecimento de atividades. Apesar dos diversos estudos sobre esse tema, um parâmetro pouco considerado é o reconhecimento da intensidade. Neste trabalho, propomos o acoplamento da atividade com a intensidade, a qual denominamos Atividade-Intensidade, em dados obtidos de acelerômetros, para melhor descrever as atividades diárias de um paciente. Adicionalmente, investigamos iniciativas de Lógica Fuzzy no reconhecimento de atividades. Os resultados mostraram a viabilidade da classificação e o bom desempenho do reconhecimento Atividade-Intensidade.*

1. Introdução

Com o avanço das tecnologias de sensores sem fio, posicionados tanto nos ambientes como no corpo das pessoas (computação vestível), está sendo possível projetar novos mecanismos de monitoramento remoto de pacientes e aperfeiçoar tanto a análise quanto a acurácia dos sistemas. No monitoramento em saúde, uma aplicação essencial no acompanhamento do dia a dia da pessoa é o reconhecimento de atividades como dormir, sentar, caminhar, varrer e etc. Uma forma de efetuar esse reconhecimento é com o auxílio dos dados fornecidos por sensores do tipo acelerômetro [Bulling et al. 2014] [Li et al. 2009] [Helmi et al., 2009].

Um aspecto que não tem sido abordado em trabalhos que envolvem reconhecimento de atividades é o reconhecimento do nível de *intensidade* da atividade em questão. Geralmente a interpretação dada para a intensidade é em relação ao esforço que a atividade proporciona. Neste trabalho, enfatizamos a necessidade de se considerar os

movimentos que são realizados durante uma atividade. Em trabalhos da área de Saúde, a intensidade é estratificada em termos de atividade leve, moderada e vigorosa [Ainsworth et al. 2000]. Nesse sentido, é importante acoplar esse parâmetro ao reconhecimento de atividades. Na área biomédica, a intensidade de movimentos tem sido determinada com métodos fisiológicos específicos que envolvem o cálculo do gasto energético. Entretanto, em determinadas abordagens, é importante ter a informação da intensidade do movimento através de métodos mais simples e sem o interesse principal, com o gasto energético.

Com esse acoplamento da atividade com a intensidade, a qual denominamos Atividade-Intensidade, é possível representar melhor o conhecimento de profissionais de saúde, na forma de sistemas de decisão com regras de especialistas. Exemplos de regras em um sistema de decisão, como apresentado em [Copetti et al. 2013], podem ser estabelecidos: a) se a *atividade* é doméstica, a *intensidade* é vigorosa e a *frequência cardíaca* é alta, então a situação do Paciente é de alerta (e não de emergência). Com isso, as alterações nos sinais vitais influenciadas por mudanças no comportamento poderiam ser toleradas com base em regras que levem esses fatores em consideração. Em outro exemplo, se a *atividade* é dormindo e a intensidade é vigorosa, então incrementa a possibilidade de insônia. Os dois exemplos mostram então uma intensidade explícita enquanto a pessoa realiza a atividade.

O presente trabalho investiga o impacto da discriminação da intensidade no reconhecimento de atividades distintas. Assim, propomos o acoplamento Atividade-Intensidade para aperfeiçoar os mecanismos de reconhecimento de atividade. Adicionalmente, investigamos iniciativas de Lógica Fuzzy no reconhecimento de atividades. A justificativa por essa técnica se dá por permitir expressar o conhecimento de especialistas no seu conjunto de regras e eventualmente beneficiar a inferência da Atividade-Intensidade.

O trabalho está estruturado como segue: a seção 2 apresenta trabalhos na área de Reconhecimento de Atividades, dando ênfase para trabalhos que já fizeram uso de Lógica Fuzzy como mecanismo de inferência ou de classificação. A seção 3 aborda o Reconhecimento de Atividades acoplado a intensidade dos movimentos. A seção 4 apresenta os testes com essa abordagem. Por fim, são apresentadas as conclusões.

2. Trabalhos Relacionados em Reconhecimento da atividade

Os sistemas de monitoramento para reconhecimento da atividade são ferramentas essenciais para os profissionais de saúde. Os sistemas podem propiciar intervenções precoces, e atuar tanto na reabilitação quanto na prevenção. Podemos imaginar sistemas de reconhecimento de atividades que analisem informações sobre quantos e em quais momentos o paciente experimenta uma alteração em sua rotina física. Uma aplicação prática pode ser a identificação e a análise de momentos de falta de mobilidade. Em uma situação contrária, pode ser identificado, por exemplo, o excesso de tremores [Palmerini et al. 2011].

Muitos trabalhos [bin Abdullah et al. 2012] [Gjoreski, H. et al. 2015] [Li 2009] já usam técnicas para reconhecimento de atividades das pessoas, principalmente, com o uso de algoritmos de aprendizado de máquina. Recentemente, algumas pesquisas têm direcionado para contextos específicos de saúde, como, por exemplo, envolvendo o autismo [Albinali 2009], reabilitação [Pärkkä 2009] e hipertensão [Copetti 2009]. Essa diversidade de áreas de aplicação na saúde faz com que a interação tenha que ser maior entre o conhecimento do especialista e o projetista do sistema. Portanto, há a necessidade de

modelos que consigam expressar conceitos e regras, com fácil interpretação por parte do especialista.

No trabalho de [Copetti et al. 2013] a Lógica Fuzzy foi usada para modelar variáveis relacionadas a dados fisiológicos associados, em alto nível, com o comportamento. Nesse trabalho, alterações nos sinais vitais (elevação da frequência cardíaca) influenciados por mudanças no comportamento (ex: realização de uma atividade como varrer) puderam ser tolerados como dentro da normalidade.

No trabalho de [Helmi 2009] foi utilizado Lógica Fuzzy para reconhecer quatro atividades distintas utilizando um acelerômetro triaxial na cintura do voluntário. As atividades estudadas foram: ir para frente, saltar, subir escada e descer escada. Nesse estudo, foi observado que a Lógica Fuzzy obteve um desempenho melhor do que os outros classificadores, tendo um total de 95% de acurácia no reconhecimento das atividades.

No trabalho de [Chiang 2012] o sistema de Lógica Fuzzy foi utilizado no reconhecimento de atividades corporais para auxiliar na fisioterapia de pacientes com AVC (acidente vascular cerebral) e na fisioterapia geriátrica. Para coletar os movimentos do corpo humano utilizou-se uma placa com acelerômetro e giroscópio no tórax e outra na coxa e as atividades avaliadas foram deitar, sentar, levantar, andar e correr. A acurácia total no reconhecimento das atividades alcançou mais de 96%.

No trabalho de [Yang 2008] foi utilizado uma rede Fuzzy-Bayesiana para reconhecer as atividades dos voluntários no seu dia-a-dia através de um acelerômetro 2D no antebraço e mais quatro sensores fisiológicos. Foram nove atividades do cotidiano analisadas: caminhar, correr, se exercitar, comer, ler, estudar, brincar, descansar e dormir. A acurácia obtida foi de 74,4% no reconhecimento das atividades, incluindo atividades dinâmicas e atividades estáveis.

Já o trabalho de [Liu e Chang 2009] realiza o reconhecimento das atividades andando, sentando e caindo. A técnica neuro-fuzzy foi utilizada e 3 acelerômetros foram posicionados no corpo no peito, na cintura e na coxa. A acurácia obtida foi de 88,7%.

Nos trabalhos discutidos de [Helmi 2009] [Chiang 2012] [Yang 2008], existe uma complexidade em definir conjuntos para as variáveis Fuzzy de entrada e de saída, pois o problema do reconhecimento da atividade depende do conhecimento dos dados (em contraposição ao conhecimento dos especialistas). Assim, é necessário criar conjuntos Fuzzy bem definidos para as variáveis de entrada, as quais são as acelerações dos eixos ou são medidas estatísticas aplicadas a essas acelerações. Já a variável de saída atividade, apresentada em [Helmi 2009] e [Chiang 2012] como uma sequência de conjuntos fuzzy - cada conjunto é uma atividade - é uma forma que viola o benefício da sobreposição de conjuntos Fuzzy. Já a técnica de Neuro-Fuzzy, utilizada no trabalho de [Liu e Chang 2009], considera o conhecimento dos dados para ajustar os pesos sinápticos da rede neural na fase de treinamento, o que é adequado para o problema. No entanto, a técnica perde a interpretabilidade das regras por extrapolar as combinações em razão das entradas.

O nosso trabalho apresenta uma nova abordagem, complementar a forma como é tratado o reconhecimento de atividades dos trabalhos discutidos nessa seção, integrando a intensidade dos movimentos à atividade sendo executada, isto é, abordagem Atividade-Intensidade. Para nosso conhecimento, essa abordagem não foi explorada até o momento, como pode ser verificado nos estudos de área, como, [bin Abdullah et al. 2012] [Bulling et al. 2014] [Gjoreski, H. et al. 2015], bem como trabalhos específicos contemplando uma

interpretação da intensidade diferente da adotada aqui [Ribeiro Filho et al. 2016] [Reiss 2014], discutidos na seção 3.

3. Reconhecimento de Atividades Acoplado à Intensidade dos Movimentos

A intensidade de uma atividade, sob a perspectiva do esforço fisiológico dispendido, é habitualmente determinada pelo Gasto Energético Metabólico ou MET. O MET (Metabolic Energy Expenditure) [Ainsworth et al. 2000] reflete o consumo energético associado à execução de uma atividade. O MET representa por definição a proporção entre a taxa metabólica durante o trabalho físico e a taxa metabólica basal padrão de 1,0. 1 MET é considerado a taxa metabólica em repouso enquanto sentado quieto. Então, de maneira organizada e sistematizada, pode-se estabelecer diferentes níveis de intensidades (leve: <3,0 METs; moderado: 3 - 6 METs e vigoroso (>6 METs). São atribuídos níveis de intensidade à cada atividade em função do gasto energético expresso em METs. Essa classificação de esforço depende do *tipo* de atividade e do grau de *intensidade* da mesma, gerando um valor representativo (ex. andar de bicicleta: 8,0 METs e meditar : 1,0 MET). Assim, a intensidade do movimento recebe geralmente uma interpretação relativa ao gasto energético [Ainsworth et al. 2000]. Observa-se aqui que a variável intensidade pode ser naturalmente modelada como variável Fuzzy, e ainda fazer parte de regras em sistemas de decisão em saúde.

A atividade física é definida como qualquer movimento corporal realizado pela musculatura esquelética que leve a algum gasto acima do repouso [Caspersen et al., 1985]. Assim, a quantidade de energia gasta na execução do movimento é o critério usado para diferenciar a atividade física. A medição do gasto energético é amplamente discutida no meio científico por meio de consumo de oxigênio, frequência cardíaca e acelerômetros [Melanson e Freedson, 1996]. Apesar da complexidade da medida da atividade física, instrumentos acessíveis e de baixo custo são de interesse de pesquisadores e clínicos. Nesse sentido, a utilização de sensores de movimento parecem ser interessantes, e se apoiam na ideia de que movimentos corporais sejam reflexo do gasto energético [Melanson e Freedson, 1996].

Na área de Computação Pervasiva, alguns trabalhos seguem essa interpretação de esforço para a intensidade. [Ribeiro Filho et al. 2015] e [Ribeiro Filho et al. 2016] propõem um sistema de reconhecimento de atividades e medição de sua intensidade. Um experimento foi desenvolvido com 3 acelerômetros posicionados na coxa, no peito e na cintura. As atividades analisadas incluem correndo, subindo e descendo escadas, andando, em pé, sentado e deitado. A taxa de acerto ficou em torno de 82%. A frequência cardíaca é usada para medir a intensidade, no entanto, o tratamento e a análise desses dados não são discutidos para uma classificação da intensidade.

Já no trabalho de [Reiss 2014], variações de Adaptive Boosting são usadas como técnicas para estimar a intensidade da atividade nas três categorias (leve, moderado e vigoroso). Usando a base de dados PAMAP2 [Reiss & Stricker 2012], a qual possui dados de atividades monitoradas com três acelerômetros e monitor de frequência cardíaca, são obtidos resultados próximos a 90% para estimar a intensidade. Por exemplo, atividades como estar sentado, trabalhar no computador e assistir TV são classificadas como de esforço leve.

No entanto, neste trabalho estamos interessados na movimentação (oscilação) da

pessoa durante uma atividade, isto é, na interação atividade x intensidade. Adotamos a mesma classificação de intensidade, mas sem associar ao gasto energético. Alguns exemplos motivadores de situações para adoção da abordagem Atividade-Intensidade são:

- movimentação excessiva durante o sono: a pessoa mesmo estando deitada, apresenta níveis diferentes de movimentação que deveriam ser identificados para saber, por exemplo, qual a intensidade dos movimentos no último minuto.
- ritmo da caminhada que difere do habitual: a atividade de andar pode apresentar características a serem exploradas por especialistas quanto a análise da caminhada.
- permanência prolongada em um sofá: estar sentado pode revelar momentos que a pessoa adormece ou que realiza tarefas manuais.

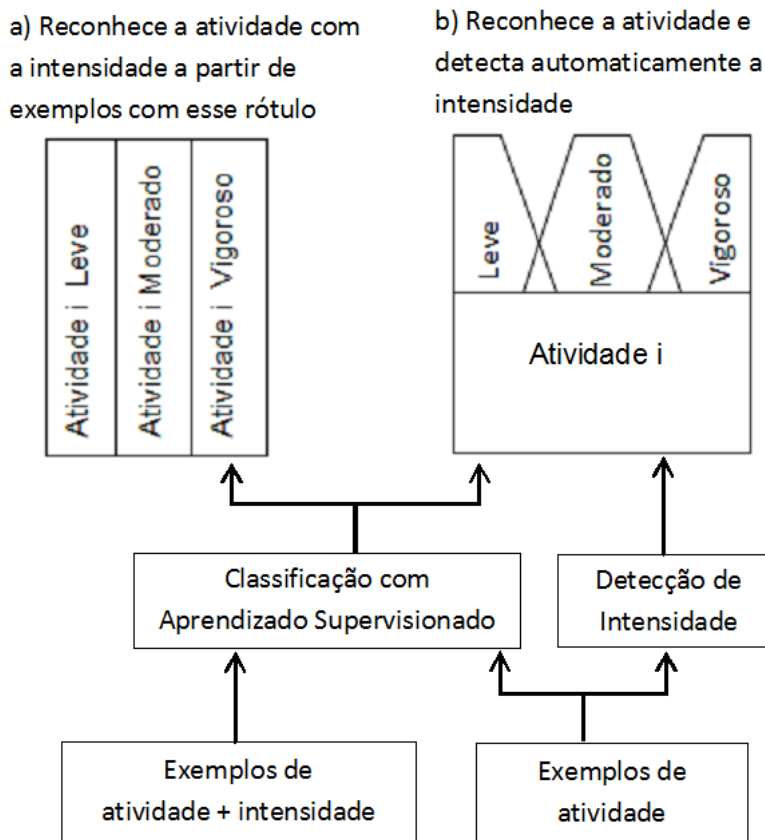


Figura 4.1 Proposta de abordagem Atividade-Intensidade

A Figura 4.1 apresenta a abordagem proposta de Atividade-Intensidade com duas opções para obter esse acomplamento. Na opção ‘a’, os exemplos usados no treinamento são rotulados com a atividade+intensidade. Essa forma é adotada nos experimentos da seção 4. Na opção ‘b’, o exemplo possui o rótulo somente da atividade. Em seguida, há a necessidade de uma técnica para detectar os limites entre intensidades, por exemplo, agrupamento em 3 clusters ou análise estatística dos dados. Os limites também podem ser usados para definir funções de pertinência para uma variável fuzzy.

4. Experimentos e Análise dos Resultados

O principal objetivo do experimento foi investigar se é possível obter bons resultados no reconhecimento da intensidade de determinadas atividades. Neste trabalho, investigamos 3 atividades: deitado, sentado e andando. As atividades foram escolhidas por serem as mais corriqueiras de um idoso em seu domicílio. Cada atividade tem 3 tipos de intensidade avaliados: leve, moderado e vigoroso. Dessa forma, totalizando 9 possíveis situações. De maneira detalhada, cada situação corresponde as seguintes exemplos do dia a dia: Deitado leve: todo corpo parado na cama; Deitado moderado: em certo tempo vira o corpo para mudar de posição; Deitado vigoroso: simulação de uma insônia crônica onde todos os membros inclusive a cabeça alteram a posição inicial; Sentado leve: membros inferiores totalmente relaxados no assento; Sentado moderado: pernas inquietas e braços praticamente relaxados; Sentado vigoroso: membros inferiores e superiores ativos; Andar leve: caminhada lenta; Andar moderado: caminhada normal; Andar vigoroso: caminhada com passos rápidos.

O experimento consistiu na coleta de dados de 5 voluntários (4 homens e 1 mulher) na faixa de 15 a 50 anos. Cada voluntário possuía um celular posicionado dentro de uma pochete fixa na cintura conforme Figura 4.2. Os voluntários foram orientados a reproduzir cada uma das situações. Durante essas reproduções, o acelerômetro 3D presente no celular registrou informações da aceleração da gravidade do dispositivo e do tempo. O celular utilizado possui processador de 1GHz, 1GB de memória RAM, com SO Android 4.1.2.

O tempo de coleta de dados no celular foi de 1 minuto para cada uma das 9 situações. A segmentação foi aplicada com uma janela de 2,5s (sem sobreposição). Na extração de características, as medidas estatísticas aplicadas incluíram: média, mediana, moda, variância, desvio padrão e RMS. Em seguida, os dados foram discretizados, possibilitando um tratamento por faixas numéricas. Devido aos motivos apresentados na seção 2 sobre Lógica Fuzzy aplicada no reconhecimento da atividade, a técnica de classificação utilizada foi a de Redes Neurais. A validação utilizou validação cruzada de 10 partições.



Figura 4.2 Posição do celular no corpo do voluntário.

O primeiro teste verifica o desempenho do reconhecimento para o acoplamento Atividade-Intensidade. A Figura 4.3 apresenta a acurácia para todas as 9 situações. Os melhores resultados são para as atividades de deitado e sentado. Nessas atividades, se analisarmos as intensidades (leve, moderado e vigoroso) vemos que não apresentam

variações significativas, além de não ser possível identificar a intensidade que obteve os piores resultados. Esses resultados mostram a viabilidade de integrar atividade com intensidade do movimento.

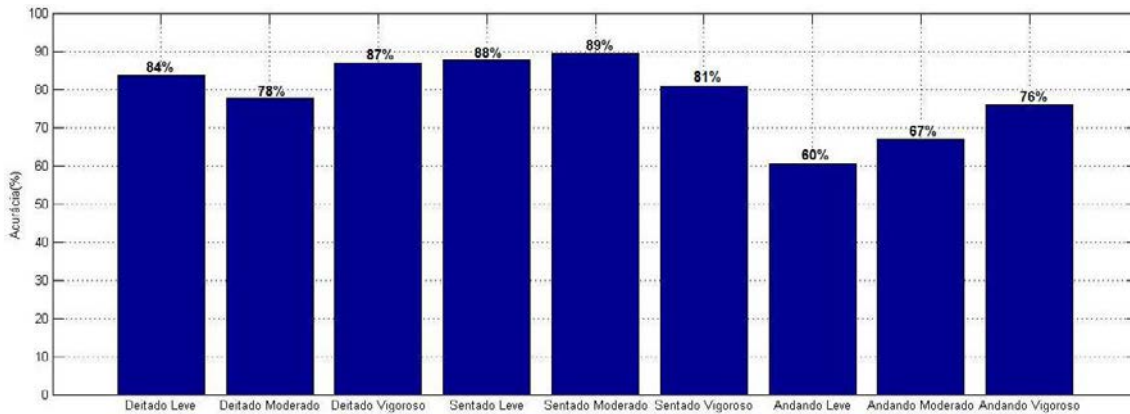


Figura 4.3 Acurácia nas 9 situações de Atividade-Intensidade.

Para avaliar mais detalhadamente a atividade de andando, a Figura 4.4 mostra a matriz de confusão obtida. Com a matriz é possível ver que a atividade de ‘andando’ não obteve resultados ótimos em comparação com as demais atividades. A diferença no ritmo da caminhada de cada voluntário (mais lento ou mais rápido) pode explicar esse resultado. Um fato que comprova essa observação na matriz de confusão é o erro sempre quantificado entre vizinhos de intensidade e nunca entre o ‘leve’ e o ‘vigoroso’, e um número reduzido de 9 casos que foram classificados erroneamente como ‘SentadoVigoroso’, mas deveriam ser ‘AndandoLeve’.

a	b	c	d	e	f	g	h	i	<-- classified as
105	15	2	0	1	0	0	0	0	a = DeitadoLeve
11	94	16	0	0	0	0	0	0	b = DeitadoModerado
3	15	104	0	0	0	0	0	0	c = DeitadoVigoroso
0	0	0	105	11	4	2	0	0	d = SentadoLeve
0	1	0	7	105	7	1	0	0	e = SentadoModerado
0	0	0	2	5	98	13	3	0	f = SentadoVigoroso
0	0	0	1	0	9	80	26	4	g = AndandoLeve
0	0	0	1	0	2	19	83	16	h = AndandoModerado
0	0	1	0	0	0	8	21	90	i = AndandoVigoroso

Figura 4.4 Matriz de confusão da classificação das 9 situações.

Uma outra razão para os resultados menores na atividade de andando é devido ao posicionamento do acelerômetro. Testes realizados com um acelerômetro no pulso (embarcada em um relógio) resultaram em uma acurácia acima de 85%.

A Figura 4.5 mostra a acurácia quando somente é considerada a atividade sem a intensidade. Esse cenário é considerado tradicional, apresentado na maior parte dos trabalhos relacionados. Vemos que a técnica obteve excelentes resultados.

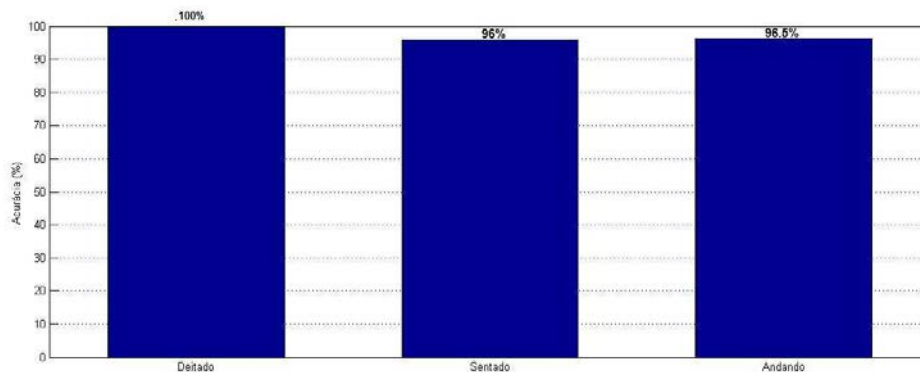


Figura 4.5 Acurácia considerando somente a atividade dos voluntários.

Para verificar os resultados com somente as situações que apresentam na prática um maior distanciamento em relação a intensidade dos movimentos, fizemos um teste somente com as situações: deitado leve, sentado moderado e andando vigoroso (Figura 4.6). Ou seja, são os exemplos mais representativos de situações mais equidistantes. O teste mostra que a integração atividade-intensidade não deteriora o resultado, ao considerar situações extremas entre as 9 trabalhadas.

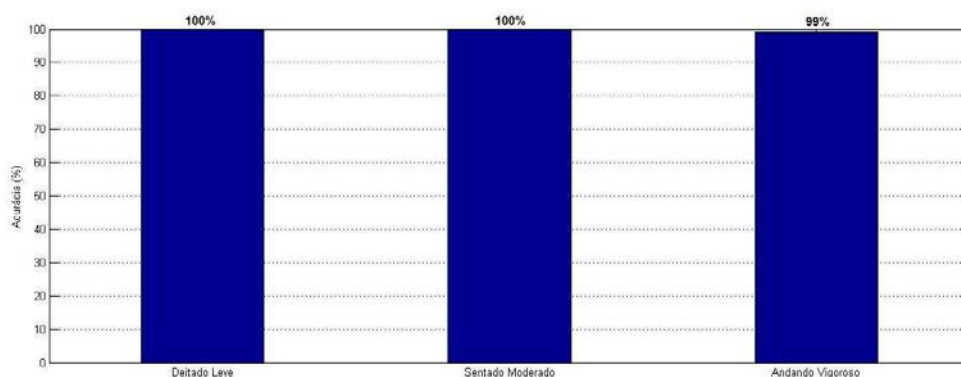


Figura 4.6 Classificação com dados de somente 3 situações que apresentam maior distanciamento em relação a intensidade dos movimentos.

Concluindo, vemos que, das três atividades que dominam o maior tempo da vida diária de idosos, a atividade de andar acoplada a intensidade é a que obteve a acurácia mais baixa (60%-76%, Figura 4.3). Esse resultado pode ser melhorado com o uso de mais acelerômetros ou o posicionamento do acelerômetro em outras partes do corpo. Ademais, mantendo nossa proposta de uso do smartphone, podemos treinar com dados somente do paciente (específico do usuário). Segundo [Bulling et al. 2014], o desempenho é geralmente mais alto do que o caso de ser independente do usuário.

5. Conclusões

O trabalho investigou a inserção do conceito da intensidade no reconhecimento de atividades, propondo a abordagem Atividade-Intensidade. A classificação utilizando essa abordagem obteve bons resultados: a acurácia média foi de 83% para atividade deitado;

86% para sentado; e 67,67% para andando. Nesse sentido, acreditamos que a abordagem Atividade-Intensidade forneça recursos para melhor descrever as atividades diárias de um paciente. A Lógica Fuzzy pode modelar de uma forma natural a variável *intensidade*, além de ser uma ferramenta interessante para sistemas de decisão envolvendo Atividade-Intensidade. Além disso, os resultados apresentados são de grande interesse pois esse mecanismo de classificação, baseado em acelerômetros, tem um custo muito baixo sendo portanto de fácil acesso.

Como trabalho futuros, pretendemos explorar ainda mais os resultados obtidos com o posicionamento de mais acelerômetros em diferentes partes do corpo, o que beneficia determinados tipos de atividades. Pretendemos, ainda, aplicar técnicas como a clusterização para agrupar os diferentes níveis de intensidade levando em conta a personalização, isto é, detectar os níveis adequados de intensidade para uma pessoa específica. De forma complementar, um Sistema Especialista Fuzzy será desenvolvido para analisar o comportamento de pacientes, agregando o classificador desenvolvido no presente trabalho.

Agradecimentos

Os autores gostariam de agradecer o apoio financeiro recebido da FAPERJ e da PROPPI/PDI/UFF.

Referências Bibliográficas

- Ainsworth, B. E. et al. (2000). “Compendium of physical activities: an update of activity codes and MET intensities”. *Medicine and science in sports and exercise*, 32(9; SUPP/1), S498-S504.
- Albinali, F., Goodwin, M. S. and Intille, S. S. (2009) “Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum”, *Ubiquitous Computing*, pp. 71-80.
- Bulling, A., Blanke, U., & Schiele, B. (2014). “A tutorial on human activity recognition using body-worn inertial sensors”. *ACM Computing Surveys (CSUR)*, 46(3), 33.
- bin Abdullah, M. F. A., Negara, A. F. P., Sayeed, M. S., Choi, D. J., & Muthu, K. S. (2012). Classification algorithms in human activity recognition using smartphones. *International Journal of Computer and Information Engineering*, 6, 77-84.
- Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports*, 100(2), 126.
- Chiang, S. Y., Kan, Y. C., Tu, Y. C. and Lin, H. C. (2012) “Activity Recognition by Fuzzy Logic System in Wireless Sensor Network for Physical Therapy”, *Intelligent Decision Technologies*, vol. 2, no. 16, pp. 191-200.
- Copetti, A., Leite, J. C. B., Loques, O., da Nóbrega, A. C. L. and Barbosa, T. P. (2009) “Chequer. Intelligent Context-Aware Monitoring of Hypertensive Patients.” In: *Situation Recognition and Medical Data Analysis in Pervasive Health Environments*.

- Pervasive Healthcare Conference, Londres, Reino Unido.
- Copetti, A., Leite, J. C. B. and Loques, O. (2013) “A Decision-making Mechanism for Context Inference in Pervasive Healthcare Environments”, *Decision Support Systems*, 55(2): 528-537.
- Gjoreski, H. et al. (2015). Competitive Live Evaluations of Activity-Recognition Systems. *Pervasive Computing, IEEE*, 14(1), 70-77.
- Helmi, M. and AlModarresi, S.M.T. (2009) “Human Activity Recognition Using a Fuzzy Inference System”. FUZZ-IEEE. Korea.
- Li, Q., Stankovic, J. A., Hanson, M. A., Barth, A. T., Lach, J. and Zhou, G. (2009) “Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information”, *Wearable and Implantable Body Sensor Networks*, pp. 138-143.
- Liu, S.H. and Chang, Y.J. (2009) “Using accelerometers for physical actions recognition by a neural fuzzy network”, *Telemedicine and e-Health*, 15(9): 867-876.
- Melanson Jr, E. L., Freedson, P. S., & Blair, S. (1996). “Physical activity assessment: a review of methods”. *Critical Reviews in Food Science & Nutrition*, 36(5), 385-396.
- Palmerini, L., Rocchi, L., Mellone, S., Valzania, F., & Chiari, L. (2011). “Feature selection for accelerometer-based posture analysis in Parkinson's disease”. *Information Technology in Biomedicine, IEEE Transactions on*, 15(3), 481-490.
- Pärkkä, J. et al. (2009) “Relationship of psychological and physiological variables in long-term self-monitored data during work ability rehabilitation program”, *Information Technology in Biomedicine*, 13(2): 141-151.
- Reiss, A., & Stricker, D. (2012). “Introducing a new benchmarked dataset for activity monitoring”. In *Symposium on Wearable Computers (ISWC)*, (pp. 108-109). IEEE.
- Reiss, A. (2014). “Personalized mobile physical activity monitoring for everyday life” (Tese de Doutorado, Technical University of Kaiserslautern).
- Ribeiro Filho, J. D. P., Silva, F. J. S., Coutinho, L. R. and Gomes, B. T. P. (2015) “MHARS: Sistema Móvel de Reconhecimento de Atividades em Ambient Assisted Living”, In: *Simpósio Brasileiro de Computação Ubíqua e Pervasiva- SBCUP*, Recife.
- Ribeiro Filho, J. D. P., Silva, F. J. S., Coutinho, L. R., Gomes, B. T. P. and Endler, M. (2016) “A Movement Activity Recognition Pervasive System for Patient Monitoring in Ambient Assisted Living”, *31st ACM Symposium On Applied Computing (SAC)*, Health Track.
- Yang, S. I. and Cho, S. B. (2008) “Recognizing human activities from accelerometer and physiological sensors” *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*. Seoul, Korea.

Avaliação de Desempenho na Comunicação em Redes Acústicas Aquáticas Utilizando Modems Reais de Baixo Custo

Thales T. Almeida¹, Lucas Bragança¹, Rafael Antônio²
Francisco Lima², Marcos A. M. Vieira², Jose G. R. Júnior³
Luiz F. M. Vieira² e José A. M. Nacif¹

¹Universidade Federal de Viçosa (UFV)
Florestal, MG – Brasil

²Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, MG – Brasil

³Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
Leopoldina, MG – Brasil

{thales.almeida, lucas.braganca, jnacif}@ufv.br
{rafael.antonio, francisco.lima, mmvieira, lfvieira}@dcc.ufmg.br
jgrjunior@leopoldina.cefetmg.br

Abstract. *Water networks have many potential applications, such as preservation of the ecosystem, and prevention of environmental disasters. As radio-frequency signals are quickly absorbed underwater, these networks use the acoustic signal as a communication system. Compared to terrestrial networks, acoustical networks have unique characteristics, such as low bandwidth, high latency and large delay spread. With so many restrictions, one should emphasize the influence exerted by acoustic modems, whose characteristics can affect network performance. Since low-cost modems are more affordable, a good understanding of its operation is necessary. This article aims to analyze the performance of acoustic communication from the comparison between the theoretical and actual transmission rates (used in simulation experiments) and the actual transmission rates obtained for a low cost modem. Experiments were performed, where it was possible to identify the deterioration in modem transmission rates, caused by the effects of propagation delays in the medium.*

Resumo. *Redes aquáticas possuem diversas aplicações potenciais, como preservação do ecossistema e prevenção de desastres ambientais. Como sinais de radio-frequência são rapidamente absorvidos debaixo d'água, estas redes utilizam o sinal acústico como sistema de comunicação. Comparada com redes terrestres, redes acústicas possuem características únicas, como baixa largura de banda, alta latência e grande atraso de propagação. Com tantas restrições, deve-se ressaltar a influência exercida por modems acústicos, cujas características podem afetar o desempenho da rede. Uma vez que modems de baixo custo são mais acessíveis, se faz necessária uma boa compreensão acerca de seu funcionamento. Este artigo tem por objetivo analisar a performance da comunicação acústica a partir da comparação entre as taxas de transmissão teóricas e reais obtidas por um modem de baixo custo. Experimentos reais foram realizados, onde foi possível identificar o comprometimento nas taxas de transmissão, deteriorada pelos efeitos dos atrasos de propagação no meio.*

1. Introdução

Redes aquáticas apresentam potencial para uso em várias aplicações, incluindo o monitoramento de derramamento de óleo, monitoramento de lagos e de reservatórios de hidrelétricas [Vieira et al. 2012]. Como os sinais de rádio-frequência são rapidamente absorvidos debaixo d'água, estas redes utilizam enlaces acústicos para se comunicar, uma vez que a propagação acústica é melhor suportada em baixas frequências. No entanto, apesar de apresentar um baixo custo de instalação e manutenção, a largura de banda disponível é extremamente limitada, apresentando baixíssima velocidade quando comparada com outros meios de propagação, como a propagação óptica e eletromagnética. Em transmissões acústicas, a velocidade de propagação do som alcança uma taxa máxima de 1.500 m/s. Características como a perda de transmissão, o alto nível de ruído, a propagação em múltiplos caminhos, o efeito *Doppler* e o atraso de propagação variável tornam a comunicação acústica ainda mais complexa [Stojanovic and Preisig 2009].

Redes acústicas aquáticas podem operar utilizando Veículos Autônomos Aquáticos (VAA) ou sensores para monitoramento, usando enlaces acústicos para comunicação. A rede pode ser ligada a uma estação central, responsável por conectá-la ao mundo externo. Esta estação pode ainda estar conectada a um *backbone*, que permite a conexão da rede à Internet. Este cenário possibilita que dados coletados sejam acessados remotamente, em tempo real, por cientistas e pesquisadores envolvidos em projetos de monitoramento. No entanto, a área de monitoramento destes equipamentos é limitada pelo raio de cobertura do modem acústico, que varia de poucos metros a poucos quilômetros [Catipovic et al. 1993].

Além da limitação da área de monitoramento, outro ponto relevante se refere à transmissão dos modems acústicos, que sofre consideravelmente os efeitos dos atrasos de propagação no meio [Brady and Catipovic 1994], afetando o desempenho da rede. Devido ao alto custo para aquisição de equipamentos mais robustos [Benson et al. 2010], modems de baixo custo são mais acessíveis para construir plataformas de testes e formar redes mais densas, sendo necessária uma boa compreensão acerca de seu funcionamento e possíveis impactos que a utilização destes equipamentos podem causar no desempenho e performance da rede. Neste contexto, este trabalho tem por objetivo analisar a performance da comunicação acústica a partir da comparação entre as taxas de transmissão teóricas (utilizadas em experimentos simulados) e taxas de transmissão reais (obtidas em testes de campo) de um modem de baixo custo. Nos experimentos, optou-se pela utilização do modem acústico SAM-1 [Star 2016], devido principalmente ao seu custo reduzido comparado a outros equipamentos [Benson et al. 2010] e possibilidade de variação da taxa de transmissão. Nos testes reais realizados, foi possível identificar o comprometimento da taxa de transmissão dos modems, a qual é severamente prejudicada pelos efeitos dos atrasos de propagação no meio.

A principal contribuição deste artigo, além de apresentar informações que auxiliem os pesquisadores com interesse em redes acústicas, mais especificamente nos seus dispositivos de monitoramento, está nos resultados obtidos nos experimentos. Tais resultados demonstram que as taxas máximas de transmissão utilizadas em simulações não refletem as taxas máximas reais quando são consideradas características como a perda de transmissão, o alto nível de ruído, a propagação em múltiplos caminhos, o efeito *Doppler* e o atraso de propagação variável, existentes em um cenário real. Ademais, os dados

apresentados possibilitam o desenvolvimento de protocolos MAC (*Median Access Control*) mais eficientes, permitindo determinar a taxa máxima de transferência de pacotes.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados, enquanto a Seção 3 apresenta a caracterização do modem SAM-1, utilizado nos experimentos. São apresentadas ainda as características de um modem real e os respectivos modelos de taxa de transferência. A Seção 4 apresenta os resultados dos experimentos realizados. Finalmente, na Seção 5, são apresentadas as conclusões e a descrição de trabalhos futuros.

2. Trabalhos Relacionados

Alguns trabalhos existentes na literatura indicam que a taxa de transmissão de um modem acústico é de aproximadamente 10 kb/s. Segundo Yackoski *et al.* [Yackoski and Shen 2008], a taxa de transmissão de um modem utilizado nos experimentos simulados é definida como sendo de aproximadamente 15 kb/s, com a velocidade de propagação de 1.500 m/s (padrão em comunicações aquáticas que utilizem o canal acústico como meio de transmissão). Vale ressaltar que esta taxa em *bits* por segundo corresponde ao valor teórico de um modem utilizado em simulações de laboratório.

O mesmo ocorre no trabalho de Ma *et al.* [Ma and Lou 2011], no qual os autores constroem um algoritmo de agendamento para lidar com as incertezas espaço-temporais geradas pelos longos atrasos de propagação em redes acústicas. No trabalho, os autores também adotam como parâmetro a taxa teórica de transferência do modem de 15 kb/s, além de outros parâmetros adicionais como velocidade de propagação do sinal acústico, tamanho do pacote de dados e duração do atraso de transmissão dos pacotes. Gan *et al.* [Guan et al. 2011] projetaram um algoritmo baseado em agendamento que leva vantagem sobre os longos atrasos dos sinais em ambiente acústico, mas também assumem a mesma taxa de transmissão do modem de 15 kb/s.

Park *et al.* [Park and Rodoplu 2007] propõem um protocolo MAC eficiente em termos energéticos, utilizando uma taxa de transmissão do modem de cerca de 5.4 kb/s. Cabe ressaltar que todos os experimentos foram realizados por meio de simulações, não utilizando modems reais.

Na realidade, estas taxas de transmissão, em sua maioria, não correspondem à taxa real de transmissão. Não se pode pré-determinar a taxa de transmissão máxima, o que pode prejudicar, entre outras situações, o funcionamento e o desempenho da rede. Além disso, o desenvolvimento de protocolos são sensíveis a essas circunstâncias e, portanto, a avaliação de modems reais é primordial para o desenvolvimento e funcionamento de aplicações aquáticas.

A evidência de que as taxas reais de transmissão, em sua maioria não correspondem com a taxa teórica foi identificada nos experimentos realizados por Zhu *et al.* [Zhu et al. 2013]. Após a realização de diversos testes de campo, as características de um modem real foram identificadas. Os autores utilizaram um dispositivo da série ATM, o modem *Teledyne Benthos* e, nos experimentos práticos realizados pelo autor, a taxa máxima de transmissão alcançou aproximadamente 800 b/s. Em condições ideais, a taxa de transmissão de modems desta categoria é 2400 b/s [Benson et al. 2010]. De acordo com os resultados de Zhu *et al.*, os longos preâmbulos e as baixas taxas de trans-

missão são determinantes no desempenho de redes aquáticas, influenciando o tempo total de transmissão e a probabilidade de colisão dos pacotes.

Preâmbulos são sequências incorporadas ao início do *frame* da camada física, com a finalidade de sincronização, detecção do sinal e estimativa do canal [Zhu et al. 2013]. Quando o atraso gerado pelo tempo de transmissão do preâmbulo é muito curto e a taxa de transmissão é alta, ambos são desconsiderados pela comunidade, como em meios de transmissão via cabo e rede sem fio. No entanto, para sistemas acústicos reais, o tempo de transmissão do preâmbulo é muito longo e não pode ser descartado.

De acordo com os experimentos realizados em [Zhu et al. 2013], o tempo de transmissão encontrado para o preâmbulo em sistemas aquáticos reais foi de 1.5 s, aumentando o tempo de transmissão de pacotes e conseqüentemente a possibilidade de colisão entre eles. De acordo com os autores, para calcular o tempo de transmissão do pacote é necessário analisar três variáveis:

1. T_{pre} : refere-se ao tempo de transmissão do preâmbulo, que é incorporado no tempo de transmissão do pacote;
2. (L): refere-se ao tamanho do pacote em *bits*;
3. (B): corresponde a taxa de transmissão efetiva do modem em *bits* por segundo.

A equação (1) [Zhu et al. 2013] apresenta o tempo de transmissão de um pacote, T_{tx} , considerando o tempo de transmissão do preâmbulo:

$$T_{tx} = \frac{L}{B} + T_{pre} \quad (1)$$

3. Caracterização do Modem SAM-1

O modem acústico SAM-1, utilizado nos experimentos realizados, é caracterizado por uma tecnologia robusta que opera em vários cenários como bacias portuárias, recifes e águas rasas. Este dispositivo trabalha em condições ideais de baixa velocidade no ambiente aquático com um alcance de até 1.000 m. Outras características importantes do modem SAM-1 referem-se à sua sensibilidade ao sinal acústico, que varia de 85 dB a 125 dB, além da velocidade de comunicação, que varia de 5 a 154 b/s. Em relação a outros modems acústicos, o SAM-1 destaca-se pelo menor consumo de energia, menor custo e capacidade de utilização em ambientes com ruídos [Star 2016]. O modem SAM-1 é ilustrado na Figura 1.

Existem diversas opções de taxas de transmissão no modem SAM-1. A velocidade de transmissão dos dados acústicos pode ser definida por dois modos distintos de configuração: modem configurado para salto de frequência ou configurado para canal único. Caso o modo de configuração seja salto de frequência, as taxas de transmissão podem ser selecionadas entre 15, 38, 77 e 154 b/s. Caso o modo escolhido seja canal único, as taxas são de 5, 13 e 38 b/s [Star 2016].

É importante ressaltar que o modo de configuração em canal único (utilizado nos experimentos práticos deste artigo) deve ser selecionado quando a comunicação for realizada entre dois módulos de modems SAM-1. As demais taxas são utilizadas quando o modem SAM-1 estiver em uma comunicação com outro modem capaz de se comunicar em saltos de frequência.



Figura 1. Modems SAM-1 [Star 2016].

A Tabela 1 apresenta as diversas taxas de transmissão configuráveis para o modem SAM-1 [Star 2016]. O fabricante fornece diretrizes para a seleção das taxas de transmissão para alguns ambientes típicos. Estas diretrizes são apresentadas na Tabela 2.

Tabela 1. Taxas de transmissão do modem SAM-1 [Star 2016].

Modo de Configuração	Taxa de Transmissão
Salto de frequência	S0: 15 b/s
Salto de frequência	S0: 38 b/s
Salto de frequência	S0: 77 b/s
Salto de frequência	S0: 154 b/s
Canal-único	S0: 5 b/s
Canal-único	S0: 13 b/s
Canal-único	S0: 38 b/s

4. Resultados Experimentais

Os experimentos apresentados nesta seção foram realizados em um cenário real, dentro de uma piscina. Foi utilizado um par de modems, modelo SAM-1, distantes 2 m um do outro, conforme pode ser observado na Figura 2. A rotina de transmissão e recepção foi realizada por ambos os modems, ora agindo como transmissor, ora como receptor.

Utilizou-se do envio de pacotes de dados de diversos tamanhos, variando de 2 a 32 *bytes* (aumentando sempre de 2 em 2 *bytes*). No que se refere às taxas de transferência do modem, foram utilizadas as taxas de 5, 13 e 38 b/s. O nível de sensibilidade à ruído foi definido em 109 dB. Para cada taxa de transferência do modem, foram realizadas quatro repetições dos experimentos.

As Figuras 3, 4 e 5 apresentam os gráficos que demonstram as médias das porcentagens de *bytes* dos pacotes entregues para cada uma das três taxas de transferência possíveis (variando o tamanho do pacote).

Nos experimentos realizados, é possível observar que quando as taxas de 5 e 13 b/s são selecionadas há alta probabilidade de a entrega do pacote ser feita com sucesso. Re-

Tabela 2. Diretrizes para seleção de velocidade para alguns ambientes típicos - Adaptado de [Star 2016].

Ambiente	Velocidade Máxima	Descrição
Lagos	Canal Único: S5-S6 Salto de Frequência: S2-S3	Um "lago típico", ou seja, enlamado e com uma depressão levemente inclinada no solo, com pouca superfície de eco. Ecos não tem duração longa, então altas velocidades podem ser usadas.
Piscinas e Tanques	Canal Único: S4 Salto de Frequência: S0	Piscinas e tanques exibem ecos muito persistentes. Somente operações em velocidades mínimas são possíveis.
Recifes Superficiais	Canal Único: S5 Salto de Frequência: S1-S2	O som tende a saltar do fundo rochoso e superfície do mar, criando ecos moderados e exigindo, portanto, a velocidade moderada.
Águas Profundas	Canal Único: S6 Salto de Frequência: S2-S3	Há poucos ecos em águas profundas, sendo assim, altas velocidades são possíveis.
Portos	Canal Único: S4-S5 Salto de Frequência: S0-S2	As superfícies lisas e resistentes de navios e docas refletem de forma eficiente sinais acústicos. Ecos são de persistência moderada à forte.



Figura 2. Cenário do Experimento.

sultados mostram uma taxa média de entrega superior a 90% na maior parte dos cenários com variação do tamanho do pacote. Já considerando a taxa máxima de transmissão do modem (38 b/s), pacotes pequenos (com 2 a 10 *bytes*) obtiveram taxa média de entrega inferior a 70%. É de consenso comum que pacotes menores tem menos chances de causar colisões no meio. Contudo, conforme análise apresentada por Le *et. al.* [Le et al. 2013], caso o tempo de início da transmissão destes pacotes não seja cuidadosamente agendado por um protocolo MAC, pode degradar gravemente a performance da rede.

Como apresentado na Seção 2, a presença de preâmbulos longos e baixas taxas de transmissão aumenta consideravelmente o tempo para transmissão de pacotes em redes acústicas aquáticas equipadas com modems reais. Nos experimentos realizados neste trabalho, foi possível observar os efeitos do atraso de propagação do meio acústico no tempo para transmissão de pacotes entre os modems. A disparidade entre as taxas teóricas estimadas e as taxas reais do modem SAM-1 acarretaram em uma substancial diferença no tempo final para entrega dos pacotes. As Figuras 6, 7 e 8 apresentam as médias dos tempos de transmissão real comparadas com os tempos de transmissão ideal de pacotes no modem SAM-1 para cada uma das taxas de transmissão configuráveis do modem (5,

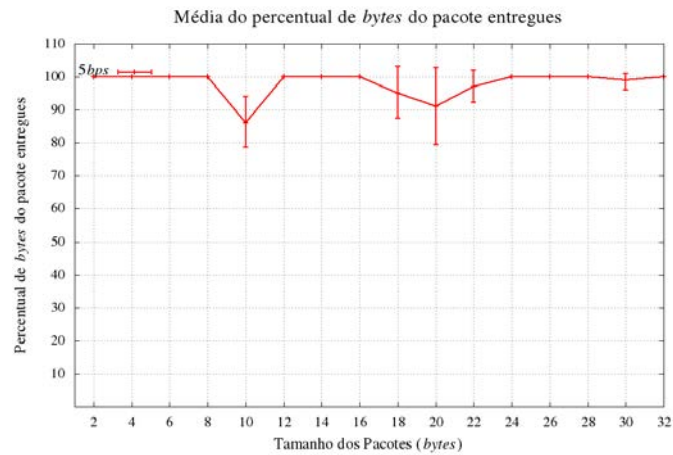


Figura 3. Média do percentual de *bytes* do pacote entregues (5 b/s).

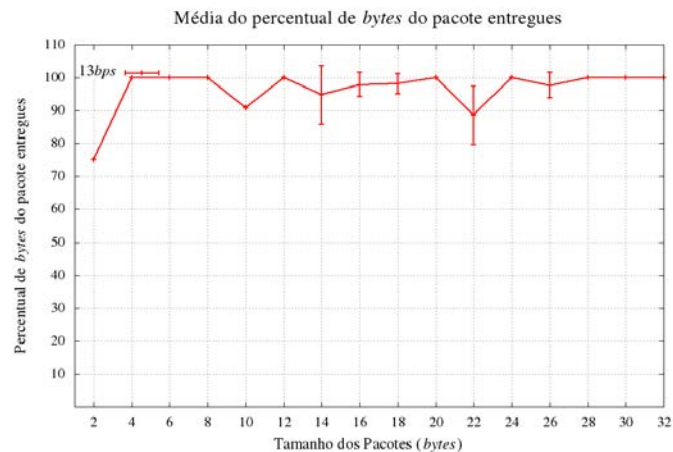


Figura 4. Média do percentual de *bytes* do pacote entregues (13 b/s).

13 e 38 b/s) e com variação do tamanho do pacote para cada velocidade. A Tabela 3 apresenta o resultado da taxa real obtida em cada transmissão realizada, considerando todos os pacotes transmitidos.

É possível observar uma grande disparidade entre o tempo de transmissão ideal e o tempo de transmissão real de pacotes, principalmente à medida que o tamanho do pacote aumenta. Como exemplo, conforme apresentado no gráfico da Figura 6, um pacote de 32 *bytes* em condições ideais demoraria aproximadamente 50 s com uma taxa de transferência de 5 b/s para ser entregue. Em condições reais, este tempo aumenta para quase 90 s. A taxa de transmissão real do modem acústico também foi aferida. O modem SAM-1 alcançou no máximo 50% de sua taxa de transmissão máxima, demonstrando claramente que as taxas máximas de transmissão utilizadas em simulações não refletem as taxas máximas reais, uma vez considerados todos os efeitos de perda de propagação inerentes aos ambientes reais.

5. Conclusão

Neste trabalho, foi realizada a avaliação de enlaces de redes acústicas aquáticas utilizando um modem de baixo custo, modelo SAM-1. O principal objetivo dos experimentos é de-

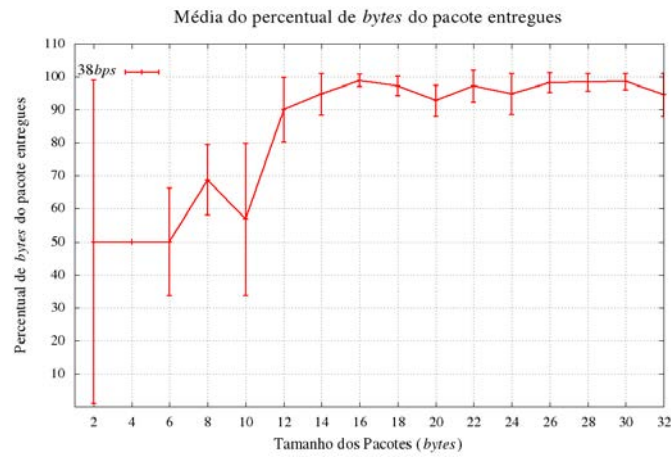


Figura 5. Média do percentual de bytes do pacote entregues (38 b/s).

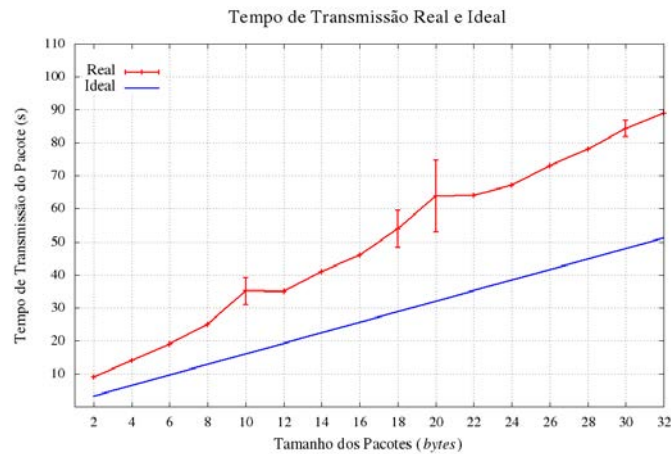


Figura 6. Tempo de Transmissão Real x Tempo de Transmissão Ideal (5 b/s).

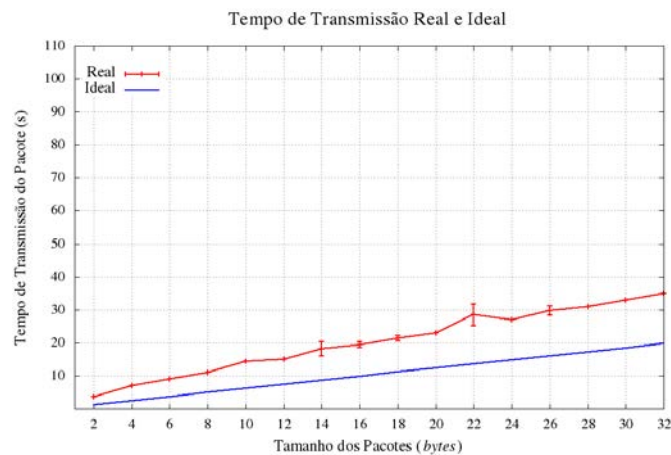


Figura 7. Tempo de Transmissão Real x Tempo de Transmissão Ideal (13 b/s).

monstrar que as taxas máximas de transmissão utilizadas em simulações não refletem as taxas máximas reais quando são consideradas características como a perda de trans-

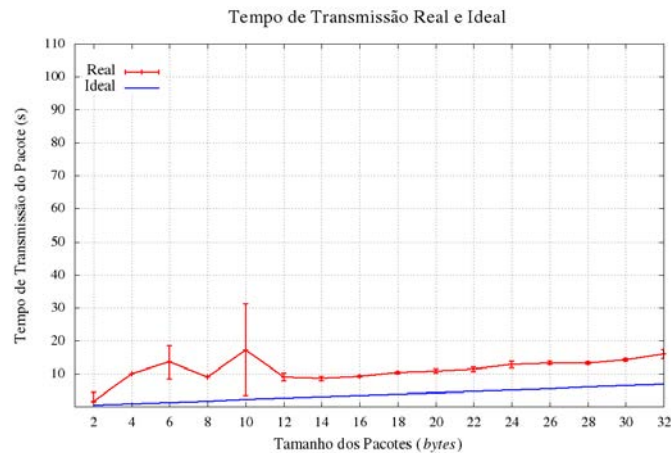


Figura 8. Tempo de Transmissão Real x Tempo de Transmissão Ideal (38 b/s).

Tabela 3. Taxa máxima de transmissão real obtida.

Pacote (em bytes)	Taxa Real (5 b/s)	Taxa Real (13 b/s)	Taxa Real (38 b/s)
2	1,78 b/s	4,57 b/s	10,67 b/s
4	2,29 b/s	4,57 b/s	3,20 b/s
6	2,53 b/s	5,33 b/s	3,56 b/s
8	2,56 b/s	5,82 b/s	7,11 b/s
10	2,29 b/s	5,59 b/s	4,63 b/s
12	2,74 b/s	6,40 b/s	10,65 b/s
14	2,73 b/s	6,18 b/s	13,19 b/s
16	2,78 b/s	6,58 b/s	14,07 b/s
18	2,67 b/s	6,72 b/s	13,98 b/s
20	2,51 b/s	6,96 b/s	14,80 b/s
22	2,75 b/s	6,17 b/s	15,51 b/s
24	2,87 b/s	7,11 b/s	15,11 b/s
26	2,85 b/s	6,99 b/s	15,71 b/s
28	2,87 b/s	7,23 b/s	16,94 b/s
30	2,85 b/s	7,27 b/s	16,88 b/s
32	2,88 b/s	7,31 b/s	16,05 b/s

missão, o alto nível de ruído, a propagação em múltiplos caminhos, o efeito *Doppler* e o atraso de propagação variável, existentes em um cenário real. Conforme demonstrado nos experimentos práticos realizados neste trabalho, a taxa de transmissão real do modem acústico SAM-1 foi, no melhor dos cenários, 50% da taxa de transmissão teórica, independente da velocidade configurada no equipamento. Como consequência levou-se, em muitos casos, o dobro do tempo que se levaria em condições normais para transmitir um pacote, independente de seu tamanho.

Acredita-se que os dados apresentados neste artigo servirão de auxílio na compreensão e na modelagem do comportamento dos modems acústicos, o que permitirá o projeto de protocolos sensíveis à necessidade de pré-determinar a taxa de transferência máxima e probabilidade de colisão na rede. Espera-se ainda que os resultados possam auxiliar o desenvolvimento de aplicações em diversas áreas, como monitoramento de lagos, exploração de petróleo e monitoramento da infra-estrutura na indústria petrolífera.

Para trabalhos futuros, planeja-se o desenvolvimento de protocolos que se baseiem nas taxas de transmissão reais, permitindo o funcionamento correto da rede.

6. Agradecimentos

À CAPES, CNPq e FAPEMIG, pelas bolsas e apoio financeiro.

Referências

- Benson, B., Li, Y., Kastner, R., Faunce, B., Domond, K., Kimball, D., and Schurgers, C. (2010). Design of a Low-cost, Underwater Acoustic Modem for Short-range Sensor Networks. In *OCEANS 2010 IEEE - Sydney*, pages 1–9.
- Brady, D. P. and Catipovic, J. A. (1994). Acoustic Local Area Networks. US Patent 5,303,207.
- Catipovic, J., Brady, D., and Etchemendy, S. (1993). Development of Underwater Acoustic Modems and Networks. *Oceanography*, 6(3):112–119.
- Guan, Y., Shen, C.-C., and Yackoski, J. (2011). MAC Scheduling for High Throughput Underwater Acoustic Networks. In *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, pages 197–202. IEEE.
- Le, S. N., Zhu, Y., Peng, Z., Cui, J.-H., and Jiang, Z. (2013). PMAC: a Real-world Case Study of Underwater MAC. In *Proceedings of the Eighth ACM International Conference on Underwater Networks and Systems*, page 2. ACM.
- Ma, J. and Lou, W. (2011). Interference-aware Spatio-temporal Link Scheduling for Long Delay Underwater Sensor Networks. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*, pages 431–439. IEEE.
- Park, M. K. and Rodoplu, V. (2007). UWAN-MAC: An Energy-efficient MAC Protocol for Underwater Acoustic Wireless Sensor Networks. *Oceanic Engineering, IEEE Journal of*, 32(3):710–720.
- Star, D. (2016). Manual do Fabricante. Acessado em fevereiro de 2016. Disponível em: <http://desertstar.com/product/sam-1/>.
- Stojanovic, M. and Preisig, J. (2009). Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization. *Communications Magazine, IEEE*, 47(1):84–89.
- Vieira, L. F., Vieira, M. A., Pinto, D., Nacif, J. A. M., Viana, S. S., and Vieira, A. B. (2012). HydroNode: An Underwater Sensor Node Prototype for Monitoring Hydroelectric Reservoirs. In *Proceedings of the Seventh ACM International Conference on Underwater Networks and Systems*, page 43. ACM.
- Yackoski, J. and Shen, C.-C. (2008). UW-FLASHR: Achieving High Channel Utilization in a Time-based Acoustic MAC Protocol. In *Proceedings of the Third ACM International Workshop on Underwater Networks*, pages 59–66. ACM.
- Zhu, Y., Jiang, Z., Peng, Z., Zuba, M., Cui, J.-H., and Chen, H. (2013). Toward Practical MAC Design for Underwater Acoustic Networks. In *INFOCOM, 2013 Proceedings IEEE*, pages 683–691. IEEE.

Avaliando Algoritmos para um Elemento Sensor Autônomo

Alexandra da Silva Pereira^{1,2}, Thais R. M. B. Silva¹, Fabrício A. Silva¹,
Antonio A. F. Loureiro², Linnyer B. Ruiz³

¹Instituto de Ciências Exatas – Universidade Federal de Viçosa - *Campus Florestal* (UFV)
35.690-000 - Florestal, MG - Brasil

²Departamento de Ciência da Computação Universidade Federal de Minas Gerais (UFMG)
31.270-010 - Belo Horizonte, MG - Brasil

³Departamento de Informática – Universidade Estadual de Maringá (UEM)
87.020-900 – Maringá - PR - Brasil

alexandra.pereira@ufv.br, thais.braga@ufv.br, fabricio.asilva@ufv.br
loureiro@dcc.ufmg.br, linnyer@gmail.com

Abstract. *Wireless Sensor Networks (WSN) are used in different contexts, including monitoring of climatic aspects in a remote region, actuators for geographic, spatial events and etc. Based on this limitation scenario of WSN a generic model that seeks to use a network computational resources in a more balanced way was developed, with autonomic computing concepts that create a self configurable network. In this paper we develop instances based on this generic model to improve network performance. We have implemented two more sophisticated algorithms and analyze their behavior in experiments performed in real sensor networks, also known as testbeds. The results show that the implementations maintain or improve important aspects such as the number of sent messages, self managed and sensitivity of the application.*

Resumo. *Redes De Sensores Sem Fio (RSSF) são utilizadas em diversos contextos, que podem incluir monitoração de aspectos climáticos em uma região remota, atuadores para eventos geográficos, espaciais e etc. Diante do cenário de limitação de RSSF foi desenvolvido um modelo genérico de Elemento Sensor Autônomo (ESA) que busca utilizar de maneira mais equilibrada os recursos computacionais de um nó sensor, com conceitos de computação autônoma que tornam a rede auto-configurável. Neste trabalho, desenvolvemos instâncias baseadas no modelo genérico para melhorar o desempenho da rede. Implementamos algoritmos de análise e planejamento, e avaliamos seus comportamentos em experimentos executados em redes de sensores reais. Os resultados mostram que as implementações mantêm ou melhoram aspectos importantes como o número de mensagens enviadas, a autoconfiguração e a sensibilidade da aplicação.*

1. Introdução

As Redes de Sensores Sem Fio são um tipo específico de redes sem fio, classificadas como rede *ad-hoc* e formadas por milhares de dispositivos chamados de nós

sensores[Buratti et al. 2009]. Para a formação de uma RSSF, nós sensores são disseminados em um ambiente remoto, geralmente, com a finalidade de monitorar e coletar dados pertinentes a modelagem da rede. Os nós sensores que compõem a rede são conhecidos por possuírem restrições de hardware (i.e., processamento, memória e energia)[Rawat et al. 2014].

Para tornar o gerenciamento de uma RSSF mais inteligente, foi proposto por [Braga 2006] [Braga et al. 2006] Elemento Sensor Autônomo (ESA). O ESA é relevante pela sua estrutura genérica, o que permite incluir e experimentar diversos algoritmos em cada uma de suas etapas. Ele é um modelo desenvolvido com base na proposta de tornar os sistemas de RSSF melhores no aspecto de auto-gestão. O seu funcionamento é cíclico e suas etapas são capazes de prover à rede os seguintes serviços: monitorar, analisar, planejar e executar ações ao longo do ciclo de execução de uma aplicação.

Diante do cenário de restrições em RSSFs e da característica genérica do ESA, a motivação é observar e analisar o comportamento de algoritmos no contexto do ESA. Portanto, implementamos, além da proposta original do ESA, três novas estratégias de análise e planejamento baseadas em meta-heurísticas.

Para todas as estratégias, foram feitas análises e levantamentos do quanto eles podem impactar a RSSF. O objetivo é avaliar a capacidade da rede de monitorar eventos mais rapidamente e mais precisamente (i.e., com mais informações), com o intuito de prover mais detalhes aos responsáveis por atuar nos casos de ocorrências de eventos, como um incêndio, por exemplo.

O presente trabalho segue organizado da seguinte forma. Na Seção 2 é feita uma apresentação de trabalhos presentes na literatura que se relacionam com o tema de pesquisa desenvolvido. A Seção 3 apresenta os cenários escolhidos, bem como, os algoritmos implementados. A Seção 4 discute os cenários de cada uma das implementações, quais foram os resultados e como estes resultados podem impactar o funcionamento da RSSF. Na Seção 5 apresentamos, finalmente, expectativas de trabalhos futuros e algumas considerações finais.

2. Trabalhos Relacionados

A computação autônoma aplicada às RSSFs é um tema que vem sendo explorado na literatura nos últimos anos. [Peng et al. 2014] discutem um modelo que utiliza um autômato celular para auto-organização da rede. Para isso, o algoritmo possui rotinas que mantêm a cobertura da rede estável, isto é, um algoritmo que verifica periodicamente a conectividade com os vizinhos, caso ele tenha menos de 2 vizinhos ativos, ele sai do modo *stand-by*. Os experimentos para validação foram feitos em simulação com uma rede de sensores homogênea.

Os autores [Sadouq et al. 2014] fazem a proposta de um arcabouço que trabalha com a rede de sensores como um cluster formado por vários nós. Os nós possuem 3 funções pré definidas: sensorear, rotear dados e atuar como ponto de acesso. O principal objetivo é ter uma rede mais sistêmica, robusta e escalável.

Na proposta discutida por [Abdelaal 2015], um modelo foi desenvolvido com o intuito de melhorar as técnicas existentes em RSSFs para economizar ainda mais energia. Os três principais objetivos são: agregar dados para o sensoriamento tornar-se mais barato, otimização de hardware e auto-adaptações preditivas.

O trabalho de [Jabeur et al. 2015] propõe modelos para redes de sensores sem fio baseadas em aspectos da natureza e são chamados de bio-inspirados. Nele, a rede formada pelos nós sensores é descrita como análoga a um ecossistema. Conforme este ecossistema passa por metamorfoses a rede evolui, e essa evolução resolve problemas que são recorrentes em RSSF. Todo o modelo proposto mostrou-se bastante interessante e pode trazer melhorias para aplicações em redes de sensores sem fio.

Esses aspectos da natureza também são explorados em outros trabalhos. Já [Lakshmi and Mohan Rao 2014] propõem um algoritmo com a finalidade de proporcionar auto-cura baseada em uma técnica de otimização de enxame. A ideia é estender o tempo de vida de uma rede de sensores. Neste trabalho foi desenvolvido um algoritmo baseado em colônia de formigas, no qual os pacotes buscam por caminhos mais curtos até o ponto de acesso. [Nguyen et al. 2014] propuseram um algoritmo genético para atuar na auto-configuração e otimização de uma RSSF. A rede é definida como um cluster, existe comunicação hierárquica entre os nós, e os nós que podem se comunicar com o ponto de acesso. O algoritmo genético trabalha para encontrar os melhores líderes, esses serão definidos como clusters e os nós dos quais cada cluster é responsável.

O trabalho proposto neste artigo difere dos anteriores por apresentar uma modelagem genérica e bem dinâmica no contexto das RSSF. O ESA genérico estabelece rotinas para controle autônomo da rede, e trabalha sob rotinas de auto-cura, auto-controle e auto-configuração. Partindo de uma implementação fiel ao modelo proposto por [Braga 2006], foram criadas instâncias que utilizam algoritmos mais sofisticados, que atuam na análise e no planejamento que compõem o ESA. O objetivo é fazer com que eventos ocorridos na rede sejam mais rapidamente e mais precisamente monitorados pelos nós sensores, fazendo com que os responsáveis pela atuação possam agir de maneira mais apropriada.

3. Soluções

Buscamos dois algoritmos que pudessem contribuir de maneira positiva para a autoconfiguração da rede. Logo, poderemos fazer comparações entre as instâncias mais básicas do ESA e as três novas instâncias desenvolvidas com algoritmos mais complexos e que exigem maior processamento. Foram implementados dois algoritmos diferentes para trabalhar em conjunto com o ESA. O primeiro é baseado em um algoritmo genético (GA), o segundo é uma implementação de colônia de formigas (ACS) e o terceiro uma composição com os dois algoritmos anteriores.

O objetivo dos algoritmos implementados é atuar na autoconfiguração dos parâmetros de sensoriamento dos ESAs. Em outras palavras, ambos os algoritmos devem ajustar os intervalos de sensoriamento e de disseminação dos dados sensoreadas de acordo com as medidas coletadas. Neste caso, quando os valores sensoreados estiverem dentro de um valor considerado fixo (i.e., sem ocorrência de eventos), é possível aumentar os intervalos para economizar recursos. Por outro lado, quando os valores sensoreados apresentarem mudanças significativas, os intervalos devem ser reduzidos para que os detalhes dos eventos não sejam perdidos.

3.1. Algoritmo Genético (GA)

Os algoritmos genéticos (GAs) utilizam conceitos provenientes do princípio de seleção natural para abordar uma série ampla de problemas, em especial de

otimização [Whitley 1994].

A nova instância do ESA com o algoritmo genético se concentra na monitoração e reproduz grandezas sensoreadas, para que a aplicação tenha uma perspectiva de análise mais interessante para as etapas de planejamento e execução.

```

1 AlgoritmoGenetico()
2 t=0;
3 cria populacao P(t);
4 for t ← 1 to P(t) do
5   | avalia aptidão indivíduo(t)
6 end
7 while condição de parada não satisfeita do
8   | t=t+1;
9   | Selecionar população P(t) de P(t-1);
10  | Aplicar operadores de cruzamento sobre P(t);
11  | Aplicar operadores de mutação sobre P(t);
12  | Avaliar P(t);
13 end

```

Algorithm 1: Pseudo-código GA.

O algoritmo 1 apresenta um pseudo-código do GA implementado, no qual a população é composta por variáveis que representam os dados sensoreados pela aplicação de RSSF. Para essa implementação, a geração inicial foi determinada por dados reais sensoreados em uma RSSF real. Com isso, a linha 3 do pseudo-código representa uma função que tem a responsabilidade de percorrer um vetor com dados sensoreados e representar cada indivíduo, que é representado por uma média de 3 três ocorrências de sensoreamento que foram observadas em experimentos prévios. Na linha 5, temos uma função responsável por fazer o cálculo da aptidão, que avalia o quão bom um indivíduo é.

Na linha 7, entramos em um laço que irá determinar quando finalizar o algoritmo, de acordo com o número de gerações estabelecidas. Neste laço, são executadas as etapas de seleção, cruzamento, mutação e avaliação. Na função de seleção, representada pela linha 9, é utilizado o resultado da aptidão para estabelecer quais serão os indivíduos que farão parte do cruzamento, no caso implementado, os indivíduos, ou dados, com maior variação (i.e., maior desvio padrão) foram considerados mais valiosos. Na linha 10, a chamada de cruzamento é responsável por executar o procedimento com um ponto de corte, onde o filho recebe características genéticas dos pais escolhidos. Em seguida, a linha 11 indica a função responsável pela mutação que causará variabilidades no indivíduo, hora com perturbações de 101,15%, hora com perturbações de 98,85%, valores esses escolhidos para não causar variações muito significativas. Por fim, uma nova chamada a avaliação, que determina a aptidão e escolhe quais indivíduos farão parte das próximas iterações.

3.2. Algoritmo Baseado em Regras (ESA-REGRAS)

As regras foram implementadas para que sejam executadas as seguintes reprogramações na etapa de planejamento:

- **Condição:** Caso o intervalo de sensoriamento esteja muito alto.

Ação: Diminuir intervalo até o limite de 1 segundo.

- **Condição:** Caso o intervalo de sensoriamento esteja muito baixo.
Ação: Aumentar intervalo até o limite de 10 segundos.
- **Condição:** Caso o intervalo de disseminação esteja muito alto.
Ação: Diminuir intervalo até o limite de 1 segundo.
- **Condição:** Caso o intervalo de disseminação esteja muito baixo.
Ação: Aumentar intervalo até o limite de 20 segundos.

3.3. Algoritmo de Colônia de Formigas (ACS)

O ACS surgiu em 1992 como fruto da tese do pesquisador italiano Marco Dorigo. Para a instância seguinte, optamos por integrar o ESA ao ACS. Este algoritmo retrata o comportamento forrageiro das formigas. Quando caminha em busca de alimento, uma formiga deixa um rastro pelo caminho, chamado de feromônio. O feromônio é responsável por estabelecer uma comunicação entre as formigas, e cada formiga deposita ao longo do caminho para indicar o quão interessante foi o caminho que ela escolheu. Com o decorrer do tempo, essa substância química passa por um processo chamado evaporação, que é exatamente a perda de feromônio. [Dorigo et al. 2006].

$$\rho_{i,j}^k = \frac{(\tau_{i,j})^\alpha (\eta_{i,j})^\beta}{\sum_{I \in N_i^k} (\tau_{i,j})^\alpha (\eta_{i,j})^\beta}$$

Figura 1. Função de Probabilidade de Transição.

Na instância implementada com o ACS, a heurística foi aplicada na etapa responsável pelo planejamento da melhor estratégia de disseminação a ser adotada pela rede. O algoritmo trabalha sob uma matriz de adjacência que representa a vizinhança factível, ou seja, quais as alternativas de variação nos intervalos de disseminação existem para replanejamento da rotina da rede a partir de determinado intervalo de tempo. Para escolher qual o tempo será mais adequado, o algoritmo trabalha com uma função de probabilidade de transição que pode ser observada na Figura 1, onde:

$\tau_{i,j}$: Representa o feromônio ligado à aresta i,j .

α e β : São valores determinados aleatoriamente que representam a influência do feromônio e da informação heurística.

N_i^k : Representa a vizinhança factível da formiga.

É importante dizer que o feromônio é de natureza aleatória e para o desenvolvimento deste trabalho, ele está configurado para receber valores que variam de 0 a 1, o que pode causar variações de 0% a 100%. Os pesos α e β também são de natureza aleatória e variam entre 0 e 1. A vizinhança factível representa as variações permitidas em intervalos de disseminação, como ilustrado pela Figura 2.

No grafo de exemplo da Figura 2, temos nos vértices o atual intervalo de disseminação, em segundos. As arestas representam a vizinha factível, isto é, quais as possibilidades novas de auto-configuração a partir de um estado, a configuração é passível de

incrementar ou decrementar o tempo da aplicação de acordo com o que seja mais interessante para a configuração da aplicação (e.g., diminuir na ocorrência de eventos e aumentar caso contrário). Eles levam em consideração os dados sensoreados e sua vizinhança para fazer o cálculo de qual estratégia deve ser adotada.

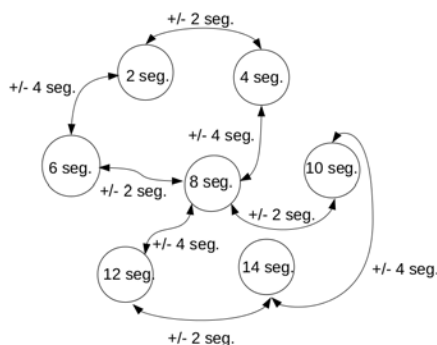


Figura 2. Grafo que ilustra intervalos de disseminação e possibilidades.

Pensando neste contexto do funcionamento básico do algoritmo de colônia de formiga, o integramos na etapa de planejamento do ESA. O algoritmo trabalha sob a ideia de um reforço positivo, que deposita ou evapora feromônios e portanto colocamos essa ideia no contexto do ESA e das RSSFs. A ideia do reforço positivo está diretamente ligada a como serão os intervalos de disseminação na RSSF. O funcionamento para o ESA é estabelecido na ideia de que enquanto o sensoramento retornar dados relevantes, será reforçada determinada estratégia de distribuição de informações sensoreadas com o objetivo de entregar mais detalhes sobre o evento para os interessados, e a partir do momento em que essa rotina torna-se desinteressante, há um reforço negativo, ou seja, evaporação de feromônio, para evitar que dados não tão relevantes sejam transmitidos com muita frequência. Com isso, haverá variações no intervalo em que as mensagens são disseminadas na rede.

3.4. Algoritmo Genético e de Colônia de Formigas

Em um outro cenário, implementamos uma aplicação que trabalha com as duas estratégias apresentadas anteriormente: GA e ACS. As mesmas configurações dos algoritmos foram mantidas. O ponto principal deste cenário é analisar o comportamento de dois algoritmos atuando em conjunto no ESA.

4. Implementação e Resultados

Todos os experimentos foram executados no ambiente real (i.e., *Testbed*) chamado Indriya [Doddavenkatappa et al. 2011], para que os resultados fossem coesos e consistentes. Para apresentar os resultados, foram feitas análises de métricas importantes para a detecção e detalhamento de eventos em uma RSSFs, sendo: quantidade de mensagens recebidas pelo ponto de acesso (PA), quantidade de mensagens enviadas pelos nós sensores ao longo do tempo, e tempo gasto para que a primeira notificação chegue ao PA. Vale destacar que, como os nós sensores do Indriya estão constantemente ligados à uma fonte de energia, não foi possível fazer análises confiáveis e coerentes sobre o consumo de energia nos nós sensores.

4.1. Cenários Implementados

Todas as aplicações foram executadas durante 10 minutos, para a obtenção de dados mais consistentes e concisos. Para todas as implementações, trabalhamos com linguagens de programação C e nesC, além de adotar TinyOS como o sistema operacional utilizado na plataforma de nós sensores.

Um nó telosB, utilizado nos experimentos, possui as seguintes características de hardware e software. O transceptor, utiliza um rádio TPR2420CA que opera em faixa de frequência de 2400 MHz a 2483.5 MHz com capacidade de transmissão de 250 kbps. O alcance está entre 75 a 100 metros para ambientes externos e de 20 a 30 metros para ambientes internos. Possui taxas de consumo de 23mA em modo de recepção, 21 μ A em modo ocioso e 1 μ A para modo sleep. O TelosB utiliza um processador Texas TI MSP430 com frequência de 8MHz e tem disponível os sensores de temperatura, luz, infravermelho e umidade. Em relação à memória, possui 1MB de memória não-volátil, 1024Kb de memória de programa e 10kB de RAM.

Para avaliar os algoritmos propostos para o ESA, foram implementados 5 cenários:

Básico (NAIVE): Não implementa nenhuma estratégia do ESA, sendo utilizada como base de comparação;

Algoritmo Baseado em Regras (ESA_REGRAS): implementa a solução que altera os parâmetros de acordo com regras definidas;

Algoritmo Genético (ESA_GA): implementa a solução que altera os parâmetros de acordo com o algoritmo genético;

Algoritmo de Colônia de Formiga (ESA_ACS): implementa a solução que altera os parâmetros de acordo com o algoritmo genético;

Algoritmos GA e ACS (ESA_GA_ACS): implementa as duas soluções, GA e ACS, em conjunto.

Ao longo dos experimentos, utilizamos 105 nós sensores. A aplicação implementada para todos os cenários coleta periodicamente (de 1 a 10 segundos) a temperatura do ambiente, e dissemina, também periodicamente (de 1 a 20 segundos), os valores coletados após a última disseminação. O intervalo de tempo de sensoriamento e disseminação é fixo em 5 segundos apenas para o cenário chamado básico (*NAIVE*). Para os demais cenários, esses valores são iniciados em 5 segundos, mas variam ao longo do tempo de acordo com a ocorrência de eventos na rede.

Para tornar possível avaliar a atuação dos cenários durante ocorrência de eventos, foram gerados eventos virtuais aleatoriamente na rede, fazendo com que aproximadamente 10 nós sensores "monitorem" valores bem superiores de temperatura, indicando, por exemplo, a ocorrência de um incêndio. Essa estratégia se fez necessária pois no *Test-bed* utilizado em um ambiente fechado, as temperaturas não variam significativamente. A Figura 3(a) representa os instantes de tempo que ocorreram eventos para os 5 cenários. Essa figura será referenciada na análise dos resultados para que seja possível verificar o momento de atuação dos ESAs implementados.

4.2. Resultados

Nessa seção discute-se algumas análises obtidas pelos dados do arquivo de saída fornecidos pelo *Testbed*.

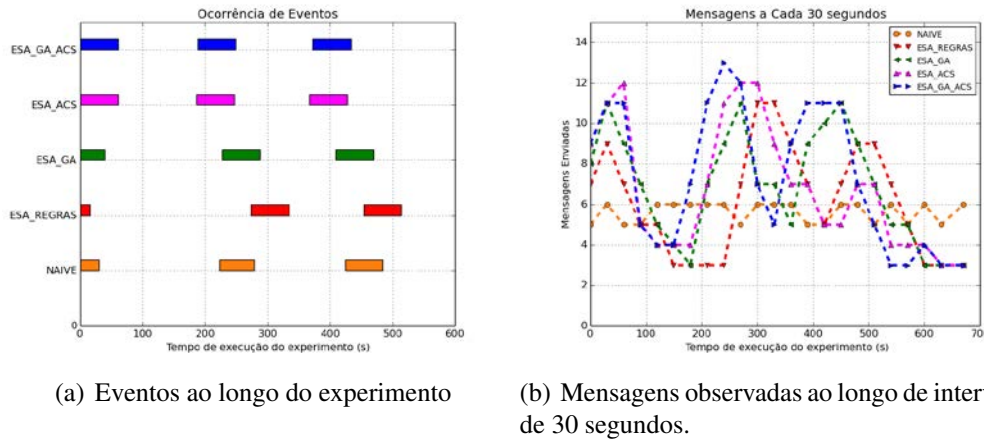


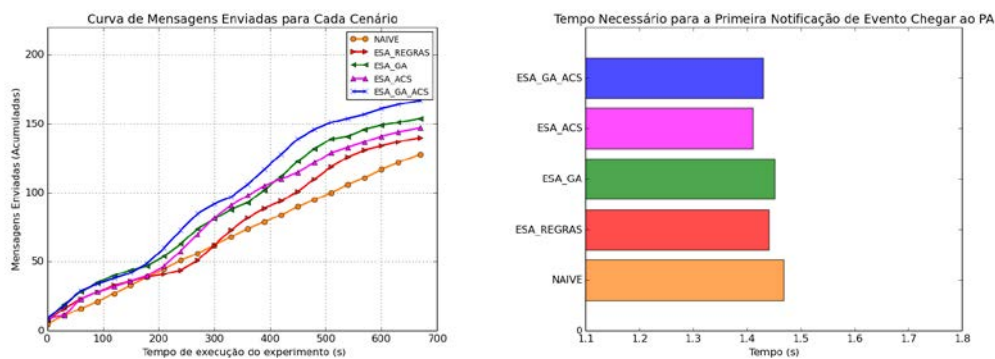
Figura 3. Eventos Virtuais e Mensagens ao Longo do Tempo.

A Figura 3(a) representa os eventos virtuais que aconteceram ao longo da execução dos experimentos para todos os cenários. Consideramos que por aspectos naturais de redes de computadores, como assincronicidade e paralelismo, não foi possível fazer com que os eventos fossem rigorosamente idênticos. Mesmo assim, foi possível observar o funcionamento do ESA nos cenários propostos e implementados, uma vez que ocorreram algumas variações bruscas nas temperaturas coletadas e analisadas.

A Figura 3(b) ilustra a quantidade de mensagens enviadas por todos os nós sensores dentro de um intervalo de 30 segundos. Em outras palavras, esses valores indicam que para cada intervalo de 30 segundos, a quantidade de mensagens que cada cenário enviou. Ao analisarmos em conjunto as figuras 3(a) e 3(b), é possível notar que os intervalos dos eventos gerados estão diretamente relacionados aos intervalos de maior envio de mensagens para cada cenário. Um detalhe que deve ser ressaltado é que a aplicação *NAIVE* manteve-se com uma quantidade de envios aproximadamente constante durante todo o experimento. Por outro lado, as implementações do ESA alteraram os intervalos de sensoriamento e disseminação durante a ocorrência dos eventos, elevando assim a quantidade de mensagens enviadas e, conseqüentemente, os detalhes dos eventos enviados aos responsáveis. Em geral, o cenário *ESA_GA_ACS* se mostrou mais atuante que os demais, enviando mais mensagens quando necessário. O cenário *ESA_REGRAS*, por outro lado, se mostrou menos atuante, enviando menos detalhes sobre o evento. Os cenários *ESA_GA* e *ESA_ACS* apresentaram resultados intermediários. Como observação geral, ficou caracterizado que as implementações do ESA levaram a uma melhor monitoração dos eventos, e que algoritmos elaborados são importantes para que a auto-configuração seja efetiva.

A Figura 4(a) apresenta as curvas de mensagens acumuladas por cada cenário ao longo dos 10 minutos de execução dos experimentos. É possível observar variações nas inclinações das curvas, e relacionar esses fatos diretamente aos eventos causados pelos sensores virtuais (veja Figura 3(a)). Praticamente em sincronia com o evento, há um aumento gradual do número de mensagens recebidas em tal intervalo de tempo, com

exceção do cenário *NAIVE*, em que a inclinação continua constante ao longo do tempo. Vale destacar também que o cenário *ESA_GA_ACS* levou a um maior número de mensagens, notificando em mais detalhes a ocorrência dos eventos ao PA. Por outro lado, o cenário *ESA_REGRAS* enviou um número menor de mensagens, enquanto os cenários *ESA_GA* e *ESA_ACS* transmitiram valores intermediários. Mais uma vez, vale destacar a importância de algoritmos mais elaborados, que são mais sensíveis à detecção e disseminação dos eventos, enviando assim mais informações sobre os mesmos. Também é importante comentar que, ao final dos eventos, é notada uma redução na inclinação das curvas, mostrando que os cenários que implementam o ESA se readaptaram com o fim dos eventos.



(a) Mensagens acumuladas durante o experimento. (b) Tempo médio de notificação de uma mensagem ao PA.

Figura 4. Total de mensagens enviadas (acumulado) e Tempo médio de notificação do evento no PA.

Por fim, foi feita uma análise do tempo necessário para que a primeira notificação da ocorrência de um evento chegue ao PA da rede. O gráfico apresentado na Figura 4(b) mostra um tempo médio, com base em 3 rodadas de experimentos, que pode ser observado para cada um dos cenários. Para essa métrica, o cenário *ESA_ACS* se mostrou mais eficiente, entregando a primeira notificação do evento mais rapidamente que as outras estratégias. A estratégia mais elaborada *ESA_GA_ACS* obteve o segundo melhor resultado, seguida pelas estratégias *ESA_REGRAS*, *ESA_GA* e *NAIVE*. Como nos resultados anteriores, essa métrica demonstra a importância de se adotar um ESA em RSSFs com algoritmos elaborados.

5. Conclusão e Trabalhos Futuros

Neste trabalho, implementamos e avaliamos novas instâncias de um ESA, com o objetivo de monitorar eventos em uma RSSF. As soluções foram implementadas em um ambiente real com mais de cem nós sensores, que tornaram os resultados coesos e consistentes. Foram avaliadas métricas que indicam a qualidade da monitoração durante a ocorrência de evento na rede. Em geral, pode-se concluir que as implementações de algoritmos elaborados para o elemento sensor autônomo levam a uma monitoração mais rápida e detalhada dos eventos. Como trabalhos futuros, esperamos implementar novas instâncias com base no ESA genérico. Essas novas instâncias podem trazer conjuntos de algoritmos das mais diversas áreas que são relevantes na literatura e que podem causar impacto positivo na RSSF. Também pretendemos avaliar outras configurações de experimentos.

6. Agradecimento

Este trabalho contou com o apoio do CNPq (573.738/2008-4 INCT NAMITEC) e CAPES.

Referências

- Abdelaal, M. (2015). **Distributed Techniques for Energy Conservation in Wireless Sensor Networks**. In *DcSensorNets*.
- Braga, T. R., Silva, F. A., Ruiz, L. B., Nogueira, J. M. S., and Loureiro, A. A. (2006). **Design and Evaluation of an Autonomic Sensor Element**. In *1st Latin American Autonomic Computing Symposium, LAACS*.
- Braga, T. R. d. M. (2006). **Um Elemento Sensor Autônomo Para Redes De Sensores Sem Fio**. In *Dissertação de Mestrado, Departamento de Ciência da Computação da Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil*.
- Buratti, C., Conti, A., Dardari, D., and Verdone, R. (2009). **An overview on wireless sensor networks technology and evolution**. volume 9, pages 6869–6896. Molecular Diversity Preservation International.
- Doddavenkatappa, M., Chan, M. C., and Ananda, A. L. (2011). **Indriya: A low-cost, 3D wireless sensor network testbed**. In *Testbeds and Research Infrastructure. Development of Networks and Communities*, pages 302–316. Springer.
- Dorigo, M., Birattari, M., and Stützle, T. (2006). **Ant colony optimization**. In *Computational Intelligence Magazine, IEEE*, volume 1, pages 28–39. IEEE.
- Jabeur, N., Sahli, N., Zeadally, S., and Yasar, A.-U.-H. (2015). **Enabling Cyber Physical Systems with Wireless Sensor Networking Technologies, Multiagent System Paradigm, and Natural Ecosystems**. volume 2015. Hindawi Publishing Corporation.
- Lakshmi, C. B. and Mohan Rao, S. (2014). **Bio-inspired self-healing routing to improve lifetime of wireless sensor networks**. In *Communication and Network Technologies (ICCNT), 2014 International Conference on*, pages 134–138.
- Nguyen, T.-T., Shieh, C.-S., Horng, M.-F., and Dao, T.-K. (2014). **A Genetic Algorithm with Self-Configuration Chromosome for the Optimization of Wireless Sensor Networks**. In *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*.
- Peng, M., Xu, K., Yu, Q., Jiang, W., Leng, S., and Mao, Y. (2014). **Cellular automata self-organization algorithm for Wireless Sensor Network**. In *Communication Problem-Solving (ICCP), 2014 IEEE International Conference on*.
- Rawat, P., Singh, K. D., Chaouchi, H., and Bonnin, J. M. (2014). **Wireless sensor networks: a survey on recent developments and potential synergies**. volume 68, pages 1–48. Springer.
- Sadouq, Z., El Mabrouk, M., and Essaaidi, M. (2014). **Conserving energy in WSN through clustering and power control**. In *Information Science and Technology (CIST), 2014 Third IEEE International Colloquium in*.
- Whitley, D. (1994). **A genetic algorithm tutorial**. In *Statistics and computing*, volume 4, pages 65–85. Springer.

BBAware - A Context-Aware Mobile and Wearable Architecture for Monitoring Beta-Blocked Cardiac Patients

Alberto Vianna Dias da Silva¹, Vaninha Vieira^{1,2}

¹Department of Computer Science – Federal University of Bahia (UFBA) – Salvador – BA – Brazil

²Fraunhofer Project Center for Software and Systems Engineering – Technological Park – Salvador – BA – Brazil

{alberto.vianna, vaninha}@ufba.br

Abstract. *Beta-blockade drugs are still in use as treatment option to lower heart rate, to improve cardiac function, and to reduce cardiovascular events. Patients who use beta-blockers usually surpass a therapeutic test full of collateral effects to adapt their organisms. Furthermore, these patients with a baseline heart rate above 70 beats per minute have a significantly higher risk of all cardiovascular events. Context-aware healthcare field arises as an alternative to monitor patients constantly. This paper introduces the Beta-Blocked Aware (BBAware) project, a pervasive solution, that uses the concepts of ubiquitous healthcare in order to help patients under beta-blockade treatments.*

1. Introduction

Ischemic heart disease was the first cause of death in the world, between 2000 and 2012 [WHO 2014]. Several efforts have been made to decrease this rate in different areas.

A particular research area on cardiologic medicine is the study of beta-blockers after myocardial infarction [Freemantle et al. 1999]. Beta-blockers are a class of drugs that are among the current treatment options to lower heart rate, improve cardiac function, and reduce cardiovascular events in postmyocardial infarction patients and in heart failure patients [Mathers and Loncar 2006, Kjekshus 1986].

One of the main determinant of a cardiac problem is the heart rate [Andrews et al. 1993]. As beta-blockers decrease heart rate, in the beginning of the treatment patients should undergo a period of the treatment that will evaluate and balance the medication dose. This period is called therapeutic test and usually causes the patient tiredness, dizziness, and sometimes fainting [de Melo 2011]. These collateral effects could be prevented or minimized if the patient could be alerted if his/her heart rate is out of range.

Furthermore, some studies support the idea that patients with a baseline heart rate above 70 beats per minute (bpm) have a significantly higher risk of all cardiovascular events as compared to those with heart rate below 70 bpm. Heart rate reduction decreases the risk of coronary events in Coronary Artery Disease (CAD) patients with heart rate above 70 bpm [Diaz et al. 2005, Fox et al. 2008, Mathers and Loncar 2006]. Therefore, a continuous way to measure the patient heart rate and guarantee that his/her heart rate does not surpass this limit will help this kind of treatment.

Nowadays, patient monitoring by doctors is held in sporadic consultations where heart rate is measured. Perhaps, at this time, the patient heart rate is between the safety

limits. However, on other time of daily life, the value of heart rate may surpass the safety limits. Hence, if there was a way to monitor a patient, to guarantee that his/her heart rate is between the limits at most of the time, it would help patients and doctors. This thought leads to the research question: "How context-aware, mobile and wearable computing can help beta-blocked patients monitoring?"

With the evolution of smartphones and wearable devices, heart rate sensors have been embedded to them to increase the use of context in the development of mobile and wearable applications. Because of that, those devices could be used to help patients and doctors on the administration of beta-blockade treatment as an ubiquitous healthcare solution [Gelogo and Kim 2013].

Thus, this paper proposes a project called Beta-Blocked Aware (BBAware). The hypothesis of this research is: if wearable and mobile devices tailored with context-awareness could be used for monitoring beta-blocked patients then these patients will have a safer heart rate monitoring and doctors will have better information to define the treatment.

BBAware uses smartwatches and smartphones to help patients, the patient family, and the doctors on monitoring heart rate. Doctors can administrate the beta-blockade treatment on patients that have some cardiac problem that demands the reduction of heart rate having precise measurement from the sensor.

This paper is organized as follows. Section 2 describes some related works on heart rate measurement. Section 3 introduces the BBAware Project. Section 4 presents some examples of use. The conclusions and future work on this study are given in section 5.

2. Related Works

According to Murnane *et al.* [Murnane et al. 2015] there are many applications for health care purposes. The most popular ones are for physical activities and education. The second ones are for medical purposes. The focus on these related works is on monitoring patients solutions that uses context-awareness.

To the best of our knowledge, there is no solution or proposal for monitoring beta-blocked cardiac patients with or without the use of context. However, there are several projects for monitoring heart rate for health concerns that use context-aware computing.

One problem of the proposed works would be if the sensor embedded on the device is not precise enough to be used in medical area. Important studies that evaluate the performance of the heart rate sensor embedded on smartwatches have been published. The results showed in Phan *et al.* [Phan et al. 2015], Jovanov [Jovanov 2015], and Lemay *et al.* [Lemay et al. 2014] have approved the use of the sensor, the photoplethysmogram (PPG), as a reliable heart rate sensor.

Studies have used devices to measure heart rate for specific purposes. One of them is presented in Rocha *et al.* [Rocha et al. 2015]. They presented a heart rate monitor model. It uses a smartphone to synchronize health care data with the cloud. A similar study was presented by Sannino and De Pietro [Sannino and De Pietro 2011], a solution that monitors patients for a nuclear medicine department. These solutions for generic

patient monitoring differ from our solution, mainly on software adaptation from surpass safe limits from heart rate.

Another interesting study, in which the authors create a device is presented in Chigira *et al.* [Chigira et al. 2014]. The authors created a heart rate monitoring through the surface of a drinkware, a glass type drinkware that have PPG sensor to monitor heart rate during beverage consumption. However, this device is not indicated to use during the daily life, because is only connected to the body when the user is drinking something.

Wearable devices were used by specific health care solutions. One example is the study published by Rubin *et al.* [Rubin et al. 2015] that used a chest-worn device that measures heart rate, respiration rate, perspiration and skin temperature. They proposed a system that combines a wearable device with a smartphone to predict and react to approaching panic attacks. The study used one specific wearable device that was created for monitoring patient proposes. Our idea is to use devices that the user already have or can by at many general stores.

Other example is in Phan *et al.* [Phan et al. 2015]. They used a smartphone and a wearable device to measure sleep quality using the heart rate readings and accelerometer readings captured from the smartwatch. A solution that can be compared to our solution, however, it was created for other scenario and cannot be applied on beta-blocked patients.

To summarize, BBAware differs from the other projects because the solution adapts considering beta-blocked patient necessities. Its adapt on the use of sensors to notify doctors, the patient and patient family members that the patient heart rate is out of safety range. The idea is to have a long period monitoring using a wearable and mobile devices that the user already have or is easy to buy, like a smartwatch and a smartphone.

Some studies argues that the best way to monitor cardiac patients is using the electrocardiography (ECG) to evaluate heart rhythm [Crawford et al. 1999]. ECG is in use for decades, and is considered by doctors to be a better and more reliable sensor than the PPG used in BBAware. Popular smartwatches presented on the market does not have ECG sensors, only PPG.

Even with studies defending the use of PPG, as mentioned in this section, it is important to mention that BBAware is not a replacement for traditional ambulatory ECG monitoring solutions. BBAware is an alternative monitoring system that uses sensors that are related to daily life of the patients, reducing the traditional sensors intrusiveness.

3. BBAware Project

This research investigates in which ways mobile and wearable computing tailored with context-aware computing techniques can contribute on beta-blocked patient monitoring. As a motivating scenario, we consider the use of a smartphone and a smartwatch to help monitoring these specific patients.

BBAware is a software suite that unites a wearable application, a mobile application and an application server (web service). The Figure 1 illustrates how these devices are combined. The arrows presented in the figure represents the data flow between the devices.

In the doctor vision, he/she can register a patient using a smartphone, including

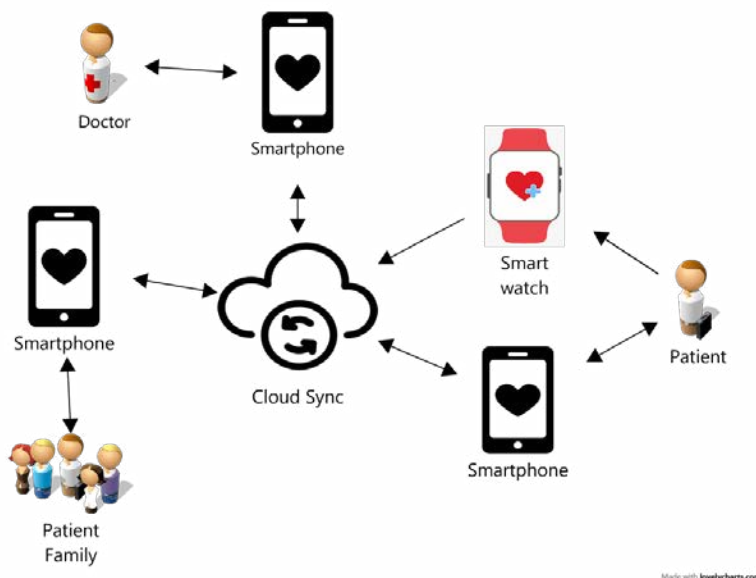


Figure 1. BBAware Overview.

patient safety range of heart rate, and synchronize within the cloud. The cloud presented in Figure 1 is an abstraction of a server application, an application that comprehends a reasoner and a knowledge data base, and a web service. The doctor has a list of patients that he/she is monitoring and can check a log from each patient. This log has data from user heart rate, time (date and hour) and if the patient had some problem, such as nausea, dizziness, and others.

Patient is monitored by smartwatch heart rate sensor. This information is synchronized with the web service. Further, patient and family members can check patient heart rate information on a smartphone application. The application of the patient connect patient with smartwatch device to inform that the data collected by a watch is from that specific patient.

3.1. Context Model and Rule Definition

In order to model a context-aware application, it is necessary to identify entities and contextual elements [Vieira et al. 2011]. The context model of BBAware is presented in Table 1. A contextual element can be static (does not change over time), dynamic (changes over time), explicit (informed by the user) and implicit (obtained without user interaction).

In the BBAware context model, one entity was identified, the patient that is monitored. The focus of this work is 4 contextual elements, maximum heart rate and minimum heart rate (safety limit range for the patient), current heart rate (patient heart rate monitoring), and "patient condition" (the patient can inform if he/she is feeling sick, dizzy, etc.). The others contextual elements are mapped for future work.

Rules are defined in a knowledge base to help reasoner processing. To define those rules, a tool, called Drools¹, was used. A rule definition is presented in Figure 2, where is possible to see a rule that verify if the maximum heart rate limit was surpassed. If this test

¹www.drools.org

Table 1. BBAware Context model

Entity	Contextual Element	Static	Dynamic	Explicit	Implicit	Context Source
Patient	Maximum Heart Rate	X		X	X	User profile or default value of 70 bpm
Patient	Minimum Heart Rate	X		X		User profile
Patient	Current Heart Rate		X		X	Heart rate provided by the smartwatch sensor
Patient	Patient Condition		X	X		Patient can inform if is feeling sick, dizzy, nausea, etc.
Patient	Activity		X		X	Data provided by the smartwatch accelerometer sensor, for example: to identify falls
Environment	Time		X		X	Time provided by the smartwatch sensor

rule is true, a notification is sent to the doctor, patient family members, and the patient. Hence, this rule will be invoked for each measured heart rate.

```

rule "Maximum Limit Surpassed"
  when
    p : PatientDataRule(
      heartRate > maximumHeartRateLimit,
      myHeartRate : heartRate )
  then
    p.sendNotification(p.MAXIMUM_LIMIT_SURPASSED);
  end
end

```

Figure 2. BBAware Rule Example.

3.2. Architecture

The BBAware architecture has 3 main parts: a remote server application with a web service for communication, a smartwatch application, and a smartphone application. A better view of these parts is in Figure 3.

The smartwatch application has the Dynamic Context Acquisition module, where current heart rate and time are obtained from the patient. This data is persisted locally inside a log database, until the device connect to internet to synchronize data with the web service.

The smartphone application has two main modules. The Context Acquisition module receive profile information about the patient from the doctor and the Notifier

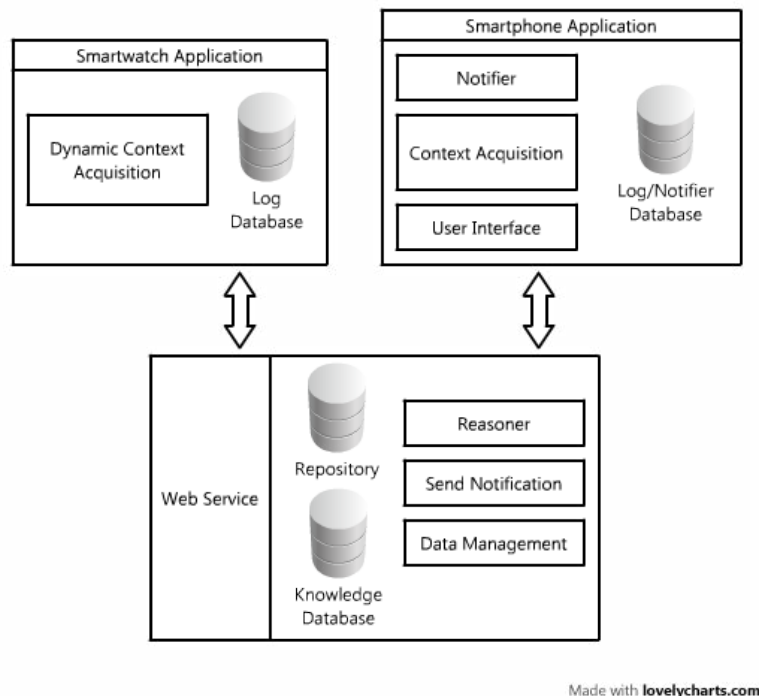


Figure 3. BBAware Architecture.

module that notifies patient, doctor and family members through software adaptation after analysis of patient heart rate and when patient explicit say that is not feeling well.

The server side contain a web service to synchronize data between smartphone and smartwatch applications. Moreover, the processing of the context rules is made by the Reasoner module. A history repository, a knowledge base and a Data Management module is also located at the remote server.

4. Example of Use

In order to illustrate the use of BBAware, this section introduces a scenario of use. Consider a patient, called Robert Smith, that just has had a heart attack. He is recovering and just left the hospital. He is using beta-blockers and he is undergoing therapeutic test. Consider another patient, called Daniel Clark, that had a heart attack several years ago and also uses beta-blockers.

Considering the ideas presented in Diaz *et al.* [Diaz et al. 2005], Fox *et al.* [Fox et al. 2008], and Mathers and Loncar [Mathers and Loncar 2006], Daniel Clark should maintain his heart rate below 70 bpm. Furthermore, his heart rate cannot be too short, or he will feel dizzy, nauseous, etc. To monitor this patient and guarantee that his heart rate is between safety limits, BBAware can be used. If heart rate is out of range, the system will adapt and notify patient family members, the patient, and his doctor. If the doctor deems necessary, he can communicate to the patient and can conduct him for an evaluation.

The other patient, Robert Smith, is undergoing therapeutic test, he is at home resting and is using BBAware to monitor his heart rate. His next evaluation with the

doctor is in next week. However, today he felt really dizzy when he went to the toilet and interact using BBAware to tell that he did not feel well. The system will adapt and notify patient family members, the patient, and his doctor. Then anyone can check the moment that the patient interact with BBAware and check his heart rate during that phase. If the doctor deems necessary, he can communicate to the patient and can conduct him for an evaluation and not to wait for next week. Family members can contact the patient to know if patient need help.

At any time, the patient, family members and the doctor can check the patient heart rate associated with the time. An example of Robert Smith heart rate is presented in Figure 4, where is possible to see a list of the patient heart rate and one interaction. The interaction has a different image (exclamation) that can be related to the low heart rate before the interaction (39).

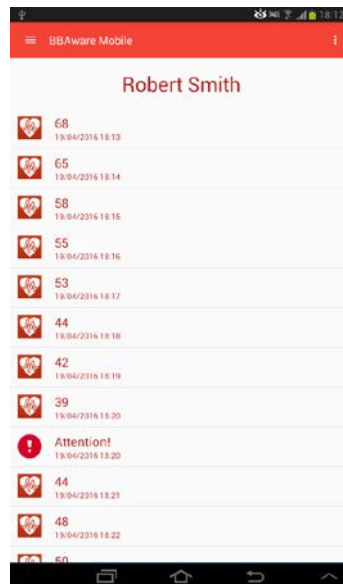


Figure 4. Heart rate list screen inside BBAware mobile application.

Is important to mention that if the doctor have many patient, BBAware could be intrusive, because the mobile application would send him/her many notification about different patients. To avoid this situation BBAware has a notifier module that can be managed by the doctor. Thus, the doctor can choose if he/she want to receive notifications or not, and can choose how many times he/she want to receive them. For example, a doctor can receive notifications once a day.

5. Conclusions and Future Work

This paper presented a smartphone and smartwatch integrated solution for monitoring patients that are under treatment using beta-blockade drugs. For this, it was developed a study on monitoring beta-blocked patients using a smartphone and a smartwatch and it was proposed BBAware, a software suite that unites a wearable application, a mobile application and a remote server application. A context model and architecture was presented to demonstrate our solution.

As ongoing work we are planning the following (i) the execution of an experiment

with real patients that are undergoing treatment using beta-blockers; (ii) the evolution and implementation of the modules described for the proposed architecture.

As opportunity for future work we indicate: (i) the use of machine learning techniques to learn, identify, and adapt to harmful changes on patient heart rate; (ii) the use of others contextual elements to identify, for instance, if a patient fell and adapt to this scenario.

References

- Andrews, T. C., Fenton, T., Toyosaki, N., Glasser, S. P., Young, P. M., MacCallum, G., Gibson, R. S., Shook, T. L., and Stone, P. H. (1993). Subsets of ambulatory myocardial ischemia based on heart rate activity. Circadian distribution and response to anti-ischemic medication. The Angina and Silent Ischemia Study Group (ASIS). *Circulation*, 88(1):92–100.
- Chigira, H., Ihara, M., Kobayashi, M., Tanaka, A., and Tanaka, T. (2014). Heart rate monitoring through the surface of a drinkware. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 685–689.
- Crawford, M. H., Bernstein, S. J., Deedwania, P. C., Dimarco, J. P., Ferrick, K. J., Garson, A., Green, L. A., Greene, H. L., Silka, M. J., Stone, P. H., Tracy, C. M., Gibbons, R. J., Alpert, J. S., Eagle, K. A., Gardner, T. J., Gregoratos, G., Russell, R. O., Ryan, T. J., Smith, S. C., and Introduction, I. (1999). *Acc/aha guidelines for ambulatory electrocardiography: executive summary and recommendations: a report of the american college of cardiology/american heart association task force on practice guidelines (committee to revise the guidelines for ambulatory electrocardiography)*. *Circulation*, (100):886–893.
- de Melo, D. S. B. (2011). *Impacto do uso rápido dos betabloqueadores sobre a mortalidade e remodelamento ventricular na insuficiência cardíaca avançada*. Doctorate, Faculdade de Medicina da Universidade de São Paulo.
- Diaz, A., Bourassa, M. G., Guertin, M.-C., and Tardif, J.-C. (2005). Long-term prognostic value of resting heart rate in patients with suspected or proven coronary artery disease. *European heart journal*, 26(10):967–74.
- Fox, K., Ford, I., Steg, P. G., Tendera, M., and Ferrari, R. (2008). Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *The Lancet*, 372(9641):807–816.
- Freemantle, N., Cleland, J., Young, P., Mason, J., and Harrison, J. (1999). Beta Blockade after myocardial infarction: systematic review and meta regression analysis. *BMJ (Clinical research ed.)*, 318(7200):1730–7.
- Gelogo, Y. E. and Kim, H.-K. (2013). Unified Ubiquitous Healthcare System Architecture with Collaborative Model. *International Journal of Multimedia and Ubiquitous Engineering*, 8(3):239–244.
- Jovanov, E. (2015). Preliminary analysis of the use of smartwatches for longitudinal health monitoring. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 865–868.

- Kjekshus, J. K. (1986). Importance of heart rate in determining beta-blocker efficacy in acute and long-term acute myocardial infarction intervention trials. *The American Journal of Cardiology*, 57(12).
- Lemay, M., Bertschi, M., Sola, J., Renevey, P., Parak, J., and Korhonen, I. (2014). Application of optical heart rate monitoring. In *Wearable Sensors: Fundamentals, Implementation and Applications*, pages 105–129. Academic Press, Oxford.
- Mathers, C. D. and Loncar, D. (2006). The Importance of Heart Rate in Coronary Artery Disease. *PLoS Medicine*, 3(11):2011–2030.
- Murnane, E. L., Huffaker, D., and Kossinets, G. (2015). Mobile Health Apps : Adoption , Adherence , and Abandonment. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 261–264.
- Phan, D., Siong, L. Y., Pathirana, P. N., and Seneviratne, A. (2015). Smartwatch: Performance evaluation for long-term heart rate monitoring. In *International Symposium on Bioelectronics and Bioinformatics (ISBB)*, pages 144–147.
- Rocha, C. C. L., da Costa, C. A., and Righi, R. d. R. (2015). Um modelo para monitoramento de sinais vitais do coração baseado em ciência da situação e computação ubíqua. In *VII Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP)*, Pernambuco.
- Rubin, J., Eldardiry, H., Abreu, R., Ahern, S., Du, H., Pattekar, A., and Bobrow, D. G. (2015). Towards a Mobile and Wearable System for Predicting Panic Attacks. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 529–533.
- Sannino, G. and De Pietro, G. (2011). An evolved ehealth monitoring system for a nuclear medicine department. In *Proceedings - 4th International Conference on Developments in eSystems Engineering, DeSE 2011*, pages 3–6.
- Vieira, V., Tedesco, P., and Salgado, A. C. (2011). Designing context-sensitive systems: An integrated approach. *Expert Systems with Applications*, 38(2):1119–1138.
- WHO, W. H. O. (2014). The top 10 causes of death. URL: <http://www.who.int/mediacentre/factsheets/fs310/en/>.

Black Box: Uma Aplicação de Gerenciamento e Entrega Tolerante a Atrasos e Desconexões de Vídeos Para Veículos

Charles T. B. Garrocho¹, Maurício J. da Silva¹, Ricardo A. R. Oliveira¹

¹ Laboratório IMóbilis

Departamento de Computação – DECOM

Universidade Federal de Ouro Preto (UFOP) – Ouro Preto, MG – Brazil

{ctgarrocho,badriciobq,rrabelo}@gmail.com

Abstract. *Develop, test and experience network applications for vehicles has become a major challenge for the scientific community, since most applications are evaluated in simulated form. Most of these applications are focused on the processing of information in real time. However, by mobility issues, it is virtually impossible distribute contents such as videos, because they are heavy, and thus unworkable in services such as mobile telephony, by resource constraints. To overcome these limitations, it's proposed a vehicular application which performs video recording, fragmenting them into smaller pieces and lower quality to be transmitted in encounters between the vehicles through an opportunistic network until they arrive at their destination, forming so a delay and disruption tolerant network. Experiments have shown that the application has successfully delivering video to the destination through the network created.*

Resumo. *Desenvolver, testar e experimentar aplicações de rede para veículos se tornou um grande desafio para a comunidade científica, uma vez que, a maioria das aplicações são avaliadas de forma simulada. A maioria dessas aplicações são focadas no processamento de informações em tempo real. Entretanto, por questões de mobilidade, é praticamente impossível distribuir conteúdos como vídeos, por serem pesados, e assim, inviáveis em serviços como telefonia móvel, por limitações de recursos. Para superar essas limitações, é proposto uma aplicação veicular que realiza o gerenciamento da gravação de vídeos, fragmentando-os em pedaços menores e com menor qualidade para serem transmitidos em encontros entre os veículos através de uma rede oportunista até que cheguem ao seu destino, formando assim uma rede tolerante a atrasos e desconexões. Experimentos demonstraram que a aplicação conseguiu com sucesso a entrega dos vídeos ao destino através da rede criada.*

Introdução

Hoje em dia, as redes de comunicação sem fio tornaram-se essenciais no cotidiano de toda sociedade [Castells et al. 2009]. As pessoas podem se conectar a redes de dados, acessar e disponibilizar conteúdos em praticamente qualquer lugar e a qualquer momento, por meio de diferentes dispositivos e tecnologias de comunicação [Ma et al. 2014]. Isso se tornou possível principalmente graças ao avanço da rede celular (por exemplo, 3G (*Third Generation of Mobile Telecommunications Technology*), e LTE (*Long Term Evolution*)), com mais locais com acessos através de torres de transmissão e o aumento da capacidade de transmissão entre os dispositivos e as torres [Dahlman et al. 2013].

A pesar da evolução da rede celular, ela ainda não está disponível em muitos lugares e nem mesmo presente em grande parte da população de países com índice de desenvolvimento alto [Rainie 2010]. A rede celular também apresenta falhas de infraestrutura de comunicação, parciais ou totais causadas por desastres naturais [Skouloudi and Karsberg 2014], censura do governo [Deibert et al. 2011] ou por até uma interrupção de Internet [Chen 2011]. Além disso, as operadoras de telecomunicações tradicionais hesitam em investir em tais áreas por não gerarem lucro, uma vez que há riscos e custos elevados [Ims 2012].

A fim de superar essa falha e falta de conectividade para permitir as comunicações entre os dispositivos e aplicações em cenários perturbadores foi proposto as redes DTNs (*Delay and Disruption Tolerant Networks*) [Fall 2003]. As DTNs são um tipo de rede com conexão intermitente na qual um caminho fim-a-fim entre os nós que estão comunicando pode nunca existir [Fall and Farrell 2008]. Este tipo de abordagem de comunicação é feita através de oportunidades de encontros entre os nós que trocam dados entre si até que a mensagem chegue ao destinatário [Galati 2010].

As redes DTNs evoluirão muito nos últimos anos produzindo diversos protocolos e algoritmos de comunicação veicular [Vasilakos et al. 2011]. Esses trabalhos surgiram por que o veículo se tornou um lugar onde as pessoas ficam muito tempo durante seu dia-a-dia [Barfield and Dingus 2014], e assim, fornecer informação ao usuário no veículo se tornou um grande desafio, que tem sido explorado através dos SIT (*Sistemas Inteligentes de Transporte*) [Dimitrakopoulos and Demestichas 2010]. Entretanto, a maioria dos SIT são analisados através de simulações, e poucas avaliações reais através de aplicações práticas foram executadas em ambientes veiculares na literatura [Pereira et al. 2012]. Assim, desenvolver, testar e avaliar aplicações reais para veículos se tornou um desafio para a comunidade científica [Karagiannis et al. 2011].

Com o objetivo de preencher essa lacuna dos SIT, este trabalho tem como principal contribuição uma aplicação veicular, o *Black Box*, que realiza a gravação dos vídeos no veículo, fragmentando essa gravação em vários pedaços de vídeos com baixa qualidade com o objetivo de gerar arquivos de vídeos menores que possibilitem serem trocados entre veículos em curtos espaços de tempo. Essa troca de vídeos é realizada em encontros não programados entre os veículos através de uma arquitetura transparente [Garrocho et al. 2015] de rede oportunista.

Resultados de experimentos realizados na aplicação *Black Box* demonstraram que a arquitetura de rede oportunista utilizada se comporta melhor em locais onde há maior quantidade de veículos (por exemplo, vias congestionadas e sinais de trânsito). Além disso, os resultados demonstraram que a medida que a quantidade de veículos aumenta a taxa de transmissão diminui, e assim, a fragmentação dos vídeos é essencial para que a rede não seja sobrecarregada com a entrega dos vídeos.

O restante do artigo está organizado da seguinte forma: na Seção 2 são apresentados os trabalhos relacionados. Na Seção 3 é apresentado uma visão geral da aplicação *Black Box*. Na Seção 4 é feita uma avaliação do aplicação, sendo que na Subseção 4.1 são apresentados os cenários e métricas avaliadas, e na Subseção 4.2 são apresentados os resultados dos experimentos. Finalmente na Seção 5 são apresentadas as conclusões e os trabalhos futuros.

Trabalhos Relacionados

Uma aplicação é uma coleção de instruções que descrevem uma tarefa a ser realizada por um dispositivo [Lloyd 2012]. Normalmente ela é desenvolvida para uma determinada tarefa gerando resultados esperados. Atualmente existem diversas aplicações que permitem a simulação de ambiente de redes DTN [Kaur and Malhotra 2015]. Essas simulações podem garantir um menor custo da avaliação de projetos e criar ambientes que são impossíveis por falta de tecnologia. Porém, os resultados de experimentos em ambientes reais continuam tendo maior impacto que ambientes simulados [Rutten et al. 2012].

Recentemente, em redes DTN, os protocolos de roteamentos foram mais abordados que aplicações práticas. Um grande número de algoritmos de encaminhamento, sem infraestruturas foram desenvolvidos. Os autores desses trabalhos baseiam-se que os dispositivos possuem a capacidade de troca e encaminhamento de mensagens através de uma rede Wireless Ad-Hoc, utilizando tecnologias sem fio [Wei et al. 2014]. Entretanto, existem na literatura algumas soluções de aplicações cuja meta é explorar as características inerentes as redes DTNs para a transmissão de mensagens, permitindo uma entrega de mensagens mesmo em falta de infraestrutura de comunicação [Vasilakos et al. 2011].

Um dos precursores de aplicações DTN foi o projeto ZebraNet [Juang et al. 2002]. Este sistema consiste em coleiras de rastreamento transportadas por zebras no âmbito de reservas da África. Os colares operam como forma de conexão oportunista para transmitir dados de volta para os pesquisadores. Isso significa que quando duas zebras se encontram, as coleiras registram as informações de encontro e trocam os dados armazenados em suas memórias locais. Através de oportunidades de encontro intermitentes, informação é gravada e depois difundida a essas zebras que transportam coleiras de rastreamento.

O sistema DakNet [Pentland et al. 2004] foi um projeto desenvolvido para criação de uma rede Ad-Hoc tolerante a atrasos e desconexões. O sistema DakNet consiste de quiosques com Wi-Fi habilitado em aldeias, e de dispositivos móveis que ao se aproximarem de quiosques de uma aldeia, detecta automaticamente as conexões sem fio e, em seguida, realiza a transmissão e recebimento de dados com os quiosques. Quando este dispositivo móvel entra no alcance de comunicação de um ponto de acesso à Internet, ele irá sincronizar automaticamente as mensagens dos quiosques das aldeias com à Internet.

Cartel [Hull et al. 2006] é um sistema de comunicação projetado para coletar, transmitir e visualizar dados a partir de sensores localizados em unidades móveis, como carros. Um nó CarTel (computador acoplado a um conjunto de sensores) coleta e processa a mensagem de sensores localizados em automóveis, e em seguida, entrega esses dados para um portal central (que funciona como ponto de controle e configuração para o sistema), utilizando CafNet. Em outras palavras, nós CarTel dependem de conectividade sem fio oportunista (por exemplo, Wi-Fi, Bluetooth) para se comunicar com um portal.

BikeNet [Eisenman et al. 2009] é um sistema de mapeamento das experiências de ciclistas. Ele usa uma série de sensores embarcados em uma bicicleta para coletar dados sobre os passeios de ciclistas. BikeNet encontra pontos de acesso sem fio de forma oportunista para troca de conteúdos, e aproveita o canal de dados celular dos ciclistas para comunicação em tempo real, se necessário. BikeNet também fornece um portal Web para o ciclista acessar seus dados, e permitir o compartilhamento desses dados (por exemplo, rotas de ciclismo favoritas, dados de poluição) com grupos interessados em ciclismo.

Como pode ser observado, existem poucas aplicações práticas voltadas ao ambiente de redes DTN. O principal objetivo das aplicações é permitir uma entrega tolerante a atrasos e desconexões de um conteúdo a um destino. O sistema DakNet [Pentland et al. 2004] é o único que permite um acesso a internet sem ser em tempo real. Os trabalhos ZebraNet [Juang et al. 2002] e BikeNet [Eisenman et al. 2009] estão focados na difusão de informações sobre zebras e ciclistas. Já o sistema Cartel [Hull et al. 2006] está voltado para o ambiente veicular, e é neste caso, o sistema mais parecido com o *Black Box*.

Cartel [Hull et al. 2006] está mais focado na distribuição de informações de tempo real dos veículos nas estradas, através de informações fornecidas pelos seus sensores. Essas informações podem ser utilizadas para evitar congestionamentos e acidentes nas estradas. Entretanto, nem todos os veículos possuem tais sensores. Já veículos pessoais, policiais e de transporte, em sua maioria já tem integrado câmeras que realizam filmagens tanto para dentro quanto para fora do veículo. A gravação de vídeo nesses veículos são utilizadas para monitoramento e detecção de movimentos de veículos em tempo real nas estradas [Jazayeri et al. 2011]. Porém os vídeos gravados por essas câmeras não só são utilizados no momento em que o veículo está em movimento, como também para uma análise posterior da conduta do motorista e dos passageiros, ou para análise de uma situação que tenha ocorrido no veículo ou próximo a ele [Noble and Alpert 2008]. Com o objetivo de auxiliar no gerenciamento de vídeos gravados nos veículos e na entrega posterior desses vídeos para análise, foi desenvolvido e avaliado o sistema *Black Box*.

O Sistema Black Box

Esta aplicação gerencia a gravação de vídeos em partes e realiza a entrega de vídeos solicitados do veículo para um servidor através de uma rede DTN formada através de uma arquitetura transparente [Garrocho et al. 2015] que fornece conectividade oportunista entre os veículos.

Organização e Implementação

A Figura 1 ilustra a organização dos componentes no veículo. Cada veículo é equipado de uma placa *PandaBoardES*¹ e uma câmera que realiza o monitoramento do motorista.

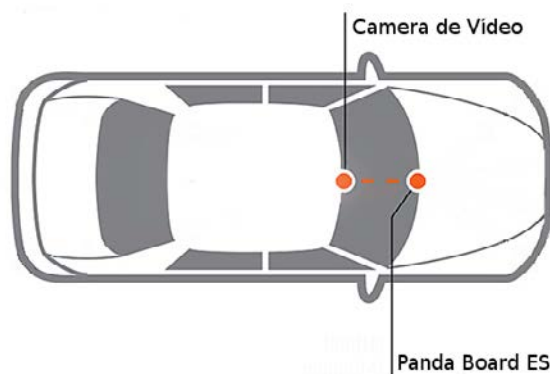


Figura 1. Organização dos componentes no veículo.

¹<http://pandaboard.org/content/pandaboard-es>

A aplicação *Black Box* foi desenvolvida sobre o SO (*Sistema Operacional*) Ubuntu 14.04, sendo utilizado uma versão totalmente customizada, que satisfaça apenas as necessidades da aplicação. A aplicação foi implementada para executar como um serviço para o SO, sendo assim, ele executa todas as suas funções em segundo plano.

Funcionamento

Quando a aplicação *Black Box* é executada na inicialização do SO móvel da placa *PandaBoardES*, é realizada uma operação de inicialização da arquitetura transparente de rede oportunista. Após isso, a aplicação executa diversas tarefas simultaneamente, realizando a gravação de vídeos da câmera instalada no veículo, transmissão dos vídeos através da rede oportunista, e gerenciamento do espaço disponível em disco apagando vídeos antigos.

A aplicação é dividida em duas *threads*, sendo que a primeira *thread* realiza a gravação dos vídeos da câmera do veículo de tempo em tempo em uma pasta do sistema. O intervalo de tempo de gravação dos vídeos é definido em um arquivo de configuração, podendo assim, ser modificado a qualquer momento. O nome de identificação dos vídeos gravados na pasta do sistema é definido através do endereço MAC da placa juntamente com a data e hora atual do sistema, criando assim, uma identificação única.

A segunda *thread* realiza o gerenciamento do espaço disponível em disco do sistema, baseado em um limite de espaço definido em um arquivo de configuração, sendo possível sua modificação. Para realizar esse gerenciamento, a aplicação *Black Box* verifica que se caso o limite máximo exceda, o vídeo mais antigo é removido da pasta. Esse processo é executado de minuto em minuto através da ferramenta *crontab*², com o objetivo de sempre estar com um espaço disponível no sistema.

Quando o veículo está desligado ou em movimento, o sistema permanece ativo, e sistematicamente tenta trocar vídeos através de contatos oportunistas. Quando uma rede é estabelecida, e um vídeo é solicitado, conforme é ilustrado na Figura 2(a), *Black Box* automaticamente distribui o vídeo entre todos os veículos próximos.

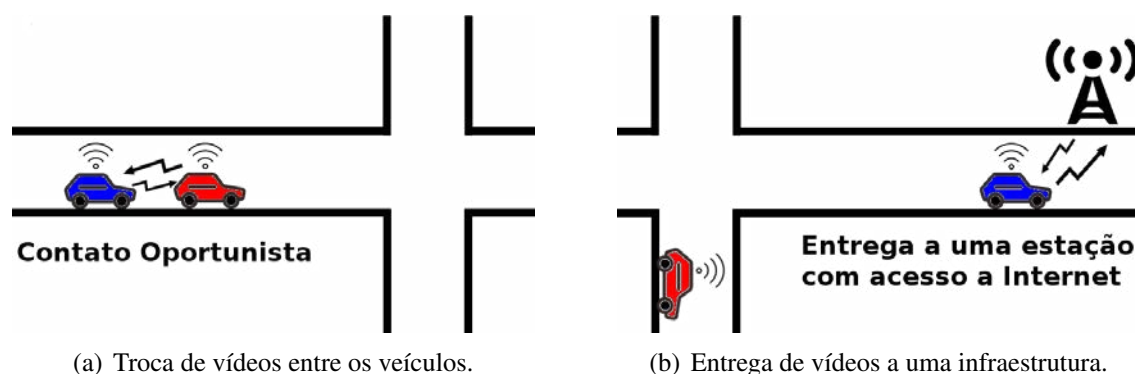


Figura 2. Funcionamento da rede DTN criada pela aplicação *Black Box*.

Quando um veículo se encontra com uma infraestrutura instalada na via, conforme é ilustrado na Figura 2(b), a aplicação *Black Box* entrega todos os vídeos armazenados no dispositivo e que estão sendo solicitados para a infraestrutura. Assim, a infraestrutura, com conexão a internet, trata de entregar esses vídeos ao servidor destino.

²<http://www.devin.com.br/crontab/>

Avaliação do Black Box

Nesta Seção são apresentados os cenários e métricas da aplicação *Black Box* para avaliação da rede DTN.

Cenários e Métricas

Em todos os experimentos que envolvem a transferência de arquivos no *Black Box*, os arquivos possuíam o tamanho de 1MB. Esse tamanho representava o tamanho dos vídeos gravados nos veículos. Eles possuíam esse tamanho, por serem gravados em pequenas faixas de tempo, estarem em preto e branco, e terem uma resolução menor.

Foram realizados experimentos para avaliar o tempo de formação topológica de rede, a taxa de transmissão, e o tempo de entrega de arquivo tolerante a atrasos e desconexões. Os experimentos foram realizados em laboratório, de maneira controlada, não sendo considerado interferências e movimentos.

Os dados dos três experimentos foram obtidos realizando cálculos na própria aplicação *Black Box*, sendo considerado um intervalo de confiança de 95%. Foi utilizado o padrão IEEE 802.11g, quadros de controle RTS/CTS e potência Tx padrão do dispositivo de 20 dBm.

No primeiro experimento, foi medido o tempo de associação topológica de rede entre os dispositivos. Cada dispositivo obtém o tempo através do Log da aplicação Linux, que registra quando o dispositivo começa o processo de escaneamento de redes, e quando o dispositivo estabeleceu uma conexão. Cada dispositivo envia seu tempo ao dispositivo AP (*Ponto de Acesso*). Ao final o dispositivo AP faz cálculo com todos os tempo (inclusive o seu). Este experimento é importante para avaliar o impacto que a quantidade de dispositivos promove no tempo de formação da rede de comunicação.

No segundo experimento, a aplicação foi alterada de modo que 4 dispositivos requisitassem um arquivo de 1MB a um dispositivo. Foi medido a taxa de transmissão de pacotes entre os dispositivos através do cálculo dos tempos e do tamanho do arquivo. O principal objetivo deste experimento é avaliar como a rede de comunicação se comporta quando um dispositivo realiza a transmissão de um arquivo para vários dispositivos (quantidades diferentes) ao mesmo tempo.

No terceiro experimento, foi avaliado o tempo de entrega de um arquivo de 1MB em uma rede DTN. Cada operação (envio e recebimento) de arquivos juntamente com o tempo que a operação foi efetivada são armazenados em um Log para que posteriormente pudesse ser lidos e gerados as estatísticas de tempos de entrega dos arquivos. As estatísticas extraídas foram calculadas através da soma do tempo médio que cada arquivo (de cada dispositivo) teve para chegar em cada dispositivo participante da rede de comunicação, sendo considerado dois cenários, o primeiro de três dispositivos, e o segundo com quatro dispositivos.

Nos dois primeiros experimentos, os dispositivos ficaram sobre uma mesa com uma média de três metros de distância. Foram utilizadas 5 placas *PandaBoardES*, e todos os experimentos foram executados 30 vezes em cada dispositivo simultaneamente. Já no terceiro experimento as placas estavam separadas por um distância de 25 metros e utilizadas três a quatro, e os testes foram executados quatro vezes em cada dispositivo simultaneamente.

Resultados

Nesta Seção são apresentados os resultados dos experimentos realizados no *Black Box*. A Figura 3 ilustra os resultados do primeiro experimento. Pode-se observar que, quando há poucos dispositivos, o tempo de formação de rede é grande, e sua faixa de erro também. Porém, quando a quantidade de dispositivos aumenta, o tempo de formação de rede diminui e a taxa de erro também. Isso acontece por que quando há poucos dispositivos, a probabilidade dos dispositivos ficarem ao mesmo tempo escaneando ou se tornando ponto de acesso é maior do que quando há uma quantidade maior de dispositivos.

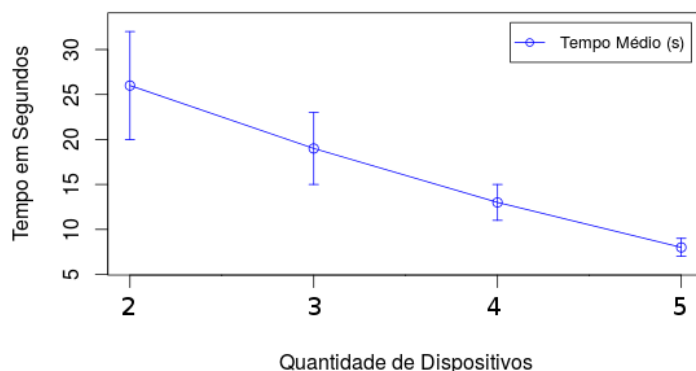


Figura 3. Tempo de formação da rede de comunicação.

A partir dos resultados do primeiro experimento, é possível concluir que o tempo de formação de rede, não é viável para comunicação com os veículos em movimento. Entretanto, esse sistema pode ser aplicado em locais onde os veículos se encontram e permanecem parados por alguns instantes de tempo (por exemplo, em um sinal de trânsito, ou em um estacionamento).

No segundo experimento, pode-se observar na Figura 4, que a quantidade de dispositivos influencia na taxa de transmissão, onde a taxa diminui de acordo que a quantidade de dispositivos que recebem um arquivo aumenta. Isso acontece por que o dispositivo servidor que está cuidando de transmitir o arquivo, tem um trabalho maior com quantidade maiores de dispositivos, e assim, a banda de conexão desse dispositivo está mais ocupada com várias conexões simultâneas e mais pacotes para serem processados.

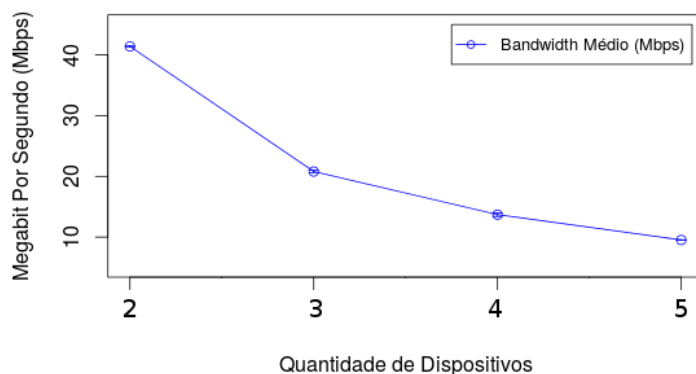


Figura 4. Taxa de transmissão de pacotes.

A partir dos resultados apresentados na Figura 4, também é possível verificar que a taxa ficou menor que o limite do padrão 802.11g e variou pouco. Isso aconteceu por causa da distância entre os dispositivos e o tamanho de arquivos utilizados em cada dispositivo. Variou menos por causa do tamanho do arquivo utilizado, no caso 1MB na aplicação. Já a taxa ficou menor por causa da distância entre os dispositivos, sendo de 15 metros.

No terceiro experimento, pode-se observar na Figura 5, que o que mais impactou no tempo de entrega de arquivos tolerante a atrasos e desconexões foi a distância entre os dispositivos. É possível visualizar que tanto para três quanto para quatro dispositivos, o tempo de entrega dos arquivos dos dispositivos que estavam nas extremidades é maior do que o tempo dos arquivos dos dispositivos que estavam no meio, servindo como ponto de comunicação entre os dispositivos das extremidades.

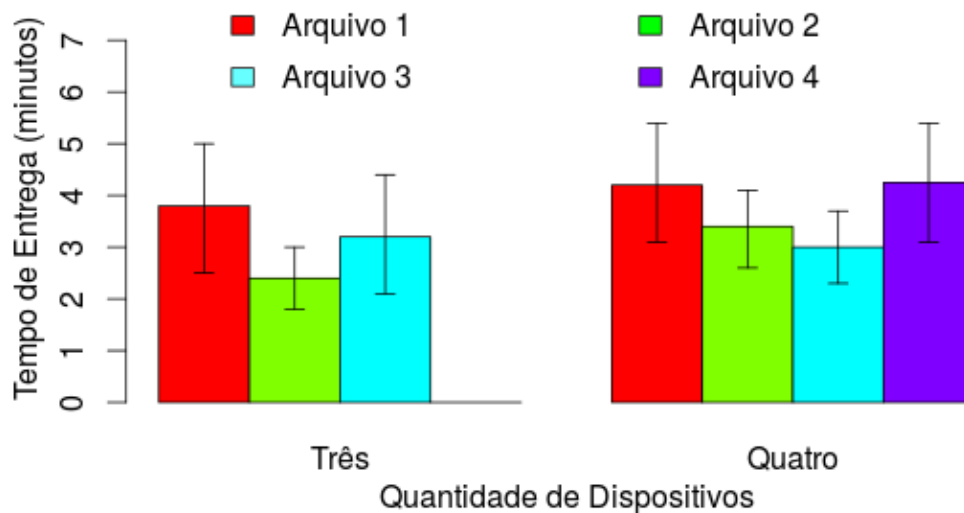


Figura 5. Tempo de entrega de arquivos em uma rede DTN.

A explicação dessa diferença de tempo deve-se principalmente ao fato de que os dispositivos que estão no meio se comunicam e entregam seus arquivos com um tempo menor aos dispositivos que estão próximos a eles, nesse caso, os dispositivos que estão na extremidade. O tempo de entrega dos arquivos dos dispositivos das extremidades ficou maior pelo fato de que para seus arquivos serem entregues de uma extremidade a outra, deve passar por um (três dispositivos) ou mais (quatro ou mais dispositivos) pontes de comunicação, e uma (três dispositivos) ou mais (quatro ou mais dispositivos) redes devem ser criadas para que isso aconteça.

Na Figura 5 também é possível visualizar uma taxa de erro grande no tempo de entrega de arquivos, principalmente nos tempos de entrega dos dispositivos das extremidades. Isso acontece por causa das redes formadas pelo middleware Net-Opp. Como o sistema baseia-se em tempos aleatórios para pesquisar e formar as redes, em alguns momentos, quando não há arquivos para serem transmitidos, o dispositivo não consegue encontrar um segundo dispositivo além do último que ele se comunicou, e ele acaba voltando a se comunicar com o primeiro, afetando assim, o tempo de entrega de seus arquivos e dos arquivos do primeiro dispositivo conectado à ele.

Conclusão e Trabalhos Futuros

Neste artigo, foi apresentado o *Black Box*, uma aplicação que realiza o gerenciamento eficiente de vídeos em veículos e permite a entrega de vídeos solicitados a um servidor através de uma rede DTN criada pela aplicação.

Os resultados apresentados neste artigo, mostraram em experiências reais a eficiência de *Black Box* na formação de rede e na troca de conteúdos entre os veículos, sendo escalável e promovendo uma comunicação transparente. Mostrando-se viável sua aplicação em locais onde os veículos se encontram e permanecem parados por alguns instantes de tempo (por exemplo, em um sinal de trânsito, ou em um estacionamento).

Como trabalhos futuros, além do aprimoramento da aplicação *Black Box*, pretende-se estender os estudos, empregando o uso da aplicação em diferentes cenários e maiores quantidades de dispositivos, com o propósito de avaliar mais profundamente o comportamento da aplicação nessa estrutura de rede.

Referências

- Barfield, W. and Dingus, T. A. (2014). *Human factors in intelligent transportation systems*. Psychology Press.
- Castells, M., Fernandez-Ardevol, M., Qiu, J. L., and Sey, A. (2009). *Mobile communication and society: A global perspective*. Mit Press.
- Chen, T. M. (2011). Governments and the executive”internet kill switch”[editor’s note]. *Network, IEEE*, 25(2):2–3.
- Dahlman, E., Parkvall, S., and Skold, J. (2013). *4G: LTE/LTE-advanced for mobile broadband*. Academic press.
- Deibert, R., Palfrey, J., Rohozinski, R., and Zittrain, J. (2011). *Access contested: security, identity, and resistance in Asian cyberspace*. MIT Press.
- Dimitrakopoulos, G. and Demestichas, P. (2010). Intelligent transportation systems. *Vehicular Technology Magazine, IEEE*, 5(1):77–84.
- Eisenman, S. B., Miluzzo, E., Lane, N. D., Peterson, R. A., Ahn, G.-S., and Campbell, A. T. (2009). Bikenet: A mobile sensing system for cyclist experience mapping. *ACM Transactions on Sensor Networks (TOSN)*, 6(1):6.
- Fall, K. (2003). A delay-tolerant network architecture for challenged internets. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 27–34. ACM.
- Fall, K. and Farrell, S. (2008). Dtn: an architectural retrospective. *Selected Areas in Communications, IEEE Journal on*, 26(5):828–836.
- Galati, A. (2010). *Delay tolerant network*. Germany: LAP Lambert Academic Publishing.
- Garrocho, C. T. B., Silva, M. J., and Oliveira, R. A. R. (2015). Transparent sharing architecture of content between mobile devices in opportunistic networks. In *Proceedings of the annual conference on Brazilian Symposium on Information Systems: Information Systems: A Computer Socio-Technical Perspective-Volume 1*, page 48. Brazilian Computer Society.

- Hull, B., Bychkovsky, V., Zhang, Y., Chen, K., Goraczko, M., Miu, A., Shih, E., Balakrishnan, H., and Madden, S. (2006). Cartel: a distributed mobile sensor computing system. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, pages 125–138. ACM.
- Ims, L. A. (2012). *Broadband access networks: introduction strategies and techno-economic evaluation*. Springer Science & Business Media.
- Jazayeri, A., Cai, H., Zheng, J. Y., and Tuceryan, M. (2011). Vehicle detection and tracking in car video based on motion model. *Intelligent Transportation Systems, IEEE Transactions on*, 12(2):583–595.
- Juang, P., Oki, H., Wang, Y., Martonosi, M., Peh, L. S., and Rubenstein, D. (2002). Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. In *ACM Sigplan Notices*, volume 37, pages 96–107. ACM.
- Karagiannis, G., Altintas, O., Ekici, E., Heijenk, G., Jarupan, B., Lin, K., and Weil, T. (2011). Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions. *Communications Surveys & Tutorials, IEEE*, 13(4):584–616.
- Kaur, A. and Malhotra, J. (2015). A survey of network simulation tools. *Wireless Communication*, 7(6):191–194.
- Lloyd, J. W. (2012). *Foundations of logic programming*. Springer Science & Business Media.
- Ma, H., Zhao, D., and Yuan, P. (2014). Opportunities in mobile crowd sensing. *Communications Magazine, IEEE*, 52(8):29–35.
- Noble, J. J. and Alpert, G. P. (2008). *Managing accountability systems for police conduct: Internal affairs and external oversight*. Waveland Press.
- Pentland, A., Fletcher, R., and Hasson, A. (2004). Daknet: Rethinking connectivity in developing nations. *Computer*, 37(1):78–83.
- Pereira, P. R., Casaca, A., Rodrigues, J. J., Soares, V. N., Triay, J., and Cervelló-Pastor, C. (2012). From delay-tolerant networks to vehicular delay-tolerant networks. *Communications Surveys & Tutorials, IEEE*, 14(4):1166–1182.
- Rainie, L. (2010). Internet, broadband, and cell phone statistics. *Pew Internet & American Life Project*, 5.
- Rutten, N., van Joolingen, W. R., and van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1):136–153.
- Skouloudi, C. and Karsberg, C. (2014). Annual incident reports 2014. *European Union Agency for Network and Information Security (ENISA)*.
- Vasilakos, A. V., Zhang, Y., and Spyropoulos, T. (2011). *Delay tolerant networks: Protocols and applications*. CRC press.
- Wei, K., Liang, X., and Xu, K. (2014). A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues. *Communications Surveys & Tutorials, IEEE*, 16(1):556–578.

BlinGui: uma solução vestível de apoio a pessoas portadoras de deficiência visual na detecção de obstáculos estáticos em ambientes internos

Elidiane P. dos Santos¹, Flávia S. Nascimento^{1,2}, Vaninha Vieira^{1,3}

¹Programa de Pós-Graduação em Computação da Universidade Federal da Bahia (UFBA)

²Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA)

³Centro de Projetos Fraunhofer da UFBA
Salvador – BA – Brasil

{elidiane.pereira,vaninha}@ufba.br, flaviamsn@ifba.edu.br

Abstract. *Frequently, visually impaired people bump into objects during their walking paths and this is the main cause of collision accidents. To minimize this problem, many researches develop technological solutions capable of detecting obstacles such as electronic canes. However this type of solution is ineffective for different scenarios with limitations in the perception of obstacles, besides occupying one hand. This article present the development of a new solution wearable, that is the BlinGui for detecting static obstacles in indoor environments. Wearable devices have the advantage to function without the use of hands, and is observable and controllable. The evaluation had on average 90% positive results.*

Resumo. *Frequentemente os portadores de deficiência visual se deparam com objetos durante seus trajetos e esses são os principais causadores de acidentes por colisão. Para amenizar esse problema diversas pesquisas desenvolveram soluções capazes de detectar obstáculos, no entanto muitas como as que usam bengala são ineficazes para diversos cenários, com limitações na percepção de obstáculos, além de ocupar uma das mãos. Este artigo apresenta o desenvolvimento de uma nova solução vestível, que é o BlinGui para detecção de obstáculos estáticos em ambientes internos. Dispositivos vestíveis tem a vantagem de funcionar sem o uso das mãos, além de ser observável e controlável. A avaliação desta solução teve em média 90% de resultados positivos.*

1. Introdução

Nos últimos anos a pesquisa científica tem criado métodos para facilitar o deslocamento dos Portadores de Deficiência Visual (PDV), que já são 285 milhões em todo o mundo segundo a Organização Mundial de Saúde [OMS 2016]. Esse número tem estimulado pesquisadores a usar tecnologias a favor desses indivíduos, que enfrentam diariamente grandes dificuldades em se locomover devido a existência de obstáculos não percebidos. Objetos estáticos como galhos de árvores, telefones públicos e fundo de caminhões podem prejudicar o deslocamento dos PDV como ilustra a Figura 1.

A bengala é uma ferramenta de baixo custo frequentemente utilizada por PDV como auxílio na locomoção, entretanto pode apresentar limitações como a falta de

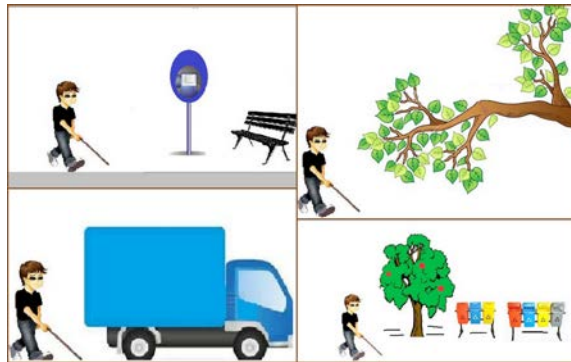


Figura 1. Cenários de interação entre PDV e obstáculos estáticos suspensos (imagens superiores) e rasteiros (imagem inferior).

percepção com objetos suspensos como, galhos de árvores e telefones públicos. Entretanto, existem soluções automatizadas acopladas a bengala [Kumar et al. 2014] e diversas outras complementares [Lee et al. 2013] que podem ser mais úteis. As chamadas bengalas eletrônicas detectoras de obstáculos que apesar de automatizadas continuam sendo uma bengala. Isso quer dizer que ainda possuem limitações, além de requerer exclusividade de uso das mãos do usuário. Todavia, existem as soluções complementares que são vestíveis e deixam o indivíduo com seus principais membros livres como boné eletrônico, cinto, mochila, colete, pulseira e colar. Essas soluções são capazes de se conectar ao usuário sem prejudicar suas atividades, não monopoliza a sua atenção e pode funcionar continuamente sem intervenções e ser controlada pelo usuário [Mann 1997].

O grande desafio observado nas soluções atuais é desenvolver ferramentas de baixo custo capazes de encontrar objetos imperceptíveis pela bengala. Assim como há muitas soluções aplicadas para detectar obstáculos, também existem diversos espaços para pesquisa nesta área. A detecção de objetos em um ambiente interno é um ponto que deve ser melhorado. Em vista disso, o objetivo geral deste trabalho é desenvolver uma solução vestível e acessível para detecção de obstáculos estáticos em ambientes interno, visando apoiar a locomoção dos PDV. Através de um protótipo eletrônico desenvolvido com as ferramentas advindas dos avanços da computação ubíqua. Uma vez que esta área promissora tem apresentado resultados de destaque no sentido da produção de protótipos computacionais miniaturizados, com boa capacidade de processamento e importantes características como mobilidade e transparência de localização [Weiser 1999]. Espera-se com essa pesquisa contribuir para os PDV com a prevenção contra colisões graves tornando a locomoção mais segura.

Além dessa introdução o restante deste artigo está organizado como segue. A Seção 2 descreve a solução proposta, a Seção 3 apresenta a avaliação da solução, a Seção 4 traz a comparação entre os trabalhos correlatos e a proposta e a Seção 5 discute as conclusões e trabalhos futuros.

2. BlinGui 2.0: Detecção de obstáculos

Este trabalho propõe uma solução vestível para detecção de obstáculos em ambientes internos, visando apoiar a locomoção de PDV, deixando-os mais seguros dos caminhos a serem percorridos, conscientes dos obstáculos existentes e conseqüentemente mais integrados com o meio em que vivem. O intuito é que seja um recurso adicional para o usuário

em combinação com os instrumentos rotineiros de locomoção (como a bengala). O nome BlinGui (lê-se blaingui) é oriundo do inglês *Guide Blind* que quer dizer guiar cegos. Essa solução está dividida em dois componentes (1) Guia e (2) Detecção de Obstáculos. O componente 2 será apresentado neste trabalho intitulado como BlinGui 2.0. A Figura 2 ilustra a arquitetura do BlinGui com seus componentes. O componente 1 é um aplicativo android que utiliza sinal Wi-Fi para localização e guia de PDV. Este pode ser utilizado em conjunto com o componente 2 que é um sistema vestível, uma pequena caixa de dimensões (103x130x50 mm) com itens eletrônicos embutidos e presa ao pescoço como um colar. Este é utilizado para detectar obstáculos por meio de sensor ultrassônico informando ao indivíduo sobre objetos em sua direção que podem causar colisão.

A ilustração da Figura 2 também destaca em duas dimensões uma pessoa PDV fazendo uso do protótipo. A imagem com a pessoa virada para o lado demonstra a varredura frontal que o sensor faz para captar os obstáculos. A ilustração também redimensiona a representação do colar para uma melhor visualização. O sensor ultrassônico é apresentado acoplado com um servo motor, que faz a movimentação deste. Os dados que entram para serem processados são ilustrados como os valores das distâncias entre o dispositivo e um obstáculo. Estes dados são processados pelo arduíno e depois de toda lógica ser executada um sistema de alerta irá atuar com sons e luzes, as luzes são para PDV que distingue cores por vultos.

O sensor ultrassônico calcula a distância entre obstáculo e usuário através da emissão de ondas sonoras, a partir da distância estabelecida é possível detectar o obstáculo. Além da emissão de alertas o BlinGui 2.0 utiliza *bluetooth*, como ilustrado na Figura 2 para enviar informações a um *smartphone* que através de um aplicativo emite alertas de voz informando a direção do obstáculo (alto ou baixo) com vibrações.

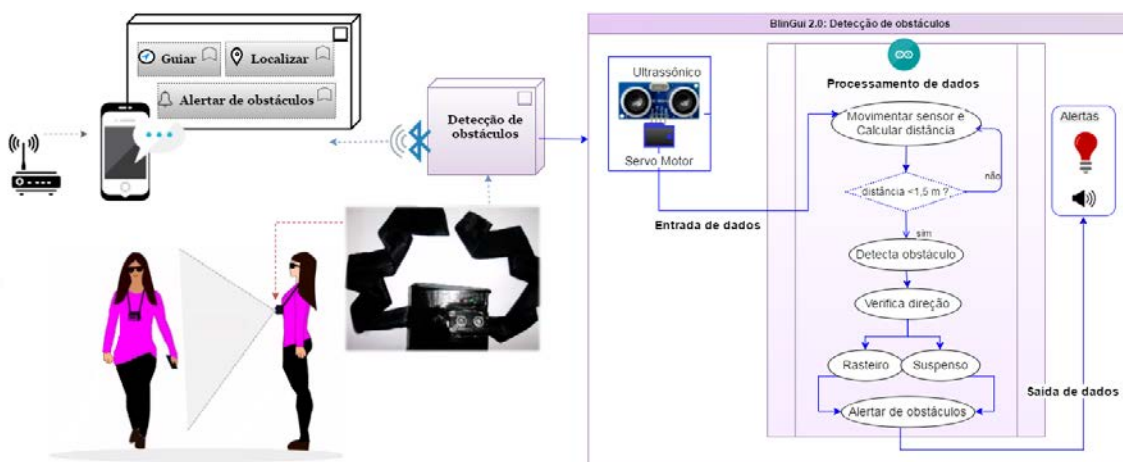


Figura 2. Arquitetura do BlinGui com o componente BlinGui 2.0 ilustrado em detalhes.

A Figura 3 ilustra o protótipo sendo utilizado por PDV em fase de testes com a detecção de obstáculos. O colar é composto por um Arduíno Uno, sensor ultrassônico, servo motor, módulo bluetooth, leds e buzzer.



Figura 3. BlinGui sendo testado por PDV na Associação Baiana dos Cegos com detecção de obstáculos.

3. Avaliação

O ambiente utilizado para avaliação foi em uma Associação para PDV. No início deste trabalho um grupo de 12 indivíduos participou da pesquisa de campo respondendo a um questionário, cujo objetivo foi conhecer as especificidades do público alvo. Na fase final de avaliação um outro grupo de 8 indivíduos participou dos testes com o protótipo, alguns faziam parte do primeiro grupo, outros não, isso devido a alguns contratempos que impediram as mesmos indivíduos de estarem presentes.

O questionário aplicado tinham perguntas referentes ao tipo de deficiência de cada pessoa, as dificuldades mais comuns enfrentadas por eles diariamente, os principais obstáculos que mais provocavam colisões, se a bengala consegue detectar todos os obstáculos, quais tipos de alertas preferiam receber para avisar de obstáculos e se usariam um dispositivo preso ao corpo para ajudar na detecção. Todas as respostas foram registradas para serem analisadas. Os participantes envolvidos tinham entre 18 e 60 anos de idade e a maioria nasceu com a deficiência.

Para estimar o impacto do colar e sua aceitação por PDV, uma avaliação foi executada com 8 participantes com o objetivo de avaliar a seguinte hipótese:

- **Hipótese:** as colisões dos PDV com obstáculos estáticos durante um percurso são minimizadas com a utilização do BlinGui para detecção de obstáculos.

Espera-se que a hipótese seja aceita, de forma que um indivíduo ao fazer uso do BlinGui para detecção de obstáculos as colisões sejam minimizadas. Para isso foi conduzido um teste em uma área reservada de $200m^2$, cujo objetivo era que os indivíduos usassem o dispositivo e trafegasse nesse espaço. Dessa forma poder avaliar se o indivíduo adaptou-se bem ao uso do dispositivo e se o dispositivo realizou as suas funções corretamente. Os testes foram divididos em duas etapas: Detecção de objetos rasteiros e Detecção de objetos suspensos. As duas etapas foram realizadas com e sem o colar eletrônico. Os indivíduos foram orientadas a percorrer uma trajetória com alguns obstáculos inseridos. Foram coletadas variáveis como tempo de cada percurso, obstáculos detectados corretamente e os não detectados. Os testes foram conduzidos com filmagem de cada etapa para uma posterior análise.



Figura 4. Execução dos testes para detecção de obstáculos rasteiros.

3.1. Primeira Etapa da Avaliação - Detecção de Obstáculos Rasteiros

Nesta etapa foram inseridos no caminho objetos rasteiros como cadeiras, lixeiras além das paredes e vasos que já continham no ambiente. Para cada objeto foi traçado um raio de $2m$, com objetivo de verificar se o BlinGui 2.0 de fato emitia o alerta a $1,5m$ do objeto. Estes testes foram feitos com e sem o uso do colar, nas duas situações os indivíduos preferiram utilizar a bengala. Porém o objetivo era que o protótipo alertasse dos obstáculos antes da bengala por mensagem de voz informando a direção do objeto. A Figura 4 exibe a realização desses testes.

Cada indivíduo fez um percurso com objetos espalhados no ambiente. A maioria dos obstáculos foi detectado de forma correta, porém teve momentos que a movimentação angular do sensor de distância era interrompida, ou tornava lenta. Nessas situações a detecção correta de obstáculos foi prejudicada, pois a direção do sensor poderia não estar voltada para o obstáculo principal a ser detectado. Um outro problema também encontrado com o dispositivo foi o excesso de alertas, que mesmo com o usuário parado ainda continuava informando excessivamente. Houve indivíduos não familiarizados com os alertas, e muitas vezes não se atentavam a tais informações e colidiram com os objetos.

3.2. Segunda Etapa da Avaliação - Detecção de Obstáculos Suspensos

Na segunda etapa dos experimentos bandeirolas suspensas na altura da cabeça do usuário foram inseridas no ambiente, simulando obstáculos suspensos para não provocar acidentes, já que a bengala não poderia ajudar neste caso. O objetivo era que o indivíduo percorresse o caminho com obstáculo e recebesse alertas com uma mensagem de voz sobre a direção do objeto detectado, se estar no alto ou baixo. Cada usuário executou este teste com e sem o uso do BlinGui, nos dois casos eles preferiram utilizar a bengala. A Figura 5 exibe a realização desses testes.

3.3. Resultados

Os resultados da pesquisa de campo constataram que todos os entrevistados concordam em utilizar o colar eletrônico de apoio à locomoção para ambientes internos não conhecidos, desses 90,0% acreditando que traria mais segurança no deslocamento. Todos os entrevistados relataram que já tiveram alguma espécie de colisão com objetos não detectados pela bengala, como fundos de caminhões, carros estacionados em locais indevidos, orlhões e postes.

Os resultados dos testes funcionais, para detecção de obstáculos rasteiros, resultaram na média de 88,0% detectados corretamente. Percebeu-se que com a movimentação



Figura 5. Execução dos testes para detecção de obstáculos Suspensos.

angular do sensor juntamente com o deslocamento do usuário o BlinGui 2.0 em casos específicos não identificou obstáculos corretamente. Nos testes executados sem o uso do colar sob as mesmas condições, foi observado que a bengala conseguiu detectar os mesmos obstáculos, porém quase colidindo. O BlinGui alertava com 1,5m do obstáculo, deixando o usuário atento sobre possíveis colisões.

Os testes executados para obstáculos suspensos resultaram na média de 90,0% detectados corretamente, o restante foram alertados com alguns atrasos, ou não alertou corretamente. Em todas as execuções dos testes a bengala não detectou os obstáculos suspensos, pelo fato do obstáculo estar a uma altura superior, o que torna difícil a detecção. A autonomia do protótipo foi satisfatória na execução dos testes com uma duração média da bateria. Baseados nesses testes iniciais o BlinGui 2.0 apresentou resultados satisfatórios com relação a detecção de obstáculos. A média e o intervalo de confiança foram calculados para cada um dos testes, considerando uma confiança de 95,0%. Os valores são representados na Tabela 1.

Tabela 1. Intervalo de Confiança dos Testes

Testes	Média	Desvio Padrão	Intervalo de Confiança
Obstáculos Rasteiros	4,4	0,55	[4,38; 4,41]
Obstáculos Suspensos	1,8	0,45	[1,78; 1,81]

Fundamentado nos testes executados pode-se aceitar a hipótese, com 95,0% de confiança, que as colisões dos PDV com obstáculos estáticos durante um percurso são minimizadas com a utilização do BlinGui 2.0. Em relação ao tempo não houve diferenças devido ao lugar ser um ambiente já conhecido pelos usuários. Pelo fato do colar ser algo novo para eles, alguns acabaram se distraindo com os alertas e avisos sonoros, ocasionando uma diminuição dos passos comprometendo uma amostragem real do tempo, já os demais permaneceram com passos normais. Opiniões dos usuários foram analisadas após a execução dos testes para verificar a aceitação do BlinGui 2.0, e todos concordaram em utilizar o colar eletrônico no dia-a-dia por ser confortável e leve, além de detectar objetos que muitas vezes não são encontrados pela bengala. Os hardwares utilizados para compor o colar têm alta disponibilidade no mercado e são de baixo custo, a fim de que o preço seja acessível para o usuário.

4. Comparação entre o BlinGui 2.0 e alguns Trabalhos Correlatos

As soluções correlatas apresentadas neste trabalho se baseiam nas vestíveis, acopladas à bengala e móveis. A Figura 6 exhibe alguns trabalhos relacionados e a Tabela 2 apresenta

a comparação entre eles. Para as soluções que têm um dispositivo acoplado a bengala foram identificadas algumas desvantagens, como:

- **Limitação aos movimentos:** obstáculos só serão encontrados onde a bengala for posicionada, por exemplo um suporte de telefone público é de difícil percepção através da bengala.
- **Não vestível:** O usuário só conta com a utilização de uma das mãos para suas demais atividades, pois a outra está exclusiva para a bengala.
- **Dependência:** priva o indivíduo de ficar independente da bengala em algum momento de sua vida, mesmo com o uso da tecnologia que deveria propiciar novas alternativas.



Figura 6. Algumas Soluções de apoio à locomoção de PDV vestíveis.

Tabela 2. Relação entre os trabalhos existentes na literatura e a solução proposta

Soluções Correlatas	Descrição das Soluções	Vestível	Bengala Acoplada	Mobile	Aéreo e Rasteiro	Externo	Interno
[Long et al. 2010]	Sistema de obstáculos baseado em visão estéreo combinado com atenção visual.	?	?	?	✓	✓	✓
[Garcia et al. 2011]	Bengala eletrônica que detecta obstáculos acima da cintura, utiliza sensor ultrassônico e alertas vibratórios.	X	✓	X	X	✓	✓
[Hernandez et al. 2012]	Usa estéreo visão, Kinect e regras de decisão fuzzy.	✓	X	X	✓	X	✓
[Bhatlawande et al. 2012]	Cinto e óculos eletrônico utiliza sensor ultrassônico.	✓	X	X	X	✓	✓
[Savvas et al. 2012]	Luva eletrônica composta de ultrassônico e alertas vibratórios, sonoros e Bússola Digital.	✓	X	X	X	✓	✓
[Lee et al. 2013]	Mochila inteligente possui sensores ultrassônicos nas alças em vários ângulos, usa redes zigbee.	✓	X	✓	✓	✓	X
[Kumar et al. 2014]	Dispositivo acoplado a bengala, informações de obstáculos pelo colete.	✓	✓	X	✓	✓	✓
[Rey et al. 2015]	Boné eletrônico utiliza sensor ultrassônico e arduino, detecta obstáculos acima da cabeça.	✓	X	X	X	✓	✓
BlinGui 2.0 (Proposta)	Colar eletrônico detecta obstáculos rasteiro e suspensos em ambientes internos. Utiliza ultrassônico e arduino e um smartphone.	✓	X	✓	✓	X	✓

Alguns trabalhos como por exemplo, *Electronic Long Cane (ELC)* [Garcia et al. 2011] e *Ultrasound Cane* [Kumar et al. 2014] decidiram automatizar o instrumento rotineiro utilizado pelos PDV. As soluções desenvolvidas foram acopladas à bengala para encontrar obstáculos de difícil acesso. A proposta do brasileiro Garcia é uma bengala eletrônica que sinaliza obstáculos acima da cintura. A proposta de Kumar é mais completa detectando obstáculos em outras direções e buracos, utiliza um colete pra receber avisos vibratórios e sonoros. Os dois realizaram experimentos, Kumar com dez voluntários vendados, Garcia com oito PDV. Os dois trabalhos não deram resultados que provem estatisticamente a validade das soluções.

Obstacle Detection Stereo vision [Long et al. 2010] e *Cognitive Guidance* [A. Landa-Hernandez 2012] utilizam visão computacional para detecção de obstáculos. Na proposta de Long não ficou claro se o sistema é vestível, porém a de Hernandez ficou explícito que é um protótipo em fase de construção. Apesar de vestível como uma espécie de capacete precisa carregar sempre um computador nas costas. Os dois declaram ter feito testes de avaliação porém, Hernandez não demonstrou resultados por meio de dados que validasse a sua proposta. Long apresentou uma tabela com alguns resultados que não foram analisados no artigo. Os dois não realizaram testes com PDV.

Waist-Belt for Visually Impaired [Bhatlawande et al. 2012] é um cinto e óculos detector de obstáculos composto com pares de sensores ultrassônicos, e *Smart Backpack* [Lee et al. 2013] é a mochila inteligente com sensores ultrassônicos acoplados em cada alça e um computador portátil. A solução de Bhatlawande detecta obstáculos na altura da cintura e cabeça, enquanto que a de Lee é mais completa calcula a posição do usuário, avalia a localização de objetos e determinar a direção da caminhada. Os testes feitos para avaliar o cinto e óculos foram executados com quatro usuários PDV treinados e quatro inexperientes. Segundo autor os resultados comprovaram uma boa precisão do dispositivo em descobrir obstáculos. Os testes de Lee com quatro pessoas de olhos vendados teve taxa de sucesso maior do que 95%.

Os dois últimos trabalhos correlatos são o CYCLOPS [Savvas et al. 2012] e o Blind Guardian [Rey et al. 2015]. O CYCLOPS, luva eletrônica, é composto de um sensor ultrassônico, uma bússola digital com acelerômetro e motor de vibração. O Blind Guardian é um boné que detecta obstáculos na altura da cabeça, o alerta é dado por vibração na mão direita do usuário. Enquanto as luvas ocupam um membro do corpo o boné possibilita liberdade as mãos. Porém não se pode dizer qual dos dois é mais útil, pois os resultados dos testes não possibilitam essa comparação. A avaliação da luva foi feita com um total de 16 testadores PDV e vendados. O Blind Guardian fez os testes com 29 indivíduos vendados onde 68,9% não colidiram com os obstáculos.

A Tabela 2 traz comparações dos trabalhos com o BlinGui 2.0, suas características, funcionalidades e tecnologias por ordem cronológica. A maioria dos trabalhos são sistemas vestíveis, os demais são acoplados a bengala ou ao computador, com detecção de obstáculos suspensos e rasteiros. Muitos trabalhos deixam a entender que funcionam em espaços internos e externos. Porém não houve relatos se aplicaram técnicas variadas para a detecção de obstáculo em ambientes distintos, visto que existem características diferentes para esses ambientes. Os trabalhos também não deixaram claro se as soluções detectam objetos em movimento com a mesma precisão dos estáticos. Esta pesquisa é ciente de que existem diferenças na detecção de obstáculos estáticos ou em movimentos

em ambientes internos ou externos, com isso deixa bem explícito em que domínio está inserida.

O BlinGui 2.0 possui independência da bengala, enquanto as soluções que se acoplam a ela, como as de Garcia [Garcia et al. 2011] e Kumar [Kumar et al. 2014] não possuem tal independência. Por ser uma solução vestível o colar está sempre conectado com o usuário, alguns trabalhos porém não se atentam a esta contribuição, como a solução de Long [Long et al. 2010]. À medida que o BlinGui é pequeno (103x130x50 mm), leve (140 gramas aproximadamente) e não proporciona desconforto, a proposta de Hernandez [A. Landa-Hernandez 2012] faz o usuário carregar alguns hardwares pesados sobre o corpo que pode atrapalhar a locomoção, a solução de Bhatlawande [Bhatlawande et al. 2012] possui mais de um dispositivo, e a mochila de Lee [Lee et al. 2013] pode ser grande e desconfortável ao usuário que deverá sempre carregá-la. O BlinGui utiliza somente um colar que permite as mãos do usuário ficarem livres, o que não acontece na luva de Savvas [Savvas et al. 2012]. O boné [Rey et al. 2015] só detecta objetos na altura da cabeça, enquanto o BlinGui em mais de uma direção.

5. Conclusões e Trabalhos Futuros

Colisões frequentes com obstáculos são reais problemas enfrentados pelos PDV, que sofrem diariamente com a possibilidade de danos fatais ou permanentes. Como exemplo, a perfuração de um dos olhos, causado pela colisão com o fundo de um caminhão estacionado em um local indevido, caso verídico relatado por um PDV. Baseado nessa problemática esse trabalho se propôs a contribuir com uma nova solução capaz de alertar ao usuário sobre possíveis colisões, evitando esses indivíduos passarem por situações temerosas como esta.

Este artigo propôs o BlinGui 2.0 como uma solução para detecção de obstáculos estáticos em ambientes internos, representado como colar eletrônico vestível, que detecta obstáculos suspensos e rasteiros. A solução foi desenvolvida e testada, os testes foram feitos com os PDV e estimaram uma média de quase 90% que o colar funciona satisfatoriamente. Os testes foram feitos em situações específicas, mas novas avaliações deverão ser feitas como trabalhos futuros, afim de testar todas as possibilidades possíveis, com variedades de objetos e em locais diferentes. Através dos testes foram reveladas algumas falhas de precisão existentes no colar. Aumentar a precisão dos métodos de Detecção de obstáculos é uma questão de muita relevância para próximas pesquisas.

Esse trabalho deixou várias lacunas que possibilitam o desenvolvimento de novos trabalhos. Uma dessas é integrar a computação sensível ao contexto [Vieira et al. 2009] ao BlinGui 2.0, para a detecção de obstáculos estáticos em qualquer ambiente com adaptação ao contexto do usuário. Projeto este que já está em fase de desenvolvimento, e será a próxima versão desse trabalho.

Os trabalhos correlatos levantaram a seguinte discussão: não deixaram claro o domínio ao qual estão inseridas suas pesquisas. Pelo fato, de não informarem se as ferramentas desenvolvidas são para qualquer ambiente(internos ou externos) e para qualquer obstáculo(estático ou em movimento). Esses critérios não foram evidenciado, deixando a entender que em alguns casos a solução engloba todo esse domínio. Porém isso só poderá ser comprovado por diversos testes aplicados em várias dessas situações, deixando claro as técnicas utilizada que resolvem esse problema.

References

- A. Landa-Hernandez, E. B.-C. (2012). Cognitive guidance system for the blind. *World Automation Congress (WAC), 2012*, pages 1–6.
- Bhatlawande, S., Mukhopadhyay, J., and Mahadevappa, M. (2012). Ultrasonic spectacles and waist-belt for visually impaired and blind person. *National Conference on Communications*, pages 8–11.
- Garcia, A., Fonseca, R., and Duran, A. (2011). Electronic long cane for locomotion improving on visual impaired people. a case study. *IEEE Pan American Health Care Exchanges*.
- Krumm, J. (2009). *Ubiquitous Computing Fundamentals*. Chapman & Hall/CRC, 1st edition.
- Kumar, K., Champaty, B., Uvanesh, K., Chachan, R., Pal, K., and Anis, A. (2014). Development of an ultrasonic cane as a navigation aid for the blind people. *IEEE, International Conference on Control, Instrumentation, Communication and Computational Technologies*.
- Lee, J. H., Kim, K., Lee, S. C., and Shin, B. S. (2013). Smart backpack for visually impaired person. *Proceedings - International Conference on ICT for Smart Society: "Think Ecosystem Act Convergence"*, pages 26–29.
- Long, C., Bao-long, G., and Wei, S. (2010). Obstacle Detection System for Visually Impaired People Based on Stereo Vision. *Fourth International Conference on Genetic and Evolutionary Computing*, pages 723–726.
- Mann, S. (1997). Wearable computing: A first step toward personal imaging. *Computer*, 30:25–32.
- OMS (2016). Organização mundial de saúde: Cegueira e deficiência visual. Disponível em: <http://www.who.int/mediacentre/factsheets/fs282/es/>.
- Rey, M., Hertzog, I., Kagami, N., and Nedel, L. (2015). Blind guardian: A sonar-based solution for avoiding collisions with the real world. *XVII Symposium on Virtual and Augmented Reality*.
- Savvas, D., Salonikidou, B., Diamantis, G., and Astaras, A. (2012). Development and evaluation of an open source wearable navigation aid for visually impaired users (CYCLOPS). *IEEE 12th International Conference on BioInformatics and BioEngineering*, pages 115–120.
- Vieira, V., Tedesco, P., and Salgado, C. (2009). Modelos e Processos para o Desenvolvimento de Sistemas Sensíveis ao Contexto. *Challenges*, 20:381–431. Disponível em: http://homes.dcc.ufba.br/vaninha/context/2009_TextoJAI_Final.pdf.
- Weiser, M. (1994). The world is not a desktop. *Interactions*, 1(1):7–8.
- Weiser, M. (1999). The computer for the 21st century. *SIGMOBILE Mobile Computing and Communication Review*, 3(3):3–11.

Como inferir o meio de transporte das pessoas em cenários urbanos a partir de smartphones

Carlos Alvaro de M. S. Quintella, Carlos Alberto V. Campos, Leila C. V. Andrade

Programa de Pós-Graduação em Informática – PPGI

Universidade Federal do Estado do Rio de Janeiro (UNIRIO) – Rio de Janeiro – Brasil

caquintella@gmail.com, leila@uniriotec.br, beto@uniriotec.br

Abstract. *This paper presents a mechanism to classify the transportation mode used by people in urban scenarios through movement traces collected by smartphone's own (mechanism user). In addition, a performance evaluation of the classifiers is conducted in order to provide valuable information for use by systems context-aware, especially through middleware with cloud based processing.*

Resumo. *Este trabalho apresenta um mecanismo para classificar o meio de transporte utilizado pelas pessoas em cenários urbanos por meio de traces (registros) de movimento coletados por smartphone da própria pessoa (usuário do mecanismo). Além disso, uma análise do desempenho dos classificadores do modo de transporte é realizada com o objetivo de fornecer informações valiosas para sua utilização por sistemas cientes de contexto, em especial através de middleware com processamento na nuvem.*

1. Introdução

Foi através dos *smartphones* que a visão da computação ubíqua introduzida nos anos 90 por Mark Weiser, pesquisador da Xerox, está sendo efetivamente realizada de acordo com [Baudalf 2007]. Assim, hoje uma boa parte da geração e do consumo de conteúdo digital passou a ocorrer com o usuário em movimento, dado tanto pela ampla disponibilidade das redes sem fio quanto dos dispositivos móveis.

Nos últimos anos, diversos estudos se voltaram para entender as informações produzidas por usuários de *smartphones* durante suas utilizações diárias. Destacam-se os estudos: [Zheng 2008a, Zheng 2009, Zheng 2010, Reddy 2010, Stenneth 2011, Shin 2015, Shafique 2016, Zhoua 2016]. Através destes estudos é possível constatar que a capacidade de identificar o contexto do meio de transporte das pessoas nas cidades é importante para diferentes tipos de aplicação, como: sistemas de *e-health*, marketing, cidades inteligentes, gerenciamento de trânsito, auditoria e de segurança em seus mais variados espectros. Cabe ressaltar que a detecção desses meios de transporte em tempo de utilização requer um alto nível de acurácia, sob pena de que os usuários percam a confiança na aplicação e, por consequência, deixem de usá-la [Reddy 2008].

Dentro desse contexto, o presente trabalho tem o objetivo de apresentar uma técnica aplicada para inferir o meio de transporte das pessoas em ambientes urbanos a partir de *traces* (registros de posicionamento) de mobilidade de *smartphones*, gerados pelo deslocamento dos usuários portadores destes *smartphones*. A partir do presente trabalho têm-se as seguintes contribuições: (i) um modelo para representação de trajetórias a partir de registros de movimentação de *smartphones*; (ii) um mecanismo

para segmentar trajetórias em agregados temporais, (iii) uma proposta para implementação de classificadores de *traces* de *smartphones* compatível com o uso de middleware de sistemas cientes de contexto, possivelmente baseados em uma nuvem computacional, (iv) uma coleta de dados de mobilidade humana em um cenário real (regiões das cidades: Rio de Janeiro e Niterói) e, (v) uma lista dos melhores classificadores para o meio de transporte de acordo com os métodos apresentados.

Este trabalho se estrutura da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3 são apresentados os principais métodos empregados a coleta de dados em um cenário real. A Seção 4 apresenta as atividades empregadas para a descoberta de conhecimento em banco de dados. A Seção 5 apresenta os resultados obtidos. Por fim, a Seção 6 apresenta a conclusão do artigo e os trabalhos futuros.

2. Trabalhos relacionados

No conjunto de trabalhos, [Zheng 2008a, Zheng 2008b, Zheng 2009], os autores abordam processos para classificar a movimentação dos usuários baseado na coleta de dados de movimentação, com uso do GPS embarcado em telefones celulares. Foi proposto um mecanismo para segmentar as trajetórias em etapas de movimentação, ou segmentos de trajetória de acordo com pontos de mudança. Os autores levaram em consideração que entre a troca de meios de transportes deve haver uma pausa, mesmo que discreta. Com os pontos de mudança detectados, os autores descobriram que quando ocorre a mudança de meio de transporte, deve haver um segmento que é do tipo andando, como apresentado em [Zheng 2010]. Apresentaram também, um mecanismo de detecção de meio de transporte, que se utiliza de limiares de distância e tempo para classificar uma movimentação, em tipos distintos: andando e não andando. Nos segmentos de trajetória, os atributos analisados foram: velocidade, aceleração, mudança de direção e paradas. Assim, para inferir o meio de transporte utilizado pelos usuários, foi usado um processo sequencial baseado em quatro etapas: segmentação, extração de características, inferência e pós-processamento.

Cabe ressaltar que inicialmente a aplicação do estudo não foi voltada para a identificação dos deslocamentos em tempo real, mas sim para os deslocamentos que já haviam sido concluídos. Entretanto, como explicado em [Zheng 2010], os autores além de apresentarem em maiores detalhes o mecanismo de classificação baseado em pontos de mudança, também introduzem um modo de inferência, funcionando próximo ao tempo real. Este é baseado no treinamento com dados históricos, que são gerados pelo método apresentado no trabalho anterior [Zheng 2008b], possibilitando assim inferir os meios de transporte ainda em andamento. Foram utilizados *logs* (registros) de GPS, portanto, sem o uso da técnica de fusão de sensores. A utilização do mecanismo de identificação de pontos de mudança caracteriza um sistema de informações geográficas, no qual os próprios usuários são usados para alimentar a base de dados. O uso dos pontos de mudança caracteriza uma abordagem composta de duas etapas, nas quais são empregadas técnicas para melhorar os resultados da classificação obtida inicialmente.

No trabalho [Reddy 2008] apresenta-se uma técnica para identificação de meio de transporte através de *smartphones* com uso de GPS e do acelerômetro. O foco adotado foi a análise da variação de velocidade, pelo GPS, com o uso do acelerômetro de forma complementar, permitindo identificar com um alto grau de precisão os deslocamentos (ente os modos: parado, andando, motorizado e bicicleta). O uso da

combinação dos sensores mostrou um aumento de 10% na precisão, em contraste com a utilização somente do GPS. A taxa de amostragem utilizada foi 1Hz. O tempo para inferência inicial do meio de transporte foi de apenas 2 segundos transcorrido da coleta de dados. Neste trabalho, os autores analisaram o desempenho de diferentes classificadores: k-Nearest Neighbor, Decision-tree, Naive Bayes, Support Vector Machines, Continuous Hidden Markov Models (CH-MM) e um classificador em dois estágios compostos de Decision Tree associado à Discrete Hidden Markov Models - onde o segundo componente é aplicado como um filtro para eliminar os “ruídos” relativos a fronteiras de trocas de meio de transporte.

No trabalho [Stenneth 2010], os autores apresentam uma abordagem para detecção de meios de transporte utilizando informações de posicionamento de *smartphones* e informações geográficas diversas, como: pontos de ônibus, de linhas ferroviárias e informação do posicionamento dos ônibus em tempo real. Diferentemente dos estudos apresentados anteriormente, neste estudo a taxa de amostragem para o posicionamento utilizada foi de 30 segundos. Para a detecção do meio de transporte próximo ao tempo real, os autores utilizaram uma janela móvel de 1 minuto (equivalente a dois registros de posicionamento), onde utilizam os seguintes atributos: velocidade média, aceleração média, média de mudança de direção.

No trabalho [Shin 2015] os autores utilizaram apenas o acelerômetro dos *smartphones* para detectar o meio de transporte utilizado pelo usuário. Sendo que também se utilizaram dos conceitos de segmentação de trajetória e da separação entre segmentos de meios de transportes distintos por períodos de caminhada, de [Zheng 2008a]. A frequência de leitura utilizada foi 5s para computar as médias de aceleração, com as quais foi possível classificar os segmentos da trajetória com um algoritmo de quatro etapas. A coleta de dados foi feita em Zurique na Suíça, com 495 amostras de 30 usuários, os meios de transportes detectados foram: trem/bonde, carro e ônibus.

Com base nesses trabalhos, nas próximas seções do presente artigo, etapas foram realizadas para explorar o problema da classificação dos *traces* e obter informações para subsidiar a proposta de um mecanismo de classificação de *traces* que possa melhorar os resultados aqui apresentados.

3. Métodos utilizados e a coleta de dados realizada

Nesta seção serão apresentados os métodos empregados para a classificação do *traces* de movimentação e a descrição de uma coleta de dados reais.

Para selecionar a plataforma a ser utilizada, primeiramente foi necessário analisar a representatividade de mercado de cada plataforma, seguida de uma análise detalhada dos aspectos tecnológicos: disponibilidade de ferramentas de desenvolvimento, sensores, APIs, documentação e suporte. Foram avaliadas as plataformas Android, Windows Phone 7 e Apple iPhone. Sendo escolhida a plataforma iPhone da Apple por ser a que melhor atendeu o conjunto dos critérios avaliados. Cabe ressaltar que durante o processo de desenvolvimento e coleta de dados, a Google ainda não havia disponibilizado o Activity Recognition na plataforma Android.

Para localização foi utilizado o sensor de localização do iPhone, acessado pelo *framework CoreLocation* do iOS, apresentado em [Apple 2012]. Trata-se de uma fusão de sensores, que utiliza os dados do GPS, o *fingerprnt* dos sinais de WiFi e da telefonia celular para estimar a localização com um grau ótimo de precisão. A seguir serão

apresentadas seções referentes à representação da trajetória analisada, as técnicas e métricas utilizadas e um detalhamento sobre a coleta de dados realizada.

3.1. Representação da trajetória

Muitas vezes, os dados brutos coletados tem pouco valor para mineração de dados, assim, é necessária uma primeira interpretação e extração de características (features) antes de usá-los. A associação de registros de posicionamento a trajetórias é uma prática comum em banco de dados de trajetórias. O processo de estimar uma trajetória através de *smartphones* consiste em um problema difícil, pois não há um sensor embarcado que seja preciso em todas as localidades, conforme apresentado por [Thiagarajan 2011].

No presente trabalho foi utilizada uma etapa de processamento aplicada após a coleta de dados para o agrupamento dos diversos registros de posicionamento em trajetórias. Tal etapa visa identificar e associar um grupo de registros de posicionamento à uma trajetória específica. Então, após esta etapa, cada trajetória passa ter seus períodos de deslocamentos e pausas identificados, com seus respectivos registros de posicionamento associados a um único determinado deslocamento ou uma única pausa. Esta etapa está fortemente baseada no trabalho de [Idrissov 2012], onde os autores apresentam um framework para agrupamento de registros de posicionamento.

3.2. Técnicas e métricas utilizadas para avaliação de classificadores

Entre as técnicas de avaliação de classificadores, este estudo usou a técnica de validação cruzada (cross-validation), conforme descrita em [Tan 2009], mais especificamente foi utilizado o método k-fold (com $k=10$). Esta técnica foi selecionada por ser um método bem conhecido e aceito, por ser usado nos trabalhos de [Zheng 2008a, Stenneth 2011] e por ser compatível com a ferramenta Weka (utilizada para esta etapa do estudo). Sua utilização se deu de forma análoga ao que foi apresentado nos dois referidos trabalhos.

Os termos *precision accuracy* e *recall accuracy*, apresentados em [Sokolova 2006], referem-se a métricas para medição de desempenho dos classificadores onde existe ocorrência de classes não balanceadas. Desta forma, *precision accuracy* determina a fração dos registros que realmente acabaram sendo positivos no grupo que o classificador declarou como positivo. Já o *recall accuracy* mede a fração do total de exemplos positivos corretamente classificados pelo classificador. Um *recall accuracy* alto reflete em que de todas as ocorrências para uma determinada classe, poucas foram as que deixaram de ser corretamente classificadas [Russel 2010].

3.3. Detalhamento sobre as ferramentas de desenvolvidas para a coleta de dados

Visando maior domínio sobre a qualidade da informação obtida, chegou-se a conclusão que coletar os próprios dados de mobilidade apresentariam oportunidades adicionais a pesquisa. Além de permitir a investigação de aspectos voltados às características de mobilidade local, levando em consideração [Noulas 2012].

Primeiramente, foi criado um framework, para suportar não só a coleta de dados, mas toda a estrutura de pesquisa. A partir deste framework foi implementada uma infraestrutura para subsidiar a coleta de dados. Também foi desenvolvida uma aplicação para iPhone, denominada CityTracks, cuja a interface é apresentada na Figura 1. Esta aplicação permite a coleta de dados de movimentação, além de outros dados relativos às pesquisas de sensoriamento participativo, como notas e fotos geo-localizadas.

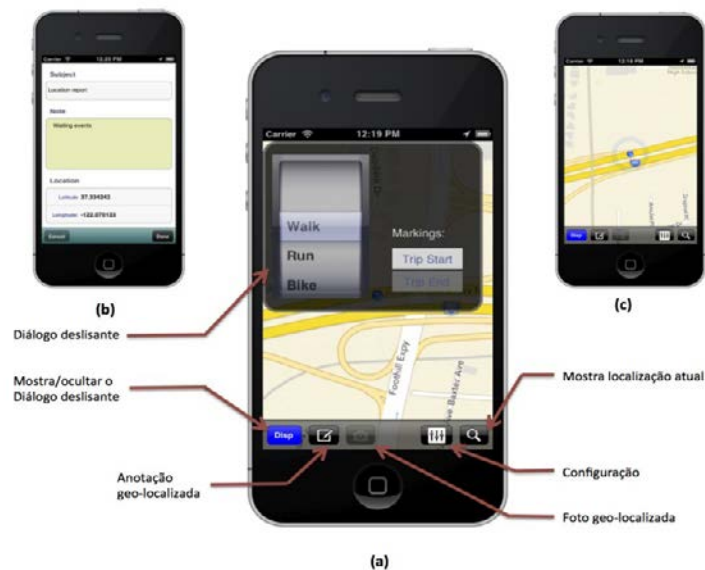


Figura 1. Interface da aplicação CityTracks. (a) tela principal, (b) tela entrada de notas, (c) tela visualização de posição.

Sobre a coleta de dados de movimento de pessoas foram utilizados nove voluntários, que disponibilizaram seus deslocamentos entre setembro/2012 e maio/2013 e, através dos quais, foram coletados mais de 880.000 registros de posicionamento. Os registros foram provenientes de 180 trajetórias desses voluntários, sendo que estas trajetórias referem-se aos seguintes meios de transporte: carro, motocicleta, ônibus, bicicleta, barca, a pé e metrô. A área geográfica abrangida pela coleta dos dados incluiu: (i) Zona Sul, Zona Norte e Centro da cidade do Rio de Janeiro e; (ii) Zona Sul, Zona Norte, Centro e Região Oceânica da cidade de Niterói. Os voluntários foram instruídos para que ao iniciar suas movimentações, executassem a aplicação *CityTracks* e informassem o início e o fim dos deslocamentos, junto com o meio de transporte utilizado. Foi através deste procedimento que viabilizou o uso de técnicas de classificação na etapa de mineração de dados. Cabe ressaltar que não foi possível trabalhar com os dados de trajetórias de metrô e barcas, devido à baixa qualidade e pequeno número de registros obtidos. Logo, as amostras relacionadas a este tipo de deslocamento foram descartadas. Assim, foram utilizados somente os dados de movimentação dos meios de transporte: ônibus, carro, moto, a pé e de bicicleta.

O formato do registro de localização coletado equivale aos dados disponibilizados pelo Corelocation framework do iOS6: *device_id*, *time_stamp*, *horizontal_accuracy*, *latitude*, *longitude*, *vertical_accuracy*, *altitude*, *speed*, *course*. Quanto a granularidade dos dados coletados, apesar de não ser possível controlar a frequência da geração dos registros no iPhone, em etapa posterior, quando a frequência ultrapassou 1 Hz, foi apenas mantido um dos registros do segundo e descartado os de menor *horizontal_accuracy*. Quando feita a comparação com os trabalhos relacionados, é possível notar as taxas de amostragem como: 30 segundos em [Stenneth 2011]; 1 segundo em [Reddy 2008] e, por fim, 2 segundos em [Zheng 2008a].

Ao longo do presente trabalho, nas etapas relacionadas à descoberta de conhecimento em banco de dados foi usado o método KDD descrito em [Fayyad 1996].

4. Análise de dados

Ao longo de todo processo apesar da utilização de algoritmos de classificação em trajetórias já concluídas, nenhum processamento ou suposição foi usado de forma que invalidasse a aplicação destas técnicas para deslocamentos em andamento. Com isso, espera-se que seu uso possa se dar para aplicações cientes de contexto e interativas.

4.1. Preparação e transformação dos dados

O processo de preparação dos dados foi feito através de scripts escritos no software *R*, sendo que um script para cada etapa. A primeira etapa correspondeu a preparação dos dados, ou pré-processamento. Nesta etapa foi feita a remoção dos registros imperfeitos, neste caso, onde o atributo *horizontal_accuracy* era maior do que 200 (metros), tendo como objetivo obter somente os registros com melhor precisão, excluindo *outliers*. A linha de corte de 200 m foi definida levando em consideração: (i) não descartar um número excessivo de registros, (ii) acomodar os diversos meios de transportes distintos, (iii) e ainda assim permitir o funcionamento em locais onde se tem mais baixa precisão de leitura, como embaixo de viadutos, áreas com arranha-céus e morros. Os registros foram então ordenados por dispositivo e tempo (*timestamp*).

A seguir, na etapa de transformação do dado foi feita a separação das trajetórias, orientada pelo framework de limpeza de trajetórias, apresentado em [Idrissov 2012], também se levando em consideração o trabalho [Zheng 2008a]. As trajetórias então foram derivadas em componentes de deslocamento ou de pausa. Sendo, que para isso foi necessário, não só a capacidade de detectar as pausas, como também o final de um deslocamento. No caso da detecção de pausa, esta foi feita pela aplicação de limiares de tempo e distância de pausa (0,4 metros e 90 segundos, respectivamente). Estes números foram obtidos pela análise visual dos deslocamentos e levaram em consideração condições de tráfego nas cidades do Rio de Janeiro e Niterói. Já para a trajetória utilizou-se o limiar de vinte minutos. Com isso, no caso deste limiar ser ultrapassado sem nenhuma ocorrência de registros de posicionamento, a trajetória corrente então passa a ser considerada encerrada, com os registros de posicionamento posteriores fazendo parte de uma nova trajetória. No caso de falta de alguns registros ao longo de uma trajetória, foi feita uma interpolação. Considerou-se a velocidade do valor anterior, considerando assim, que ocorreu um deslocamento uniforme entre o registro anterior e o registro corrente. Este tipo de procedimento produziu efeitos análogos a uma interpolação de pontos uniformemente distribuídos.

Com os segmentos de trajetória definidos foi possível analisar as informações de meio de transporte provenientes dos usuários, para então atribuir à classificação apresentada a um determinado segmento. Assim, cada registro de posicionamento, pode então ser associado a um deslocamento ou a uma pausa. Sendo que o deslocamento ou a pausa, por sua vez, são associados a uma determinada trajetória. Este procedimento é compatível ao apresentado no trabalho [Zheng 2011].

Dado que o objetivo deste trabalho é a classificação da movimentação que ainda está em andamento, foi utilizado um esquema de agregação temporal para agrupar uma série contígua de registros até que um determinado tempo seja atingido, gerando uma tupla de agregação com informações relativas ao período de tempo. Este agrupamento foi feito na última etapa de pré-processamento. Os segmentos temporais foram denominados *chunks*. Foram analisados três tamanhos de *chunks*: 60, 90 e 120 segundos. Sendo que os *chunks* de 90 segundos apresentam um melhor resultado quando se leva em consideração o ganho de precisão, inferência e o menor tempo para

início da classificação. Estes valores foram escolhidos tendo em vista não utilizar um valor pequeno demais, que poderia não incluir características de movimentação suficientemente ricas ou um valor grande demais, que poderia retardar o processo de identificação do meio de transporte.

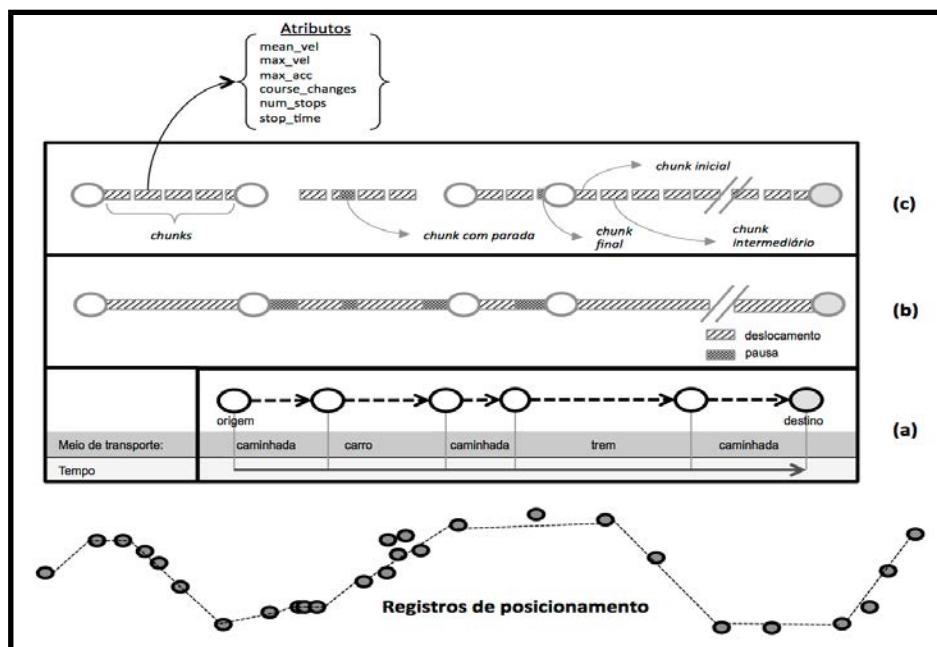


Figura 2. Representação da informação do movimento coletado passo a passo no processo de transformação e sumarização dos dados.

A partir dos *chunks* as informações de movimentação foram extraídas. Estas por sua vez podem ser traduzidas nos atributos: velocidade média; velocidade máxima, aceleração máxima, número de paradas e número de mudanças de direção, ilustrados na Figura 2. Para obtenção da velocidade foi utilizada a leitura de velocidade instantânea do GPS e a aceleração foi obtida através da computação da variação de velocidade ponto a ponto: $(v_{p2} - v_{p1})/t$, onde t é o intervalo de tempo e v_{p1} e v_{p2} , são respectivamente, velocidade reportada no ponto imediatamente anterior e a no ponto atual. Já quanto à mudança de curso (que é medido em graus radianos), este foi dividido por dez, onde somente a parte inteira da divisão foi considerada visando diminuir a sensibilidade a mudança de direção. Para o tempo de parada e número de paradas levou-se em conta a distância do deslocamento ponto a ponto dada pela fórmula de Haversine conforme descrita em [Cram 2016] e deslocamentos com velocidade menor que 0,4m/s foram considerados estacionários.

Na Figura 2 é mostrado como os registros de movimento são usados pela técnica proposta para inferir o meio de transporte. Na parte (a) desta figura é ilustrada a inferência já realizada dividindo o trajeto em diferentes meios de transporte. Na parte (b) a trajetória foi separada em porções de deslocamento e pausa. Na parte (c) podemos observar os *chunks* e seus atributos sendo identificados e representados.

4.2. Seleção de atributos e algoritmos

A etapa de seleção de atributos levou em conta tanto a análise visual da distribuição de frequência de cada um dos atributos e classes, quanto uma análise do ganho

informacional feita com o *Weka Experimenter*. Dessa forma, os seguintes atributos foram selecionados para utilização na etapa de mineração de dados: velocidade máxima, aceleração máxima, número de mudanças de direção.

A escolha de algoritmos foi feita com base nos trabalhos [Zheng 2010], [Reddy 2010], [Stenneth 2010] e [Wu 2008], onde o critério adotado para a seleção foi investigar tanto os algoritmos que apresentaram bons resultados em trabalhos anteriores, quanto outros que ainda não haviam sido verificados. Com isso, a lista de algoritmos analisados foi: Redes Bayesianas, Naive Bayes, SVM, Multilayer Perceptron, Decision Tree, Random Forest, Random Trees, K-Means, K-NN e Ada boost.

4.3. Avaliação dos classificadores

O processo de avaliação dos algoritmos de classificação foi feito com uso da ferramenta *Weka*. Para tanto, utilizou-se o método de *cross-validation 10-fold*. Iterações de análise sucessivas foram feitas, onde as etapas anteriores forneceram os fundamentos para as decisões de análise. Através desta abordagem foi possível realizar uma análise exaustiva e chegar aos melhores classificadores, inclusive para casos específicos de classificação.

Para treinamento foi utilizada uma aleatória amostragem equivalente a 20% dos dados, os demais 80% foram utilizados para a avaliação. Com isso, o número de *chunks* utilizados para o treinamento foi: carro=528; moto=144; andando=230; bicicleta=177, ônibus=54. Foi feita uma avaliação do impacto da amostra heterogênea pela aplicação de *oversampling* usando técnica de SMOTE, apresentado em [Tan 2009]. Assim, foi possível constatar a ausência de melhoria significativa para a classificação.

5. Resultados obtidos

A classificação dos *chunks*, segundo o meio de transporte, foi analisada com múltiplas abordagens e algoritmos: (i) classificar entre as classes andando, bicicleta, moto, carro e ônibus; (ii) classificar entre os meios motorizados e não motorizados; (iii) classificar apenas entre os não motorizados, ou seja: andando e bicicleta, (iv) classificar apenas entre os motorizados, ou seja, carro, moto, ônibus; (v) classificar entre andando, bicicleta e motorizado e, por fim; (vi) classificar em duas etapas, na primeira retirando algumas classes com alta taxa de classificação e, posteriormente, em uma segunda etapa fazendo a identificação das classes específicas. Com isso, foi possível identificar o conjunto de algoritmos com melhor resultado em cada caso, como descrito na Tabela 1.

Com base nos resultados obtidos na Tabela 1 foi possível chegar as seguintes conclusões: (i) - detectar um usuário andando não é tarefa difícil, pois diversos algoritmos conseguiram com precisão próxima a 100%; (ii) - detectar somente usuários andando e de bicicleta também não é difícil, pois os resultados ficaram acima de 90%; (iii) - separar as movimentações de usuários entre carro, moto e ônibus, é difícil (64%), pois os modos apresentaram perfis de movimentação muito parecidos; (iv) - separar usuários motorizados dos andando e de bicicleta, apresenta bons resultados (acima de 90%). Além disso, é possível utilizar a classificação em duas etapas, onde primeiramente extrai-se a possibilidade de alguns modos de movimentação. Depois, esta informação é usada de forma estocástica para a classificação na segunda etapa (vide Tabela 1). Os resultados se apresentaram mais promissores, subindo a precisão da classificação de 69% de acerto para 79% de acerto; quando se separando em três

classes: andando, bicicleta e motorizado na primeira etapa e, em seguida, aplicando o classificador para separar os motorizados se necessário.

Tabela 1. Resultado da análise dos meios de transporte por meio da acurácia de cada classificador utilizado para as oito classes de tipos de transporte

Classes	Abordagem	Classificador	acurácia
andando, bicicleta, moto, carro e ônibus	simples	DT	69%
andando, não-andando	simples	todos	> 98%
andando, bicicleta, motorizado		J.48	90%
motorizado, não motorizado	simples	SMO	92%
andando, bicicleta	simples	DT ou BN	94%
moto, carro, ônibus	simples	DT	64%
1) andando; bicicleta, motorizado; 2) carro, moto, ônibus	em duas etapas	J.48 e DT	79%
1) motorizado, não motorizado; 2a) andando, bicicleta; ou 2b) moto,	em duas etapas	SMO e DT	73%*

BN: Bayes Net; DT: Decision Table

6. Conclusão e trabalhos futuros

Foi apresentado um mecanismo para transformar dados brutos provenientes de traces de *smartphones*, em informação para subsidiar a extração do meio de transporte. Este mecanismo foi proposto com base na aplicação de aprendizado de máquina, e com base nos resultados de um processo de descoberta de conhecimento em banco de dados. Processo este que culminou com uma análise detalhada da performance de diversos classificadores; disponibilizando importantes informações para subsidiar a construção de um sistema de classificação da movimentação dos usuários. Tal mecanismo não necessita de etapas adicionais de pós-processamento e apresenta resultado superior a outras soluções abordadas entre os trabalhos relacionados, quando se compara as que somente utilizam o sensor de localização para fazer a inferência do meio de transporte.

Em relação aos trabalhos relacionados, é possível indicar um aumento próximo a 5% na precisão da detecção do meio de transporte sem o uso de sensores ou informações adicionais. Embora o conjunto de meios de transporte analisados e classes de inferência sejam diferentes, o presente trabalho apresenta o mesmo número que foi apresentado no trabalho com melhor resultado [Reddy 2008]. Cabe ressaltar ainda que o mecanismo aqui proposto utilizou dados para treinamento uma região geográfica específica, como em [Reddy 2008]. No caso de utilização do mecanismo aqui proposto com o uso de *middleware* para sistemas cientes de contexto baseados em nuvem, existe a possibilidade de utilizar o sistema de classificação apenas de acordo com o meio de interesse de cada aplicação. Com isso, a aplicação ciente de contexto poderia receber (através de uma assinatura) informações, somente para os meios de transporte de seu interesse. Podendo-se assim, potencializar o aumento da qualidade da identificação pela exclusão um ou mais meios de transporte; como por exemplo, uma aplicação de *e-health* poderia estar apenas interessada em diferenciar movimentações: andando, bicicleta e motorizadas. Outro exemplo seria uma aplicação de monitoração de pegadas de carbono, somente interessada nas movimentações motorizadas [Shin 2015].

Para trabalhos futuros existe a possibilidade de aumentar a precisão da inferência com o uso dos dados históricos de cada usuário ou analisar a possibilidade de aumentar a precisão à medida que o deslocamento vai acontecendo, ou seja, baseado nas medições anteriores. Além disso, existe também a possibilidade de se efetuar uma análise para bases de dados obtidas em diferentes cidades e comparar os resultados.

7. Referências

- Apple Inc., Location Manager Class Reference, disponível em: https://developer.apple.com/library/ios/#documentation/CoreLocation/Reference/CLLocationManager_Class/CLLocationManager/CLLocationManager.html#//apple_ref/doc/uid/TP40007125, acessado em 20/04/2016.
- Cram.org, Spherical trigonometry for geographic applications, página 16, disponível em: <https://cran.r-project.org/web/packages/geosphere/geosphere.pdf>, acessado em 01/06/2016.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., “The KDD process for extracting useful knowledge from volumes of data”, *Communications of the ACM*, v. 39, n. 11, 1996.
- Idrissov, A., Nascimento, M., Trajectory Cleaning Framework, Mobile Data Challenge (by Nokia) Workshop, 2012.
- Noulas, A., Scellato, S., Lambiotte, R., et. al., “A tale of many cities: universal patterns in human urban mobility”, *PLoS one*, Public Library of Science, v. 7, n. 5, 2012.
- Reddy, S., Burke, J. Estrin D., Et al. “Determining Transportation Mode On Mobile Phones”, *IEEE ISWC*, page 25-28, 2008.
- Reddy, S., Mun, M., Burke, J., et al., “Using mobile phones to determine transportation modes,” *ACM Transactions on Sensor Networks*, v. 6, n. 2, p. 13, 2010.
- Russel, S., Norvig, P., *Artificial Intelligence: A Modern Approach*, 3rd edition, New Jersey - EUA, Prentice Hall, 2010.
- Shafique, M. A., Hato E., “Travel Mode Detection with Varying Smartphone Data Collection Frequencies”, *Sensors*, 16(5), MDPI, p. 1-24, 2016.
- Shin, D., Aliaga, D., Tunçer, B., et al., “Urban sensing: Using smartphones for transportation mode classification”, *Computers, Environment and Urban Systems*, v. 53, p. 76–86, 2015.
- Sokolova, M., Japkowicz, N., Szpakowicz, S., “Beyond Accuracy, F-score and ROC: A Family of Discriminant Measures for Performance Evaluation”, *AAI 2006*, p. 1015-1021, Springer.
- Stenneth, L., Wolfson, O., Yu, P., et al., “Transportation mode detection using mobile phones and GIS information”, *19th ACM SIGSPATIAL*, p. 54-63, 2011.
- Tan, P., Steinbach, M., Kumar, V., *Introdução ao Data Mining, Mineração de Dados*, Rio de Janeiro - Brasil, Editora Ciência Moderna Ltda., 2009.
- Thiagarajan, A. “Probabilistic Models for Mobile Phone Trajectory Estimation”, Tese de doutorado, Massachusetts Institute of Technology, Setembro de 2011.
- Wu, X., Kumar, V., Quinlan, J. R., et al., “Top 10 algorithms in data mining”, *Knowledge and Information Systems*, 14(1), 1-37, 2008.
- Zheng, Y., Liu, L., Wang, L., et al., “Learning transportation mode from raw GPS data for geographic applications on the web”, *17th ACM WWW*, p. 247, 2008.
- Zheng, Y., Li, Q., Chen, Y., Xie, X. and Ma, W., “Understanding mobility based on GPS data”, *Proceedings of the 10th ACM Ubicomp*, p. 312-321, 2008.
- Zheng, Y., Zhang, L., Xie, X., and Ma, W., “Mining interesting locations and travel sequences from GPS trajectories”, *Proc. of the 18th ACM WWW*, p. 791-800, 2009.
- Zheng, Y., Chen, Q., Xie, X., et al., “Understanding transportation modes based on GPS data for web applications,” *ACM Transactions on the Web*, v. 4, no. 1, p. 1-36, 2010.
- Zhoua, X., Yub, W., Sullivan W. “Making pervasive sensing possible: Effective travel mode sensing based on smartphones”, *Computers, Environment and Urban Systems*, v. 58, 2016.

Controle de consumo de água baseado em reconhecimento facial

Pedro E. Santos¹, Roubert Edgar S. Silva¹,
Dalmy F. Carvalho Jr¹, Ilo A. S. Rivero¹, Raquel A. F. Mini²

¹Faculdade de Engenharia – Universidade de Itaúna (UI)
Rodovia MG 431 km 45, s/n - Campus Verde, Itaúna - MG, 35.680-142 - Brasil

²Departamento de Informática – Pontifícia Universidade Católica de Minas Gerais
Rua Walter Ianni, 255 - São Gabriel - Belo Horizonte/MG - Brasil

Abstract. *The ubiquitous computing aims to be invisible. From that, we can identify some individuals in an environment, allowing context aware and personalized services. Technologies like that can help water demand control, avoiding domestic and industrial water waste. This paper proposes a home water consumption control, based on facial recognition and wireless sensor networks (WSN) controlled by a mobile device. The water flow can be controlled and monitored after user's facial recognition. An experiment was developed to monitor and control water flow using real time face recognition. The results show that the proposed system can save up to 50% water consumption.*

Resumo. *A computação ubíqua possui como objetivo principal a presença uniforme e imperceptível na vida cotidiana do ser humano. A partir disso, é possível, além de outros recursos, identificar indivíduos em um ambiente, permitindo serviços personalizados e sensíveis ao contexto. Tecnologias como essa podem auxiliar o controle da demanda de recursos hídricos, afetados diretamente pelo desperdício contínuo domiciliar e industrial. Este trabalho propõe uma solução para o controle do consumo de água residencial, baseado em reconhecimento facial. Utilizando uma rede de sensores sem fio (RSSF) interligada a um dispositivo móvel, um usuário é identificado por meio da face para regular a vazão e monitorar a quantidade de água utilizada por ele. Desenvolveu-se um protótipo para monitorar e controlar esse consumo, que é ajustado em tempo real de acordo com a face do usuário identificado. Os resultados mostram que é possível economizar até 50% do consumo de água com a utilização do sistema proposto.*

1. Introdução

A utilização dos recursos hídricos cresceu mais que o dobro do crescimento populacional nas últimas décadas [WWAP 2015]. Esse crescimento se relaciona diretamente a práticas pouco estruturadas de consumo no meio agrícola, industrial e urbano. Consequentemente, há um aumento do estresse hídrico em regiões onde a demanda por água é maior do que a sua disponibilidade e capacidade de renovação [FAO 2011].

O desperdício de água é vivenciado em todo o Brasil. Segundo a Organização das Nações Unidas (ONU), cada indivíduo necessita de aproximadamente 110 litros de água por dia para atender suas necessidades de consumo e higiene [Morais 2015]. De

acordo com o relatório fornecido pela Secretaria Nacional de Saneamento Ambiental [SNSA 2014], o país possui uma média de consumo de 166,3 litros de água por habitante ao dia, podendo chegar a 253,5 litros em regiões onde a utilização do recurso não é devidamente controlada.

O objetivo deste trabalho é apresentar um sistema de monitoramento e controle do fluxo de água orientado à identificação facial de indivíduos em tempo real, desenvolvido com a utilização dos conceitos de visão computacional, microcontroladores e redes de sensores sem fio (RSSF). Com a integração do reconhecimento facial em um dispositivo móvel, interligado a uma central controladora de uma RSSF, é possível automatizar o controle da utilização da água dentro de uma residência pelos indivíduos residentes. Em relação à automação domiciliar, [Bolzani 2004] coloca que ela deve ser utilizada para ajudar o usuário a cumprir suas tarefas diárias, mas não a pensar por ele. As informações devem estar precisamente dispostas no tempo e no lugar necessários auxiliando na tomada de decisões e, através de aprendizado e adaptação, fornecer sugestões para o gerenciamento dos sistemas.

2. Reconhecimento facial

Na linguagem da Teoria da Informação, a extração de informações relevantes em uma imagem de face, codificação eficiente dos dados e uma precisa comparação de uma face com outros modelos codificados em uma base de dados são pontos importantes para que a técnica de reconhecimento seja praticável. O corrente projeto utiliza como algoritmo de reconhecimento facial a técnica de extração de texturas LBP (*Local Binary Patterns*). Nesse algoritmo, as faces podem ser vistas como uma composição de micropadrões que são bem descritos por esta técnica [Ojala et al. 1996].

O algoritmo LBP se baseia na intensidade dos *pixels* de uma imagem. A intensidade v_c de cada pixel de uma imagem é substituída por um vetor binário b , determinado pela Equação 1, que mostra a comparação entre a intensidade do pixel e as intensidades dos pixels vizinhos [do Amaral et al.].

$$b_{ij} = \begin{cases} 0, & v_{ij} < v_c \\ 1, & v_{ij} \geq v_c \end{cases}, \quad (1)$$

Na Equação 1, v_{ij} corresponde ao valor de um pixel na posição (i, j) , sendo $1 \leq i, j \leq N$. N corresponde ao tamanho da vizinhança, também chamada de máscara. Os valores obtidos para cada vizinho são, então, concatenados, gerando um número binário. Esse número obtido é convertido para sua base decimal substituindo o valor central v_c .

Os descritores de textura do LBP são obtidos isoladamente de regiões da face, comumente chamadas de malhas regulares. Após essa extração, são calculados os histogramas de intensidade dos *pixels* da imagem, resultando em valores que são concatenados e um único vetor de histogramas, facilitando a comparação e análise de amostras.

O processo chave do algoritmo se encontra nas classificações das regiões da imagem. No processo de classificação utilizam-se, basicamente, métricas que permitam explorar a importância de certas características faciais em relação a outras [Zhao et al. 2003]. Portanto, para cada região j da malha de histogramas, é possível

atribuir-se pesos específicos. Um exemplo de definição de pesos em regiões pode ser visto na Equação 2 de distância Qui-Quadrado [Ahonen et al. 2006].

$$X_w^2(x, y) = \sum_{i,j} w_j \frac{(x_{i,j} - y_{i,j})^2}{x_{i,j} + y_{i,j}}, \quad (2)$$

Na Equação 2, x e y são matrizes de histogramas a serem comparadas, i como um elemento pertencente a j -ésima região e w_j representando o peso pré-definido para a região j .

O algoritmo foi escolhido por haver melhores documentações a seu respeito, além de estar disponível em código aberto por meio da biblioteca OpenCV de visão computacional. O LBP vem sendo estudado bastante nos últimos anos, como pode ser visto em [Chan et al. 2007] e [Zhang et al. 2010]. O fato do algoritmo se apresentar de forma relativamente simples quanto à sua implementação também contribui para sua utilização.

3. Redes de sensores sem fio

Uma rede de sensores sem fio (RSSF) é definida por uma infraestrutura composta de sensoriamento, processamento e intercomunicação, realizada por um número considerável de dispositivos sem fio (nós sensores) situados em uma região de interesse, que pode ser utilizada para instrumentar, observar e reagir a eventos e fenômenos em ambientes específicos [Sohraby et al. 2007]. Uma RSSF foi utilizada em [Suprotim Sinha Majumdar and Alamelu 2015] para monitorar mudanças climáticas em plantações, permitindo um melhor controle do consumo de recursos, como água e energia.

As RSSF podem ser utilizadas para monitoração residencial, foco deste trabalho. [Rizzoli 2015] discute a importância no desenvolvimento de soluções que possam reduzir e controlar o consumo de água, e, principalmente, a importância de se conhecer o comportamento do indivíduo ao utilizar água. Este trabalho apresenta uma solução para monitorar o consumo individual, baseado na identificação visual de cada usuário de uma residência.

O nó sensor é desenvolvido com base na plataforma de prototipagem eletrônica *Arduino*, utilizando o microcontrolador *ATMega 328 AVR 8-bit*, sob um conjunto de instruções RISC [Atmel 2014]. Escolheu-se a plataforma com intuito de agilizar o processo de prototipagem dos nós sensores. A comunicação entre os nós da RSSF é realizada por meio do padrão ZigBee implementado sobre o padrão IEEE 802.15.4.

3.1. Sensoriamento e Atuação

Para as atividades de monitoramento do fluxo de água são utilizados sensores de fluxo de efeito hall. Os sensores utilizados possuem $\frac{1}{2}$ polegada de diâmetro e trabalham tensionados por 5V. Cada sensor possui apenas três fios para alimentação (*VCC*), aterramento (*GND*) e transmissão dos pulsos gerados durante a utilização de um dispositivo hídrico. O cabo condutor de pulsos deve ser ligado a uma entrada digital do *Arduino*, para que os pulsos gerados sejam recebidos e contabilizados. Para isso, são utilizadas as funções de interrupção do *Arduino* (*PCINT-Pin Change Interrupt*), disponíveis nos pinos 2 e 3.

O fluxo de água é controlado com a utilização de um servo motor atuando sobre um registro esférico de $\frac{1}{2}$ polegada e permite ajustes quantitativos de seu volume utilizado durante determinado instante de tempo. O protótipo elaborado permite ajustes quantitativos do fluxo de água em litros por minutos. A vazão em cada nó sensor é definida pelo usuário com a utilização da central de controle desenvolvida para tratamento e armazenamento dos dados oriundos do monitoramento, além do controle dos nós sensores.

3.2. Aplicação móvel de reconhecimento

O processo de detecção facial ocorre em um programa móvel, podendo ser utilizado em qualquer lugar, desde que haja luminosidade o suficiente para que um rosto seja detectado ou reconhecido. A aplicação de detecção facial de um indivíduo é uma das etapas do sistema completo. Ao haver a detecção de um membro domiciliar, suas atividades envolvendo recursos hídricos são enviadas a uma central de controle de dados, como descrito melhor na Seção 4.3. Foi desenvolvido um aplicativo para dispositivos Android. Ao iniciar a aplicação, é possível detectar uma face e realizar o cadastramento do indivíduo, reconhecê-lo após o cadastramento e visualizar os cadastros realizados.

3.3. Detecção Facial

O sistema de detecção e reconhecimento facial é composto pelas seguintes etapas: aquisição da imagem; pré-processamento da imagem adquirida; extração das características da imagem e classificação e verificação de reconhecimento da imagem. O diagrama da arquitetura do sistema de detecção e reconhecimento facial pode ser vista na Figura 1.



Figura 1. Arquitetura do sistema de reconhecimento de face

O processo do sistema inicia-se pela arquitetura do reconhecimento facial, na aquisição das imagens a serem analisadas. Uma imagem é capturada em tempo real, por meio da câmera do dispositivo móvel. Após a captura, a imagem é processada para servir de entrada para o reconhecimento e classificação final das faces. O resultado final consiste em uma imagem recortada e em escala de cinza, contendo apenas a face do indivíduo, para um melhor funcionamento dos métodos subsequentes da arquitetura. A captura e extração de imagens faciais para persistência são realizadas utilizando a biblioteca OpenCV (*Open Source Computer Vision Library*) de processamentos gráficos.

OpenCV [Bradski and Kaehler 2008] é uma biblioteca voltada à visão computacional utilizada neste trabalho. O OpenCV possui módulos de processamento de imagens

e vídeo, estrutura de dados, álgebra linear, interface gráfica do usuário (GUI), controle de mouse e teclado, além de mais de 2500 algoritmos, muitos dos quais são considerados estado da arte, tais como segmentação, detecção de faces (método Viola-Jones [Viola and Jones 2004]), aprendizado de máquinas, filtragem de imagens, rastreamento de movimento, entre outros métodos [Diniz et al. 2013].

3.4. Interface de reconhecimento de usuários

O usuário interage com o sistema apenas utilizando um dispositivo móvel, capturando informações faciais de indivíduos e posicionando o aparelho móvel em locais de fácil captura de rostos. A Figura 2 ilustra a interface principal da plataforma móvel, contendo botões de captura de faces e disparo da rotina de reconhecimento facial.

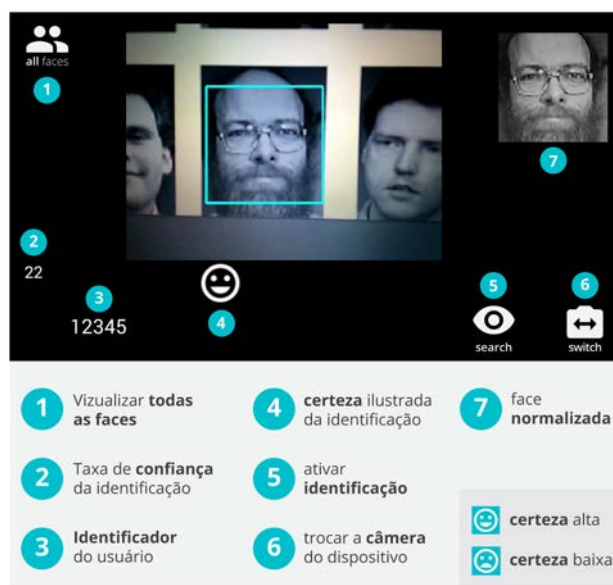


Figura 2. Interface do protótipo de reconhecimento facial móvel

As características da imagem obtidas pelo método anterior são, então, extraídas e comparadas com as imagens do conjunto de treinamento. Desse modo, a partir da comparação, é possível saber se uma face existe ou não na base de dados.

4. Monitoramento e controle do fluxo de água

4.1. Monitoramento

Os módulos da RSSF são compostos por um sensor de fluxo, um controlador de fluxo e um rádio transceptor. Durante o monitoramento do fluxo de água tem-se a execução das seguintes tarefas: coleta de dados referentes ao fluxo de água coletados pelos sensores de fluxo; os dados são transmitidos pela rede de sensores à uma central de controle; os dados são tratados e armazenados na central de controle e, por fim, os dados são apresentados ao usuário sob demanda.

O sensor de fluxo utilizado na atividade de monitoramento foi programado para fornecer a quantidade total de pulsos gerados durante a utilização de um dispositivo

hídrico. Ao tornar a geração de pulsos nula, ou seja, quando não há mais fluxo corrente no sensor, o número de pulsos contabilizado é enviado à central de controle.

O tratamento dos dados recebidos na central de controle é realizado com a divisão do número total de pulsos gerados durante uma atividade doméstica pela quantidade de pulsos por litro de água. A quantidade de pulsos por litro de água é obtida por meio da calibragem do sensor de fluxo, realizada com a contagem de pulsos durante o preenchimento de um recipiente até a marca de 1 litro. Após uma sequência de dez experimentos, obteve-se uma média de 378,9 pulsos por litro de água, com um coeficiente de variação menor que 3%, dentro dos limites definidos pelo fabricante.

A mensagem recebida na central de controle é tratada com a divisão do valor recebido (número total de pulsos gerados) por 380, o que define a quantidade de litros de água utilizados. Após o processo de tratamento, os dados são armazenados com a inclusão do nome do nó que originou a mensagem, seu endereço de *64bits* e a data e hora do registro.

4.2. Controle do fluxo de água

O controle de fluxo é realizado com a utilização de um controlador de fluxo que atua sobre a vazão de água em um dispositivo hídrico. Essa vazão é determinada na central de controle e definida no destino após a transmissão da mensagem do nó coordenador ao nó sensor. As etapas do controle de fluxo são:

- Definir o fluxo máximo de água em um nó sensor utilizando a central de controle;
- Transmitir a mensagem contendo o valor do fluxo definido ao nó destino;
- Analisar o fluxo corrente no sensor e realizar ajustes no fluxo de água em tempo de utilização.

O algoritmo desenvolvido para o controle de fluxo é integrado ao algoritmo de monitoramento. Como apresentado na sessão 4.1, apenas o número total de pulsos gerados durante toda a atividade doméstica foi utilizado, com o incremento progressivo a cada segundo. A necessidade de um comparador para que fosse medido o volume de água corrente no sensor é suprida com a integração de um contador de pulsos. O contador é zerado a cada segundo e fornece a vazão em litros de água por minuto utilizando a equação, disponibilizada pelo fabricante do sensor de fluxo de água, apresentada na Equação 3.

$$\text{litros}/\text{minuto} = \frac{P}{7,5} \quad (3)$$

Na Equação 3, P é o número de pulsos gerados durante o *delay* de um segundo dividido por 7,5, constante definida pelo fabricante para o sensor de fluxo de água de $\frac{1}{2}$ polegada.

Com a base de comparação definida, é possível realizar comparações entre o volume de água que está sendo utilizado e o valor definido na central de controle, permitindo ajustes do fluxo de água em tempo de utilização de determinado dispositivo hidráulico por um indivíduo.

4.3. Integração intersistêmica

O processo final de elaboração do sistema proposto consiste em integrar o sistema de reconhecimento facial ao sistema de monitoramento e controle do fluxo de água. A junção

dos dois sistemas ocorre na etapa de tratamento dos dados na central de controle. Durante a etapa de processamento, que precede a etapa de armazenagem dos dados, o usuário cadastrado que teve sua face detectada no sistema, é vinculado a todos os registros originados no local onde o reconhecimento foi realizado. A partir do momento em que o usuário finaliza seu uso em um dispositivo hídrico, sendo ele uma torneira ou chuveiro, os dados obtidos a partir do fluxo são enviados à central de controle para a exibição dos gastos, onde é possível verificar se há desperdício no local monitorado.



Figura 3. Processo de reconhecimento de indivíduos e monitoramento do fluxo hídrico

5. Resultados e validações

5.1. Resultados do Monitoramento

O sistema de monitoramento do fluxo de água apresentou resultados satisfatórios durante sua utilização, visto que a obtenção, transmissão, tratamento e armazenagem dos dados foram realizadas com sucesso. A obtenção dos dados foi realizada em uma residência com 5 moradores. Instalou-se o nó sensor em uma torneira localizada na cozinha da residência, com o sistema em funcionamento por aproximadamente 4 horas. Ao término dos experimentos de monitoramento, foi possível obter os resultados apresentados na figura 4

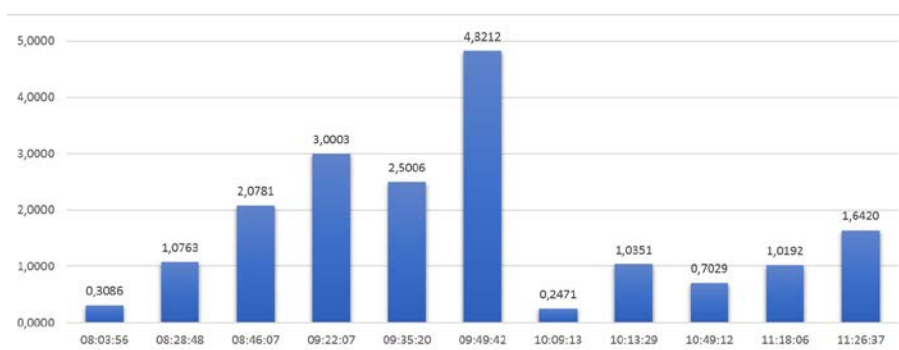


Figura 4. Resultados do monitoramento em um intervalo temporal

O gráfico apresentado na figura 4 mostra o monitoramento do fluxo de água e apresenta o eixo Y como sendo a quantidade d'água utilizados e o eixo X representando os horários das medições. O sistema trabalha com instantes de utilização, sendo possível visualizar a hora do recebimento da mensagem contendo a quantidade de litros gastos durante determinado intervalo de tempo.

5.2. Resultados do controle de fluxo

Durante o controle de fluxo, nenhum dado referente ao volume de água foi transferido à central de monitoramento, visto que toda persistência é voltada ao monitoramento. Os resultados foram obtidos através de uma aplicação desenvolvida em C++. Os dados apresentados foram registrados com a utilização do monitoramento serial por meio do método *Serial.print()*. A cada segundo, o valor obtido pelo cálculo de litros por minuto é impresso no monitor Serial. O histórico de medição é registrado no terminal de monitoramento, permitindo a extração dos resultados, sendo desnecessária a persistência de tais dados no banco de dados.

Foram realizados três experimentos de redução do fluxo de água. O primeiro experimento não conta com qualquer controle do fluxo, fornecendo a base de cálculo para a eficiência das reduções realizadas na sequência. Durante os experimentos, sem qualquer limitação de fluxo, foram obtidas 26 amostras. As quatro primeiras e quatro últimas amostras de dados foram descartadas, visto que as quatro primeiras registram a abertura da torneira e as quatro últimas registram seu fechamento. A média obtida durante os experimentos foi de 12 litros/minuto com a torneira totalmente aberta.

O segundo experimento foi realizado com a limitação do fluxo de água em 9 litros/minuto. Os resultados se mostram satisfatórios, apesar da baixa eficiência do algoritmo em tratar a redução de fluxo, visto que, se o fluxo é menor do que o valor determinado, entende-se que o controlador deve aumentar sua vazão até atingir seu ponto máximo (definido como 105 graus). O controlador de fluxo aumenta em 5 graus a abertura da válvula de fluxo de água progressivamente, até atingir o momento em que o fluxo de água se encontra próximo ou é igual ao valor delimitado. O controlador de fluxo é, então, desabilitado com a utilização do comando *detach(PIN)* no pino 11. Sua ativação é somente realizada novamente quando uma mensagem para limitação de fluxo é recebida.

O terceiro experimento de controle de fluxo foi realizado com o fluxo de água limitado a 6 litros/minuto. As diferenças no consumo de água dos três experimentos são apresentados na figura 5, onde é possível observar o problema do algoritmo de redução do fluxo. Na delimitação do fluxo em 9 litros por minuto, é possível observar que o fluxo ultrapassou o limite e só depois foi reduzido. Durante a redução, o fluxo chegou a ser menor do que o valor máximo e estabilizado no segundo seguinte, quando a diferença entre o valor delimitado (9 litros/minuto) e o valor corrente no sensor era menor ou igual a $\frac{1}{2}$. O ocorre quando o valor máximo definido para o fluxo de água foi de 6 litros/minuto.

6. Considerações Finais

6.1. Conclusões

O objetivo principal da prototipagem de reconhecimento de indivíduos foi alcançado. O protótipo realiza o reconhecimento facial e armazena seus dados, para posterior uso domiciliar e monitoramento individual dos recursos hídricos. O usuário é pré-cadastrado

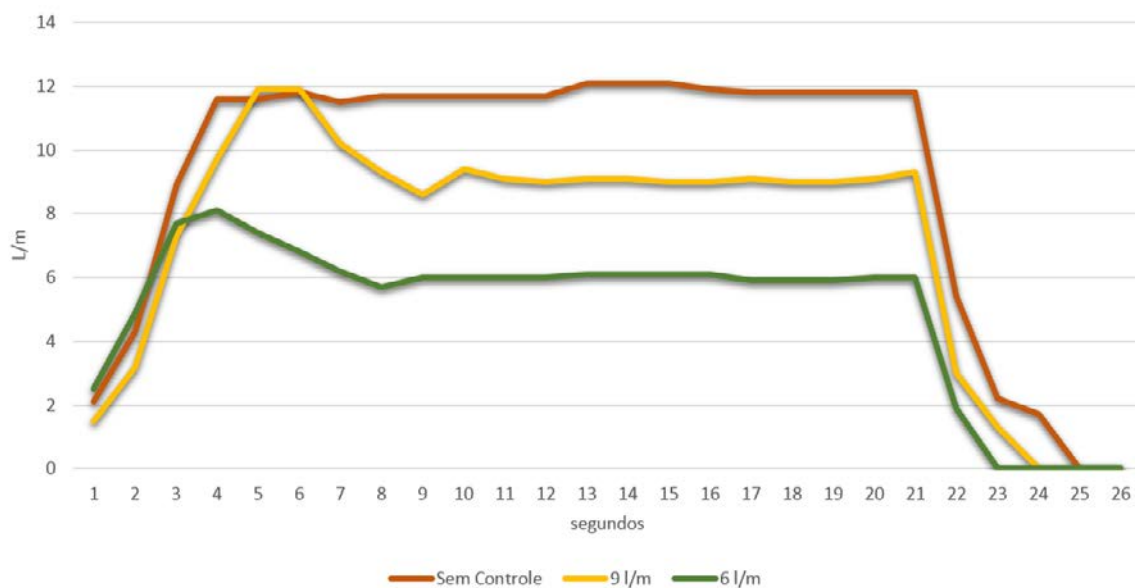


Figura 5. Experimentos de redução do controle de fluxo de água

no protótipo, por meio da detecção facial, para que posteriormente seja possível monitorá-lo e decisões sejam tomadas a respeito do possível uso excessivo da água na residência.

Desenvolveu-se o protótipo do controlador de fluxo com o objetivo de analisar a viabilidade da construção do sistema que permitisse controlar, de forma dinâmica, o fluxo de água. No resultado final do controlador, verificou-se que o protótipo apresentou folgas que não permitiu a relação direta entre a vazão e a posição do servo motor (0 a 180 graus), no entanto, o software desenvolvido para sanar o problema permitiu que o controlador atuasse de acordo com a proposta.

Os experimentos com o controlador de fluxo mostraram que é possível que o usuário tenha total controle sobre o fluxo de água em um dispositivo hídrico qualquer. Nos experimentos, definiu-se a vazão de 6 litros/minuto como o valor mínimo aceitável para que uma tarefa doméstica não seja prejudicada. Com esse valor, obteve-se uma economia de 50% em uma torneira que tinha como média de vazão 12 litros/minuto.

Dentre os trabalhos futuros está a integração desse sistema com o controle térmico da temperatura de um recurso, como temperatura da água ou do ambiente, permitindo que utilizadores diferentes tenham diferentes perfis de utilização. Um exemplo dessa personalização é o ajuste da temperatura do chuveiro e torneiras de acordo com o utilizador, bem como o ajuste da temperatura do ar condicionado no local onde a pessoa se encontra.

Referências

- Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041.
- Atmel (2014). Atmel 8-bit microcontroller with 4/8/16/32 kbytes in-system programmable flash datasheet.

- Bolzani, C. A. M. (2004). *Residências inteligentes*. Editora Livraria da Física.
- Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. "O'Reilly Media, Inc."
- Chan, C.-H., Kittler, J., and Messer, K. (2007). *Multi-scale local binary pattern histograms for face recognition*. Springer.
- Diniz, F. A., Neto, F. M. M., Júnior, F. d. C. L., and Fontes, L. M. O. (2013). Redface: um sistema de reconhecimento facial baseado em técnicas de análise de componentes principais e autofaces. *Revista Brasileira de Computação Aplicada*, 5(1):42–54.
- do Amaral, V., Giraldi, G. A., and Thomaz, C. E. Lbp estatístico aplicado ao reconhecimento de expressões faciais.
- FAO (2011). *The state of the world's land and water resources for food and agriculture (SOLAW) - Managing systems at risk*. Rome and Earthscan, London.
- Morais, R. (2015). Morador do df usa 67% mais água ao dia do que o indicado pela ONU.
- Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59.
- Rizzoli, A. C. M. G. D. P. A. C. A. (2015). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling and Software*, 72:198214:62.
- SNSA (2014). *Diagnóstico dos Serviços de Água e Esgoto - 2013*. SNSA/MCIDADES, Brasília.
- Sohraby, K., Minoli, D., and Znati, T. (2007). *Wireless sensor networks: technology, protocols, and applications*. John Wiley & Sons.
- Suprotim Sinha Majumdar, M. P. and Alamelu, J. (2015). Energy efficient wireless sensor network for polyhouse monitoring. *European Journal of Advances in Engineering and Technology*, (2.6):77–82.
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.
- WWAP (2015). *The United Nations World Water Development Report 2015 - Water for a sustainable world*. UNESCO, Paris.
- Zhang, B., Gao, Y., Zhao, S., and Liu, J. (2010). Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor. *Image Processing, IEEE Transactions on*, 19(2):533–544.
- Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458.

Controle de potência de transmissão em VANETs: Uma abordagem utilizando Teoria dos Jogos

Thiago Montenegro Uchôa¹, Heitor S. Ramos², Stênio Flávio de Lacerda Fernandes¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE) – Recife – PE – Brazil.

²Instituto de Computação – Universidade Federal de Alagoas (UFAL) – Maceió – AL – Brazil.

tmu@cin.ufpe.br, heitor@ic.ufal.br, sflf@cin.ufpe.br

Abstract. *This work proposes a power control algorithm to reduce the impact of high power transmission in Vehicular Ad Hoc Networks (VANETs). Based on game theory, we propose GRaPhic, a technique that encourages the devices connected to a VANET to reduce their transmission power. This algorithm provides enough incentive for a network does not deliberately increase its transmission power. In GRaPhic, we propose a utility function capable of reducing the transmission power of vehicles.*

Resumo. *Este trabalho propõe um algoritmo de controle de potência para reduzir o impacto de altas potências de transmissão em VANETs. Baseado em Teoria dos Jogos, propomos GRaPhic, uma técnica que estimula os dispositivos conectados à VANET a reduzir suas potências de transmissão. Este algoritmo provê incentivo suficiente para esta rede não aumentar deliberadamente suas potências de transmissão. Para isto, é proposta uma função utilidade capaz de reduzir a potência de transmissão de veículos, se um conjunto de condições for satisfeito.*

1. Introdução

A comunicação em VANETS se dá entre veículos V2V (*Vehicle-to-Vehicle*) ou veículos e infraestrutura V2I (*Vehicle-to-Infrastructure*). Ambas as formas de comunicação têm atraído o interesse de pesquisas, principalmente devido à mobilidade que se tem nestas redes (WANG; FAN; LETAIEF, 2012). Em contrapartida, a mobilidade aliada à alta concentração de transmissores, instalados nos veículos, pode gerar um problema relacionado à interferência entre os nós da rede. Em um ambiente de grande concentração veicular as taxas de transmissão podem ser reduzidas devido à alta interferência entre os nós da rede. Para reduzir este efeito, a abordagem escolhida foi a redução de potência de transmissão em cenários de grande concentração veicular. Para modelar o algoritmo de redução de potência de transmissão, foi escolhida a teoria dos jogos, como descrita abaixo.

A natureza da escolha da potência de transmissão em nós de uma rede pode ser modelada matematicamente com a Teoria dos Jogos não cooperativos. Isto se dá porque, neste caso específico, os nós da rede tentarão obter o maior *payoff* possível, ou seja, a melhor relação SINR (*Signal to Interference plus Noise Ratio*). Porém, com o uso de teoria dos jogos não cooperativos, as escolhas são feitas sem levar em

consideração os efeitos do SINR nos outros nós. Ou seja, as escolhas dos nós serão baseadas somente na maximização de seu SINR, independentemente se isto irá piorar, ou não, o SINR dos nós vizinhos.

Em contrapartida, a teoria dos jogos cooperativos é uma ferramenta poderosa na análise de comportamento de jogadores, como o exemplo de nós de uma rede, garantindo que os mesmos não tenham incentivos a desviar unilateralmente nos seus comportamentos. A teoria dos jogos cooperativos é um importante arcabouço no desenvolvimento de mecanismos descentralizados, nos quais o sistema é capaz de se auto organizar de maneira mutuamente satisfatória (CHEN, 2014). O diferencial deste trabalho é o de propor um algoritmo, o GRaPhic, que modela matematicamente a escolha da potência de transmissão utilizando teoria dos jogos cooperativos. O objetivo é aumentar a relação de sinal ruído interferência (SINR), modelando matematicamente o comportamento de veículos que trocam informações do tráfego veicular em um ambiente de grande concentração veicular, estimulando a redução da potência de transmissão dos módulos de comunicação.

As demais seções estão organizadas da seguinte forma: A Seção 2 introduz os principais trabalhos que envolvem telecomunicações utilizando teoria dos jogos. A Seção 3 detalha a metodologia proposta por este trabalho. A Seção 4 mostra os principais resultados obtidos com as simulações. A Seção 5 conclui o trabalho e apresenta os trabalhos futuros.

2. Trabalhos Relacionados

As principais abordagens envolvendo teoria dos jogos são subdivididas em teoria dos jogos: não cooperativo e cooperativo. A primeira se utiliza quando se quer modelar cenários onde os jogadores tomam decisões de forma a não se preocupar com os efeitos nos outros jogadores. Na teoria dos jogos não cooperativos o objetivo de cada jogador é maximizar sua função utilidade sem levar em consideração os efeitos de suas jogadas no *payoff* dos outros jogadores. No caso específico do cenário de comunicação, será mostrado que, apesar de largamente utilizada, a teoria dos jogos não cooperativos leva, muitas vezes, a resultados de uso da máxima potência de transmissão. Dependendo do cenário estudado, este resultado pode ou não ser prejudicial.

Por outro lado, quando utilizado teoria dos jogos cooperativos, é estimulado que os jogadores não tenham incentivos a desviar unilateralmente. Para isto, duas metodologias são utilizadas: (i) baseada em reputação, (ii) em preço. A primeira metodologia tem por filosofia isolar os nós que não cooperam. Em um cenário de comunicação, os nós que retransmitem mais informações, são classificados como cooperadores e seus pacotes são repassados por outros nós. Já no caso de nós que não cooperam, seus pacotes não são repassados. Para a metodologia baseada em preço, cada nó da rede tem associado a ele um preço pago para repassar pacotes de outros nós e o preço pago para toda a rede por repassar seus pacotes. Quando repassar pacotes de outros nós, ganha créditos para quando for necessário passar seus pacotes.

O controle de potência de transmissão se mostra eficiente na redução da interferência e aumento da capacidade na transmissão de dados em redes sem fio (DOUROS; POLYZOS, 2011). Neste trabalho, são mostrados exemplos de aplicação do controle de potência de transmissão. Apesar do foco desta seção ser a revisão de

trabalhos realizados em VANETs, muitos resultados referentes às MANETs e redes celulares são mostrados.

2.1. Aplicações de Teoria dos Jogos Não Cooperativos

Femtocell (FUJITSU, 2013) é um conceito de células de comunicação utilizando estações base (BS). Elas tipicamente são diferentes das BS tradicionais pelo seu reduzido alcance, 10m em média. Em (HAIDER et al., 2015) é proposto o uso de Mobile Femtocell (*MFemtocell*) para VANETs como trens, ônibus, táxis ou veículos particulares. Um dos maiores desafios para a aplicação das *femtocells* e *mfemtocells* está na interferência mútua entre as mesmas. Com o intuito de resolver o problema da interferência entre células, (HA; LE, 2014) propõe algoritmos de controle de potência distribuído e associação de distribuição de BS que possam manter a SINR em um nível determinado como bom. Um usuário terá a sua disposição células de vários tipos como: macrocélulas, microcélulas, picocélulas e femtocélulas. O algoritmo proposto ajuda na escolha de qual célula se conectar, baseado na qualidade do canal de comunicação.

O cenário de múltiplos nós de origem e destino, em uma rede sem fio, foi modelado utilizando teoria dos jogos não cooperativos repetitivos. Uma abordagem que visa otimizar a taxa de entrega de pacotes e SINR. Em (SMITH et al., 2014) são utilizados jogos com informações incompletas para implementações baseadas em IEEE 802.11.

Em (HAN; LIU, 2005) é proposto um algoritmo de controle de potência de transmissão e *throughput* não cooperativo para comunicações sem fio. A metodologia proposta é centralizada, possui limitações de valores de SINR e visa maximizar a taxa de transmissão dos veículos que estão na rede. Algumas limitações foram notadas, como a falta de detalhamento sobre as simulações, assim como o detalhamento a respeito dos algoritmos propostos.

2.2. Aplicações de Teoria dos Jogos Cooperativos

Em (KHAYATIAN; SAADAT; ABOUEI, 2013) foi proposto um método que utiliza a teoria dos jogos no controle da potência de transmissão associado ao roteamento cooperativo das informações. Deixa-se de trabalhar isoladamente com roteamento, sem a análise da camada física. Os nós são divididos em três categorias: origem, *relays* e destino. Os últimos são considerados as estações rádio base. O sistema de controle cooperativo é do tipo descentralizado. A transmissão da origem ao destino é feita utilizando dois *slots*. O primeiro é um *broadcast* da origem para todos os nós em sua área de alcance. O segundo é a transmissão do nó origem e *relay* para as BS (*Base Station*). A comunicação é feita em três etapas: 1) formação da coalizão, 2) alocação cooperativa de potência e 3) transmissão de dados.

Em (TINGTING CHEN, 2011) os autores estudaram uma forma de estimular que os nós em uma VANET cooperem, transmitindo os pacotes para o destinatário, obedecendo à sua tabela de roteamento. Os principais mecanismos de recompensa para roteamento sem fio são: 1) Sistemas de reputação; e 2) Sistema de crédito. É argumentado que o primeiro sistema de recompensa não é adequado para aplicações de VANETs por causa da dinâmica das VANETs. Como nas VANETs a arquitetura de rede é bastante dinâmica, é argumentado que o sistema de crédito é mais eficiente em termos de resultados que o sistema de reputação. Neste estudo só existe um protocolo de roteamento. Não são estudados cenários com mais de um protocolo de roteamento

visando a simplificação do cenário. A ideia é modelar a transferência de mensagens da origem ao destino como um jogo de coalizão de encaminhamento de informações. Para isto, são definidas funções de *payoff* para cada um dos nós integrantes da coalizão. Estas funções são projetadas levando em consideração a formação de um jogo que seja cooperativo e estável, estando este jogo modelado com teoria dos jogos utilizando a definição de *core*.

Em outros estudos como (YUYI LI, 2012), os autores desenvolveram mecanismos que permitem a seleção de *gateway* e formação da coalizão. Este trabalho, em específico, foca em teoria dos jogos cooperativos. O critério para escolha dos *gateways* é a máxima taxa de transmissão que o candidato pode receber da BS (estação base). Para a formação da coalizão, a função utilidade foi desenvolvida visando o incentivo de outros veículos a estabelecer conexões extras com outros nós. A função utilidade é variada baseada no número de nós de comunicação. Neste trabalho foi considerada uma área de 4km X 4km. Os veículos estão localizados de forma aleatória nesta área. Este estudo envolveu a análise tanto de V2V quanto de V2I. As principais limitações do trabalho são que os autores não utilizam padronização IEEE 802.11p, que é desenvolvida especificamente para o uso em VANETs e está amplamente utilizada em pesquisas no meio acadêmico.

O foco deste trabalho está no esforço em reduzir a interferência entre veículos concentrados em uma determinada área, urbana ou não, através da diminuição da potência de transmissão entre os veículos. As abordagens já utilizadas de controle de potência de transmissão normalmente não utilizam teoria dos jogos, mas outras ferramentas para modelar matematicamente o algoritmo de controle de potência de transmissão. Com a finalidade de reduzir a interferência entre os veículos, os cenários estudados utilizam jogos cooperativos de controle de potência de transmissão e potência de transmissão fixa. Estas duas abordagens são comparadas, o que é um diferencial em relação a outros trabalhos já desenvolvidos. Ao contrário dos trabalhos já desenvolvidos, onde as simulações não consideram cenários de concentração veicular, foram conduzidas simulações considerando cenários de concentrações veiculares variadas e com diferentes valores de potência de transmissão inicial. Assim, pretende-se analisar a influência da concentração veicular e potência de transmissão nos resultados do algoritmo proposto.

3. GRaPhiC: Um algoritmo de controle de potência de transmissão utilizando Teoria dos Jogos

3.1. Algoritmo Cooperativo

O cenário cooperativo consiste de veículos que se comunicam utilizando de teoria dos jogos cooperativos para decidir se devem diminuir ou não a potência de transmissão. O objetivo de cada nó é maximizar sua função utilidade, como pode ser observado na Figura 1. Porém o algoritmo induz os jogadores, transmissores veiculares, a decrementar sua potência de transmissão, considerando a potência do transmissor que enviou o pacote.

Antes de analisar a lógica do algoritmo, se faz necessário apresentar a sua função utilidade, u_i , representada em (1).

$$u_i = A \ln(1 + \text{SINR}) - B(\text{txPower}) \quad (1)$$

Para o cálculo da função utilidade acima, foi considerada a taxa máxima de transmissão de Shannon, primeiro termo da equação, e um fator decrementador em função da potência de transmissão do veículo. Se a função utilidade fosse composta apenas do primeiro termo de (1), com a maximização da função utilidade, nenhum nó iria tentar reduzir sua potência de transmissão. Tentaria apenas aumentar sua potência de transmissão, com o intuito de aumentar a SINR e, conseqüentemente a função utilidade. Para induzir um comportamento cooperativo, foi introduzida a segunda parcela do cálculo de u_i . Este segundo termo é o responsável por controlar a potência de transmissão do veículo. Os fatores que multiplicam o logaritmo natural e a taxa de transmissão (A e B) foram encontrados de forma empírica, onde $A = 1$ e $B = 0,009$. A escolha foi feita baseada nos resultados do SINR encontrados nas simulações, onde os SINRs se tornaram máximos para estes valores de fatores multiplicativos. Simulações foram realizadas variando os valores de A e B . Para cada simulação realizada com variações de A e B foram analisados os resultados de SINR. Os melhores resultados de SINR foram os encontrados para os valores escolhidos de 1 e 0,009, respectivamente.

Explicada a função utilidade, se faz necessária a explanação do algoritmo implementado no GRaPhiC. Alguns passos como calcular a função utilidade e algumas comparações da potência de transmissão e função utilidade são mostrados no diagrama abaixo. Como o objetivo é maximizar a função utilidade, o valor da função utilidade é sempre comparado com um valor máximo armazenado em memória. Se o valor atual for menor que o máximo armazenado, a potência de transmissão é reduzida para tentar aumentar o valor da função utilidade.

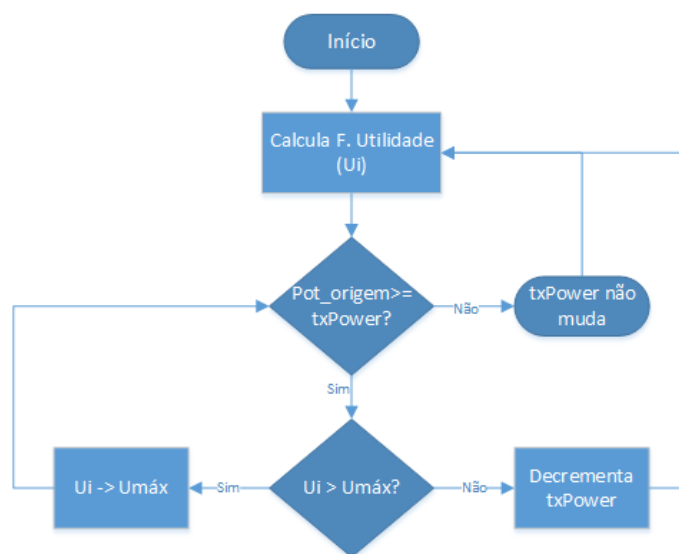


Figura 1: Algoritmo Cooperativo.

3.2 Métricas do Sistema

Para provar que o algoritmo proposto é efetivo, foram realizadas simulações levando em consideração condições realistas. Comparando com (LI *et al.*, 2014), o sistema proposto é modelado e simulado inicialmente com áreas de $8 \times 8 \text{ km}$ e $10 \times 10 \text{ km}$. Para analisar cenários com diferentes concentrações de veículos, foram simuladas rotas com 250, 500, 1000 e 2000 veículos. Os veículos se comunicam através

de *broadcasts* para todos os seus vizinhos em suas áreas de cobertura. Assim como em (YERRAMALLI; JAIN; URBASHI, 2014), a relação entre o número de veículos e a SINR média é analisada. As variáveis dependentes são: (i) potência de transmissão instantânea, (ii) SINR. Já as independentes são: (i) potência de transmissão inicial, (ii) número de veículos na simulação, (iii) área de interesse.

A influência das variáveis independentes sobre as variáveis dependentes é analisada. Espera-se que com o aumento na potência de transmissão, a SINR decresça, como pode ser observado na equação em (1). O aumento no número de nós na simulação deve aumentar a interferência na área de interesse, diminuindo a SINR e a taxa de transmissão. Todas estas análises devem ser mostradas com mais detalhes na seção de resultados das simulações.

3.3 Métodos de Avaliação

Para avaliar a metodologia proposta, GRaPhiC, se fez necessário simular os veículos gerando tráfego de dados na rede formada pelos transmissores veiculares. Para simular o tráfego de informações, foram gerados *broadcasts* quando os veículos se envolviam em acidentes. Todas as simulações foram realizadas utilizando o Veins. Nas simulações realizadas, os veículos que se envolvem nestes acidentes, enviam um *broadcast* para seus vizinhos (V2V) e Base Stations (V2I). Nas simulações as colisões são induzidas. Assim, nos instantes onde não há acidente, não há *broadcast*, a não ser aqueles de retransmissões de acidentes anteriores através de comunicação V2V. Antes de cada transmissão, quando um *airframe* é recebido, o algoritmo proposto é executado, analisando o ambiente para analisar a possibilidade de redução da potência de transmissão da unidade de bordo (OBU) do veículo em questão. O algoritmo proposto não incrementa a potência de transmissão, apenas reduz ou a mantém constante, desta forma foi possível analisar o cenário de uma forma simplificada.

O cenário de simulação estuda as comunicações tipo V2V e V2I. Para isto, são considerados *grids* com blocos de $100 \times 100 \text{ m}^2$. Os últimos simulam uma área formada por quarteirões. Para avaliar o efeito da concentração veicular no GRaPhiC foram criados dois tipos de *grid*. O primeiro tipo tem 80 cruzamentos na horizontal e 80 na vertical. O segundo tem 100 cruzamentos na vertical e 100 na horizontal. A BS é simulada no centro do cenário e os veículos se locomovem pelo cenário.

Por outro lado, cada veículo tem que ter uma rota pré-selecionada. Cada rota tem no mínimo 100 m de distância. Todas as rotas são simuladas utilizando o *randomTrips.py* (SUMO, 2013). Este *script* permite que o número de veículos na área de interesse seja gerado de forma constante ou aleatória, seguindo uma distribuição binomial. Para avaliar o efeito da quantidade de veículos em uma determinada área, foram simulados cenários com até 2000, 1000, 500 e 250 veículos. Para realizar as simulações, o cenário inicialmente é considerado vazio e novos veículos são inseridos na área de interesse a cada 0,5, 1, 2 e 4 segundos. Quando maior este intervalo, menor a quantidade de veículos no final da simulação, que se manteve constante para todos os cenários.

Para simular os acidentes, determinados nós são escolhidos e enviam *broadcasts* simulando uma mensagem de acidente. Nas simulações conduzidas, cada cenário tem 25 simulações de acidente. Em todos os cenários, os mesmos nós simulam acidentes. O número de veículos que simulam acidentes é proporcional ao número de veículos simulados. Com o número de veículos aumentando, também será maior o número de

veículos simulando acidentes. Este número varia entre 24 e 199, dependendo do número de veículos na simulação.

Foram realizadas diversas simulações, levando em consideração variações das variáveis independentes, como mostrado na Figura 2. Para analisar as simulações realizadas, inicialmente foi simulado um cenário de *grid* sem obstáculos. Por fim, também foram simulados veículos com diferentes valores iniciais de potência de transmissão: 30, 50, 80 e 100mW. A escolha destes valores de potência de transmissão se basearam em trabalhos já realizados anteriormente e são especificados em equipamentos utilizados em simulações reais.

O objetivo das simulações é analisar a razão SINR para os diferentes cenários. Para obter significância estatística, todos os cenários foram simulados por 30 vezes cada, com coeficiente de confiança de 95%. As simulações são realizadas, cada uma, simulando rotas veiculares distintas. Dependendo do cenário, várias simulações, como em (RAWAT; POPESCU; OLARIU, 2011), têm o tempo de simulação de 300s, inferior ao tempo das simulações deste trabalho, 1000s. A Figura 2 sumariza todos os cenários investigados. Como exemplo de aplicação, um cenário de simulação de 250 veículos, em uma área de 8km x 8km é destacado.

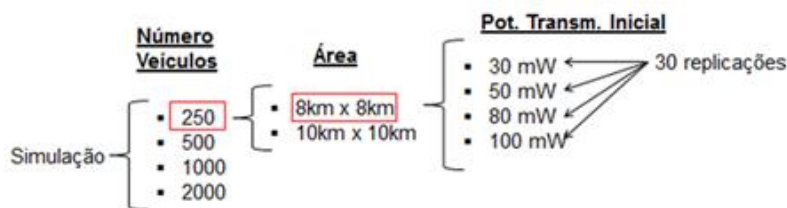


Figura 2: Cenários de Investigação.

4. Avaliação de desempenho e Discussão dos Resultados

4.1 Avaliação das Potências de Transmissão Médias

Neste tópico são analisados os resultados de potência de transmissão média. Foram simulados quatro cenários: (i) 250, (ii) 500, (iii) 1000 e (iv) 2000 veículos. Para cada um dos cenários acima, foram variados os seguintes elementos: potência de transmissão inicial entre 30 e 100mW e sua área de interesse, entre 80x80 quadras e 100x100 quadras. Por questão de simplicidade, as áreas de simulação de 80x80 quadras são representadas por áreas 80 e as áreas de interesse de 100x100 quadras são representadas por áreas 100.

Como pode ser observado na Tabela 1, a potência de transmissão média dos cenários de área de interesse de 80x80 quadras apresentam valores de potência de transmissão média inferiores que dos cenários de área de interesse de 100x100 quadras. Isto mostra que o algoritmo reduz de forma um pouco maior a potência de transmissão em áreas onde a concentração de veículos é maior.

Tabela 1: Resultados da Potência de Transmissão para simulação com 250 veículos para potências iniciais de: (i) 30mW, (ii) 50mW, (iii) 80mW e (iv) 100mW.

Potência Inicial (mW)	Área de Interesse	Potência de Transmissão Média (mW)	Redução da Potência de Transmissão (%)	Intervalo de Confiança Inferior (mW)	Intervalo de Confiança Superior (mW)
30	80	29,53	1,58	29,49	29,56
30	100	29,68	1,07	29,64	29,72
50	80	49,52	0,95	49,48	49,56
50	100	49,70	0,59	49,67	49,74
80	80	79,46	0,68	79,42	79,49
80	100	79,69	0,39	79,66	79,72
100	80	99,52	0,48	99,49	99,56
100	100	99,70	0,30	99,66	99,74

Nas simulações com 250 veículos, a maior redução na potência de transmissão se dá no cenário de potência de transmissão inicial de 30mW.

Por fim, a Tabela 2 mostra os resultados de potência média dos veículos para os diversos cenários de simulação com 2000 veículos. Como pode ser observado nos resultados obtidos de potência de transmissão, as maiores reduções de potência de transmissão se dão para a potência de transmissão inicial de 30mW. Como era de se esperar, a maior redução de potências de transmissão se dá para 30mW de potência de transmissão inicial e no cenário de 2000 veículos simulados em um *grid* de 80x80 quadras.

Tabela 2: Resultados da Potência de Transmissão para simulação com 2000 veículos para potências iniciais de: (i) 30mW, (ii) 50mW, (iii) 80mW e (iv) 100mW.

Potência Inicial (mW)	Área de Interesse	Potência de Transmissão Média (mW)	Redução da Potência de Transmissão (%)	Intervalo de Confiança Inferior (mW)	Intervalo de Confiança Superior (mW)
30	80	22,70	24,32	22,69	22,72
30	100	24,41	18,63	24,39	24,43
50	80	41,15	17,69	41,13	41,18
50	100	43,86	12,28	43,84	43,88
80	80	70,84	11,45	70,81	70,87
80	100	73,78	7,78	73,75	73,80
100	80	90,90	9,10	90,87	90,93
100	100	93,84	6,16	93,82	93,87

4.2 Avaliação dos valores médios de SINR

Nesta análise, são comparados os valores médios de SINR para os diversos cenários. Na simulação de 250 veículos, é sutil, porém perceptível, a diferença entre os valores de SINR dos cenários de área de interesse de 80x80 quadras e 100x100 quadras. No caso de maior concentração, foi possível notar que os valores de SINR são superiores. Além disso, é possível notar que para mesmos valores de concentração veicular, porém maior potência de transmissão inicial, os valores de SINR se mantêm superiores, à medida que a potência de transmissão é aumentada.

Uma redução na SINR nos cenários onde se tem 100x100 quadras mostrou que os veículos estão afastados a ponto de a SINR ser reduzida. O algoritmo proposto então reduz a potência de transmissão sem reduzir a SINR para este cenário. Ou seja, com uma maior concentração veicular, o que resulta em maior interferência, obtém-se uma redução na potência de transmissão média sem reduzir a SINR médio dos veículos.

Os resultados para a simulação com 500 veículos e potências fixas mostraram que os valores de SINR para o cenário de área de interesse de 100x100 quadras é bem próximo, mostrando que não há uma diferença substancial na redução da potência de transmissão em termos de SINR. Porém, para os cenários de área de interesse de 80x80 quadras houve uma diferença na SINR encontrada. Foram obtidos valores inferiores de SINR nos casos do algoritmo proposto.

Por fim, foi notória a redução da potência de transmissão entre os cenários de potência fixa e controlada pelo algoritmo. O que mostra que apesar de uma redução na potência de transmissão média, que pode chegar a 8% nestes cenários, não interfere nos resultados de SINR do cenário proposto.

5. Conclusão e Trabalhos Futuros

O estudo de VANETs está cada vez mais presente em pesquisas que envolvem desde aplicações para VANETs a soluções para roteamento eficiente, eficiência energética com controle de potência de transmissão, controle de mensagens através de QoS aplicado a VANETs e muitos outros. A teoria dos jogos está presente em vários trabalhos que não só se limitam à economia. São muitos trabalhos em redes de sensores, redes móveis celulares, VANETs, robótica, mobilidade urbana e outros.

Este trabalho propõe um algoritmo que modela matematicamente a escolha da potência de transmissão de nós de uma VANET utilizando teoria dos jogos cooperativos. Como diferencial, esta abordagem leva em consideração os efeitos da mudança de taxa de transmissão em uma VANET em alta concentração veicular, estimulando os nós da rede a reduzir a potência de transmissão. Foram analisados cenários com 250, 500 e 1000 veículos simulados e foi mostrado que em cenários de alta concentração veicular o algoritmo proposto reduz a potência de transmissão sem reduzir os valores médios de SINR.

Os resultados mostram que, em um cenário de alta concentração veicular, há uma redução de potência de transmissão média nos módulos de comunicação que pode chegar a uma média de 24,32% da potência de transmissão inicial. Em um cenário de veículos elétricos, há todo o interesse em se reduzir qualquer tipo de consumo elétrico,

umentando assim a eficiência energética do veículo como um todo, sem perder em qualidade de comunicação entre os veículos pertencentes a uma VANET.

6. Referências

- CHEN, X. Decentralized Computation Offloading Game for Mobile Cloud Computing. *IEEE Transactions on Parallel and Distributed Systems*, v. 26, n. 4, 2014.
- DOUROS, V. G.; POLYZOS, A. G. C. Review of some fundamental approaches for power control in wireless networks. *Computer Communications*, 34, n. 13, 2011. 1580–1592.
- FUJITSU. FUJITSU Network BroadOne LTE Femtocell Systems , 2013. Disponível em: <<http://www.fujitsu.com/global/products/network/products/femto/>>. Acesso em: 2014.
- HA, V. N.; LE, L. B. Distributed Base Station Association and Power Control for Heterogeneous Cellular Networks. *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, v. 63, n. 1, p. 282-296, 2014.
- HAIDER, F. et al. Spectral-Energy Efficiency Trade-off of Cellular Systems with Mobile Femtocell Deployment. *IEEE Transactions on Vehicular Technology*, v. PP, n. 99, p. 1-12, 2015.
- HAN, Z.; LIU, K. J. R. Noncooperative Power-Control Game and Throughput Game Over Wireless Networks. *IEEE TRANSACTIONS ON COMMUNICATIONS*, v. 53, n. 10, p. 1625-1629, 2005.
- KHAYATIAN, H.; SAADAT, R.; ABOUEI, J. Coalition-Based Approaches for Joint Power Control and Relay Selection in Cooperative Networks. *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, v. 62, n. 2, p. 835 - 842, 2013.
- RAWAT, D. B.; POPESCU, D. C.; OLARIU, G. Y. A. S. Enhancing VANET Performance by Joint Adaptation of Transmission Power and Contention Window Size. *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, v. 22, n. 9, p. 1528-1535, 2011.
- SMITH, D. B. et al. Multi-Source–Destination Distributed Wireless Networks: Pareto-Efficient Dynamic Power Control Game With Rapid Convergence. *Vehicular Technology, IEEE Transactions on*, v. 63, n. 6, p. 2744 - 2754, 2014.
- TINGTING CHEN, L. Z. F. W. S. Z. Stimulating Cooperation in Vehicular Ad Hoc Networks: A Coalitional Game Theoretic Approach. *IEEE Transactions on Vehicular Technology*, -, v. 60, n. 2, p. 566 - 579, 2011.
- WANG, Q.; FAN, P.; LETAIEF, K. B. On the Joint V2I and V2V Scheduling for Cooperative VANETs With Network Coding. *IEEE Transactions on Vehicular Technology*, v. 61, n. 1, p. 62-73, 2012.
- YERRAMALLI, S.; JAIN, R.; AND URBASHI MITRA. Coalitional Games for Transmitter Cooperation in MIMO Multiple Access Channels. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 62, n. 4, 2014. 757-771.
- YUYI LI, K. Y. P. C. H. Y. A. H. L. Cooperative Data Dissemination in Cellular-VANET Heterogeneous Wireless Networks , 2012.

Explorando Modelos Contextuais Híbridos: uma Abordagem Composicional

Roger Machado¹, Ricardo Almeida¹, Diórgenes da Rosa¹,
João Lopes², Ana Pernas¹, Adenauer Yamin¹

¹Universidade Federal de Pelotas (UFPel)

²Universidade Federal do Rio Grande do Sul (UFRGS)

{rdsmachado, rbalmeida, diorgenes, marilza, adenauer}@inf.ufpel.edu.br

jlblopes@inf.ufrgs.br

Abstract. *The Current proposals of hybrid context modeling brings with it new challenges, among which can cite the form of access and processing of data by applications. Thinking in that, this paper proposes an alternative to solve this problem through a compositional approach to context information in the hybrid models, titled EXEHDA-HM. The proposed approach stands out for the conception of a compositional processing strategy based on rules, which enables the application to combine contextual information stored in different models in a single rule, with the intention to facilitate the identification of situations of interest. To evaluate the approach was developed a case study using a repository composed of a non-relational and triple model, the results achieved stimulate continuation of the research.*

Resumo. *As atuais propostas de modelagem híbrida de contexto trazem consigo novos desafios, dentre os quais pode-se citar a forma de acesso e processamento destes contextos por parte das aplicações. Neste enfoque, o presente trabalho propõe uma alternativa para solucionar este problema por meio de uma abordagem composicional de informações contextuais em modelos híbridos, intitulada EXEHDA-HM. A abordagem proposta se destaca pela concepção de uma estratégia de processamento composicional baseada em regras, a qual possibilita à aplicação combinar informações contextuais armazenadas em diferentes modelos em uma única regra, tendo como intuito facilitar a identificação de situações de interesse. Para avaliação da abordagem foi desenvolvido um estudo de caso utilizando um repositório composto de um modelo não relacional e de triplas, os resultados atingidos estimulam a continuidade da pesquisa.*

1. Introdução

Ao se construir e executar aplicações ubíquas cientes de contexto há uma série de funcionalidades que devem ser providas, envolvendo desde a aquisição de informações contextuais, a partir de fontes heterogêneas e distribuídas, até a representação dessas informações, seu processamento, armazenamento e a realização de inferências para seu uso em tomadas de decisão [Bellavista et al. 2012].

Conhecer o contexto em que ocorre uma interação mostra-se uma tarefa complexa em sistemas computacionais. A construção de sistemas cientes de contexto exige,

inicialmente, a definição do que considerar como contexto, onde este se aplica e que informações são necessárias para descrevê-lo. É preciso viabilizar formas de realizar a aquisição do contexto de forma mais automática possível. Após a coleta de dados brutos a partir do sensoriamento, tornam-se necessários mecanismos de processamento de contexto que tratem as informações coletadas, produzindo informações contextualizadas [Bauer et al. 2014].

Atualmente, as aplicações cientes de contexto tendem a tratar com dados de diferentes naturezas, onde os mesmos podem ser modelados utilizando diversas abordagens. Nesta perspectiva se observa uma tendência na utilização de repositórios com diferentes modelos de armazenamento, chamados de modelos híbridos [Kotenko et al. 2013] [Perera et al. 2014], pois muitas vezes a utilização de somente um modelo para armazenamento das informações contextuais acaba não sendo o ideal, por questões ligadas a desempenho, utilização de disco, entre outras [Carvalho 2014].

O crescente uso de modelos híbridos traz consigo uma preocupação relacionada a forma com que as aplicações realizam o acesso aos dados. Tendo em vista o acesso e processamento de dados provenientes de diferentes modelos a proposta desenvolvida, denominada EXEHDA-HM (*Execution Environment for Highly Distributed Applications-Hybrid Modelling*), tem como premissa oferecer suporte à utilização de modelagem híbrida de contexto. Para tanto a EXEHDA-HM provê: repositório único, que permite o armazenamento das informações contextuais em diferentes modelos; uma estratégia de processamento composicional de contexto, a qual possui como diferencial a capacidade de combinação das informações presentes nos diferentes modelos de contexto utilizados. A abordagem proposta é integrada ao *middleware* EXEHDA [Lopes et al. 2014], contribuindo com o seu Subsistema de Adaptação e Reconhecimento de Contexto.

Este artigo está dividido da seguinte forma: a Seção 2 discute a Ciência de Contexto e as etapas envolvidas para a sua obtenção; a Seção 3 apresenta a abordagem EXEHDA-HM, mostrando a sua arquitetura e suas principais funcionalidades; a Seção 4 explora um estudo de caso, o qual demonstra o funcionamento da abordagem proposta; a Seção 5 descreve os trabalhos relacionados; e, por fim, a Seção 6 discute as considerações finais e os trabalhos futuros.

2. Ciência de Contexto

A ciência de contexto é um dos requisitos para obtenção da computação ubíqua, sendo vista como a capacidade de um sistema em usar o contexto para prover serviços e/ou informações relevantes para o usuário [Dey 2001] [Knappmeyer et al. 2013].

De forma resumida, a construção de um sistema ciente de contexto se dá por meio de três etapas: aquisição, a qual refere-se ao processo de monitorar e capturar informações contextuais; modelagem, que consiste na concepção de um modelo de entidades do mundo real, suas propriedades, estado de seu ambiente e situações que podem ser usadas como referência para a aquisição, interpretação e raciocínio de informações contextuais [Knappmeyer et al. 2013]; e processamento, o qual pode ser definido como um mecanismo de raciocínio para inferir novos conhecimentos e melhorar a compreensão dos contextos adquiridos [Neto et al. 2015].

Com relação a estas etapas, observa-se nos trabalhos a adoção de diferentes es-

estratégias provenientes de campos da inteligência artificial e de sistemas baseados em reconhecimento, sendo a estratégia baseada em regras a mais utilizada [Perera et al. 2014].

Já relacionado à etapa de modelagem de contexto, diversas abordagens são encontradas na literatura, cada uma com suas vantagens e desvantagens, não havendo ainda consenso a respeito de qual seria a ideal. Dadas as diferentes necessidades de representação de dados em aplicações ubíquas, abordagens híbridas tem sido consideradas promissoras, pois combinam duas ou mais formas de modelagem com variados níveis de interpretação e diferentes aspectos de uso.

Com a utilização de modelos híbridos de contexto novos desafios se mostram presentes, dentre os quais destacam-se: (i) o gerenciamento das informações contextuais; (ii) o armazenamento em diferentes modelos, proporcionando a persistência das informações contextuais; e (iii) o processamento destas informações, já que com o emprego de modelos híbridos as aplicações podem ter os contextos armazenados em diferentes modelos e, muitas vezes, é importante poder combinar estas informações visando facilitar a identificação das situações de interesse.

3. EXEHDA-HM

Tendo em vista os desafios de utilização das abordagens híbridas, o presente trabalho propõe a abordagem EXEHDA-HM, a qual possibilita o emprego de modelos híbridos de contexto, onde é concebido um repositório híbrido para armazenamento das informações contextuais, provendo as facilidades particulares de cada modelo utilizado e uma estratégia para processamento dos contextos adquiridos com base em regras composicionais. As regras composicionais possuem como diferencial a capacidade de combinar em uma mesma regra as informações presentes nos diferentes modelos de armazenamento, onde para isso são utilizadas *tags* de marcação na regra criada, sendo estas *tags* substituídas pelas informações que serão buscadas no modelo desejado.

A arquitetura da abordagem EXEHDA-HM é apresentada na Figura 1, onde pode ser visualizada na parte superior da Figura a camada de processamento, responsável pelo tratamento das regras composicionais, e o repositório desenvolvido, intitulado de RHIC (Repositório Híbrido de Informações Contextuais), o qual é dividido em duas camadas: a camada de serviço, responsável por disponibilizar métodos que permitem o acesso às informações presentes no repositório; e a camada de armazenamento, para provimento de formas diferenciadas de armazenamento.

A abordagem proposta foi integrada ao *middleware* EXEHDA, de forma a se beneficiar do suporte à aquisição, processamento e armazenamento de informações contextuais, assim como da arquitetura distribuída do *middleware*, características oportunas às funcionalidades da EXEHDA-HM. Para realizar esta integração modificações se fizeram necessárias na arquitetura de software do EXEHDA, as principais foram realizadas no Subsistema de Adaptação e Reconhecimento de Contexto, sendo garantida a interface desse subsistema com os outros existentes no *middleware*.

A seguir são apresentadas as três camadas presentes na EXEHDA-HM.

3.1. Camada de Armazenamento

Como também pode ser observado na Figura 1, o RHIC provê um repositório com três modelos de armazenamento:



Figura 1. Arquitetura da EXEHDA-HM

- modelo relacional: modelo já utilizado pelo RIC (Repositório de Informações Contextuais) do EXEHDA para armazenar informações contextuais. A sua utilização é mantida, pois se comporta de forma satisfatória para diversas situações, como por exemplo, armazenar os dados dos sensores que estão sendo monitorados por cada servidor de borda presente no ambiente ubíquo, onde no repositório é mantida a relação do servidor de borda com o sensor, facilitando o acesso à esta informação como também a sua modificação caso seja necessário;
- modelo de triplas (sujeito, predicado, objeto): empregado para ser utilizado na persistência de ontologias. Com a inserção do modelo de armazenamento de dados no formato de triplas, espera-se prover facilidades similares às experimentadas nos sistemas gerenciadores dos modelos relacionais, onde o acesso e a manipulação das informações presentes nas ontologias é feito de forma automatizada. Exemplos de manipulação seriam a inserção de novas instâncias e a criação de novas classes no modelo;
- modelo não relacional: utilizado tendo como premissa o suporte ao tratamento de grandes volumes de dados. Atualmente, tem surgido um aumento na quantidade e variedade de dados de contexto que devem ser tratados pelo *middleware* EXEHDA, como por exemplo, arquivos de log de aplicações, os quais são analisados visando, dentre outras finalidades, melhorar a segurança do ambiente computacional. Para implementação deste modelo foi escolhida a categoria orientada a documentos, a qual possui como característica ignorar campos vazios, o que otimiza o espaço em disco utilizado, sendo recomendada para o tratamento de dados semiestruturados, os quais possuem uma estrutura irregular, em que não existe um esquema padrão dos dados [Prمود J Sadalage 2013].

Na abordagem proposta fica a cargo da aplicação determinar onde prefere armazenar as informações contextuais, sendo indicada a utilização do modelo relacional para dados que possuem relação entre eles, e assim, aproveitando-se das características do modelo. O modelo de triplas para armazenar as ontologias, de forma a propiciar facilidades para o tratamento das informações presentes nas mesmas. E recomenda-se o armazenamento no modelo não relacional quando é necessário o tratamento de grande volume de dados e/ou ainda, quando os dados armazenados possuem variedade de formatos.

3.2. Camada de Serviço

A camada de serviço desenvolvida para o RHIC é responsável pela disponibilização de métodos que permitem a busca, inserção e exclusão de informações contextuais na camada de armazenamento.

As consultas realizadas na camada de armazenamento podem ser executadas diretamente pela interface web disponibilizada para o gerenciamento da EXEHDA-HM, onde podem ser visualizadas as informações presentes no repositório, ou utilizando o mecanismo de gerenciamento das informações contextuais, o qual permite o acesso às informações de forma que as aplicações possam realizar as consultas em nível de programação utilizando o método de consultas oferecido pelo mecanismo na linguagem Java.

As ontologias são processadas antes de serem armazenadas no repositório de triplas, desta forma, no momento em que as aplicações desejarem consultar as informações presentes no modelo ontológico, o processo de inferência já estará realizado, podendo resultar em uma melhora do tempo de resposta do sistema, já que o processo de inferência pode se tornar lento dependendo do tamanho da ontologia a ser processada.

3.3. Camada de Processamento

A EXEHDA-HM oferece o suporte ao processamento das informações contextuais utilizando uma estratégia baseada em regras composicionais. Para permitir à aplicação a composição das regras usando dados dos diferentes modelos, as regras definidas devem utilizar *tags* de marcação que comecem com o símbolo “#”, o qual é substituído pela consulta responsável por buscar esta informação no RHIC.

A camada de processamento realiza a execução da regra composicional seguindo um fluxo específico de execução, conforme pode ser visualizado na Figura 2. Primeiramente é buscada a regra desejada no RHIC, após, são identificadas as consultas auxiliares que devem ser realizadas, tendo como intuito buscar os dados necessários no RHIC para substituição das *tags* de marcação presentes na regra buscada. Posteriormente, com o retorno das consultas auxiliares, a regra é composta pelos valores necessários e sua avaliação é realizada, tendo como retorno o comando desejado a ser executado, cujo identificador encontra-se na própria regra. Com isso é necessário buscar o comando desejado no RHIC e, então, executar o comando retornado.

Destaca-se que as regras composicionais podem ser compostas de dados armazenados nos diferentes modelos, tornando a consulta mais expressiva e, em consequência, facilitando à aplicação identificar situações de interesse. Caracteriza-se ainda a melhora na atualização e manutenção das regras criadas, pois caso mais de uma aplicação utilize as mesmas regras só é necessário atualizar seu conteúdo em um determinado campo e todas estarão atualizadas. Além disso, os comandos que determinam as ações decorrentes das situações não são definidos nas regras, mas em uma tabela do RHIC, o que facilita a reutilização e manutenção dos mesmos.

4. Cenário de Uso

Durante os últimos anos, o Departamento de Segurança da Informação e Comunicações (DSIC) da Presidência da República do Brasil estabeleceu diversos decretos relacionados à Segurança da Informação e Comunicações no âmbito da Administração Pública

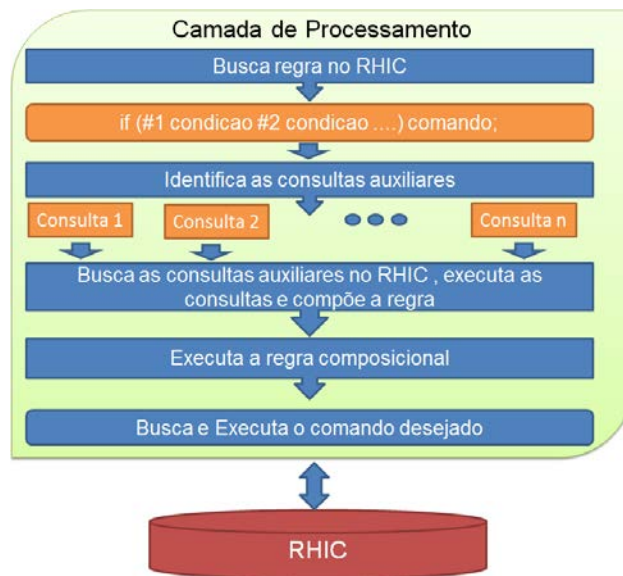


Figura 2. Fluxo de execução da regra composicional

Federal (APF). Além disso, o DSIC desenvolveu algumas normas complementares com diretrizes para ações a serem desenvolvidas pela APF, sendo que algumas destas normas referenciam a ABNT NBR ISO/IEC 27001:2013 e a ABNT NBR ISO/IEC 27002:2013.

Tendo isto em vista, que um dos requisitos a ser cumprido pelas instituições Federais é a elaboração da Política de Segurança da Informação e Comunicações (PO-SIC) [BRASIL 2000]. Sabe-se que um dos desafios relacionados à POSIC é a sua implementação e, mais especificamente, a verificação da sua conformidade. Por exemplo, definir como verificar se os usuários estão em conformidade com a política de senhas (documento geralmente derivado da POSIC).

Desta forma, este estudo de caso explora a ISO 27002 que apresenta um código de prática para a gestão da segurança da informação. Mais especificamente, este estudo de caso demonstra a utilização da abordagem EXEHDA-HM empregando o modelo não relacional e o modelo de triplas, onde são combinadas as informações presentes nos dois modelos com a aplicação da estratégia de processamento composicional.

Para implementação deste estudo de caso foi utilizada uma ontologia baseada em [Alcázar and Fenz 2012], a qual é responsável por mapear a estrutura da ISO 27002, incluindo seus controles, ativos, ameaças e vulnerabilidades. Nesta ontologia foram adicionadas novas classes de forma a facilitar o acesso às informações de interesse, sendo que optou-se por adicionar como subclasses as diretrizes relacionadas a “password policy”, tais como, “complexity requirements” e “clear text”. A ontologia foi armazenada no modelo de triplas presente no RHIC.

Como fonte para a aquisição de contexto, foi explorado o monitoramento dos logs do *Web Application Firewall* modsecurity [ModSecurity 2016] que foi devidamente configurado para identificar quebras na política de senhas de aplicações web. Após o tratamento do log é realizada a identificação da situação de interesse caracterizada pela mensagem “password does meet complexity requirements”, e por último a situação identificada é armazenada no modelo não relacional do RHIC, já que as situações possuem

um formato semiestruturado.

Com a situação já identificada e armazenada no RHIC, pode ser utilizada a estratégia de processamento composicional conforme apresentado na Figura 3, com a execução da regra *if(#3 && #4) "b"*. A tag #3 é substituída pela consulta `find(eq("subcategory", "ComplexityRequirements"))` a qual busca no modelo não relacional as situações com o atributo “subcategory” igual a “ComplexityRequirements”. A tag #4 representa a consulta SPARQL [W3C 2008] que é apresentada na Figura 4, a qual busca os controles a serem sugeridos que estão relacionados a ComplexityRequirements de acordo com a ontologia da ISO 27002 e as propriedades `controlStatement`, `implementationGuidance`, `otherInformation` onde estão inseridas as informações a respeito de cada controle que foi sugerido.

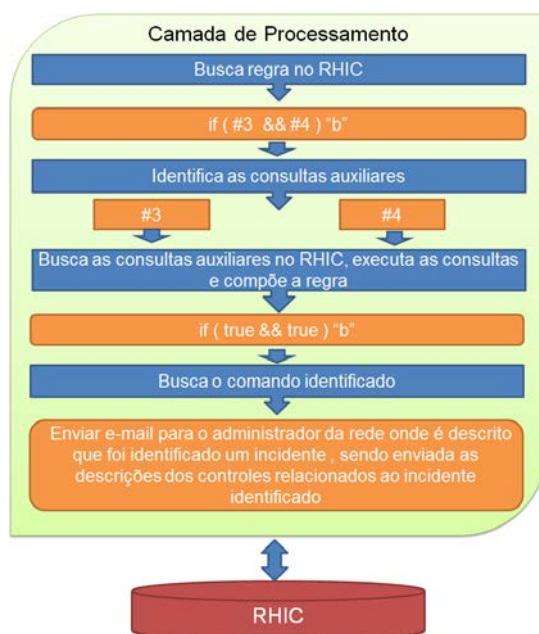


Figura 3. Funcionamento da regra para o reforço de políticas de senha

```

SELECT ?guideline ?controlStatement ?implementationGuidance ?otherInformation
WHERE {
    ?guideline rdfs:subClassOf myonto:ComplexityRequirements.
    ?guideline myonto:controlStatement ?controlStatement.
    ?guideline myonto:implementationGuidance ?implementationGuidance.
    ?guideline myonto:otherInformation ?otherInformation.
}
  
```

Figura 4. Consulta referente à busca pelos controles sugeridos relacionados à situação “complexity requirements”

Como condição a ser executada na regra composicional, foi utilizado o método que verifica se foi retornado algum resultado na consulta, o qual retorna “true”, caso positivo, e “false”, caso não tenha nenhum retorno. O comando previamente configurado para ser executado, caso seja detectada a situação, é representado por “b”, sendo este referente ao envio de e-mail para os analistas de segurança da informação. Este e-mail é composto de

uma mensagem de notificação de incidente, junto aos controles que constam na ontologia da ISO 27002 que servirão como sugestões de melhorias a serem empregadas, sendo os analistas responsáveis por avaliar a necessidade de tomada de ações para reforçar a política de senhas.

5. Trabalhos Relacionados

[Ejigu et al. 2007] propôs um modelo híbrido de contexto utilizando o modelo relacional e o baseado em ontologia para o *middleware* CoCA (*Collaborative Context-Aware Service Platform*), o qual realiza as tarefas de representação, raciocínio, agregação e interpretação de dados de contexto. As ações e decisões são realizadas com base nos dados de contexto adquiridos, e a plataforma suporta a colaboração e o compartilhamento de recursos computacionais entre os dispositivos.

Em [Kotenko et al. 2013] é apresentada uma abordagem de modelo híbrido de contexto para sistemas SIEM (*Security Information and Events Management*) que combina os modelos relacionais, baseados em XML e em ontologias. Os autores justificam que a representação relacional não permite especificar uma consulta parametrizada, por exemplo, com os nomes dos produtos e versões, para a análise de vulnerabilidades e processá-las no programa, o que torna este processo demorado. Com isso, uma solução alternativa na representação de dados em sistemas de processamento de informações com estruturas complexas de dados (tais como sistemas SIEM) é uma abordagem ontológica, o que facilita as expressões de relações complexas entre entidades e possibilita a formulação de conceitos e de relacionamentos envolvidos.

Em [VANATHI 2013] é apresentado um sistema ciente de contexto que baseia-se em uma modelagem híbrida de contexto, a qual combina os modelos objeto relacional e ontológico. O modelo objeto relacional é criado a partir dos dados de contexto estáticos, e são atualizados em tempo de execução com os novos dados de contexto capturados. O modelo baseado em ontologia *Context-Onto* é composto de três componentes básicos: o esquema da ontologia, dados de contexto estático e dados de contexto dinâmicos. O esquema da ontologia representa as hierarquias de conceito, conceito de relações, axiomas e informações de metadados, que representam o conhecimento.

Analisando a etapa de processamento dos trabalhos relacionados, apesar de utilizarem diferentes estratégias para modelagem de contexto, o acesso aos modelos é realizado de forma individual, não possibilitando a combinação das informações presentes nos diferentes modelos. Por outro lado, a EXEHDA-HM foi desenvolvida de forma a prover uma estratégia de regras composicionais que permita à aplicação combinar os contextos de diferentes modelos, viabilizando a utilização de forma combinada das informações contextuais.

Destaca-se ainda a utilização de diferentes formas de armazenamento da EXEHDA-HM, a qual possibilita explorar as vantagens de cada um dos três modelos de armazenamento: relacional; de triplas; e não relacional.

6. Considerações Finais

Neste trabalho foram tratados os desafios enfrentados pelas aplicações cientes de contexto com o emprego de modelos híbridos, onde destaca-se a ausência de uma estratégia

que permita utilizar os contextos presentes nos diferentes modelos utilizados. Com estes desafios em vista, as seguintes contribuições foram alcançadas com o desenvolvimento deste trabalho: (i) a concepção de uma abordagem com suporte a modelagem híbrida de contexto; (ii) uma camada de armazenamento que provê a persistência em três modelos distintos; (iii) uma camada de serviço que permite o acesso as informações contextuais por parte das aplicações cientes de contexto; e (iv) a concepção de uma estratégia de processamento composicional de contexto, a qual possui como diferencial a possibilidade de combinar dados presentes nos diferentes modelos de armazenamento utilizados.

Por meio da EXEHDA-HM foi realizada a concepção de um novo repositório para armazenamento de informações contextuais, o RHIC, dividido em duas camadas:

- camada de serviço: responsável por gerenciar as informações contextuais presentes na camada de armazenamento. Esta camada disponibilizou métodos para consulta, inserção e remoção de dados na camada de armazenamento. Foram disponibilizados métodos para tratar as informações contextuais nos três modelos de armazenamento presentes no RHIC;
- camada de armazenamento: distinguiu-se pela utilização de três modelos para o armazenamento de informações contextuais tratadas pelo *middleware* EXEHDA: relacional, modelo que vinha sendo utilizado pelo *middleware*; não relacional, ofereceu o suporte ao armazenamento eficiente para dados semiestruturados; e de triplas, que foi escolhido para realizar o armazenamento das informações contextuais presentes nas ontologias, oferecendo mecanismos que facilitam o acesso a estas informações.

Além da criação do RHIC, a abordagem EXEHDA-HM possui como principal contribuição a camada de processamento desenvolvida, onde se utiliza de uma estratégia baseada em regras composicionais, permitindo combinar nas regras criadas dados de diferentes modelos de armazenamento. Com a utilização da abordagem EXEHDA-HM aumenta-se a flexibilidade e a possibilidade de expressividade da aplicação, permitindo que sejam utilizados em uma mesma regra dados armazenados nos diferentes modelos, melhorando a identificação das situações de interesse das aplicações.

O estudo de caso explorou a aplicação da EXEHDA-HM em um escopo de atual relevância devido ao aumento das preocupações com segurança da informação nas instituições. Neste cenário foi explorado um problema comum e de considerável risco as instituições, onde a solução proposta possibilitou a integração das situações armazenadas no modelo não-relacional com as informações presentes na ontologia da ISO 27002, tendo como resultado a indicação de controles de segurança aos analistas.

Dentre os aspectos levantados para continuidade do trabalho, podem ser citados: (i) desenvolver interfaces que utilizem técnicas de visualização de dados, com o intuito de facilitar a interpretação das informações contextuais armazenadas e das situações identificadas; (ii) aplicar a abordagem EXEHDA-HM em diferentes cenários de utilização, por exemplo, em ambientes educacionais, os quais costumam utilizar tanto os modelos baseados em ontologias quanto os modelos relacionais.

Referências

Alcázar, F. and Fenz, S. (2012). Mapping iso 27002 into security ontology. *Treball final de grau – Universitat Politècnica de Catalunya. Escola d’Enginyeria de Telecomunicació*

i Aeroespacial de Castelldefels.

- Bauer, J. S., Newman, M. W., and Kientz, J. A. (2014). Thinking about context: Design practices for information architecture with context-aware systems. In *iConference 2014 Proceedings*, pages 398–411.
- Bellavista, P., Corradi, A., Fanelli, M., and Foschini, L. (2012). A survey of context data distribution for mobile ubiquitous systems. *ACM Comput. Surv.*, 44(4):24:1–24:45.
- BRASIL (2000). Decreto lei nº 3505 – institui a política de segurança da informação nos órgãos e entidades da administração pública federal. *Diário Oficial da União*.
- Carvalho, A. G. (2014). Interface nosql integrada a banco relacional para gerenciamento de dados em nuvem privada. Monografia bacharelado em engenharia da computação, Centro Universitário de Brasília Faculdade de Tecnologia e Ciências Sociais Aplicadas.
- Dey, A. K. (2001). Understanding and using context. *Personal Ubiquitous Computing*, 5(1):4–7.
- Ejigu, D., Scuturici, M., and Brunie, L. (2007). Coca: A collaborative context-aware service platform for pervasive computing. In *Information Technology, 2007. ITNG '07. Fourth International Conference on*, pages 297–302.
- Knappmeyer, M., Kiani, S., Reetz, E., Baker, N., and Tonjes, R. (2013). Survey of context provisioning middleware. *Communications Surveys Tutorials, IEEE*, 15(3):1492–1519.
- Kotenko, I., Polubelova, O., Chechulin, A., and Saenko, I. (2013). Design and implementation of a hybrid ontological-relational data repository for siem systems. *Future Internet - Special Issue Security of Systems and Software Resiliency*.
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014). A middleware architecture for dynamic adaptation in ubiquitous computing. *Journal of Universal Computer Science*, 20(9):1327–1351. http://www.jucs.org/jucs_20_9/a_middleware_architecture_for.
- ModSecurity (2016). Acesso em 26 de maio de 2016. Disponível em: <<https://www.modsecurity.org/>>.
- Neto, A., Azevedo, B., Bouefleuer, R., Lima, J., and Augustin, I. (2015). Modelo para predição de ações e inferência de situações de risco em ambientes sensíveis ao contexto. In *Simpósio Brasileiro de Computação Ubíqua e Pervasiva*.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *Communications Surveys Tutorials, IEEE*, 16(1):414–454.
- Pramod J Sadalage, M. F. (2013). *NoSQL Essencial, Um Guia Conciso para o Mundo Emergente da Persistência Poliglota*. Novatec.
- VANATHI, B. (2013). *Context Management Using Ontology and Object Relational Database Management System*. Doctor of philosophy, Faculty of Information and Communication Engineering Anna University, Chennai.
- W3C (2008). Acesso em 26 de maio de 2016. Disponível em: <<https://www.w3.org/TR/rdf-sparql-query/>>.

From a Smart House to a Connected City: Connecting Devices Services Everywhere

João Paulo Cardoso de Lima¹, Leandro Buss Becker², Frank Siqueira³,
Analucia Schiaffino Morales¹, Gustavo Medeiros de Araujo¹

¹Department of Computing
Federal University of Santa Catarina (UFSC)
Araranguá, SC – Brazil

²Department of Automation and System
Federal University of Santa Catarina (UFSC)
Florianópolis, SC – Brazil.

³Department of Informatic and Statistics
Federal University of Santa Catarina (UFSC)
Florianópolis, SC – Brazil
joao.pcl@grad.ufsc.br,

{leandro.becker, frank.siqueira, analucia.morales, gustavo.araujo}@ufsc.br

Abstract. *The growing development of smart devices makes it possible to create new distributed applications targeted for smart spaces. The design of intelligent spaces assumes that there is an infrastructure to support the applications requirements. Many academic works have proposed middlewares that provide an abstraction for the use of network services. The network services of a smart space, such as an automated home, can have different communications interfaces. Accordingly, we developed a middleware called UDP4US (Universal Device Pipe for Ubiquitous Services) which was designed to abstract different patterns of communication, keeping the discovery of devices on a local network services. In this paper, we present a new UDP4US architecture component that aims to expose the local network devices services to the Internet. The new component was developed with the REST technology, thus the devices services can be discovered and accessed over the Internet. The new component was exhaustively tested in order to find the limits of its effectiveness. The evaluation of the new component was performed by measuring its discovery and execution times plus the success rate of the services execution exposed over the Internet. The results from the present work are important to guide a better design of distributed applications for smart places.*

1. Introduction

Ubiquitous computing requires binding inexpensive computing device to larger computers and others resources using, preferably, wireless medium. For example, a smart house controlled with smart devices shall have remote control of home lighting, integrated entertainment systems, messaging services, and should be able to monitor the health conditions of the house inhabitants. Mark Weiser in [Weiser 1991] proposed the idea of injecting computation capacity into the physical world with high spatial density and invisibility, by having nodes and collectives of nodes operating autonomously, i.e., the technology

should be part of everyday life until they are indistinguishable from it. In general, the ubiquitous computing devices are composed of hardware platforms, operating systems, network protocols, interaction substrates, applications, privacy, and computational methods. The main concern in this area has been the integration of the digital artifacts with the physical world [Kumar 2009].

In the context of ubiquitous computing, devices interaction has been traditionally addressed within a local network [Gubbi et al. 2013]. However, nowadays, this concept is extended in a way that devices can be reached through the Internet. This new concept is known as the Internet-of-Things (IoT) [Atzori et al. 2010]. A wide variety of devices is available as a consequence, a great amount of different communication interfaces are under use. This is, in fact, a barrier to provide proper interaction between different devices. To solve this problem, efforts have been made to describe protocols that allow different classes of devices to operate together.

Some protocols have been employed in the ubiquitous computing, such as DPWS (Device Profile for Web Service)¹ which is designed to simplify the integration between web services and device-provided services. DPWS brought SOA concepts to the device level [Cândido et al. 2010], allowing devices to expose their services. This enables a wide applicability, ranging from industrial to home automation [Sleman and Moeller 2008]. The DLNA (Digital Living Network Alliance)² is a collaborative standard organization that defines interoperability guidelines to enable devices to communicate in a home network. The DLNA aims the sharing of multimedia content between producer and consumer devices. However, the DLNA is not limited to sharing multimedia content, as it can be used with the same purpose as DPWS. Given that both DPWS and DLNA use UPnP as underlying protocol, they can be combined in order to allow interoperation between heterogeneous devices and the development of several application scenarios.

Additionally, the REST (REpresentational State Transfer) technology has been widely used to integrate systems. The REST is an architectural style [Fielding 2000], which is being adopted in the context of Web-based services and applications as a strategy to provide interoperation system. The REST technology can also be useful in the context of ubiquitous systems, due to its lightweight and stateless nature, REST can be seen as a suitable solution for device integration [Guinard et al. 2010, Riedel et al. 2010, Moritz et al. 2010].

Firstly, this paper presents the Universal Device Pipe for Providing Ubiquitous Services (UDP4US) [Felisberto et al. 2015] emphasizing its new component. The new component is the REST module, which is the main contribution presented in this paper. The REST component is responsible for exposing and execution device's services over the Internet. The remainder of this paper is structured as follows. Section 2 outlines some issues about REST technologies. Section 3 presents the most relevant related work in this field. Section 4 provides an overview of the proposed UDP4US and its components. Section 5 outlines the implementation of UDP4US REST module and the experimental testbed employed. Finally, Section 6 presents the conclusions and the future work ideas.

¹<http://docs.oasis-open.org/ws-dd/ns/dpws/2009/01>

²<http://www.dlna.org/>

2. REST Technology

The REST was proposed by [Fielding 2000] aiming to simplify the Hypermedia Distributed Architecture. His work introduces the term REST, which stands for Representational State Transfer. REST is an architectural style that became an alternative to the standard Web Service technology. The main difference between traditional Web Services and REST is that the former is based on the SOAP protocol, which demands a more complex protocol stack than REST. REST is based just on HTTP operations, which scale better because of the independence from the SOAP protocol. The simplicity of REST makes it an ideal candidate to build an “universal” API for embedded devices. This concept is often referred to as “Web of things” [Guinard et al. 2010]. The main features of REST are the use of URIs and of resource entities. Four HTTP methods can be employed to provide services, as shown in figure 1. These methods are described as follows:

- *GET*: Requests data from a specific resource. It is defined as a safe method and should not be used to trigger an action.
- *POST*: The information sent in the request body is used to create a new resource instance.
- *DELETE*: Remove a resource instance. Should return the 204 status if there is no resource associated with the specified URI.
- *PUT*: Updates a resource at the specified URI. If the resource does not exist, it does create one. The main difference between POST and PUT is that the former can deal not only with resources, but can also process information.

RESTful services can exchange messages using several data formats, such as XML, JSON and HTML. The use of the HTTP protocol allows it to easily cross network boundaries. Since RESTful services have a stateless behavior, the overall system scalability is highly improved.

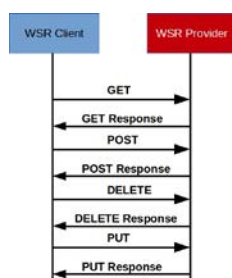


Figure 1. Message Exchanges in REST

3. Related Works

Some related works address devices integration at the protocol level, using DPWS or DLNA as a part of the proposed solution. For instance, in [Samaras et al. 2013] it is proposed a DPWS modified stack to be applied to 6LoWPAN-Based Wireless Sensor Networks (WSN). Therefore, the DPWS implementation was enhanced in order to consume less memory and decrease the processing overhead. The work shows the advantages to add SOA concepts into WSN by applying DPWS, in spite of increasing the message overhead.

A different integration approach was presented in [Dohndorf et al. 2010]. The proposal was to integrate DPWS with the OSGi platform. Therefore, all DPWS devices could be seen and controlled over the Internet, instead of being managed just in a local network. The authors intention was to place devices into the internet of things.

Some works have shown the role that REST plays on the integration of IoT systems that uses DPWS standard [Han et al. 2015, Guinard et al. 2010, Riedel et al. 2010, Moritz et al. 2010]. For instance, in [Han et al. 2015] it is presented an extension for DPWS using a REST proxy for IoT applications. Although REST or HTTP-based implementations cannot support event-driven models as DPWS, the work has shown improvements in latency and overhead, as well as simplifying the global topology.

In [Riedel et al. 2010] it is presented an approach for handling the task of integrating multiple concurrent IoT systems using automatic generation of DPWS gateways for sensor nodes. The study has shown that Web service gateways have great scalability in local area networks, besides being flexible enough to support Industrial applications.

Since SOA and REST are not conflicting, combining them, provides features to meet requirements of single applications (LAN) and IoT scenarios (WAN) for smart cooperating objects. Therefore, in [Moritz et al. 2010] it is discussed the drawbacks of using DPWS along with REST, which may require more resources and implementation efforts to carry out eventing and discovery features.

4. Universal Device Pipe for Providing Ubiquitous Services

Considering an ubiquitous environment, one can imagine several services being provided. However, making service composition using current SOA facilities is quite complicated for several reasons. The most difficult is the wide variety of technologies available to describe, publish, and compose services. Depending on the target application, different protocols and mechanisms should be used to provide the required services.

In order to illustrate this problem, let us suppose an application scenario that adopts three different technologies, such as: DPWS, DLNA, and REST. Consider a person located in a public transportation vehicle (bus, train, etc). To make good use of its time, the person can watch a movie or listen to the radio using his/her tablet or smartphone along the journey. A given software running in the tablet or smartphone could keep updating the person location to the smart home system (e.g. by REST) so that it could forecast the person arrival time in order to properly maintain a comfortable temperature in the house. When the person arrives at home, he/she might like to continue watching the video on the TV (without having to make setups). At the moment that the smart home system realizes that the person is located in the living room it would transfer the video from the tablet to the TV in a synchronized manner (e.g. by DLNA). The identification of the person location inside the house could be done, for example, using a DPWS enabled RFID reader.

However, performing such operations using current technologies is a difficult task (not to say impossible). These protocols are incompatible with each other, making it hard to perform communication between heterogeneous devices. In order to provide a contribution for solving this problem, we have analyzed which would be the requirements necessary to allow the integration of heterogeneous devices, as follows:

- The first requirement is to expose the device and its services on the network. Each device has a specific metadata which describes its features, such as manufacturer, model, version, and details about its services. For instance, a sensor node that senses ambient temperature could provide the service *getCurrentTemperature*. This service should also be detailed, with all the necessary parameters, to allow the interaction with the client.
- The second requirement is to discover the device and its services. The clients should easily request any device on the network, regardless if the device is within a local or a remote network. Likewise, the devices should reply to discovery requests sent by clients.
- The third requirement is to perform service requests. With the metadata device, the clients must be able to run the service without the knowledge of the specific communication interface from the device.

In order to provide a solution for accomplishing the mentioned requirements, we propose UDP4US (Universal Device Pipe for Ubiquitous Services). The UDP4US was designed to be an integration channel at application level. The UDP4US works as a pipe that connects the different technological solutions, such as DPWS, DLNA, and REST. The details about UDP4US architecture and its operations are described next.

4.1. UDP4US Architecture

The UDP4US was designed to abstract services provided in different network protocols. It plays the role of a *middleware* on top of the supported protocols, allowing to provide abstract services and properly transporting messages. UDP4US provides a specific communication interface for each supported protocol, as illustrated in figure 2. This is provided by the *Service Broker* component, which in fact is a service with a published operation. The Service Broker hides the complexity to deal with incoming requests from different protocols. It listens the requests and translates them through the pipe. Currently, the Service Broker can handle requests from DPWS, DLNA, and RESTful clients.

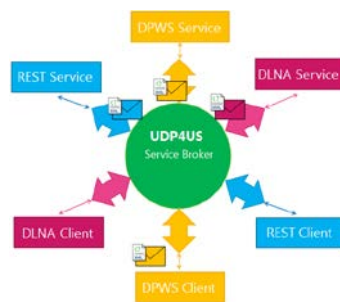


Figure 2. The Service Broker

The UDP4US stack protocol is presented in figure 3. From top-down, the topmost layer is the Service Broker Integrator (SBI). The SBI is responsible for implementing interfaces for each related protocol, in order to translate the requests. The SBI has a device for each supported protocol. For instance, the SBI has a device for DPWS that can be reached by any DPWS enable device, but cannot be directly reached by DLNA or REST enabled devices. Therefore, if a DLNA-enabled device needs to discover a

device or perform a service that is not DLNA-enabled, but that is a DPWS-enabled device, the SBI has a DLNA-enabled device that listens to requests and translates to the DPWS protocol. The devices implemented in the SBI are in fact virtual devices that simply listen to incoming requests.

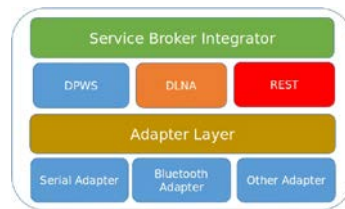


Figure 3. Proposed Protocol Stack

Below SBI there are the supported application protocols: DPWS, DLNA, and REST. The DPWS and DLNA engines are responsible for discovering and processing their requests independently. The REST interface plays the role of exposing DPWS and DLNA devices over the Internet.

The next layer is the Adapter Layer, which can be seen as a driver for the operating system. It acts like any off-the-shelf device that must implement a specific device-driver to support a given operating system. For devices that are neither DPWS nor DLNA enabled, different interfaces can be used. Currently it provides interfaces for serial communication and Bluetooth. Such requests are later translated for DPWS or DLNA. Our proposal was designed to be extended to others Adapter Layers, such as 802.15.4.

4.2. UDP4US Operation

The UDP4US can act as a Client Device and a Provider Device at the same time. For instance, if a Service Provider is not reachable using a given protocol, such request is translated to one of the other supported protocols. For example, if a DLNA client does not get an answer for a request, UDP4US translates the request to DPWS or REST in order to try to find a service. Furthermore, if the device provider is neither DLNA nor DPWS enabled, the Adapter Layer plays the role of a request translator. It is important to highlight that the requests could be a discovery or a service invocation. For example, if a Bluetooth device needs to interact with a serial device, i.e., both are not DPWS- nor DLNA-enabled devices, the requests must be translated to both technologies.

5. Experimental Evaluation

In order to validate the new component from UDP4US project, some experiments were performed to measure the bounds of the REST component. The experimental setup is illustrated in figure 4. From the right side, the DPWS-enabled client is requesting a service to its local network. However, the requested service was not presented in the same local network. Then the UDP4US listening the requests, translates to REST component and it dispatches over the internet. On the left side, a Service Provider DPWS enabled was listening incoming request from its local network. All incoming requests from internet were managed by UDP4US middleware. The incoming requests from the internet are listened by REST component, which gets the request and translate to UDP4US format.

The Service Broker component will translate the request to the specific format, which was in this case to DPWS metadata. Finally, the Service Provide DPWS enabled can process the request and reply to the DPWS-enabled client.

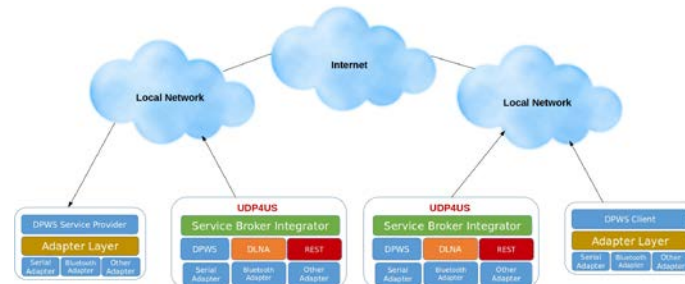


Figure 4. Experimental Setup Over the Internet

The main goal of the experiments were to find the boundaries of this new UDP4US component. Finding the limits of our approach might help to better guide the applications design, with a clear understanding of the viability the applying the concept of IoT. We performed the experiments with a high number of clients requesting one device's service. The number of clients ranged from 50 to 500, increasing with a step of 50, which means that UDP4US must handle with a high number of requests. The client and service provider were hosted in different networks. In order to reduce the statistical bias, each obtained result refers to the average of 10 executions. Furthermore, all experiments were carried out with clients performing discovery and execution of a service. The service execution performed by means of two-way requesting, such as a system requesting a value from a sensor. In the last experiment, the goal was to measure the cost of the solution over the internet, that way we have a sense of performance on IoT solutions. To oversee the performance analysis from all protocols, three metrics were chosen: i) number of successfully executed requests, ii) service discovery time, and iii) service execution time.

5.1. Implementation

The UDP4US prototype implementation was built on top of tree frameworks, the WS4D-JMEDS³, Cling⁴, and REST services, which were hosted by an Apache Server. All frameworks were developed using the Java SE platform. All source code generated in this project is publicly available in the github repository⁵. The equipment used in all experiments are described as follow:

- **Service Provider Side:**
 - *Service Provider* A notebook with Intel(R)Pentium(R) CPU P6100 @ 2.00 GHz, 2 GB of RAM and Ubuntu 13.10 Operating System hosting the service provider. An Arduino UNO R3 was attached to it by USB.
 - *UDP4US* A notebook with Intel(R) Core(TM) i3-M330 @ 2.13GHz, RAM 4 GB of RAM and Windows 10 hosting the REST Service Provider.
- **Client Side:**

³<http://ws4d.e-technik.uni-rostock.de/>, DPWS Implementation

⁴<http://4thline.org/projects/cling/>, DLNA Implementation

⁵<https://github.com/UFSCAraSistemasUbiquos>

- *UDP4US* A notebook with Intel(R) Core(TM) i5-2410M CPU @ 2.30 GHz, 6 GB of RAM and MS Windows 7 OS hosting the REST Client.
- *Client* A notebook Intel(R)Core(R) CPU i3-370M @ 2.53 GHz, 4 GB of RAM and Arch Linux Operating System hosting the client DPWS enabled.

5.2. Obtained Results

This section presents the results obtained in the performance experiments, where a set of REST Clients requested a virtual device hosted on DPWS through UDP4US. Figure 5 illustrates the Number of Requests successfully issued over. We can observe that the amount of successful requests during 500 concurrent requests was nearly 84%. Although the tests have been executed in extreme conditions, one can say that UDP4US can deal with 500 concurrent requests over the Internet.

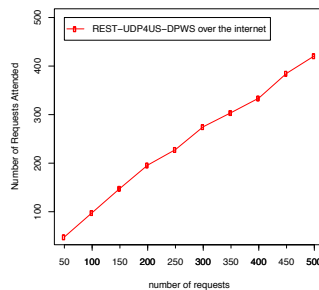


Figure 5. Number of Requests Attended Over the Internet

In figure 6 it is shown the Service Execution Time over the Internet. As expected, the response time was higher compared with requests on the local network [Felisberto et al. 2015]. This due to the fact that the Internet has a higher latency than the local network. The Service Discovery Time is also shown in figure 6. The Service Discovery Time follows a similar behavior of the Service Execution Time, also due to the high traffic over the Internet.

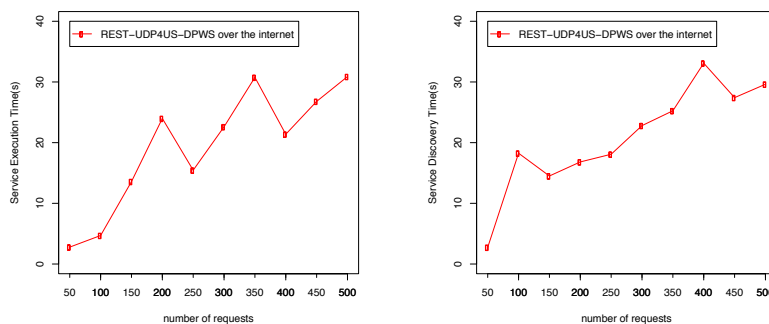


Figure 6. Service Discovery and Execution Times Over Internet

The first observation we pointed out is that, for all experiments, as the number of clients increases, the execution and discovery times also increases. In addition, the number of successfully issued requests decreases. The behavior was a bit irregular, because

the requests do not have any QoS schema in order to prioritize network traffic. Then, all requests compete for the access to the network. In spite of the high response time of the requests, the experiments have clarified the boundaries of response time and the number of possible concurrent requests for IoT applications using REST and DPWS technologies.

6. Conclusions and Future Works

A lot of research has been made to provide a baseline protocol to connect heterogeneous devices. Such protocols can be applied in different market segments such as home automation, industries, health care and many others. Most of automation segment aims the use of a local network to expose device's services. Our previous work complies with the requirements to integrate multiple protocols such as DPWS, DLNA, and Bluetooth. In this paper, we presented the REST module for the UDP4US architecture. The REST module is responsible to expose the local network services to the internet. Now the UDP4US architecture brought new possibilities to distributed applications for smart spaces. A smart space such as a house can be connect to other internet services, such as public transport services. The new opportunities for IoT applications can extend from a smart house to a connected city.

To make this new trend possible, it is important to establish the limits of the protocols in order to better guide the implementation of each application scenario. The presented work shows that the performance of the proposed solution has been satisfactory, since the tests were performed with the intent of applying the maximum stress. Besides, the integration was successful and allowed the different devices attached to incompatibles protocols to reach each other and exchanged messages over the internet.

In a near future, we intend to extend the UDP4US architecture to provide a QoS schema for client requests. The QoS will be provided by a M,K-Firm scheduler, that will prioritize the requests that will be near to dynamic failure [Hamdaoui and Ramanathan 1995]. Furthermore, others adapter layers can also be extended, such as an adapter for IEEE 802.15.4 technology.

References

- Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15):2787–2805.
- Cândido, G., Jammes, F., de Oliveira, J. B., and Colombo, A. W. (2010). Soa at device level in the industrial domain: Assessment of opc ua and dpws specifications. In *Industrial Informatics (INDIN), 2010 8th IEEE International Conference on*, pages 598–603. IEEE.
- Dohndorf, O., Kruger, J., Krumm, H., Fiehe, C., Litvina, A., Luck, I., and Stewing, F.-J. (2010). Towards the web of things: Using dpws to bridge isolated osgi platforms. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*, pages 720–725.
- Felisberto, T. Z., Tramontin, E. D., da Cunha dos Santos, F., Morales, A. S., Siqueira, F., and de Araújo, G. M. (2015). Universal device pipe for ubiquitous services. In *SBESC Brazilian Symposium on Computing Systems Engineering*. IEEE.

- Fielding, R. (2000). Representational state transfer. *Architectural Styles and the Design of Network-based Software Architecture*, pages 76–85.
- Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660.
- Guinard, D., Trifa, V., Karnouskos, S., Spiess, P., and Savio, D. (2010). Interacting with the SOA-based internet of things: Discovery, query, selection, and on-demand provisioning of web services. *IEEE Transactions on Services Computing*, 3(3):223–235.
- Hamdaoui, M. and Ramanathan, P. (1995). A dynamic priority assignment technique for streams with (m, k)-firm deadlines. *Computers, IEEE Transactions on*, 44(12):1443–1451.
- Han, S. N., Park, S., Lee, G. M., and Crespi, N. (2015). Extending the Devices Profile for Web Services Standard Using a REST Proxy. *Internet Computing, IEEE*, 19(1):10–17.
- Kumar, S. (2009). Challenges for ubiquitous computing. In *Networking and Services, 2009. ICNS '09. Fifth International Conference on*, pages 526–535.
- Moritz, G., Zeeb, E., Prüter, S., Golasowski, F., Timmermann, D., and Stoll, R. (2010). Devices profile for web services and the REST. *IEEE International Conference on Industrial Informatics (INDIN)*, pages 584–591.
- Riedel, T., Fantana, N., Genaid, A., Yordanov, D., Schmidtke, H. V., and Beigl, M. (2010). Using web service gateways and code generation for sustainable iot system development. In *Internet of Things (IOT), 2010*, pages 1–8. IEEE.
- Samaras, I., Hassapis, G., and Gialelis, J. (2013). A modified dpws protocol stack for 6lowpan-based wireless sensor networks. *Industrial Informatics, IEEE Transactions on*, 9(1):209–217.
- Sleman, A. and Moeller, R. (2008). Integration of wireless sensor network services into other home and industrial networks; using device profile for web services (dpws). In *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*, pages 1–5. IEEE.
- Weiser, M. (1991). The computer for the 21st century. *Scientific american*, 265(3):94–104.

Internet das Coisas Aplicada nas Redes Elétricas Inteligentes Brasileiras*

Efrem E. O. Lousada¹, Raquel A. F. Mini²

¹Instituto de Informática – Instituto Federal do Norte de Minas Gerais (IFNMG)
Pirapora – MG – Brasil

²Departamento de Ciência da Computação
Pontifícia Universidade Católica de Minas Gerais – Belo Horizonte – Brasil

efrem.lousada@ifnmg.edu.br, raquelmini@pucminas.br

Abstract. *Advances in the areas of embedded systems, computing and networking have created thousands of heterogeneous devices. Electronic electricity meters are examples of these devices that interconnected can be part of a larger network called Smart Grids. Applying the concept of Internet of Things in these networks, meters should provide the collected information to all system users through servers in the cloud. This work specifies and implements a Brazilian smart meters network and proposes a technique for calculating the duty cycle of the network dynamically in order to maximize the network lifetime. Experimental results show that the proposed solution, using dynamic duty cycle, increases 17 times the network lifetime.*

Resumo. *Avanços nas áreas de sistemas embarcados, computação e redes têm criado milhares de dispositivos heterogêneos. Os medidores eletrônicos de energia elétrica são exemplos desses dispositivos que interconectados podem fazer parte de uma rede maior chamada de redes elétricas inteligentes. A partir da aplicação de internet das coisas nessas redes, os medidores passam a disponibilizar as informações coletadas a todos os usuários do sistema através de servidores na nuvem. Este trabalho especifica e implementa uma rede de medidores inteligentes brasileiros além de propor uma técnica para cálculo do ciclo de trabalho da rede de forma dinâmica com o objetivo de maximizar o tempo de vida. Resultados experimentais mostram que a proposta de utilização de ciclo de trabalho dinâmico aumenta em 17 vezes o tempo de vida da rede.*

1. Introdução

O crescente desenvolvimento das aplicações embarcadas tem proporcionado inteligência computacional a diversos tipos de dispositivos. Os medidores de energia elétrica, por exemplo, são microprocessados e deixaram de registrar somente o consumo de energia do consumidor. A possibilidade de comunicação remota com os medidores de energia elétrica e a criação de redes entre eles foram os primeiros passos para a criação das redes elétricas inteligentes ou *smartgrids*. Utilizando o conceito de geração e compartilhamento de informações relevantes para a sociedade, a partir de aparelhos computacionais interligados, chamado de Internet das Coisas (IoT - *Internet of Things*), a aplicabilidade das

*Este trabalho foi parcialmente financiado pela empresa Nansen S/A, CNPq e Fapemig.

redes elétricas inteligentes pode ser expandida. As informações e serviços gerados pelos medidores eletrônicos de energia elétrica podem ser compartilhados com outros dispositivos ou pessoas interessadas nas informações [Bello and Zeadally 2014].

O grande desafio no desenvolvimento de uma rede inteligente de medidores eletrônicos de energia elétrica é atender à classe de medidores de BT (Baixa Tensão) e as normas que a define. Nessa classe, os recursos de hardware são limitados o que diminui as possibilidades de soluções. As fontes desses medidores conseguem prover no máximo 30 mA e qualquer solução implantada no medidor não poderá consumir mais do que isso. A rede deve ser capaz de funcionar mesmo sem a alimentação direta da rede elétrica, pois os nós devem enviar alarmes de falta de energia para as concessionárias atuarem de forma mais assertiva na solução do problema, atendendo às aplicações de IoT. Por isso, a alimentação por baterias é necessária quando faltar energia à rede de medidores.

Devido a utilização de baterias, é importante a aplicação de técnicas para maximizar o tempo de vida da rede. Entende-se como tempo de vida o tempo até a morte do primeiro nó. Uma das formas mais eficientes para maximizar o tempo de vida é desligar parte dos nós que não estão sendo utilizadas. Atualmente, os fabricantes de soluções para medição devem manter os parâmetros e registradores de seus aparelhos por no mínimo 120 horas, quando estão sem energia e alimentados por baterias, conforme norma ABNT 14519. Nesse caso, é necessário que o ciclo de trabalho seja o menor possível para aumentar o tempo de vida da rede [ABNT 2011]. A norma ABNT 14522 é extremamente rígida quanto ao protocolo de comunicação utilizada pelos medidores e soluções para coleta dos dados de medição. Comandos não podem sofrer qualquer alteração durante a transmissão. A agregação de dados, por exemplo, não é prevista e não pode ser utilizada [ABNT 2008].

O objetivo deste trabalho é projetar e desenvolver uma rede sem fio de medidores de energia que colete de forma automática e simples as informações de medições de consumo e alarmes, disponibilizando em um servidor para acesso externo, atendendo as regulamentações brasileiras. Além disso, a solução proposta realiza de forma dinâmica o cálculo do ciclo de trabalho de acordo com a formação e alterações na topologia, atendendo ao tempo mínimo de 120 horas de vida da rede, quando alimentada por baterias. Avaliações comprovaram um aumento de 17 vezes em relação a pilha Zigbee original.

O restante deste trabalho está organizado da seguinte forma: na seção 2, são apresentados os principais trabalhos relacionados à IoT aplicada as redes elétricas inteligentes. A seção 3 apresenta a descrição da solução proposta, especificando hardware, software e a técnica para cálculo dinâmico do ciclo de trabalho. A seção 4 apresenta as avaliações e resultados experimentais da pesquisa e, por último, na seção 5, são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Na medição de energia elétrica, os primeiros trabalhos relacionando IoT e as *smartgrids* focaram em entender as características das duas tecnologias e as possibilidades da interação entre elas. Em [Yun and Yuxin 2010], os autores analisam as tecnologias e apontam a necessidade de uma comunicação nos dois sentidos pelas *smartgrids* para que os serviços atendam aos requisitos da IoT.

Em [Chen et al. 2011], os serviços e os benefícios que as *smartgrids* podem prover foram especificadas considerando IoT. Segundo os autores, a automação da coleta de

dados de faturamento permitirá aos usuários um maior controle dos gastos com energia elétrica. Aplicações de automação residencial poderão ser incorporadas às redes elétricas inteligente. As linhas de transmissão podem ser monitoradas em tempo real e os veículos elétricos poderão ser monitorados com maior assertividade.

No trabalho desenvolvido em [Arif et al. 2013], os autores projetam e implementam um protótipo de medidor inteligente que utiliza GSM (*Global System for Mobile Communications*) e ZigBee para criar uma aplicação de *smartgrids* na Arábia Saudita. O medidor aproveita o amplo uso da rede GSM, com a sua capacidade de envio de SMS (*Short Message Service*), e da tecnologia ZigBee para enviar e coletar os dados de consumo de energia. No mesmo ano, a empresa de energia elétrica sueca Vattenfall apresentou um projeto para monitoramento em tempo real de interrupções de energia e eventos de qualidade de energia, descrito em [Garpetun 2013], com o objetivo de aprimorar o atendimento a seus consumidores. Leituras entre 1 e 10 minutos são realizadas e, caso um medidor não responda, o sistema considera que o mesmo está sem energia elétrica.

As tecnologias de comunicação para medidores eletrônicos de energia elétrica aplicadas em IoT são apresentadas em [Arnachalam et al. 2014]. Os autores comparam as tecnologias considerando frequência utilizada, taxa de transmissão, área de cobertura, aplicações suportadas pelas tecnologias e suas limitações. A pesquisa sobre tecnologias de comunicação para medidores eletrônicos aplicadas em IoT é importante como referência para escolha em implementações práticas. O trabalho não define uma tecnologia mais adequada para aplicação de IoT em *smartgrids*.

Em [Spano et al. 2014], são propostos e desenvolvidos protótipos para medidores inteligentes de energia elétrica, roteadores, servidores distribuídos para IoT e uma interface de usuário acessível por *smartphones*. O medidor inteligente proposto pelos autores não faz parte da *smartgrid*, ele é um módulo a parte que pode ser conectado às tomadas das residências. Para comunicação entre o medidor e o servidor IoT os autores utilizam um *gateway* ZigBee. A aplicação de *smartphone* acessa esses servidores IoT para coletar as informações geradas pelos medidores inteligentes.

Os trabalhos da literatura não apresentam solução para aumentar a vida útil da rede quando utiliza-se baterias. Além disso, as aplicações são voltadas para mercados específicos e, por isso, não atendem as normas que regem as *smartgrids* brasileiras.

3. Arquitetura da Rede Elétrica Inteligente

3.1. Funcionamento

Nesta seção, a arquitetura da rede inteligente para medição de energia elétrica é apresentada. A figura 1 apresenta a arquitetura da rede que se divide em três módulos: módulo concentrador de dados (1), módulo de comunicação sem fio (2) e módulo de medição de energia elétrica (3).

O módulo concentrador de dados disponibiliza uma interface desenvolvida em LabView para controle das leituras de consumo de energia elétrica dos clientes conectados à rede, além de disponibilizar os dados em servidores na nuvem. O módulo de comunicação sem fio é responsável pelo tráfego de dados na rede, executando roteamento, controle de fluxo e criptografia dos dados. Dispositivos *transceiver* XBee foram utilizados para implementação. A prevenção de colisão é realizada através de CSMA/CA. O

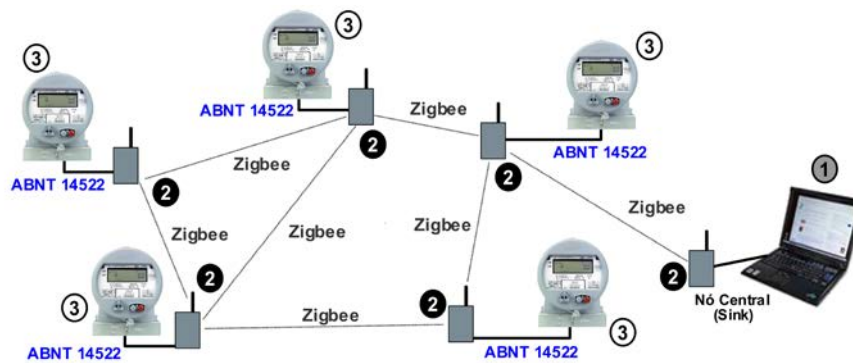


Figura 1. Rede sem fio de medidores eletrônicos de energia elétrica

módulo de medição de energia é responsável pela medição dos dados de consumo dos clientes. Para este módulo, utilizou-se o medidor eletrônico de energia Spectrum S, fabricado pela empresa Nansen S/A. O Spectrum S utiliza o DSP (*Digital Signal Processor*) TMS320C203 da Texas Instruments. As implementações desenvolvidas no sistema embarcado do medidor foram realizadas utilizando o compilador Code Composer 4.12.

Os módulos medição de energia e concentrador de dados comunicam-se através do módulo de comunicação sem fio. O módulo de medição de energia envia os dados de consumo e parametrização para a rede. O módulo de comunicação sem fio roteia esses dados, passando a informação por outros nós da rede, quando necessário, até o concentrador que disponibilizará os dados na internet pelo sistema *Metering Center* da Siemens S/A. As comunicações entre os módulos de medição de energia elétrica e o módulo de comunicação sem fio são realizadas utilizando a interface de comunicação serial RS232 dos medidores, conforme norma ABNT 14522 [ABNT 2008]. A comunicação entre os módulos de comunicação sem fio é feita utilizando ZigBee. A figura 2 apresenta as velocidades de transmissão entre os módulos.

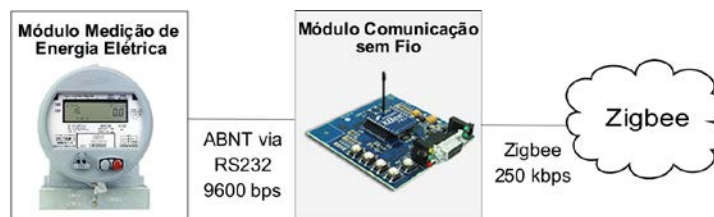


Figura 2. Velocidade de transmissão entre os módulos

Os registradores do medidor são atualizados a cada minuto e considerando a regulamentação da ANEEL que exige a atualização do mostrador de consumo do cliente sempre que o registrador de consumo do medidor for alterado, uma rede inteligente de medidores necessita de leituras a cada minuto. Quanto ao regime de operação, deve-se levar em consideração o envio de 2 blocos de 258 bytes (comandos 21 e 23), informando medição e configuração dos medidores, a cada minuto por nó da rede. Através da análise conjunta desses comandos, as concessionárias verificam se o medidor está enviando o consumo zerado devido a falta de energia, falha do medidor ou alteração indevida de terceiros. Para isso, é requisito que o medidor continue enviando o consumo mesmo durante a falta de energia. Também utiliza-se o envio a cada minuto para atendimento de norma.

3.2. Ciclo de Trabalho Dinâmico

Esta seção apresenta o ciclo de trabalho dinâmico, solução proposta para maximizar o tempo de vida da rede para medição de energia, considerando o processo de formação da rede e as alterações na topologia. Inicialmente, todos os nós possuem ciclo de trabalho de 100%. O concentrador é parametrizado na concessionária com os nós componentes de sua rede, pois os clientes devem ser identificados para serem faturados. Assim checa-se, durante a formação, se todos os nós estão ativos. Caso um nó não entre na rede, a concessionária deverá verificar o problema e corrigi-lo. Assim que todos os nós entram na rede, o primeiro ciclo de trabalho é calculado e enviado para a rede. A rede passa a trabalhar dentro do ciclo de trabalho calculado quando ocorre uma falta de energia e os módulos passam a ser alimentados por baterias. No início de cada minuto, os nós acordam, transmitem as informações de consumo de energia e voltam ao estado inativo. Para ativar ou desativar o módulo de comunicação sem fio, o módulo de medição envia um comando via RS232 para o mesmo.

Sempre que a rede volta a ser alimentada pela rede elétrica o módulo concentrador checa o sincronismo dos nós. Para isso, ele envia comando para leitura de hora e data para os módulos de medição de energia. Se necessário, um novo ciclo de trabalho é calculado considerando as diferenças de sincronismo. Medidores Spectrum S possuem relógio com precisão de 30 PPM (Partes por Milhão)/ano. Se a diferença for maior que a precisão esse medidor provavelmente está com problemas e um alarme é gerado no concentrador. Após esse alarme a concessionária pode realizar maiores checagens para entender o motivo da diferença.

A metodologia para cálculo do ciclo de trabalho deve ser aplicável a todas as configurações possíveis para redes de medidores. Como exemplo, considere uma rede de medidores formada de acordo com a topologia ilustrada na figura 3. Para cada nó da rede, um número de saltos (d) até o concentrador de dados (*sink*) está apresentado. Essa profundidade é identificada durante a formação da rede.

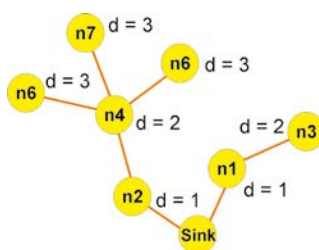


Figura 3. Exemplo de rotas entre nós

Quanto ao mecanismo de gestão dinâmica do ciclo de trabalho, sempre que um novo nó entra na rede um comando em difusão é enviado pelo mesmo, avisando a todos de sua entrada. Após receber o comando, o módulo concentrador calcula o novo ciclo de trabalho e envia o valor em difusão para todos da rede. Para cálculo do valor do ciclo de trabalho, é necessário conhecer o tempo total para que todos os nós da rede enviem seus dados. O tempo total necessário para enviar os dados é diretamente proporcional à quantidade de nós da rede (N), a profundidade de cada nó (d_i) e ao tempo de transmissão em um enlace (t).

Para que todas as informações cheguem até o concentrador, serão necessárias e transmissões, onde $e = \sum_{i=1}^N d_i$. Considerando que transmissões simultâneas não acontecem e também que todas as transmissões são bem sucedidas, o tempo total para transmissão seria: $T = t \times e$. Entretanto, colisões e erros podem acontecer durante as transmissões. Considere p a probabilidade de sucesso de uma transmissão e Z uma variável aleatória com distribuição geométrica que representa o número de tentativas para envio com sucesso de um pacote em um determinado enlace, tem-se: $Z \sim Geo(p)$. O valor esperado da variável aleatória Z é: $E(Z) = \frac{1}{p}$. Sendo assim, no pior caso, o tempo total necessário para enviar todos os pacotes será definido por: $T = \frac{1}{p} \times t \times e$

Portanto, na solução proposta, no início de cada minuto, os nós ficam acordados durante T segundos para transmitirem as informações de consumo de energia e voltam ao estado inativo. O ciclo de trabalho deve ser atualizado sempre que um novo nó entrar na rede ou uma manutenção de rota for executada. Para realização da manutenção da rota, todos os nós continuam informando a sua profundidade durante o funcionamento, essa informação é registrada no concentrador e sempre checada. Caso ocorra uma alteração nesse valor, causado pela mudança de rotas, um novo ciclo de trabalho é calculado e enviado aos nós da rede.

4. Avaliações e Resultados Experimentais

Esta seção apresenta as avaliações e resultados experimentais do sistema proposto neste trabalho. Os experimentos dividem-se em quatro grupos: medição do consumo de energia dos modos de funcionamento (seção 4.1), probabilidade de sucesso em cada transmissão (seção 4.2), tempo de vida da rede (seção 4.3) e o percentual de sucesso no recebimento das informações de consumo (seção 4.4). Todos os experimentos foram realizados na empresa Nansen S/A utilizando a plataforma XBee Série 2 - XB24BPKD juntamente com o amperímetro Minipa ET-1110 DMM para medição de consumo dos nós. Em todos os experimentos, utilizou-se a bateria Duracell Alcalina de 9 V padrão 6LR61 de 580 mAH como fonte de alimentação para os módulos de comunicação sem fio.

4.1. Medição do Consumo de Energia dos Modos de Funcionamento

O primeiro experimento foi realizado com o intuito de verificar o consumo de energia nos modos de funcionamento e o tempo de vida de um nó da rede em cada modo. Para este experimento, os nós 4 e 5, apresentados na figura 5-(b) foram utilizados. Um responsável pela transmissão e outro pela recepção. A distância entre os nós não foi alterada pois o ZigBee não possui controle de potência [Farahani 2008].

Para a coleta dos dados de consumo de energia, um circuito foi montado entre o módulo comunicação sem fio e sua alimentação. O circuito consiste em inserir um amperímetro em sequência com a bateria para medir o consumo da mesma. A cada cinco minutos, registrou-se o consumo do módulo de comunicação. A figura 4 ilustra o sistema montado e utilizado para medição das informações. Percebe-se na foto o módulo de medição de energia elétrica (1), medidor Spectrum S, o módulo comunicação sem fio ZigBee (2) e o amperímetro registrando o consumo (3).

Os experimentos foram realizados durante o período de 165 minutos, contabilizando 33 registros. Três modos de operação do módulo XBee foram avaliados: inativo, recepção e transmissão. No modo inativo, o rádio encontra-se desligado de forma que

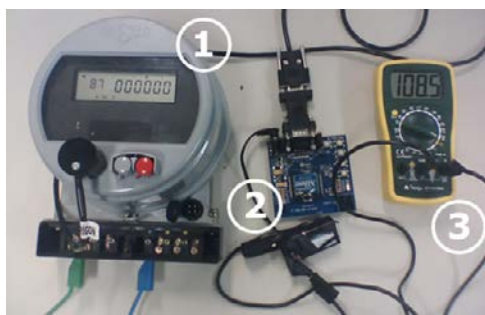


Figura 4. Registro de consumo do módulo de comunicação sem fio

apenas a placa consome energia. O consumo do modo de funcionamento inativo é de aproximadamente 4,5 mA, pois é necessário manter os *leds* de funcionamento e a porta de comunicação RS232 para ativar ou destivar o módulo. Para execução do experimento de medição do consumo durante a recepção de dados, o módulo comunicação sem fio permanece, durante todo o tempo, recebendo dados. Nesse modo, o consumo permanece constante na faixa de 87 mA. Durante a transmissão, o consumo aumenta consideravelmente se comparado com os modos anteriormente avaliados. O consumo para transmissão na placa de desenvolvimento permaneceu próximo a 110 mA.

A tabela 1 mostra o tempo de vida de um nó considerando seu modo de operação. O tempo de vida de um nó que permanece no modo inativo durante todo o período é 580 mAH (energia fornecida pela bateria) / 4,5 mA (consumo do modo de funcionamento), o que resulta em 128 horas e 53 minutos. Assim calcula-se o tempo de vida do nó para os demais modos de funcionamento.

Tabela 1. Tempo de vida de um nó

Modos de Operação	Tempo
Inativo	128 horas e 53 minutos
Recepção (Rx)	6 horas e 44 minutos
Transmissão (Tx)	5 horas e 16 minutos

A corrente necessária para transmissão do módulo de comunicação, 110 mA, é superior a fornecida pela fonte do medidor de energia elétrica que é 30 mA. Sendo assim, é impossível alimentar o módulo de comunicação com a fonte do medidor.

4.2. Probabilidade de Sucesso em cada Transmissão

O objetivo desse experimento é encontrar o valor da probabilidade de sucesso em cada transmissão (p) definido na seção 3.2. A figura 5-(a) ilustra a rede implementada, que utiliza a topologia *mesh*. Escolheu-se os nós 11 e 5 para esse experimento devido a maior quantidade de saltos para comunicação. O módulo de comunicação sem fio foi configurado para realizar apenas uma tentativa de envio do pacote pela rede. Caso ocorra algum problema durante o envio, não é realizada uma nova tentativa de transmissão. Além disso, o módulo permanece com o rádio ligado durante todo o experimento. Esse experimento foi realizado durante o intervalo de 5 horas. A cada 5 segundos, o nó 11 envia os dados de consumo e configuração. Ao final, totalizou-se o envio de 7.198 comandos pelo nó 11 dos quais 7.133 foram recebidos pelo nó 5. Sendo assim, o percentual de sucesso em cada

transmissão foi $p = 0,991$. Esse valor será utilizado nas próximas seções para o cálculo do valor do ciclo de trabalho.

4.3. Tempo de Vida da Rede

O objetivo dessa seção é avaliar o tempo de vida da rede para pilha de protocolos ZigBee e pilha de protocolos ZigBee com o ciclo de trabalho dinâmico proposto neste trabalho. Em todos os experimentos, a cada 60 segundos, o módulo de medição de energia envia os dados de consumo e configuração através do módulo de comunicação sem fio.

Para execução dos experimentos uma mesma tabela de roteamento foi utilizada. A figura 5-(b) ilustra as rotas definidas. Escolheu-se o nó 7 para medir o tempo de vida da rede devido a sua maior quantidade de filhos na rede, o que acarreta em um número maior de envio e recepção de dados. Os dados foram coletadas a cada 30 minutos, através da montagem especificada na figura 4. O tempo de vida da rede, para tecnologia ZigBee original, foi de 433 minutos ou 7 horas e 13 minutos. Um tempo de vida pequeno, mas dentro das expectativas devido às medições realizadas na seção 4.1.

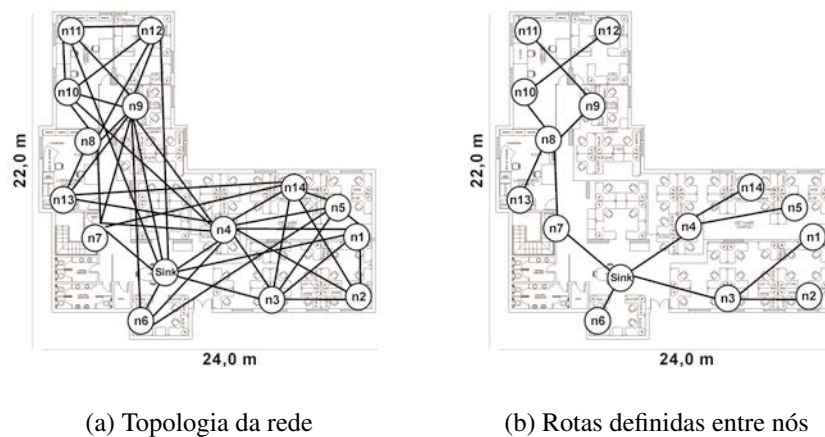


Figura 5. Topologia e Rotas

Baseando-se na topologia apresentada na figura 5-(a) e na proposta para maximizar o tempo de vida da rede, calcula-se o valor do ciclo de trabalho. Considerando a velocidade de transmissão de 250 kbps do ZigBee, 516 bytes transmitidos, o valor do tempo (t) para cada transmissão realizada por enlace da rede será 0,016 segundos. Conforme apresentado na seção 3, o número total de transmissões realizadas na rede (e), considerando o pior caso, no qual não acontece transmissão simultânea entre os nós, é um somatório da profundidade de todos os nós da rede. Considerando a topologia apresentada na figura 5, o valor de e é igual a 31. De acordo com o experimento realizado na seção 4.2, a probabilidade de sucesso em cada transmissão é de $p = 0,991$. O tempo total necessário para enviar todos os pacotes será: $T = \frac{1}{p} \times t \times e = \frac{1}{0,991} \times 0,016 \times 31 = 0,50$ segundos.

Ao tempo T calculado é necessário acrescentar o tempo gasto para comunicação entre o módulo de medição de energia elétrica e o módulo de comunicação sem fio, pois conforme especificado na seção 3.1, a porta de comunicação RS232 dos medidores possui velocidade de 9600 bps. A comunicação entre os dois módulos acontece de forma

simultânea em toda a rede, então basta acrescentar o tempo necessário para uma transmissão dos 516 bytes entre o módulos. De acordo com os dados apresentados, o valor de T será: $T = 0,50 + \left(\frac{516 \times 8}{9600}\right) = 0,93 \text{ segundos}$.

Como cada nó deverá permanecer ligado durante 0,93 segundos em cada 60 segundos, conclui-se que o ciclo de trabalho para a rede analisada é de 1,55%. Estudos mostram que o valor de ciclo de trabalho deve ser próximos a 1% para que redes de sensores sejam funcionalmente viáveis [Guo et al. 2009].

O gráfico da figura 6 apresenta a comparação entre os tempos de vida das duas tecnologias analisadas, percebe-se que ao atingir 4,8 V o nó da rede para de comunicar, pois a plataforma XBee Série 2 - XB24BPK necessita de uma alimentação de no mínimo 4,8 V para o funcionamento de seus componentes. Observa-se também que o mecanismo para maximizar o tempo de vida da rede aumenta o tempo de vida em torno de 17 vezes. O tempo máximo de vida da rede será 126 horas e 30 minutos para o ZigBee com ciclo de trabalho dinâmico, atendendo ao requisito da norma NBR14519:2011 [ABNT 2011].

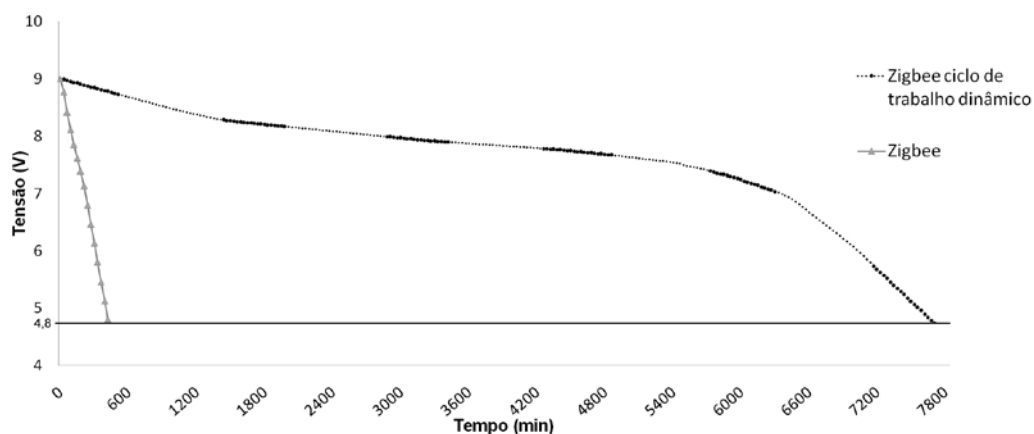


Figura 6. Tempo de vida das tecnologias analisadas

4.4. Percentual de Sucesso no Recebimento das Informações de Consumo

Para cálculo do percentual de sucesso no recebimento de informações de consumo utilizou-se a mesma topologia de rede dos experimentos de tempo de vida para o ZigBee com ciclo de trabalho dinâmico, ilustrada na figura 5-(b). Compara-se os dados recebidos pelo módulo concentrador de dados com os dados registrados nos módulos de medição de energia elétrica. Esse experimento foi realizado durante o período de 8 horas.

Os comandos trafegados na rede são compostos por 258 bytes. Os últimos dois bytes são utilizados para informar o CRC (*Cyclic Redundancy Check*) de 16 bits que é calculado sobre os demais 256 bytes. O módulo concentrador só considera um comando recebido como válido se o cálculo de seu CRC estiver correto.

Após as 8 horas de testes, o módulo de medição de energia interrompe o envio de comandos e o registro do envio dos mesmos. Em cada módulo de medição de energia da rede, registrou-se o envio de 958 comandos. Ao todo, 13.412 comandos foram enviados pela rede. O módulo concentrador de dados registrou o recebimento de 13.239 comandos corretos. Assim tem-se um percentual de sucesso de 98,7% no recebimento das informações.

5. Conclusões e Trabalhos Futuros

O trabalho apresentou uma aplicação prática para IoT nas redes elétricas inteligentes brasileiras utilizando as normas vigentes. Foi apresentada também uma metodologia para cálculo do ciclo de trabalho dinâmico cujo objetivo foi manter um ciclo de trabalho mínimo para leitura dos dados de consumo, mesmo ocorrendo alterações na topologia da rede. A principal contribuição foi o sensível aumento do tempo de vida da rede. Com a utilização do ciclo de trabalho dinâmico, a rede passou a atender as normas vigentes. O ciclo de trabalho manteve-se dentro dos padrões definidos pela literatura, próximo à 1%. Além disso, a rede proposta mostrou-se eficiente quanto ao consumo máximo de um conjunto de medição definido nas normas vigentes para medição de energia elétrica. Considerando o percentual de sucesso nas transmissões realizadas, a ciclo de trabalho dinâmico mostrou-se eficiente atingindo um percentual de 98,7%, o que comprova a confiabilidade do sistema proposto.

Importante ressaltar que, assim como as tecnologias, as normas brasileiras precisam avançar. Os protocolos definidos na norma ABNT 14522 são extremamente ultrapassados e rígidos. Propostas simples para diminuir o tráfego de dados na rede podem ser implementadas sem perda de confiabilidade. Agregação de dados e pacotes com tamanho variável, sem a obrigatoriedade de 258 bytes fixos, são exemplos.

Como trabalho futuro, pretende-se considerar, durante o cálculo do tempo necessário para enviar todos os pacotes, as transmissões simultâneas que poderão acontecer nos diversos enlaces da rede. Assim, o tempo necessário para enviar todos os pacotes pode ser diminuído. Outro trabalho futuro seria calcular o p dinamicamente. Esse valor seria utilizado no cálculo do ciclo de trabalho de forma que a solução proposta fosse adaptável às condições do meio de transmissão sem fio.

Referências

- ABNT (2008). *NBR 14522: Intercâmbio de informações para sistemas de medição de energia elétrica - Padronização*. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS ABNT, Rio de Janeiro.
- ABNT (2011). *NBR 14519: Medidores eletrônicos de energia elétrica (estáticos) - Especificação*. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS ABNT, Rio de Janeiro.
- Arif, A., Al-Hussain, M., Al-Mutairi, N., Al-Ammar, E., Khan, Y., and Malik, N. (2013). Experimental study and design of smart energy meter for the smart grid. In *Renewable and Sustainable Energy Conference (IRSEC), 2013 International*, pages 515 – 520, Ouarzazate. IEEE.
- Arnachalam, V., Jain, S., Kumar, N., Paventhan, A., Chinnaiyan, V., and Pradish, M. (2014). Survey on smart grid technologies- smart metering, iot and ems. In *Electrical, Electronics and Computer Science (SCEECS), 2014 IEEE Students' Conference on*, pages 1 – 6, Bhopal. IEEE.
- Bello, O. and Zeadally, S. (2014). Intelligent device-to-device communication in the internet of things. *Systems Journal, IEEE (Issue: 99)*, PP:1 – 11.
- Chen, X., Liu, J., Li, X., Sun, L., and Zhen, Y. (2011). Integration of iot with smart grid. In *Communication Technology and Application (ICCTA 2011), IET International Conference on*, Beijing. IET - IEEE.
- Farahani, S. (2008). *Zigbee Wireless Networks and Transceivers*. Elsevier, Oxford, UK.
- Garpetun, L. (2013). Smart meters in operation center. In *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on*, pages 1 – 3, Stockholm. IET - IEEE.
- Guo, S., Gu, Y., Jiang, B., and He, T. (2009). Opportunistic flooding in low-duty-cycle wireless sensor networks with unreliable links. In *MobiCom '09: Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 133–144, New York, NY, USA. ACM.
- Spano, E., Niccolini, L., Pascoli, S., and Iannaccone, G. (2014). Last-meter smart grid embedded in an internet-of-things platform. *Smart Grid, IEEE Transactions on (Volume:PP , Issue: 99)*, PP:1.
- Yun, M. and Yuxin, B. (2010). Research on the architecture and key technology of internet of things applied on smart grid. In *Advances in Energy Engineering, 2010 International Conference on*, Beijing. IEEE.

Lotus@Runtime: Uma Ferramenta para Monitoramento e Verificação em Tempo de Execução para Sistemas Autoadaptativos

Davi Monteiro Barbosa¹, Paulo Henrique Mendes Maia¹, Evilásio Costa Júnior¹

¹Universidade Estadual do Ceará (UECE)

Centro de Ciências e Tecnologia - CCT

Av. Dr. Silas Munguba, 1700, Campus do Itaperi – Fortaleza – CE

davi.monteiro@aluno.uece.br, pauloh.maia@uece.br, junior.facanha@gmail.com

Abstract. *Self-adaptive systems have the ability to adapt their behavior during its execution in response to changes in the environment where it operates. To ensure the success of an adaptation, traditional approaches such as testing and software verification are insufficient. Therefore, it should adopt monitoring and runtime verification approaches to ensure that adaptations have the expected result. In this paper, we propose the Lotus@Runtime to perform, in an integrated manner, monitoring, verification, and violations notification found during execution of a self adaptive system.*

Resumo. *Sistemas autoadaptativos possuem a capacidade de adaptar seu comportamento durante sua execução em resposta às mudanças no ambiente onde está inserido. Para garantir o sucesso de uma adaptação, abordagens tradicionais como testes e verificações de software são insuficientes. Por isso, deve-se adotar abordagens de monitoramento e verificação em tempo de execução para assegurar que as adaptações tenham o resultado esperado. Neste trabalho, é proposta uma ferramenta para realizar, de forma integrada, o monitoramento, a verificação e a notificação de violações encontradas durante a execução de um sistema autoadaptativo.*

1. Introdução

O crescimento da complexidade de sistemas, junto com o custo para mantê-los após o seu desenvolvimento, foram fatores apontados por Kephart e Chess (2003) que serviram de motivação para a proposta de sistemas autoadaptativos. Para Huebscher e McCann (2008), características como autoconfiguração, auto-otimização, autocorreção e autoproteção fazem parte da essência dos sistemas autoadaptativos, fazendo com que essa nova classe de sistemas tenha autonomia suficiente para operar com o mínimo de intervenção humana.

Um sistema autoadaptativo possui a capacidade, em tempo de execução, de modificar o seu comportamento ou sua estrutura através de adaptações em resposta a mudanças no ambiente, que podem representar uma simples reconfiguração ou até uma modificação no modelo do sistema, resultando em uma alteração de sua arquitetura [Chen et al. 2014]. A adaptação pode acontecer para que uma funcionalidade do sistema não deixe de ser executada (por exemplo, a troca de um serviço por outro similar caso o primeiro tenha

ficado indisponível) ou para garantir que requisitos não funcionais sejam atendidos, como níveis de confiabilidade, segurança ou disponibilidade, dentre outros.

Uma técnica bastante utilizada no desenvolvimento de sistemas autoadaptativos é a construção de modelos arquiteturais ou de comportamento em tempo de projeto para que sejam atualizados durante a execução do sistema. Tais modelos são conhecidos como *models@runtime* [Blair et al. 2009]. Esse cenário acontece, por exemplo, em sistemas que devem garantir requisitos expressos em termos de propriedades e que são altamente influenciados pela forma como o ambiente se comporta (por exemplo, levando em conta o perfil de comportamento do usuário). Com isso, é possível verificar propriedades utilizando uma técnica quantitativa, como a checagem de modelos (*model checking*), para prevenir e identificar violações dessas propriedades, bem como planejar os passos da adaptação para prevenir ou recuperar o sistema das violações [Calinescu et al. 2012].

Para endereçar esses desafios, trabalhos como Goldsby *et al.* (2008), Arcaini *et al.* (2012) e Calinescu *et al.* (2013) apresentam propostas para realizar o monitoramento e a verificação em sistemas autoadaptativos. Porém, essas soluções geralmente são específicas para uma categoria de sistemas, como os baseados em serviço, o que torna difícil sua reutilização para outros domínios. Além disso, a maioria faz uso de uma linguagem formal para especificar o comportamento do sistema e para definir as propriedades de interesse, o que também dificulta sua aplicação por usuários que não têm familiaridade com métodos formais. Por fim, esses trabalhos não disponibilizam um ambiente integrado, onde o usuário possa, na mesma ferramenta, realizar a modelagem, o monitoramento e a verificação do sistema.

Este trabalho apresenta Lotus@Runtime, uma ferramenta para realizar, de forma integrada, o monitoramento, a verificação e a notificação de violações encontradas durante a execução de um sistema autoadaptativo. O Lotus@Runtime realiza o monitoramento dos rastros de execução (*traces*) gerados por um sistema autoadaptativo para anotar o modelo do sistema baseado em *Labelled Transition System* (LTS) probabilístico. Em seguida, as verificações em tempo de execução são realizadas a partir do modelo atualizado e um conjunto de propriedades de alcançabilidade. Caso uma propriedade seja violada, o sistema autoadaptativo será notificado para realizar uma adaptação. A ferramenta foi utilizada em um sistema autoadaptativo existente que não fazia o uso de modelos em tempo de execução. Por isso, foi necessário customizar o estudo de caso para que a aplicação pudesse ser adaptada de acordo com violações encontradas pelo Lotus@Runtime. Por fim, por ser de código-fonte aberto, ela permite que a comunidade possa contribuir com melhorias e ajustes.

O restante deste artigo encontra-se organizado da seguinte forma: a seção 2 apresenta o referencial teórico. Em seguida, a seção 3 descreve a ferramenta proposta, sua arquitetura e seu funcionamento. Posteriormente, a seção 4 mostra um estudo de caso que foi realizado na ferramenta Lotus@Runtime. A seção 5 apresenta os principais trabalhos relacionados. Por fim, a seção 6 apresenta as considerações finais.

2. Fundamentação Teórica

Sistemas de software autoadaptativos modificam seu próprio comportamento em resposta às mudanças de contexto [Oreizy et al. 1999]. Pelo contexto, entende-se o ambiente no qual o sistema está inserido, ou seja, quaisquer itens observáveis do sistema, tais como:

entradas de usuário, sensores, outras aplicações, entre outros. No âmbito de sistemas autoadaptativos, as adaptações têm papel fundamental no sucesso de tais aplicações. Por isso, adaptações devem ocorrer dentro de um ciclo de vida interminável durante a execução do sistema para que possam implantar as solicitações de mudanças requeridas, tanto pelo ambiente quanto pelos usuários.

O *loop* de controle MAPE-K (do inglês, *Monitor, Analyze, Plan, Execute over Knowledge base*) [Huebscher and McCann 2008, Cheng et al. 2009] tem sido utilizado como uma alternativa para viabilizar a autoadaptação em sistemas de software [Salehie and Tahvildari 2009]. Em resumo, na proposta apresentada em [Oreizy et al. 1999], um sistema de software adquire comportamento autoadaptativo através das adaptações providas pelo *loop* de controle. Os sensores são componentes de software ou hardware responsáveis por coletar informações de um sistema autoadaptativo. Os atuadores também são componentes de software ou hardware, que são responsáveis por aplicar as adaptações provenientes do loop de controle diretamente em um sistema autoadaptativo. Por fim, Salehie e Tahvildari (2009) apresentam uma taxonomia sobre duas abordagens de adaptação: interna e externa. Na primeira a lógica de adaptação está embutida no sistema de software. Na segunda, a lógica de adaptação está isolada do sistema de software.

Uma vez que uma adaptação modifica o comportamento de um sistema autoadaptativo, verificações realizadas durante a fase de desenvolvimento são insuficientes para garantir a conformidade no comportamento desses sistemas. Por isso, deve-se realizar verificações em tempo de execução com o objetivo de identificar violações em propriedades para que se possa planejar adaptações que previnam ou recuperem um sistema de tais violações [Calinescu et al. 2012]. Desse modo, faz-se necessário representar um sistema autoadaptativo em abstrações que possam acompanhar sua evolução no decorrer do tempo. Para tanto, modelos em tempo de execução são fortes candidatos para representar a evolução de tais sistemas [Bencomo et al. 2013]. Um modelo em tempo de execução pode ser definido como uma abstração de uma representação de um sistema, incluindo sua estrutura, comportamento e objetivos, que visa atender um propósito específico durante a execução do sistema [Blair et al. 2009, Bencomo et al. 2013].

3. Lotus@Runtime

Lotus@Runtime foi desenvolvido com o objetivo de estender as funcionalidades do LoTuS¹, ferramenta para modelagem gráfica e análise de comportamento de software utilizando LTS. LoTuS fornece aos usuários um mecanismo de *drag and drop* para a criação dos modelos, o que torna a modelagem mais fácil e intuitiva. Além disso, a ferramenta também disponibiliza algumas técnicas de análise de modelo, como detecção de *deadlocks*, simulação e execução, além da verificação probabilísticas de propriedades de alcançabilidade (*reachability properties*), que são especificadas através do estado de origem e estado destino no modelo que se deseja alcançar.

Lotus@Runtime aproveita esses benefícios do LoTuS e fornece suporte para o monitoramento e verificação de software utilizando modelos probabilísticos em tempo de execução.

¹<http://www.larces.uece.br/gesad/ferramentas/lotus/>

3.1. Arquitetura da ferramenta

O Lotus@Runtime foi projetado com uma arquitetura extensível baseada em componentes, possibilitando, dessa forma, um maior desacoplamento do seu código. A interface de programação disponibilizada pela ferramenta permite que componentes sejam instalados, removidos, ou recuperados por outro componente. A Figura 1 ilustra os principais componentes que fazem parte da arquitetura do Lotus@Runtime: MonitorComponent, LotusModelComponent, ModelCheckerComponent, NotifierComponent e ConfigurationComponent. A seguir, cada componente é melhor detalhado.

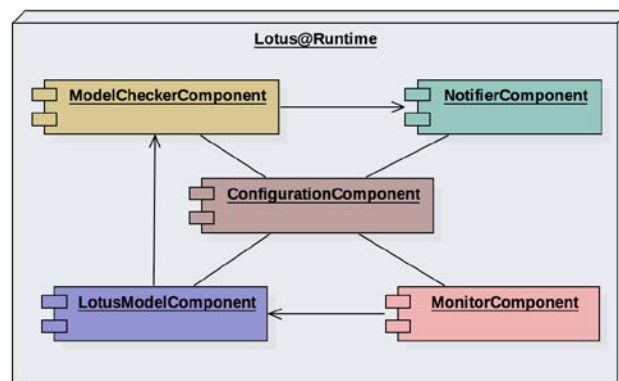


Figura 1. Arquitetura do Lotus@Runtime

3.1.1. MonitorComponent

Para realizar o monitoramento do sistema, optou-se por utilizar uma abordagem de monitoramento dos rastros de execução gerados pela aplicação [Maoz 2009]. Dessa forma, pode-se extrair informações relevantes que serão essenciais para manter o modelo em tempo de execução sempre atualizado em relação à evolução do sistema. Cada rastro é formado por uma sequência de ações visíveis que representam eventos responsáveis pela mudança de estado do sistema. Assim, um *log* é um conjunto finito de rastros que uma aplicação pode gerar durante sua execução.

O `MonitorComponent` é o componente responsável por monitorar os rastros que são gerados por um sistema autoadaptativo. Dessa forma, o monitoramento da aplicação é realizado através da leitura do arquivo de *log*, o qual deve estar no formato CSV. Cada linha no arquivo representa um rastro da aplicação e cada ação de mudança de estado é separada por uma vírgula em seus respectivos rastros. Existem duas estratégias para monitorar os rastros: na primeira, verifica-se a existência de um novo rastro de acordo com uma periodicidade definida pelo usuário (em milisegundos), enquanto na segunda verifica-se apenas a existência de um novo rastro caso ocorra uma alteração no arquivo de *log*.

3.1.2. LotusModelComponent

Para atender à necessidade de manter o modelo de um sistema autoadaptativo atualizado em relação às constantes adaptações que ocorrem ao longo do tempo, foi criado o compo-

nente `LotusModelComponent`, o qual é responsável por manter o modelo atualizado em relação às mudanças no comportamento da aplicação e oferecer serviços para recuperar o modelo atualizado para outros componentes do `Lotus@Runtime`.

O `Lotus@Runtime` usa o modelo do sistema representado como um LTS e criado de forma gráfica através da ferramenta `LoTuS`. A partir das informações extraídas pelo `MonitorComponent` dos rastros de execução da aplicação, o `LotusModelComponent` atualiza as probabilidades de ocorrência de cada ação do modelo e as anota em suas respectivas transições. Dessa forma, o modelo de comportamento do sistema passa a ser um LTS probabilístico, que pode ser visto como uma Cadeia de Markov de Tempo Discreto (DTMC) rotulada.

3.1.3. ModelCheckerComponent

Uma vez que o modelo de uma aplicação autoadaptativa esteja atualizado, pode-se então realizar verificações em tempo de execução com o objetivo de encontrar possíveis violações de alguma propriedade especificada pelo usuário. No `Lotus@Runtime`, uma propriedade é representada pela classe `Property`, composta por atributos que identificam um estado de origem e estado de destino de um modelo LTS, uma condição lógica que se deseja verificar e uma probabilidade de ocorrência.

Após definir um conjunto de propriedades que devem ser satisfeitas, o componente `ModelCheckerComponent` pode realizar verificações probabilísticas no modelo. Em uma verificação, o componente utiliza um algoritmo de alcance probabilístico para calcular a probabilidade do sistema ir de um estado inicial para um estado final, onde cada estado está definido em uma propriedade.

Em seguida, a probabilidade que foi calculada será comparada com a probabilidade que foi especificada na propriedade. Caso o resultado não satisfaça o valor por ela esperado, o que constitui uma violação, o `ModelCheckerComponent` deve invocar o serviço de notificação do `NotifierComponent`.

3.1.4. NotifierComponent

As possíveis violações de propriedades são publicadas pelo `NotifierComponent` em um *EventBus*, seguindo o padrão de arquitetura *publish subscriber* [Eugster et al. 2003]. Em seguida, para receber as violações que foram publicadas, deve-se implementar a interface `ViolationHandler` e sobrescrever o método `handler(property)`. Recomenda-se que a implementação concreta de um `ViolationHandler` faça parte da etapa de planejamento, pois, dessa forma, o planejador terá informações para projetar adaptações que possam evitar possíveis danos ao sistema.

O componente `ModelCheckerComponent` faz o uso do serviço disponibilizado pelo `NotifierComponent` toda vez que uma verificação em tempo de execução encontra uma violação no modelo de uma aplicação autoadaptativa.

3.1.5. ConfigurationComponent

O componente `ConfigurationComponent` foi desenvolvido com o objetivo de centralizar os parâmetros de configuração do `Lotus@Runtime` e fornecer, aos outros componentes, um mecanismo para recuperar tais parâmetros. Pode-se configurar o `Lotus@Runtime` através da interface de programação oferecida pelo `ConfigurationComponent` ou com o auxílio do *plugin Lotus@Runtime Configuration*, que está disponível para a ferramenta LoTuS e permite a configuração desses parâmetros através de uma interface gráfica. Por intermédio do *plugin* pode-se adicionar, remover, importar e exportar as configurações em formato JSON que serão lidas no momento da inicialização do `Lotus@Runtime`.

Os parâmetros de configuração do `Lotus@Runtime` são: (i) nome do arquivo de configuração; (ii) caminho do arquivo de *log* da aplicação autoadaptativa. (iii) caminho do arquivo do modelo LoTuS. (iv) tempo, em milissegundos, para verificar a existência de novos rastros de execução; (v) lista de propriedades que deseja-se verificar em tempo de execução.

3.2. Fluxo de uma adaptação

No `Lotus@Runtime`, o fluxo de uma adaptação, ilustrado na Figura 2, inicia-se pelo monitoramento dos rastro de execução gerados por um sistema autoadaptativo. Os rastros de execução são capturados durante a fase de monitoramento pelo *MonitorComponent*. Em seguida, durante o processo de atualização do modelo, o *LotusModelComponent* calcula as probabilidades de cada transição de estado e atualiza o modelo com as novas probabilidades em tempo de execução. Após a atualização do modelo, *ModelCheckerComponent* realiza as verificações em tempo de execução durante o processo de verificação. Caso uma propriedade seja violada, então *NotifierComponent* deve notificar a etapa de planejamento sobre a violação que ocorreu. Em seguida, independente de ocorrer ou não uma violação, o processo de monitoramento deve ser executado novamente de forma contínua. Para que as violações sejam recebidas, deve-se implementar a interface *ViolationHandler* na etapa de planejamento, como descrito na seção 3.1.4. Caso contrario, as notificações serão perdidas.

Como pode ser visto na Figura 2, em comparação com o ciclo MAPE-K, os componentes do `Lotus@Runtime` implementam apenas as fases de monitoramento e análise, fornecendo uma notificação em caso de violação de propriedades. As outras fases (planejamento e execução) devem ser implementadas pelo usuário fora da ferramenta.

4. Estudo de Caso

O estudo possui o objetivo de avaliar o uso da ferramenta proposta neste trabalho através da implementação das fases de monitoramento e análise utilizando o `Lotus@Runtime` em uma aplicação que não foi desenvolvida pelos autores deste artigo. A aplicação escolhida para realizar o estudo de caso foi a implementação de referência Tele Assistente System (TAS) proposta por Weyns and Calinescu (2015). A aplicação segue os princípios propostos na arquitetura de referência MAPE-K, o que motivou a escolha desta aplicação no estudo de caso. O material utilizado nesse estudo de caso está disponível na conta de um dos autores autor no GitHub².

²<https://github.com/davimonteiro>

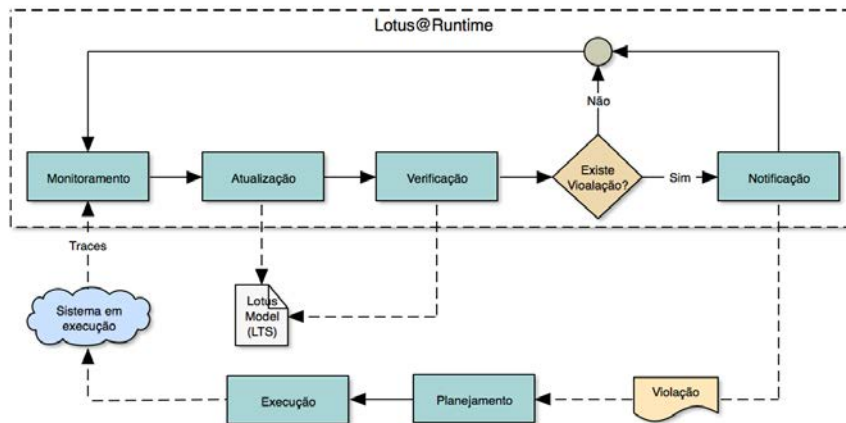


Figura 2. Fluxo de uma adaptação

A implementação de referência da aplicação TAS foi desenvolvida utilizando a plataforma ReSeP [Weyns and Calinescu 2015], que seleciona serviços durante a etapa de planejamento de acordo com uma política de menor custo ou maior confiabilidade que deve ser selecionada antes da inicialização de uma aplicação autoadaptativa. Assim, caso uma aplicação autoadaptativa seja executada utilizando a política de maior confiabilidade, o planejador deverá selecionar os serviços com menor taxa de falhas.

4.1. Implementação do estudo de caso

O presente estudo de caso concentra seus esforços em realizar as etapas de monitoramento e verificação na construção do TAS utilizando o Lotus@Runtime. Dessa forma, para validar o estudo de caso da ferramenta proposta, foi necessário realizar as seguintes atividades: (i) criar o modelo probabilístico com o auxílio da ferramenta LoTuS; (ii) configurar o Lotus@Runtime através do *plugin* de configuração; (iii) adicionar ao TAS a capacidade de gerar rastros de execução; (iv) receber as notificações sobre as violações na etapa de planejamento e modificar a estratégia de seleção dos serviços.

Como resultado da primeira atividade, foi criado na ferramenta LoTuS o modelo LTS probabilístico, ilustrado na Figura 3. Esse modelo foi baseado no modelo em DTMC apresentado pelos autores Calinescu *et al.* (2013). Em seguida, a configuração do Lotus@Runtime foi feita com o auxílio do *plugin* Lotus@Runtime Configuration. O arquivo de configuração é composto pelo nome da configuração, o caminho onde está localizado o arquivo de *logs* do sistema, o caminho onde está localizado o arquivo contendo o modelo probabilístico e as 2 propriedades que se desejam verificar, que são: (P1) a probabilidade de uma falha após o serviço de alarme deve ser menor que 20%, representado por alcançar o estado 5 a partir do estado 0, e (P2) a probabilidade de uma falha acontecer após o serviço de farmácia deve ser menor que 10%, o que é representado pelo alcance do estado 11 a partir do estado 0.

Posteriormente, para adicionar ao TAS a capacidade de gerar rastros de execução, foi necessário instrumentar o código fonte para que cada funcionalidade do TAS que corresponde a uma ação no modelo de comportamento da aplicação fosse gravada num *log*. Um exemplo de rastro de execução gerado pela aplicação é *start, vitalParamMsg, analyzeData, changeDrug, notifyPA, stopMsg, exit*. O arquivo de *logs* gerado foi utilizado como entrada para a o *plugin* Lotus@Runtime Configuration, conforme descrito anteriormente.

A estratégia de monitoramento adotada no estudo de caso foi a checagem de novos rastros a cada alteração no arquivo de *logs*.

Por último, para receber as notificações sobre as violações, foi implementada a classe `LotusRuntimeHandler` dentro da etapa de planejamento. Para realizar a troca de uma estratégia de seleção de serviços dentro do TAS, foi necessário alterar o comportamento do *framework* ReSeP. Assim, foi possível selecionar qual estratégia de seleção de serviços deve ser utilizada, de acordo com as informações recebidas após a etapa de análise realizada pelo Lotus@Runtime. A estratégia de seleção de serviços utilizada foi: caso P1 seja violada, então será selecionado o serviço de menor custo; caso P2 não seja satisfeita, então será selecionado o serviço com maior confiabilidade. É importante ressaltar que o presente trabalho não se propõe a comparar a estratégia de seleção de serviços adotada neste estudo de caso com a estratégia adotada no trabalho de Weyns e Calinescu (2015).

Fazendo uma análise sobre o estudo de caso, podemos relatar que não foi complicado adaptar o código-fonte do TAS para utilizar o Lotus@Runtime nas fases de monitoramento e análise. Além disso, pelo fato das estratégias de planejamento e execução das adaptações já estarem implementadas no *framework* ReSeP, o esforço para também adequar essas etapas ao estudo de caso foi pequeno. Contudo, vale ressaltar que o estudo de caso introduziu duas técnicas que não haviam na aplicação original: o monitoramento em tempo real da aplicação através dos rastros de execução e o uso de um modelo atualizado em tempo de execução para verificação de propriedades.

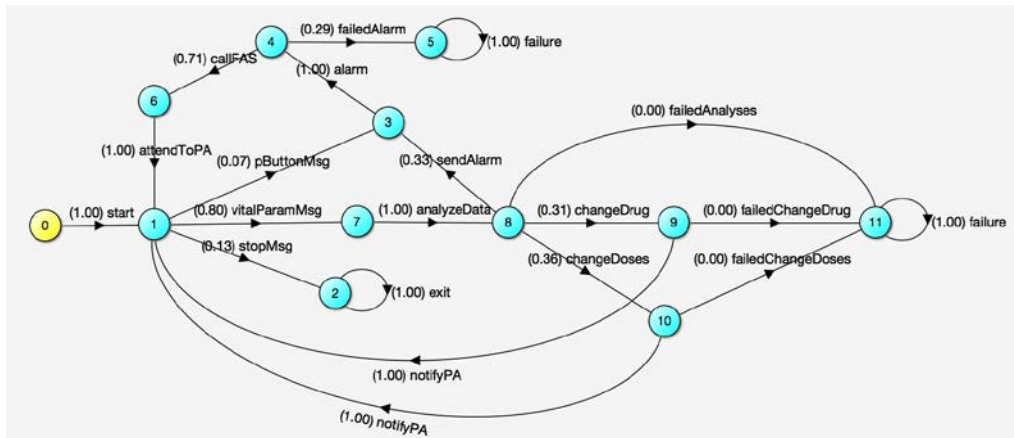


Figura 3. Modelo do Tele Assistance System (TAS)

5. Trabalhos Relacionados

No trabalho de Goldsby *et al.*(2008), os autores descrevem a ferramenta AMOEBA-RT que foi desenvolvida com o objetivo de realizar verificações em tempo de execução de sistemas autoadaptativos. A ferramenta faz o uso de uma abordagem não intrusiva baseada em programação orientada a aspectos para coletar informações sobre o estado atual do sistema. Após serem coletadas, essas informações são enviadas para um servidor remoto, onde um verificador de modelo confere se o estado atual do sistema satisfaz as propriedades que foram especificadas em formalismo baseado em LTL. Em resposta às violações detectadas, a ferramenta grava o caminho de execução das violações em um relatório de erros no qual é processado *offline*.

Em Arcaini *et al.*(2012), os autores apresentam o CoMA, uma ferramenta para monitoramento em tempo de execução de aplicações Java. Nessa proposta, são utilizadas anotações Java para associar o código-fonte do sistema monitorado com o modelo descrito no formalismo máquinas de estados abstratos (ASM). O monitor da ferramenta consegue extrair as informações sobre o estado do sistema em tempo de execução utilizando a ferramenta AspectJ. Após a extração dessas informações, o monitor verifica se o comportamento do sistema está em conformidade com o comportamento esperado que foi especificado no formalismo ASM. Caso o monitor encontre uma não conformidade, será reportado um relatório de erro que poderá ser analisado quando o sistema estiver *offline*.

No trabalho de Filieri *et al.*(2011) é proposta uma abordagem para verificação de modelo em tempo de execução mais eficiente do que as soluções tradicionais como o PRISM. O autor propõe uma abordagem que utiliza DTMC para expressar o modelo do sistema e representá-lo através da linguagem formal PCTL (*Probabilistic Computation Tree Logic*) para expressar as propriedades do sistema. Os resultados apresentados pelo autor mostram que sua abordagem, diferente das ferramentas PRISM e MRMC, obteve uma performance constante mesmo com o aumento do número de estados do DTMC.

No trabalho de Calinescu *et al.*(2013), os autores apresentam o *framework* COVE para criar sistemas autoadaptativos baseados em serviços. Os sistemas autoadaptativos construídos com base no COVE devem, de preferência, ter serviços redundantes com diferentes taxas de confiabilidade e custo. Dessa forma, o COVE pode realizar verificações formais no modelo a fim de garantir que o sistema utilize serviços confiáveis com o mínimo de custo possível. Para tal, o framework realiza verificações utilizando a ferramenta PRISM [Hinton et al. 2006] para analisar o modelo do sistema definido em DTMC.

Os trabalhos acima utilizam verificadores externos e formalismos para representar o comportamento de sistemas. Diferentemente, o presente trabalho realiza a verificação em um ambiente integrado que abrange desde a construção do modelo de comportamento de forma gráfica até a notificação de violações encontradas em tempo de execução.

6. Considerações Finais

Este trabalho apresentou o Lotus@Runtime para realizar, de forma integrada, o monitoramento, a verificação e a notificação de violações encontradas durante a execução de um sistema autoadaptativo. Para validar a ferramenta proposta, foi desenvolvido um estudo de caso para a aplicação de referência Tele Assistance System [Weyns and Calinescu 2015]. Uma limitação para adoção da ferramenta proposta é a capacidade do sistema autoadaptativo gerar rastros de execução que permitam popular o modelo em tempo de execução. Outra limitação atualmente é necessidade de implementar as etapas de planejamento e execução para completar o ciclo MAPE-K. Como trabalhos futuros, pretende-se adicionar o monitoramento para aplicações em tempo real e suportar outras etapas do ciclo MAPE-K. Além disso, pretende-se desenvolver outros estudos de caso e realizar comparativos de desempenho, custo e confiabilidade com outras abordagens.

Referências

- Arcaini, P., Gargantini, A., and Riccobene, E. (2012). Coma: conformance monitoring of java programs by abstract state machines. In *Runtime Verification*, pages 223–238. Springer.

- Bencomo, N., Bennaceur, A., Grace, P., Blair, G., and Issarny, V. (2013). The role of models@ run. time in supporting on-the-fly interoperability. *Computing*, 95(3):167–190.
- Blair, G., Bencomo, N., and France, R. B. (2009). Models@ run. time. *Computer*, 42(10):22–27.
- Calinescu, R., Ghezzi, C., Kwiatkowska, M., and Mirandola, R. (2012). Self-adaptive software needs quantitative verification at runtime. *Communications of the ACM*, 55(9):69–77.
- Calinescu, R., Johnson, K., and Rafiq, Y. (2013). Developing self-verifying service-based systems. In *Automated Software Engineering (ASE), 2013 IEEE/ACM 28th International Conference on*, pages 734–737. IEEE.
- Chen, B., Peng, X., Yu, Y., Nuseibeh, B., and Zhao, W. (2014). Self-adaptation through incremental generative model transformations at runtime. In *36th International Conference on Software Engineering, Hyderabad*. ACM/IEEE.
- Cheng, B. H. C., de Lemos, R., Giese, H., Inverardi, P., et al. (2009). *Software Engineering for Self-Adaptive Systems*, chapter Software Engineering for Self-Adaptive Systems: A Research Roadmap, pages 1–26. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Eugster, P. T., Felber, P. A., Guerraoui, R., and Kermarrec, A.-M. (2003). The many faces of publish/subscribe. *ACM Computing Surveys (CSUR)*, 35(2):114–131.
- Filieri, A., Ghezzi, C., and Tamburrelli, G. (2011). Run-time efficient probabilistic model checking. In *Proceedings of the 33rd international conference on software engineering*, pages 341–350. ACM.
- Goldsby, H. J., Cheng, B. H., and Zhang, J. (2008). Amoeba-rt: Run-time verification of adaptive software. In *Models in Software Engineering*, pages 212–224. Springer.
- Hinton, A., Kwiatkowska, M., Norman, G., and Parker, D. (2006). Prism: A tool for automatic verification of probabilistic systems. In *Tools and Algorithms for the Construction and Analysis of Systems*, pages 441–444. Springer.
- Huebscher, M. C. and McCann, J. A. (2008). A survey of autonomic computing—degrees, models, and applications. *ACM Computing Surveys (CSUR)*, 40(3):7.
- Kephart, J. O. and Chess, D. M. (2003). The vision of autonomic computing. *Computer*, 36(1):41–50.
- Maoz, S. (2009). Using model-based traces as runtime models. *Computer*, 42(10):28–36.
- Oreizy, P., Gorlick, M. M., Taylor, R. N., Heimbigner, D., Johnson, G., Medvidovic, N., Quilici, A., Rosenblum, D. S., and Wolf, A. L. (1999). An architecture-based approach to self-adaptive software. *IEEE Intelligent systems*, 14(3):54–62.
- Salehie, M. and Tahvildari, L. (2009). Self-adaptive software: Landscape and research challenges. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 4(2):14.
- Weyns, D. and Calinescu, R. (2015). Tele assistance: a self-adaptive service-based system exemplar. In *Proceedings of the 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pages 88–92. IEEE Press.

(Meta)Modelagem de Espaços Inteligentes Pessoais e Espaços Inteligentes Fixos para Aplicações Ubíquas

Marcos Alves Vieira¹, Sergio T. Carvalho²

¹Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano)
Iporá – GO – Brasil

²Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

marcos.vieira@ifgoiano.edu.br, sergio@inf.ufg.br

Abstract. *User mobility and the increasing number of smart spaces, fostered also by the Internet of Things (IoT) and the Web of Things (WoT), can lead to smart spaces overlap, where a certain smart object is configured in different smart spaces, whether fixed or personal. In addition, smart spaces are complex and difficult to model and maintain, as, among other factors, they have to deal with different smart objects. This paper proposes the use of Model-Driven Engineering techniques to enable modeling of ubiquitous computing scenarios, considering the coexistence between fixed and personal smart spaces. The validation of the proposal was carried out by modeling a scenario with the aid of a graphical modeling tool, constructed to produce models conforming to the proposed metamodel.*

Resumo. *A mobilidade dos usuários e o crescente número de espaços inteligentes tanto pessoais quanto fixos, fomentados também pela Internet das Coisas e Web das Coisas, podem levar à sobreposição de espaços inteligentes, onde um determinado objeto inteligente pode ser utilizado em diferentes espaços inteligentes. Além disso, espaços inteligentes são complexos e difíceis de modelar e manter, pois, entre outros fatores, precisam lidar com diferentes objetos inteligentes. Este trabalho apresenta o uso de técnicas de Engenharia Dirigida por Modelos para possibilitar a modelagem de cenários de computação ubíqua, considerando a coexistência entre espaços inteligentes fixos e pessoais. A validação da proposta se deu pela modelagem de um cenário com o auxílio de uma ferramenta gráfica de modelagem construída para produzir modelos em conformidade com o metamodelo proposto.*

1. Introdução

O conceito de ubiquidade vislumbrado por Mark Weiser [Weiser 1991] está sendo concretizado pela recente convergência, disseminação e popularização das tecnologias de rádio, dos microprocessadores e dos dispositivos eletrônicos digitais pessoais, aliadas aos novos paradigmas computacionais, como a Internet das Coisas (*Internet of Things* - IoT) e a Web das Coisas (*Web of Things* - WoT). Esses dispositivos podem ser integrados aos espaços físicos cotidianos, transformando-os, assim, em espaços inteligentes.

Em complementação aos espaços inteligentes tradicionais, que em geral são fixos e confinados a uma determinada área, surgiu recentemente o conceito de espaços inteligentes pessoais [Dolinar et al. 2008]. Um espaço inteligente pessoal (do inglês, *Personal Smart Space* - PSS) é composto pelo conjunto de objetos inteligentes que um usuário carrega consigo e o acompanha durante sua mobilidade. O PSS está sempre disponível e atua como uma interface entre o usuário e os serviços disponíveis em seu próprio espaço inteligente pessoal e entre este e os demais espaços inteligentes, sejam fixos ou pessoais [Dolinar et al. 2008].

Uma abordagem que vem sendo considerada no desenvolvimento de sistemas ubíquos é a adoção de conceitos da Engenharia Dirigida por Modelos (do inglês, *Model-Driven Engineering* - MDE). A MDE considera os modelos como os principais artefatos no desenvolvimento de um sistema. Assim, além de descrever ou documentar um software, os modelos também atuam no seu desenvolvimento, manutenção e operação [Schmidt 2006]. Os modelos são geralmente construídos utilizando-se linguagens de modelagem específicas de domínio (do inglês, *Domain-Specific Modeling Language* - DSML), que, por sua vez, são definidas por um meta-modelo [López-Fernández et al. 2015].

Propõe-se neste trabalho um metamodelo construído sobre os conceitos da MDE, para modelagem de cenários de computação ubíqua compostos de espaços inteligentes pessoais e fixos, objetos inteligentes, aplicações ubíquas e políticas de acesso aos objetos inteligentes e aplicações ubíquas. O metamodelo proposto visa (i) facilitar a modelagem de espaços inteligentes; (ii) permitir tratar os usuários como entidades de primeira classe, desacoplados do espaço fixo e acompanhados de seu próprio PSS; e (iii) possibilitar vocabulário e representação gráfica comuns entre os pesquisadores da área para os conceitos específicos do domínio de espaços inteligentes.

No presente trabalho, o termo *cenário de computação ubíqua* refere-se a:

Cenário de computação ubíqua. *Um conjunto modelado de espaços inteligentes tanto pessoais quanto fixos, pessoas, aplicações ubíquas, objetos inteligentes e suas inter-relações.*

O restante do trabalho está organizado da seguinte forma: a Seção 2 apresenta um cenário para facilitar o entendimento da proposta e detalha o metamodelo proposto para modelagem de cenários de computação ubíqua; a Seção 3 apresenta uma ferramenta de modelagem gráfica e, em seguida, demonstra sua utilização para modelagem do cenário descrito; a Seção 4 apresenta e discute sobre os trabalhos relacionados; por fim, a Seção 5 traz as conclusões acerca do presente trabalho e indica possíveis trabalhos futuros.

2. Proposta

Como forma de facilitar o entendimento da proposta e seu contexto, esta seção traz primeiramente a descrição de um cenário, em sua Subseção 2.1, o qual também serve de base para a validação da proposta, detalhada na Seção 3. Na Subseção 2.2 encontra-se o conceito de sobreposição de espaços inteligentes utilizado neste trabalho e, na Subseção 2.3, é apresentado o metamodelo para modelagem de cenários de computação ubíqua.

2.1. Cenário

A seguir é descrito o cenário de uma casa inteligente na qual residem dois pacientes idosos que necessitam de cuidados médicos. Este cenário possui considerável relevância, que é corroborada pela preocupação de diversos autores em fornecer ferramentas de suporte à assistência domiciliar e permitir o monitoramento remoto de pacientes (e.g., [Carvalho et al. 2011], [Carvalho et al. 2012], [Vieira and Carvalho 2015]).

O Senhor Genaro e a Dona Rute formam um casal de idosos que vivem juntos. Sr. Genaro possui hipertensão arterial, e, por conta disso, desenvolveu recentemente arritmia cardíaca. D. Rute também é hipertensa, além de sofrer de osteoporose. Ambos tomam medicação de forma contínua e recebem recomendação médica para realizar exercícios físicos diários. Na residência do casal foram instalados vários objetos inteligentes, tais como sensores e atuadores fixos, conforme ilustrado na Figura 1. Dispositivos médicos vestíveis são usados por eles de acordo com suas enfermidades. Esses dispositivos enviam informações para uma aplicação ubíqua de saúde pessoal, instalada em seus smartphones. Tanto o Sr. Genaro quanto a D. Rute possuem os seus próprios espaços inteligentes pessoais, compostos por seus dispositivos vestíveis, seus smartphones e a aplicação ubíqua de saúde pessoal. Demais objetos inteligentes são adicionados ao espaço inteligente pessoal sob demanda, como é o caso da TV da sala de estar, que é usada para exibir informações e alertas da aplicação ubíqua de saúde pessoal de cada um dos moradores, tais como mensagens com o horário para tomar as medicações e lembretes para a realização de exercícios físicos ou para aferir seu peso na balança digital, que é compartilhada entre o casal. Além disso, há ainda na casa três outras aplicações ubíquas (de segurança, de bem-estar, e de economia de energia), que estão hospedadas em um espaço inteligente fixo que abrange todos os cômodos da casa.

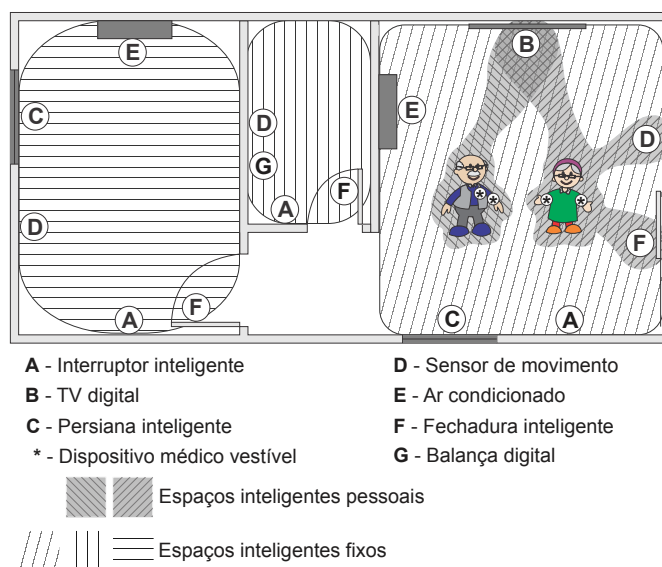


Figura 1. Casa inteligente, na qual dois usuários compartilham objetos inteligentes simultaneamente configurados em diferentes espaços inteligentes.

2.2. Sobreposição de Espaços Inteligentes

Apesar de ser um problema típico e facilmente observável da área de espaços inteligentes, não há uma definição clara sobre o conceito de sobreposição de espaços inteligente. Di-

ante disso, este trabalho propõe e usa a seguinte definição:

Sobreposição de espaços inteligentes. *A sobreposição de espaços inteligentes ocorre quando um ou mais objetos inteligentes estão configurados em espaços inteligentes distintos, sejam estes pessoais ou fixos.*

Por “configurado” considera-se que o objeto inteligente está pronto para ser utilizado pela aplicação ubíqua. Em outras palavras, “configurado” significa que o objeto inteligente foi localizado, instalado e configurado pela aplicação, tornando seus serviços disponíveis perfeitamente utilizáveis.

A sobreposição de espaços inteligentes pode ocorrer pela mobilidade de um usuário que carrega consigo seu espaço inteligente pessoal. Dessa forma, ele pode interagir com outros espaços, tanto fixos quanto pessoais de outro usuário, requisitando acesso a dispositivos que estão configurados em outros espaços inteligentes. Outra forma de sobreposição ocorre quando um único dispositivo está configurado em diferentes espaços inteligentes, novamente, fixos ou pessoais.

2.3. Metamodelo para Cenários de Computação Ubíqua

Esta seção detalha o metamodelo proposto para configuração de cenários compostos por espaços inteligentes pessoais e espaços inteligentes fixos, além de considerar a questão da sobreposição de espaços inteligentes. A arquitetura, suas abstrações e relacionamentos expressam sua sintaxe abstrata, apresentada na Figura 2. Como forma de facilitar a modelagem de instâncias do metamodelo, sua sintaxe concreta foi definida como um conjunto de ícones gráficos. A sintaxe concreta do metamodelo é apresentada na Tabela 1.

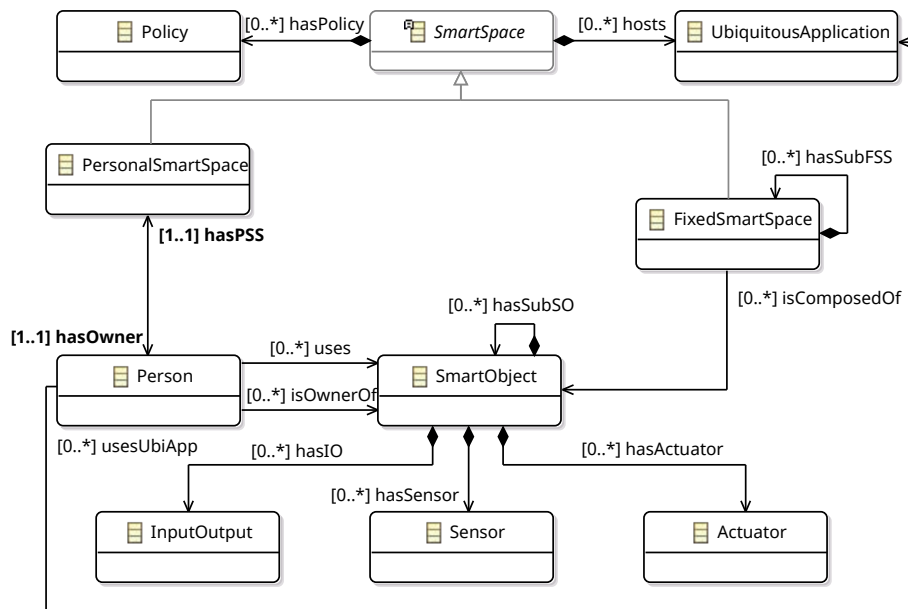


Figura 2. Metamodelo proposto para modelagem de cenários compostos por espaços inteligentes pessoais e fixos.

O metamodelo foi desenvolvido no *Eclipse Modeling Framework (EMF)*¹ sendo, portanto, uma instância do meta-metamodelo Ecore. A seguir apresenta-se a descrição de cada um dos componentes do metamodelo:

¹<http://www.eclipse.org/modeling/emf/>

Tabela 1. Representação gráfica da sintaxe concreta do metamodelo para cenários de computação ubíqua.

Entidade	Representação Gráfica
Policy	
UbiquitousApplication	
PersonalSmartSpace	
Person	
FixedSmartSpace	
SmartObject	
Sensor	
Actuator	
InputOutput	
Associações: hasPolicy, hosts, hasPSS, hasOwner, hasSubFSS, hasSubSO, hasIO, hasSensor, hasActuator	
Associações: uses, isOwnerOf, isComposedOf	

- SmartSpace: um espaço inteligente é uma generalização de espaços inteligentes pessoais (PersonalSmartSpace) e espaços inteligentes fixos (FixedSmartSpace), que, por sua vez, hospedam as aplicações ubíquas (UbiquitousApplication). Essa é uma metaclassse abstrata.
- Policy: um espaço inteligente pode possuir políticas (Policy) para a utilização dos objetos inteligentes (SmartObject) pelas aplicações ubíquas e pelas pessoas (Person) ou para a utilização das aplicações ubíquas pelas pessoas. Estas políticas podem ser, por exemplo, do tipo evento-condição-ação (do inglês, *event-condition-action* - ECA). O detalhamento desta metaclassse foge do escopo deste trabalho. Contudo, é possível adaptar e usar abordagens já existentes em conjunto com o metamodelo.
- UbiquitousApplication: representa uma aplicação projetada para ser executada em um espaço inteligente pessoal (PersonalSmartSpace) ou fixo (FixedSmartSpace). Uma aplicação ubíqua é naturalmente sensível ao contexto, ou seja, ela pode reagir a mudanças “sensoriadas” no ambiente. Essas informações são repassadas à aplicação por meio de um serviço de contexto externo, como o *Hermes* [Veiga et al. 2014] ou *Context Toolkit* [Dey et al. 1999]. A aplicação ubíqua pode se basear no contexto obtido para executar alguma ação e essa ação pode levar em consideração as políticas (Policy) definidas.
- PersonalSmartSpace: define um espaço inteligente pessoal. Cada pessoa, representada por uma entidade Person, possui seu próprio espaço inteligente pessoal e este pode ser estendido, na medida em que novos objetos inteligentes, pertencentes ou não a outros espaços inteligentes, são nele configurados.
- Person: define uma pessoa. As associações possíveis entre uma pessoa e um objeto inteligente são:
 - uses: define que um objeto inteligente pode eventualmente ser compartilhado e usado simultaneamente entre espaços inteligentes, configurando

na sobreposição de espaços inteligentes.

- `isOwnerOf`: define que um determinado objeto inteligente é de uso particular de uma pessoa. Dessa forma, por padrão, um objeto inteligente ligado a uma pessoa por meio da associação `isOwnerOf` não pode ser compartilhado com (*i.e.*, utilizado por) outra pessoa que não seja seu dono.
- `FixedSmartSpace`: define um espaço inteligente fixo. Um espaço inteligente fixo pode ser constituído por outros espaços inteligentes fixos. Por exemplo: cada sala do Instituto de Informática de uma determinada universidade pode ser um espaço inteligente fixo (ou *Fixed Smart Space* - FSS). O prédio do Instituto de Informática, por sua vez, pode ser um FSS, constituído por suas várias salas. E cada instituto de uma universidade pode fazer parte do FSS de um de seus câmpus.
 - `isComposedOf`: um espaço inteligente é composto por objetos inteligentes, que são compartilhados (ou não) entre as pessoas que nele entram e saem.
- `SmartObject`: representa um objeto inteligente. Um objeto inteligente pode ser composto por sensores (`Sensor`), atuadores (`Actuator`) ou interfaces de entrada e saída (`InputOutput`). Um *smartphone*, por exemplo, é um objeto inteligente que possui diversos sensores, tais como acelerômetro e bússola; atuadores, como *vibracall* e alto-falantes; e interfaces de entrada e saída, como tela e botões físicos (liga/desliga, aumenta/diminui volume, etc.). Além disso, um objeto inteligente pode ser constituído por outros objetos inteligentes. Pode-se ainda desejar modelar cada um destes dispositivos como um objeto inteligente independente, por exemplo, um sensor de temperatura. Dessa forma, é possível criar modelos em diferentes níveis de granularidade.

3. Validação da Proposta

Como forma de identificar como os pesquisadores da área de metamodelagem têm validado e avaliado suas propostas de metamodelos, foi conduzida uma Revisão Sistemática da Literatura (RSL). Foram identificados 438 trabalhos nas bibliotecas digitais das bases de pesquisa definidas no protocolo da RSL, dos quais 60 foram detectados como duplicados e 281 foram removidos pelos critérios de exclusão. Dentre os 97 trabalhos incluídos, constatou-se que a forma mais comum de validação dos metamodelos é a construção de uma ferramenta para produzir modelos em conformidade com o metamodelo proposto, em conjunto com a sua utilização para modelagem de cenários reais ou fictícios.

A validação do metamodelo apresentado neste trabalho seguiu, portanto, as tendências adotadas pelos pesquisadores da área. Por limitações de espaço, apenas o cenário apresentado na Subseção 2.1 foi modelado como uma instância do metamodelo proposto. Sendo assim, um editor gráfico foi implementado utilizando o *Eclipse Graphical Modeling Framework* (GMF)². Essa ferramenta foi incrementada com o uso de linguagens da família *Epsilon*³ e sua construção foi apoiada pela ferramenta *Eugenia*, que também integra a família *Epsilon*.

A Figura 3 apresenta a ferramenta de modelagem sendo utilizada para a configuração de um cenário de computação ubíqua. A janela da ferramenta é dividida

²<https://www.eclipse.org/gmf-tooling/>

³<https://www.eclipse.org/epsilon/>

em duas partes: (i) e (ii). Em (i), há a área reservada para a construção do modelo. Em (ii), estão os ícones que representam os conceitos do metamodelo, agrupados em uma barra de ferramentas (ou paleta). A barra de ferramentas é dividida em duas partes: (a) ferramentas para criação de elementos, ou seja, instâncias dos tipos definidos no metamodelo; e (b) ferramentas para criação de associações entre os elementos.

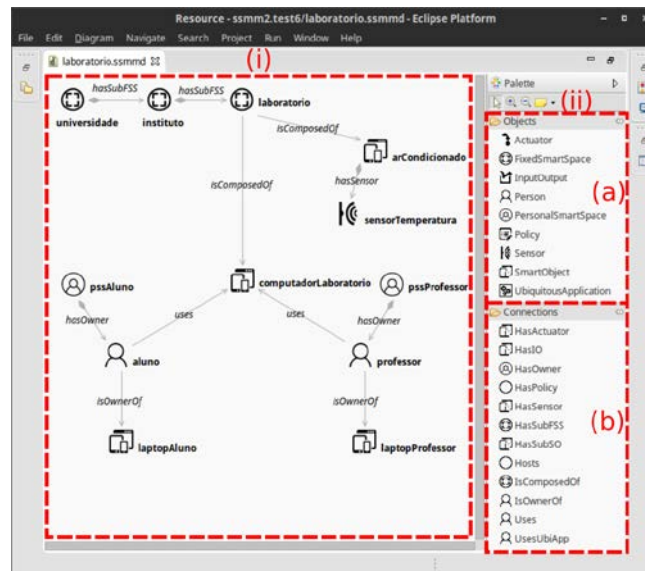


Figura 3. Ferramenta de modelagem sendo utilizada para modelagem de um cenário de computação ubíqua.

A ferramenta de modelagem permite validar o modelo construído para garantir sua conformidade com o seu metamodelo. Caso alguma inconsistência seja encontrada, ela é apontada, permitindo sua correção. Para estender a validação dos modelos, foram definidas regras por meio da linguagem EVL (*Epsilon Validation Language*)⁴, permitindo incrementar as regras sintáticas definidas pelo metamodelo com restrições adicionais, tais como impedir o autorreferenciamento para certos tipos de elementos.

O Código 1 apresenta trechos do código do modelo do cenário, construído por meio da ferramenta de modelagem. Esse código pode ser utilizado, por exemplo, como base para a geração automática de código em alguma linguagem de programação ou ainda como base para sua transformação em outro modelo, este por sua vez baseado em metamodelo próprio. A Figura 4 ilustra a representação gráfica do cenário, tal qual este é modelado na ferramenta. Os rótulos em **negrito** representam os nomes dos elementos na forma de instâncias de tipos do metamodelo. Os rótulos em *itálico* representam as associações entre esses elementos.

Código 1. Instância do metamodelo de cenários de computação ubíqua, representando a configuração do cenário (trechos).

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <ssmm:SmartSpaceDiagram xmi:version="2.0" xmlns:xmi="http://www.omg.org/XMI" xmlns:xsi="http
  ://www.w3.org/2001/XMLSchema-instance" xmlns:ssmm="http://www.example.org/ssmm">
3   <SmartSpace xsi:type="ssmm:PersonalSmartSpace" hasOwner="//@Person.0" name="pssP1"/>
4   <SmartSpace xsi:type="ssmm:PersonalSmartSpace" hasOwner="//@Person.1" name="pssP2"/>
5   <SmartSpace xsi:type="ssmm:FixedSmartSpace" name="house"/>
6   /* suprimido */

```

⁴<https://www.eclipse.org/epsilon/doc/evl/>

```

7 <SmartSpace xsi:type="ssmm:FixedSmartSpace" isComposedOf="//SmartObject.4 //SmartObject.5
  //SmartObject.8 //SmartObject.9" name="bathRoom"/>
8 <UbiquitousApplication name="personalHealth"/>
9 /* suprimido */
10 <SmartObject name="smartPhone"/>
11 <SmartObject name="ecg"/>
12 /* suprimido */
13 <Person uses="//SmartObject.6 //SmartObject.4" isOwnerOf="//SmartObject.0 //SmartObject
  .1 //SmartObject.2" hasPSS="//SmartSpace.0" name="patient1"/>
14 <Person uses="//SmartObject.6 //SmartObject.4" isOwnerOf="//SmartObject.0 //SmartObject
  .3 //SmartObject.2" hasPSS="//SmartSpace.1" name="patient2"/>
15 <Hosts source="//SmartSpace.0" target="//UbiquitousApplication.0"/>
16 /* suprimido */
17 <Hosts source="//SmartSpace.2" target="//UbiquitousApplication.1"/>
18 <HasSubFSS source="//SmartSpace.2" target="//SmartSpace.5"/>
19 <HasSubFSS source="//SmartSpace.2" target="//SmartSpace.4"/>
20 <HasSubFSS source="//SmartSpace.2" target="//SmartSpace.3"/>
21 </ssmm:SmartSpaceDiagram>

```

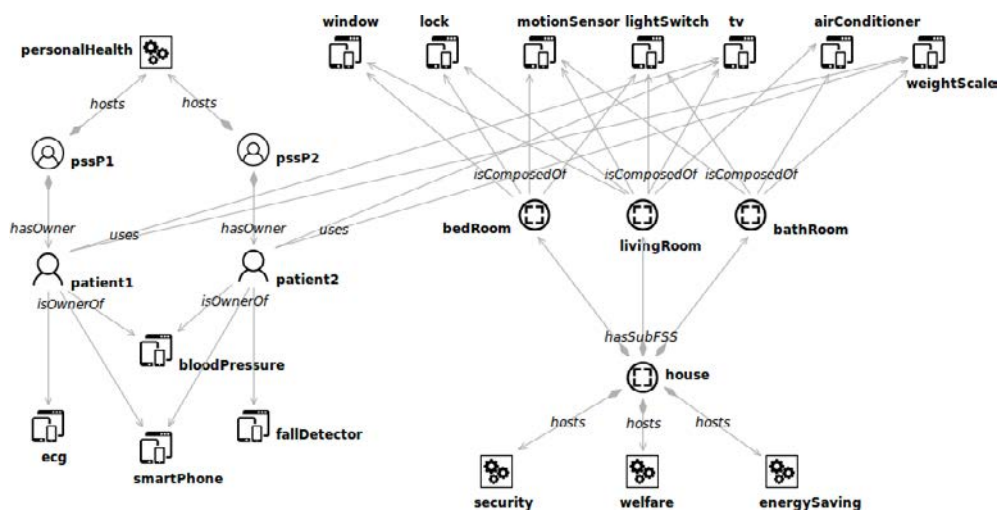


Figura 4. Modelagem do cenário utilizando a ferramenta de modelagem.

Na Figura 4, o usuário identificado por *patient1*, possui configurados em seu espaço inteligente pessoal um sensor de eletrocardiograma (*ecg*), um monitor de pressão arterial (*bloodPressure*) e um aparelho *smartphone*, todos do tipo *SmartObject* e definidos como de uso privado, conforme indicado pela associação *isOwnerOf*. Além disso, o usuário eventualmente utiliza uma balança (*weightScale*) para medir seu peso, a qual está localizada no espaço inteligente fixo *bathRoom* e é compartilhada pelo usuário *patient2*. Um televisor (*tv*), localizado no espaço inteligente fixo *livingRoom*, é utilizado para exibir as informações e alertas enviados pela aplicação ubíqua de saúde pessoal (*personalHealth*). Essas duas últimas associações são do tipo *uses*.

A configuração do espaço inteligente pessoal do usuário *patient2* é composta também por um *smartphone* e um monitor de pressão arterial, além de um sensor de quedas (*fallDetector*). O usuário ocasionalmente confere seu peso na balança compartilhada pelo casal. Semelhante ao que acontece com *patient1*, sua aplicação ubíqua de saúde pessoal utiliza a *SmartTV* da sala de estar para exibir informações e alertas.

Por fim, as aplicações ubíquas de segurança (*security*), de bem-estar (*welfare*) e de economia de energia (*energySaving*) são hospedadas por um espaço inteligente fixo (*house*), que compreende todos os cômodos da casa, abrangendo os espaços inteligentes fixos da sala de estar (*livingRoom*), banheiro (*bathRoom*) e quarto do casal (*bedRoom*).

4. Trabalhos Relacionados

O trabalho de Corredor *et al.* [Corredor et al. 2012] propõe uma metodologia chamada de *Resource-Oriented and Ontology-Driven Development* (ROOD), que disponibiliza um conjunto de ferramentas de modelagem e se baseia nos conceitos de ontologias e arquitetura dirigida a modelos (do inglês, *Model-Driven Engineering* - MDA) para apoiar o desenvolvimento de espaços inteligentes.

Em [Freitas et al. 2014] é apresentada a linguagem 2SML, que permite modelar espaços inteligentes em alto nível, usando a abordagem de modelos em tempo de execução. O 2SML divide a programação do espaço inteligente em dois modelos: modelo do engenheiro e modelo do usuário. Cada um desses modelos possui sua própria linguagem de modelagem, com sintaxes diferentes. Os modelos gerados devem ser processados por uma máquina própria, chamada 2SVM, que possui arquitetura em camadas, inspirada pela CVM (*Communication Virtual Machine*) [Deng et al. 2006].

O projeto PERSIST (*Personal Self Improving Smart Spaces*) [Dolinar et al. 2008] introduz a visão de que os espaços inteligentes pessoais fornecem uma interface entre o usuário e os vários serviços e objetos inteligentes que estão disponíveis, interagindo com outros espaços inteligentes pessoais para criar um ambiente poderoso e flexível.

A nossa proposta difere dos trabalhos apresentados, pois considera que cada pessoa pode formar seu próprio *espaço inteligente pessoal*, e que em determinados momentos pode haver a sobreposição deste com o espaço inteligente configurado em outra aplicação ubíqua (*e.g.*, segurança, bem-estar, economia de energia). O metamodelo para cenários de computação ubíqua proposto apresenta formas de modelar espaços inteligentes, os quais hospedam aplicações ubíquas capazes de lidar com estes fatores.

5. Conclusão

Neste trabalho foram utilizados conceitos de Engenharia Dirigida a Modelos para propor um metamodelo que permite a modelagem em alto nível de um conjunto de espaços inteligentes pessoais e fixos e suas entidades constituintes, aqui denominado de cenário de computação ubíqua. O metamodelo possibilita modelar as formas de interação entre os usuários, representados por seus espaços inteligentes pessoais, e as aplicações ubíquas, objetos inteligentes e demais espaços inteligentes pessoais ou fixos, por meio de uma notação gráfica. Os modelos gerados podem ser utilizados para se ter uma visão geral em alto nível do cenário de computação ubíqua e identificar possíveis inconsistências antes de iniciar o processo de codificação das aplicações ubíquas. Por essa ótica, esses modelos gráficos agem também como forma de documentar o cenário de computação ubíqua.

A validação da proposta se deu por meio da modelagem de um cenário de computação ubíqua, na forma de uma instância dos metamodelo para modelagem de cenários de computação ubíqua proposto, utilizando uma ferramenta de modelagem gráfica desenvolvida para construir modelos em conformidade com esse metamodelo.

O principal trabalho futuro consiste na implementação de transformações M2T (*model-to-text*) na ferramenta de modelagem para permitir a geração de código em nível M0 a partir dos modelos nela construídos. A modelagem e tratamento das políticas de adaptação das aplicações ubíquas com base em informações de contexto obtidas do ambiente também constituem trabalhos futuros.

Agradecimentos

Os autores agradecem ao Instituto Federal de Educação, Ciência e Tecnologia Goiano (IF Goiano) pelo apoio financeiro recebido.

Referências

- Carvalho, S. T., Copetti, A., and Loques Filho, O. G. (2011). Sistema de computação ubíqua na assistência domiciliar à saúde. *Journal Of Health Informatics*, 3(2).
- Carvalho, S. T., Murta, L., and Loques, O. (2012). Variabilities as first-class elements in product line architectures of homecare systems. In *Software Engineering in Health Care (SEHC), 2012 4th International Workshop on*, pages 33–39.
- Corredor, I., Bernardos, A. M., Iglesias, J., and Casar, J. R. (2012). Model-driven methodology for rapid deployment of smart spaces based on resource-oriented architectures. *Sensors*, 12(7):9286–9335.
- Deng, Y., Sadjadi, S. M., Clarke, P. J., Zhang, C., Hristidis, V., Rangaswami, R., and Prabakar, N. (2006). A communication virtual machine. In *Computer Software and Applications Conference, 2006. COMPSAC'06. 30th Annual International*, volume 1, pages 521–531. IEEE.
- Dey, A. K., Salber, D., Abowd, G. D., and Futakawa, M. (1999). The conference assistant: Combining context-awareness with wearable computing. In *The Third International Symposium on Wearable Computers*, pages 21–28. IEEE.
- Dolar, K., Porekar, J., McKitterick, D., Roussaki, I., Kalatzis, N., Liampotis, N., Papaioannou, I., Papadopoulou, E., Burney, S. M., Frank, K., Hayden, P., and Walsh, A. (2008). PERSIST Deliverable D3.1: Detailed Design for Personal Smart Spaces. <http://www.ict-persist.eu/?q=content/persist-deliverables-and-publications>. [Acessado em Abril-2015].
- Freitas, L. A., Costa, F. M., Rocha, R. C., and Allen, A. (2014). An architecture for a smart spaces virtual machine. In *Proceedings of the 9th Workshop on Middleware for Next Generation Internet Computing*, page 7. ACM.
- López-Fernández, J. J., Cuadrado, J. S., Guerra, E., and de Lara, J. (2015). Example-driven meta-model development. *Software & Systems Modeling*, 14(4):1323–1347.
- Schmidt, D. C. (2006). Guest editor’s introduction: Model-driven engineering. *Computer*, 39(2):0025–31.
- Veiga, E. F., Melo e Maranhão, G., and Bulcão Neto, R. F. (2014). Apoio ao Desenvolvimento de Aplicações de Tempo Real Sensíveis a Contexto Semântico. In *IX Workshop de Teses e Dissertações do XX Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, pages 1–4, João Pessoa-PB.
- Vieira, M. A. and Carvalho, S. T. (2015). Configuração de Espaços Inteligentes para Sistemas Ubíquos de Monitoramento de Pacientes Domiciliares. In *Anais da III Escola Regional de Informática de Goiás (ERI-GO 2015)*, pages 19–30, Goiânia-GO, Brazil. Sociedade Brasileira de Computação (SBC).
- Weiser, M. (1991). The computer for the 21st century. *Scientific american*, 265(3):94–104.

Modelagem Genética para Estender o Tempo de Vida de Redes de Sensores Sem Fio com Sorvedouro Móvel*

Sérgio de Oliveira, Pedro Mitsuo Shiroma, Marconi de Arruda Pereira, Rone Ilídio da Silva, Cristiano Maciel da Silva

Departamento de Tecnologia – Campus Alto Paraopeba
Universidade Federal de São João Del Rei – Ouro Branco – MG

{sergiool, pshiroma, marconi, rone, cristiano}@ufsj.edu.br

Abstract: *Mobile Sinks has been used in Wireless Sensor Networks (WSNs) to increase network lifetime, especially to distribute, in homogeneous way, the energy consumption among all nodes. The development of low cost multicopters, aerial vehicles capable of autonomous flights, turn this type of vehicle a good solution to be used as mobile sinks. This work presents a new approach to the problem of how to define multicopter tours to collect data in WSNs, in order to extend the network lifetime. We consider a drone that can stop in a fixed predefined number of points inside a monitored area to collect sensor data. We propose a genetic model that shows how the distribution of the nodes' workload can extend the network lifetime. Simulation were performed to adjust the parameters of the model and to evaluate its performance.*

Resumo: *Sorvedouros móveis têm sido usados em Redes de Sensores Sem Fios (RSSF) visando aumentar o tempo de vida da rede, especialmente por distribuir, de forma mais homogênea, o consumo de energia por todos os nós. O desenvolvimento de drones multirotores, veículos aéreos não-tripulados capazes de voos autônomos, de baixo custo, os tornam aptos a serem usados como sorvedouros móveis. Este trabalho considera que o drone pode parar em um número fixo de pontos da rede, característica dos multirotores, coletando dados de sensoriamento em um número fixo predefinido de pontos da rede. Este trabalho propõe um modelo genético que mostra como a distribuição da carga de trabalho dos nós pode estender o tempo de vida da rede. Simulações foram executadas para ajustar os parâmetros do modelo e avaliar seu desempenho.*

1. Introdução

Redes de Sensores Sem Fio (RSSF) são redes *ad hoc* compostas por nós sensores espalhados em uma área geográfica de monitoramento, com o objetivo de coletar dados diversos e encaminhar esses dados para os nós sorvedouros. O maior problema das RSSF é prolongar o tempo de vida da rede explorando todas as suas funcionalidades. Nos últimos anos, foram apresentadas várias abordagens para o uso de sorvedouros móveis em RSSF visando o aumento do tempo de vida da rede. Visto que os nós mais próximos do sorvedouro tendem a apresentar um consumo maior de energia por sua função de roteamento, eles tendem a se exaurir mais rapidamente na configuração de sorvedouro fixo. O uso de sorvedouros móveis possibilita reduzir essa sobrecarga, distribuindo a função de roteamento pela rede (MA et al, 2013), (BASAGNI

*Este trabalho foi parcialmente financiado pelo CNPQ

et al, 2008) e (WANG et al, 2014). Para tanto, é preciso identificar o melhor caminho e posicionamento para a coleta de dados.

Esse trabalho apresenta uma proposta que considera: i) a possibilidade de que o sorvedouro móvel possa parar sobre os pontos de coleta e aguardar que os dados se sensoriamiento da região sejam enviados até ele; ii) os nós sorvedouros móveis tem energia limitada para percorrer a área monitorada; iii) os nós sensores tem energia limitada e posições, previamente conhecidas, e gastem energia de forma previsível, em função do número de mensagens transmitidas; iv) o raio de comunicação é simétrico e igual para todos os sensores. A partir destas premissas, este trabalho visa identificar qual o melhor posicionamento para que o drone pare e faça a coleta de dados, a cada viagem, até a exaustão da energia da rede, visando maximizar a mesma. Para tanto, apresenta um método para encontrar os melhores pontos de parada através do uso de algoritmos genéticos, visando maximizar o tempo de vida da rede.

Os resultados indicam que é possível estender o tempo de vida da rede significativamente. Comparando com as soluções aleatórias, o incremento no tempo de vida chega a 200%. As soluções obtidas também foram comparadas em função no número de gerações genéticas, representadas por conjuntos de operações do algoritmo genético, indicando um aumento de cerca de 20% após algumas centenas de gerações genéticas. Também foram realizadas comparações com soluções baseadas em centroides. Os resultados, para essas comparações, indicaram a necessidade de uma energia inicial até 500% maior para manter o mesmo tempo de vida da rede em ambas as simulações.

O artigo está organizado da seguinte forma: a seção 2 apresenta os trabalhos relacionados. A seção 3 formaliza o problema. A seção 4 apresenta a estratégia de solução. A seção 5 apresenta os resultados. A seção 6 conclui o documento.

2. Trabalhos Relacionados

A maioria dos trabalhos considerados baseia-se somente no custo de comunicação entre nós sensores em suas abordagens, partindo do pressuposto que maximizar o tempo de vida da rede é equivalente a minimizar o consumo de energia (ROMAO et al. 2013, TASHTARIAN et al., 2012; ALMI'ANI et al, 2010). Embora esta correlação exista, este não é o único fator envolvido, também é preciso considerar a energia residual dos nós e a distribuição do consumo pelos nós da rede ao longo do seu tempo de vida. Existem alguns poucos trabalhos que consideram a energia residual nos nós sensores (LUO e JEAN-PIERRE, 2005; LIANG et al., 2011).

Embora o problema considerado seja NP-difícil, muitas vezes é possível compará-lo com o ótimo para pequenas instâncias. Normalmente ele é formulado como um problema de programação linear inteira (ALMI'ANI et al., 2010; LIANG et al., 2011; MA et al., 2013). Uma exceção é o trabalho de Luo e Jean-Pierre (2005) onde se resolve analiticamente o problema, encontrando a solução exata para o problema de determinação da trajetória do sorvedouro móvel. Entretanto, tal solução só é válida se os nós estiverem distribuídos em torno de um círculo de raio conhecido e seguindo uma distribuição de Poisson. Em seguida, uma heurística é utilizada para encontrar uma solução para o problema de roteamento.

SHI e HOU (2012) apresentam uma modelagem matemática para o problema de sorvedouro móvel em redes de sensores sem fio. Entre as contribuições, há uma redução

do problema, com uma transformação tempo para espaço, de forma a tratar o problema apenas no domínio do espaço, ao invés de tratá-lo no domínio do tempo e espaço. Essa redução tem por objetivo facilitar o processo de otimização. Além disso, mostraram também que é possível discretizar o espaço de soluções, tornando-o finito, visto que, a princípio, seria infinito. Essas transformações aproximam-se do nosso trabalho e premissas, visto que, para o drone multirrotor, a autonomia é função do tempo de voo e não da distância percorrida.

Diversas heurísticas foram propostas para resolver o problema nos casos onde o número de nós sensores é grande o suficiente para invalidar a busca por uma solução exata. ALMI'ANI et al. (2010) propõem uma clusterização usando o algoritmo *k-means* para definir os pontos de coleta. Embora tal clusterização seja um bom critério para minimizar a quantidade média de saltos na rede como um todo, dependendo da topologia da rede é possível que o *k-means* crie *clusters* desbalanceados, deixando poucos nós com um número elevado de saltos que podem, conseqüentemente, causar a desconexão de um número significativo de nós.

ROMAO et al. (2013) propõem um algoritmo genético para determinar o conjunto de *líderes de grupo* que concentrarão os dados coletados, o roteamento de cada nó sensor para o respectivo *líder de grupo*, e um caminho visitando todos os *líderes de grupo* com tamanho menor que o máximo estabelecido. Para evitar um número excessivo de cálculos, o algoritmo genético foca apenas na seleção de quais nós sensores serão considerados como *líderes de grupo*, enquanto as outras etapas são feitas posteriormente.

Alguns trabalhos da literatura demandam prévio conhecimento, pelos nós, do caminho dos sorvedouros móveis (WANG et al, 2014), considerando que o sorvedouro não pode parar até que os dados cheguem até ele. Visto que o controle de posicionamento dos VANTs é eficaz, esses veículos são capazes de parar no ar com precisão para realizar a coleta dos dados de todos os nós de uma região específica. Diferenciam-se, ainda, de outros (SILVA e NASCIMENTO, 2016) por ter como objetivo maximizar o tempo de vida da rede e considerar isso no modelo.

Este trabalho desenvolve uma solução integrada onde o roteamento dos pacotes na RSSF e a determinação das trajetórias dos sorvedouros móveis são tratados de forma conjunta, e não isoladamente. A trajetória do sorvedouro móvel é restrita apenas pela autonomia de voo do drone e os pontos de coleta escolhidos podem estar em qualquer posição do espaço. O roteamento dos pacotes considera não apenas o custo de comunicação entre nós, mas também a energia residual.

3. Formalização do Problema

A rede de sensores sem fio com sorvedouro móvel pode ser representada pela tupla (S, A, r, m) , no qual:

- $S = \{s_1, s_2, s_3, \dots, s_n\}$, o conjunto de n nós sensores, espalhados de forma aleatória sobre a área sensorizada;
- A , a área monitorada, quadrada, com lado de L metros;
- r - alcance do rádio dos nós sensores, fixo e idêntico para todos os nós;

- m – número de posições que o sorvedouro móvel assume para coletar dados.

O resultado do problema deve indicar uma sequência J , ordenada, de conjuntos D_i , de tamanho m , os quais representam as posições assumidas pelo sorvedouro móvel a cada passada i , de forma a maximizar o tempo de vida da rede, indicado como o momento em que 1% dos nós da rede tem sua bateria exaurida (ALMI'ANI et AL, 2010).

O problema consiste em encontrar a melhor sequência J , que resulta no tempo de vida máximo para a rede. Esse problema envolve um número infinito de soluções, visto que o drone pode parar em qualquer ponto na área dentro da qual se encontram os nós sensores. Uma possível discretização dos pontos candidatos a receber o drone, que tornaria os conjuntos de possíveis soluções finitos, pode ser modelada pelo Problema da Mochila (*Knapsack problem*), notadamente NP completo.

4. Estratégia de Solução

Algoritmos genéticos são algoritmos evolucionários, isto é, partem de uma população de soluções viáveis (indivíduos) para o problema, onde cada solução é construída de forma aleatória. Essa população é submetida a operações genéticas (cruzamento, mutação e seleção) de forma que aqueles indivíduos mais aptos (apresentam maior adequação ao problema) são selecionados para as próximas gerações, além de serem mais utilizados nas operações de cruzamento. A avaliação do desempenho do indivíduo no problema (i.e. o grau de “aptidão”) é medido pela função *fitness*. Assim, quando se utiliza um algoritmo genético, faz-se necessário identificar quais são os aspectos mais relevantes da solução a serem considerados, de tal maneira a serem otimizados pelo algoritmo em questão. Particularmente neste trabalho, algoritmos evolucionários podem melhorar a solução final de problemas de posicionamento do sorvedouro móvel em redes de sensores, desde que se conheça previamente o estado atual dos nós, tanto em termos de posicionamento quanto energia.

A modelagem do problema para que seja solucionado com um algoritmo genético deve considerar o conceito de **população**, que representa um conjunto de soluções viáveis (possíveis) para o problema. Cada solução viável é chamada de indivíduo; a **avaliação** da população se dá pela avaliação de cada indivíduo, pela qual cada uma dessas soluções viáveis é avaliada com uma pontuação, que representa a qualidade da solução para o problema. O valor dessa avaliação serve de base para que um indivíduo participe de operações genéticas como **mutação** e **cruzamento**.

A **população**, para este problema, será representada por um conjunto de soluções instantâneas para o posicionamento do sorvedouro móvel. Cada solução permitirá determinar qual o melhor posicionamento para o sorvedouro móvel para a configuração atual da rede, incluindo sua disponibilidade de energia e conectividade. A cada ciclo de sensoriamento, o algoritmo genético será executado, visando identificar os pontos de parada do drone para coletar os dados de sensoriamento. A **avaliação** da qualidade da solução vai considerar, prioritariamente, o impacto da solução no tempo de vida da rede. Essa avaliação vai considerar a energia gasta no roteamento e a energia residual dos nós após o processo. Os nós com menos energia devem ser poupados da função de roteamento, garantindo seu funcionamento por mais tempo.

A operação de **mutação** consistirá no deslocamento de um ponto de coleta do sorvedouro móvel, gerando uma solução diferente da solução usada para a mutação. O deslocamento será de um valor aleatório entre zero e o alcance do rádio, intervalo que se espera alterações efetivas na solução, visto que deslocamentos pequenos não alteram a solução. Vale destacar que os indivíduos resultantes da operação de mutação continuam válidos, isto é, continuam como uma solução viável para o problema. A operação de **cruzamento** será realizada considerando um particionamento geográfico dos pontos de coleta. As novas soluções são geradas tomando metade dos pontos de coletas mais à direita de uma solução e metade dos pontos de coleta mais à esquerda de outra solução.

4.1. Exaustão da energia da rede

Após definida a modelagem, o algoritmo genético é executado de forma exaustiva, até exaurir a energia de 1% dos nós da rede, considerado então o limite para o tempo de vida. O conjunto solução, nesse caso, inclui configurações diversificadas para o posicionamento dos sorvedouros móveis a cada fase de coleta, pois, do contrário, os nós próximos ao sorvedouro móvel teriam sua energia exaurida mais rapidamente.

4.2. Balanceamento da energia

Dois princípios foram aplicados para balancear a energia da rede e, conseqüentemente, prolongar o tempo de vida na rede: 1) roteamento dos pacotes por caminhos de maior energia residual; e 2) consumo mínimo de energia para coleta dos dados. Estes princípios foram usados em dois momentos: 1) na forma como os nós encaminham os dados até o sorvedouro móvel; e 2) como é calculada a função *fitness*. Para tanto, dois parâmetros foram considerados: a energia gasta na fase de coleta e a distribuição da energia residual dos nós. A relação entre a energia gasta e a energia residual em cada fase de coleta deve ser tal que a energia residual deve ser tão balanceada quanto possível após cada fase de coleta.

A escolha do caminho a ser percorrido, considerou também o custo de execução na definição do roteamento, de forma a criar uma solução já balanceada, isto é, evitando sobrecarregar um ou mais nós da rede, de tal maneira a evitar que esses nós se tornassem um gargalo. A disponibilidade de outro caminho com maior energia residual faz com que ele seja escolhido para o encaminhamento dos pacotes, resultando em uma solução mais balanceada. A Figura 1 apresenta um exemplo do funcionamento da escolha do caminho para o roteamento.

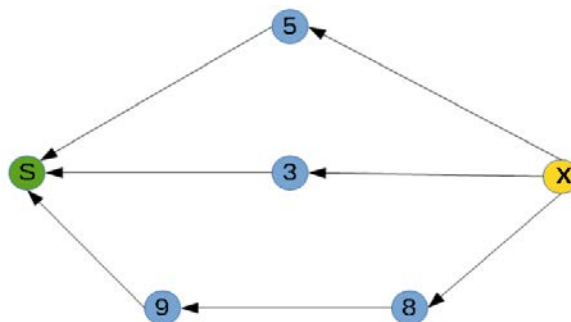


Figura 1 - Exemplo de balanceamento de energia no roteamento.

A escolha da formulação adequada para definir as rotas, bem como a função *fitness*, iniciou-se somando o custo relativo de cada salto do roteamento, definido como

a relação entre o custo do salto, sempre considerado como uma unidade, e sua energia residual ($1/E_i$). Nas primeiras simulações usando essa relação, foi percebida a importância de reforçar a energia residual nessa formulação, razão pela qual foi atribuído, a ela, um fator denominado **fator de apreciação da energia residual α** , chegando então à soma dos custos relativos elevados ao fator α : $(1/e_i)^\alpha$. Para o cálculo da função *fitness*, simplesmente foram somados todos os custos relativos de transmissão de todos os nós da rede, resultando na seguinte formulação:

$$F_i = \sum_{j=0}^n \frac{e_j}{E_{ji}^\alpha}$$

Onde: F_i é o resultado da função de avaliação *fitness* após a fase de coleta i , com n nós sensores, e_j representa a energia gasta pelo nó j na fase i e E_{ji} representa a energia residual do nó j após a fase i .

Encontrar o melhor valor de α que otimizasse o problema aqui apresentado passou a ser parte do problema, de forma que foram executadas várias simulações com esse objetivo, como pode ser visto na seção 5. Foi identificado que $\alpha = 3$ apresenta bons resultados para a heurística proposta.

Para melhor entendimento, seja o exemplo da Figura 1. Supondo o nó S , como nó sorvedouro, ponto onde o drone faz a coleta de dados, e o nó X a ser inserido na coleta, ele tem três possibilidades de encaminhamento de dados ao sorvedouro. Em cada nó, está representada sua energia residual. Na primeira alternativa, superior, como a energia residual do nó é 5, o custo desse caminho é $1/5^3$, igual a 0,008. Na segunda, pelo nó com energia residual 3, o custo é $1/3^3$, igual a 0,037; e, finalmente, na última, o custo seria $1/9^3 + 1/8^3$, resultando 0,003; se tornando, então, a melhor solução, apesar do caminho utilizar de duas unidades de energia para realizar os dois saltos até o sorvedouro.

5. Resultados

Usando o ambiente de programação *Anaconda* para a linguagem *Python*, foi implementado um simulador usando o modelo genético para obter os resultados esperados para este trabalho.

Como massa de testes, foram gerados vários arquivos, variando o número de nós de 50 a 1000, e variando algum parâmetro de entrada, como a função *fitness*, número de nós e número de gerações para o modelo genético. As simulações foram executadas 30 vezes, tomando a média e o desvio padrão, que só não foi registrado neste artigo na seção 5.1, dado o grande volume de dados a ser apresentado.

5.1. Avaliando a Energia por Fase

Com o objetivo de verificar se cada fase de coleta tem um consumo de energia compatível com outras soluções, um conjunto de simulações foi realizado, verificando-se apenas a energia média gasta pelos nós para aquela fase. Foram avaliados uma distribuição uniforme de nós “Grid”, além dos algoritmos de localização de centroides conhecido como K-Means (ALMI’ANI et al., 2010), e o modelo genético proposto neste trabalho com 20, 40 e 100 gerações.

A Figura 2 apresenta os resultados. Como apresenta a energia média para apenas uma fase de coleta, os valores são próximos, mas é possível identificar uma ordem decrescente de energia: grid, genético com 20 gerações, K-Means, genético com 40 gerações e genético com 100 gerações.

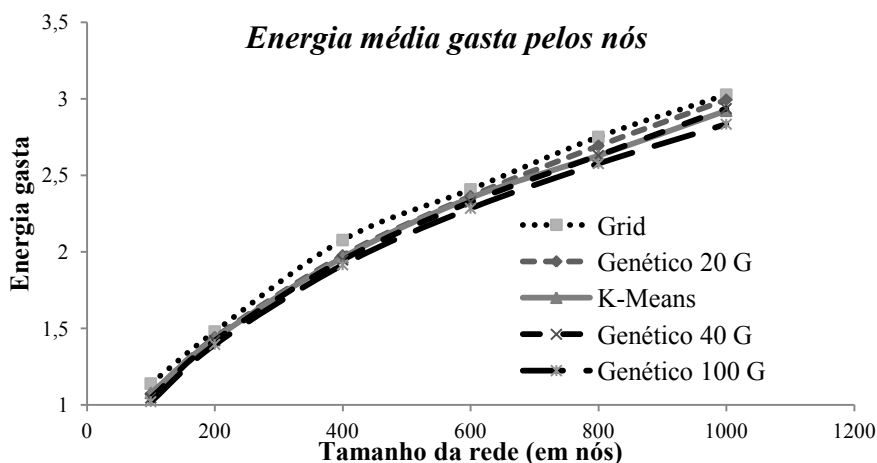


Figura 2 - Comparação da energia gasta com outras soluções

O modelo genético utilizado está configurado para balancear a energia da rede e não apenas minimizar o consumo de energia para uma única fase de coleta. Essa simulação teve por objetivo apenas mostrar que a solução obtida é compatível com outras soluções. É possível configurar o algoritmo genético para minimizar apenas a energia consumida em uma fase única, o que resultaria em resultados melhores, mas não faz parte dos objetivos deste trabalho.

5.2. Cálculo da Função Fitness

A próxima simulação foi realizada com o objetivo de identificar qual seria a melhor função *fitness*. Essa função deverá maximizar o tempo de vida da rede, reduzindo o consumo dos nós e o promovendo de forma balanceada. Partindo da consideração que cada transmissão consuma sempre uma unidade de energia e que a manutenção da energia residual torne sua conservação fundamental, foi escolhida a função $1/E^a$, sendo E a energia residual do nó, como custo relativo de cada transmissão. Para essa simulação foram considerados 50 nós, em uma área 1000x1000 e a execução ocorreu até o limite definido para o tempo de vida da rede.

Tabela 1 - Resultados da seleção da função fitness

Função fitness	Média	Desvio Padrão
$1/x$	29,30	0,92
$1/x^2$	32,87	0,63
$1/x^3$	33,80	0,61
$1/x^4$	33,33	0,88
$1/x^5$	33,40	0,86

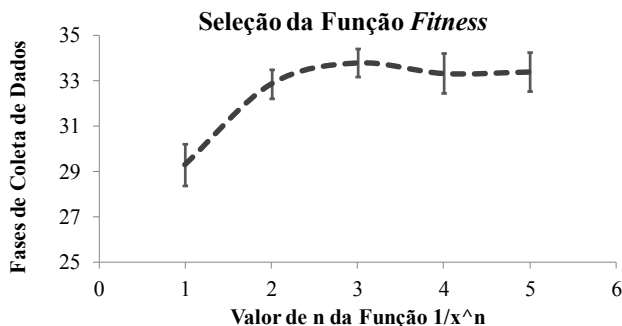


Figura 3 - Resultados da seleção da função fitness

A tabela 1 apresenta os resultados obtidos. A figura 3 apresenta os resultados graficamente. Observa-se que a função $1/x^3$ apresenta os melhores resultados para a massa de testes considerada, com pequena variação para as funções $1/x^2$, $1/x^4$ e $1/x^5$, que também poderiam ser consideradas. A partir da escolha da função, iniciamos os testes com o objetivo de validar a modelagem genética como método para alcançar uma solução boa para o problema de escolher o posicionamento do sorvedouro móvel.

5.3. Estendendo o tempo de vida da rede

O principal resultado esperado nos testes foi o número de operações de sensoriamento e coleta que a rede poderia fazer durante seu ciclo de vida. Foram realizados dois conjuntos de simulações com essa finalidade.

O primeiro conjunto tem por objetivo comparar o modelo genético com outras soluções e, para tanto, foi usada como referência a implementação do *k-means*, como em ALMI'ANI et al. (2010). Para esse primeiro conjunto, o número de nós foi usado como parâmetro, variando entre 100 e 500 nós, mas mantendo fixa a densidade de nós por área, ou seja, a área foi aumentada proporcionalmente ao número de nós para manter a conectividade média, como pode ser visto na figura 4.

Tabela 2 – Tempo de vida em função do número de gerações

K-Means		
N	Média	Desvio Padrão
100	12,00	1,08
200	46,70	9,60
300	92,20	10,17
400	108,63	12,49
500	152,73	23,02
Genético 40 G		
N	Média	Desvio Padrão
100	3,13	0,43
200	12,50	1,81
300	18,63	2,41
400	26,23	3,22
500	34,67	3,92

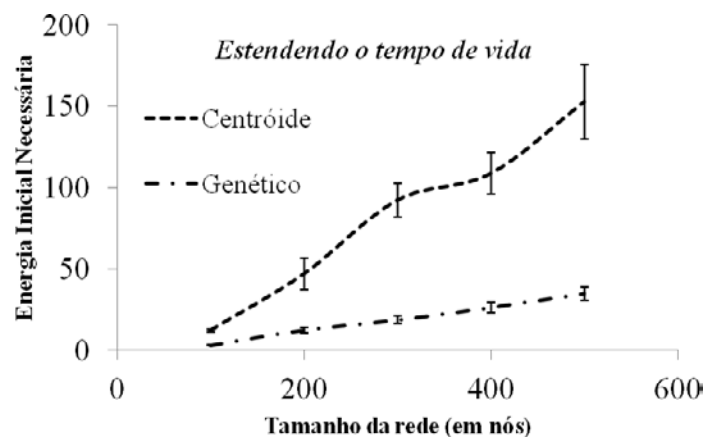


Figura 4 - Energia necessária para tempo de vida

O segundo conjunto tem por objetivo avaliar como o número de gerações do algoritmo genético interfere na qualidade da solução encontrada. Para tanto, foi considerada uma simulação com o número fixo de 50 nós, em uma área 1.000×1.000 , 4 pontos de parada para o nó sorvedouro e energia inicial igual a 100 unidades, executando até que os parâmetros de fim do tempo de vida da rede sejam alcançados (exaustão de energia de 1% dos nós da rede), como visto na Figura 5.

Foi adotado o princípio de que, quanto maior a quantidade de gerações do algoritmo genético, mais a solução encontrada se aproximam da solução ótima. O caso base considerado foi um caso no qual nenhuma operação genética havia sido realizada, e o posicionamento do sorvedouro móvel é realizado de forma aleatória.

Observa-se uma melhora significativa do tempo de vida da rede, em mais de 10 fases de sensoriamento, comparando com uma primeira solução com apenas uma

geração de operações genéticas, para uma energia inicial de 100 unidades. Comparando com a solução aleatória, o ganho é de quase 200%, registrada na Tabela 3 como “zero” gerações genéticas.

Tabela 3 – Tempo de vida em função do número de gerações

Gerações	Média	Desvio padrão
0	27,53	3,95
1	63,27	1,01
2	64,17	0,87
4	65,13	1,01
8	66,63	0,81
16	68,83	0,83
32	70,90	0,71
64	72,43	0,73
128	73,77	0,73
256	74,43	0,63

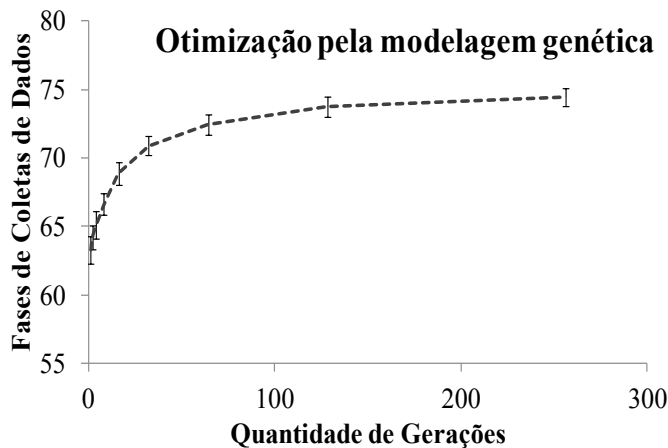


Figura 5 - Tempo de vida em função do número de gerações

Observa-se, ainda, que o tempo de vida da rede estabiliza-se próximo a 75 fases de coleta, indicando um possível limite de uma solução ótima. Considerando as 100 unidades de energia dos nós, o custo de uma unidade para o repasse de cada mensagem e a necessidade de encaminhar algumas mensagens em múltiplos saltos para atingir alguma posição do sorvedouro, essa solução pode ser considerada um resultado bem próximo do esperado.

O máximo teórico de 100 fases só poderia ser atingido se a cobertura do alcance do sorvedouro incluísse todos os nós em comunicação direta, o que não é possível para o exemplo citado, bem como não representaria um problema a ser tratado, visto que, nesse caso, seria fácil obter a solução ótima, na qual todos os nós enviam diretamente ao sorvedouro e o tempo de vida seria igual à energia inicial dos nós.

6. Conclusões e Trabalhos Futuros

Este artigo apresentou a implementação de uma modelagem genética para localizar os pontos de parada de um sorvedouro móvel, implementado por um drone, em uma rede de sensores sem fio, visando equilibrar o consumo de energia da rede e, assim, prolongar o tempo de vida da rede. A abordagem utilizada se difere das principais referências da literatura por considerar um drone multirrotor, capaz de permanecer em uma determinada posição pelo tempo necessário por realizar a coleta dos nós da região onde se encontra.

O deslocamento do drone entre os pontos de parada foi desprezado, uma vez que ele pode ser feito na velocidade máxima de deslocamento do drone, que tem sua autonomia dada em tempo, e não em distância percorrida. Assim, a limitação imposta foi o número de paradas do drone. Os resultados mostram que a modelagem genética é uma boa alternativa para definir os pontos de parada do drone, visando maximizar o consumo de energia da rede. Definindo a função *fitness* de forma a balancear a energia da rede, é possível estender o tempo de vida de forma a exaurir toda a energia da rede uniformemente. Comparando com outras soluções não balanceadas, como *K-Means*, foi

possível identificar um aumento do tempo de vida da rede em algumas vezes, identificando, aí, a maior contribuição deste trabalho.

Como trabalho futuro é possível obter um método alternativo para o cálculo da solução ótima para uma instância mais reduzida de testes, para comparar com a solução genética, com o objetivo de apurar a proximidade da solução alcançada com a solução ótima. Também é interessante considerar nós sensores capazes de variar a potência do rádio durante a transmissão, visando economizar energia nas transmissões para os nós mais próximos. Essa abordagem apresentaria resultados interessantes, visto que privilegiaria transmissões para nós mais próximos, ao mesmo tempo em que deve privilegiar nós com maior disponibilidade de energia.

Referências

- BASAGNI, S., et al. Controlled sink mobility for prolonging wireless sensor networks lifetime. *Wireless Networks*, vol. 14, no. 6, pp. 831–858, 2008.
- ALMI'ANI, J. et al. Energy-efficient data gathering with tour length-constrained mobile elements in wireless sensor networks. In: *IEEE CONFERENCE ON LOCAL COMPUTER NETWORKS*, 2010, Denver. pp.582-589.
- LIANG, W.; LUO, J. Network lifetime maximization in sensor networks with multiple mobile sinks. In: *IEEE CONF. ON LOCAL COMPUTER NETWORKS*, 2011, Bonn. pp.350-357.
- LUO, J.; JEAN-PIERRE, H. Joint mobility and routing for lifetime elongation in wireless sensor networks. In: *IEEE CONFERENCE ON COMPUTER COMMUNICATIONS (INFOCOM)*, 2005, Miami. pp.1735-1746.
- MA, M. et al. Tour planning for mobile data-gathering mechanisms in wireless sensor networks. *IEEE Trans. Vehicular Technol*, vol 62, no. 4, 1472–1483, 2013.
- ROMAO, O. C. et al. Lifetime maximization of hop-and-delay constrained wireless sensor networks with mobile agents. In: *IEEE CONGRESS ON EVOLUTIONARY COMPUTATION (CEC)*, 2013. pp.1083-1090.
- SHOLES, E. Evolution of a UAV Autonomy Classification Taxonomy. In: *IEEE Aerospace Conference*, Big Sky, MT, 2007
- SHI, Y. e HOU, Y. T. Some Fundamental Results on Base Station Movement Problem for Wireless Sensor Networks. *IEEE/ACM Transactions on Networking*, vol. 20, no.4, pp.1054-1067, Aug. 2012
- SILVA, R. I. e NASCIMENTO, M. A. - “On best drone tour plans for data collection in wireless sensor network,” in *Proceedings of the 31st ACM Symposium on Applied Computing*, SAC 2016, Pisa, Italy, April 4-8, 2016, 2016.
- TASHTARIAN, F. et al. Energy efficient data gathering algorithm in hierarchical wireless sensor networks with mobile sinks. In: *INTERNATIONAL CONFERENCE ON COMPUTER AND KNOWLEDGE ENGINEERING*, 2012. pp.350-357.
- WANG, J. et al. A Survey about Routing Protocols with Mobile Sink for Wireless Sensor Network. *International Journal of Future. Generation Communication and Networking*, vol.7, no.5, pp.221-228, 2014.

Planejamento Regional Adaptativo em Sistemas Self-Adaptive de Larga Escala

Eudes S. Andrade, Gabriel F. Sousa, Sandro S. Andrade

Grupo de Pesquisa em Sistemas Distribuídos, Otimização, Redes e Tempo-Real (GSORT)
Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA)
Av. Araújo Pinho, no 39 – Canela – Salvador/BA – CEP: 40.110-150

{eudesandrade, gabrielferreira, sandroandrade}@ifba.edu.br

Resumo. *Com a crescente adoção de sistemas distribuídos de larga escala, tais como clusters para cloud computing, as atividades de implantação e configuração destas soluções se tornam consideravelmente mais complexas. Uma solução promissora é dotar tais sistemas com capacidades de autogerenciamento e diversos padrões de projeto para Sistemas Self-Adaptive estão disponíveis na literatura. Alguns são mais efetivos para operação em ambientes que sugerem a adoção de arquiteturas mais centralizadas. Outros apresentam melhores resultados ao operar de forma mais distribuída. Em ambientes altamente dinâmicos e incertos, no entanto, é difícil encontrar uma única solução efetiva para todos os cenários experimentados. Este trabalho apresenta um modelo adaptativo para implementação de Sistemas Self-Adaptive de larga escala que realizam autogerenciamento com diferentes graus de centralização. Os resultados obtidos mostram que o modelo apresenta desempenho satisfatório tanto em situações onde o sistema monitorado e o ambiente sugerem a adoção de topologias mais centralizadas, quanto em situações onde padrões mais descentralizados apresentam melhor custo-benefício.*

Abstract. *As large-scale distributed systems such as clusters for cloud computing become increasingly adopted, new challenges are posed when deploying and configuring such systems. A promising solution is to endow such systems with self-management capabilities and a number of design patterns for Self-Adaptive Systems is currently available. Some of those patterns are effective when operating in environments that require the adoption of centralized architectures, while others yield better results when controlling large-scale distributed systems. For highly dynamic and uncertain environments, though, a single solution is unlikely to provide effective control in every distinct scenario experienced by the system. This work presents an adaptive model for endowing large-scale Self-Adaptive Systems with self-management capabilities operating with a varying centralization degree. Results show that the proposed model presents improved performance in situations demanding both centralized and decentralized control topologies.*

1. Introdução

Uma série de requisitos e características associados aos sistemas distribuídos modernos tem modificado a forma como tais aplicações são projetadas, desenvolvidas e avaliadas.

Dentre tais características, destacam-se: a utilização de dispositivos multiprocessados, a disponibilidade de paralelismo em soluções altamente distribuídas, a necessidade de integração com outros sistemas e a capacidade de autogerenciamento em ambientes incertos [Northrop 2013]. Apesar da disponibilidade de diversas estratégias de Engenharia de Software para gerência da complexidade gerada por estes desafios, acredita-se que a capacidade humana de compreensão e manipulação de artefatos de software será um fator limitante em um futuro próximo [Huebscher and McCann 2008, Kephart and Chess 2003].

Tendo em vista que a maior parte dos desafios acima apresentados impactam diretamente os requisitos não-funcionais da aplicação – e que estes são amplamente influenciados por decisões arquiteturais tomadas em tempo de projeto – pesquisas recentes indicam que a transferência de decisões de projeto para o tempo de execução (*runtime*) é uma solução promissora no gerenciamento desta complexidade. Para tanto, os sistemas computacionais precisam apresentar alguma capacidade de autogerenciamento [Huebscher and McCann 2008, Kephart and Chess 2003, Salehie and Tahvildari 2009]. Adicionalmente, características de Computação Ubíqua comumente presentes, por exemplo, em aplicações da Internet das Coisas e *Cyber-Physical Systems*, frequentemente demandam a necessidade de autogerenciamento. Um Sistema *Self-Adaptive* (SSA) é aquele que avalia o seu próprio comportamento e o modifica quando esta avaliação indica que: *i*) o seu propósito principal não está sendo efetivamente cumprido; ou *ii*) uma melhor funcionalidade e/ou desempenho pode ser alcançado [DARPA 1997].

Muitas arquiteturas para autogerenciamento estão atualmente disponíveis na literatura [Patikirikoral et al. 2012]. Ainda assim, a maioria delas apresenta desempenho satisfatório apenas em sistemas centralizados [Weyns 2010]. A dificuldade em alcançar autogerenciamento efetivo em ambientes descentralizados decorre do *trade-off* comumente encontrado em sistemas distribuídos: conhecimento do estado global do sistema vs. *overhead* de comunicação. Pode-se ilustrar esse *trade-off* analisando *clusters* para suporte a *cloud computing*. Em tais situações, quanto maior o *overhead* de comunicação (e, portanto, o conhecimento sobre o estado global do sistema), maior será a capacidade do sistema coordenar adaptações em caso de mudanças no ambiente. Por outro lado, um maior consumo de recursos pode diminuir a utilização útil do *cluster* para fins de computação. O objetivo da adoção de arquiteturas para autogerenciamento descentralizado, em tais cenários, é reagir ao *trade-off* que envolve a capacidade de atendimento da *cloud* frente às diversas configurações de ambiente apresentadas ao longo do tempo.

Este trabalho apresenta o projeto e avaliação de um modelo adaptativo para autogerenciamento em SSA de larga escala. Esse modelo prevê a reconfiguração dinâmica da topologia e forma de comunicação entre os múltiplos componentes locais de controle. Com esta abordagem, decisões de projeto acerca da topologia de controle a ser adotada são levadas para o tempo de execução. Com isso, obtém-se desempenho satisfatório de controle tanto em situações onde o sistema monitorado e o ambiente sugerem a adoção de topologias mais centralizadas, quanto em situações onde padrões completamente descentralizados apresentam um melhor *trade-off*. O mecanismo de reconfiguração dinâmica da arquitetura de controle aqui proposto foi modelado e avaliado via simulação de eventos discretos. Resultados indicam que houve uma economia estatisticamente significativa de recursos computacionais ao utilizar o modelo proposto.

O restante deste trabalho está organizado como segue. A Seção 2 apresenta os

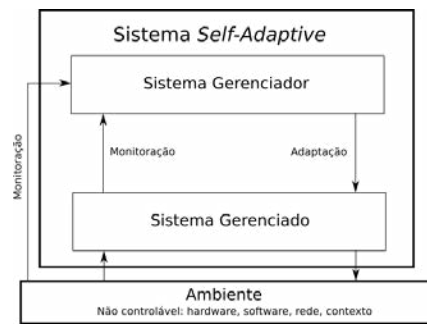


Figura 1. Elementos arquiteturais principais de um Sistema Self-Adaptive.

fundamentos sobre SSA e seus principais desafios. A Seção 3 explica o mecanismo para reconfiguração dinâmica de arquiteturas de controle aqui proposto. A Seção 4 apresenta detalhes sobre as avaliações via simulação realizadas, enquanto a Seção 5 discute e analisa os resultados obtidos. A Seção 6 apresenta os principais trabalhos relacionados e, finalmente, as conclusões e sugestões de trabalhos futuros são discutidos na Seção 7.

2. Sistemas Self-Adaptive de Larga Escala

O autogerenciamento é uma estratégia de projeto utilizada para os mais diversos fins. A despeito disto, a maior parte destas soluções realiza a migração de decisões – anteriormente tomadas em tempo de projeto – para o tempo de execução. Com este artifício, busca-se minimizar possíveis inflexibilidades/ineficiências decorrentes de comprometimentos prévios com determinadas decisões arquiteturais. A ocorrência de situações não previstas pelo arquiteto de *software* é consequência, sobretudo, do dinamismo dos ambientes de execução das aplicações.

Uma definição de SSA comumente adotada na literatura é aquela proposta em [DARPA 1997]: um SSA é aquele que avalia o seu próprio comportamento e o modifica quando esta avaliação indica que: *i*) o seu propósito principal não está sendo efetivamente cumprido; ou *ii*) uma melhor funcionalidade e/ou desempenho podem ser alcançados [DARPA 1997]. Um SSA é geralmente caracterizado pela presença de dois elementos arquiteturais principais, ilustrados na Figura 1. O primeiro – denominado **sistema gerenciado** – é o elemento responsável pela implementação das regras de negócio da aplicação. O segundo elemento – denominado **sistema gerenciador** – é responsável pela implementação das ações necessárias para a realização das adaptações.

Embora diferentes mecanismos para autogerenciamento tenham sido propostos nos últimos anos, a maior parte deles baseia-se na utilização de algum tipo de *loop* de adaptação. Uma arquitetura de referência amplamente adotada no projeto destes *loops* de adaptação é a arquitetura MAPE-K [Kephart and Chess 2003]. Esta arquitetura define a execução, em sequência, das seguintes fases: **M**onitoramento, **A**nálise, **P**lanejamento e **E**xecução – com o auxílio de uma base de conhecimento (**K**nowledge Base). A fase de monitoramento é responsável pela obtenção de informações sobre o estado atual do sistema/ambiente. A atividade de análise realiza a avaliação das informações obtidas na fase de monitoramento. Após a avaliação destes dados, delibera-se pela necessidade ou não da realização de uma adaptação. A fase de planejamento é responsável pela definição de quais atividades de adaptação devem ser realizadas, com o objetivo que o sistema retorne a

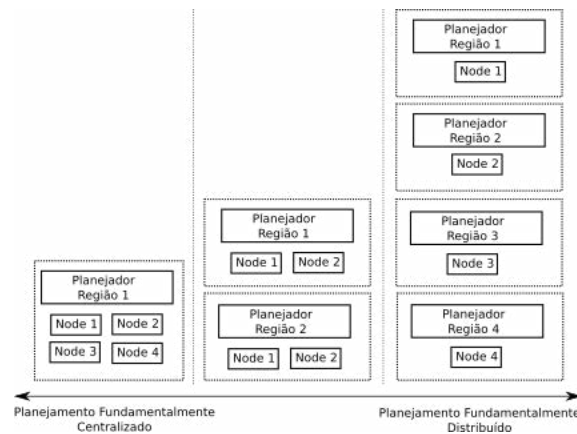


Figura 2. Diferentes instâncias do padrão *Regional Planning*.

um estado particular de interesse. Por fim, a atividade de execução realiza a implantação, no sistema gerenciado, das ações de adaptação definidas na fase de planejamento.

Em ambientes altamente distribuídos, heterogêneos e complexos, um único *loop* de controle pode não ser suficiente para gerenciar toda a sua estrutura [Weyns 2010]. Nestes cenários, portanto, múltiplos *loops* devem ser empregados para controlar as diferentes partes (múltiplos sistemas gerenciados) que compõem o sistema. Esta associação de *loops* de controle aos diversos nós do sistema distribuído traz uma complexidade a mais durante o projeto da solução: a coordenação e interação entre estes múltiplos *loops*.

Em [Weyns 2010], um conjunto de seis padrões arquiteturais para implementação de autogerenciamento descentralizado é apresentado. Para cada padrão, são apresentados os atributos de qualidade decorrentes da sua adoção e as restrições impostas pelas decisões que compõem a solução. Um dos padrões apresentados é o *Regional Planning* – recomendado para situações onde diversas partes integradas (regiões) de um sistema demandam não apenas adaptações locais (dentro das regiões) mas também adaptações globais (que extrapolam os limites das regiões). A solução proposta pelo *Regional Planning* prevê um único componente *P* (*planning*) para cada região. Este planejador regional é responsável por coletar as informações de todos os componentes que fazem parte da sua região, para que o planejamento seja realizado. Planejadores regionais coordenam entre si (sobre múltiplas regiões) para a realização do planejamento da adaptação global.

A Figura 2 apresenta diferentes instâncias do padrão *Regional Planning*, com diferentes tamanhos de região. Na figura, dois tipos de nós de controle são apresentados. O primeiro deles apresenta apenas o componente *P* (planejador da região) e cada região é controlada por apenas um nó do tipo *P*. O segundo tipo apresenta os outros três componentes MAPE-K restantes: MAE (Monitoramento, Análise e Execução). Em cada região, existem múltiplos nós de controle do tipo MAE, que realizam o monitoramento do seu sistema local e do seu ambiente de execução. O componente *A* realiza uma análise local dos dados monitorados e reporta resultados para o planejador *P* da sua região. O planejador regional *P*, por sua vez, planeja a adaptação local da região após cooperar com os planejadores *P* das outras regiões. Com este arranjo, a adaptação realiza a busca por adaptações globais, extrapolando os limites do monitoramento e análise locais.

Como consequência da adoção do padrão *Regional Planning*, espera-se uma

redução na quantidade de dados trafegados, uma vez que o monitoramento e análise são realizadas localmente. A análise local dos dados monitorados reduz a quantidade de dados e a frequência com que interações serão realizadas com o nó central da região. Por outro lado, a necessidade de agregação local – no nó de planejamento – dos resultados das diversas operações de análise pode implicar em *overhead* excessivo. Outra desvantagem é a necessidade de uma fase de planejamento demasiadamente detalhada, uma vez que as fases de execução não coordenam. Finalmente, a adoção do padrão *Regional Planning* requer a definição – em tempo de projeto – do tamanho de região com o qual o sistema será configurado. Consequentemente, ambientes diferentes podem demandar diferentes tamanhos de região para que o autogerenciamento apresente desempenho satisfatório. Em ambientes altamente dinâmicos e incertos, autogerenciamento satisfatório só é obtido com a variação – em tempo de execução – do tamanho de região adotado.

3. O Mecanismo de Planejamento Regional Adaptativo

Este trabalho apresenta o projeto, implementação e avaliação de um modelo adaptativo para autogerenciamento em SSA de larga escala. O modelo permite a reconfiguração dinâmica da topologia do sistema gerenciador e da forma de comunicação entre os múltiplos componentes locais do controle. Com esta abordagem, decisões referentes à topologia de controle a ser adotada (tamanho da região de planejamento) são levadas para o tempo de execução. Com isso, espera-se obter desempenho satisfatório de autogerenciamento tanto em situações onde o sistema gerenciado/ambiente sugerem a adoção de topologias mais centralizadas, quanto em situações onde padrões completamente descentralizados são mais adequados. Em ambientes altamente dinâmicos e incertos, é comum que uma determinada topologia – adequada para um dado cenário – não seja mais satisfatória após a ocorrência de mudanças no ambiente.

Este trabalho viabiliza o projeto de sistemas gerenciadores de tal modo que eles mesmos sejam tratados como SSA. Dessa forma, decisões acerca dos limites das regiões em arquiteturas que seguem o padrão *Regional Planning* não mais são realizadas *off-line*. Com o modelo proposto, em função do seu estado interno e das condições do ambiente, a arquitetura do sistema gerenciador será reconfigurada para adotar desde abordagens mais centralizadas até os extremos providos pelos padrões mais distribuídos.

3.1. Motivação

Conforme discutido na Seção 2, a adoção do padrão *Regional Planning* apresenta um *trade-off* que envolve a qualidade da adaptação vs. a quantidade de dados trafegados (e localmente mantidos) nos nós responsáveis pelo planejamento. Este *trade-off* é comumente administrado pelo arquiteto de software, em tempo de projeto. Para tanto, os tamanhos das regiões são fixos e definidos antecipadamente, sobretudo considerando as situações de maior sobrecarga previstas para a operação do sistema.

Visto que no padrão *Regional Planning* o tamanho das regiões é definido em tempo de projeto, os sistemas gerenciadores tendem a operar de forma não-ótima quando as condições ambientais se desviam daquelas consideradas em tempo de projeto. Para situações de sobrecarga do sistema – onde um máximo de controle se faz necessário – os controles regionais atuam de maneira satisfatória. Por outro lado, em situações onde a utilização de tamanhos menores de região seria suficiente para garantir a qualidade de operação do sistema, um superprovisionamento de recursos estará sempre presente.

A possibilidade de utilização de um modelo onde a decisão acerca do tamanho da região não mais seja uma definição a ser administrada em tempo de projeto – mas sim uma decisão levada para o tempo de execução – é o que motiva a adoção do Planejamento Regional Adaptativo proposto neste trabalho. Ao utilizar um tamanho variável de região, o sistema gerenciador terá a possibilidade de operar de maneira ótima, independente dos diferentes estados apresentados pelo ambiente e sistema gerenciado ao longo do tempo.

3.2. O Mecanismo Proposto

O Planejamento Regional Adaptativo aqui proposto implementa uma variação do padrão *Regional Planning*. Diferente do *Regional Planning* convencional, o Planejamento Regional Adaptativo transfere para o tempo de execução a decisão acerca do tamanho das regiões. Dessa forma, o sistema gerenciador passa a ser, ele mesmo, um SSA que rejeita perturbações no ambiente através da variação no tamanho das regiões de planejamento.

Para isso, um segundo nível de controle (metacontrole) é inserido na arquitetura. Os tamanhos de região que antes eram fixos – definidos *off-line* de maneira a suportar os cenários de pior caso – passam a ser agora mantidos em tempo de execução pelo metacontrole. Com isso, espera-se obter desempenho satisfatório de autogerenciamento tanto em situações onde o sistema gerenciado/ambiente sugerem a adoção de regiões maiores (planejamento global), quanto em situações onde o uso de pequenas regiões (planejamento local) já é suficiente. Trata-se, portanto, de adicionar ao padrão *Regional Planning* um segundo nível de adaptação.

Em momentos onde o sistema gerenciado e o ambiente demandem menos da adaptação, o tamanho das regiões será diminuído pelo metacontrole. Dessa forma, recursos serão poupados. Por outro lado, em momentos onde o metacontrole verifica que o componente MAPE-K responsável pelo planejamento regional necessita de monitoramento e análises mais globais – de forma a garantir os requisitos de qualidade acordados – o tamanho das regiões é aumentado. A fundamentação para tal investigação envolve os impactos decorrentes da variação do tamanho das regiões. A hipótese aqui investigada considera que a adoção de regiões grandes traz como vantagem a maior presença de informações globais, o que resulta na possibilidade de realização de adaptações mais efetivas. Por outro lado, a adoção de regiões pequenas beneficia aspectos tais como disponibilidade e escalabilidade. Sendo assim, o objetivo geral da solução aqui apresentada é regular, de forma automática, o *trade-off* existente entre a efetividade das adaptações vs. a disponibilidade e escalabilidade da solução.

Algumas premissas foram consideradas nesta proposta. Primeiro, assume-se que o *overhead* introduzido pelo metacontrole é desprezível. Visto que as operações de reconfiguração geralmente envolvem a modificação de parâmetros, tais como tamanho de *buffers* ou *thread pools*, o custo do trânsito de tais parâmetros na rede é consideravelmente menor que transitar os dados da aplicação. Segundo, todas as regiões possuem sempre o mesmo tamanho. Embora tal premissa eventualmente não seja verificada em *clusters* com bastante heterogeneidade, suportar diferentes tamanhos de região traria uma complexidade maior ao modelo e será analisado em trabalhos futuros. Por fim, considera-se que o *cluster* possui *jobs* sendo demandados a todo momento, o que viabiliza o contínuo monitoramento do desempenho das aplicação e do autogerenciamento.

4. Avaliação

O mecanismo de Planejamento Regional Adaptativo aqui proposto foi modelado e avaliado via simulação de eventos discretos. Os resultados da simulação foram avaliados considerando dois grupos de controle: o primeiro, decorrente da execução da simulação utilizando o Planejamento Regional Adaptativo proposto neste trabalho; e o segundo, utilizando o padrão *Regional Planning* convencional, com o uso de um tamanho fixo de região de planejamento. As seguintes hipóteses foram investigadas no experimento:

- Hipótese alternativa (H_1): mantida a mesma qualidade do serviço, o tamanho médio da região (e, portanto, o consumo de recursos) ao utilizar o Planejamento Regional Adaptativo aqui proposto (T_{PRA}) é menor que o tamanho médio da região ao utilizar o *Regional Planning* convencional (T_{PR}).
- Hipótese nula (H_0): mantida a mesma qualidade do serviço, o tamanho médio da região (e, portanto consumo de recursos) ao utilizar o Planejamento Regional Adaptativo aqui proposto (T_{PRA}) não é menor que o tamanho médio da região ao utilizar o *Regional Planning* convencional (T_{PR}).

4.1. Simulação

O modelo da simulação define o sistema gerenciado, gerenciador e metacontrole como um único sistema distribuído. Este sistema distribuído é modelado como um grafo $E(N, C)$. O grafo E é composto por um conjunto finito de nós, denominado $N = \{N_1, N_2, \dots, N_n\}$ e um conjunto finito de canais de comunicação $C = f(N_i, N_j) | N_i, N_j \in N$. Vértices representam nós (sistemas gerenciados e gerenciadores) e as arestas representam canais de comunicação. O grafo E implementa os detalhes referentes ao sistema simulado, tais como os componentes que simulam os nós do sistema, informações acerca das regiões e eventuais perturbações (P) presentes no ambiente.

Uma vez que os nós N têm o seu tempo de serviço afetado diretamente pela qualidade da adaptação, tem-se que o tempo de serviço das operações (TS) é função do tamanho das regiões (R). Além do tamanho das regiões, nota-se que o tempo de serviço das operações é diretamente influenciado pela perturbação (P) presente no ambiente. Dessa forma, tem-se que TS é função tanto do tamanho das regiões quanto da perturbação existente no ambiente: $TS(R, P)$. A perturbação (P) foi modelada através de uma variável aleatória, cuja intensidade é obtida utilizando a distribuição normal, ao passo que sua frequência de ocorrência é denotada por uma distribuição exponencial.

Além das relações definidas na função $TS(R, P)$, uma variável aleatória K foi inserida no cálculo do tempo de serviço das operações para representar componentes estocásticos internos ao nó. Os valores de K são obtidos de acordo com a distribuição *Erlang* [Jain 1991]. E armazena ainda a fila de operações a ser executada. O modelo proposto considera a existência de uma fila de operações constantemente alimentada e representa *jobs* demandados por usuários. Os nós presentes no sistema atuam consumindo as operações que aguardam por processamento. O tempo de chegada de novas operações na fila foi modelado através de uma variável aleatória com distribuição exponencial.

Dois níveis de adaptação são observáveis no modelo proposto. O primeiro nível implementa o padrão *Regional Planning* e é responsável por garantir a adaptação nos nós. Esse controle é implementado na função que calcula o tempo de serviço das operações

$TS(R, P)$. O tamanho das regiões (R) é a variável que representa a contribuição, do controle de primeiro nível, no tempo de serviço das operações. O segundo nível implementa a operação de metacontrole (MC), responsável por adaptar os controladores de primeiro nível, sobretudo atuando no tamanho das suas regiões. O MC atua no tamanho das regiões da seguinte forma: se a quantidade de ciclos *idle* realizados é maior que um determinado valor de referência, o tamanho das regiões é reduzido. Por outro lado, caso a quantidade de ciclos *idle* mantenha-se abaixo do valor de referência, o MC atuará aumentando o tamanho das regiões. Apesar de o modelo não depender diretamente da lei de controle, para que a simulação pudesse ser executada fez-se necessário optar por uma implementação concreta. Nesta simulação, optou-se por utilizar controladores PID (Proporcional-Integral-Derivativo) [Hellerstein et al. 2004] para realização da adaptação.

A simulação foi implementada utilizando a linguagem de programação *Python* e o *framework* para simulação *SimPy* (<http://simpy.readthedocs.org>). Pacotes adicionais foram utilizados para geração dos números aleatórios de acordo com as distribuições definidas no modelo. A análise dos resultados foi realizada com a ferramenta *R* (<http://www.r-project.org>), com o auxílio de módulos para execução dos testes estatísticos.

4.2. Análise dos Resultados

A análise dos dados foi realizada com base em um estudo comparativo com dois diferentes grupos de dados. O primeiro grupo representou a execução da simulação utilizando o padrão *Regional Planning* convencional. O segundo grupo, por sua vez, representou a execução da simulação com a utilização do Planejamento Regional Adaptativo proposto neste trabalho. O experimento apresentou um único fator: o tamanho das regiões em tempo de execução. Esse único fator apresenta dois diferentes níveis: Planejamento Regional Adaptativo (*PRA*) ou *Regional Planning* convencional (*PR*).

Para cada nível foram executadas cinquenta replicações utilizando as mesmas sementes de geração de valores para as variáveis aleatórias. A quantidade de ciclos *idle* foi medida para cada uma das replicações. Os resultados das cinquenta replicações foram inicialmente analisados com os testes de *Anderson-Darling* e *Levene*, com o objetivo de analisar a normalidade e homoscedasticidade dos resultados. Com base nos resultados destes testes, decidiu-se pela avaliação da hipótese nula usando ou testes paramétricos (*t-test*) ou testes não-paramétricos (*Wilcoxon*).

5. Resultados

Para realização dos testes considerando o nível *PR* (tamanho fixo das regiões) foi necessário definir um tamanho fixo para as regiões. Este tamanho foi o mínimo necessário para que o controle tivesse condições de adaptar o sistema satisfatoriamente na situação de pior caso. Uma vez que o tempo de serviço TS é função do tamanho da região e do nível de perturbação verificado em dado instante, calculou-se o tamanho mínimo de região necessário para suportar o ambiente nos momentos de maior perturbação possível. A perturbação gerada no ambiente é uma variável aleatória com distribuição normal. Logo, considerou-se como perturbação máxima o seu valor médio somado ao erro máximo (valor de referência – valor observado). A segunda execução da simulação foi realizada considerando o nível *PRA* (tamanho variável das regiões). A configuração inicial da simulação inicia o tamanho da região com o mesmo valor utilizado no nível *PR*, porém, com o decorrer da simulação, o tamanho da região é administrado pelo metacontrole.

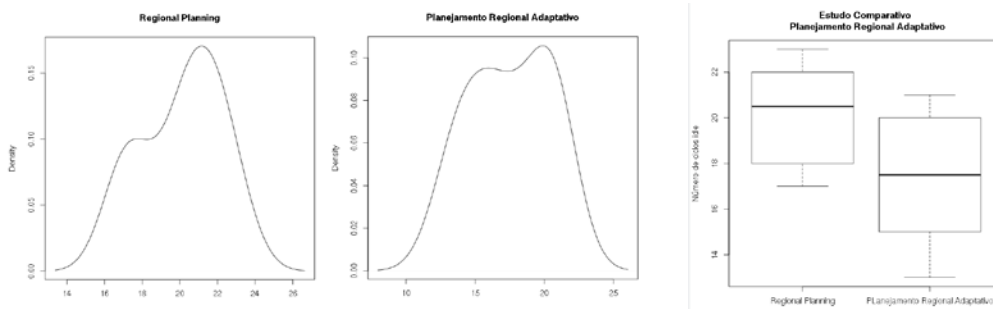


Figura 3. Distribuição dos dados resultantes da simulação.

Uma investigação acerca da normalidade e homoscedasticidade dos dados foi realizada, com o objetivo de decidir se a hipótese nula do experimento poderia ser analisada através de um teste paramétrico (*t-test*). O resultado da análise de normalidade, com um nível de significância $\alpha = 0.01$, apresentou um *p-value* igual a 0.85. A análise de homoscedasticidade apresentou um *p-value* igual a 0.05. Dos resultados anteriores, concluiu-se que ambos os testes falharam e, portanto, testes não-paramétricos devem ser utilizados. A Figura 3 apresenta os resultados obtidos.

Considerando as características de normalidade e homoscedasticidade dos dados, concluiu-se que o *t-test* não poderia ser utilizado como ferramenta para investigação da hipótese nula. Optou-se então pela utilização do teste de *Wilcoxon* (não-paramétrico). O resultado do teste de *Wilcoxon* para as amostras anteriores apresentou, com um nível de significância $\alpha = 0.01$, um *p-value* igual a 0.85. O resultado obtido indica que a hipótese nula H_0 pode ser rejeitada e, portanto, que as amostras avaliadas apresentam melhora significativa dos resultados. Uma representação gráfica do estudo comparativo dos resultados é apresentada na Figura 3.

6. Trabalhos Relacionados

Diferentes padrões de projeto podem ser encontrados na literatura para tentar resolver o problema do planejamento [Abuseta and Swesi 2015, Weyns 2010]. Um destes padrões de projeto é o "*Information Sharing*", que restringe a comunicação entre os componentes presentes no padrão descentralizado. Nesta organização, apenas o componente de monitoramento *M* pode se comunicar com outros componentes externos. Esta abordagem resulta em soluções com melhor escalabilidade de comunicação do que o controle totalmente descentralizado, agilizando o processo de execução da adaptação. Porém, este padrão não é indicado para ambientes altamente dinâmicos pois, como as análises e planejamentos são feitos localmente, as adaptações acabam sendo sub-ótimas.

Outra possibilidade é o padrão de projeto denominado controle hierárquico. Esta abordagem define uma separação de *concerns* por camadas, onde cada nível executa seu próprio *loop* MAPE-K. Os níveis inferiores, com seus componentes de monitoramento e execução, lidam diretamente com o subsistema gerenciado. A camada mais alta se preocupa com a adaptação geral do sistema. O padrão de controle hierárquico permite gerenciar a complexidade do SSA. Porém, esta abordagem não é indicada para ambientes em constante mudança pois a divisão em camadas permite que existam dois planejamentos de adaptação, sendo eles global e específico.

7. Conclusão

Este trabalho apresentou o Planejamento Regional Adaptativo – uma extensão do padrão arquitetural *Regional Planning* que promove uma economia de recursos computacionais e menor necessidade de provisionamento de ambiente em SSA de larga escala. Diferente do padrão *Regional Planning*, o modelo proposto dotou, tanto o sistema gerenciador quanto o sistema gerenciado, de capacidades de autogerenciamento. Com isso, viabiliza-se a execução do sistema gerenciador utilizando tamanhos reduzidos de regiões, o que implica em uma economia de recursos. O mecanismo foi modelado e avaliado via simulação de eventos discretos. Dois grupos foram comparados: controle utilizando o padrão *Regional Planning* convencional e controle utilizando o Planejamento Regional Adaptativo proposto. A análise dos resultados indica que, com um nível de significância $\alpha = 0.01$, houve um aumento de desempenho do sistema gerenciador ao utilizar o Planejamento Regional Adaptativo.

Referências

- [Abuseta and Swesi 2015] Abuseta, Y. and Swesi, K. (2015). Design patterns for self adaptive systems engineering. *CoRR*, abs/1508.01330.
- [DARPA 1997] DARPA, B. A. A. (1997). Self-adaptive software. Technical Report 98-12, DARPA (Defense Advanced Research Projects Agency).
- [Hellerstein et al. 2004] Hellerstein, J. L., Diao, Y., Parekh, S., and Tilbury, D. M. (2004). *Feedback Control of Computing Systems*. John Wiley & Sons.
- [Huebscher and McCann 2008] Huebscher, M. C. and McCann, J. A. (2008). A survey of autonomic computing - degrees, models, and applications. *ACM Comput. Surv.*, 40(3).
- [Jain 1991] Jain, R. (1991). The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling. *SIGMETRICS Performance Evaluation Review*, 19:5–11.
- [Kephart and Chess 2003] Kephart, J. O. and Chess, D. M. (2003). The vision of autonomic computing. *IEEE Computer*, 36(1):41–50.
- [Northrop 2013] Northrop, L. M. (2013). Does scale really matter? ultra-large-scale systems seven years after the study (keynote). In Notkin, D., Cheng, B. H. C., and Pohl, K., editors, *35th International Conference on Software Engineering, ICSE '13, San Francisco, CA, USA, May 18-26, 2013*, page 857. IEEE / ACM.
- [Patikirikorala et al. 2012] Patikirikorala, T., Colman, A., Han, J., and Wang, L. (2012). A systematic survey on the design of self-adaptive software systems using control engineering approaches. In *Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*, pages 33–42.
- [Salehie and Tahvildari 2009] Salehie, M. and Tahvildari, L. (2009). Self-adaptive software: Landscape and research challenges. *TAAS*, 4(2).
- [Weyns 2010] Weyns, D. (2010). On patterns for decentralized control in self-adaptive systems. In de Lemos, R., Giese, H., Müller, H. A., and Shaw, M., editors, *Software Engineering for Self-Adaptive Systems II*, volume 7475 of *Lecture Notes in Computer Science*, pages 76–107. Springer.

Projetando a Infraestrutura de Comunicação em Redes Veiculares para a Distribuição de Conteúdos Demandando Níveis Distintos de Desempenho

Cristiano M. Silva¹, Cristiano G. Pitangui¹, Alex V. Bastos¹, Rone I. Silva¹,
Fabrício A. Silva²

¹Universidade Federal de São João Del-Rei (UFSJ) – Minas Gerais – Brasil

²Universidade Federal de Viçosa (UFV) – Minas Gerais – Brasil

{cristiano, pitangui.cristiano, alexvbh, rone}@ufs.j.edu.br

fabricio.asilva@ufv.br

Abstract. *We investigate the development of Mobile Content Delivery Vehicular Networks. We consider the availability of several content types that must be made available for vehicles at specific locations of the road network. We also assume each content type demanding distinct levels of performance from the network, accordingly to the location where it must be made available. We devise a novel metric for evaluating the performance of vehicular networks, and we use the metric for devising a novel infrastructure deployment strategy. We compare our deployment strategy to locating roadside units at the densest locations of the road network. Results demonstrate our strategy requiring much less roadside units than the baseline to achieve similar levels of performance.*

Resumo. *Este trabalho investiga a distribuição de conteúdos em Redes Veiculares. Considera-se a disponibilidade de vários tipos de conteúdos que devem ser disponibilizados para os veículos em locais específicos da malha viária. Assume-se que cada tipo de conteúdo exige diferentes níveis de desempenho da rede. O trabalho propõe uma nova métrica para avaliar o desempenho de redes veiculares, e utiliza essa métrica para elaborar uma nova estratégia de implantação de infraestrutura para redes veiculares. A estratégia proposta é comparada com o método intuitivo de implantar as unidades de comunicação nos locais de tráfego mais intenso. Os resultados demonstram que a estratégia proposta exige menos unidades de comunicação para atingir a mesma qualidade de serviço.*

1. Introdução

O rápido crescimento dos centros urbanos aumenta a demanda de transportes para a população, tornando o tráfego num problema central da sociedade moderna. A impossibilidade de alterar constantemente a malha viária motiva os governos a buscarem soluções para maximizar a capacidade das vias já existentes através do desenvolvimento de Sistemas Inteligentes de Transportes (SITs) que se baseiam na comunicação veicular para o gerenciamento da mobilidade urbana. A grande mobilidade dos veículos impõe grandes desafios de comunicação, visto que o canal deve lidar com vários nós (veículos) enviando e recebendo dados simultaneamente. Além disso, diferentemente das redes

celulares, veículos não exigem cobertura contínua, podendo realizar a troca de dados ao passarem por unidades de comunicação¹.

Percebe-se que as aplicações veiculares têm evoluído de simples alertas de tráfego a transmissões em tempo real de altíssima qualidade, demandando novas estratégias de gerenciamento das redes de comunicação. Ainda que os veículos possam se comunicar de maneira *ad-hoc*, o uso de unidades de comunicação alocadas ao longo das vias pode melhorar significativamente o desempenho da rede. Por outro lado, a implantação em larga escala de uma infraestrutura de comunicação pode demandar grandes investimentos [Cheng et al., 2015].

Este trabalho investiga uma rede veicular de distribuição de conteúdos. Já que algumas informações de tráfego são importantes apenas para uma determinada região, assume-se que cada tipo de conteúdo está relacionado a uma região específica da cidade. Além disso, dada a ampla gama de aplicações veiculares vislumbradas, também se considera que cada conteúdo exige diferentes níveis de desempenho. Inicialmente, deriva-se uma métrica para medir a qualidade de serviço oferecida pela infraestrutura suportando a comunicação veicular chamada de *Deposição*² *Sigma* (Σ). A *Deposição Sigma* permite que o projetista verifique se uma rede veicular fornece suporte adequado para a entrega de vários tipos de conteúdo, cada um deles exigindo níveis específicos de desempenho da rede, de acordo com a sua localização na cidade.

Em seguida, o trabalho propõe *Sigma-g*, uma heurística gulosa para resolver a alocação de infraestrutura de forma a se atingir uma *Deposição Sigma*. *Sigma-g* é avaliada usando-se o rastro (*trace*) de mobilidade realístico da cidade de Colônia³, Alemanha, composto de 75.515 veículos. Os resultados são comparados com a estratégia intuitiva de se alocar as unidades de comunicação nos locais de maior tráfego, chamada DL (*Densest Location*). Os resultados demonstram que a heurística proposta reduz de 20%-40% a quantidade de unidades de comunicação preservando a mesma qualidade de serviço.

Diferentemente das abordagens anteriores, a *Deposição Sigma* é uma métrica orientada a conteúdos para avaliar o desempenho de uma dada deposição de infraestrutura suportando redes veiculares. Este trabalho encontra-se organizado da seguinte forma. A seção 2 apresenta os trabalhos relacionados. A seção 3 apresenta a *Deposição Sigma*. A seção 4 apresenta uma heurística gulosa para solucionar uma *Deposição Sigma*. A seção 5 apresenta os experimentos. A seção 6 conclui o trabalho.

2. Trabalhos Relacionados

Alguns estudos abordam o problema de distribuição de conteúdo no contexto de redes veiculares [Silva et al., 2014]. Em [Gossa et al., 2008], os autores propõem uma solução de replicação de dados proativa para selecionar veículos como réplicas. Por outro lado, a solução descrita no [Li et al., 2015], replica o conteúdo em unidades de comunicação com base na análise do rastro (*trace*) de mobilidade real. Da forma

¹Esse trabalho utiliza a expressão "unidades de comunicação" para se referir ao termo "roadside units", cujo significado literal é "unidades de (comunicação) de beira de estrada".

²O termo "deposição" possui a conotação de implantação ou colocação (de infraestrutura de comunicação para redes veiculares). Assim, uma estratégia de deposição é aquela que identifica onde as unidades de comunicação devem ser implantadas na rede veicular.

³Disponível em: <http://kolntrace.project.citi-lab.fr/>

semelhante, o trabalho [Bruno et al., 2015] investiga um modelo de otimização para replicar os conteúdos nas unidades de comunicação, considerando sua popularidade. O presente trabalho difere das abordagens anteriores no sentido de que os trabalhos anteriores consideram cobertura total da malha viária.

A literatura apresenta várias estratégias para a alocação de unidades de comunicação em redes veiculares. Zheng et al. [2010] apresentam a avaliação da estratégia de implantação considerando a oportunidade de contato. Lee and Kim [2010] propõem uma heurística gulosa para alocar a infraestrutura com o objetivo de melhorar a conectividade entre veículos. Xie et al. [2013] abordam o problema de alocação de unidades de comunicação baseando-se em um modelo probabilístico para inferir as localizações das unidades de comunicação.

Trullols et al. [2010] propõem a modelagem de alocação de unidades de comunicação como um problema de cobertura máxima, assumindo-se conhecimento prévio das trajetórias dos veículos. Essa premissa é relaxada nos trabalhos [Silva et al., 2013, 2015b] através do uso da taxa de migração de veículos entre regiões da área urbana.

De forma geral, o presente trabalho se diferencia da literatura ao propor uma estratégia de implantação que fornece garantia de desempenho mínima especificada em termos da probabilidade de contatos e duração do mesmos, para vários tipos de conteúdo, e de acordo com a localização do mesmo.

3. Métrica *Deposição Sigma*

O planejamento de redes veiculares de distribuição de conteúdos requerem o planejamento da localização física das unidades de comunicação onde se espera que os potenciais clientes passem. Além disso, diferentes tipos de conteúdo podem exigir níveis distintos de desempenho da rede. Assim, este trabalho propõe a *Deposição Sigma* (Σ) como uma estratégia para o planejamento da localização das unidades de comunicação em redes veiculares. Essa estratégia de implantação permite diferentes níveis de disponibilidade de conteúdo ao longo da malha viária.

A principal característica da *Deposição Sigma* é o fato de considerar as unidades de comunicação armazenando conteúdos diferentes, sendo que cada conteúdo exige níveis distintos de qualidade de serviço especificados através das oportunidades de contatos entre veículos e a infraestrutura, e a duração desses contatos. Uma dada deposição de unidades de comunicação é considerada $\Sigma \binom{R_c}{c} \binom{\rho_1}{\rho_2}$ sempre que ρ_2 por cento dos veículos que viajam por R_c estão conectados a unidades de comunicação tendo o conteúdo c durante (pelo menos) ρ_1 por cento da duração da viagem ao longo de R_c , onde R_c é uma região contida na malha rodoviária R . Os parâmetros ρ_1 e ρ_2 indicam garantias de desempenho para veículos encontrando o conteúdo c nas regiões R_c da malha rodoviária. Formalmente:

Definição 1 (Deposição: $\Sigma \binom{R_c}{c} \binom{\rho_1}{\rho_2}$) Assuma que $R_c \subseteq R$ represente um subconjunto da malha rodoviária e que V represente o conjunto de veículos que viaja em R_c . Considere que $V' \subset V$ represente o conjunto de veículos conectados às unidades de comunicação possuindo o conteúdo c durante, ao menos, ρ_1 por cento da duração da viagem. Uma implantação de unidades de comunicação é considerada $\Sigma \binom{R_c}{c} \binom{\rho_1}{\rho_2}$ quando $\frac{|V'|}{|V|} \geq \rho_2$.

Nas redes veiculares de distribuição de conteúdos, as unidades de comunicação atuam como servidores que fornecem conteúdo para os veículos. O modelo assume que

existem diferentes tipos de conteúdos ao longo da rede veicular (tais como informações de trânsito, vídeos, anúncios, etc.). Cada tipo de conteúdo deve estar fisicamente localizado em um subconjunto de células urbanas (regiões da malha rodoviária). Este é um requisito muito importante, pois alguns conteúdos podem ter significado apenas local (por exemplo, informações de trânsito e propagandas não necessariamente precisam ser transmitidos a todos os veículos). Além disso, cada tipo de conteúdo exige níveis distintos de disponibilidade representada em termos da *Deposição Sigma*, e tais níveis podem variar de acordo com a região (i.e., um subconjunto de células urbanas).

Para resolver a *Deposição Sigma*, assume-se o particionamento da área urbana em um conjunto de $\psi \times \psi$ células urbanas de mesmo tamanho. As células urbanas podem ter tamanhos arbitrários, a fim de atender às necessidades de projeto da rede. Quando se exige maior/menor precisão, simplesmente aumenta-se/diminui-se o número de células urbanas que cobrem a malha rodoviária. Particionando-se a malha rodoviária, diminui-se o número de locais a serem avaliados para receberem unidades de comunicação. Tal particionamento acarreta, portanto, uma redução do esforço computacional para a obtenção das melhores localidades.

4. *Sigma-g*: Uma Heurística para a Deposição de Unidades de Comunicação

Nesta seção, define-se *Sigma-g*, uma heurística gulosa para resolver o problema de *Deposição Sigma*. A heurística considera a malha rodoviária dividida em $\psi \times \psi$ células urbanas. Sem perda de generalidade, este trabalho assume que cada célula urbana é coberta pela implantação de uma única unidade *lógica* de comunicação. Assim, considera-se que uma unidade *lógica* é capaz de cobrir completamente uma célula urbana.

Sigma-g recebe como entrada: (a) o subconjunto de células urbanas onde a implantação deve ocorrer (M); (b) o conjunto de veículos viajando por M (V); (c) a trajetória de cada veículo (T); e, (d) os parâmetros de desempenho ρ_1 e ρ_2 , onde ρ_1 indica a duração de contato V2I, enquanto o ρ_2 indica a fração de veículos.

Nota-se que o algoritmo 1 não recebe o tipo de conteúdo como entrada. Assim, deve-se executar *Sigma-g* para cada tupla $\langle \text{tipo de conteúdo}, \text{localização do conteúdo} \rangle$. A localização do conteúdo é dada por M , enquanto que o desempenho requerido para o conteúdo é dado por (ρ_1, ρ_2) . O conjunto final de unidades de comunicação é simplesmente a união dos conjuntos de unidades retornados por *Sigma-g* para cada tupla $\langle \text{tipo de conteúdo}, \text{localização do conteúdo}, \rho_1, \rho_2 \rangle$.

Algoritmo 1 *Sigma-g*

Entrada: M, V, T, ρ_1, ρ_2 ;

Saída: Γ

1: $\Gamma \leftarrow \emptyset$

2: $M' \leftarrow M$

3: **Enquanto** $\frac{|C|}{|V|} < \rho_2$ **Faça**

4: $\varphi \leftarrow \text{TEMPOMAXABS}(M', V-C)$;

5: $\Gamma \leftarrow \Gamma \cup \varphi$;

6: $M' \leftarrow M' - \varphi$;

7: $C \leftarrow \text{CONTATO}(M, V, T, \Gamma, \rho_1)$;

8: **Fim Enquanto**

9: **Retorne** Γ ;

▷ Localizações recebendo unidades de comunicação.

▷ Conjunto solução.

▷ Salva uma cópia da malha rodoviária.

▷ Laço até alcançar a cobertura desejada.

▷ Selecione uma célula urbana.

▷ Adiciona a célula urbana selecionada ao conjunto solução.

▷ Remove a célula urbana selecionada.

▷ Confere os veículos que realizam contatos ρ_1 .

Sigma-g seleciona células urbanas do subconjunto (M) e as coloca no conjunto solução (Γ). Cada M_{ij} indica o número de veículos que atravessam a célula urbana (i, j).

A escolha gulosa baseia-se no tempo de contato absoluto que é fornecido por cada célula urbana ao receber uma unidade de comunicação. *Sigma-g* calcula um índice para cada célula urbana. Este valor é a soma do tempo de contato V2I total obtido pela alocação de uma unidade na célula. Para calcular este valor, são considerados todos os veículos que satisfizeram o critério ρ_1 (duração da conexão V2I).

A primeira unidade de comunicação é alocada na célula urbana que possui maior tráfego. A célula urbana selecionada é removida do conjunto de candidatos (M'), e o algoritmo armazena em C os veículos que realizam conexão durante ρ_1 por cento da duração da viagem. Tal computação é realizada pela função *Contato()*. A função *Contato()* usa o subconjunto original de células urbanas (M), o conjunto de veículos (V), a informação de trajetórias (T), as células urbanas recebendo unidades (Γ), e ρ_1 . Enquanto a fração de ρ_2 veículos cobertos não for alcançada, a heurística seleciona de forma iterativa a célula urbana que apresenta a maior contribuição em termos de tempo de contato absoluto para os veículos (ainda) descobertos, usando a função *TempoMaxAbs()*.

Sigma-g possui uma propriedade importante: ela permite que se desconsidere o tipo de conteúdo. O conjunto solução fornecido pela *Sigma-g* é incremental. Assuma duas deposições: Σ' e Σ'' . Considere que a implantação Σ' apresente garantias de desempenho inferiores à implantação Σ'' . Assim, o conjunto de solução de Σ' demanda menos unidades de comunicação que o conjunto solução de Σ'' . No entanto, *Sigma-g* garante que a solução Σ' é um subconjunto de Σ'' . Assim, $\Gamma' \subset \Gamma''$.

5. Experimentos

Este trabalho compara *Sigma-g* com DL, uma estratégia intuitiva que aloca as unidades de comunicação nos locais de maior tráfego da malha viária. DL funciona da seguinte forma: enquanto não se atinge a fração de ρ_2 veículos cobertos durante ρ_1 por cento da viagem, DL permanece alocando unidades de comunicação, iniciando a partir do local de maior tráfego de veículos.

Os experimentos consideram o rastro (*trace*) realístico de mobilidade da cidade de Colônia (Alemanha), que é particionada em uma grade (*grid*) de 100×100 células urbanas a fim de se obter células capazes de serem cobertas por apenas uma unidade de comunicação. Tal particionamento gera células urbanas de 270m x 260m, um raio típico de cobertura para a comunicação veicular [Teixeira et al., 2014].

Ao se analisar a qualidade de uma deposição de unidades de comunicação, duas medidas requerem maior atenção: a) investimentos; b) cobertura. Os investimentos são estimados através da análise do número de unidades *lógicas* requeridas por cada estratégia de deposição, enquanto a cobertura é analisada utilizando-se a *Deposição Sigma*.

5.1. Análise de Investimentos: Número de Unidades de Comunicação

Para estimar os custos financeiros para a criação da rede, a figura 1(a) apresenta a área necessária a ser coberta para que cada estratégia de deposição atinja vários níveis de desempenho especificados por $\{\rho_1, \rho_2\}$. O eixo x indica ρ_2 (fração de veículos), enquanto que o eixo y indica a quantidade de unidades de comunicação que são implantadas por cada estratégia de forma a se obter as Deposições Sigma indicadas no grafico. São plotados cinco conjuntos de dados ρ_1 ($\{0.1, 0.3, 0.5, 0.7, 0.9\}$).

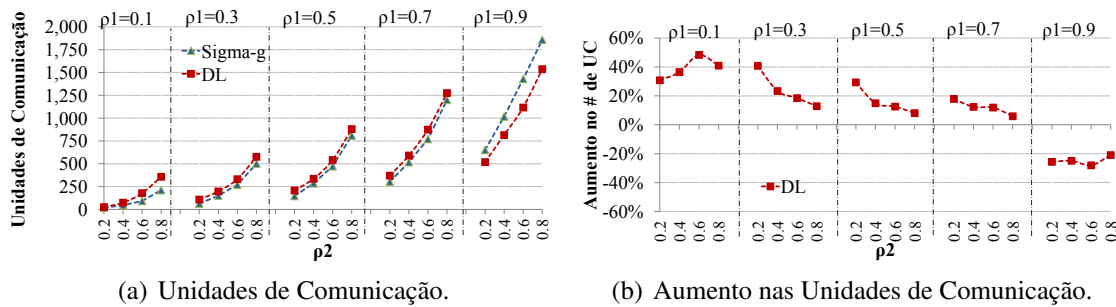


Figura 1. A fig. (a) apresenta a quantidade de unidades de comunicação demandadas por cada estratégia. A fig. (b) apresenta a variação no número de unidades de comunicação demandadas por DL em termos de Sigma-g.

A quantidade de unidades de comunicação demandadas por Sigma-g é indicada na cor azul, enquanto que DL é indicado na cor vermelha. A figura demonstra que Sigma-g demanda menos unidades que DL para todas as configurações, exceto quando $\rho_1=0,9$. No entanto, por razões de escala, o gráfico não apresenta com clareza quantas unidades de comunicação Sigma-g economiza em relação à DL.

Com esse intuito, apresenta-se a figura 1(b). Essa figura apresenta a proporção de unidades de comunicação que DL demanda a mais que Sigma-g. Essa proporção varia conforme os valores $\{\rho_1, \rho_2\}$. Para $\rho_1=0,1$, DL demanda de 25% até 50% mais unidades de comunicação que Sigma-g. Conforme aumenta-se ρ_1 (percentual da viagem que veículos devem ficar em contato com unidades de comunicação), essa relação diminui progressivamente. Para $\rho_1=0,9$ DL utiliza menos unidades de comunicação que Sigma-g. Tal nível de desempenho exige uma implantação massiva de unidades de comunicação e a estratégia de cobrir os locais de maior tráfego se mostra bastante útil.

De forma geral, DL apresenta uma curva decrescente para cada conjunto de dados ρ_1 . Aumentando-se ρ_2 , aumenta-se o número de veículos que devem satisfazer o critério de duração de conectividade (ρ_1), o que exige um aumento no número de células urbanas cobertas. À medida que se cobre um maior número de células urbanas, diminuem-se as possíveis combinações de disposições das unidades de comunicação, tornando o problema mais fácil de ser resolvido.

Em última análise, ao se especificar que 100% dos veículos realizem contatos durante 100% do tempo de duração da viagem, i.e., uma Deposição $\Sigma \binom{R_c}{c} \binom{1,0}{1,0}$, cada célula urbana que apresente um nível de tráfego maior que zero irá receber uma unidade de comunicação, e ambas as estratégias irão implantar 4.219 unidades lógicas para cobrir todas as células urbanas que apresentam tráfego⁴, e não haverá diferença entre as estratégias.

5.2. Relacionando Investimentos e Cobertura

Visto que as estratégias Sigma-g e DL incorrem na implantação de quantidades diferentes de unidades de comunicação, esse experimento caracteriza como a variação no número de unidades de comunicação impacta a cobertura atingida por cada estratégia. Na figura 2(a), a linha azul representa o aumento percentual no tempo de contato (V2I) alcançado por

⁴Os traços de Colônia possuem 4.219 células urbanas com tráfego.

DL quando comparado a *Sigma-g*. Quando a coordenada y é maior do que zero, DL proporciona mais tempo de contato que *Sigma-g*. Pode-se notar que DL supera em 11% o tempo de contato fornecido por *Sigma-g* para o $\Sigma_c^{(R_c)}(0.1)$ e em 32% para $\Sigma_c^{(R_c)}(0.3)$.

No entanto, deve-se lembrar que DL requer mais unidades de comunicação que *Sigma-g*. Assim, para uma comparação justa, indica-se também o aumento no número de unidades de comunicação para DL em relação a *Sigma-g* (linha vermelha). Enquanto DL aumenta em 11% o tempo de contato de $\Sigma_c^{(R_c)}(0.1)$, ela exige 38% mais unidades de comunicação que *Sigma-g*. De forma similar, quando se considera a Deposição $\Sigma_c^{(R_c)}(0.9)$, DL exige +18% mais unidades de comunicação que *Sigma-g*, mas reduz em 3% o tempo de contato. De forma geral, *Sigma-g* apresenta um melhor custo-benefício entre investimentos e serviço ofertado.

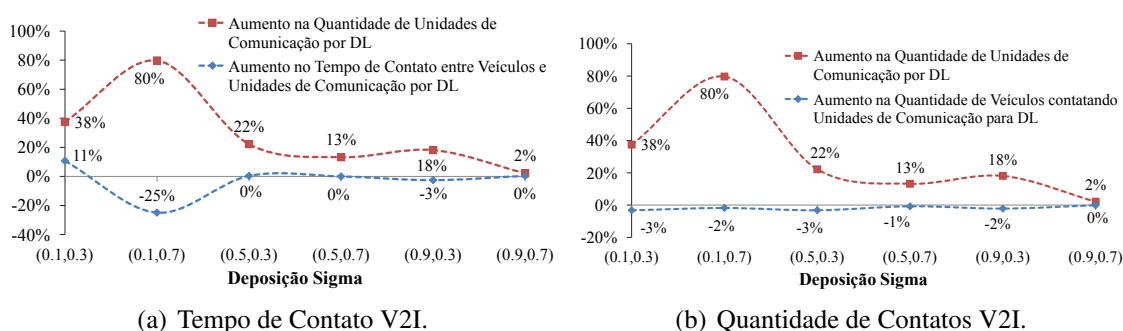


Figura 2. Distribuição de tempo "entre-contatos" para algumas Deposições Sigma. A fig. (a) apresenta a variação percentual no tempo de contato V2I de DL em relação à Sigma-g. A fig. (b) apresenta a variação na quantidade de contatos de DL em relação à Sigma-g.

Na figura 2(b), a linha azul indica o aumento no número de contatos V2I, em termos percentuais, de DL comparado com Sigma-g. Nota-se um desempenho similar entre as estratégias, com *Sigma-g* proporcionando um pouco mais de contatos V2I. Mais uma vez, a linha vermelha indica o aumento percentual do número de unidades de comunicação requerido por DL em relação a *Sigma-g*. É fácil perceber que o aumento no número de unidades de comunicação demandado por DL não retorna em termos de mais veículos contatando unidades de comunicação.

5.3. Intervalo "Entre-Contatos" do Veículo com Unidades de Comunicação

O intervalo "entre-contatos" mede o tempo entre duas conexões consecutivas dos veículos a unidades de comunicação. As figuras 3(a)-(b) apresentam o intervalo entre-contatos para *Sigma-g* e DL considerando diferentes parâmetros Sigma. Para ser justo, calcula-se o tempo entre-contatos em relação à duração da viagem para cada veículo (percentual do tempo de duração da viagem), e a figura apresenta a fração de veículos com um determinado intervalo de tempo entre-contatos (indicado na legenda).

Para melhor entendimento da figura, observe a primeira barra. Ela representa a distribuição dos tempos entre-contatos V2I considerando a Deposição $\Sigma_c^{(R_c)}(0.9)$. As cores representam intervalos entre-contatos. A cor verde indica um tempo entre-contatos muito baixo (0%-10% da duração da viagem). Azul indica um tempo entre-contatos médio (20%-30% da duração da viagem). Amarelo indica alto tempo entre-contatos

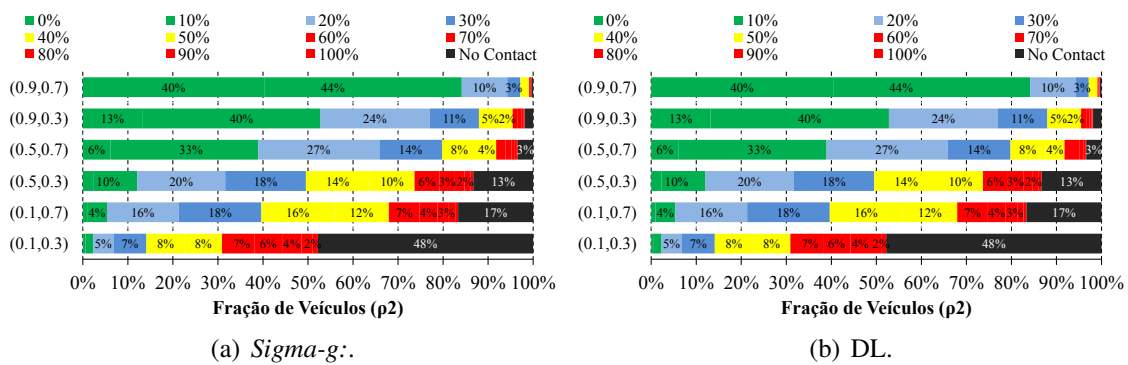


Figura 3. Distribuição de tempo "entre-contatos" para algumas deposições Sigma. A figura (a) apresenta Sigma-g. A figura (b) apresenta DL.

(40%-50% da duração da viagem). Vermelho indica tempo entre-contatos muito elevado (> 70% da duração da viagem), enquanto que a cor preta indica a fração de veículos que não realizam contatos com unidades de comunicação durante sua viagem. À medida que se aumentam os níveis de desempenho da implantação, aumenta-se a fração de veículos que apresentam baixos tempos entre-contatos.

5.4. Análise de Cobertura

Sempre que se resolve a Deposição $\Sigma_c^{(R_c)}(\rho_1, \rho_2)$ utilizando alguma estratégia, tem-se, como saída, o número de unidades de comunicação necessário (e suas localizações) na malha rodoviária. Essa disposição de unidades de comunicação permite que ρ_2 por cento dos veículos realizem contatos V2I durante ρ_1 por cento da duração da viagem. Ao se implantar as unidades de comunicação e realizar o rastro (*trace*) de mobilidade, pode-se traçar a relação $\rho_1 \times \rho_2$ (chamada de Rede de Comunicação *Delta* [Silva et al., 2015a]). Ao se traçar a Rede Delta, caracteriza-se toda a gama de operação da rede, permitindo que o projetista possa medir e comparar o desempenho das redes veiculares.

A heurística *Sigma-g* é apenas uma estratégia possível para resolver a Deposição $\Sigma_c^{(R_c)}(\rho_1, \rho_2)$. Já que estratégias distintas tendem a selecionar células urbanas distintas para receber as unidades de comunicação, pode-se inferir que existem diferenças de desempenho da rede como um todo. Por exemplo, uma dada solução para a Deposição *Sigma* pode permitir mais oportunidades de contato para veículos que viajam distâncias menores, enquanto outras podem beneficiar veículos em viagens mais longas.

Como exemplo, a figura 4(a) plota a Rede Delta para várias Deposições (na legenda). O eixo x indica a duração do contato (ρ_1), enquanto o eixo y indica a porcentagem de veículos (ρ_2). Cada ponto indica a porcentagem de veículos (eixo y) tendo contatos V2I durante (pelo menos) ρ_1 por cento da duração da viagem (eixo x). Pode-se notar que à medida que aumentam-se os parâmetros de desempenho (ρ_1 e ρ_2), também aumenta-se o desempenho da rede (cada *Sigma* representa uma rede) ao longo de toda a curva, já que utiliza-se uma maior quantidade de unidades de comunicação que fornecem suporte à rede veicular.

Na figura 4(b) usa-se a Rede *Delta* para comparar *Sigma-g* e DL. No entanto, em vez de plotar uma curva para *Sigma-g*, e outra para a DL, plota-se apenas uma curva que representa as diferenças entre ambas as estratégias (a Rede *Delta* Diferencial). Para cada

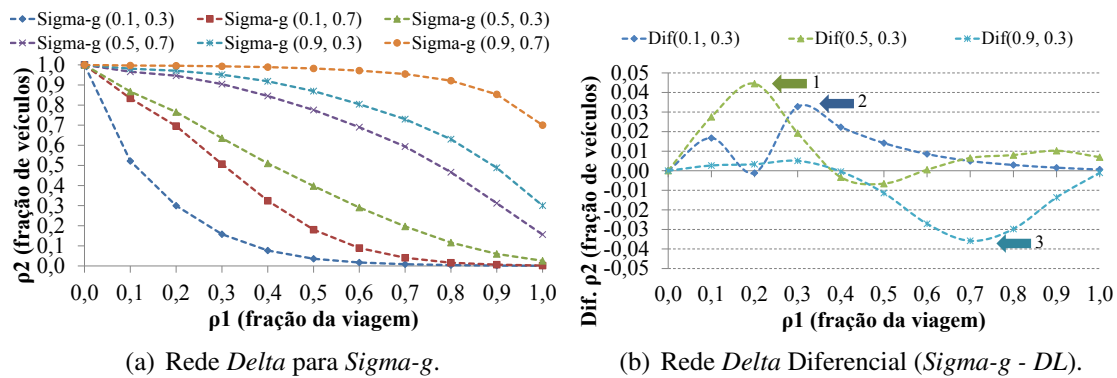


Figura 4. A fig. (a) apresenta a relação entre ρ_1 e ρ_2 ao se resolver a Deposição *Sigma* usando-se *Sigma-g*. A fig. (b) apresenta ρ_2 de *Sigma-g* menos o ρ_2 de DL.

ρ_1 , marca-se o ρ_2 de *Sigma-g* menos o ρ_2 de DL. Essa figura traça as diferenças em ρ_2 para cada ρ_1 (*Sigma-g* menos DL). Sempre que a coordenada y é positiva, *Sigma-g* tem melhor desempenho em termos de número de veículos que realizam contato durante o valor dado no eixo x . Quando o segmento de curva tem uma coordenada y negativa, DL tem mais veículos que realizam conexões.

A linha azul compara *Sigma-g* e DL para uma Deposição $\Sigma \binom{R_c}{c} \binom{0.1}{0.3}$. *Sigma-g* supera DL quando a coordenada x é maior do que 0,2. A seta azul (#2) indica que *Sigma-g* apresenta 3% mais veículos que realizam contato durante 30% do tempo de duração da viagem. Da mesma forma, a linha verde indica uma Deposição $\Sigma \binom{R_c}{c} \binom{0.5}{0.3}$, e a seta verde (#1) indica que *Sigma-g* apresenta 4% mais veículos que realizam contato durante 20% do tempo de duração da viagem. A curva ciano indica a Deposição $\Sigma \binom{R_c}{c} \binom{0.9}{0.3}$, e que *Sigma-g* apresenta -4% veículos realizando conexões durante 70% da duração da viagem⁵.

6. Conclusão

Este trabalho apresenta uma investigação sobre o planejamento de redes veiculares para a distribuição de conteúdos. Propõe-se a Deposição *Sigma* como métrica para avaliar se uma determinada distribuição de unidades de comunicação é capaz de suportar a distribuição de conteúdos. Considera-se que a rede deve gerenciar a distribuição de diversos conteúdos que devem ser oferecidos para os veículos em locais específicos da malha viária sob qualidades de serviço distintas. Para resolver a Deposição *Sigma*, propõe-se uma heurística gulosa *Sigma-g*. Comparou-se a eficiência de *Sigma-g* com a estratégia intuitiva de implantar unidades de comunicação nos locais de maior tráfego da malha rodoviária. Os resultados demonstram que *Sigma-g* requer menos unidades de comunicação que DL para um mesmo patamar de qualidade de serviço ofertada.

Agradecimentos

Esse trabalho foi parcialmente financiado por CNPq, CAPES e FAPEMIG.

⁵Ressalta-se que esses resultados aplicam-se apenas ao rastro de mobilidade em análise.

Referências

- Bruno, F., Cesana, M., Gerla, M., Mauri, G., and Verticale, G. (2015). Optimal content placement in icn vehicular networks. In *5th IEEE International Conference on Network of the Future*.
- Cheng, H., Fei, X., Boukerche, A., and Almulla, M. (2015). Geocover: An efficient sparse coverage protocol for rsu deployment over urban vanets. *Ad Hoc Networks*, 24, Part B(0):85 – 102. Modeling and Performance Evaluation of Wireless Ad-Hoc Networks.
- Gossa, J., Janecek, A. G., Hummel, K., Gansterer, W. N., Pierson, J.-M., et al. (2008). Proactive replica placement using mobility prediction. In *Mobile Data Management Workshops, 2008. MDMW 2008. Ninth International Conference on*, pages 182–189. IEEE.
- Lee, J. and Kim, C. (2010). A roadside unit placement scheme for vehicular telematics networks. In Kim, T.-h. and Adeli, H., editors, *Advances in Computer Science and Information Technology*, volume 6059 of *Lecture Notes in Computer Science*, pages 196–202. Springer Berlin Heidelberg.
- Li, Y., Jin, D., Hui, P., and Chen, S. (2015). Contact-aware data replication in roadside unit aided vehicular delay tolerant networks. *Mobile Computing, IEEE Transactions on*, PP(99):1–1.
- Silva, C. M., Andre L. L. Aquino, and Wagner Meira Jr (2015a). "avaliando o desempenho de redes veiculares heterogêneas". In *SBCUP 2015 (Simpósio Brasileiro de Computação Ubíqua)*.
- Silva, C. M., Sergio Oliveira, Andre L. L. Aquino, and Fernando Augusto Teixeira (2013). PMCP: Uma Heurística Probabilística para Otimizar a Instalação de Pontos de Disseminação em Redes Veiculares. In *SBCUP 2013 (Simpósio Brasileiro de Computação Ubíqua)*.
- Silva, C. M., Wagner Meira Jr, and Joao F. M. Sarubbi (2015b). Non-Intrusive Planning of Roadside Infrastructure for Vehicular Networks Without Tracking Individual Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, -(–):–.
- Silva, F., Boukerche, A., Silva, T. R. M., Ruiz, L. B., Cerqueira, E., and Loureiro, A. A. (2014). Content replication and delivery in vehicular networks. In *ACM International Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications, DIVANet '14*, pages 127–132, New York, NY, USA. ACM.
- Teixeira, F., Silva, V., Leoni, J., Macedo, D., and Nogueira, J. M. S. (2014). "vehicular networks using the ieee 802.11p standard: An experimental analysis". *Vehicular Communications*, 1(2):91 – 96.
- Trullols, O., Fiore, M., Casetti, C., Chiasserini, C., and Ordinas, J. B. (2010). Planning roadside infrastructure for information dissemination in intelligent transportation systems. *Computer Communications*, 33(4):432 – 442.
- Xie, B., Xia, G., Chen, Y., and Xu, M. (2013). Roadside infrastructure placement for information dissemination in urban its based on a probabilistic model. In *Network and Parallel Computing*, volume 8147 of *Lecture Notes in Computer Science*, pages 322–331. Springer Berlin Heidelberg.
- Zheng, Z., Lu, Z., Sinha, P., and Kumar, S. (2010). Maximizing the contact opportunity for vehicular internet access. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9.

Rede de Sensores para Controle Inteligente de Ambientes

Artejose Revoredo da Silva¹, João C. Xavier-Júnior¹,
Ivanovitch Medeiros Dantas da Silva¹

¹Instituto de Metrópole Digital – Universidade Federal do Rio Grande do Norte – (UFRN)
Av. Senador Salgado Filho, 3000 – 59.078-970 – Natal – RN

{ivan,jcxavier}@imd.ufrn.br, artejose.rs@gmail.com

Abstract. *In general, laboratories or labs are workplaces that provide controlled conditions for experiments and measurements to be performed. For this reason, controlling temperature and humidity is an important requirement that needs to be achieved in order to guarantee the reproducibility of processes carried out in labs. Aiming to introduce efficient environmental controlling mechanisms, we present in this paper an intelligent environmental control system based on sensors network and pattern classification. Our system uses the information generated by the sensors to make decisions for constant controlling of temperature and relative humidity within the lab.*

Resumo. *Laboratórios são ambientes que necessitam ser controlados, principalmente devido ao fato de que nesses ambientes são realizados processos químicos. Controlar a temperatura e a umidade relativa do ar é tarefa primordial para que os processos desenvolvidos dentro de tais laboratórios possam ser repetidos dentro dos parâmetros exigidos. Buscando propor mecanismos de controle para ambientes, neste artigo, apresentamos um sistema baseado em rede de sensores e classificação de padrões para controle inteligente de ambientes. Nosso protótipo utiliza as informações gerados pelos sensores para tomar decisões relativas ao controle constante da temperatura e da umidade relativa do ar.*

1. Introdução

A evolução exponencial das tecnologias tem provocado o surgimento de soluções que não eram pensadas apenas há alguns anos atrás. Entre as áreas emergentes destaca-se a Internet das Coisas. Em um futuro muito breve a maioria dos dispositivos que nos cercam estarão conectados na Internet através de uma grande rede de sensores. Esse cenário irá provocar o surgimento de uma quantidade colossal de dados. Esses por sua vez, precisarão ser processados para gerar informação útil. Nesse contexto, surge uma nova e promissora área de pesquisa, a Ciência dos Dados. Essa nova frente de pesquisa irá impactar diretamente a indústria e especificamente, a automação. Técnicas de classificação de dados e tomada de decisões podem ser empregadas como uma ferramenta da Ciência dos Dados para gerar sistemas automatizados de controle e monitoramento de ambientes.

Baseado no cenário anterior, o respectivo trabalho busca desenvolver um sistema inteligente para automatizar o controle e monitoramento térmico de labo-

ratórios de análises químicas a partir de uma rede de sensores e algoritmos de classificação de dados. Em laboratórios químicos, controlar as variações de temperatura e umidade resulta diretamente na melhoria da qualidade das amostras trabalhadas, assim como a economia da própria energia elétrica utilizada para abastecer aparelhos de ar-condicionados e desumificadores. As ferramentas da Ciência dos Dados serão utilizadas para auxiliar a tomada de decisão no controle de tais equipamentos.

O restante deste artigo está dividido em 5 seções. A Seção II descreve alguns conceitos teóricos importantes e alguns estudos relacionados ao tema deste artigo. A seção III discute os conceitos relacionados. Na Seção IV, a metodologia utilizada no trabalho experimental do presente trabalho é apresentada, enquanto que uma análise dos resultados fornecidos pela análise empírica é mostrada na Seção V. Finalmente, a Seção VI apresenta as considerações finais deste trabalho.

2. Trabalhos Relacionados

Os sistemas de controle e monitoramento de ambientes inteligentes apresentam diversos desafios, englobando aspectos de desempenho (rede, latência, confiabilidade), segurança e de tomada de decisão. O referido trabalho está focado nos desafios de tomada de decisão a partir de uma rede de sensores para o controle térmico de laboratórios químicos. Na literatura, as principais contribuições utilizam aspectos de automação *on/off*, ligando ou desligando luzes e ar-condicionados. Aspectos de tomada de decisão embarcadas em redes de sensores são pouco explorados.

No trabalho apresentado em [da Silva 2013], uma solução de automação residencial baseado na plataforma Arduino foi implementada. Controle de iluminação, abertura e fechamento de portas, alarmes e câmeras foram utilizadas para validação da proposta. Em outro trabalho [Maycon J. C. Mesquita 2014], um procedimento de controle manual de equipamentos residenciais através de um sistema web foi implementado. No entanto, em ambos os trabalhos citados não houve a exploração dos aspectos de tomada de decisão em conjunto com rede de sensores para uma automação mais eficiente.

Um sistema para monitoramento de concreto baseado em redes de sensores de temperatura e umidade foi apresentado em [Barroca et al. 2013]. Na mesma linha de pesquisa, foi desenvolvido em [Abraham and Li 2014] uma rede de sensores para o controle da qualidade do ar em ambiente fechados. Ambos não descreveram técnicas de controle inteligente (classificação de padrões) na tomada de decisões.

Outros trabalhos que, de forma geral utilizaram sensores de temperatura e que implementaram algum tipo de inteligência, focaram em aplicações diversas. [Vera-Repullo et al. 2015] utilizou sensores para capturar informações de temperatura, umidade e solo. A partir de tais informações, aplicou regras pré-definidas para controlar a quantidade e duração da irrigação de vegetais. Em outro trabalho [Yuwono et al. 2015], foi utilizado seleção não supervisionada de atributos e um algoritmo de agrupamento inspirado no Enxame de Partículas, do inglês *Particle Swarm Optimization (PSO)*, para detecção automática de falhas e diagnóstico em sistemas de ventilação e de ar-condicionado.

A partir da discussão acima, é importante salientar que nenhum dos trabalhos encontrados na literatura utilizou algoritmos de agrupamento para encontrar

os grupos que são particulares de um ambiente específico tampouco algoritmos de classificação para tomar decisões. Além disso, poucos utilizam rede de sensores de temperatura e umidade.

3. Controle Térmico em Laboratórios

Os laboratórios de análise química têm características bem específicas, pois podem variar de acordo com o prédio que guarda suas instalações. Adicionalmente, existem exigências que precisam ser seguidas, nomeadamente o controle da temperatura e umidade relativa do ar. O controle dessas variáveis é de suma importância haja vista que influencia diretamente no resultado das amostras assim como danificar os equipamentos do laboratório. O objeto principal deste trabalho, o controle de ambientes, está sendo testado em um dos laboratórios de análises químicas da UFRN.

Devido a fatores tão importantes, é essencial distribuir vários sensores pelo laboratório, de modo que todo o espaço esteja coberto por eles. A Figura 1 descreve o cenário utilizado no trabalho. Perceba que cinco sensores (azul claro) foram distribuídos pelo espaço físico, sendo que três ficaram na parte à esquerda (maior número de equipamentos) e dois ficaram na parte à direita (menor número).

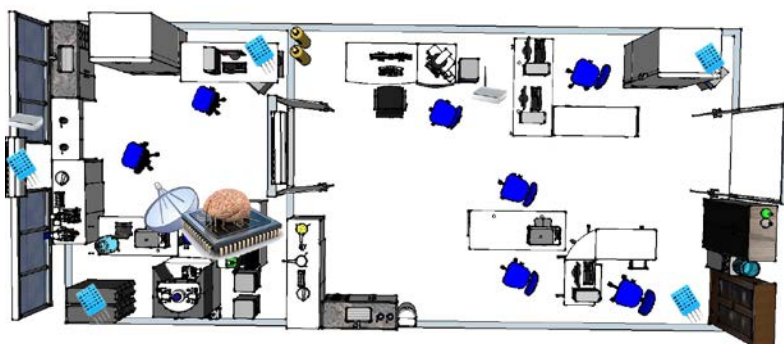


Figure 1. Distribuição dos sensores no laboratório.

Em relação aos aspectos de hardware, existem diversos sensores de fácil aquisição que podem ser utilizados para o monitoramento de temperatura e umidade [Margolis 2011, McRoberts 2011]. A Figura 2 descreve os sensores mais utilizados para esse fim [McRoberts 2011], nomeadamente, o LM35 (sensor de precisão de temperatura) e o DHT11 (sensor de umidade e temperatura). Em testes realizados no laboratório, o sensor de temperatura LM35 apresentou melhor precisão que o sensor DHT11. Por essa razão, o sensor DHT11 foi utilizado apenas para monitorar a umidade.

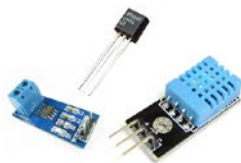


Figure 2. Sensores mais utilizados para medição de temperatura e umidade.

Um aspecto também importante a ser mencionado é a questão da comunicação entre os sensores. No cenário descrito na Figura 1, os cinco sensores de

temperatura e umidade são conectados em uma topologia em estrela a um centralizador ou comumente chamado na área de redes de sensores de *sink*. Tal dispositivo irá coletar todos os dados e compartilhar através de um enlace Wifi a um banco de dados remoto. O centralizador é um hardware dedicado e construído a partir de microcontroladores ATMEGA e com software embarcados. Essa solução é uma das contribuições do trabalho. A Figura 3 descreve uma visão do centralizador.



Figure 3. Protótipo de monitoração e controle de temperatura e umidade.

4. Metodologia de Experimentos

Como pode ser visualizado na Figura 1, os sensores foram distribuídos em cinco pontos estratégicos do laboratório, visando desta forma, cobrir as áreas mais críticas com relação a temperatura e umidade. O tempo de aferição foi setado em 10 minutos, ou seja, a cada 10 minutos os sensores enviavam dados para o controle que os armazenava em uma base de dados. Também utilizamos tempos de amostra iguais a 5 e 15 minutos, porém verificamos que não houve alteração significativa de temperatura e umidade. Depois disso, o protótipo ficou em execução por 50 horas, e com isso armazenou-se um total de 300 registros. Cada registro gerado é composto por 10 (dez) atributos, sendo cinco referente as temperaturas e cinco as umidades.

O objetivo principal desse primeiro experimento era gerar uma base de dados, para que a partir dela, se descobrisse algum padrão de comportamento de temperatura e umidade dentro do laboratório de análises químicas escolhido como caso de uso. O método utilizado para descobrir padrões na base de dados foi o de agrupamento ou *clustering*. Nesse método, os valores dos atributos são utilizados conjuntamente com uma medida de distância, para dessa forma, separar os registros ou instâncias em grupos ou *clusters*.

Na fase de agrupamento, escolhemos três métodos de agrupamentos bastante conhecidos na literatura, e a partir deles criamos partições de k grupos variando de dois a seis grupos. Dessa forma, gerou-se cinco bases clusterizadas (com k grupos) para cada método de agrupamento, perfazendo um total de vinte bases (4 métodos x 5 partições), já que um dos métodos foi usado duas vezes. Mudando-se para isso, alguns parâmetros probabilísticos. O passo seguinte foi avaliar as partições criadas. Para essa tarefa, escolheu-se dois índices internos de validação de agrupamentos, já que as bases não possuíam classes ou rótulos. Os índices geraram valores para cada uma das vinte bases. A partir de tais valores, escolheu-se as três melhores bases geradas por cada um dos quatro métodos de agrupamento. Essas bases foram usadas na fase seguinte.

Na fase de classificação, escolheu-se seis métodos de classificação bastante conhecidos na literatura. Além da melhor base (partição), de acordo com os índices, pegou-se as três melhores bases resultantes de cada método de agrupamento, perfazendo um total de 12 (doze) bases. Aplicamos validação cruzada para cada uma das bases. O objetivo principal dessa fase era verificar o desempenho dos métodos de classificação, considerando o percentual de acerto, aqui chamado de acurácia. Baseado na acurácia de cada um dos seis classificadores sobre cada uma das 12 bases resultantes da fase anterior, calculou-se a média e o desvio padrão para que fosse possível averiguar se os desempenhos eram significativamente diferentes. A partir do teste estatístico realizado, chegou-se ao método de classificação mais adequado para classificar (rotular) os registros futuros.

Na fase de controle, identificamos o conjunto de ações possíveis de controle para o ambiente em questão, e as variações de estados (possíveis valores da classe). Desse modo, o controle passou a executar as ações relacionadas aos estados, assim como, a desfaz as ações anteriores. Dentre os valores da classe, identificamos cinco possíveis valores, que são: frio, normal, calor1, calor2 e perigo. Para cada valor, há pelo menos uma ação vinculada, que visa controlar os aparelhos de ar-condicionado e desumificador.

As subseções seguintes mostrarão em detalhes os métodos utilizados nas fases de agrupamento e classificação relatadas acima.

4.1. Algoritmos de Agrupamentos

Existe uma grande diversidade de algoritmos de agrupamento que podem ser encontrados na literatura [Xu and Wunsch 2005]. Nesta subseção serão apresentados os algoritmos utilizados na fase de agrupamento deste trabalho.

- **k-Médias:** o algoritmo k-Médias é comumente utilizado em tarefas de agrupamento, principalmente por de fácil implementação. Ele se baseia na ideia de centroides, onde o centróide representa a instância média de um grupo, ou seja, a instância que é a média de todas as outras dentro de um mesmo grupo. Os elementos são agrupados em k grupos com base em alguma medida de similaridade. Geralmente, é utilizado a distância Euclidiana.
- **Expectância Máxima:** o algoritmo Expectância Máxima, do inglês *Expectation-Maximization* - EM, é uma técnica usada para estimar funções de máxima verossimilhança a partir de dados incompletos, de modo que se os dados estão incompletos, pode-se utilizar os casos em que foram observadas as variáveis para aprender a prever seus valores quando não observados [Faceli et al. 2011].
- **Hierárquico Aglomerativo:** os algoritmos de agrupamento hierárquicos geram uma sequência de partições aninhadas, semelhante a uma estrutura hierárquica do tipo árvore [Faceli et al. 2011]. Há dois tipos de algoritmos: os divisivos e os aglomerativos. Na primeira, os objetos são alocados em um único grupo, sendo divididos sucessivamente até o número de grupos escolhido previamente. Já no outro tipo, inicialmente, cada objeto representa um grupo, sendo aglomerados até o número de grupos desejado.

4.2. Medidas para Validação

Como forma de validar as 20 bases criadas a partir da utilização dos algoritmos de agrupamentos citados na subseção anterior, utilizou-se os índices Davies-Bouldin (DB) e Silhueta [Halkidi et al. 2002].

Índice Davies-Bouldin (DB): esta medida calcula a compacticidade dos grupos formados, através da dispersão intra-grupos e entre os grupos. Dado uma partição $C = \{C_1, C_2, \dots, C_k\}$, o índice DB pode ser definido da seguinte forma:

$$DB(C) = \frac{1}{k} \sum_{i=1}^k \max_{j=1, \dots, k | j \neq i} \left\{ \frac{\Delta(C_i) + \Delta(C_j)}{\delta(C_i, C_j)} \right\} \quad (1)$$

em que k é o número de *clusters*; $\Delta(C_i)$ é a distância intra-*cluster* para o *cluster* i e $\delta(C_i, C_j)$ é a distância entre os centróides dos *clusters* C_i e C_j .

A distância intra-*cluster*, $\Delta(C_i)$, representa a distância média de todos os elementos do *cluster* para o seu centróide e pode ser definida como sendo:

$$\Delta(C_i) = \frac{1}{|C_i|} \sum_{x \in C_i} (x - z_i)^2 \quad (2)$$

em que x é uma instância de i e z_i é o centróide do grupo i .

Silhueta: a medida de silhueta mede a qualidade dos *clusters* com base na proximidade entre as instâncias de um *cluster* e na distância das instâncias de um *cluster* com relação aos outros *clusters*. As silhuetas mostram quais instâncias estão bem situados dentro dos seus *clusters* e quais não estão. Seja $a(x_i)$, a distância média da instância x_i em relação a todas as outras instâncias do *cluster* C_i ; $d(x_i, C_j)$, a distância média da instância x_i em relação as instâncias do *cluster* C_j ; e $b(x_i)$, a menor distância média de x_i em relação a todos os demais *clusters*, pode ser definida como sendo:

$$b(x_i) = \lim_{C_i \neq C_j} d(x_i, C_j) \quad (3)$$

Desse modo, a silhueta de uma instância, $s(x_i)$, pode ser definida como sendo:

$$s(x_i) = \begin{cases} 1 - \frac{a(x_i)}{b(x_i)}, & a(x_i) < b(x_i) \\ 0, & a(x_i) = b(x_i) \\ \frac{b(x_i)}{a(x_i)} - 1, & a(x_i) > b(x_i) \end{cases} \quad (4)$$

4.3. Algoritmos de Classificação

Os algoritmos de classificação, também chamados de algoritmos preditivos, são aplicados em problemas onde as instâncias de uma base de dados, representados por um

conjunto de atributos, precisam ser enquadrados em um conjunto pré-definidos de possíveis rótulos (classes). Há um bom número de algoritmos de classificação propostos na literatura. Os algoritmos utilizados neste trabalho são descritos a seguir.

- *k*-Nearest-Neighbor (*k*-NN): o algoritmo *k*-NN, também conhecido como *k* vizinho mais próximo, é um método baseado em instância que classifica uma instância de teste baseada na a classe referente as instâncias mais próximas pertencentes ao conjunto de treinamento [Mitchell 1997].
- Árvore de decisão (AD): esse algoritmo utiliza a técnica de dividir para conquistar na solução de problemas de classificação. Seu funcionamento consiste em, dado um conjunto de treinamento, gerar uma árvore de decisão como saída. Essa árvore será utilizada para classificar as instâncias de teste [Berry and Linoff 2004].
- Rede Neural Artificial: o funcionamento das redes neurais artificiais é baseado no sistema nervoso. Uma RNA é composta por neurônios artificiais que estão conectados em camadas. Os neurônios podem estar organizados em três camadas: de entrada, escondida e de saída. O Multilayer Perceptron (MLP) é um dos tipos de RNAs mais conhecidos na literatura [Berry and Linoff 2004].
- Máquina de vetor de suporte (SVM): o algoritmo SVM é embasado na teoria de aprendizado estatístico, que permite a obtenção de classificadores com boa capacidade de generalização [Lorena and de Carvalho 2007].
- Naive bayes (NB): é um algoritmo probabilístico que é baseado no teorema de Bayes, originado na Estatística. Para que esse método funcione corretamente é necessário garantir que os dados sejam independentes entre si [Mitchell 1997].

5. Resultados dos Experimentos

Nesta seção, apresentamos os resultados experimentais obtidos nas fases de agrupamento e de classificação. É importante lembrar que a base de dados utilizada foi gerada pelo próprio protótipo, que recebeu dados de cinco sensores de temperatura e umidade. Cada sensor gera dois atributos, totalizando 10 (dez) atributos. O protótipo funcionou por 50 horas, totalizando 300 registros. Os valores de temperatura variaram de 15 a 47 graus Celcius. Já os valores de umidade variaram de 20 a 60%.

Como já foi mencionado anteriormente, utilizamos dois índices internos para avaliar as partições criadas na fase de agrupamento. A Tabela 1 mostra os valores do índice DB correspondente a cada uma das bases (partições) criadas, assim como também a média e o desvio padrão dos referidos valores. Observe que para os valores do índice DB, quanto menor for o valor gerado, melhor será a partição. Os experimentos mostraram, que de acordo com o índice DB, o algoritmo Hierárquico Aglomerativo gerou a melhor partição com 5 (cinco) grupos.

A Tabela 2 mostra os valores do índice Silhueta correspondente a cada uma das bases (partições) criadas, e também a média e o desvio padrão dos referidos valores. É importante enfatizar, que para os padrões do índice Silhueta, quanto maior for o valor, melhor será a partição. Os experimentos mais uma vez mostraram, que de acordo com o referido índice, o algoritmo Hierárquico Aglomerativo gerou a melhor partição com 5 (cinco) grupos.

Table 1. Resultados experimentais obtidos para o índice DB.

Índice DB							
Grupos (k)	2	3	4	5	6	Média	D. Padrão
EM	37,5822	42,1109	48,3575	45,4935	37,2736	42,1635	4,8568
HA	43,4795	39,3663	40,7793	33,5883	34,2792	38,2985	4,2566
<i>k</i> -Média01	38,3968	41,1507	48,3575	46,2688	42,6041	43,3555	3,9850
<i>k</i> -Média02	48,9702	38,2874	41,4610	40,7931	41,9508	42,2924	3,9902

Table 2. Resultados experimentais obtidos para o índice Silhueta.

Índice Silhueta							
Grupos (k)	2	3	4	5	6	Média	D. Padrão
EM	0,4756	0,3500	0,3083	0,3462	0,4024	0,3765	0,0647
HA	0,4601	0,5126	0,5069	0,5599	0,5310	0,5141	0,0366
<i>k</i> -Média01	0,4683	0,3663	0,3083	0,2487	0,2798	0,3343	0,0865
<i>k</i> -Média02	0,3275	0,2423	0,3409	0,2349	0,3819	0,3055	0,0643

Analisando as melhores partições, percebeu-se que houve um padrão similar de comportamento entre os grupos gerados. Devido a esse fato, decidiu-se utilizar as três melhores partições geradas por cada algoritmo de agrupamento, e não somente a melhor, de acordo com os índices de validação. Assim, as partições com 4, 5 e 6 grupos foram utilizadas nessa fase (classificação), ilustrada na Tabela 3. Na primeira linha depois do cabeçalho desta tabela, onde se lê EM_4k, na verdade significa: partição com 4 grupos gerada pelo algoritmo EM (Expectância Máxima). Ou seja, primeiro o nome do algoritmo de agrupamento, depois o parâmetro probabilístico (semente), se for o caso, e por último, o número de grupos. Como foram utilizadas as três melhores de cada algoritmo, temos um total de 12 bases (3 bases x 4 algoritmos).

Além das bases, a referida tabela ainda mostra os cinco algoritmos de classificação (colunas) utilizados nos experimentos. Vale salientar que o algoritmo k-NN foi utilizado duas vezes, com $k = 1$ e $k = 3$. Dessa forma, para cada uma das 12 bases, anotou-se o comportamento de cada classificador, usando-se para tal a acurácia (percentual de acerto). Ainda na mesma tabela, pode-se verificar a média geral da acurácia de cada classificador, assim como o desvio padrão. Note, que o classificador MLP obteve a melhor média entre os classificadores. Em segundo lugar ficou o 1k-NN com 97,01149 de média. A Árvore de Decisão (AD) obteve a pior média entre todos.

A fim de comparar o desempenho dos algoritmos de classificação, aplicou-se um teste estatístico, chamado teste de hipótese (t-test unicaudal), com um nível de confiança de 95% ($\alpha = 0,05$) [Fisher 1925]. Através deste teste estatístico, foi possível analisar o desempenho de cada classificador, comparando-os entre si. Dessa forma, concluiu-se que o desempenho do melhor (MLP) não foi estatisticamente significativo, se comparado com o segundo melhor (1k-NN). Na verdade, o MLP foi estatisticamente melhor apenas que o AD. Logo, qualquer um dos quatro classificadores (MLP, k-NN, SVM e NB) podem ser implementado no protótipo, para quando em funcionamento, possa vir a classificar as novas instâncias (registros) geradas pelos sensores de temperatura e umidade.

Contudo, como é de conhecimento geral, alguns classificadores são mais com-

Table 3. Resultados experimentais obtidos para os seis classificadores.

Acurácia (%) dos Classificadores						
Bases	1k-NN	3k-NN	AD	MLP	SVM	NB
EM_4k	95,1724	92,7586	95,1724	96,2069	93,1034	97,9310
EM_5k	97,2414	97,2414	93,7931	95,1724	87,9310	98,9655
EM_6k	96,2069	96,8966	94,1379	96,5517	94,8276	99,6552
HA_4k	98,9655	98,9655	98,9655	99,3103	98,9655	97,9310
HA_5k	98,9655	98,9655	97,5862	99,3103	98,9655	97,5862
HA_6k	98,9655	98,9655	96,8966	99,3103	98,9655	97,9310
kM10s_4k	96,5517	96,2069	92,7586	97,2414	97,5862	94,1379
kM10s_5k	96,2069	96,5517	95,1724	96,8966	97,2414	92,4138
kM10s_6k	96,2069	96,8966	94,1379	95,1724	95,5172	93,1034
kM12000s_4k	97,5862	96,2069	96,5517	97,2414	98,2759	94,4828
kM12000s_5k	96,2069	96,5517	91,0345	96,8966	97,2414	95,1724
kM12000s_6k	95,8621	95,1724	93,7911	97,2414	96,2069	91,7241
Média	97,01149	96,78161	94,99983	97,21264	96,23563	95,91953
Desvio Padrão	1,32603	1,76033	2,20807	1,45153	3,18582	2,72644

plexos que outros. Além disso, alguns são computacionalmente mais custosos que outros. Logo, para utilização em nosso protótipo, o k-NN nos parece bastante apropriado, considerando a facilidade de implementação e o tamanho inicial da base (número de instâncias e de atributos).

Depois de finalizadas as fases de agrupamento e classificação utilizadas para criar a base de treinamento e para escolher o classificador mais adequado para ser implementado, chega o momento de pensarmos nos controles que serão vinculados aos valores da classe. Dessa forma, o sistema irá classificar as novas instâncias em função do aprendizado durante o treinamento, e executará as ações vinculadas aos possíveis valores da classe (frio, normal, calor1, calor2 e perigo).

6. Conclusão e Trabalhos Futuros

A principal contribuição deste trabalho é propor um mecanismo de controle para ambientes, baseado em rede de sensores e classificação de padrões. A ideia principal é utilizar as informações geradas pelos sensores para tomar decisões relativas ao controle constante da temperatura e da umidade relativa do ar. Como caso de uso, escolhemos um dos laboratórios de análises químicas da UFRN.

A fim de provar que nosso trabalho é factível, dividimos os experimentos em duas fases. Na primeira, também chamada de fase agrupamento, usamos três algoritmos de agrupamento para gerar a base de treinamento e teste. As partições criadas foram avaliadas a partir de dois índices de validação (DB e Silhueta). Já na fase de classificação, utilizamos cinco algoritmos de classificação para avaliar seus desempenhos sobre as três melhores partições criadas por cada algoritmo de agrupamento. Depois de analisados os resultados de ambas as fases, identificamos o conjunto de ações de controle relacionado a cada valor de estado (classe) e suas variações.

Como trabalhos futuros, precisamos comparar o desempenho de nosso sistema de controle de ambientes com outros sistemas de controle que utilizam soluções clássicas.

References

- Abraham, S. and Li, X. (2014). A cost-effective wireless sensor network system for indoor air quality monitoring applications. *Procedia Computer Science*, 34(0):165 – 171.
- Barroca, N., Borges, L. M., Velez, F. J., Monteiro, F., Górski, M., and Castro-Gomes, J. (2013). Wireless sensor networks for temperature and humidity monitoring within concrete structures. *Construction and Building Materials*, 40(0):1156 – 1166.
- Berry, M. J. and Linoff, G. S. (2004). *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management*. Wiley Publishing, Inc, Indianapolis, Indiana, USA, 2nd edition.
- da Silva, L. T. (2013). Uma proposta de automação residencial utilizando a plataforma de prototipagem eletrônica arduino. *Revista Tecnologias em Projeção*.
- Faceli, K., Lorena, A. C., Gama, J., and de Carvalho, A. C. (2011). *Inteligência Artificial: uma Abordagem de Aprendizado de Máquina*. LTC - Livros técnicos e científicos Editora Ltda.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.
- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002). Clustering validity checking methods: part ii. *SIGMOD Rec.*, 31 (3):19–27.
- Lorena, A. C. and de Carvalho, A. C. P. L. F. (2007). Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada (RITA)*, 14(2):43 – 67.
- Margolis, M. (2011). *Arduino Cookbook*. O’Reilly Media, USA, 2nd edition.
- Maycon J. C. Mesquita, Jose C. G. Neto, W. R. M. A. (2014). Sistema de controle e supervisão residencial utilizando plataforma arduino.
- McRoberts, M. (2011). *Beginning Arduino*. Apress, USA, 1st edition.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1st edition.
- Vera-Repullo, J., Ruiz-Peñalver, L., Jiménez-Buendía, M., Rosillo, J., and Molina-Martínez, J. (2015). Software for the automatic control of irrigation using weighing-drainage lysimeters. *Agricultural Water Management*, 151(0):4 – 12. New proposals in the automation and remote control of water management in agriculture: agromotic systems.
- Xu, R. and Wunsch, II, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678.
- Yuwono, M., Guo, Y., Wall, J., Li, J., West, S., Platt, G., and Su, S. W. (2015). Un-supervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in heating ventilation and air conditioning systems. *Applied Soft Computing*, 34(0):402 – 425.

Resolvendo conflitos em aplicações de distribuição de conteúdo em redes veiculares

Ronan D. Mendonça^{1,2}, Thais R. M. Braga Silva², Fabrício A. Silva², Linnyer B. Ruiz^{1,3}

¹Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais (UFMG), MG – Brasil

²Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa (UFV) – Campus Florestal, MG – Brasil

³Departamento de informática
Universidade Estadual de Maringá (UEM) – Maringá, PR – Brasil

{ronan.dutra, thais.braga, fabricio.asilva}@ufv.br, linnyer@gmail.com

Abstract. *The distribution of entertainment content in vehicular networks consume a large volume of bandwidth. When solicitations are not attended, it occurs conflicts caused by the dispute for resources, reducing the users' satisfaction. This paper proposes a solution to solve these conflicts, considering the best algorithm according to environmental contexts and the application, through the CReMe methodology. The obtained results show that was possible use the methodology to solve those conflicts, and still be adaptative to the context, balancing the use of resources and the users' satisfaction.*

Resumo. *A distribuição de conteúdo de entretenimento em redes veiculares consome um volume significativo de largura de banda. Quando as solicitações de conteúdos não são totalmente atendidas, ocorrem conflitos de interesses na disputa pelos recursos, reduzindo a satisfação dos usuários. Este trabalho propõe uma solução para resolver estes conflitos, considerando o algoritmo mais adequado ao contexto do ambiente e da aplicação, por meio da metodologia CReMe. Os resultados obtidos mostraram que foi possível utilizar a metodologia para resolver os conflitos, de modo adaptativo ao contexto, balanceando o consumo de recursos e a satisfação do usuário.*

1. Introdução

As redes veiculares têm como função viabilizar uma conexão para a troca de mensagens e compartilhamento de recursos entre os próprios veículos e ou entre os veículos e estações fixas. Aplicações de entretenimento para as redes veiculares têm a função de distribuir conteúdos como músicas, vídeos, páginas web e são de grande importância e utilidade por atuarem em um ambiente cada vez mais caótico, implicado por problemas gerados pelo próprio trânsito dos veículos [Costa-Montenegro et al. 2012]. Um dos grandes desafios inerentes às aplicações de entretenimento é a alta demanda por largura de banda [Jiau et al. 2015].

Para as aplicações de uso coletivo disponibilizarem seus recursos, com todos os seus benefícios, são imprescindíveis a utilização dos dados de contexto dos usuários e o ambiente. Os dados de contexto são informações relevantes ao sistema e referem à

situação em que se encontra o ambiente e o próprio usuário [Perera et al. 2014]. As aplicações veiculares cientes de contexto são também coletivas, uma vez que os recursos utilizados são compartilhados por vários veículos simultaneamente. No instante, vários dos recursos disponibilizados pelas aplicações são limitados e podem se tornar escassos, comprometendo a utilização da aplicação e gerando uma disputa coletiva por estes recursos [Jiau et al. 2015]. Para não deixar que os recursos se esgotem e os usuários fiquem completamente insatisfeitos, torna-se necessário a detecção e o tratamento dos conflitos ocorridos por metodologias e procedimentos específicos ao problema.

O objetivo deste trabalho é propor uma solução para identificar e tratar conflitos em aplicações de entretenimento para redes veiculares. Para isso, será utilizada a metodologia CReMe (Conflict Resolution Methodology) [Silva. 2010], proposta para resolver conflitos em sistemas coletivos, ubíquos e cientes de contexto. Foram agregados a esta metodologia, as particularidades das redes veiculares e algoritmos específicos para a resolução dos conflitos pertinentes ao ambiente veicular. Por meio desta metodologia, espera-se que o consumo de recursos seja balanceado com a satisfação dos usuários, fazendo com que os usuários sejam atendidos sem que os recursos sejam degradados.

As próximas seções deste artigo estão organizadas da seguinte forma: na Seção 2 são apresentados os trabalhos relacionados; na Seção 3 é apresentada a solução proposta por este trabalho; na Seção 4 são descritos os detalhes da simulação e apresentados os resultados obtidos juntamente com as análises dos mesmos; e a Seção 5 apresenta as considerações finais e os trabalhos futuros.

2. Trabalhos relacionados

A discussão dos trabalhos relacionados abrange os estudos que envolvem conteúdo de entretenimento, utilização de dados de contexto e taxa de transmissão em redes veiculares.

Com relação à entrega de conteúdo de entretenimento, o trabalho de [Sarakis et al. 2016] propõe uma avaliação da capacidade das redes veiculares na transmissão de conteúdos multimídia. A avaliação incide na camada de aplicação para *streaming* de vídeos. Os autores realizaram testes para diferentes tipos de vídeos, velocidades e ambientes, concluindo que uma qualidade aceitável de transmissão de conteúdos multimídia pode ser alcançada em redes veiculares. A abordagem proposta por [Yang et al. 2012] tem como objetivo principal a codificação de dados para distribuição de fluxo contínuo de conteúdo multimídia em redes veiculares. Os autores relatam o problema de se transmitir conteúdo de entretenimento em redes veiculares com baixa largura de banda disponível. Propuseram então, um algoritmo de escalonamento de transmissão oportunista para melhorar o desempenho em diferentes densidades da rede. [Jiau et al. 2015] apresentam e avaliam arquiteturas para serviços multimídia em redes veiculares baseadas em nuvens. Por meio de uma comparação da estabilidade, capacidade de extensão, gestão e outras características, os autores concluem que a forma híbrida, onde um veículo faz o papel da infraestrutura e de um usuário final ao mesmo tempo, é a mais indicada para utilização de serviços multimídia.

Com relação aos dados contextuais, existem trabalhos que focam na utilização destes para uma distribuição eficiente de recursos. Em [Wan et al. 2014], os autores apresentam uma proposta de arquitetura em camadas baseada em dados de contexto para utilização de redes sociais e aplicações de segurança em redes veiculares. As camadas têm

como objetivos compartilhar conexão e recursos de informação de tráfego e de entretenimento aos veículos, fornecer informações de contexto e interação dos serviços disponíveis e disponibilizar aplicações e serviços em nuvem. [Sepulcre et al. 2015] propõe uma técnica de uso de informações de contexto para decidir qual tecnologia de comunicação é mais adequada ao uso em cada momento. As decisões são feitas diretamente por cada veículo com a assistência da infraestrutura na aquisição dos dados de contexto. Os autores apresentam os resultados obtidos, por meio de simulações, que demonstram benefícios na utilização de dados contextuais na tomada de decisão em redes heterogêneas.

Com relação à taxa de transmissão em redes veiculares, o trabalho de [Shankar et al. 2008] propõe a seleção de taxa de transferência baseando-se em informações de contexto para redes veiculares a partir da camada de aplicação. Por meio de uma adaptação, pela qualidade da conexão, é realizada a seleção da taxa de transferência. Em [Campolo and Molinaro 2010] e [Nwizege et al. 2013] os autores avaliam o desempenho e propõem algoritmos de seleção de taxas de transmissão. Em ambos os trabalhos a inferência da taxa de transmissão é realizada por meio dos dados de contexto da distância da posição dos veículos em relação ao outro ponto de comunicação.

Diferentemente dos estudos da literatura, este trabalho considera os contextos individuais e coletivos dos usuários, além da disponibilidade de recursos para atender às demandas da melhor maneira coletiva, sem degradar o consumo de recursos. Para isso diferentes algoritmos são usados intercaladamente de acordo com os contextos atuais.

3. Descrição da Solução

A solução proposta baseia-se na metodologia de resolução de conflitos coletivos disponibilizada na literatura por [Silva. 2010]. Foram agregados a esta metodologia as particularidades das redes veiculares e algoritmos específicos para a resolução dos conflitos pertinentes à disputa por estes recursos. Neste modelo de solução, as Unidades de Acomodamento (*RSUs*) atuam na disponibilização da aplicação com a finalidade de distribuição de conteúdo de entretenimento e no controle da distribuição de largura de banda. As Unidades de Bordo (*OBUs*) disponibilizam os seus dados de contexto e solicitam os recursos de entretenimento às *RSUs*, que por sua vez verificam e atendem as solicitações de acordo com suas disponibilidades. A insuficiência de recursos disponibilizados pode gerar uma insatisfação no usuário não atendido. Esta solução busca, por meio da escolha de algoritmos mais adequados e mediante o contexto atual da rede e dos veículos, balancear o consumo de recursos (como a largura de banda), o tempo consumido para a resolução dos conflitos e a satisfação geral dos usuários.

3.1. Instância da Metodologia CReMe para Redes Veiculares

Neste trabalho, foi desenvolvida uma instância da metodologia CReMe com a finalidade de resolver os conflitos de interesses coletivos ocorridos em aplicações de entretenimento para redes veiculares. O controle na distribuição de recursos acontece devido a quantidade limitada da largura de banda disponibilizada pela infraestrutura por meio de um repositório de algoritmos com opções que podem manter o equilíbrio entre a satisfação dos usuários, o tempo de resposta e o consumo dos recursos disponíveis.

Os elementos que compõem o modelo da aplicação são divididos entre os módulos de *Solicitação de Conteúdo* que é disponibilizado nas *OBUs* e o da *Instância da metodologia CReMe* que estará disponível nas *RSUs*.

A aplicação veicular atua de acordo com o seguinte modelo conceitual:

- **Inserção de solicitações:** Cada *OBU* pode solicitar um conteúdo de entretenimento à *RSU* que está conectada, durante uma determinada *JTS* (Janela de Tempo de Solicitações), que são intervalos de tempo nas quais as *RSUs* aguardam por solicitações.
- **Liberação de largura de banda:** A *RSU* analisa as solicitações feitas durante a *JTS*. As solicitações recebidas por uma *RSU* são devidamente tratadas pela abordagem de resolução de conflitos.
- **Execução da tarefa:** Uma quantidade de largura de banda é atribuída a cada *OBU* solicitante, e assim elas estarão autorizadas a obter o conteúdo.

Todo esse processo, à partir da solicitação até a obtenção do conteúdo, é considerado como sendo uma rodada da aplicação. Para a liberação de banda, a *RSU* depende das demandas imediatas pelo recurso solicitado e da taxa de transferência de dados residual na mesma. O valor que será concedido à *OBU* resultará, nos casos em que houver conflitos, do processo de resolução de conflitos gerado na disputa pelo recurso. A disponibilidade de largura de banda de cada *RSU*, em um determinado momento, é dada pela sua taxa de transferência de dados, definida pela tecnologia de comunicação utilizada, subtraindo o valor já alocado aos veículos e que ainda estão em uso do serviço.

3.2. Detecção e Resolução de Conflitos

Baseado no modelo de aplicação, um conflito ocorre sempre que a soma da quantidade de largura de banda solicitada por todos os veículos, à partir de uma *RSU* e em uma determinada rodada, superar a quantidade de banda disponível naquele momento. Ou seja, um conflito coletivo ocorre quando a soma das solicitações ultrapassa a largura de banda disponível. Caso isso ocorra, o módulo de Resolução de Conflitos é acionado. Caso contrário, as solicitações são atendidas e a rodada é encerrada.

O foco da resolução de conflitos é equilibrar o consumo de recursos com a satisfação dos usuários, mantendo transparente para eles a existência do conflito. Os usuários da aplicação almejam receber o conteúdo solicitado antes de chegar ao seu destino final. Sabe-se que os veículos deslocam a partir de um ponto de origem até um ponto de destino e a uma determinada velocidade média. A disponibilização dos recursos requeridos deve ser suficiente, ou estar o mais próximo de conseguir satisfazer os usuários na suas solicitações. Os recursos que não podem ser utilizados indiscriminadamente pelos algoritmos de resolução de conflitos são o tempo e a largura de banda. Assim sendo, caso o esquema de resolução de conflitos leve muito tempo para resolver a disputa, ou seja, o uso excessivo do recurso tempo, uma *OBU* pode não estar mais interessada no conteúdo solicitado ou ficar fora de alcance da *RSU*. Além disso, o algoritmo de resolução não deve utilizar muitas trocas de mensagens para resolver os conflitos pois sobrecarregaria ainda mais a rede.

Os algoritmos específicos para resolução de conflitos são armazenados em um repositório de algoritmos conforme regras da metodologia CReMe, e são indicados para uso em um dado momento por meio do cálculo da *Distância Euclidiana* dos meta-dados correspondentes aos algoritmos recuperados a partir dos valores definidos previamente. Para a proposta de aplicação deste trabalho foram utilizados os seguintes algoritmos:

- **Negociação:** a *RSU* troca mensagens com as *OBUs* com o intuito de negociar e, conseqüentemente, reduzir a demanda por largura de banda;

- **Divisão Igualitária:** a *RSU* divide igualmente entre todas as *OBUs* solicitantes a sua largura de banda residual.

Uma simples verificação nos procedimentos do *Algoritmo de Negociação* mostra claramente que ele utiliza um maior tempo e maior consumo dos recursos de rede. Porém, ele tende a proporcionar melhores níveis de satisfação dos usuários, uma vez que irá fornecer a largura de banda necessária para o usuário receber todo o conteúdo negociado. Por outro lado, o algoritmo de *Divisão Igualitária* atua mais rapidamente que o de *Negociação* e aloca a mesma fração da largura de banda para todas as *OBUs* com solicitações em conflito. Porém isso pode resultar para alguns um valor insuficiente, levando à insatisfação. Desta forma, a metodologia *CReme* visa utilizar um dos dois algoritmos de acordo com os contextos atuais, visando o equilíbrio entre a satisfação coletiva dos usuários e o consumo de recursos como tempo de resposta e largura de banda.

4. Análise dos Resultados

4.1. Ambiente de simulação

O ambiente de simulação utilizado para a implementação do modelo proposto foi a ferramenta de simulação chamada *SIAFU* [Martin and Nurmi 2006]. A simulação deste modelo utiliza o mapa da cidade de Valência na Espanha, e conta com 300 veículos se locomovendo de um ponto inicial a um destino final. Três *RSUs* são distribuídas na abrangência do mapa com uma capacidade em largura de banda de 60 Mbps cada. Os veículos se locomovem a velocidades que variam entre 5 e 20 metros por segundo (m/s). Cada simulação transcorre por um período de 12 horas e durante este tempo, os veículos têm uma possibilidade aleatória, gerada por uma distribuição uniforme, de solicitar um conteúdo à aplicação de entretenimento. Este conteúdo é diferenciado pelo seu tamanho em megabytes (MB) que pode variar de 5 a 6 MB. Estes valores são gerados também aleatoriamente por uma distribuição uniforme, obtendo assim tamanhos diferentes de conteúdos mas dentro do intervalo estipulado. Foram criados três *cenários* para a proposta de resolução de conflitos neste modelo e são apresentados na Tabela 1.

Tabela 1. Abordagens de cenários para avaliação

Cenário	Descrição
Cenário 1 (CReme)	Solução proposta usando a metodologia CReme
Cenário 2 (Negociação)	Usa somente o algoritmo de negociação
Cenário 3 (Igualitário)	Usa somente o algoritmo de divisão igualitária

Os Cenários 2 (Negociação) e 3 (Igualitário) foram criados para serem comparados com a avaliação do Cenário 1 (CReme) que utiliza a metodologia CReme. No Cenário 2 (Negociação), somente o algoritmo de negociação é utilizado para ser priorizado um maior índice de satisfação coletiva. Já para o Cenário 3 (Igualitário), foi utilizado o algoritmo de divisão igualitária que prioriza o consumo de recursos.

A CReme realiza a escolha do algoritmo mais indicado neste modelo por meio da observação de dois parâmetros: a quantidade de largura de banda residual na *RSU* e a velocidade média dos veículos solicitantes. A razão entre a largura de banda disponível (L) e a velocidade média (Vm) dos veículos foram utilizados na obtenção de um índice. Este índice foi chamado de índice de vazão (iQ) e é responsável por representar as condições

em que o ambiente se encontra naquele instante. Ao detectar estas condições, a CReMe escolhe o algoritmo de resolução de conflitos mais indicado para o índice determinado. Conforme os resultados obtidos pelo cálculo do índice iQ , pode-se concluir que quanto maior a largura de banda disponível e menor a velocidade dos veículos, maior será a quantidade de dados possíveis de se transmitir em um determinado intervalo de tempo.

4.2. Análise das Simulações

A análise sobre os dados apresentados nesta seção tem como objetivo avaliar a natureza dinâmica da metodologia CReMe para resolução de conflitos coletivos em ambientes pervasivos e cientes de contexto. Por meio da utilização de cada estratégia de resolução de conflitos, apresentadas na Tabela 1, foram comparados os resultados para apresentar a estratégia do Cenário 1, que mantém um equilíbrio entre o tempo gasto para a resolução dos conflitos com a satisfação da resolução em detrimento do recurso de rede. Foram capturados para cada rodada com conflitos, os dados das métricas *Consumo de rede*, *Tempo de resposta* e *Satisfação Coletiva*. Em todos os gráficos, são mostrados os valores médios e os respectivos desvios padrão de 33 rodadas de simulação. Os resultados apresentados na Figura 1 mostram os valores das métricas, obtidos na simulação para cada cenário.

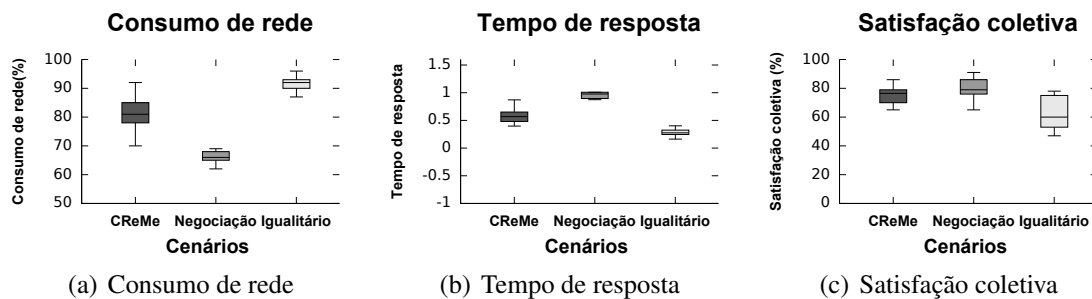


Figura 1. Métricas - Consumo de rede, satisfação coletiva e tempo de resposta

O **consumo de rede** é medido rodada a rodada pela porcentagem utilizada do total da largura de banda disponível da RSU requisitada. A Figura 1a apresenta o consumo para os três cenários avaliados. O cenário 1, que visa o equilíbrio no uso dos recursos, consegue manter o uso controlado da rede sem atingir o seu limite máximo e não deixando o recurso ocioso. Já o cenário 2, no qual é utilizada sempre uma negociação com o veículo, o consumo de largura de banda é menor, porém são feitas negociações com o usuário, para manter a satisfação, utilizando muito do recurso tempo. No cenário 3, o recurso de rede é utilizado indiscriminadamente, deixando-o muitas vezes escasso.

O **tempo de resposta** é o valor dado ao tempo gasto pelo algoritmo de resolução de conflitos para responder às requisições por largura de banda dos veículos em uma rodada. Este tempo é medido pela diferença entre o instante da requisição e o da resposta. Como se pode observar na Figura 1b, o resultado no cenário 1 situa-se entre o cenário 3 que resolve o conflito em menos tempo e o cenário 2 que consome um tempo maior para proporcionar uma melhor satisfação. Este resultado é esperado, uma vez que, o cenário 1 alterna entre as duas estratégias dos para cada mudança de contexto, visando o equilíbrio entre a satisfação e uso de recursos.

A **satisfação coletiva** foi coletada com o intuito de verificar o quão satisfeitos ficaram os usuários com a resolução do conflito existente. Foi capturada, a cada rodada,

a satisfação individual de cada veículo solicitante, e a média destas satisfações resultou na métrica de satisfação coletiva. Na Figura 1c é possível verificar a média de satisfação coletiva dada por cada cenário, sendo que o cenário 2 apresenta uma satisfação coletiva maior, por priorizar sempre esta métrica. No cenário 3, a satisfação é baixa em relação aos outros cenários, mostrando que seu objetivo é de apenas resolver o conflito em um menor tempo. Já o cenário 1 apresenta uma satisfação equilibrada entre os outros dois cenários, pois utiliza a sua estratégia para manter os usuários satisfeitos com um baixo consumo de tempo e consumo controlado de rede.

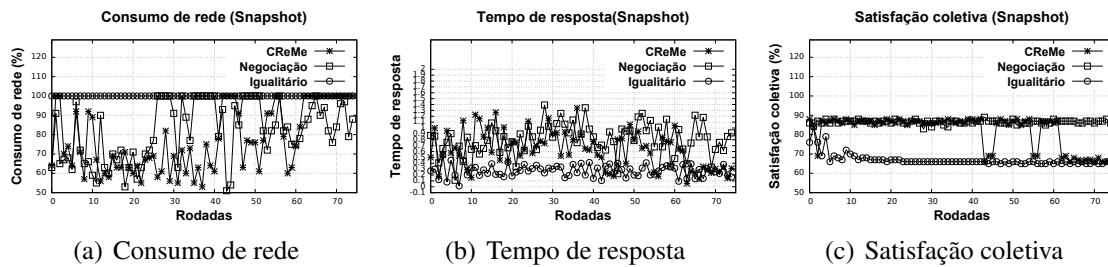


Figura 2. Snapshots - Consumo de rede, satisfação coletiva e tempo de resposta

Por meio dos snapshots da Figura 2, observa-se o comportamento das métricas avaliadas durante todo o tempo de uma simulação. Na Figura 2a é possível observar comportamentos distintos, de acordo com as características de cada um dos cenários. No Cenário 3, o consumo de recursos alcança sempre o máximo disponível para aquela rodada. Isto ocorre porque o algoritmo utilizado neste cenário divide, igualmente, o total do recurso para todos os solicitantes, conforme a Tabela 1. Já no Cenário 2, o consumo da largura de banda é atenuado por meio da negociação realizada pelo algoritmo para a redução do tamanho do conteúdo e aumento da satisfação. Para o Cenário 1, o consumo de rede oscila entre um valor total, quando é utilizado o algoritmo de divisão igualitária, e valores intermediários, quando o algoritmo indicado para ser utilizado é o de negociação, mostrando assim o dinamismo da CReMe na escolha da estratégia de resolução. Ainda na Figura 2a, é possível observar que, em algumas rodadas sequenciais, o consumo de rede permanece no máximo disponível, mostrando que naquele momento o contexto indica a escolha desta estratégia.

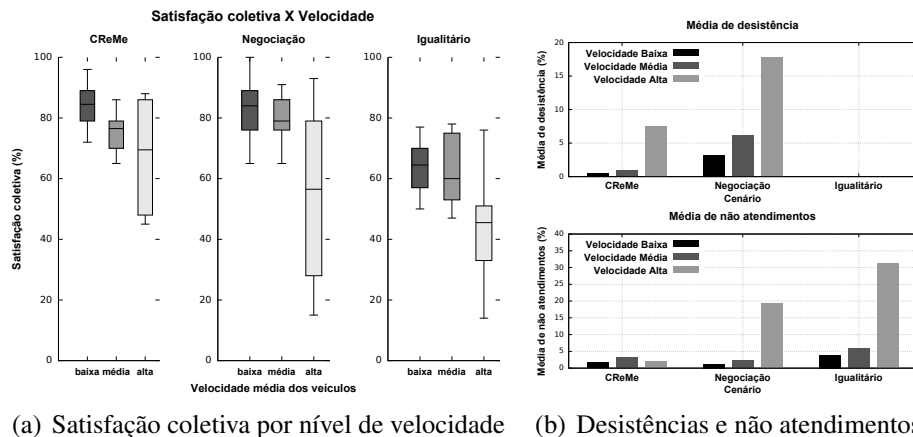
A Figura 2b apresenta os valores capturados do tempo de resposta para os três cenários. O Cenário 1 apresenta alguns valores próximos ao do Cenário 2 e outros próximos do Cenário 3, o que mostra a mudança de estratégia da CReMe durante o decorrer da simulação. Os valores apresentados para o Cenário 3 são menores devido o algoritmo utilizado consumir menos tempo de processamento do que o do Cenário 2 que consome tempo para analisar as solicitações, negociar e resolver cada situação. A característica de flexibilidade da CReMe pode ser observada na Figura 2c por meio do comportamento do nível de satisfação do Cenário 1, ao obter satisfações maiores e menores de acordo com a utilização do algoritmo mais indicado para aquela rodada.

4.3. Análise das simulações com variações de parâmetros

Análise 1: Alteração na velocidade média dos veículos. Com o objetivo de avaliar e comparar as abordagens de resolução de conflitos em relação à variação da média da velocidade dos veículos, foram utilizados três níveis de velocidades. Os níveis definem

que os veículos, em média, transitem em uma velocidade próxima ao valor indicado como o mais provável para o mesmo. Os valores foram definidos como 5 m/s para o primeiro nível de velocidade “Baixa”, 10 m/s para o segundo nível definido como “Média” e 20 m/s para o terceiro nível definido como “Alta” velocidade.

A satisfação coletiva, obtida e apresentada na Figura 3a, tende a ser reduzida quando a velocidade média está no nível “Alta”. Este fato ocorre devido à disponibilidade de tempo de comunicação entre as *OBUs* e as *RSUs* se tornar menor. A redução dos valores médios da satisfação ocorre devido ao impacto causado pelo aumento das desistências e não atendimentos quando os veículos estão trafegando em velocidades altas. A Figura 3b superior mostra o aumento do número de desistências em função do aumento da velocidade. Este fato é evidenciado principalmente para a abordagem do Cenário 2, explicando assim a queda na satisfação neste cenário. O aumento no número de desistências é afetado pela combinação do maior tempo de resposta do algoritmo de negociação e o fato dos veículos chegarem ao seu destino mais rapidamente.



(a) Satisfação coletiva por nível de velocidade (b) Desistências e não atendimentos

Figura 3. Satisfação coletiva, desistências e não atendimentos em função da alteração na velocidade dos veículos

Para a abordagem de divisão Igualitária no Cenário 3, não há desistência devido ao baixo tempo de resposta, como é apresentado na Figura 3b com valores zerados. Porém, ocorre um número maior de não atendimentos, mostrado na Figura 3b inferior, pelo fato da utilização excessiva de largura de banda por rodada, deixando assim a *RSU* sobrecarregada para a próxima rodada. Contudo, com velocidades altas, é necessário uma maior vazão de dados para que os veículos consigam receber todo o conteúdo no tempo disponível de conexão.

A abordagem utilizando a metodologia CReMe, o Cenário 1, apresenta um equilíbrio entre as outras duas abordagens, uma vez que utiliza a velocidade como um dos parâmetros para a escolha estratégica do algoritmo de resolução de conflitos. Portanto, com a variação da velocidade entre alta, média ou baixa observa-se como resultado uma satisfação coletiva estável, conforme a Figura 3a. A utilização da metodologia CReMe consegue manter um patamar baixo de desistências mesmo em altas velocidades, como é apresentado na Figura 3b. O fato de acontecer o não atendimento, mostrado na Figura 3b, é atenuado mediante a redução da velocidade dos veículos.

Análise 2: Alteração no tamanho do conteúdo solicitado pelos veículos. Foi re-

alizada uma comparação do uso da abordagem de resolução de conflitos com diferentes valores para o tamanho do conteúdo solicitado pelos veículos. Os valores possíveis para o tamanho do conteúdo foram medidos em faixas de intervalos definidos como pequeno, médio e grande, sendo os valores de 1 a 1,2 MB para “pequeno”, de 5 a 6 MB para “médio”, e de 10 a 12 MB para “grande”. A satisfação coletiva, apresentada na Figura 4a, foi obtida por meio destas variações e conforme pode ser observado, tende a ser reduzida quando o tamanho do conteúdo aumenta. Isto ocorre devido ao fato que, com um conteúdo maior, é necessário mais recursos para sua utilização.

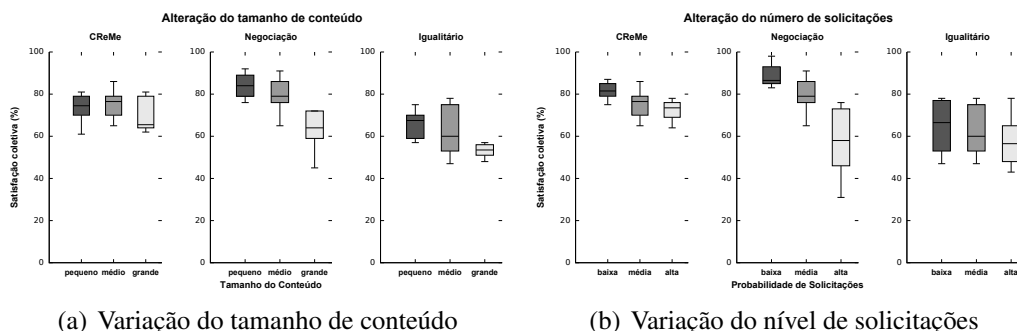


Figura 4. Satisfação coletiva em função da variação de parâmetros

Análise 3: Alteração na probabilidade de solicitações pelos veículos. Esta análise apresenta uma comparação do uso da abordagem de resolução de conflitos com diferentes probabilidades para o veículo solicitar um conteúdo. Para diferenciar as probabilidades de solicitações, foi criado o nível de taxa de solicitação “baixa”, na qual uma média de 20% dos veículos solicitam um conteúdo, o nível “média”, em que 50% dos veículos solicitam conteúdos, e o nível “alta” na qual 100% dos veículos solicitam conteúdos. O gráfico da Figura 4b apresenta, para os três cenários, uma queda na satisfação em função do aumento do número de solicitações. Este resultado mostra que é crítica a escalabilidade em redes veiculares, quando se trata da distribuição de conteúdos de entretenimento. Mais uma vez, o Cenário 1 conseguiu balancear a satisfação dos usuários, mantendo praticamente estáveis esses valores mesmo quando 100% dos veículos solicitam conteúdos.

5. Conclusão

Este trabalho propôs e avaliou uma solução para resolver os conflitos que ocorrem em redes veiculares quando os veículos exigem, simultaneamente, mais largura de banda do que a disponível para a aplicação. Os resultados obtidos mostram que este tipo de conflito deve ser identificado e resolvido de acordo com o contexto, para satisfazer os usuários e manter um consumo eficiente dos recursos. Esta abordagem é baseada em uma metodologia chamada CReMe, que equilibra a satisfação dos usuários com o consumo de recursos. Os resultados revelam a natureza flexível e dinâmica da metodologia CReMe. Os benefícios apresentados pelo comportamento dinâmico e adaptativo da metodologia CReMe são evidentes, na qual são escolhidas as formas de resolução do problema de acordo com as características apresentadas pelo contexto atual da aplicação. Por meio desta solução, foi possível balancear o nível de satisfação dos usuários com um consumo controlado de rede e tempo de resposta. Os próximos passos para este trabalho envolvem a utilização da metodologia CReMe de forma distribuída, fazendo com que os veículos resolvam, por si próprios, os conflitos existentes na disputa pelos recursos.

6. Agradecimento

O presente trabalho foi realizado com o apoio financeiro da CAPES - Brasil e do CNPq (573.738/2008-4 INCT NAMITEC).

Referências

- Campolo, C. and Molinaro, A. (2010). Data rate selection in wbs-based iee 802.11p/wave vehicular ad hoc networks. In *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, pages 412–416.
- Costa-Montenegro, E., Quinoy-Garcia, F., Gonzalez-Castano, F., and Gil-Castineira, F. (2012). Vehicular entertainment systems: Mobile application enhancement in networked infrastructures. *Vehicular Technology Magazine, IEEE*, 7(3):73–79.
- Jiau, M. K., Huang, S. C., Hwang, J. N., and Vasilakos, A. V. (2015). Multimedia services in cloud-based vehicular networks. *IEEE Intelligent Transportation Systems Magazine*, 7(3):62–79.
- Martin, M. and Nurmi, P. (2006). A generic large scale simulator for ubiquitous computing. In *Mobile and Ubiquitous Systems - Workshops, 2006. 3rd Annual International Conference on*, pages 1–3.
- Nwizege, K. S., He, J., Kim, K. S., and Iqic, P. (2013). Performance evaluation of adaptive context aware rate selection algorithm (acars) for road safety applications in vehicular network. In *Modelling Symposium (EMS), 2013 European*, pages 640–646.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys Tutorials*, 16(1):414–454.
- Sarakis, L., Orphanoudakis, T., Leligou, H. C., Voliotis, S., and Voulkidis, A. (2016). Providing entertainment applications in vanet environments. *IEEE Wireless Communications*, 23(1):30–37.
- Sepulcre, M., Gozalvez, J., Altintas, O., and Kremo, H. (2015). Context-aware heterogeneous v2i communications. In *Reliable Networks Design and Modeling (RNDM), 2015 7th International Workshop on*, pages 295–300.
- Shankar, P., Nadeem, T., Rosca, J., and Iftode, L. (2008). Cars: Context-aware rate selection for vehicular networks. In *Network Protocols, 2008. ICNP 2008. IEEE International Conference on*, pages 1–12.
- Silva., T. R. d. M. B. (2010). *Tratamento de conflitos coletivos em sistemas ubíquos cientes de contexto*. PhD thesis, Universidade Federal de Minas Gerais.
- Wan, J., Zhang, D., Zhao, S., Yang, L., and Lloret, J. (2014). Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions. *Communications Magazine, IEEE*, 52(8):106–113.
- Yang, Z., Li, M., and Lou, W. (2012). Codeplay: Live multimedia streaming in vanets using symbol-level network coding. *Wireless Communications, IEEE Transactions on*, 11(8):3006–3013.

Sistema embarcado para o controle de acesso em áreas restritas de prédios inteligentes

Hyuri S. Maciel, David F. Silva, Clóvis G. M. do Nascimento e Andre L.L Aquino

¹ Instituto de Computação – Universidade Federal de Alagoas – Maceió – AL – Brasil

{smhyuri, firminosilva.david, clovisgabrielm, alla.lins}@gmail.com

Abstract. *This paper presents an embedded system for access control in restricted area through the integration of two authentication strategies: facial recognition and RFID TAGs. Once that face recognition, in an embedded system, is susceptible to failures and RFIDs TAGs can be used by other users, the integration of these two solutions aims to increase the robustness of access control system. This system can be easily deployed into enterprise environments, meetings and offices rooms. Additionally the system proposed does not require the need to have the system connected to the Internet. Through the evaluations we observed an acceptable execution time and better robustness in access control system, even using simple techniques of face recognition.*

Resumo. *Este trabalho apresenta um sistema embarcado para o controle de acesso em áreas restritas por intermédio da integração de reconhecimento facial e autenticação com RFID. A escolha da integração das duas soluções tem como objetivo aumentar a robustez do sistema para o controle de acesso, uma vez que o reconhecimento facial, quando concebido num sistema embarcado, é suscetível a falhas no reconhecimento e as TAGS RFIDs podem facilmente ser utilizadas por outros usuários. Esse sistema pode ser concebido facilmente em ambientes empresariais, salas de reuniões e repartições públicas e não exige a necessidade de termos o sistema conectado a Internet. Por intermédio das avaliações realizadas observamos um tempo de resposta aceitável e uma maior robustez no controle de acesso, mesmo usando técnicas simples de reconhecimento de faces.*

1. Introdução

Uma cidade inteligente é um sistema urbano que utiliza tecnologia de informação e comunicação para trazer mais interatividades tanto aos aspectos de infraestrutura como aos serviços públicos em geral. Essa interatividade visa a acessibilidade e eficiência sob o ponto de vista dos cidadãos. Ademais, em uma cidade inteligente é de se esperar que a mesma esteja comprometida com o meio ambiente e com os elementos históricos e culturais. Nesse cenário a infraestrutura pode ser equipada com as mais avançadas soluções tecnológicas com o intuito de facilitar a interação do cidadão com os elementos urbanos e os diversos ambientes que eles estão inseridos [Pellicer et al. 2013].

Numa cidade inteligente é possível observar diversas possibilidades que as novas tecnologias oferecem para o fortalecimento do sistema urbano em geral. Como exemplo podemos citar: o monitoramento de poluentes tanto do ar como de rios e monitoramento e alarme sobre as condições climáticas de centros urbanos [Han et al. 2013]; as

diferentes soluções para a problemática de economia dos recursos energéticos em grandes cidades, por intermédio de *smart grids* [Weixiao et al. 2014]; a integração de veículos “inteligentes” capazes de interagir entre si para compartilhar informações de acidentes ou congestionamentos [Dias et al. 2014]; e a utilização de sensores inteligentes e sistemas embarcados com comunicação sem fio para automação de ambientes controlados, como prédios empresariais ou repartições públicas, permitindo assim a concepção de ambientes inteligentes [Sadri 2011]. Apenas delimitando o foco de nossos esforços, neste artigo, estaremos contribuindo apenas nesse último cenário apresentado.

Neste trabalho apresentamos um sistema embarcado para o controle de acesso em áreas restritas por intermédio da integração de reconhecimento facial e autenticação com RFID (*Radio Frequency Identification-Identificação por Rádio frequência*) [Weis 2016]. Esse sistema visa aumentar a robustez no controle de acesso de prédios inteligentes.

A escolha da autenticação com RFID vem da sua grande utilização em diferentes aplicações, como por exemplo, na indústria, em prédios e em lojas comerciais, tornando essa tecnologia objeto das mais diversas aplicações. RFID é um sistema sem contato, com o sistema de reconhecimento facial. Combinando o reconhecimento facial com o sistema de RFID irá adicionar mais robustez para qualquer sistema de segurança [Affandi et al. 2013]. Existem diferentes tecnologias de RFID que consideram uma comunicação de longo, médio ou curto alcance. Na concepção desse trabalho utilizamos um leitor de médio alcance.

A etapa de reconhecimento facial é implementada usando técnicas leves de visão computacional que podem facilmente ser embarcadas em, por exemplo e no nosso caso, uma placa Beaglebone Black ¹. Para a leitura, controle e tratamentos dos dados dos sensores RFID e para decidir sobre a permissão ou não do acesso aos ambientes utilizamos uma placa Arduino UNO ². O Arduino UNO é uma plataforma de prototipagem eletrônica de hardware livre e possui baixo custo. Utiliza um microcontrolador da Atmel, com entradas/saídas digitais e analógicas, possui uma linguagem de programação semelhante ao C/C++ e foi projetado para facilitar a criação de protótipos.

As principais contribuições desse trabalho são: i. avaliação de técnicas de visão computacional convencionais em sistemas embarcados; ii. integração de duas soluções para controle de acesso num mesmo arcabouço; e iii. aumento na robustez dos sistemas de controle de acesso em ambientes inteligentes. Vale destacar que não é do escopo desse trabalho a proposição de novas técnicas de visão computacional para tais aplicações. Especificamente para as técnicas de visão computacional, estamos apenas interessados em avaliar a viabilidade da utilização dessas técnicas em ambientes embarcados voltados para a aplicação de controle de acesso em prédios inteligentes. Esse último ponto é importante para que, dado um cenário específico, o projetista possa avaliar e, consequentemente, identificar rapidamente quais as melhores técnicas que podem ser utilizadas. Por intermédio das avaliações realizadas observamos um tempo de resposta aceitável e uma maior robustez no controle de acesso, mesmo usando técnicas simples de reconhecimento de faces.

Este artigo está organizado como segue: Seção 2 apresenta o embasamento teórico

¹<https://beagleboard.org/>

²<https://www.arduino.cc/>

e os trabalhos correlatos; Seção 3 mostra o funcionamento do sistema proposto para o reconhecimento de faces; Seção 4 apresenta os principais resultados; e por fim, Seção 5 conclui o trabalho.

2. Embasamento teórico e trabalhos relacionados

Para o reconhecimento facial, o método mais utilizado na literatura é o *Eigenfaces* ou autofaces [Turk and Pentland 1991]. Este método retorna um conjunto de vetores e seu fundamento básico é a utilização de vetores de distribuição probabilística para a geração de dados da face. Ele utiliza PCA (*Principal Componente Analysis*) [Jun Hu et al. 2015] para projetar e computar um subespaço para reconhecimento facial, o espaço de faces, através do treinamento de uma base de imagens, transformando as informações visuais em vetores. Esse espaço é definido pelos vetores autofaces que consistem na combinação linear dos pontos mais relevantes das imagens faciais originais, os quais são os autovetores da matriz covariância correspondente às imagens faciais originais. A Figura 1 mostra exemplos de autofaces.

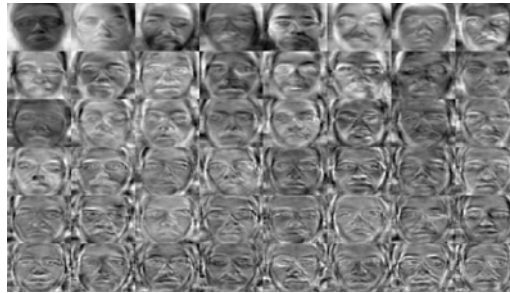


Figura 1. Conjunto de autofaces – imagens de faces após aplicar o método *Eigenfaces*.

Para compreendermos matematicamente o algoritmo de *Eigenfaces*, seja um conjunto de faces de treinamento $T = (T_1, T_2, \dots, T_M)$. Seja M a quantidade de faces presentes no conjunto T , e m a face média (um vetor que representa a face média da base de imagens de faces),

$$m = \frac{1}{M} \sum_{i=1}^M T_i. \quad (1)$$

Seja $\Phi_i = T_i - m$ a subtração da face média de cada imagem T_i , e a matriz $A = [\Phi_1 \Phi_2 \dots \Phi_n]$ onde cada coluna representa uma imagem Φ_i . A partir disso calculamos a matriz de covariância $C = A A^T$. Como a tarefa de encontrar autovalores e autovetores para a matriz C é custosa computacionalmente, então reduzimos a dimensão da matriz utilizando $C' = A^T A$. Utiliza-se a matriz de transformação $U = \Phi C'$ para encontrar os autovetores e autovalores de C' . Em seguida, para projetar cada face no espaço das faces, aplica-se $\Omega = U^T \Phi$.

Com o objetivo de reconhecer uma face de teste T_1 do banco de imagens, calcula-se a subtração da face média da face de teste $\Phi_1 = T_1 - m$, para então projetar a face de teste no espaço das faces $\Omega = U^T \Phi_1$. Por fim, calcula-se a distância euclidiana $\varepsilon_i = \|\Omega_1 - \Omega_i\|^2$ para cada face conhecida no banco de imagens, com $i = 1, 2, \dots, M$.

A face do banco de imagens que possuir menor distância euclidiana com a face de teste permite fazer a conclusão de que são imagens bastante semelhantes ou iguais.

Outro método bastante utilizado para o reconhecimento facial é o *Haar Cascade* [Viola and Jones 2004]. Esse método é definido como uma estrutura contendo um encadeamento de classificadores do mais genérico ao mais específico [dos Reis 2014] e é baseado em aprendizagem de máquina, pois se utiliza de características previamente extraídas de um objeto para sua detecção. A função é treinada a partir de imagens positivas (imagens de faces) e imagens negativas (imagens sem faces). Utiliza retângulos como *features* (características) para localizar faces, de forma que possuam um valor único obtido subtraindo a soma dos *pixels* sob a região branca retangular da soma dos *pixels* sob a região preta retangular. Diferentes *features* podem representar um mesmo espaço na imagem, então são treinados diversos classificadores considerados fracos e depois combinados para gerar classificadores fortes, esses classificadores serão usados em sua cascata, a qual efetua uma busca por negativos. Ao serem encontrados, eles são descartados. Caso passe por todos os níveis da cascata, a região analisada possui uma face. A Figura 2 mostra alguns exemplos de *features*.

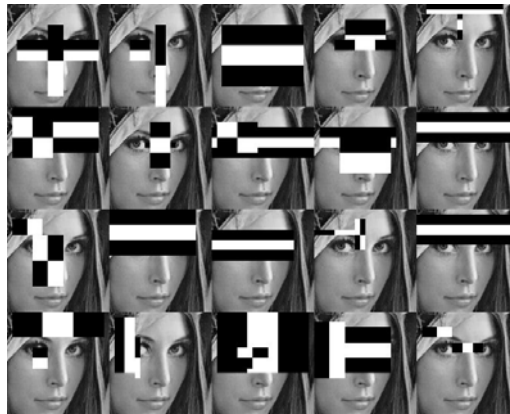


Figura 2. Modelos de *features* do *Haar Cascade*

As partes pretas e brancas das *features* representam a diferença de intensidade das cores entre as regiões da imagem, por exemplo: regiões mais escuras são representadas pela parte preta e as mais claras pela parte branca. Para o cálculo do *Haar Cascade*, precisa-se de uma imagem integral, isto é, uma matriz de mesma dimensão da imagem de entrada, onde (x, y) armazena a soma de todos os *pixels* do retângulo entre $(0, 0)$ e (x, y) . Sendo $a(x, y)$ a imagem de entrada, a imagem integral $a'(x, y)$ é calculada através da seguinte recorrência:

$$S(x, y) = S(x, y - 1) + a(x, y)$$

e

$$a'(x, y) = a'(x - 1, y) + S(x, y),$$

onde $S(x, y)$ é a soma acumulada na linha, $S(x, -1) = 0$ e $a'(-1, y) = 0$. Após o cálculo da imagem integral, o valor relativo a cada retângulo da *feature* pode ser obtido através de quatro acessos a memória, isto é, quatro pontos (x, y) de *pixels*. Suponhamos uma subregião retangular D da *feature*: ponto (x_1, y_1) superior à esquerda; ponto (x_2, y_2)

superior à direita; ponto (x_3, y_3) inferior à esquerda e ponto (x_4, y_4) inferior à direita. Calculamos essa região como segue:

$$D = a'(x_4, y_4) + a'(x_1, y_1) - (a'(x_2, y_2) + a'(x_3, y_3)).$$

Para utilizar as técnicas de visão computacional, tratamento e processamento da imagem da face, utilizamos de forma adaptada para o nosso sistema embarcado o OpenCV (*Open Source Computer Vision library*)³ que é um biblioteca que possui módulos de processamento de imagem e vídeo que contemplam as técnicas acima mencionadas. Além disso, o OpenCV possui suporte para as linguagens C/C++, JAVA, PYTHON e *Visual Basic*, o que facilita a portabilidade e difusão das soluções apresentadas.

Dentre os diversos trabalhos relacionados [Kail et al. 2007] e os citados no restante do trabalho sobre reconhecimento de faces temos o estudo apresentado por Hu et al. [Jun Hu et al. 2015] que compararam dois métodos, *Principal Component Analysis* (PCA) e *Two-Dimensional Principal Component Analysis* (2DPCA), utilizando quatro base de dados de faces famosas com o objetivo de descobrir qual apresentava melhores resultados. Outro trabalho interessante é o apresentado por Kamencay et al. [Kamencay et al. 2014] que apresenta uma metodologia utilizando um algoritmo de *Principal Component Analysis* (PCA) combinado com *Canonical Correlation Analysis* (CCA) para aprender o mapeamento entre uma imagem facial 2D e dados de face 3D.

3. Sistema para controle de acesso

O sistema aqui proposto considera o controle de acesso em áreas restritas em, por exemplo, prédios públicos, laboratórios, arquivos e até salas de reuniões. Em suma o controle de acesso aqui proposto, consiste da utilização integrada de uma placa Beaglebone, para o reconhecimento de faces, e uma placa Arduino, para a autenticação via RFID, decisão da abertura da fechadura magnética e interação com o indivíduo via display.

O sistema inicia quando o acesso ao ambiente é solicitado, ou seja, quando a câmera identificar um indivíduo se aproximando da sala. Em seguida a própria câmera, por intermédio do microcontrolador embarcado nela (Beaglebone), aciona o módulo de reconhecimento facial. Ao mesmo tempo, o indivíduo pode autenticar sua chave RFID por intermédio de um chaveiro. Assim que a face é reconhecida e a chave RFID autenticada, a porta é aberta e um registro de acesso é armazenado, informando a entrada do indivíduo e o horário em que o mesmo entrou. Em caso de acesso negado, o sistema informa, via *display*, que o acesso foi negado. O gerente do sistema possui uma interface de acesso para que ele insira novos usuários, ou configure acessos temporários e programados.

As duas autenticações ocorrem de forma paralela, uma vez que temos um microcontrolador para o reconhecimento facial e outro para a autenticação da TAG RFID. O sistema de autorização de acesso tem os seguintes passos: i. a câmera captura uma imagem digital, onde é efetuada uma busca por uma região onde possa ter uma face; ii. a região é selecionada e passa por um processo de extração de dados relevantes; iii. os dados extraídos serão comparados com uma base de dados a fim de fazer o reconhecimento; iv. retorna para a placa Arduino o resultado se a face é conhecida; v. o leitor RFID coleta os dados do chaveiro; vi. é verificado se a face reconhecida corresponde a chave RFID

³<http://opencv.org/>

lida; e vii. se o usuário estiver cadastrado a placa Arduino abre a fechadura magnética, senão mostra uma mensagem de acesso negado no *display*.

A detecção da face, realizada pela Beaglebone, é obtida pelo método *Haar Cascade*. Após a detecção, a área da imagem que contém a face é normalizada usando funções disponibilizadas pelo OpenCV ⁴, isto é, a face é alinhada e são removidas imperfeições causadas pelo meio que as imagens são adquiridas. Após normalizar as imagens, é então utilizada a técnica *Eigenface*. O objetivo da aplicação desta técnica é extrair e codificar as características relevantes para reconhecimento das faces. Com esses dados é possível cadastrar um novo indivíduo no sistema ou realizar o reconhecimento utilizando como comparativo informações presentes na base de imagens do sistema.

Para conceber o sistema proposto, utilizamos uma placa Arduino UNO para receber e processar todos os dados e tomar a decisão de acesso a área restrita. O Arduino UNO possui um micro-controlador baseado no ATmega328, possuindo 14 pinos de entradas e/ou saídas digitais, das quais 6 podem ser usadas como PWM, um cristal oscilador de 16MHz, uma conexão USB, 6 entradas analógicas, uma entrada de alimentação e uma conexão ICSP. Para a montagem do protótipo Arduino, temos: um leitor RFID, um *display* LCD 2X16, um trafo 220Vac 12Vac, um relé 5V, um fecho magnético e um potenciômetro 10k. O *display*, relé e o leitor de RFID são ligados as portas de entrada e saída do Arduino. O usuário utiliza um chaveiro RFID de médio alcance que permite que a TAG seja lida através de barreiras e objetos, não necessitando o contato direto. A Figure 3 mostra o protótipo Arduino desenvolvido.

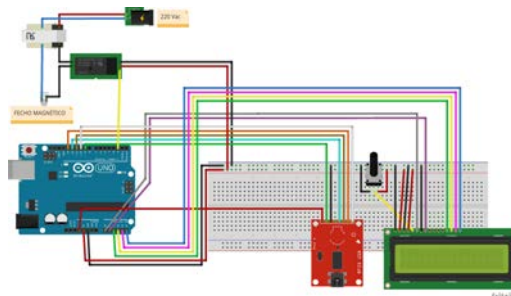


Figura 3. Protótipo do controle de acesso

Nesse protótipo, quando uma TAG é lida pelo leitor RFID o Arduino interpreta seus dados e verifica se a TAG está habilitada para acessar o ambiente, se estiver e a face foi reconhecida o Arduino aciona a porta que está ligado o relé, fazendo com que o fecho magnético receba um pulso elétrico e destrave a porta, assim dando acesso ao usuário.

A comunicação entre a Beaglebone e o Arduino é feita pela porta serial. O Arduino cria uma comunicação serial com o microcontrolador e através dessa porta serial o sistema envia a informação do reconhecimento da face. Essa comunicação pode ser feita de outras maneiras, por exemplo, via bluetooth ou transmissor de radio frequência. Como a Beaglebone possui um sistemas operacionais embarcado optamos por usar a comunicação serial, pois o tempo de envio e recebimento de informação é mais rápido. Ademais, ao utilizar a Beaglebone com o modulo de radio frequência algumas funções,

⁴<http://opencv.org/>

como a saída de vídeo, é desabilitada o que dificulta o seu uso integrado com uma câmera para o reconhecimento de face.

Para o processamento das imagens, utilizamos a Beaglebone Black que possui um processador ARM Cortex A8 1GHz, com 512M de memória RAM, saída de vídeo HDMI, 2GB de eMMC de memória flash, Kernel do Linux 3.8 e sistema operacional Debian 7.8.

Acoplada a Beaglebone realizamos testes com duas câmeras. A primeira delas, bem sofisticada, é a D-link – HD Wireless N Cube Network Camera – capaz de capturar imagens em alta resolução 1 megapixel HD-720P garantido a captura dos detalhes que necessitamos, usa uma compressão de vídeo H.264, MPEG-4 e MJPEG. A câmera possui a seguinte especificação: 1/4 1 megapixel progressiva CMOS sensores, iluminação mínima de 1.0 lux, zoom digital 10x, comprimento fixo 3.45 mm, abertura de F2.0, ângulo de visão de H 57,8°, V 37,8°, D 66°. A outra câmera, uma simples web Cam, é a Goldship hd, capaz de gerar imagem com qualidade HD e que possui a seguinte especificação: conexão com o computador via USB, classe de vídeo USB (UVC), resolução Máxima em HD 720p e taxa de quadros acima de 30fps.

4. Resultados e discussões

Para identificar o impacto do uso do reconhecimento facial de forma embarcada e considerando, em especial, o tempo de processamento das técnicas utilizadas realizamos experimentos com 10 faces, 5 cadastradas na base de imagens e 5 não cadastradas. A base continha 12 faces diferentes cadastradas: 100 fotos foram tiradas de cada face para efetuar o cadastro, resultando em um total de 1.200 imagens. Ressaltando que foram testes iniciais, nos trabalhos futuros será incluída um base de dados maior e os testes serão realizados com mais faces.

Os testes visam o desempenho da placa embarcada na detecção da face e no seu reconhecimento. Salientando que os testes não são para analisar o desempenho dos algoritmos de detecção e reconhecimento da face e, sim a placa embarcada. Buscar por algoritmo que possuam o melhor desempenho qualitativo esta fora do escopo desse trabalho.

Os testes foram efetuados em tempo real utilizando a câmera D-Link com e resolução 352×288 . As faces utilizadas nos testes foram postas em 3 posições: frontal, alta e lateral. A posição frontal refere-se à câmera posicionada de frente à face; na posição alta a câmera é posta em um local acima da face, porém o indivíduo deve olhar para a câmera (simula-se uma câmera na parede em nível mais alto que o indivíduo); e a posição lateral é semelhante à posição alta, mas forma um ângulo diferente de 90 graus com a face, escondendo parcialmente um dos lados da face. Para cada posição foram efetuadas 30 capturas de imagens e repetido o processo para 3 graus de confiança diferentes, 0.975, 0.980 e 0.985, resultando em 300 análises para cada grau, com um total de 900 análises.

A Tabela 1 exhibe o resultado com faces existentes no banco de imagens. Configurando o nível de confiança em 0.975, foram obtidos resultados muito bons para a posição frontal e alta, basta verificar a taxa de acerto, 94% e 74%, respectivamente. Já na posição lateral o reconhecimento correto ficou abaixo de 50% e mais próximo dos resultados errados e em 15.3% das imagens não foi obtido êxito na detecção da face.

Tabela 1. Indivíduos presentes na base.

Confiança	0.975	0.980	0.985	
Frontal	0.940	0.794	0.173	Certo
	0.060	0.180	0.053	Errado
	0	0.026	0.774	Desconhecido
Alta	0.740	0.507	0	Certo
	0.253	0	0	Errado
	0.007	0.493	1	Desconhecido
Lateral	0.467	0.300	0	Certo
	0.380	0.100	0	Errado
	0	0.520	0.967	Desconhecido

Tabela 2. Indivíduos não presentes na base.

Confiança	0.975	0.980	0.985	
Frontal	0	0.613	1	Certo
	1	0.387	0	Errado
	0	0	0	Desconhecido
Alta	0.060	0.947	0.987	Certo
	0.840	0.007	0	Errado
	0	0	0	Desconhecido
Lateral	0.007	0.920	0.980	Certo
	0.867	0	0	Errado
	0	0	0	Desconhecido

Aumentando o nível de confiança para 0.980 os resultados se mantiveram consistentes. Na posição alta, mesmo com reconhecimento correto um pouco acima de 50%, os resultados errados foram zerados, melhorando a eficiência do algoritmo. Com o nível de confiança em 0.985, só houve reconhecimento correto para a posição frontal; contudo, com taxa de acerto baixa e recebendo como saída muitas vezes “face desconhecida”.

A Tabela 2 mostra o resultado dos testes efetuados com faces não cadastradas. Para o nível de confiança 0.975, as saídas do reconhecimento facial foram em sua grande maioria errados; contudo, ao aumentar o nível de confiança, o retorno do algoritmo foi também em sua grande maioria correto.

Nas duas tabelas citadas foi observado que o resultado bom na primeira refletia em um resultado ruim na segunda. Por esse motivo a Tabela 3 foi criada, unificando os resultados das tabelas em uma só. Os valores para o nível de confiança 0.980 para a posição frontal e alta de forma geral ficaram bastante satisfatórios, alcançando um bom índice de acerto e baixos índices de erros. O resultado para a posição lateral ficou com uma boa porcentagem de acerto, o motivo é a alta taxa de acerto das faces não cadastradas, embora a taxa de acerto para faces cadastradas seja baixo.

Tabela 3. Indivíduos presentes e não presentes na base.

Confiança	0.975	0.980	0.985	
Frontal	0.470	0.703	0.587	Certo
	0.530	0.283	0.027	Errado
	0	0.013	0.386	Desconhecido
Alta	0.400	0.727	0.494	Certo
	0.547	0.003	0	Errado
	0.003	0.247	0.500	Desconhecido
Lateral	0.237	0.610	0.490	Certo
	0.623	0.050	0	Errado
	0	0.260	0.483	Desconhecido

Para comparar o tempo de resposta da placa, executamos em tempo real o módulo de reconhecimento facial também em um computador. O computador utilizado possui um sistema operacional linux ubuntu 14.04.LTS, memória RAM de 8G, processador intel Core i7-3770 CPU@3.40GHz x 8, placa gráfica AMD Radeon HD 6450 e disco rígido de 1T.

Os gráficos da Figura 4 ilustram o tempo de execução em ambos os casos, vale destacar que, com o objetivo melhorar a visualização, as escalas das figuras no eixo-y não são as mesmas. Pode-se ver que apenas quando o sistema esta executando sem ter nenhuma face a frente da câmera é que o processamento é muito alto. No entanto, nos

outros cenários o tempo de processamento ao se utilizar a placa se assemelha a de um servidor. Um fato importante é que estamos considerando a execução isolada de apenas uma requisição. No caso de termos várias requisições, certamente um maior poder computacional será requerido para o processamento no servidor. Vale salientar que nossa proposta é que o sistema seja desconectado a Internet, logo para utilizarmos um PC teríamos que ter uma máquina exclusiva para isso.

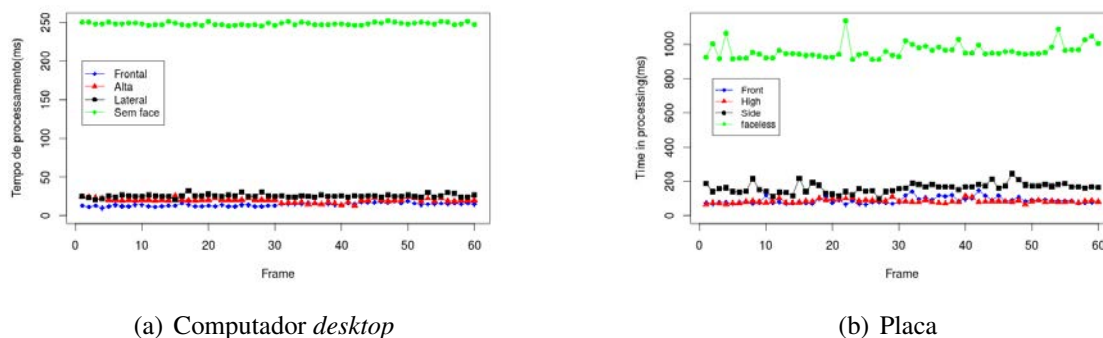


Figura 4. Tempo de resposta referente ao reconhecimento facial.

Uma vez que foi possível identificar o baixo impacto do uso do reconhecimento facial de forma embarcada realizamos testes para identificar a robustez a validade de nosso sistema completo, ou seja, considerando o reconhecimento de face e a autenticação via RFID simultaneamente.

Foram realizados testes com quatro pessoas, das quais três possuíam suas faces cadastradas no sistema de reconhecimento facial e uma não possuía, foi utilizado dois chaveiros RFIDs, um com acesso permitido e outro sem acesso.

Foi realizado quatro tipos de teste: i. faces cadastradas e o chaveiro RFID valido, nesse teste o sistema funcionou corretamente para todas as tentativas reconhecendo a face e validando com o chaveiro RFID assim liberando o acesso do usuário ao ambiente; ii. as faces cadastradas e o chaveiro RFID sem acesso permitido ao ambiente, neste caso sistema reconheceu as faces mas não liberou o acesso, pois o chaveiro não validou o acesso; iii. com a face não cadastrada e o chaveiro RFID valido, neste caso o sistema algumas vezes a face, gerando um falso positivo então o sistema liberava o acesso; iv. face não cadastrada e o chaveiro RFID sem acesso permitido, neste teste mesmo gerando algum falso positivo o acesso ao ambiente era negado pois a etiqueta não era valida para acessar o ambiente. Para a detecção e reconhecimento facial, foram utilizadas as técnicas *Eigenfaces* e *Haar Cascade* descritas na sessão 2 deste trabalho.

Os teste foram feitos tanto no PC como na placa, em ambos os casos, foram observados os mesmos resultados para os testes, só que com um tempo maior para fazer o reconhecimento facial na placa, pois o tempo de processamento e acesso a memória são superiores dos que observados no computador *desktop*.

5. Conclusão e trabalhos futuros

Este artigo apresentou uma proposta de um sistema embarcado o controle de acesso considerando o reconhecimento de faces e autenticação via RFID. O estudo aqui realizado

mostra a viabilidade do sistema embarcado, sem a dependência de uma central. Como o sistema é executado na própria câmera, sem acesso a Internet, aumentando a segurança do sistema. Mesmo com um atraso na detecção, quando comparado com uma execução num computador *desktop*, o sistema funcionou como esperando só liberando o acesso para pessoas autorizadas.

Para trabalhos futuros, pretendemos fazer o rastreamento de pessoas dentro do prédio. Planejamos inserir outros elementos de sensoriamento ao sistema, no caso, sensores de presença e temperatura, sensor de luminosidade e controlador das lâmpadas e ar-condicionado, para assim realizarmos o monitoramento completo do sala ou área considerada. Pretendemos também usar uma rede e protocolo de comunicação para que a base de dados de reconhecimento seja compartilhada entre várias áreas do prédio.

Referências

- Affandi, A., Awedh, M., Husain, M., and Alghamdi, A. (November 2013). Rfid and face recognition based security and access control system. *International Journal of Innovative Research in Science, Enginnering and Technology*.
- Dias, J. A. F. F., Rodrigues, J. J. P. C., and Zhou, L. (2014). Cooperation advances on vehicular communications: A survey. *Vehicular Communications*, 1(1):22 – 32.
- dos Reis, W. A. R. (2014). Detecção de sinais de trânsito através do método de classificação adaboost. *UNOPAR Científica Ciências Exatas e Tecnológicas*, 12(1).
- Han, G., Shu, L., Pathan, A. S. K., Rodrigues, J. J. P. C., and Mellouk, A. (2013). Wireless sensor networks based on environmental energy harvesting. *The International Journal of Distributed Sensor Networks*, 2013(816063):2.
- jun Hu, J., zheng Tan, G., Ogang Luan, F., and Libda, A. S. M. L. (2015). 2DPCA versus PCA for face recognition. *Journal of Central South University*, 22(5):1809–1816.
- Kail, K., Williams, C., and Kail, R. (2007). Access control system with rfid and biometric facial recognition. US Patent App. 11/790,385.
- Kamencay, P., Hudec, R., Benco, M., and Zachariasova, M. (2014). 2D-3D face recognition method based on a modified ccapca algorithm. *Int J Adv Robot Syst*, 11:36.
- Pellicer, S., Santa, G., Bleda, A. L., Maestre, R., Jara, A. J., and Skarmeta, A. G. (2013). A global perspective of smart cities: A survey. In *7th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*.
- Sadri, F. (2011). Ambient intelligence: A survey. *ACM Computing Surveys*, 43(4):36:1 – 36:66.
- Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86.
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.
- Weis, S. A. (2016). Rfid (radio frequency identification): Principles and applications.
- Weixiao, M., Ma, R., and Chen, H. H. (2014). Smart grid neighborhood area networks: A survey. *IEEE Network*, 28(1):24 – 32.

Smart Station: Um Sistema Pervasivo de Notificação em Paradas de Ônibus para Pessoas com Deficiência Visual

Lucas Agostini¹, Alex Braga¹, Julio Machado¹, Alysso Nogueira¹, Tatiana Tavares², Adenauer Yamin⁴

¹CDTec - Centro de Desenvolvimento Tecnológico - Universidade Federal de Pelotas (UFPel)
Pelotas - RS - Brasil

{lbagostini, abbraga, jmdsneto, anrodrigues, tati, adenauer}@inf.ufpel.edu.br

Abstract. *Vision is one of the most important channels for the individual's relationship with the world. Also vision is our main "guidance system". Visual impairment people do not have this "guidance system" as their main orientation. A natural form for adaptation is to take advantage of other senses, as hearing and touch, to provide an independent way to accomplish daily activities. An example is the mobility, especially for accessing urban public transportation. Thus, we propose a solution for notification at bus stops based on pervasive computing, using microprocessors and speakers, centered in visual impairment people.*

Resumo. *A visão é um dos canais mais importantes de relacionamento do indivíduo com o mundo, sendo nosso principal "sistema-guia" de orientação. As pessoas com deficiência visual não possuem esse "sistema-guia" como sua orientação principal. Uma forma de adaptação natural é se beneficiar de sentidos alternativos, como audição e tato, para oferecer uma forma independente de realização das atividades do cotidiano. Um exemplo é a mobilidade, especialmente no acesso ao transporte público urbano. Dessa forma, propõe-se uma solução para notificação em paradas de ônibus baseada em computação ubíqua, usando microprocessadores e caixas de som, voltado a pessoas com deficiência visual.*

1. Introdução

Segundo Garlan et al (2002) o recurso mais precioso em um sistema computacional já não é mais o processador, memória, disco ou rede, mas sim o recurso limitado da atenção humana. A atenção humana é um recurso escasso pois o usuário está muitas vezes focado em atividades do seu cotidiano como andar, dirigir ou falar enquanto acessa algum sistema computacional. Quando o sistema computacional não compete pela atenção do usuário o qual pode disfrutar da informação de forma transparente temos então um ambiente genuinamente pervasivo, ou um ambiente pervasivo *distraction-free*.

Nas atividades do nosso cotidiano que podem ser facilitadas por ambientes dessa natureza, destacamos o transporte público. Além disso, o transporte público é uma condição necessária para a concretização das relações econômicas e sociais mais

desafiadoras da nossa Sociedade e um dos desafios para promoção da mobilidade urbana. Para Vasconcellos (2001) existem várias visões da mobilidade, no entanto, essas visões partem do princípio que a mobilidade está relacionada com a capacidade e/ou habilidade de movimentar-se. No entendimento do Ministério das Cidades (2004): “a mobilidade está vinculada à qualidade de vida dos locais onde as pessoas moram e para onde se deslocam, devendo estar articulada com o plano de desenvolvimento da cidade e com a democratização dos espaços públicos conferindo prioridades às pessoas e não a veículos”. Assim, um plano de desenvolvimento urbano para mobilidade deve abordar aspectos que envolvam atributos de qualidade voltados à sustentabilidade, tais como: (i) minimizar o consumo de espaço e recursos naturais; (ii) racionalizar e gerenciar eficientemente os fluxos urbanos; (iii) proteger a saúde da população urbana; (iv) assegurar igualdade de acesso a recursos e serviços; (v) manter a diversidade social e cultural.

Dentre os atributos supracitados, destacamos a igualdade de acesso a recursos e serviços. No Brasil, o ônibus é o veículo mais utilizado para o transporte público de passageiros, no entanto, a acessibilidade ainda é muito restrita no transporte público, especialmente, para os usuários cegos ou de baixa visão. Quando estas pessoas precisam utilizar o transporte público, elas enfrentam grandes dificuldades, que, segundo depoimentos, são ocasionadas devido a falta de dispositivos adaptados nas paradas de ônibus. Segundo Vasconcellos (2001) essa deficiência do serviço de transporte público muitas vezes coloca seus usuários em situação de desconforto e sem autonomia para realizar sua atividade de movimentação urbana como qualquer outro passageiro. Segundo Henry (2014) um dos pontos principais a acessibilidade é o design de produtos capazes de tornar mais autônomas ou independentes pessoas que possuem algum tipo de deficiência, seja de modo direto (sem ajuda de ninguém) ou indireto (com a ajuda de alguém ou alguma coisa/dispositivo). Nesse sentido uma solução pervasiva que permita a igualdade de acesso a recursos e serviços no transporte urbano pode contribuir efetivamente para promoção da acessibilidade.

Neste artigo é apresentado o *Smart Station*, um sistema pervasivo de notificação capaz de alertar os usuários sobre o *status* de uma parada de ônibus utilizando avisos ou mensagens sonoras. Como contribuições da proposta, este sistema utiliza um solução de baixo custo e baixo consumo energético capaz de comunicar informações imediatas sobre os ônibus de uma parada. Nossa intenção é, então, oportunizar maior independência aos usuários com deficiência visual no usufruto do serviço de transporte coletivo.

2. Trabalhos Relacionados

O aplicativo CitaMobbi Acessibilidade foi desenvolvido por um usuário de ônibus cego e tem como finalidade avisar em tempo real informações para embarque e desembarque de passageiros. Este aplicativo é uma extensão do CitaMobbi, aplicativo mobile que oferece informações em tempo real para transporte público com ônibus em 15 cidades brasileiras. O CitaMobbi Acessibilidade utiliza mensagens de voz e sistema de vibração do celular. Luiz Eduardo Porto, desenvolvedor da versão acessível,

acrescenta: *“De casa já é possível acionar o aplicativo, programando até a parada final da viagem. E não precisa levar o celular na mão, virando alvo de ladrões, pois dá para perceber ele vibrando dentro do bolso do casaco quando o ônibus está chegando”* (Uehara, 2015).

Em (Willis, 2005) é apresentada uma solução que utiliza um sistema de grade RFID, como solução para o problema de acessibilidade de alunos cegos. As RFIDs seriam programadas com informações sobre o ambiente ao seu redor, permitindo assim que não seja necessário um sistema centralizado com os dados armazenados ou uma rede *wireless* para troca de informação. Esse sistema em grade de RFID permitirá também avanços na robótica para o conhecimento preciso de lugares. Para utilização desse sistema, cada usuário levará em seu sapato um leitor de RFID que se comunicará com o usuário pelo celular através de conexão Bluetooth. Para dar assistência a locomoção, o usuário usa um cinto que emite sinais sonoros e vibra, indicando se o mesmo está indo na direção desejada.

De acordo com Scavin (2016) para atender a lei de mobilidade urbana 12.587/12 uma empresa mineira criou um sistema para disponibilizar e equipar ônibus para a maior acessibilidade aos seus usuários cegos. Ao usuário coube adquirir e portar um controle remoto com o sistema embarcado. Essa empresa atua nas cidades de Jáu/SP Araucária/PR e Limeira/SP. No interior de São Paulo foi idealizado pelo grupo criar BUSALERT que tem como o objetivo um sistema que informa a distância, tempo estimado de chegada e o número de paradas até o ponto solicitado pelo usuário. O diferencial para a melhor acessibilidade é que o sistema funciona em dispositivos móveis, tanto por texto como por áudio, neste último caso, facilitando o manuseio ao deficiente visual.

Na mesma linha de desenvolvimento de software para acessibilidade, Bianchi (2014) apresenta o desenvolvimento de um sistema móvel para o controle de transporte coletivo, com o fim de auxiliar a utilização deste por pessoas com deficiência visual. O funcionamento do sistema conta com recursos disponíveis na plataforma Android: como a utilização de GPS, conectividade entre dispositivos e o servidor de dados, sintetização de voz e mapas da plataforma. A execução do sistema é basicamente um aplicativo localizador, em que o aplicativo cliente envia e recebe informações do servidor, executando suas funcionalidades de forma a auxiliar os usuários no transporte coletivo com ou sem deficiência visual. Em síntese o sistema desenvolvido permite que o usuário saiba quando está se aproximando do seu destino através de um aviso sonoro e também indicia o ponto de parada mais próximo além de permitir o recurso de acessibilidade por reconhecimento de voz para que o deficiente visual possa ativar e interagir com o sistema. O intuito do sistema desenvolvido é facilitar o acesso e a autonomia para locomoção dos portadores de deficiência visual.

A Tab. 1 apresenta um resumo dos trabalhos supracitados em função dos aspectos: forma de interação, tipo de informação e dispositivos usuais ou dedicados (hardware específico). Em dispositivos distinguimos os dispositivos ligados ao

oferecimento da infraestrutura e os necessários para que o usuário tenha acesso ao serviço ou aplicação.

Tabela 1. Comparativo das Soluções Correlatas.

	Forma de Interação	Tipo de Informação	Dispositivos	
			Infraestrutura	Usuário
Portal do CittaMobi (2016)	Mensagens de voz e Vibração	Embarque e desembarque	Servidor de Aplicação Web	Dispositivos Móveis
Willis et al. (2005)	Sinais Sonoros e Vibração	Ambiente ao seu redor	Dispositivos Móveis + RFID	Leitor de RFID (sapato) + Dispositivos Móveis
Scavin (2016)	Áudio	Distância, tempo e o número de paradas de ônibus	Sistema Embarcado	Controle Remoto + Sistema Embarcado + Dispositivos Móveis
Bianchi (2014)	Síntese e reconhecimento de voz	Ponto de parada mais próximo	Servidor de Aplicação Web	Dispositivos Móveis

2. Smart Station: Da concepção ao Desenvolvimento

Segundo Rocha et. al (2013) sistemas de apoio ao cidadão são de considerável importância. Nesse cenário, a computação ubíqua e ciente de contexto, que se caracteriza pela presença de dispositivos portáteis, aparece como peça fundamental. Nogueira et. al (2012) acrescenta que visando a melhoria desse cenário, o conceito de computação pervasiva tem sido apontado como promissor, pela integração da computação com o cotidiano focada na melhoria da qualidade de vida, saúde pessoal e capacitação de usuário. Aliado a isso, ter em mente o requisito de não colocar na responsabilidade dos usuários usar ou carregar ou iniciar um aplicativo ou dispositivo para acesso ao serviço ou informação nos leva à concepção do *Smart Station*.

O *Smart Station* é focado no acesso do usuário com deficiência visual ao transporte público. Diferentemente de soluções baseadas em dispositivos móveis (Portal do CittaMobi (2016), Bianchi (2014)), nosso intuito é prover uma solução que não necessita de gatilho ou dispositivo para sua ativação. A ideia é bastante intuitiva: o usuário que está na parada do ônibus escuta áudio que identifica o ônibus que se aproxima da zona de embarque, ou o usuário que está dentro do ônibus escuta o áudio que identifica a próxima parada. Para tanto, dotamos o ônibus e a parada de microprocessadores ESP8266 e comunicação Wi-Fi conforme ilustrado na Fig.01.

2.1. Metodologia de Desenvolvimento

Como dito anteriormente a abordagem "*distraction free*" imprime um ritmo de desenvolvimento focado no usuário. Este fato também está presente no conceito de

experiência de usuário (do inglês: "*user experience ou UX*") e foi aplicado neste trabalho como uma forma mais ampla de analisar a relação usuário-dispositivo-sistema através da compreensão da interação na perspectiva do usuário. Arhippainen (2003) caracteriza *UX* pela experiência que o usuário obtém quando interage com um produto. Segundo o autor a *UX* é definida como as emoções e expectativas do usuário e sua relação com outras pessoas e com o contexto de uso, onde destacam-se aspectos afetivos e experiências pessoais.



Figura 1. Visão Geral da Solução.

A Fig.2 apresenta as principais atividades da metodologia de desenvolvimento utilizada no *Smart Station*. A definição do problema compreende o entendimento do cenário de uso e de soluções correlatas. E, então, é necessário definir o público-alvo através da identificação e descrição das personas. O refinamento dos requisitos nos permite compreender as funcionalidades que neste caso contou a participação dos usuários reais. O design da interação consiste na elaboração de um modelo de execução do sistema como também os protótipos necessários à sua validação. Por fim, o produto final é avaliado. Para a avaliação da interface de áudio utilizamos o sistema MOS (*Mean Opinion Score*) que dita regras de como avaliar um arquivo de áudio (ou vídeo) através da percepção dos usuários.



Figura 2. Metodologia de Desenvolvimento Baseada em UX.

2.2. Personas do *Smart Station*

A definição e descrição das personas foi essencial no entendimento da dinâmica de uso do sistema, especialmente, na compreensão da percepção dos elementos físicos responsáveis pela prover a ubiquidade da solução. Nesta etapa foi delineado dois perfis de usuário como principais personas desejáveis:

- Pessoa com deficiência visual que deseja utilizar o transporte público.
- Pessoa que não sabe ler, e que deseja saber o ônibus que está passando.

O entendimento da persona - pessoa com deficiência visual - necessitou de uma aproximação com público-alvo para que pudessemos entender melhor a situação atual e projetar uma solução para a situação desejada (com uso do sistema proposto). Dessa forma, os usuários reais do sistema participaram ativamente do processo de

desenvolvimento, desde a etapa de entendimento do problema até a etapa de avaliação do produto final. Na Fig.3 (a) é ilustrado um *storyboard* que representa situações do cotidiano de pessoas com deficiência visual que utilizam transporte público, onde observa-se a total dependência de uma terceira pessoa (motorista ou passageiro) para que a informação de qual ônibus esteja acessível para esses usuários. Já a Fig.3 (b) representa o cenário desejado, onde uma solução pervasiva integrada ao ônibus e a parada de ônibus é capaz de notificar ao usuário com deficiência visual qual o ônibus está chegando na parada.

É interessante destacar, que na análise de requisitos foram levantadas outras funcionalidades como tempo de espera ou frequência de parada, no entanto, na reunião com os usuário a funcionalidade desejada foi: "Eu quero saber qual o ônibus está se aproximando da parada!".



Figura 3. Storyboard com detalhamento das personas.

2.3. Smart Station: Tecnologias Utilizadas e Prototipação

A implementação do *Smart Station* foi projetada a partir de uma arquitetura estilo cliente-servidor conforme ilustra a Fig.4. O módulo servidor - *Server* - pode ser detalhado em quatro componentes: *AudioFile*, *Receiver*, *IPGenerator* e *Sender*. O módulo cliente - *Client* - possui três componentes: *AudioFile*, *Receiver* e *Sender*. A conexão entre os módulos *Server* e *Client* é estabelecida via UDP.

Os módulos servidor e cliente são executados em microprocessadores ESP8266. O microprocessador ESP8266 (Espressif Systems, 2015) possui uma interface de comunicação USB e um módulo de comunicação Wi-Fi, e com este módulo ela consegue fazer comunicação ad-hoc (sem infraestrutura adicional). A comunicação entre duas placas é feita através da criação de um servidor (com uma placa) e um cliente (com a outra placa). Quando elas se aproximam (segundo o *datasheet* a distância é de 50m) a conexão entre elas é realizada e as informações trocadas. As informações trocadas pelas placas identificam os áudios, ou seja, a placa que esta no ônibus diz a linha do ônibus na qual ela se encontra e a placa que está na parada informa ao ônibus sua identificação.

Na Fig. 4 temos um diagrama de componentes que ilustra a arquitetura principal do *Smart Station*. O componente *AudioFile* designa os arquivos de áudio, cada parada de ônibus tem um arquivo de áudio com seu nome, e cada ônibus possui um arquivo com o nome da linha. O componente *Sender* é responsável por enviar o *AudioFile* para o outro ponto da conexão. O componente *Receiver* é responsável por enviar o áudio

que está na ESP8266 via conexão sem fio. O componente *IPGenerator* fica na parte do servidor e é responsável por determinar IPs para os clientes. Os componentes *AudioFile*, *Receiver*, *IPGenerator* e *Sender* descrevem o servidor. Os componentes *AudioFile*, *Receiver* e *Sender* descrevem o cliente. Com isso, os ônibus possuem uma ESP8266 que funciona como servidor. Enquanto nas paradas dos ônibus temos outra ESP8266 funcionando como cliente.

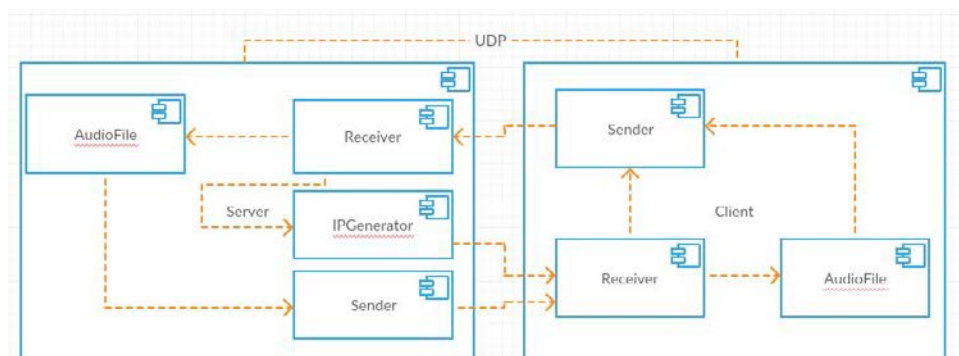


Figura 4. Arquitetura do Smart Station.

Os primeiros testes realizados com protótipos funcionais foram essenciais para conhecer as fronteiras da tecnologia adotada. A primeira preocupação foi entender a relação entre tempo de resposta e tempo de reação dos usuários. Lembrando que a intenção é oferecer um sistema de notificação que além de informar aos usuários qual o ônibus que se aproxima ofereça tempo para que o usuário possa tomar a decisão de solicitar ou não a parada do veículo.

Para execução dessa primeira bateria de testes foi utilizada: (i) para representar o ônibus: a placa conectada a entrada de USB de um veículo de passeio; (ii) para representar a parada: uma placa conectada a um notebook em uma parada de ônibus. Através desses testes foi constatado que: (i) o tempo médio de resposta do sistema é cerca de 3,5 segundos; (ii) a distância de conexão sem o uso de antenas externas é de aproximadamente 40 metros; (iii) a velocidade de deslocamento do veículo não interfere no tempo de resposta.

2.4. Avaliação

Como dito anteriormente o envolvimento com usuários reais é uma característica deste projeto. Para avaliação dos protótipos foi utilizada a metodologia MOS (Rec, I. T. U. T., 2006) procurando identificar clareza, velocidade, volume e o tempo de resposta do sistema. Na escala de avaliação percebida sugerida pela metodologia MOS são apresentados para os voluntários seis opções para medição do grau de satisfação. Para realização dos testes para avaliar clareza, velocidade e volume foram envolvidos 23 voluntários. Para realização dos testes de tempo de resposta foram envolvidos 06 voluntários.

O resultado referente a clareza do áudio percebido pelos voluntários ouvintes obteve uma classificação unânime de *perfeito*. A Fig.5 (a) denota a avaliação de velocidade e sua relação com a compreensão dos voluntários ouvintes. Podemos

observar que a maioria (>75%) atribuiu conceito *perfeito*. Foram testados nomes de linhas simples (por exemplo: Pycrespo) e nomes compostos (por exemplo: Santa Terezinha).

A Fig.5 (b) mostra a avaliação do volume a qual também obteve maioria de respostas relacionadas ao conceito *perfeito*. Esse primeiro conjunto de testes foi realizado em uma reunião do grupo de cegos da Escola Luís Braille. Para se aproximar do contexto real optou-se por fazer uma avaliação coletiva e sem solicitar silêncio na sala, procurando assim garantir algum ruído. A distância máxima de um voluntário ao ponto de emissão do áudio foi de aproximadamente 5 metros e a distância mínima 1.

Adicionalmente foi avaliado o tempo de resposta de seis usuários em ambiente real, isto é, na parada do ônibus e os resultados obtidos podem ser visualizados na Fig.5 (c). Após o teste coletivo (Fig.5 (a) e (b)), convidamos usuários voluntários para simular o ambiente da parada de ônibus e cronometramos o tempo de tomada de decisão de cada usuário. Entende-se por tomada de decisão fazer sinal para que o ônibus possar parar. Percebemos que o tempo médio foi de 0,6 segundos e que não ultrapassamos 1,2 segundos.

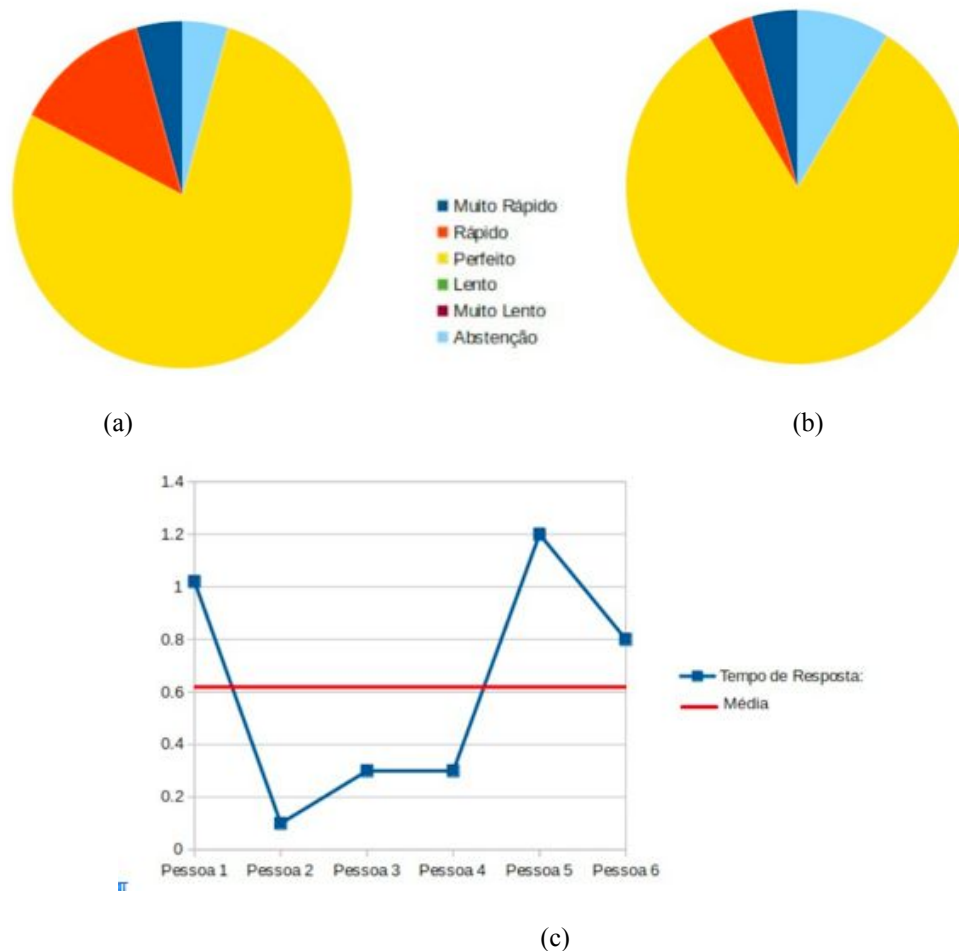


Figura 5. Resultados.

3. Consideração Finais

Este trabalho apresentou uma experiência de desenvolvimento de uma solução pautada em dois conceitos desafiadores e complementares: tecnologia assistiva e computação pervasiva. Enquanto para o desenvolvimento voltado à tecnologia assistiva a autonomia é uma premissa, para a computação pervasiva exigir menos comprometimento da atenção do usuário, ou seja, atribuir mais autonomia, é mais.

Dentro dessa perspectiva, a Tab.2 apresenta a análise do *Smart Station* segundo os mesmos aspectos com os quais analisamos os trabalhos relacionados (vide Tab.1). A diferença essencial do *Smart Station* em relação aos demais é não necessitar que o usuário final vista, use ou acesse nenhum dispositivo ou aplicativo. Essa característica é fundamental para a autonomia percebida pelos usuários finais segundo a perspectiva de tecnologia assistiva. A mesma característica é o que torna a solução *distraction-free*, em outras palavras, genuinamente pervasiva uma vez que não implica na adição de nenhuma atividade de gatilho ou manipulação de aplicativo ou dispositivo por parte do usuário no ambiente de execução.

Tabela 2. Análise da Solução *Smart Station*.

<i>Smart Station</i>	Forma de Interação	Tipo de Informação	Dispositivos	
			Infraestrutura	Usuário
	Áudio (mensagens sonoras)	Próximo ônibus e próxima parada	Servidor + Cliente + placas + auto-falante	Não necessita

Dessa forma, o resultado alcançado neste trabalho corrobora com a visão de Vanderheiden (2008), que destaca que o objetivo de uma solução em tecnologia assistiva é habilitar qualquer um, em qualquer lugar, para ser capaz de utilizar os serviços que necessitam em qualquer computador/dispositivo que encontrarem. Objetivo inerente a soluções em computação pervasiva. Portanto, a tecnologia assistiva se beneficia da computação pervasiva no intuito de oferecer serviços cada vez mais transparentes e integrados ao cotidiano dos usuários.

8. Agradecimentos

Nosso agradecimento especial aos alunos, professores e colaboradores da Escola Especial Louis Braille pela sua contribuição essencial para realização deste projeto.

Referências

- Henry, Shawn Lawton, Shadi Abou-Zahra, and Judy Brewer. "The role of accessibility in a universal web." Proceedings of the 11th Web for All Conference. ACM, 2014.
- Vasconcelos, Eduardo de Alcântara. Transporte Urbano, espaço e Equidade: análise das políticas públicas, São Paulo: Annablume, 2001.
- Nogueira, Carolina, Hyggo Almeida, and Angelo Perkusich. "Arcabouço para o Desenvolvimento de Aplicações de Monitoramento Remoto e Auxílio de Pessoas com Doença de Alzheimer." In: Anais do Simpósio Brasileiro de Computação Proceedings of the SBCUP'2016 (Julho 2016)

- Ubíqua e Pervasiva (SBCUP), Congresso da Sociedade Brasileira de Computação (CSBC). 2012.
- Pansanato, Luciano TE, Christiane E. Silva, and Luzia Rodrigues. "Uma Experiência de Inclusão de Estudante Cego na Educação Superior em Computação." XX Workshop sobre Educação em Computação. 2012.
- Rocha, A., et al. "YouOnAlert: Um Sistema para Alertar Cidadãos Comuns Acerca de Problemas do Cotidiano das Cidades." In: Anais do V Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP), Maceió-Brazil, 2013.
- Willis, S., & Helal, S. (2005, October). RFID information grid and wearable computing solution to the problem of wayfinding for the blind user in a campus environment. In IEEE International Symposium on Wearable Computers (ISWC 05).
- Bianchi, Evaldo Augusto. "Sistema para controle de frotas do transporte coletivo com acessibilidade para deficientes visuais." (2014) Disponível em <http://repositorio.roca.utfpr.edu.br/jspui/handle/1/2002>.
- Espresif Systems. "ESP8266EX Datasheet Version 4.3." (2015) Disponível em http://www.esp8266.com/wiki/lib/exe/fetch.php?media=0a-esp8266_datasheet_en_v4.3.pdf.
- Arhippainen, L. Capturing user experience for product design. IRIS26, 26º Seminário de investigação de sistemas de informação na Escandinávia, Porvoo, Finlândia 2003. Disponível em: <http://www.vtt.fi/virtual/adamos/material/arhippa2.pdf>
- Uehara, Fernanda. App ajuda deficiente visual a andar de ônibus: Desenvolvido por cego, CittaMobi Acessibilidade usa comando de voz para embarque de passageiro. Disponível em: <http://www.diariosp.com.br/noticia/detalhe/85852/app-ajuda-deficiente-visual-a-andar-de-onibus> Data de Publicação: 05/09/2015 Acessado em: 05/03/2016.
- Portal do CittaMobi. Disponível em: <http://www.cittamobi.com.br/>. Acessado em: 05/03/2016.
- Scavasin, Flávio. "Aplicativos para Segurança e Acessibilidade no Transporte Público" *In: Caderno de Cidades - Acessibilidade 2016 - Revista Nacional de Reabilitação* Edição número 90, 2016.
- Bianchi, E. A. (2014). Sistema para controle de frotas do transporte coletivo com acessibilidade para deficientes visuais. (Monografia de Conclusão de Curso).
- Rec, I. T. U. T. (2006). P. 800.1, Mean opinion score (MOS) terminology. International Telecommunication Union, Geneva.
- D. Garlan, D. P. Siewiorek, A. Smailagic and P. Steenkiste, "Project Aura: toward distraction-free pervasive computing," in IEEE Pervasive Computing, vol. 1, no. 2, pp. 22-31, April-June 2002.
- Vanderheiden, Gregg C. "Ubiquitous accessibility, common technology core, and micro assistive technology: Commentary on "computers and people with disabilities"." ACM Transactions on Accessible Computing (TACCESS) 1.2 (2008): 10.

The classification of algorithms for Privacy Management in Ubiquitous Environments

Valderi R. Q. Leithardt^{1,2}, Carlos O. Rolim¹, Anubis G. M. Rossetto¹,
Guilherme A. Borges¹, Jorge Sá Silva², Claudio F. R. Geyer¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

² Universidade de Coimbra, Sistemas Largamente Distribuídos, Departamento de
Engenharia Informática (DEI) _ Pólo II - Coimbra - Portugal

{vrqleithardt, carlos.oberdan, agmrossetto, gaborges,
geyer}@inf.ufrgs.br, sasilva@dei.uc.pt

***Abstract.** Control and data management in ubiquitous environments is not a trivial activity owing to the heterogeneity of the users, applications and devices, required to exchange information. However, various problems have been found in the literature with regard to privacy information and related to the data used in ubiquitous environments. This paper offers a solution by means of statistical classification algorithms that can be used for control and privacy management. On the basis of the algorithms used in the tests, it proved to be possible to control and manage information by providing definitions of the variables and parameters for users, devices, and ubiquitous environments.*

1. Introduction

In the last ten years, advances in mobile communication technologies have led to a change in the computing paradigm. The traditional model is static and relatively predictable with based workstations, and has created a highly dynamic environment with constant changes caused by user mobility. This feature is enhanced by the use of multifunctional mobile devices such as cell phones and smartphones, as well as educational environments such as interactions Teleduc, Moodle, etc. This change can be seen as another stage towards the concept of Ubiquitous Computing (Ubiquitous/Pervasive Computing) introduced by Mark Weiser (1991).

We are now living in an inter connected society, with e-mails, cell phones, palms, chats, information search engines, news sites, online communities, SMS, VoIP and other tools that until recently were not part of our daily routine either at work or leisure. According to Abech et al. (2012), the popularity of mobile devices to access the internet makes it feasible to obtain educational content regardless of time or place. In this new scenario of technological changes, there are new challenges and new forms of relationships that affect human behavior and hence all social factors involving education. Among these challenges is the question of privacy control which is of great importance since data and a shared location are unavailable without prior knowledge and authorization. This is a considerable problem given the increasing ease of access to computing resources. This kind of information can be best managed by the ubiquitous environment.

Thus it is also necessary to have a control of privacy, since the user may not need or want to locate or share his/her data at all times. The shared information can be best managed by the pervasive or ubiquitous environment since this is a means, for example, of reducing unnecessary data processing, and increasing the level of security and management of services. According to Cristiano Andre da Costa et al. (2008), the popularity of mobile devices to access the internet makes it feasible to obtain educational content regardless of time or place. In this new scenario of technological changes, there are new challenges and new forms of relationships that affect human behavior and hence all social factors involving education.

Thus, a model of ubiquitous privacy control is needed that meets as many requirements as possible related to the physical and virtual environment. In the literature, several studies can be found that address the question of privacy control and these are aimed at the user and the devices, services or communications employed. With regard to these features, this paper seeks to make a proposal for a privacy model that is designed for a ubiquitous academic environment, related to educational computing in an ubiquitous real world. Security issues will not be addressed in this ubiquitous computing, as there are already techniques to prevent attacks or disclosure of encryption information. Nor will this study address the question of the restrictive controls of users and devices, or its services and forms of communication.

The main concern of this work is the privacy model proposed for ubiquitous environments that provides definitions of parameters and criteria for an individual environment. Appropriate data classification algorithms were used for the control and privacy management environment and these were based on rules.

2. Related Word

In ubiquitous environments there are many problems and issues that need to be discussed, in particular, the control and management of privacy. According to Warren and Brandeis (1890), privacy is intrinsically linked to the perception of each individual about what it represents, such as a threat to their personal property or physical or moral integrity. Thus, it can be inferred that the privacy setting is something very abstract and subjective, and takes account of the diverse needs of each individual. These needs are not homogenous and may depend on cultural areas such as religion, tradition, customs, education or politics, or more subjective concerns such as user privacy or everyday factors such as age, health status, job responsibilities, mood, and leisure activities.

According to Cristiano et al. (2008) data that characterize a context may range from the physical world to the virtual world, and sometimes the two are merged. People often do not think of physical environments (e.g. an office, shop floor, stadium, and classroom) and virtual environments (e.g. a desktop computer, or the features of a mobile phone) as separate areas [Rodrigues et al. 2006]. Thus, the problem becomes even greater owing to the interaction between devices, users, applications, and communications between these environments. The work of involves the transfer of the control of music files based on the location of WIFI points. This author offers a different interpretation to what is given in this study, he seeks to define the types and sizes of information that must be transferred with regard to section and location. However, in our view, the question does not concern the environment itself but the point

of access to it. Thus, it is necessary to couple several other systems to supply information about the ubiquitous environment, and hence how it should proceed.

Henricksen et al. (2005) describe the hierarchy of control based on facts and user preferences. They also describe a context model of the application that controls the facts and individual occurrences, by seeking information on various ubiquitous sources; thus, it is not the privacy control in the ubiquitous environment. The work described by Lachello and Hong (2007), conducts a survey of several privacy issues addressed in the context of human computer interaction (HCI); the work also provides an overview of several points that should be tackled such as trends in the field and research being carried out. The main contribution of this work is that it addresses several key issues including the protection of the pervasive environment. In Bardram, Kjaer and Pedersen (2003) an authentication based solution is provided that is based on several examples of communication such as Radio Frequency Identification (RFID), and offers a single mechanism to manage different authentication protocols in ubiquitous environments. Despite carrying out authentication iterations with the pervasive system, there is not a change of perspective in the environment, nor any attempt to address issues related to individual privacy.

The work of Santarosa and Comfort and Basso (2010) seeks to support digital social inclusion in the technological dimension including principles of accessibility this exposes many points of weakness in the virtual learning environments, in particular the privacy control system in the devices used. In Tao and Peiran (2010) there is a research inquiry into the protection of data for individual transactions between users and "things" (Internet of things, IOT), based on the use of cards, tags and other devices used every day. It also outlines some specific situations in which IOT is used with categories and applications where concepts are defined in terms of a specific situation: for example, medical treatment is defined as private identification, but only based on the user's location and restricted to the pervasive goal. In the work carried out by Gotthard and Zorzo (2007), there is an examination of the technologies that assist in the process of teaching and learning which are being discussed in various fields of knowledge where the issue of privacy is handled by a user agent. On the basis of these research studies, a comparison is made in Table 1, where the existing solutions are displayed. The left column has abbreviations and the number of references that are cited. The first line of this comparative table outlines the approaches required for privacy management in pervasive environments and these were set up on the basis of the literature. The following definitions are used:

- > Addresses (A): the work deals with the Question addressed;
- > It does not address (NA): The work does not deal with the Question addressed;
- > Describes not (ND): information not found to address the Question;
- > Developing (D): The item is still being developed; it is often pointed out in tests, validations, obtained results or future work.

Several studies describe the particular solution that applies to a pervasive or ubiquitous environment, but do not state clearly how to go about the management and control of ubiquitous environments. In the next section, there will be an examination of the application scenario where the privacy management model in ubiquitous

environments is applicable. Based on the comparison with and related work, we define a privacy model for ubiquitous environments that will be outlined in Section 3.

Table 1. Comparative State-of-the-art Privacy Management in Ubiquitous Environments.

Approach solution	User	Device	Applications	Services	Communication	Environment	Privacy
Leithardt et al 2010	A	NA	A	D	A	D	NA
Rodrigues 2006	A	NA	A	A	A	NA	NA
Görlach, Heinemann e Terpstra 2005	A	A	A	NA	D	NA	NA
Kalempa 2009	D	NA	A	A	A	ND	D
Henricken et al 2005	D	D	A	A	D	NA	NA
Iachello e Hong 2007	NA	ND	D	D	A	NA	NA
Bardram, Kjaer e Pedersen 2003	A	D	A	A	A	NA	NA
Santarosa, Conforto e Basso 2010	A	D	D	NA	NA	D	NA
Tao e Peiran 2010	D	A	D	A	D	NA	D
Gotardo e Zorzo 2007	A	D	A	A	D	NA	D
Proposed model	A	A	A	A	A	A	A

3. Proposed Model

The use of a privacy management mechanism for ubiquitous environments that meets the requirements of a given scenario has been identified as a problem. Together with this, there is a need to collect personal information to operate these systems within ethical and legal constraints, because the privacy of people is involved. A proposed model for privacy management for ubiquitous environments that complies with these requirements is illustrated in Figure 1.

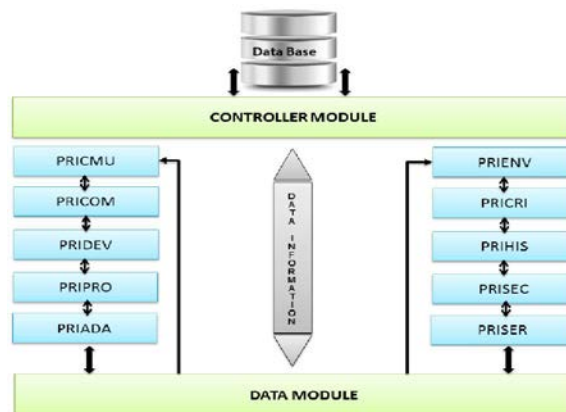


Figure 1. Model for a Privacy Manager adapted from [Leithardt et al., 2013].

The proposed privacy manager model consists of several components that are necessary to control ubiquitous environments. They can be individually described as follows:

Data Base: Rules for Information Storage and settings for users, devices, and the ubiquitous communication environment.

Controller module: the purpose of this is to receive access requests and perform the database control of the tables directly, with necessary information in accordance with the requests and access settings by maintaining control of the ubiquitous environment.

Data Module: In this module, calculations will be made of all the variables and parameters received from the other modules.

PRICRI: this module contains the rules and definitions of criteria and environmental settings such as access, use, sharing, location and other variables that can be manipulated or replaced in accordance with a) the environment settings and b) established criteria and standards;

PRICMU: Module management and privacy control of user information, which will handle definitions of related features for individual user preferences such as: temperature, light, authorized shares (such as information that someone wishes to share with other users and with his/her own environment), location data and other user preferences.

PRIDEV: management module and privacy control devices. This module will handle the data on the devices if these devices are in the environment itself and will then interact with it.

PRICOM: management module and communications privacy control. This concerns which forms of communication will be employed within the ubiquitous environment and how they will be used.

PRIADA: management and adjustment control module. This module will handle the information related to the adaptation of software and hardware in ubiquitous environments.

PRISER: environmental services management module. This module is responsible for the availability of services that can be used individually in each environment such as, shared information from other environments, devices, communications, location of users, environmental availability and its components that interact with users.

PRIHIS: this module will store and handle information on the historical user, environment, devices and other variables that may include other factors depending on the context.

PRIPRO: this module will carry out transactions of controls related to the user profile management.

PRISEC: this module will carry out the controls and management with regard to the safety of both the user and environment.

PRIENV: this module will register the attributes related to the environment.

All the modules operate independently and have their own characteristics and features that may vary according to the rules that have been previously established registered and enforced. Once these rules have been set out, each module sets its parameters based on the settings of the previous module. Thus, it is possible to have multiple environments with different rules and definitions for the same ubiquitous environments and the same user can use one or more different environments each with a

defined criterion. This can change depending on the device used in the communication as well as other factors that will be calculated on the data module. A preliminary strategy to address this need is to use algorithms for processing and classification of data used to manage the criteria set out in privacy management model. Therefore, some algorithms used in the literature were used and compared with the initial results of the tests described in the next section.

4. Prototype Testing and Preliminary Results

The implemented prototype, illustrated in Figure 2 shows the scenario used where the privacy settings server (4), (which acts as an authority for the mobile devices (2) and environments), receives the inferred symbolic locations through physical locations (1) and the mobile devices that the users upload. Each mobile device belongs to one user at a time while in use, and the same device is used by more than one user at different times. On receiving the symbolic location for a medium processed by the user's mobile location, the server passes by the model described in the previous section, and checks the changes that need to be adapted/ suited to the new context of the user with regard to environmental privacy and directs them via communication (3).



Figure 2. Architecture implemented

The server and the model are implemented with the aid of the Java EE programming language, which currently supports communications WebService Rest, Google Cloud Messaging (GCM) and Serial Communication. A PostgreSQL relational database was also employed. The mobile clients have been trained for the Android platform, and their native APIs were used for access to information about location, the prototype currently:

- > Makes environmental changes using different locations such as GPS and NFC Tags;
- > Identifies what types of access each user profile possesses when entering an environment;
- > Enables or disables the smartphone functionality in accordance with the privacy required by the environment.

One problem was how to classify the degree of access that a user has to the environment. This problem was solved by identifying the variables that provide the level of Access required by the user to the environment, which are as follows: Profile and user frequency in the environment, an environment, weekday, shift and working Day.

The first three profiles (Unknown, Transient and User) are automatically identified and allowed to proceed through the system (evolutionary profiles), while the last three (Responsible, Student and Manager) are assigned manually by the environmental manager. This configuration is necessary because a large number of users will probably not access all the environments known by the system. In this way, the system itself can distinguish between ordinary and new users while it is running, thus dispensing with the settings of the system manager. By contrast, ordinary users will not be able to access all the resources, especially in private settings, or automatically be allowed to proceed for security reasons. For example, the customer in a cafeteria might not be able to access the box, as such, since there are profiles that have to be assigned to the user manually.

The rules for the progression of evolutionary profiles are configurable in the system. In this study the frequency (F) of the user (u) is used to determine when it should evolve into a profile (R) before it can advance through the environment or be resolved as a less permissive profile. In this case, for each environment, the implemented rules define, (a) the location of the lower frequency ranges (I) and upper (S) at which a change of profile should be made. If the environment has no evolutionary profiles, the equation below expresses the progression rule that is implemented.

$$P_{u,a} = \{P_{u,a} + 1, \text{if } F_{u,a} > S_a; P_{u,a} - 1, \text{if } F_{u,a} < I_a; P_{u,a}, \text{otherwise}\} \forall P \in [1, n]$$

In the case of non-evolutionary profiles, the frequency is also used to increase or reduce the level of access to users. However, the user profile remains the same and only the type of access is changed. In both cases, it was assumed that the frequency can take on three distinct levels: frequent, normal and infrequent.

Three types of environment were also taken into account: restricted, private and public; it is assumed that the binary value is true for weekdays, and false, for weekends; There were two shifts, day and night; Finally, the variable working day indicates whether a day is useful or not with regard to the location based on the day of the week or holidays since there are no days which have working hours for certain environments. The combination of all the variables that are possible in the case scenario studied resulted in a rating with 383 possibilities. In the second instance, after being identified, the variables were assigned to the combinations of the following types of access: Locked, Guest, Basic, Advanced and Administrative. These data were used for training and testing in seven different classification algorithms (Table 2), to determine the one with the highest degree of accuracy.

These experiments were used to select the algorithm that could be used by the server to automatically classify the user access level in unfamiliar surroundings and that had not been configured in the system, or in other words, where all the rules of a well-defined environment can be found. The comparative experiment between the classification algorithms was carried out by Weka tool by Hall, Mark, et al. (2009). The Table with the rules (a combination of all the attributes and their types that result in access) was divided into training and testing sets, through a cross-validation technique with ten subsets (10-fold cross-validation) where 90% of the data is used for training the classification algorithms, and the remaining 10% is used to check the results of these rules (unknown to the classifier). In addition, with this method, the test set is varied

among all the possible data training subsets. The final degree of accuracy shown in Table 2 is obtained from the average of the tests.

Table 2: Comparison of Classification Algorithms

Algorithm Classification	Precision	Correct instances	Incorrect instances
DecisionTable	0.887	343	40
Bayes Network	0.814	322	61
J48	0.887	341	42
Best-First Decision Tree (BT-Tree)	0.871	336	47
RandomTree	0.861	332	51
Nearest Neighbor With Generalization (NNge)	0.848	326	57
MultilayerPerceptron	0.888	341	42

An ontology was designed to formally represent the UbiPri model, [Leithardt et al., 2013]. According to [Bouiadjra et al., 2011], ontologies can be used to represent the context, provide inferences and share the knowledge generated by the application. Similarly, X states that axioms, bodies and vocabulary can be shared with the scientific community. FOEval assessment and evaluation scenarios were used to validate the ontology. FOEval allows users to select a set of metrics that can assist in the evaluation of ontologies. For this study, metrics were chosen with a level of detail and computational efficiency recommended by. Three calculations are made to assess wealth. First the wealth ratio (RR) measures the range of relationships, and assumes that the higher the number of non-hierarchical relationships, the richer the ontology. Another calculation performed is the wealth attribute (RA) which is the average number of attributes that are defined for each class. This may indicate the amount of information in the instance data, since the more attributes are set, the greater the amount of knowledge the ontology conveys. Finally, the rich ontology is calculated (RO) from the RR and RA.

According to [Bouiadjra et al., 2011], RO is set to the sum of RR and RA. RR is defined as the ratio between the numbers of non-hierarchical relationships defined in the ontology, divided by the number of all the (R) attribute wealth relations. RA, in turn, is defined as the number of attributes defined for all classes divided by the number of ontology classes. To make the calculations of the ontology [Gruber, T. R. et al., 1993], data were used (as summarized in Table 3). The results obtained for the ontology were UbiPri [Leithardt et al., 2013] points for RR 1.27; 0.31 points for RA; and 1.58 points for RO. These results demonstrate that UbiPri ontology is richer in relationships than in attributes. Furthermore, there is, on average, 15 classes by attribute and a relationship of 3 classes. The overall level of detailed calculation is defined by the average number of subclasses divided by the number of ontology classes. When we make to obtain the total number of subclasses, including 45 classes. We come to near zero, 0.02 ($(48/45) / 45 = 0.02$). UbiPri ontology is divided in terms of classes and subclasses, and is very close to midway between the vertical and horizontal types of the taxonomy.

5. Conclusion and Suggestions for Future Work

Several studies have been conducted que have led to the survey. With the results, we intend that they be used for future research. The scientific investigations showed that it is possible to set criteria, parameters and variables que comply with the particular rules

of each environment. Thus, it can be concluded that it is also the basis of information que is handled automatically. In this study, a comparison preliminary was made of classification algorithms for data privacy and treatment.

These were focused on the environment. However, a good deal of further research and implementations are required in the future, due to the large amount of information que needs to be controlled and managed in ubiquitous environments. The results were generated from real time and in real world scenarios, and were based on information about students at an academic institution. That involves listening to lectures and taking part in other events in classrooms and the auditorium. We carried out a simulation to test the same settings with thousands of students, but due to several factors such as time and physical resources, it was not possible to simulate all profiles and managements. For this reason this fact can be listed among many others the areas for the continuation of research into data privacy in the future which shouldn't be of value in the area of computer studies. Figure 3 shows the location, identification and classification of the environment used in the tests. It was concluded from this que the behavior of algorithms chosen for the sample of 500 users achieved the expected results.

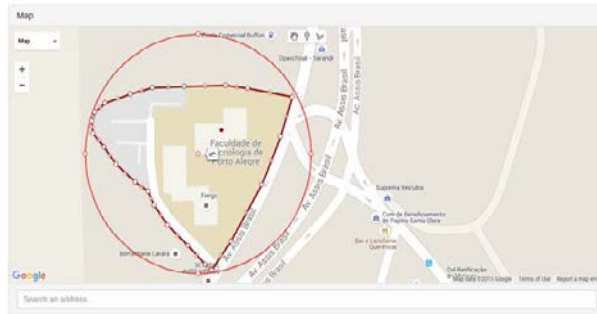


Figure 3. User identification and environment

In future work we plan to simulate larger environments, with a greater number of users, devices and criteria definitions than what was used to obtain the results of this work. For this reason, two projects have been submitted to funding agencies of Brazil requesting financial assistance for further studies. This will enable us to identify and develop more robust algorithms for the handling of private data on a large scale. Other important factors also need to be considered for future work, including the following:

- Definition of data processing techniques on a large scale;
- Devising computational metrics for a database;
- Preparation of other environmental ratings, for ex-ample, those restricted to individuals;
- Implementation of distributed algorithms for processing information;
- Definition of bug tracking techniques, as well as others that will be determined in the course of the research.

Acknowledgment

This work forms a part of the UbiArch project - Ubiquitous Architecture for Context Management and Application Development of the Federal University of Rio Grande do

Sul. We would like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [National Council of Scientific and Technological Development].

References

- Bardram, J. E., Kjær, R. E., & Pedersen, M. (2003) “Context-aware user authentication—supporting proximity-based login in pervasive computing”. In *UbiComp 2003: Ubiquitous Computing*, pp. 107-123. Springer Berlin Heidelberg.
- Abech, M., da Costa, C. A., Barbosa, J., Rigo, S., and Cambuzzi, W. (2012) “Um Modelo de Adaptação de Objetos de Aprendizagem com foco em Dispositivos Móveis”. In *Anais do Simpósio Brasileiro de Informática na Educação*.
- Bouiadjra A B. (2011) FOEval: Full Ontology Evaluation In proceeding of: 7th International Conference on Natural Language Processing and Knowledge Engineering, NLPKE, Tokushima, Japan.
- Cristiano Andre da Costa, Adenauer Correa Yamin, and Claudio Fernando Resin Geyer. 2008. Toward a General Software Infrastructure for Ubiquitous Computing. *IEEE Pervasive Computing* 7, 1 (January 2008), 64-73.
- Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5, (2):199-220, 1993.
- Hall, Mark, et al. "The WEKA data mining software: an update." *ACM SIGKDD explorations newsletter* 11.1 (2009): 10-18.
- Henricksen, K., Wishart, R., McFadden, T., & Indulska, J. (2005) “Extending context models for privacy in pervasive computing environments”. In *Pervasive Computing and Communications Workshops, 2005. PerCom 2005 Workshops*. pp. 20-24. IEEE.
- Iachello, G., and Hong, J. (2007) “End-user privacy in human-computer interaction”. *Foundations and Trends in Human-Computer Interaction*, v. 1, n. 1, pp. 1-137.
- Leithardt, Valderi R. Q., Guilherme A. Borges, Anubis G. M. Rossetto, Carlos O. Rolim, Claudio Geyer, Luiz H. A. Correia, David Nunes and Jorge Sa Silva: A Privacy Taxonomy for the Management of Ubiquitous Environments. *Journal of Communication and Computer* Volume 10, Number 12, December 2013.
- Rodrigues, Vagner J.d.S. (2006) “Privacy Management to Context-Aware Applications on Mobile Networks”, <http://www.inf.ufg.br/~vagner/publications/Tese-2006-Vagner.pdf>. Accessed in January 2016.
- Santarosa, L. M. C., Conforto, D., & Basso, L. d. O. (2010) “Eduquito: Ergonomia Cognitiva para a Diversidade Humana”. In: *Educação, Formação & Tecnologia*, v. 3, p. 4-13. ISSN 1646-933X.
- Tao, H., & Peiran, W. (2010) “Preference-Based Privacy Protection Mechanism for the Internet of Things”. In *Information Science and Engineering (ISISE), 2010 International Symposium on*, pp. 531-534. IEEE.
- Warren, S. D., and Brandeis, L. D (1890) “The right to privacy”. *Harvard law review*, 193-220.
- Weiser, M. “The computer for the twenty-first century.” *Scientific American*, vol. 265, no. 3, pp. 94–104, September 1991.

Um mecanismo baseado em sensibilidade à situação para otimizar a transmissão de dados de aplicações de sensoriamento urbano

Carlos O. Rolim¹, Anubis G. Rossetto¹, Valderi R. Q. Leithardt¹
Guilherme A. Borges¹, Tatiana F. M. dos Santos²
Adriano M. Souza², Claudio F. R. Geyer¹

¹Institute of Informatics – Federal University of Rio Grande do Sul (UFRGS)
91.501-970 – Porto Alegre – RS – Brazil

²Postgraduate Program in Production Engineering
Federal University of Santa Maria (UFSM)
Santa Maria, RS, Brazil

carlos.oberdan@inf.ufrgs.br, taty.nanda@gmail.com

Abstract. *One challenge in urban sensing applications is related to the transmission of data when network infrastructure is intermittent or unavailable. For such task, the Opportunistic Networks emerged as a promising alternative. The main problem in this context is how to improve messages routing between network nodes. Thus, this work presents a engine that employs situation awareness in conjunction with uncertainty treatment to overcome such limitation. The experiments demonstrate that the proposed engine was an acceptable performance and can be used as a basis for opportunistic data transmission in a urban sensing architecture called UrboSenti.*

1. INTRODUÇÃO

O conceito de Cidades Inteligentes foi utilizado pela primeira vez por [Gibson et al. 1992] para expressar como o desenvolvimento urbano estava passando a utilizar a tecnologia, inovação e globalização em assuntos cotidianos. Para autores como [Hernández-Muñoz et al. 2011], objetos, serviços e pessoas em uma Cidade Inteligente estão interconectados por uma infraestrutura de rede. Como resultado, pesquisadores estão procurando alternativas para proporcionar aos habitantes da cidade novos serviços e aplicações que melhorem a qualidade de vida e ao mesmo tempo atendam os critérios de eficiência energética e sustentabilidade.

Uma das alternativas mais promissoras e que já possui aplicações reais é o uso de pessoas, dispositivos móveis e sensores para a coleta de informações a respeito de diferentes aspectos da cidade em uma nova área da computação ubíqua chamada de Sensoriamento Urbano. Uma limitação para aplicações de sensoriamento urbano que utilizam protocolos Internet para transmissão está relacionada a como transmitir os dados de sensoriamento quando a infraestrutura de rede for intermitente ou indisponível (devido aos protocolos tradicionais de roteamento necessitarem de um caminho fim-a-fim entre a origem e o destino [Cao and Sun 2013]). As Redes Oportunistas surgem então como uma alternativa para suprir essas limitações [Boldrini et al. 2008]. Entretanto, no contexto de Redes Oportunistas, um problema que pesquisadores da área ainda buscam resolver está

relacionado a como otimizar o roteamento e a disseminação de mensagens (aumentando sua taxa de entrega e diminuindo sua latência e overhead) em cenários heterogêneos com alta dinamicidade, como no caso de Cidades Inteligentes.

Assim, o presente trabalho tem por objetivo apresentar um mecanismo que emprega sensibilidade à situação juntamente com tratamento de incertezas para otimizar o roteamento de mensagens em Redes Oportunistas. Parte-se da hipótese de que derivar os dados de contexto dos nodos de uma Rede Oportunista para uma visão de mais alto nível proporcionada pela sensibilidade à situação melhora as tomadas de decisões de roteamento. Como contribuições desse trabalho destaca-se: (i) é o primeiro trabalho que demonstra resultados de aplicação de sensibilidade à situação em Redes Oportunistas; (ii) apresentação da modelagem conceitual de um mecanismo de sensibilidade à situação capaz de lidar com incertezas. Além disso, os conceitos desse mecanismo serão aplicados no módulo de sensoriamento de uma arquitetura ubíqua orientada a serviços para o sensoriamento urbano chamada UrboSenti [Rolim et al. 2014]. Desde já ressalta-se que esse artigo apresenta uma evolução do trabalho apresentado anteriormente em [Rolim et al. 2015]. Assim, os maiores diferenciais em relação ao trabalho anterior estão relacionados aos modelos e técnicas empregadas para o tratamento de situações e o tratamento de incertezas.

O trabalho está organizado da seguinte forma: na próxima seção é apresentado o cenário problema que serve de motivação para o desenvolvimento desse trabalho; a seção 3 descreve a modelagem do cenário-problema e também a arquitetura do mecanismo de situação proposto; a seção 4 detalha as configurações empregadas nas simulações e também os resultados experimentais e, por fim, a seção 5 apresenta as conclusões e trabalhos futuros.

2. CENÁRIO PROBLEMA

Esse artigo faz o uso de um cenário-problema de uma Cidade Inteligente que utiliza nodos fixos e móveis para efetuar o sensoriamento urbano. Veículos e pessoas carregam seus dispositivos para obter dados do trânsito, da qualidade do ar, níveis de ruído, concentração de CO_2 , etc. Em locais estratégicos da cidade existem Pontos de Acesso sem fios (Access Points – APs) que são responsáveis por receberem os dados de sensoriamento coletados e os encaminharem para serem processados por um sistema central. Além de dados de sensores, redes sociais são usadas para que seja possível compreender o comportamento da cidade e assim, através de uma visão holística, fornecer melhores serviços aos seus habitantes.

Todo o processo de coleta, processamento, análise e retorno de informações aos moradores da cidade é efetuado por uma arquitetura ubíqua orientada a serviços para o sensoriamento urbano chamada UrboSenti. Um dos módulos da UrboSenti é chamado de *Sensing module* e é responsável pela coleta de dados de sensoriamento de forma intencional ou não-intencional. Ele é executado em dispositivos móveis (como por exemplo, smartphones, veículos, etc) ou até mesmo em sensores fixos que são dispersos ao longo da cidade. Uma das características desse módulo é o suporte a diferentes métodos para transmissão dos dados coletados. É possível usar o padrão IEEE 802.11b/g/n (em modo estruturado ou ad hoc), GPRS/EDGE/3G ou mesmo rede cabeada como base para a comunicação utilizando TCP/UDP. Em locais onde a infraestrutura física não estiver

presente ou for intermitente, o *Sensing module* proporciona que a comunicação ocorra através de formas alternativas, não baseadas em caminhos fim-a-fim, como acontece na pilha TCP/IP.

Essa forma de comunicação alternativa emprega o paradigma de Redes Oportunistas para permitir que os dados de sensoriamento sejam roteados entre os nodos até serem entregues aos APs que os enviarão para o processamento. As Redes Oportunistas (ou OppNets, do inglês *Opportunistic Networks*) são consideradas uma evolução das MANETs e um subcategoria de Redes Tolerantes a Atrasos e Desconexões. As OppNets herdaram todas as características das DTNs, porém elas exploram a mobilidade e contatos dos nodos como forma de proporcionar a entrega de dados. Dessa forma, mobilidade, desconexões, atrasos, particionamentos e demais aspectos considerados desafios em redes móveis são vistos como características das Redes Oportunistas. Pelo fato de não necessitar da existência de um caminho direto entre a origem e o destino, é empregado o conceito de armazena-carrega-encaminha (*Store-Carry and Forward*) para troca de mensagens. A cada contato entre os nodos é decidido quais mensagens devem sair do buffer e serem encaminhadas ao nodo encontrado e quais devem permanecer. Se o nodo encontrado recebeu mensagens, ele irá armazenar localmente e em um novo contato com outro nodo deverá decidir o que deve ser encaminhado ou permanecer consigo. Esse processo se repete em todos os encontros entre nodos da rede até que a mensagem seja entregue ao seu destino.

O desafio existente nessa forma de comunicação está relacionado à estratégia de seleção de boas “mulas de dados” (do inglês *data mule* – um termo utilizado em Redes Oportunistas para designar nodos com boa capacidade de carregar uma mensagem ao seu destino ou o mais próximo deste). Diferentes propostas foram feitas por pesquisadores para superar tal desafio. Como exemplos, podem ser citados: Epidemic, Prophet, BubbleRap, MobySpace, AFRON, CAR, HiBOp, Propicman, DRAFT entre outros. Não está no escopo do presente artigo analisar a cada uma das propostas (maiores informações podem ser obtidas em [Boldrini et al. 2008]). Entretanto, acredita-se que essas propostas existentes não capazes de lidar de forma adequada com os requisitos do cenário-problema apresentado nesse trabalho. Assim, busca-se verificar a validade da hipótese de que a aplicação de sensibilidade à situação é uma forma promissora de superar o desafio de escolha de boas mulas dados em cenários de sensoriamento urbano.

3. SOLUÇÃO PROPOSTA

3.1. Modelagem do cenário

A modelagem conceitual do cenário-problema apresentado anteriormente é representada na Figura 1 e assim descrita: cada nodo do cenário possui dados de contexto chamados de **contexto de baixo nível**. Ele é definido pelo conjunto de dados de contexto interno e de contexto externo. O **contexto interno** envolve os dados que caracterizam o nodo em um determinado instante corrente ou passado. Como exemplos de atributos de contexto interno podem ser considerados: nome, posição, velocidade, energia da bateria, nodos próximos, distância percorrida total, distância percorrida desde o último encontro, capacidade de processamento, número de mensagens no buffer, espaço disponível no buffer para armazenamento, tráfego na rede e também outros atributos suportados pelo dispositivo. O contexto interno também é formado por valores que provavelmente irão caracterizar o

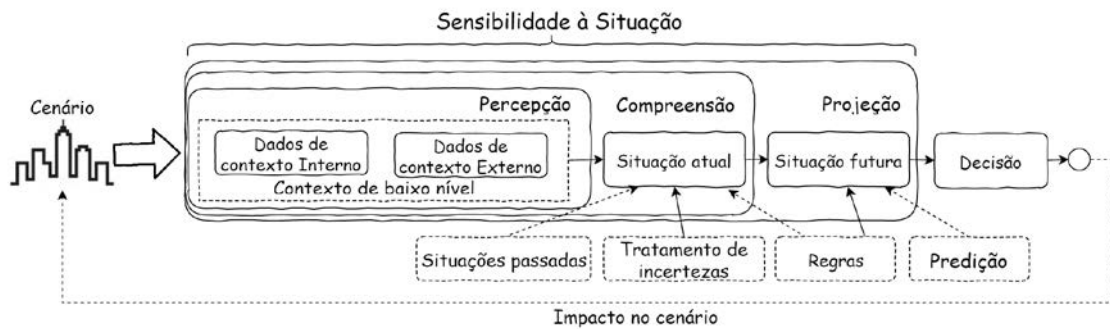


Figura 1. Modelo do cenário-problema empregando sensibilidade à situação

nodo em um momento futuro. Esses são chamados de **valores preditos** e são gerados a partir da aplicação de uma técnica de predição.

Além do contexto interno, cada nodo é capaz de obter informações dos nodos vizinhos e também do ambiente que o circunda. Esses dados são chamados de **contexto externo**. Como contexto externo são considerados os atributos dos nodos vizinhos como: velocidade, capacidade de processamento, espaço em buffer, energia disponível, etc. O ambiente possui atributos que podem ser coletados pelos sensores disponíveis no nodo, como por exemplo: ruído, luminosidade, qualidade do ar e também outros atributos que possam ser coletados pelos dispositivos.

O contexto de baixo nível do nodo é utilizado para auxiliar na tomada de decisão de encaminhamento de mensagens e também nas adaptações das funcionalidades internas do componente. Porém, conforme apontado por [Ye et al. 2011], utilizar partes isoladas de dados de contexto pode não apresentar significado útil para a tomada de decisões. Assim, com a finalidade de agrupar os dados de contexto de baixo nível e então derivar uma nova visão de alto-nível que facilite a tomada de decisão é empregado o conceito de *sensibilidade à situação* que originalmente foi proposto por [Endsley 1995].

Os 3 níveis desse modelo de sensibilidade à situação foram adaptados para o mecanismo proposto da seguinte forma: **Nível 1 - Percepção:** é considerado o nível mais baixo da sensibilidade à situação. Aqui, ele tem por finalidade perceber o status e valores dos atributos dos elementos e interações do contexto de baixo nível e então agrupa-los para serem utilizados pelas camadas superiores. É importante notar que alterações nos valores do contexto de baixo nível poderão caracterizar uma situação atual diferente da anterior, tornando assim o sistema dinâmico; **Nível 2 - Compreensão:** utiliza os dados de contexto de baixo nível advindos do Nível 1 e deriva a visão de alto nível chamada de situação. Em outras palavras, a principal função do nível 2 é de fornecer a situação dos nodos. Um repositório de *situações passadas* é usado juntamente com um conjunto de *regras* para caracterizar situações. Um fato importante que deve ser considerado é que os dados advindos do Nível 1 nem sempre serão dados precisos. Dessa forma, o Nível 2 deve ser capaz de lidar com a imprecisão dos dados, preferencialmente, através de uma abordagem baseada em Inteligência Computacional. **Nível 3 - Projeção:** tem por finalidade projetar ações futuras dos nodos. Os valores de contexto interno preditos são utilizados pelo nível 2 para caracterizar uma provável situação futura. O conjunto de situações passadas, corrente e futura são utilizados por este nível para estimar uma ação que deverá

ser executada em um futuro próximo. O acerto ou erro na estimativa da projeção futura influencia na decisão tomada. O impacto disso no cenário é mensurado para melhorar as novas projeções que serão efetuadas.

É importante ressaltar que devido às restrições de energia e processamento dos nodos existentes no ambiente, as atividades dos níveis 2 e 3 devem ser desenvolvidas com custo computacional que não comprometa a precisão das ações e não deteriore a bateria dos nodos.

3.2. Arquitetura do mecanismo

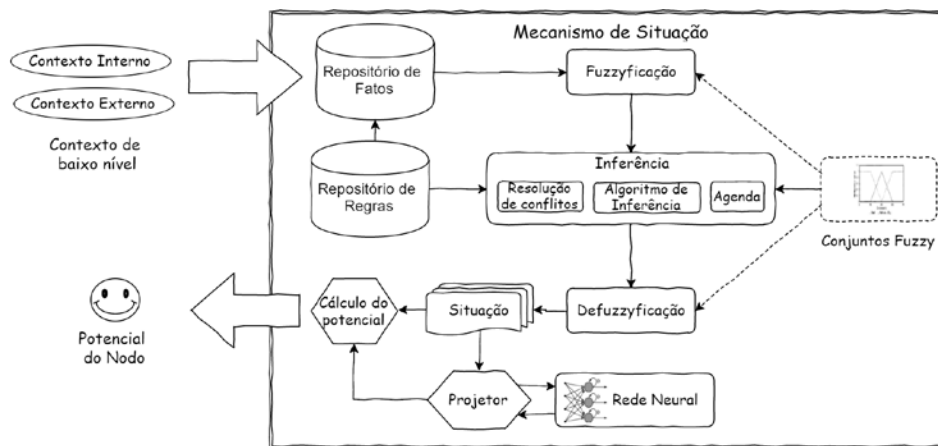


Figura 2. Arquitetura do Mecanismo de Situação

A arquitetura do Mecanismo de Situação é apresentada na Figura 2 e assim descrita: Os dados de *Contexto de baixo nível* (contexto interno e contexto externo) são fornecidos como entrada para o mecanismo. Esses dados, denotados por C_n onde $n = \{power, buffer, distance_travelled, node_type, \#of_messages, \dots\}$, são estruturados de forma adequada e então armazenados no *Repositório de fatos*. Cada conjunto de dados de Contexto de baixo nível é considerado um fato. Dessa forma, no *Repositório de fatos* existem fatos atuais e também fatos históricos que são utilizados para proporcionar a percepção da situação (nível 1 do modelo de Endsley). Um fato passa então pelo processo de *fuzzyficação* onde seus valores são convertidos para valores fuzzy de acordo com funções (membership functions) definidas no *Conjunto Fuzzy*. Os valores convertidos representam o grau de pertinência daquele valor a determinado grupo (como por exemplo, carga da bateria cheia, pela metade, ou vazia). Tais valores fuzzy servem de entrada para o processo de *Inferência*. Nesse processo, são aplicadas regras armazenadas no *Repositório de Regras* juntamente com um algoritmo de inferência baseado em encadeamento progressivo ou regressivo (forward chaining ou backward chaining) para determinar quais regras são aplicadas aos dados fuzzy. Se forem verificadas regras conflitantes, é aplicado a *Resolução de conflitos* que determina quais regras devem ser disparadas. A *Agenda* então permite priorizar, dentre as regras quem combinam com os dados (considerado um match), a ordem de disparo de cada uma. A saída do processo de inferência é então *Defuzzyficada* e são obtidos valores V_{s_n} que caracterizam a situação de cada uma das variáveis de Contexto de baixo nível (nível 2 do modelo de Endsley). Os valores V_{s_n} são encaminhados para o *Projektor* que fornece os V_{s_n} atuais e anteriores (histórico) para

uma *Rede Neural*. Como saída são gerados valores P_{s_n} que representam a predição da caracterização de situações futuras (nível 3 do modelo de Endsley). Além de serem utilizados para a projeção de situação, os valores V_{s_n} são empregados para o *Cálculo do potencial* do nodo. Nesse processo é atribuído um peso W_n para cada C_n de acordo com sua importância para o nodo ser considerado uma boa “mula de dados”. Então é feita a normalização desses pesos utilizando uma função na forma $W'_n = f(x) = w_{ni} / \sum_{i=1}^n w_{ni}$. Uma vez que os pesos estão normalizados, é calculado o *Potencial do Nodo* com a equação $P_{node} = V_{s_{n1}} * W'_{n1} + V_{s_{n2}} * W'_{n2} + \dots + V_{s_{nN}} * W'_{nN}$. Todo esse processo é executado ciclicamente em períodos determinados de tempo.

Quando um nodo n_j encontra um nodo n_k é usado P_{node} para decidir se as mensagens que estão no buffer local devem ser encaminhadas para o nodo encontrado ou devem permanecer no buffer local. Para essa decisão n_j aplicada a equação:

$$forward = \begin{cases} true, & \text{if } P_{node}^{n_j} > P_{node}^{n_k} \\ false, & \text{otherwise} \end{cases}$$

Ressalta-se que ainda não foi incorporada à equação de definição do potencial do nodo P_{node} as situações previstas P_{s_n} . Tal aspecto encontra-se em análise e será publicado em artigos futuros.

4. SIMULAÇÕES E RESULTADOS EXPERIMENTAIS

4.1. Configurações dos cenários

Para verificar a funcionalidade e a performance do Mecanismo de Situação foram executadas simulações utilizando o ONE (Opportunistic Network Environment) Simulator [Keränen et al. 2009]. Sua escolha deve-se a ser de código aberto, extensível e amplamente utilizado por pesquisadores da área de Redes Oportunistas. As simulações foram executadas durante um período de 6 horas em um cenário da cidade de Porto Alegre/RS. A cidade teve as suas principais avenidas mapeadas e convertidas para o formato interno aceito pelo simulador. Foram então criados 2 grupos de nodos: pessoas e veículos.

Em cada uma das simulações foi adotada uma quantidade crescente de nodos (10 nodos de cada grupo na primeira simulação, 25 de cada grupo a segunda, 50, 75, 100, 500 e 1000 nodos de cada grupo na última simulação). Ambos os grupos se movimentavam no mapa utilizando o modelo de mobilidade *Shortest Path Map Based Movement* (nesse modelo, os nodos utilizam caminhos aleatórios disponíveis no mapa e utilizam o algoritmo de Dijkstra para encontrar o caminho mais curto entre os nodos e Pontos de Interesse). Além disso, aos dois grupos foram aplicadas as seguintes restrições de energia: todos os nodos iniciaram com uma carga completa de bateria de 19080 Joules (que corresponde a uma bateria comumente encontrada em celulares de 5.3Wh com 3.7V) e uma recarga sendo efetuada aleatoriamente em intervalos entre 4 horas e 4 horas e meia. A energia gasta para a busca de outros nodos foi definida em 0.092 mW/s e o gasto para enviar e receber mensagens definido em 0.08 mW/s. O grupo formado por pessoas se movimentavam com velocidade entre 0.5 e 1.5 Km/h e tinham um celular equipado com uma interface Bluetooth com alcance de 5 metros e velocidade de transmissão de 2Mbps e uma interface WiFi com alcance de 20 metros e 10Mbps de velocidade de transmissão. O grupo de carros se movimentava com velocidades entre 10-50 km/h e possuía uma interface WiFi com as mesmas características do grupo de pessoas. Tanto o grupo de carros

quanto o de pessoas possuía um buffer para armazenamento de mensagens com 50M de tamanho.

Em locais estratégicos da cidade foram distribuídos 7 APs responsáveis por receber os dados de sensoriamento coletados pelos nodos móveis e então os encaminhá-los via rede cabeada para serem processados por um sistema centralizado. Para isso, eles foram equipados com uma interface WiFi com alcance de 30 metros e 10Mbps de velocidade de transmissão e uma interface cabeada com 1Gb de velocidade. Por serem equipamentos dedicados, esses APs estavam conectados permanentemente na energia, não precisando de recarga de elétrica, e possuíam o buffer para armazenamento de mensagens com 512Mb de tamanho. Cada nodo do cenário foi equipado com uma aplicação responsável por simular o sensoriamento de eventos ao longo da cidade. Os eventos de sensoriamento ocorreram em intervalos de 100 a 200 segundos, sendo que as mensagens foram geradas com tamanho variando segundo uma distribuição uniforme de 100k a 2M. Cada mensagem foi definida com um TTL de 300 segundos (5 horas). Esse é o tempo máximo para que uma mensagem seja roteada no cenário e então entregue ao primeiro APs que o nodo que carrega tal mensagem tiver contato.

4.2. Estratégia de implementação

Para implementar o Mecanismo de Situação foi utilizado como base o JBoss Drools. O Drools é uma de sistema de inferência híbrido baseado em regras. Seu motor de inferência permite a construção de regras que serão avaliadas através de algoritmos baseados na combinação de encadeamento progressivo e regressivo, como RETE networks [Forgy 1982] (no Drools chamado de ReteOO para indicar suporte à orientação a objetos) e também uma evolução desse algoritmo chamado de PHREAK.

Para definição de regras, o Drools utiliza uma linguagem própria chamada DRL (Drools Rule Language). Um arquivo DRL, basicamente, contém um conjunto de regras do tipo “when LHS then RHS” onde LHS (Left Hand Side) ou antecedente corresponde aos elementos condicionais da regra e RHS (Right Hand Side), também chamado consequente, corresponde as ações que serão executadas em caso de disparo da regra. Tanto o LHS e o RHS podem ser formados por conjuntos de zero, um ou mais itens. Como o Drools nativamente somente é capaz de utilizar regras sem tratamento de incertezas, foi utilizado uma extensão do sistema chamada de Drools Chance [Sottara et al. 2008, Wulff and Sottara 2009]. O Chance permite que as regras sejam definidas e avaliadas utilizando Lógica Fuzzy. Dessa forma, a partir das definições de funções de pertinência e da aplicação dos algoritmos de inferência Fuzzy consegue-se efetuar o tratamento de incertezas conforme requerido pelo modelo de situação que está sendo proposto e foi apresentado na seção 3.1. O Chance segue a mesma lógica de declaração de variáveis e regras de um arquivo DRL normal. Entretanto, deve-se declarar a forma como os valores serão manipulados pela Lógica Fuzzy. Para isso, as variáveis linguísticas são definidas através de funções de pertinência e então podem ser utilizadas nas regras. No mecanismo de situação, cada uma dos atributos de contexto de baixo nível teve uma variável linguística vinculada com três funções triangulares com valores variando em escala de “pouco”, “médio” e “alto” de acordo com sua magnitude. Essas variáveis linguísticas foram empregadas em regras que fornecem como saída a situação de cada atributo de contexto de baixo nível. Por questão de espaço, não serão fornecidos exemplos de sintaxe das regras utilizadas.

Conforme apresentado na seção 3.2, uma vez definida a situação de cada atributo de contexto de baixo nível é calculado o potencial do nodo. Como parte da equação de cálculo do potencial, foi definida uma tabela contendo os pesos para cada um dos dados de contexto de baixo nível. Lembrando que esses pesos são estipulados de acordo com a importância do dado para o nodo ser considerado uma boa mula de dados. Foi definido então a energia da bateria e o tipo de nodo (se estático ou móvel) com peso 1.0, o espaço em buffer com peso 0.8, distância percorrida com peso 0.4 e a quantidade total de mensagens que o nodo já carregou com peso 0.2. Esses pesos então foram normalizados e aplicados na equação que como saída indicou o potencial momentâneo de cada um dos nodos e que define se uma mensagem deve permanecer no buffer local ou ser encaminhada ao nodo encontrado. Ressalta-se que o mecanismo de situação proposto no momento utiliza somente situações atuais, sem empregar situações projetadas (preditas). Dessa forma, não foi explicado nesse trabalho o funcionamento da Rede Neural que será utilizada (maiores detalhes da Rede podem ser encontrados em [Rolim et al. 2015]).

4.3. Resultados experimentais

Para verificar a performance do mecanismo proposto, foram realizados experimentos em diferentes cenários utilizando as configurações de simulação detalhadas acima. Os resultados são apresentados na Figura 3.

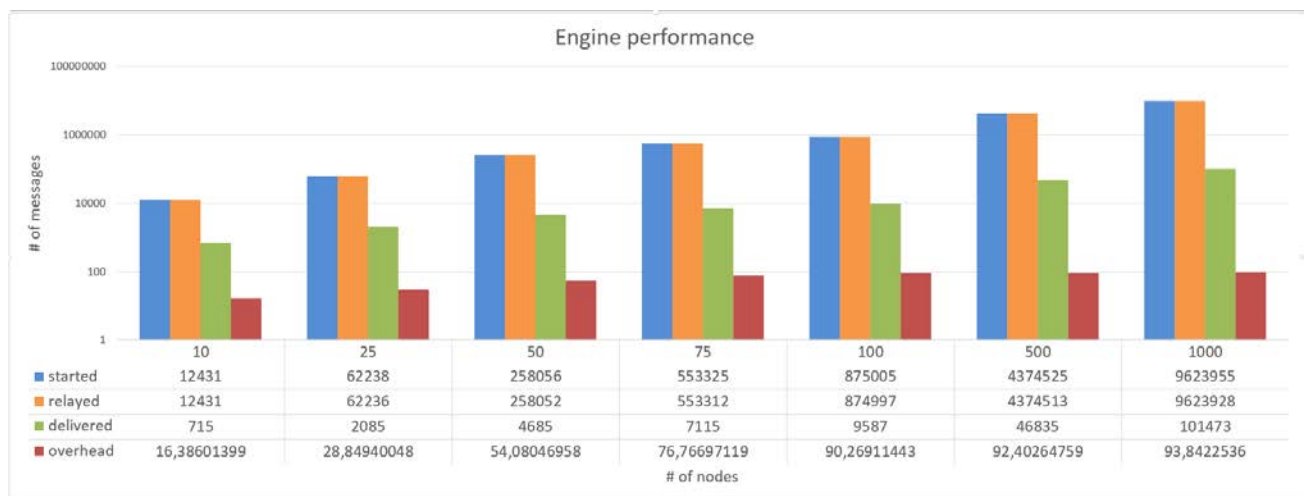


Figura 3. Performance do mecanismo proposto

Os resultados demonstraram que no geral há um incremento no número de mensagens entregues (delivered) conforme o aumento de nodos. Comparando os cenário de 10 nodos com o cenário de 50 nodos (crescimento de 5 vezes o número de nodos), ocorreu um aumento de mais de 500% na quantidade de mensagens entregues, porém o overhead foi de apenas 29%. No comparativo de 100 nodos com 500 nodos o aumento do overhead foi de apenas 2% e o aumento da quantidade de mensagens entregues foi de quase 400%. Ou seja, cada vez que o número de nodos duplica praticamente duplicam as mensagens entregues. O overhead não acompanha essa taxa de crescimento, mantendo-se baixo. Isso demonstra que o uso do potencial do nodo baseado na situação definida está sendo efetiva nas tomadas de decisões de encaminhamento.

Um fato que chamou a atenção nas simulações foi a quantidade de mensagens

criadas (started) e encaminhadas (relayed) entre os nodos. Isso ocorre devido ao aumento de contatos a medida que são inseridos mais nodos no cenário. Pode-se notar que algumas vezes uma mensagem saiu do buffer de um nodo e algum tempo depois, após outros contatos, acabou retornando ao mesmo nodo. Ao mesmo tempo que isso aumenta a probabilidade de entrega de uma mensagem ao destino, também aumenta o overhead para entrega da mensagem. Uma solução é fazer com que uma mensagem não possa retornar a um nodo pelo qual ela já passou. Nos testes isso reduziu em torno de 30% a quantidade de mensagens encaminhadas e também o overhead, mas também diminuiu a quantidade de entregas em torno de 20%. Essa solução precisa ser melhor analisada.

Outro aspecto interessante que pode-se observar nas simulações e que não está representado nos gráficos é o impacto do TTL das mensagens. Foram efetuados testes com TTL de 60 minutos e com 300 minutos (5 horas). Com poucos nodos do cenário a quantidade de mensagens expiradas foi maior com um TTL de 60 minutos. Isso também era esperado devido ao pouco contato entre nodos. Porém pode-se notar que a partir de 50 nodos no cenário o TTL de 60 minutos gerou um impacto menor no overhead e manteve praticamente a mesma taxa de mensagens entregues. Ou seja, pode-se pensar que o uso de um TTL dinâmico, definido de acordo com alguma variável de contexto (como por exemplo quantidade de contatos do nodo) pode ser uma boa alternativa. Essa opção também será analisada melhor no futuro.

Além disso, foram realizados comparativos de performance com algumas das principais iniciativas de Redes Oportunistas. Essas iniciativas foram escolhidas devido a representarem diferentes “classes de protocolos” estudadas por pesquisadores da área [Boldrini et al. 2008]. Foram efetuadas simulações com o Epidemic, Prophet, Bubble-Rap e Draft (devido a questões de espaço não serão fornecidas maiores informações das configurações usadas e o detalhamento dos resultados). Os experimentos demonstraram que o mecanismo proposto teve a quantidade de mensagens entregues apenas inferior ao Epidemic, porém com overhead menor em todos os cenários. Com isso, pode-se concluir que o uso de sensibilidade à situação está sendo uma boa estratégia para guiar as decisões de roteamento.

5. Conclusões e trabalhos futuros

Esse artigo apresentou um mecanismo que utiliza sensibilidade à situação para otimizar o roteamento de mensagens em Redes Oportunistas aplicadas ao sensoriamento urbano. Além do uso de sensibilidade à situação, um dos diferenciais desse mecanismo é o suporte ao tratamento de incertezas para definição da situação dos nodos. Os resultados experimentais demonstraram que a caracterização de situação proporciona melhorias na tomada de decisões de roteamento, validando assim a hipótese inicial.

O mecanismo proposto já implementa os níveis 1 e 2 do modelo de situação. Porém ainda não está fazendo projeções de situações futuras (nível 3). O suporte a essa característica está em fase de desenvolvimento com o uso de Redes Neurais e acredita-se que os resultados poderão ser melhorados quando a codificação estiver finalizada.

Como trabalhos futuros, pode-se apontar a finalização do suporte a projeção de situação, o desenvolvimento de novas simulações para verificação da performance do componente e também estudos do impacto da recusa de um nodo receber mensagens que já passou por ele anteriormente e o uso de TTL definidos dinamicamente de acordo

com alguma variável de contexto do nodo. Por fim, espera-se que com a conclusão do mecanismo o mesmo possa ser incorporado ao módulo Sensing da UrboSenti para então iniciarem pequenos testes de prova de conceito em ambientes reais.

Referências

- Boldrini, C., Boudec, J. Y. L., Chaintreau, A., and Conti, M. (2008). Deliverable 2.2: Final specification of forwarding paradigms in Haggle. Technical report.
- Cao, Y. and Sun, Z. (2013). Routing in Delay/Disruption Tolerant Networks: A Taxonomy, Survey and Challenges. *IEEE Communications Surveys & Tutorials*, 15(2):654–677.
- Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems.
- Forgy, C. L. (1982). Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial Intelligence*, 19(1):17–37.
- Gibson, D. V., Kozmetsky, G., and Smilor, R. W., editors (1992). *Technopolis Phenomenon: Smart Cities, Fast Systems, Global Networks*. Rowman & Littlefield Publishers, Inc.
- Hernández-Muñoz, J. M., Vercher, J. B., Muñoz, L., Galache, J. A., Presser, M., Hernández Gómez, L. A., and Pettersson, J. (2011). Smart cities at the forefront of the future internet. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6656:447–462.
- Keränen, A., Ott, J., and Kärkkäinen, T. (2009). The ONE Simulator for DTN Protocol Evaluation. In *SIMUTools '09: Proceedings of the 2nd International Conference on Simulation Tools and Techniques*, New York, NY, USA. ICST.
- Rolim, C. O., Rossetto, A. G., Leithardt, V. R., Borges, G. A., dos Santos, T. F., Souza, A. M., and Geyer, C. F. (2014). Towards Predictive Routing Agents in Opportunistic Networks. In *Fifth International Workshop on Collaborative Agents Research & Development, CARE for Intelligent Mobile Services - Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Paris, France. Springer-Verlag Berlin Heidelberg.
- Rolim, C. O., Rossetto, A. G., and Leithardt, V. R. Q. (2015). Putting Opportunistic , Situational and Smart Approaches to Underlie the Data Transmission of Social Urban Sensing Applications. In *7o Simpósio Brasileiro de Computação Ubíqua e Pervasiva, Anais do XXXV Congresso da Sociedade Brasileira de Computação*, pages 191–200, Recife, PE.
- Sottara, D., Mello, P., and Proctor, M. (2008). Adding uncertainty to a Rete-OO inference engine. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5321 LNCS:104–118.
- Wulff, N. and Sottara, D. (2009). Fuzzy reasoning with a rete-oo rule engine. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5858 LNCS:337–344.
- Ye, J., Dobson, S., and McKeever, S. (2011). A review of situation identification techniques in pervasive computing. *Pervasive and Mobile Computing*, In Press,(0).

Um modelo de monitoramento de pacientes utilizando a ciência de situação

Lucas Closs, Cristiano André da Costa, Rodrigo da Rosa Righi

Programa de Pós-Graduação em computação Aplicada
Universidade do Vale do Rio dos Sinos (UNISINOS)
Avenida Unisinos, 950 – 93022-750 – São Leopoldo – RS - Brasil

lucas.closs@gmail.com, {cac, rrrighi}@unisinos.br

Abstract. *The constant advances in mobile and ubiquitous computing enable new developments and features in several areas, health being one of them. From this, arise the concept of ubiquitous care, where new ways of medical area applied like patient monitoring. In view of this feature, the SAC - Situation-Aware Care model proposes the use of patient monitoring together with situation awareness to perceive, comprehend and project deviations and degradations in the health of patients, providing the best medical information to doctors. To assess the viability of the SAC model was applied an evaluation by scenario, which is supported by the development of a prototype.*

Resumo. *Os constantes avanços na computação móvel e ubíqua possibilitam novos desenvolvimentos e recursos em diversas áreas, sendo a saúde uma destas. A partir disto, surge então o conceito de cuidados ubíquos, onde são aplicadas novas formas cuidados médicos, como o monitoramento de pacientes. Tendo em vista este recurso, o modelo SAC - Situation-Aware Care propõe o uso do monitoramento de pacientes em conjunto da ciência de situação para perceber, compreender e projetar desvios e degradações na saúde, disponibilizando aos médicos melhores informações para atendimento. Para avaliar a viabilidade do modelo SAC foi aplicado um estudo de caso, baseado em cenário, sendo este apoiado pelo desenvolvimento de um protótipo do modelo.*

1. Introdução

A comercialização de dispositivos móveis (*Mobile devices*), principalmente de *smartphones*, segundo a Gartner, no ano de 2015 foram comercializadas um total de 1.917 bilhões de telefones móveis, onde os *smartphones* representam 70% do total [MEULEN; WOODS, 2016]. Para o ano de 2016, a Gartner projeta um aumento de 1,4% no comércio de telefones móveis – 1.943 bilhões unidades – e também um aumento de 7% na representação de *smartphones* – atingindo 77% [MEULEN; WOODS, 2016]. Tal circunstância, acaba por incentivar e influenciar no desenvolvimento e nos estudos de novas tecnologias e novos recursos para a área de Computação Móvel (*Mobile Computing*). Outro nicho de dispositivos que vem se destacando tanto para os consumidores quanto para a comunidade acadêmica são os dispositivos vestíveis (*Wearable devices*) [PATEL; ASCH; VOLPP, 2015]. Estes, são capazes de coletar diversos dados de seu utilizador. Indo além desta avaliação individual

dos conceitos acima destacados, a aplicação combinada dos mesmos disponibiliza novos recursos para a Computação Ubíqua (*Ubiquitous Computing*), onde temos a tecnologia fortemente integrada ao cotidiano da sociedade [WEISER, 1991] [FRIEDEWALD; RAABE, 2011].

Conforme mencionado, os novos desenvolvimentos e os novos recursos disponíveis para a computação móvel e ubíqua, estão permitindo a manifestação de oportunidades em diferentes áreas de aplicação, como: saúde, educação, comércio e transportes [FELTES; BARBOSA, 2014]. Na área da saúde, destaca-se o segmento de Cuidados Ubíquos que intentam prover melhores condições, informações e recursos aos hospitais e aos médicos, tendo como consequência a melhora no atendimento e no tratamento dos pacientes [GELOGO; KIM, 2013]. Um destes recursos disponibilizados seria o monitoramento de pacientes, que tem como objetivo o acompanhamento diário do paciente registrando os seus sinais vitais. Entretanto, este recurso pode ser ampliado com a aplicação de uma camada de inteligência. Essa camada atua sobre estes dados detectando possíveis desvios ou problemas relacionados com a saúde do paciente. Tendo em vista esta lacuna, esse trabalho propõe o emprego da Ciência de Situação (*Situation Awareness*) [ENDSLEY, 1995] sobre estes dados transformando-os em informações que representam a situação atual do paciente e a projeção de futuros estados.

O modelo SAC – Situation-Aware Care tem como foco a implementação dos três níveis da ciência de situação, onde apresenta projeções de futuros estados como resultado e não somente compreensões sobre a situação atual, projetando assim possíveis desvios, degradações ou problemas na saúde deste paciente. Tais inferências permitem controles de registros médicos mais apurados sobre as condições do paciente, melhores diagnósticos e principalmente o monitoramento do paciente. Sendo está uma das principais contribuições do modelo SAC. Para este artigo foram avaliados outros cinco modelos, que também utilizam a ciência de situação, mas com temáticas distintas. Porém, dentre os cinco modelos avaliados, somente dois deles tiveram a aplicação das três fases da ciência de situação. Outro aspecto não abordado pelos trabalhos relacionados considerados é quanto a padronização dos dados e das informações processadas pelos mesmos. A padronização garante uma melhor estruturação assim como a interoperabilidade com aplicações terceiras e até mesmo com instituições médicas. O modelo SAC irá utilizar as definições propostas pelo openEHR [GONZÁLEZ-FERRER, 2015].

O artigo apresenta seis seções. A segunda seção fundamenta sobre os principais conceitos utilizados. A terceira seção apresenta as análises sobre os trabalhos relacionandos. Na quarta seção é descrito o modelo SAC. A quinta seção descreve sobre a avaliação realizada. E por fim, apresenta a conclusão e os futuros trabalhos.

2. Cuidados Ubíquos e Ciência de Situação

Cuidados ubíquos surgem com a intenção aumentar a eficiência, a acuracidade e disponibilidade de tratamentos médicos [GELOGO; KIM, 2013]. Uma das abordagens de cuidados ubíquos seria o monitoramento de pacientes, que possibilita que hospitais e médicos tenham conhecimento sobre sinais vitais de seus pacientes no decorrer do dia [GELOGO; KIM, 2013]. Tal conhecimento permite: um melhor monitoramento sobre

os tratamentos médicos ou estado do paciente; e diagnósticos mais acurados [GELOGO; KIM, 2013]. Os cuidados ubíquos, além de disponibilizarem mais dados e informações, possibilitam que o próprio paciente tenha melhor conhecimento sobre a sua saúde, verificando variações e degradações em seus sinais vitais, servindo assim como um indicador, para que o mesmo tenha maiores e melhores cuidados com a sua saúde.

A ciência da situação, originalmente utilizada pelo setor da aviação, foi definida por Endsley (1995) como a percepção dos elementos de um ambiente dentro de um período de tempo, como a compreensão sobre o significado dos mesmos e assim como a projeção de seus estados em um futuro próximo. Endsley (2015) descreve o processo de análise da situação possuindo três fases. A primeira fase, atua na percepção e reconhecimento de atributos e estados de elementos relevantes do ambiente (localização, altitude e destino próprio e de outras aeronaves). Na segunda fase, compreensão sobre a situação, tem-se a síntese dos elementos percebidos pela primeira (tempo e distância a ser percorrida com o combustível disponível). E a terceira fase, baseia-se na habilidade de projetar ações e prever estados futuros dos elementos do ambiente (projeção de desvios no cronograma; projeção da trajetória própria e de outras aeronaves). Estas projeções e predições são recursos valiosos nas tomadas de decisões. A ciência de situação vem sendo aplicada em diversas áreas da saúde, tais como: monitoramento de pacientes com problemas cardiológicos [DA ROCHA; DA ROSA; RIGHI, 2015], monitoramento de pacientes anestesiados [FIORATOU et al., 2010], apoio a decisão em diagnósticos de pacientes [SINGH et al., 2012], dentre outras.

3. Trabalhos Relacionados

Foram avaliados cinco modelos que utilizam o conceito de monitoramento de pacientes e ciência de situação. O artigo de Don et al. (2013) apresenta uma proposta de modelo para realizar o monitoramento de pacientes remotamente, tendo como foco a utilização da ciência da situação. O modelo toma decisões de modo declarativa, ajudando a equipe médica a compreender a situação.

O trabalho de Delir Haghghi et al. (2014) propõe um modelo de monitoramento móvel da saúde utilizando ciência de situação. O modelo proposto integra o conhecimento de alto nível com os dados sensoriais de baixo nível, utilizando como mecanismo de raciocínio, uma adaptação da Lógica Fuzzy, chamada pelos autores de Inferência Fuzzy de Situação (*Fuzzy Situation Inference*).

O artigo escrito por Wai et al. (2010) apresenta um modelo que utiliza o sensoriamento do ambiente aliado a ciência de situação para monitorar possíveis eventos perigosos, quando pacientes estão deitados ou ao redor da cama ao redor. Para isso, são utilizados sensores de pressão e vídeo. O modelo aplica o conceito de sistema especialista, para realizar as inferências sobre o paciente. A proposta principal do modelo é permitir um monitoramento sobre o bem-estar dos pacientes.

Os autores Rocha, Costa e Righi (2015) apresentam o modelo UbHeart, que tem como proposta o monitoramento de pacientes com insuficiência cardíaca utilizando a ciência de situação para identificar possíveis riscos à saúde do paciente. A recorrência desta patologia, acaba por diminuir a qualidade de vida do paciente assim como influência em maiores gastos para o sistema de saúde, sendo estes alguns dos motivos pela definição da temática do artigo.

O trabalho de Pereira, Costa e Almeida (2013) apresenta um modelo de plataforma para gerenciamento de situação. Plataforma chamada de SCENE, permite a especificação de regras baseadas em situação. O modelo proposto, de acordo com o conceito de ciência de situação de Endsley (1995), encontra-se no nível de Avaliação da Situação, no qual engloba a percepção e a compreensão sobre a situação.

A ciência da situação, conforme apresentado anteriormente, possui três fases: percepção, compreensão e projeção. Nos trabalhos estudados, foi indicado pelos autores a aplicação deste conceito, porém, ao realizar a análise os modelos propostos, se identificou que o terceiro nível da ciência de situação – projeção de futuros estados – não foi implementado em todos os trabalhos. Exceto os artigos apresentados por Rocha, Costa e Right (2015) e Delir Haghighi et al. (2014), apresentaram o terceiro nível da ciência de situação, os outros três trabalhos tiveram a implementação somente dos dois primeiros níveis.

4. Modelo SAC

O modelo SAC – Situation-Aware Care propõe o monitoramento de pacientes, utilizando os sensores e recursos disponíveis nos dispositivos móveis e nos dispositivos vestíveis. Coletando sinais vitais e outros dados, que combinados permitirão ao modelo acompanhar a saúde do paciente, possibilitando assim um maior e melhor conhecimento sobre a saúde do paciente monitorado. O modelo SAC aplica a combinação de lógica Fuzzy e de Redes Neurais Artificiais como método para o raciocínio da ciência de situação. O uso dessas técnicas permite ao modelo identificar e projetar desvios, degradações e problemas na saúde do paciente. O objetivo do modelo é disponibilizar recursos para que hospitais e médicos tenham registros mais apurados sobre o estado de saúde de seus pacientes, permitindo aos mesmos realizarem diagnósticos mais precisos. A Figura 1, expõe a visão geral do modelo.

Foi definida para o modelo SAC uma arquitetura baseada em serviços, conforme pode ser vista pela Figura 2. A escolha por este tipo de arquitetura se deu pelo fato de que a mesma permite a criação de serviços interoperáveis que podem ser facilmente reutilizados e compartilhados entre aplicações [ARSANJANI, 2004]. A seguir serão apresentados cada um dos serviços – Serviços SAC e Serviços Públicos SAC, ambos em nuvem computacional – e clientes disponibilizados pelo modelo SAC, descrevendo sobre as suas funcionalidades e responsabilidades.

4.1. Clientes SAC4Mobile e SAC4Monitors

A aplicação SAC4Mobile, disponibilizada para pacientes e médicos através dos dispositivos móveis, tem como responsabilidade prover as seguintes funcionalidades: coleta dos dados pacientes; cadastro usuários; comunicação entre paciente, médico e hospital; monitoramento dos sinais vitais (pacientes); transmitir dados coletados (pacientes); relatar situação atual (pacientes); e solicitar emergência. Este cliente, tem como principal funcionalidade a captura e envio dos dados coletados através dos dispositivos vestíveis e sensores. Após a coleta os dados são enviados para o serviço público SAC, Concentrador de Dados. Os usuários da aplicação SAC4Mobile, possuem acessos específicos, sendo vinculados ao tipo de usuário – paciente ou médico.

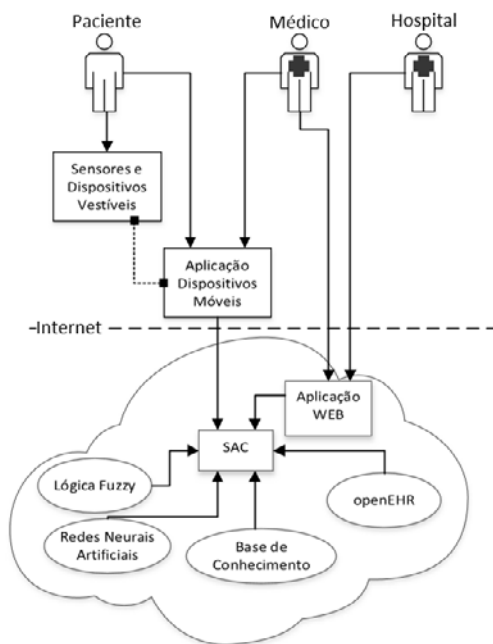


Figura 1. Visão geral modelo SAC

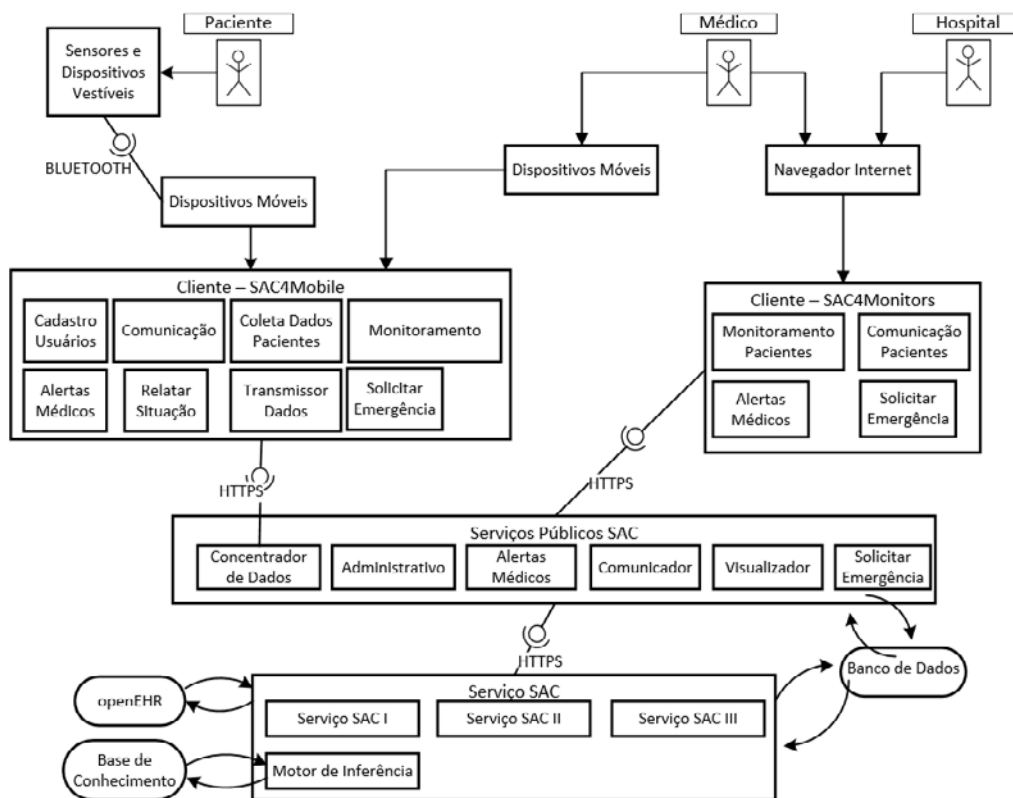


Figura 2. Arquitetura modelo SAC

A aplicação SAC4Monitors, para médicos e hospitais por uma aplicação *web*, permite o acompanhamento dos pacientes, tendo acesso a visualização dos dados coletados e das inferências obtidas. Permite também a comunicação direta com os pacientes, criação de alertas, que são refletidas em notificações, onde utilizam como

referência as projeções inferidas pelo serviço SAC III e também o recurso de acionar o resgate médico para um determinado paciente.

4.2. Serviço SAC I

O serviço SAC I, em conjunto com a aplicação SAC4Mobile e o componente Concentrador de Dados dos serviços públicos SAC, representam a primeira fase da ciência de situação [ENDSLEY, 1995], a percepção dos elementos. Possui as seguintes funcionalidades: **Receptor**, aguarda o envio dos dados pelos Serviços Públicos SAC; **Validador**, valida a integridade dos dados recebidos, desconsiderando ruídos e dados incompletos; **Organizador**, organiza e padroniza os dados segundo openEHR; **Persistência**, armazena os dados, tendo como referência a data e hora da coleta de cada dado.

4.3. Serviço SAC II

O serviço SAC II, proposto pelo modelo SAC, representa a segunda fase da ciência de situação [ENDSLEY, 1995], compreensão sobre a situação atual. Para atingir esta fase da compreensão da situação atual do paciente, foi escolhida a Lógica Fuzzy. A escolha feita devida ao tratamento de imprecisões, a especificação das regras mais próximo a linguagem natural e simplificação na construção da base de conhecimento [ZADEH, 1996]. O serviço possui as seguintes responsabilidades: **Fuzzyficador de Dados**, aplica a fuzzyficação nos dados coletados, convertendo-os em expressões linguísticas; **Regras de Inferência**, constituem a Base de Conhecimento, ou seja, representam o conhecimento do especialista médico. Como resultado, retorna à situação atual do paciente; **Defuzzyficador de Resultados**, componente realiza a fase final da Lógica Fuzzy, convertendo os resultados fuzzyficados em valores numéricos; **Analisador da Situação**, analisa os resultados obtidos, e conforme o risco determinado, realiza ação.

4.4. Serviço SAC III

A terceira fase da ciência de situação, projeção sobre estados futuros, é aplicada pelo serviço SAC III. O modelo SAC utiliza o conceito de predição de Série de Dados Temporais por Redes Neurais Artificiais para atingir o objetivo proposto pela terceira fase da ciência de situação. O treinamento da Rede Neural se dá com a utilização do histórico do paciente observado, permitindo assim aprender sobre como é o comportamento – tendências, sazonalidades, limites, constâncias, dentre outros padrões – dos dados coletados [MITCHELL, 1997]. Para tal, os seguintes componentes foram especificados: **Prognosticador de Dados**, responsável pela implementação da Rede Neural, ou seja, realizará a predição da série de dados temporais referente aos dados dos pacientes; **Projetor de Saúde**, a partir das informações inferidas pelo serviço SAC II, este componente fica responsável por projetar as possíveis degradações na saúde do paciente; **Harmonizador de Dados**, ao fim das inferências sobre as predições, é realizada uma calibração sobre os dados paciente conforme histórico do mesmo.

4.5. Serviço Públicos SAC

Para o acesso aos dados e as informações inferidas, o modelo SAC disponibiliza um pacote de serviços, chamado de Serviços Públicos SAC. Este pacote permite o consumo dos dados e das informações tanto para as aplicações SAC quanto para as aplicações

terceiras. Este pacote de serviços foi criado para separar as responsabilidades do modelo. Atividades que não são vinculadas diretamente com a ciência de situação foram alocadas para os serviços públicos, deixando assim o serviço SAC responsável apenas por aplicar a ciência de situação. Os recursos disponibilizados pelos serviços públicos SAC seriam: **Administrativo**, realiza o gerenciamento administrativo do modelo SAC; **Comunicador**, permite a comunicação direta entre os envolvidos; **Concentrador de dados**, responsável por receber os dados coletados pela aplicação SAC4Mobile e enviar para o serviço SAC I; **Alertas Médicos**, envia uma notificação a partir de um alerta gerado por uma das aplicações; **Visualizador**, permite a visualização dos dados coletados assim como as informações inferidas sobre os pacientes observados; **Solicitar Emergência**, permite a solicitação de resgate médico em caráter de emergência para o paciente observado.

5. Avaliação

A aplicação de cenários, como avaliação de modelos ubíquos vem sendo utilizada pela comunidade científica [QUEVEDO; COSTA; RIGHI; RIGO, 2015] [ROCHA; COSTA; RIGHI, 2015]. É uma ferramenta para visualização de um possível caminho de ação para uma determinada situação, ou seja, ajudam na compreensão do fluxo das aplicações e dos serviços de novas tecnologias imaginadas. Avaliações de cenários também são uma maneira flexível e prática na descoberta de possíveis atitudes e necessidades do usuário. O valor dos cenários está na representação dos conceitos em descrições mais sólidas para efeitos de análise e apresentação [IKONEN, 2002]. Para o modelo SAC foi realizada a seguinte aplicação de cenário:

“Maria, jovem de 24 anos, mora sozinha e tem o hábito de praticar corridas. Maria trabalha como modelo fotográfica, logo a boa aparência e saúde são seus aliados. Pensando na sua saúde, Maria começou a utilizar a aplicação SAC4Mobile para monitorar o seu dia a dia, em conjunto com um dispositivo vestível para monitorar a temperatura e uma smartband capaz de coletar a frequência cardíaca. Maria indicou o seu médico particular como referência médica em casos de alerta, dessa forma ele recebe toda e qualquer notificação referente a sua saúde. Durante a prática das corridas, o SAC4Mobile reconhece que ela está se exercitando e por isso considera que a elevação nos batimentos cardíacos está de acordo, assim como o aumento da temperatura. Maria, há dois atrás realizou a sua corrida diária, porém, desta vez foi surpreendida por uma chuva durante a atividade. Após dois dias, o SAC4Mobile apresenta a notificação de um aumento na temperatura de Maria, e durante o dia a mesma reportou estar sentindo dores no corpo e na cabeça. Com essas novas informações enviadas ao serviço SAC, o mesmo indica possível gripe, como projeção para a saúde de Maria. Médico recebe alerta referente a alteração na saúde de Maria, e o mesmo lhe encaminha mais questionamentos sobre o estado atual e também sobre a rotina de Maria durante os dias anteriores, a mesma reporta que foi surpreendida pela chuva. Em posse das inferências apresentadas pelo serviço SAC, o médico recomenda a compra de antigripais e repouso, pois, ele tem a mesma opinião que o SAC, de que Maria pegou uma gripe. Maria atua conforme recomendado, pois, em quatro dias precisará estar em plena saúde devido a um trabalho fotográfico”.

Tendo em vista o cenário acima apresentado, foi então desenvolvido o protótipo da aplicação SAC4Mobile para a plataforma Android, assim como os Serviços Públicos

SAC e os Serviços SAC, estes dois hospedados na nuvem, Amazon EC2. O monitoramento dos sinais vitais é realizado através de um dispositivo vestível que monitora a temperatura corporal do paciente – desenvolvido utilizando Arduino LilyPad e sensores de temperatura – e de uma *smartband* que monitora a frequência cardíaca – FitBit Charge HR. Nos protótipos desenvolvidos, o modelo SAC objetivou o menor desconforto ao utilizador da solução. A Figura 3, apresenta o protótipo SAC4Mobile.



Figura 3. Protótipo SAC4Mobile

Durante as práticas das corridas de Maria, o modelo SAC, compreendeu e projetou, que a situação dela se tratava de uma atividade física e que o aumento na frequência cardíaca e na temperatura estavam de acordo. Entretanto conforme foi apresentado, a paciente realizou atividades em um período chuvoso, durante o monitoramento dos dias seguintes, foram compreendidos e projetados desvios – aumento – na temperatura. Esta projeção aliada aos reportes de situação da paciente, onde a mesma indicava dores no corpo e na cabeça, inferiram na situação de gripe. Ao projetar o desvio na saúde da paciente, é disparada uma notificação tanto para o médico quanto para a paciente monitorada. Com a notificação, permite hospitais e médicos estarem cientes dos desvios e também a realizar alguma forma de intervenção médica. E com a intervenção realizada pelo médico de Maria, permitiu que a mesma iniciasse o seu tratamento sem mesmo ter realizado uma consulta médica presencial.

Com o desenvolvimento dos protótipos acima descritos e avaliação de cenário, foi então solidificada as expectativas quanto ao modelo SAC e a utilização de ciência de situação para a projeção de futuros estados da saúde. Onde o modelo SAC, através de suas aplicações, monitoramento e serviços, perceberam os desvios nos sinais vitais monitorados, e em conjunto com os reportes da paciente, pôde projetar a tendência dos desvios e tendo por conclusão inferido a situação da paciente. Se apresentando então, como uma solução relevante em relação aos cuidados ubíquos disponibilizando recursos – para médicos e para pacientes – que possibilitam melhores cuidados para com a saúde dos pacientes monitorados e também tomadas de decisões mais acertivas.

6. Conclusão

O modelo SAC realiza o monitoramento de sinais vitais de pacientes aliado a aplicação das três fases da ciência de situação sobre dados obtidos. A ciência de situação permite

ao modelo realizar inferências mais acertivas quanto ao estado ou a situação em que o paciente se encontra, pois, compreende os elementos e suas relações, assim como projeta futuros estados do paciente. Sendo estas inferências, recursos que permitem um maior e um melhor conhecimento sobre a saúde dos pacientes monitorados, possibilitando melhores condições para as tomadas de decisão dos médicos. E o protótipo desenvolvido solidificou as expectativas quanto ao modelo, identificando as situações dos pacientes e realizando as devidas notificações aos envolvidos, permitindo assim que os médicos tenham um tempo de resposta muito rápido em casos de desvios e degradações da saúde. Demonstrando a sua relevância para o domínio dos cuidados ubíquos e também como sendo uma solução de aplicação possível.

Para trabalhos futuros pretende-se incorporar mais dados dos pacientes e consequentemente incluir novas regras de inferências, permitindo assim uma maior abrangência do modelo SAC. Outra perspectiva para o futuro é avaliar o modelo proposto quanto a usabilidade das aplicações e o quanto ao desempenho das inferências, buscando realizar melhorias no modelo SAC.

Agradecimentos

Os autores gostariam de agradecer o CAPES e ao CNPq pelo apoio a esta pesquisa.

Referências

- ARSANJANI, Ali. Service-oriented modeling and architecture. IBM developer works, p. 1-15, 2004.
- DELIR HAGHIGHI, Pari et al. Situation-aware mobile health monitoring. Em: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014. p. 248-256.
- DON, S. et al. An awareness approach to analyze ECG streaming data. Journal of medical systems, v. 37, n. 2, p. 1-12, 2013.
- ENDSLEY, Mica R. Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, v. 37, n. 1, p. 32-64, 1995.
- ENDSLEY, Mica R. Situation awareness misconceptions and misunderstandings. Journal of Cognitive Engineering and Decision Making, v. 9, n. 1, p. 4-32, 2015.
- FIORATOU, E. et al. Beyond monitoring: distributed situation awareness in anaesthesia. British journal of anaesthesia, v. 105, n. 1, p. 83-90, 2010.
- FRIEDEWALD, Michael; RAABE, Oliver. Ubiquitous computing: An overview of technology impacts. Telematics and Informatics, v. 28, n. 2, p. 55-65, 2011.
- GELOGO, Yvette E.; KIM, Haeng-Kon. Unified Ubiquitous Healthcare System Architecture with Collaborative Model. International Journal of Multimedia and Ubiquitous Engineering, v. 8, n. 3, p. 239-244, 2013.

- GONZÁLEZ-FERRER, Arturo; PELEG, Mor. Understanding requirements of clinical data standards for developing interoperable knowledge-based DSS: A case study. *Computer Standards & Interfaces*, v. 42, p. 125-136, 2015.
- IKONEN, Veikko; RENTTO, Katja. Scenario Evaluation for Ubiquitous Computing-Stories Come True. Em: Position paper for the Ubicomp 2002 conference workshop Evaluation Methods for Ubiquitous Computing. Goteborg, Sweden. 2002.
- MEULEN, Rob van der; WOODS, Viveca. Gartner Says Global Smartphone Sales to Only Grow 7 Per Cent in 2016. 2016. Gartner. Disponível em: <<http://www.gartner.com/newsroom/id/3270418>>. Acesso em: 11 abril 2016.
- MITCHELL, Tom M. *Machine Learning*. 1997.
- PATEL, Mitesh S.; ASCH, David A.; VOLPP, Kevin G. Wearable devices as facilitators, not drivers, of health behavior change. *Jama*, v. 313, n. 5, p. 459-460, 2015.
- PEREIRA, Isaac SA; COSTA, Patrícia Dockhorn; ALMEIDA, João Paulo A. A rule-based platform for situation management. Em: Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2013 IEEE International Multi-Disciplinary Conference on. IEEE, 2013. p. 83-90.
- QUEVEDO, N. M. M. ; COSTA, C. A. da ; RIGHI, R. R. ; RIGO, Sandro J. . Um Modelo Ubíquo de Detecção de Alergia Baseado em Ciência de Situação. Em: 7o Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2015), 2015, Recife. Anais do 7o Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2015). Porto Alegre: SBC, 2015. p. 31-40.
- ROCHA, C. C. L. ; COSTA, C. A. da ; RIGHI, R. R. . Um modelo para monitoramento de sinais vitais do coração baseado em ciência da situação e computação ubíqua. Em: 7o Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2015), 2015, Recife. Anais do 7o Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP 2015). Porto Alegre: SBC, 2015. p. 51-60.
- SATYANARAYANAN, Mahadev. Mobile computing: the next decade. Em: Proceedings of the 1st ACM workshop on mobile cloud computing & services: social networks and beyond. ACM, 2010. p. 5.
- SINGH, Hardeep et al. Exploring situational awareness in diagnostic errors in primary care. *BMJ quality & safety*, v. 21, n. 1, p. 30-38, 2012.
- WAI, Aung Aung Phyo et al. Situation-aware patient monitoring in and around the bed using multimodal sensing intelligence. Em: Intelligent Environments (IE), 2010 Sixth International Conference on. IEEE, 2010. p. 128-133.
- WEISER, Mark. The computer for the 21st century. *Scientific american*, v. 265, n. 3, p. 94-104, 1991.
- ZADEH, Lotfi A. Fuzzy logic = computing with words. *Fuzzy Systems, IEEE Transactions on*, v. 4, n. 2, p. 103-111, 1996.

Um modelo proativo de antecipação de ações de times de resposta rápida baseado em análise preditiva

Fabio de Oliveira Dias^{1,2}, Cristiano André da Costa¹, Rodrigo da Rosa Righi¹

¹Programa de Pós-Graduação em Computação Aplicada (PIPCA)
Universidade do Vale do Rio do Sinos (UNISINOS)

²Instituto Federal Sul-rio-grandense (IFSul)

fabio@sapucaia.ifsul.edu.br, {rrrighi, cac}@unisinis.br

Abstract. *A Rapid Response Team intends to prevent deaths in patients who have clinical deterioration outside of intensive care units in hospitals environments. The Predictvs model seeks to anticipate actions of rapid response teams, through the analysis of vital signs of patients with the use of early warning scores. The scientific contribution of the presented model is that we could better predict possible collapse situations of patients, through the monitoring and analysis of vital signs. The Predictvs evaluation was made using scenarios and their analysis showed encouraging results.*

Resumo. *Um Time de Resposta Rápida busca prevenir mortes de pacientes que tenham piora clínica fora de ambientes de Unidades de Tratamento Intensivo em hospitais. O modelo Predictvs busca antecipar ações dos times de resposta rápida, através da análise dos sinais vitais dos pacientes com o uso de escores de alerta precoce. A contribuição científica do modelo é dada em virtude da possibilidade de efetuar a predição de possíveis situações de colapso dos pacientes através do monitoramento e análise dos sinais vitais. A avaliação do Predictvs foi efetuada com a utilização de cenários e a sua análise apresentou resultados que motivam a continuidade da pesquisa.*

1. Introdução

Em dezembro de 2004, o fundador e CEO do *Institute for Healthcare Improvement* (IHI), uma organização independente e sem fins lucrativos que busca promover melhorias nos cuidados médicos, Dr. Don Berwick, lançou um desafio na reunião anual da entidade, ocorrida em Orlando. Impaciente com o ritmo das mudanças relacionadas à segurança dos pacientes e com objetivos vagos da área da saúde, ele anunciou uma campanha para salvar 100 mil vidas nos hospitais nos próximos 18 meses [Watcher and Pronovost 2006]. Para atingir esse audacioso objetivo, foi proposta a adoção de seis práticas pelas instituições de saúde. Dentre essas práticas estava a sugestão de implantação de Times de Resposta Rápida (*Rapid Response Teams*, RRTs) nos hospitais. Passados os 18 meses, o IHI anunciou que o objetivo inicial havia sido ultrapassado, tendo sido salvas 122.300 vidas [Watcher and Pronovost 2006]. Posteriormente, uma extensão dessa campanha foi lançada, propondo salvar 5 milhões de vidas entre dezembro de 2006 e dezembro de 2008 [McCannon et al. 2007].

Um Time de Resposta Rápida consiste em uma equipe de profissionais de saúde liderada por um médico, com a finalidade de prevenir mortes de pacientes que tenham piora clínica fora de um ambiente preparado para atendê-los em uma eventual situação crítica (por exemplo, pacientes em enfermarias). Para tanto, é necessária a análise de

sinais vitais dos pacientes, de forma a prevenir ou detectar um eventual colapso dos mesmos [Berwick et al. 2006].

A computação móvel e ubíqua tem propiciado o advento de soluções que permitem o monitoramento em tempo real de sinais advindos de sensores e o seu processamento por aplicações que podem executar ações de acordo com as condições encontradas. Esta característica possibilita o uso da computação móvel e ubíqua para o monitoramento de condições de saúde de pacientes, denominado de cuidados ubíquos (*ubiquitous healthcare*) [Gelogo et al. 2013]. Indo além, provê, inclusive, condições para que possa ser efetuada a predição de eventuais pioras clínicas e o acionamento de profissionais de saúde para prestar o adequado socorro, permitindo o salvamento de vidas.

Este trabalho se insere nestas condições, estando voltado para a antecipação de ações de times de resposta rápida baseado em análise preditiva, com o uso de escores de alerta precoce, propondo o modelo Predictvs. A Figura 1 apresenta o cenário em que o modelo proposto se encaixa, demonstrando a transmissão dos dados advindos dos sensores e o envio de sinais de alerta aos profissionais de saúde e ao Time de Resposta Rápida.

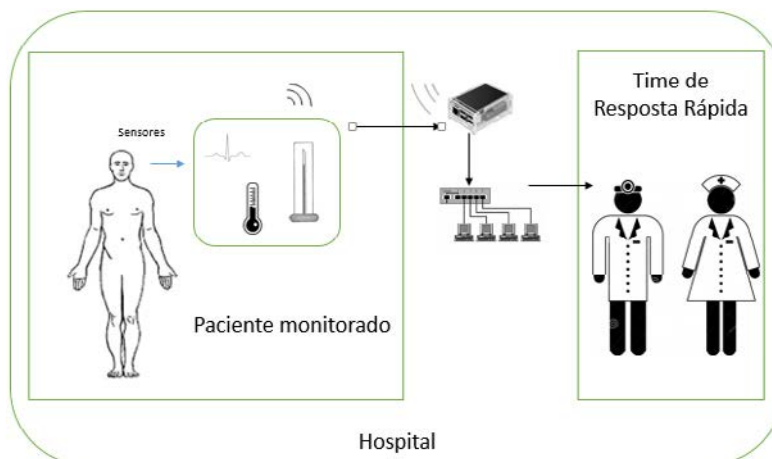


Figura 1. Cenário de utilização do modelo Predictvs

A principal contribuição científica do trabalho decorre da possibilidade de, através do monitoramento dos sinais vitais com o uso de escores de alerta precoce, prever possíveis situações de colapso dos pacientes e antecipar o acionamento dos times de resposta rápida. Não foram encontrados, durante as pesquisas de trabalhos relacionados, modelos com essa característica. Os trabalhos relacionados demonstraram que o uso de escores de alerta precoce são uma importante ferramenta utilizada pelos profissionais de saúde para efetuar a medição do estado clínico dos pacientes, mas as totalizações geralmente ocorrem de forma manual, em formulários de papel situados junto ao prontuário dos pacientes [Morgan et al. 1997; Kyriacos et al. 2014]. O modelo propõe a automatização da totalização da pontuação do escore de alerta precoce, possibilitando a antecipação do acionamento dos times de resposta rápida.

Este trabalho foi dividido em seis seções. Após a introdução, a seção 2 descreve, brevemente, a relação proposta entre escores de alerta precoce com os times de resposta rápida. A seção 3 descreve a arquitetura do modelo Predictvs. A seção 4 apresenta uma avaliação do modelo baseada em cenários. Trabalhos relacionados são apresentados na seção 5. Finalmente, as conclusões são apresentadas na seção 6.

2. Escores de Alerta Precoce

Em 1997, Morgan, Williams e Wright [Morgan et al. 1997] foram os primeiros a desenvolver e publicar um escore de alerta precoce (EWS) no Reino Unido. Este EWS fazia referência a cinco parâmetros fisiológicos: frequência cardíaca, pressão arterial sistólica, frequência respiratória, temperatura e nível de consciência. Cada parâmetro possui uma gama de pontos de corte com gatilhos correspondentes, não para efetuar a predição, mas para servir como um sistema de rastreamento e gatilho para identificar os primeiros sinais de deterioração dos pacientes. Por exemplo, uma frequência cardíaca entre 111 e 129 bpm (batimentos por minuto) recebe um escore de 2 pontos, indicando a necessidade de escalada para intervenção pela equipe médica.

Desde então, os sistemas de EWS foram modificados (sendo denominados MEWS, escore de alerta precoce modificado) e padronizados por todo o Reino Unido. Em adição aos cinco parâmetros fisiológicos citados, outros foram incorporados na maioria dos MEWS, tais como nível de saída da urina e oxigenação do sangue. Como não existe um MEWS ideal, a multiplicidade destes sistemas resultou em uma falta de consistência no reconhecimento e na resposta a deterioração clínica, tornando necessária a sua padronização. Em julho de 2012 foi implementado um escore de alerta precoce nacional (NEWS), com o monitoramento de seis parâmetros fisiológicos (frequência respiratória, saturação de oxigênio, temperatura, pressão arterial sistólica, frequência cardíaca e nível de consciência), porém a adoção deste não se tornou obrigatória nos hospitais [Kyriacos et al. 2014]. A Tabela 1 mostra um exemplo de tabela de escore de alerta precoce [Kyriacos et al. 2014].

Sinal fisiológico/Escore	3	2	1	0	1	2	3
Frequência respiratória/min		≤ 9		9-14	15-20	21-29	≥ 30
Saturação O ₂	< 85	85-89	90-92	≥ 93			
Frequência cardíaca/min		≤ 40	41-50	51-100	101-110	111-129	≥ 130
Pressão sistólica	≤ 70	71-80	81-100	101-199		≥ 200	
Temperatura °C		≤ 35		35-38.4		≥ 38.5	
Estado neurológico Escala de Coma Glasgow				15	14	13-9	≤ 8
Estado neurológico Escala AVPU				Alerta	Reage à voz	Reage à dor	Sem resposta
Urina ml/kg/hora	0.5ml/kg /1 hora		≤ 1 ml/kg/1 hora		≥ 3 ml/kg/1 hora		

Tabela 1. Exemplo de escore de alerta precoce [Kyriacos et al. 2014]

A análise de sinais vitais de pacientes permite monitorar possíveis ocorrências para acionar os times de resposta rápida. As ocorrências normalmente são classificadas em duas chamadas: Código Amarelo para urgências e Código Azul para emergências - PCR (Parada Cardiorrespiratória). Estudos comprovam a eficiência da abordagem, reduzindo a quantidade de óbitos nessas condições de forma significativa [Offner et al. 2007; Gonçalves et al. 2012]. Abaixo são descritos alguns exemplos de sinais de alerta que podem disparar as ações dos Times de Resposta Rápida:

- Alteração súbita na frequência cardíaca, ficando abaixo de 40 ou acima de 130 bpm;

- Alteração súbita na pressão arterial sistólica, ficando menor que 90 mmHg (milímetros de mercúrio);
- Alteração súbita na frequência respiratória, estando abaixo de 8 ou acima de 28 ipm (inspirações por minuto), ou com comprometimento de via aérea;
- Alteração súbita de nível de consciência [Offner et al. 2007].

3. Modelo Predictvs

O modelo proposto permite o acionamento do Time de Resposta Rápida do hospital em virtude da predição da deterioração do estado clínico do paciente através do uso de escores de alerta precoce. Conforme [Bailey et al 2013], o tempo é um fator essencial para intervenções médicas em situações de urgência e emergência, de forma a prevenir o agravamento do estado clínico dos pacientes. Partindo dessa premissa, o modelo Predictvs foi planejado para ter uma arquitetura simples, de forma a dispender o tempo mínimo no tratamento dos sinais fisiológicos, porém eficiente, provendo os alertas necessários e adequados para os times de resposta rápida. A Figura 2 exibe a arquitetura do modelo Predictvs.

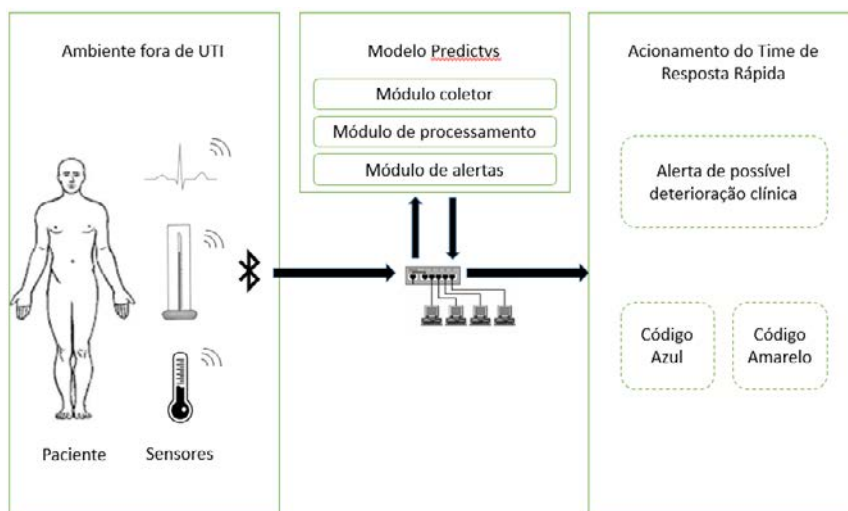


Figura 2. Arquitetura do modelo Predictvs

Inicialmente, o perfil do paciente deve ser informado ao sistema, bem como o tipo de escore de alerta precoce que será utilizado. O modelo permite o uso de múltiplos escores de alerta precoce, que devem ser cadastrados previamente pelo médico responsável. Os *thresholds* dos sinais vitais presentes no escore de alerta precoce atribuído pelo médico também podem ser personalizados, de forma a refletir eventuais peculiaridades no estado clínico do paciente, permitindo uma predição mais apurada.

Sensores acoplados aos pacientes efetuam a leitura de seus sinais fisiológicos (tais como frequência cardíaca, frequência respiratória, pressão arterial sistólica e temperatura) e os transmitem para um módulo coletor via Bluetooth. O módulo coletor consiste em um dispositivo, conectado à rede local da instituição, que possui a finalidade de transmitir os dados lidos ao módulo de processamento do modelo. O modelo suporta múltiplos módulos coletores, provendo escalabilidade e redundância ao sistema.

O módulo de processamento, através da função de *análise da leitura efetuada*, verifica se o sinal fisiológico foi detectado corretamente ou se houve alguma discrepância,

que leva ao descarte do dado e à solicitação de nova leitura. Em seguida, a função de *pontuação do escore de alerta precoce* faz o cálculo da pontuação dos sinais vitais lidos, com base no escore de alerta precoce configurado pelo médico. Escores de até 3 pontos geram alertas do tipo *warning* para o Time de Resposta Rápida, indicando possível necessidade de intervenção; escores acima deste valor geram alertas do tipo *critical* contendo um Código Azul (se houve parada cardiorrespiratória) ou um Código Amarelo (se houve outro tipo de colapso), obrigando o acionamento da equipe. A Figura 3 exibe a arquitetura do módulo de processamento.

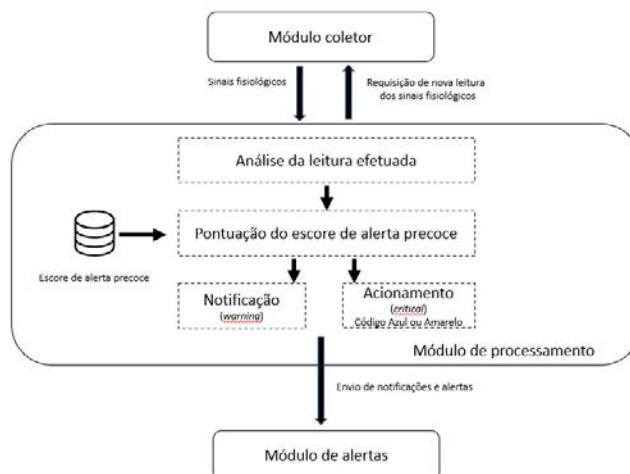


Figura 3. Módulo de processamento

O módulo de alertas é responsável por transmitir estes avisos, através da função *envio de alertas*. De forma sucinta, o modelo prevê que os profissionais de saúde devem assinar uma fila de tópicos de mensagens para receber os alertas do Predictvs. Assim, o módulo de alertas efetua a transmissão da mensagem e, de acordo com as condições do paciente, os integrantes do time de resposta rápida recebem as notificações e os acionamentos.

4. Avaliação

A utilização de cenários tem sido empregada pela comunidade científica para a avaliação de aplicações ubíquas [Satyanarayanan, 2011; Rocha, 2015]. Desta forma, foi construído um cenário para a avaliação do modelo apresentado na seção anterior. O cenário demonstra a atuação do Predictvs na predição da deterioração clínica dos pacientes, com base nos sinais vitais coletados. O cenário utilizado foi:

“Júlia, por requisição de seu médico, em virtude de seu histórico de problemas cardíacos, vai ao centro clínico de seu hospital para que seja realizado um ecodopplercardiograma bidimensional com mapeamento de fluxo em cores, um importante exame para pacientes que possuem doenças relacionadas ao coração. Ao chegar, Júlia recebe uma série de sensores, que são acoplados ao seu corpo e que se comunicam com um dispositivo coletor Raspberry Pi conectado à rede local do hospital. Tais sensores efetuam, periodicamente, a leitura de sua frequência cardíaca, frequência respiratória, pressão arterial sistólica e temperatura e repassam os dados ao dispositivo coletor, que os retransmite ao módulo de processamento do sistema Predictvs. Júlia realiza o exame e é deixada em repouso na enfermaria. Em determinado momento, o Predictvs detecta que seus sinais vitais começam a se deteriorar, com seu escore de alerta

precoce atingindo a pontuação 3. Com base nisto, o módulo de alertas do sistema envia um alerta do tipo *warning* ao Time de Resposta Rápida, indicando tal situação. No entanto, a deterioração de seus sinais vitais torna-se ainda mais aguda (com a pontuação de seu escore de alerta precoce chegando a 7) e o sistema Predictvs emite um novo alerta, agora do tipo *critical*, contendo um Código Azul, acionando os integrantes do Time de Resposta Rápida do hospital. Assim, os devidos cuidados são aplicados à paciente.”

Para que a avaliação do cenário exposto pudesse ser efetuada, um protótipo baseado em Python, com uma interface web JavaScript, foi desenvolvido. O módulo de processamento recebe os sinais vitais através do uso das filas de tópicos do middleware RabbitMQ. Após, efetua a análise dos sinais e atribui um escore ao paciente. Quando necessário, o módulo de alertas envia mensagens, também através do uso das filas de tópicos do middleware RabbitMQ, ao Time de Resposta Rápida do hospital, avisando-o sobre possível colapso ou acionando-o em caso de ocorrência de um Código Azul ou Código Amarelo.

Predictvs

Mensagens

```
18/04/2015 07:41:02 - Paciente: Julia - Sinal fisiologico: Pressao arterial sistolica - Medicao: 120 - MEWS: 0
18/04/2015 07:41:03 - Paciente: Julia - Sinal fisiologico: Temperatura - Medicao: 36.7 - MEWS: 0
MEWS acumulado: 0

18/04/2015 07:42:00 - Paciente: Julia - Sinal fisiologico: Frequencia cardiaca - Medicao: 49 - MEWS: 1
18/04/2015 07:42:01 - Paciente: Julia - Sinal fisiologico: Frequencia respiratoria - Medicao: 12 - MEWS: 0
18/04/2015 07:42:02 - Paciente: Julia - Sinal fisiologico: Pressao arterial sistolica - Medicao: 110 - MEWS: 0
18/04/2015 07:42:03 - Paciente: Julia - Sinal fisiologico: Temperatura - Medicao: 36.7 - MEWS: 0
MEWS acumulado: 1
Notificacao Time de Resposta Rapida - warning

18/04/2015 07:43:00 - Paciente: Julia - Sinal fisiologico: Frequencia cardiaca - Medicao: 51 - MEWS: 0
18/04/2015 07:43:01 - Paciente: Julia - Sinal fisiologico: Frequencia respiratoria - Medicao: 18 - MEWS: 1
18/04/2015 07:43:02 - Paciente: Julia - Sinal fisiologico: Pressao arterial sistolica - Medicao: 81 - MEWS: 1
18/04/2015 07:43:03 - Paciente: Julia - Sinal fisiologico: Temperatura - Medicao: 36.6 - MEWS: 0
MEWS acumulado: 2
Notificacao Time de Resposta Rapida - warning

18/04/2015 07:44:00 - Paciente: Julia - Sinal fisiologico: Frequencia cardiaca - Medicao: 35 - MEWS: 2
18/04/2015 07:44:01 - Paciente: Julia - Sinal fisiologico: Frequencia respiratoria - Medicao: 21 - MEWS: 2
18/04/2015 07:44:02 - Paciente: Julia - Sinal fisiologico: Pressao arterial sistolica - Medicao: 70 - MEWS: 3
18/04/2015 07:44:03 - Paciente: Julia - Sinal fisiologico: Temperatura - Medicao: 36.6 - MEWS: 0
MEWS acumulado: 7
Acionamento Time de Resposta Rapida - critical
ALERTA - Frequencia cardiaca BAIXA - CODIGO AZUL
```

Logs

```
Opening Web Socket...
Web Socket Opened...
>>> CONNECT login:rrt passcode:unisinos host:/ accept-version:1.1,1.0 heart-beat:0,0
<<< CONNECTED session:cmk-2gR7viyv7QpOP2f0u heart-beat:0,0 server:RabbitMQ/3.5.6 version:1.1
connected to server RabbitMQ/3.5.6
>>> SUBSCRIBE id:sub-0 destination:/exchange/rrt/#
<<< MESSAGE subscription:sub-0 destination:/exchange/rrt/Julia message-id:T_sub-0@session-cmk-2gR7viyv7QpOP2f0u@1 redelivered:false content-length:37
Julia:18/04/2015:07:41:00;fc:1;60;0;0
<<< MESSAGE subscription:sub-0 destination:/exchange/rrt/Julia message-id:T_sub-0@session-cmk-2gR7viyv7QpOP2f0u@2 redelivered:false content-length:37
Julia:18/04/2015:07:41:01;fr:1;12;0;0
<<< MESSAGE subscription:sub-0 destination:/exchange/rrt/Julia message-id:T_sub-0@session-cmk-2gR7viyv7QpOP2f0u@3 redelivered:false content-length:39
Julia:18/04/2015:07:41:02;pas:1;120;0;0
```

Figura 4. Protótipo do Predictvs em ação

A utilização do Predictvs requer o cadastro de algumas informações relacionadas ao paciente. Basicamente deve ser informada a identificação do paciente e o perfil de escore de alerta precoce a ser utilizado. O Predictvs é extremamente customizável. Além do escore de alerta precoce ser adaptável, permitindo a criação e utilização de múltiplos escores de alerta precoce modificados (MEWS), os profissionais de saúde podem atribuir os *thresholds* que julgarem apropriados em cada item do escore. A Figura 4 apresenta a tela do módulo de processamento em ação, demonstrando o escore MEWS calculado a cada ciclo de leitura dos sensores, bem como as notificações e o acionamento do time de resposta rápida. Os parâmetros de escore de alerta precoce utilizados são os descritos na Tabela 1.

Assim que foram acoplados em seu corpo, os sensores passaram a reportar, via Bluetooth, os dados de frequência cardíaca, frequência respiratória, pressão arterial sistólica e temperatura de Júlia. Em determinado momento, logo após Júlia ter realizado o exame e ter sido colocada em repouso, sua frequência cardíaca estava em 51 bpm, sua frequência respiratória estava em 18 ipm e sua pressão arterial sistólica estava em 81 mmHg. O módulo de pontuação do Predictvs atribuiu 2 pontos a essa condição clínica, o que disparou um alerta de possível colapso ao Time de Resposta Rápida. Entretanto, poucos minutos depois, as condições clínicas de Júlia deterioraram-se ainda mais: sua frequência cardíaca baixou para 35 bpm, sua frequência respiratória subiu para 21 ipm e sua pressão arterial sistólica baixou para 70 mmHg. Neste caso, a condição clínica de Júlia recebeu 7 pontos do módulo de pontuação, o que gerou o acionamento do Time de Resposta Rápida do hospital com um Código Azul. Assim, Júlia recebeu os devidos cuidados médicos, sendo transferida para a Unidade de Tratamento Intensivo do hospital e teve a sua vida salva.

5. Trabalhos relacionados

Foram considerados cinco trabalhos que possuem características semelhantes ao modelo proposto. O trabalho de [Kirkland et al. 2013] propõe o desenvolvimento de uma ferramenta para predição da deterioração clínica de pacientes não monitorados durante seu tempo de permanência em um hospital, utilizando, para tanto, variáveis procedentes de análise de regressão. A análise final de regressão multivariada demonstrou que as variáveis Escala de Braden, frequência respiratória, saturação de oxigênio arterial e índice de choque hipovolêmico foram associadas de forma independente aos eventos. Frequência cardíaca e temperatura foram eliminadas do modelo final, apesar da primeira já estar representada através do índice de choque hipovolêmico.

O artigo de [Kyriacos et al. 2014] procura desenvolver e validar, por consenso entre especialistas da área de saúde, a construção e o conteúdo de uma tabela de observações para enfermeiros, incorporando um escore de alerta precoce modificado (MEWS) para o monitoramento de pacientes em um hospital público na África do Sul. O MEWS se baseou em sete parâmetros fisiológicos (frequência respiratória, frequência cardíaca, pressão arterial sistólica, temperatura, nível de consciência, saturação de oxigênio – SaO₂ e nível da saída de urina) e em uma tabela de observações.

[Brady et al. 2013] apresentam um sistema que busca identificar e mitigar riscos de colapsos de pacientes utilizando princípios de organizações de alta confiabilidade, como os empregados em usinas nucleares e empresas de aviação comercial. Estas organizações lidam com riscos constantes e mantêm recordes de segurança exemplares. Foram analisados eventos adversos sérios mais recentes e as transferências para a UTI. Cinco fatores de risco foram associados com cada evento: familiares, terapias de alto risco, presença de alto escore de alerta precoce (EWS), problemas intestinais e de comunicação. Usando o modelo como base para melhorias, uma intervenção foi desenvolvida e testada para identificar de forma proativa o risco de deterioração do estado clínico dos pacientes. Após a intervenção, as taxas de transferência para a UTI em virtude de eventos de falha de contextos não reconhecidos reduziu de forma significativa de 4,4 para 2,4 durante o período do estudo.

O objetivo do trabalho de [Badriyah et al. 2013] é comparar a performance de um escore de alerta precoce nacional (NEWS), gerado por humanos e otimizado por tentativa

e erro, com um escore de alerta precoce baseado em árvore de decisão (DTEWS), gerado totalmente por meios computacionais. Os autores avaliaram a habilidade do DTEWS de discriminar pacientes com risco de parada cardíaca, admissão imprevista em UTI ou morte, cada um dentro de 24 horas a partir de uma dada observação de sinais vitais. Foram comparadas as performances de DTEWS e NEWS utilizando a área sob a curva característica de operação do receptor (curva ROC), tal como no primeiro artigo estudado. Os autores concluem o trabalho afirmando que a técnica de árvore de decisão valida de forma independente o padrão NEWS. A abordagem por árvore de decisão provê rapidamente um escore de alerta precoce quase idêntico ao NEWS.

O artigo de [Bailey et al. 2013] mostra um ensaio de predição em tempo real com pacientes em enfermarias. O objetivo central do estudo era validar um algoritmo preditivo da piora clínica em pacientes da ala médica geral e conduzir um teste de alertas em tempo real baseados neste algoritmo. Esses alertas em tempo real foram gerados por um algoritmo projetado para prever a necessidade de transferência para UTI usando dados eletronicamente disponíveis. Os alertas eram enviados ao enfermeiro chefe da ala. Para tanto, foi desenvolvida uma ferramenta de predição em tempo real, utilizando o particionamento recursivo da análise da árvore de regressão para identificar iminentes infecções. Essa implementação com base na ferramenta resultou em aumento de intervenções antecipadas, incluindo escalada de antibióticos, fluídos intravenosos, oxigenoterapia e diagnósticos em pacientes identificados como de risco.

Tabela 2. Comparação de características entre os trabalhos relacionados

Características		Trabalhos				
		Kirkland et al., 2013	Kyriacos et al., 2014	Brady et al., 2013	Badriyah et al., 2013	Bailey et al., 2013
Técnica utilizada		Escala de Braden Semi-Markov	Sistema de escores <i>Thresholds</i>	Algoritmo de ciência de situação	Árvore de decisão	Regressão logística
Variáveis utilizadas no escopo do trabalho	Pressão sistólica	•	•		•	•
	Pressão diastólica	•			•	•
	Frequência cardíaca	•	•		•	•
	Índice de choque hipovolêmico	•				•
	Frequência respiratória	•	•		•	•
	Frequência do pulso				•	
	Temperatura	•	•		•	
	Saturação O ₂	•	•		•	•
	Nível de consciência		•		•	
	Cinco fatores de risco: familiares, terapias de alto risco, presença de alto EWS, problemas intestinais e de comunicação.			•		
	Escala de Braden	•				
	Medição do nível de urina		•			
Suporte computacional descrito no trabalho		Não	Não	Sim	Sim	Sim
Faz uso de ciência de situação		Não	Não	Sim	Não	Não
Padrões utilizados		Baseado em EWS	MEWS	EWS e PEWS	NEWS e DTEWS	Atribuição de escores baseados em EWS
Elaboração de protótipo		Não	Não	Não	Não	Sim
Modelo proposto relaciona-se com Times de Resposta Rápida		Não	Não	Não	Não	Não

A Tabela 2 apresenta um comparativo entre os trabalhos relacionados. O modelo proposto neste trabalho procura preencher as lacunas encontradas, permitindo o uso de praticamente qualquer medição de sinais fisiológicos que possa ser efetuada com o uso de sensores, sendo compatível com múltiplos escores de alerta precoce e propiciando, como explanado nas seções anteriores, o acionamento de times de resposta rápida.

6. Conclusão

O presente trabalho apresentou o modelo Predictvs, que permite antecipar ações dos times de resposta rápida dos hospitais através do monitoramento dos sinais vitais dos pacientes utilizando escores de alerta precoce. O protótipo apresentado mostrou, no cenário efetuado, um competente instrumento na tarefa de efetuar a análise de sinais vitais advindos de sensores dos pacientes, de forma a prevenir ou detectar um eventual colapso dos mesmos e, nos casos pertinentes, enviar os devidos alertas ao Time de Resposta Rápida.

Como trabalhos futuros, pretende-se desenvolver um protótipo que possa ser aplicado em diferentes situações reais. Para tanto, planeja-se a realização de estudo de caso em ambiente monitorado, de forma a analisar e comprovar a eficiência da modelo.

Agradecimentos

Os autores gostariam de agradecer ao CNPq pelo apoio a esta pesquisa.

Referências

- Ameyed D., Miraoui M., Tadj C. (2015). “A Survey of Prediction Approach in Pervasive Computing”. *International Journal of Scientific & Engineering Research*.
- Berwick D. et al. (2006). “The 100,000 Lives Campaign: Setting a goal and a deadline for improving health care quality”. *The Journal of the American Medical Association, JAMA*, 295:324–327.
- Bailey, T. C. et al. (2013). “A Trial of a Real-Time Alert for Clinical Deterioration in Patients Hospitalized on General Medical Wards”. *Society of Hospital Medicine*. DOI 10.1002/jhm.2009.
- Badriyah, T. et al. (2013). “Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS)”. *Resuscitation*. Elsevier Ireland.
- Brady, P. W. et al. (2013). “Improving Situation Awareness to Reduce Unrecognized Clinical Deterioration and Serious Safety Events”. *Pediatrics*. 131:e298 DOI: 10.1542/peds.2012-1364.
- Burbey I. and Martin T. (2012). “A survey on predicting personal mobility”. *International Journal of Pervasive Computing and Communications*. Emerald.
- Costa, C. A, Yamin, A. Geyer, C. (2008). “Towards a General Software Infrastructure for Ubiquitous Computing”. *IEEE Pervasive Computing*. V. 7, p. 64-73.
- Elgazzar K., Aboelfotoh M., Martin P. (2012). “Ubiquitous Health Monitoring Using Mobile Web Service”. *Procedia Computer Science*. V. 10, n. 0, p. 332 – 339.
- Gelogo, Y. E., and Kim, H. K. (2013). “Unified Ubiquitous Healthcare System Architecture with Collaborative Model”. *International Journal of Multimedia and Ubiquitous Engineering*, 8(3).

- Gonçalves P. et al. (2012). “Reduced frequency of cardiopulmonary arrests by rapid response teams”. *Einstein*. SciELO.
- Hendrich, A. L. (2007). “Predicting Patient Falls”. *American Journal of Nursing*. 107(11): 50-58.
- Kirkland, L. et al. (2013). “A Clinical Deterioration Prediction Tool for Internal Medicine Patients”. *American Journal of Medical Quality*. 28(2) 135–142.
- Kitchenham, B. (2004). “Procedures for Performing Systematic Reviews”. *Joint Technical Report*. Keele University.
- Kyriacos, U. et al. (2014). “Monitoring Vital Signs: Development of a Modified Early Warning Scoring (Mews) System for General Wards in a Developing Country”. *PLoS ONE*. 9(1): e87073. doi:10.1371/journal.pone.0087073
- Konig I., Voigtmann C., Klein B., David K. (2011). “Enhancing alignment based context prediction by using multiple context sources: experiment and analysis”. In *Modeling and Using Context*. Springer-Verlag, p. 159–172.
- McCannon C., Hackbarth A., Griffin A. (2007). “Miles to Go: An Introduction to the 5 Million Lives Campaign”. *The Joint Commission Journal on Quality and Patient Safety*, Volume 33, Number 8, pp. 477-484(8).
- Morgan R. J. M., Williams F., Wright M. M. (1997). “An early warning scoring system for detecting developing critical illness”. *Clinical Intensive Care*. 8: 100.
- Offner P., Heit J., Roberts R. (2007). “Implementation of a Rapid Response Team Decreases Cardiac Arrest Outside of the Intensive Care Unit”. *Journal of Trauma-Injury Infection & Critical Care*. OVID.
- Patterson D., Liao L., Fox D., Kautz H. (2003). “Inferring High-Level Behavior from Low-Level Sensors”. In *Ubiquitous Computing*, p. 73–89.
- Rocha, C., Costa, C. A., Righi, R. R. (2015). Um modelo para monitoramento de sinais vitais do coração baseado em ciência da situação e computação ubíqua. *SBCUP 2015*.
- Rosa J. H., Barbosa J. L.V., Ribeiro G. D. (2015). “ORACON: An Adaptive Model for Contexts Prediction”. *Expert Systems With Applications*, Elsevier.
- Satyanarayanan M. (2001). “Pervasive computing: vision and challenges”. *IEEE Personal Communications*. IEEE, 2001.
- Sigg S. (2008). “Development of a Novel Context Prediction Algorithm and Analysis of Context Prediction Schemes”. *Kassel University Press*.
- Watcher R., Pronovost P. (2006). “The 100,000 Lives Campaign: A Scientific and Policy Review”. *Journal on Quality and Patient Safety*.
- Weiser M. (1991). “The computer for the twenty-first century”. *ACM SIGMOBILE Mobile Computing and Communications Review - Special issue dedicated to Mark Weiser*. ACM Digital Library.
- Weiss G., Hirsh H. (1998). “Learning to Predict Rare Events in Categorical Time-Series Data”. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, CA.

Um modelo ubíquo para retriagem de pacientes baseado na Internet das coisas

Guilherme Wunsch¹, Cristiano André da Costa¹, Rodrigo da Rosa Righi¹

¹Programa de Pós-Graduação em Computação Aplicada
Universidade do Vale do Rio dos Sinos (UNISINOS)

email@guiwunsch.com, {cac, rrrighi}@unisinis.br

Abstract. *Triage is a process which aims to sort patients according to their need for care in emergency department. Within this process there is re-triage, which occurs when another assessment is performed if the patient has clinical needs for it or on a fixed period defined by triage protocols. This study aimed to develop a computer model which use the concepts of mobile and ubiquitous computing focused on healthcare in patients re-triage stage. The main scientific contribution of this study was infer the correct re-triage moment collecting the vital signs of the patients with the use of Internet-of-things. For the evaluation of the proposed model was developed a prototype trying to analyze its application in a created scenario. Thus, we concluded that the developed model is viable to use in emergency department.*

Resumo. *Triagem é um processo que visa ordenar os pacientes de acordo com sua necessidade de cuidados em emergências de hospitais. Contida nesse processo existe a retriagem, que ocorre quando uma outra avaliação é efetuada caso haja necessidades clínicas para isso ou em um período fixo e recorrente definido pelo protocolo do triagem. Esse trabalho se propôs a desenvolver um modelo computacional que visa utilizar conceitos de computação móvel e ubíqua focados nos cuidados de saúde durante a etapa de retriagem de pacientes. A principal contribuição científica desse trabalho foi inferir o momento correto para efetuar a retriagem através da coleta de sinais vitais dos pacientes fazendo o uso da Internet das coisas. Para a avaliação do modelo proposto foi desenvolvido um protótipo buscando analisar sua aplicação em um cenário criado. Sendo assim, nós concluímos que o modelo desenvolvido é viável para uso nas emergências de hospitais.*

1. Introdução

Triagem é um processo realizado em hospitais para classificar os pacientes de acordo com sua necessidade de cuidados [Jentsch et al. 2013]. Atualmente, protocolos robustos de triagem vem sendo amplamente utilizados nos hospitais do mundo todo [Christ et al. 2010]. A retriagem é uma etapa importante contida nesse processo e ocorre quando uma nova avaliação do paciente é efetuada caso haja necessidades clínicas para isso [Rahmat et al. 2013]. Segundo [Christ et al. 2010], três dos quatro principais protocolos de triagem - Australasian Triage Scale (ATS), Canadian Triage And Acuity Scale (CTAS), Manchester Triage System (MTS) e Emergency Severity Index (ESI) - inserem

a retriagem em seu fluxo. Sendo assim, para essas etapas serem bem sucedidas, necessitamos de profissionais treinados e artefatos que auxiliem e proporcionem maior agilidade e precisão, tornando este processo efetivo e funcional.

Em apenas um dos principais protocolos de triagem é definida uma periodicidade para a retriagem dos pacientes. Entretanto, os demais protocolos, que aplicam essa etapa durante o processo de triagem, apenas adotam que a retriagem deve ser utilizada quando necessário [Christ et al. 2010]. Sendo assim, por um lado, pode-se estar fazendo verificações desnecessárias junto aos pacientes caso adotada uma periodicidade fixa para efetuar a retriagem. Por outro lado, aplicá-la quando necessário é, de certa forma, muito subjetivo. Nesse aspecto é importante conseguir determinar o momento correto da retriagem visando eliminar esses problemas mencionados.

A Computação ubíqua (ou computação pervasiva) é usada para descrever a onipresença dos computadores no cotidiano das pessoas [Weiser 1991]. Assim, tem se desenvolvido aplicações cada vez mais adaptáveis, capazes de interpretar um contexto e antecipar futuras ações que devem ser efetuadas [Schilit et al. 1994, Marilza Pernas et al. 2012]. Em função disso, esse trabalho se propõe a estudar a computação ubíqua com o objetivo de propor um modelo computacional para auxiliar os enfermeiros na retriagem de pacientes em emergências de hospitais usando também a Internet das coisas afim de detectar o melhor momento para efetuar a retriagem.

Este artigo está estruturado nas seções a seguir. A segunda apresenta a fundamentação teórica utilizada. A terceira traz o modelo da solução proposta. Na sequência, a quarta seção apresenta a metodologia de avaliação. A quinta descreve os trabalhos relacionados. Por fim, a conclusão deste trabalho é apresentada, na sexta seção.

2. Fundamentação teórica

A ideia da tecnologia estar inserida de forma natural no cotidiano das pessoas foi visionada em 1991 por Weiser. O autor defende que o conceito, denominado computação ubíqua, tornaria a interação entre o ambiente, usuário e tecnologia invisíveis [Weiser 1991]. Segundo [Satyanarayanan 2011], esse conceito só pôde ser alcançado nos dias de hoje através de inovações na tecnologia de redes sem fio, maior eficiência energética e hardwares cada vez mais portáteis com softwares cada vez mais adaptáveis.

Essa adaptação de softwares pôde ser alcançada através do estudo de ciência de contexto. Conforme [Schilit et al. 1994], com a computação móvel os dispositivos estão constantemente trocando seu ambiente de execução. A quantidade de processadores disponíveis para determinada tarefa, dispositivos de entrada acessíveis para o usuário, capacidade de rede e custos de conexão podem mudar ao longo do tempo e lugar. Explorar essa troca é o objetivo da ciência de contexto [Schilit et al. 1994]. Ampliando esse conceito, [Marilza Pernas et al. 2012] define que, através de ciência de situação, é possível proporcionar que aplicativos tenham a capacidade de perceber e compreender a situação atual, com a finalidade de antecipar ou prever sua demanda em um futuro próximo.

Com a finalidade de antecipar essa demanda, pode-se fazer o uso de dispositivos inteligentes interconectados para auxiliar no entendimento de um determinado contexto. Isso vai ao encontro do uso de Internet das coisas. Esse conceito promete criar um mundo onde todos os objetos ao redor são conectados a Internet e se comunicam uns

com os outros sem a necessidade de intervenção humana [Atzori et al. 2010]. Segundo [Blake 2015], a Internet das coisas para cuidados de saúde tem emergido recentemente como uma abordagem para cuidados de saúde personalizados.

Esse é o objetivo dos cuidados de saúde de forma ubíqua. Ou seja, aproximar os serviços médicos em ambos os lados, tanto para pacientes quanto para profissionais da saúde. Um sistema para cuidados de saúde de forma ubíqua implica em um ambiente que pacientes possam receber, a qualquer momento, tratamento, assim como, também, possam transmitir remotamente informações pertinentes à sua saúde. [Gelogo and Kim 2013].

3. Modelo proposto

A arquitetura geral do modelo proposto buscou integrar todos os conceitos definidos anteriormente em uma arquitetura de sistemas distribuídos, denominada *Service-oriented architecture* (SOA), ou arquitetura orientada à serviços. O principal fundamento da SOA é a disponibilização das funcionalidades de um sistema na forma de serviços [Krafzig et al. 2005]. Sendo assim, como observado na Figura 1, as regras de negócio e dados armazenados pelo modelo estão localizados nesse serviço, que será descrito posteriormente na Subseção 3.1. Por outro lado, os clientes, futuramente descritos na Subseção 3.2, são aplicativos distribuídos em dispositivos móveis que podem ser utilizados tanto pelos enfermeiros quanto pelos pacientes (enquanto aguardam o atendimento).

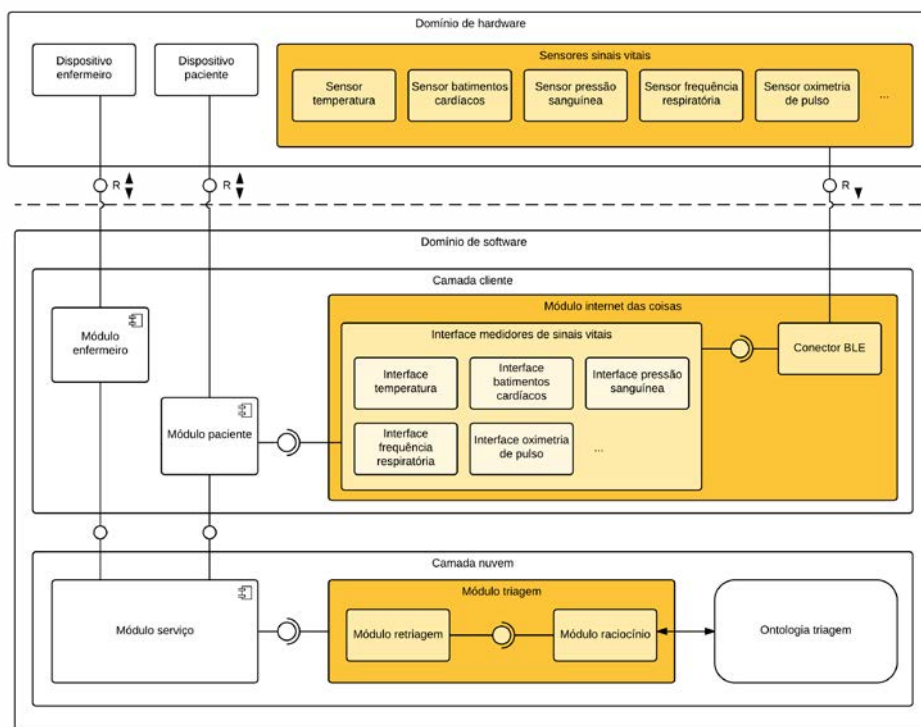


Figura 1. Visão geral do modelo proposto.

3.1. Camada nuvem

A camada nuvem é onde se localiza o serviço do modelo. Esse serviço disponibiliza uma interface REST para que todos os clientes possam acessá-la de forma fácil, rápida e eficaz.

Nessa camada, as regras de negócio do modelo estão organizadas de forma hierárquica, tendo como principal objetivo prover informações voltadas ao processo de triagem a todos os clientes da solução. Nesse caso, ela é responsável por receber as informações coletadas diretamente dos pacientes através dos dispositivos vestíveis, alimentar um subsistema conectado a uma ontologia do protocolo de triagem utilizado e, através de uma técnica de raciocínio, inferir a triagem do paciente de acordo com os dados coletados. O módulo mais importante dessa camada é o responsável pela retriagem dos pacientes. Seu papel é interpretar todos os dados de sinais vitais coletados em tempo real, analisá-los e, novamente, através de raciocínio na ontologia do protocolo avaliar se houve alguma mudança na triagem do paciente para então disparar uma notificação ao enfermeiro responsável caso necessário. Essa ontologia pode ser personalizada de acordo com o protocolo de triagem utilizado pelo hospital. Um modelo dela pode ser observado na Figura 2.

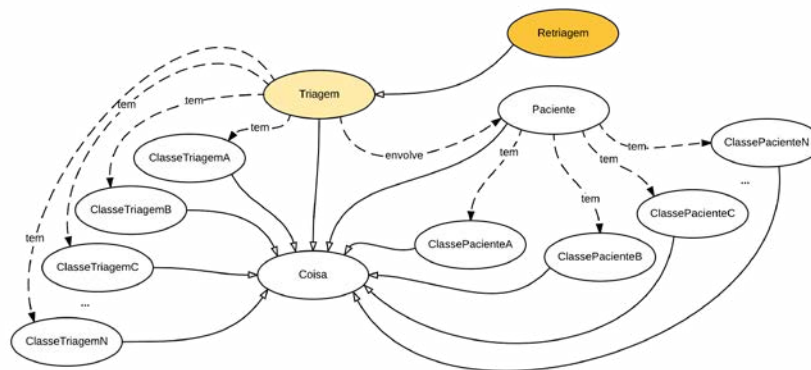


Figura 2. Modelo da ontologia para definir a triagem e retriagem do paciente.

Note que no diagrama desenvolvido na Figura 2, as classes da ontologia são representadas por figuras ovais. Setas sólidas representam a hierarquia entre classes e subclasses, enquanto setas pontilhadas são utilizadas para representar as propriedades entre as instâncias dessas classes. Os indivíduos dessa ontologia são, por exemplo, apresentações, discriminadores e sintomas dos pacientes que estão inseridos em suas respectivas classes relacionadas à triagem (ClasseTriagemA, B, C, ... e N). Eles também são, por exemplo, grupos de idade, gênero e demais informações relacionadas ao paciente através de suas respectivas classes (ClassePacienteA, B, C, ... e N). Os indivíduos dessa ontologia devem estar relacionados através de regras e propriedades de transitividade com as suas respectivas categorias de triagem para que um raciocinador consiga inferir a triagem de um paciente com base nas informações obtidas do paciente.

3.2. Camada cliente

A camada cliente proposta nesse modelo, no entanto, é constituída por aplicativos desenvolvidos para dispositivos móveis. Esses aplicativos devem auxiliar tanto enfermeiros como os pacientes durante o processo de triagem e enquanto o paciente está aguardando atendimento médico. No caso dos enfermeiros, o aplicativo auxilia em toda a gestão do fluxo de pacientes na emergência do hospital e agiliza o processo de coleta de dados na triagem. Esses dados vão desde informações pessoais até os sinais vitais, que podem ser obtidos de dispositivos vestíveis instalados diretamente nos pacientes através de uma ponte criada a partir do aplicativo do paciente. Além disso, os enfermeiros também

podem receber notificações (disparadas pelo serviço da solução) quando for necessário efetuar a retriagem de algum paciente. Por outro lado, os pacientes dispõem apenas de informações voltadas ao seu processo de triagem, como, por exemplo, o tempo em fila, sua posição atual, estimativa de tempo para o atendimento, etc. Esse aplicativo do paciente também servirá como agregador dos dispositivos vestíveis instalados em seu corpo que visam monitorar seu estado de saúde através dos sinais vitais.

Esses dispositivos vestíveis, usados para aferir os sinais vitais dos pacientes, conectam-se com o aplicativo do paciente (no momento da triagem e enquanto aguarda pelo atendimento) através de redes de conectividade de baixo alcance como ZigBee, Bluetooth, *Wireless Body Area Networks* (WBAN), etc [Atzori et al. 2010]. O modelo se propõe a integrar apenas dispositivos compatíveis com *Bluetooth Low Energy* (BLE) visto que essa é uma das principais tecnologias no mercado para esses sensores vestíveis e é compatível também com *smartphones*, *tablets* e *phablets* que estão sendo vendidos. Sendo assim, um módulo importante da camada cliente é o de Internet das coisas. Esse módulo possui um conector BLE responsável por receber a comunicação vinda dos diversos dispositivos vestíveis conectados e modulariza-la, através de uma interface para cada tipo de sensor, de modo que o restante do sistema possa entender qual tipo de sinal vital esta sendo recebido em dado momento e em qual unidade de medida pré-definida ele deve ser interpretado. Esse módulo é uma peça fundamental, pois existem hoje no mercado muitos tipos de dispositivos dos mais diversos fabricantes.

4. Avaliação

A comunidade científica tem empregado o uso de cenários para avaliação de aplicações ubíquas e sistemas sensíveis ao contexto [Satyanarayanan 2011, da Rocha et al. 2015]. Segundo os autores [Westergren et al. 2014], com o uso dessa metodologia na área da saúde é possível tomar uma decisão mais consciente, pois eliminamos o fator humano e focamos apenas nos reais problemas envolvidos. Além disso, ela proporciona uma economia de custo e tempo e, também, não são envolvidas questões éticas como, por exemplo, aplicar dois processos de triagem ao mesmo paciente [Worster et al. 2007, Jobé et al. 2014]. Com o objetivo de avaliar o modelo descrito anteriormente, foi criado um cenário que descreve uma situação fictícia de uma paciente que chegou à emergência do hospital e foi devidamente triada. Após algum tempo na fila de espera, apresentou alterações significativas em seus sinais vitais e precisou ser retriada. A seguir é descrito o cenário utilizado para a avaliação do modelo proposto:

“Uma menina de dezesseis anos de idade, na qual foi enviada para casa, pois não estava se sentindo bem na escola, chegou com sua mãe na emergência. Ela é asmática há cinco anos. O hospital possui enfermeiros treinados e aptos a utilizar o protocolo de Manchester no processo de triagem. Durante a triagem, o enfermeiro verificou que a paciente está completamente lúcida e mencionou que tem estado cada vez mais ofegante a dois dias. Seus batimentos cardíacos e a saturação de oxigênio, aferidos no momento da triagem através de dispositivos vestíveis são respectivamente 100 bpm e 92%. O enfermeiro utilizou o fluxograma para asma do protocolo de Manchester e definiu a classificação amarela à paciente. A menina foi dirigida até a sala de espera ainda com os dispositivos vestíveis para o monitoramento, em tempo real, dos seus sinais vitais. Durante a espera sua pulsação sofreu uma alteração para 110 bpm e a saturação de oxigênio para 90%. Automaticamente o sistema enviou uma notificação para o enfermeiro responsável pela

triagem para que fosse efetuada a retriagem da paciente. Então, baseado nas alterações nos sinais vitais, a paciente foi reclassificada com a cor laranja, ganhando maior prioridade no atendimento de acordo com o protocolo aplicado pelo hospital.”

Complementando a avaliação proposta para o modelo, foi desenvolvido um protótipo de aplicação, para iOS, que se comunica com um serviço na nuvem, também desenvolvido para esse processo. A elaboração do protótipo foi efetuada com base no modelo proposto e dividida em três etapas de desenvolvimento: criação do serviço, desenvolvimento do aplicativo para dispositivos móveis e desenvolvimento do módulo de integração com os sensores vestíveis que monitoram os sinais vitais do paciente.

O serviço foi desenvolvido utilizando a linguagem PHP em conjunto com a *framework* Symfony¹. A camada de interface do serviço, responsável por disponibilizar os recursos que são acessados pelos aplicativos dos pacientes e enfermeiros, foi construída utilizando o padrão REST para transferir objetos *JavaScript Object Notation* (JSON) sobre o protocolo *Hyper Text Transfer Protocol Secure* (HTTPS). Já para a persistência dos dados foi utilizado o Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL. Em conjunto com o banco de dados, foi utilizada a *Application Programming Interface* (API) de mapeamento objeto relacional para PHP denominada Doctrine², que é distribuída junto com a *framework* Symfony. Esse serviço é responsável por gerenciar todo o processo de triagem, que é atribuída ao paciente através de um processo de raciocínio efetuado em uma ontologia, e, além disso, notificar aos enfermeiros o momento que se deve efetuar a retriagem dos pacientes com base em alterações significativas nos sinais vitais ou no tempo máximo definido pelo protocolo para fazer a retirada.

O protótipo desenvolvido utilizou uma ontologia do protocolo de Manchester, que é um dos mais utilizados no mundo e muito comum nas emergências de hospitais brasileiros, durante o processo de avaliação. Apesar do protocolo de Manchester possuir mais de cinquenta fluxogramas [Mackway-Jones et al. 2014], que devem ser utilizados de acordo com a apresentação de cada paciente, na avaliação proposta foi utilizado apenas o fluxograma para Asma e os discriminadores voltados a esse fluxograma. Sendo assim, após a triagem do paciente, o serviço começa a escutar modificações nos sinais vitais enviadas pela interface do paciente e alerta o enfermeiro responsável pela triagem quando, no caso de pacientes com asma, houverem alterações significativas na temperatura corporal, pulsação, espirometria ou na oximetria de pulso do paciente. Segundo o protocolo de Manchester, alguns outros discriminadores também afetam a triagem do paciente, como o nível de consciência do paciente e a capacidade de falar pequenas sentenças sem a necessidade de respirar [Mackway-Jones et al. 2014]. Entretanto, esses discriminadores não foram utilizados na avaliação proposta.

O aplicativo que faz interface com o paciente e enfermeiro foi desenvolvido para a plataforma de dispositivos móveis iOS. As telas desse protótipo podem ser observadas na Figura 3, que apresenta as visões do (A) enfermeiro e do (B) paciente. Sua construção utilizou a linguagem nativa para desenvolvimento nessa plataforma, denominada Objective-C. O aplicativo que faz interface com o paciente fica conectado ao servidor enviando todas as alterações nos sinais vitais do paciente. Sendo assim, sempre que for identificada alguma mudança significativa, o enfermeiro responsável pela triagem do paciente é

¹Disponível em <http://symfony.com>.

²Disponível em <http://www.doctrine-project.org>.

notificado para que efetue a retriagem.

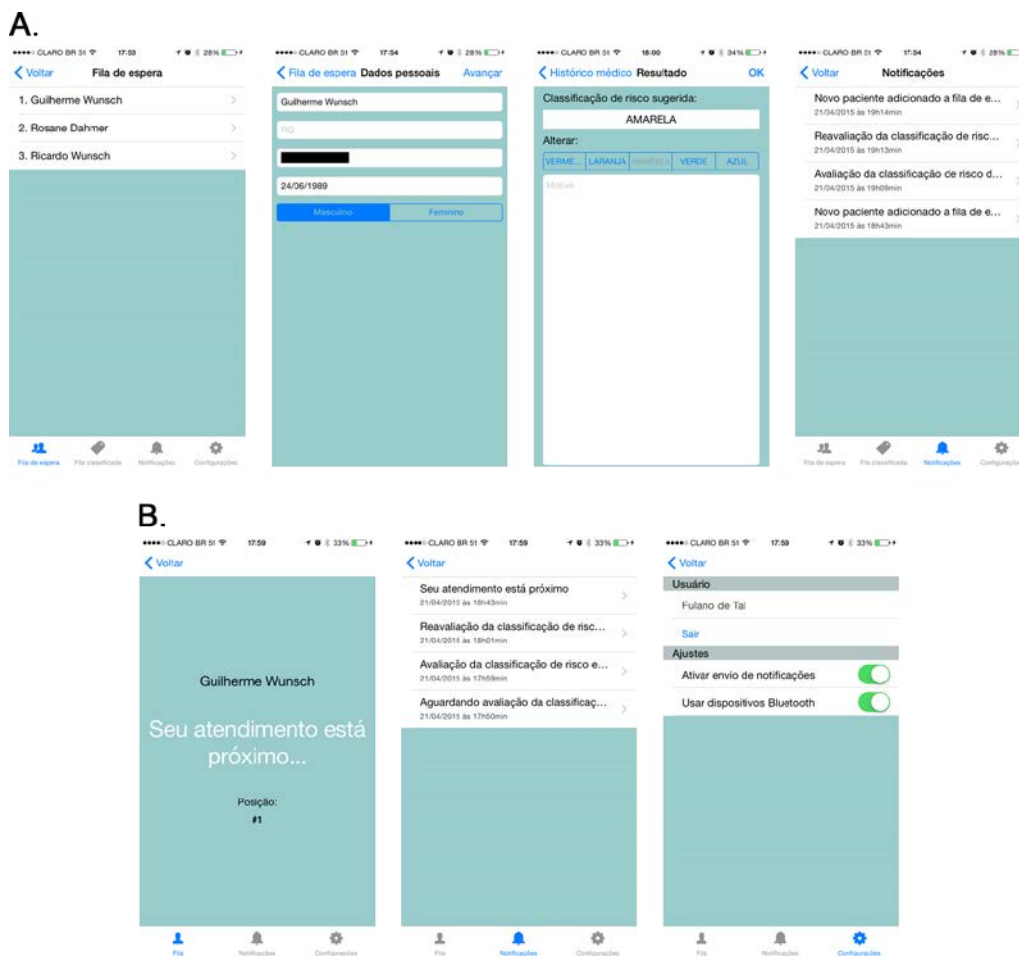


Figura 3. Telas do protótipo desenvolvido: Interface dos (A) enfermeiros e dos (B) pacientes.

No cenário criado para a avaliação do modelo, a paciente inicialmente foi triada com a cor laranja, entretanto durante a avaliação as medições de oximetria de pulso e da pulsação do paciente alteraram. Essas alterações foram enviadas ao serviço que detectou uma retriagem do paciente. Nesse mesmo momento o enfermeiro responsável recebeu uma notificação para que fosse efetuada essa retriagem. Para que esse resultado fosse alcançado, foram desenvolvidos, também na linguagem Objective-C, simuladores dos sensores citados anteriormente, que foram acoplados ao módulo responsável pela integração com dispositivos vestíveis para monitoramento dos sinais vitais. Esse módulo não foi desenvolvido por completo, pois foi o meio encontrado para conseguir chegar mais próximo da situação criada no cenário proposto para a avaliação.

Com isso, foi possível verificar que o modelo se comportou de maneira adequada, garantindo que a retriagem da paciente fosse efetuada em seu devido momento, ou seja, quando, através da ciência de situação, foi detectado que as informações de sinais vitais, aferidas através de sensores vestíveis, sofreram alterações significativas. Sendo assim, o protótipo confirmou que o modelo foi capaz de detectar, de forma ubíqua, alterações no quadro dos pacientes notificando enfermeiros quando fosse necessário efetuar a retriagem.

5. Trabalhos relacionados

Alguns modelos, que se propõem a monitorar os sinais vitais de pacientes, podem ser encontrados na literatura. Entretanto, esses modelos não estão direcionados diretamente ao processo de triagem de pacientes nas emergências de hospitais. Dentre os modelos encontrados, foram escolhidos os mais representativos em relação ao modelo proposto.

O trabalho de [Borodin et al. 2015] se propõe a desenvolver uma arquitetura, utilizando o conceito de Internet das coisas, para um sistema de monitoramento de saúde usando diferentes fontes para monitorar dados voltados à saúde. O modelo prevê suporte a eletrocardiograma, batimentos cardíacos e outros, como, espirômetro, estetoscópio e tonômetro. A arquitetura ainda define o uso de serviços na nuvem para armazenar e tratar registros remotamente.

Já [Hassanalieragh et al. 2015] se preocupa em definir como deve ser uma arquitetura para monitoramento de saúde usando Internet das coisas e processamento em nuvem. O modelo desenvolvido trata desde a aquisição de dados voltados à saúde dos usuários, através de sensores conectados a um concentrador (que pode estar localizado em um smartphone), até a transmissão desses dados de forma segura e modularizada para posterior armazenamento e análise com uma interface de visualização intuitiva que facilita a leitura dos dados.

Uma arquitetura com uma camada de hardware que suporta múltiplos protocolos para conectar dispositivos clínicos, sensores de ambiente e sistemas de diferentes fontes de dados foi proposta por [Jara et al. 2012]. Ela transmite de forma contínua todos os dados coletados para uma outra camada, já em software, que é responsável por modularizar e proteger esses dados para serem transmitidos, em tempo real, à uma plataforma capaz de monitorar, de forma proativa, os dados recebidos. Gerando alertas e possibilitando sua utilização para o diagnóstico remoto.

Por fim, o modelo proposto em [da Rocha et al. 2015], permite o monitoramento da saúde utilizando dispositivos conectados através de uma rede Bluetooth ao smartphone do paciente. Esse modelo analisa o contexto do paciente, com base nos sinais fisiológicos recebidos e envia alterações da sua saúde a um serviço na nuvem capaz de efetuar o processamento de situação para tomar ações de acordo com a classificação de risco definida pela análise efetuada previamente. Essas ações vão desde enviar orientações ao paciente, até notificar o seu cuidador e, posteriormente, um centro de telemedicina.

Comparando os trabalhos relacionados, podemos observar que nenhum dos modelos propostos faz uso de informações pertinentes a saúde dos pacientes para auxiliar no processo de triagem. Outra lacuna que foi considerada muito importante, e, também, não explorada pelos modelos mencionados, é retriagem de pacientes. A Tabela 1 facilita a compreensão dessa comparação e apresenta as principais características dos trabalhos relacionados, incluindo o modelo proposto pelo presente trabalho.

6. Conclusão

O presente trabalho teve como objetivo elaborar um modelo computacional para auxiliar os enfermeiros no processo de retriagem em emergências de hospitais usando os conceitos da computação móvel e ubíqua e Internet das coisas. A crescente ascensão dos dispositivos móveis e a necessidade de maior agilidade e menor subjetividade nas funções

Tabela 1. Comparação de características dos trabalhos relacionados em relação ao modelo proposto.

Característica	[Borodin et al. 2015]	[Hassanalieragh et al. 2015]	[Jara et al. 2012]	[da Rocha et al. 2015]	Modelo proposto
Uso de dispositivos móveis	Sim	Sim	Não	Sim	Sim
Uso de sensores	Sim	Sim	Sim	Sim	Sim
Serviço na nuvem	Sim	Sim	Não	Sim	Sim
Análise de contexto	Não	Sim	Não	Sim	Sim
Análise de situação	Não	Não	Não	Sim	Sim
Uso na triagem de pacientes	Não	Não	Não	Não	Sim
Uso na retriagem	Não	Não	Não	Não	Sim

desempenhadas pelos profissionais da área de saúde foram as principais motivações que ocasionaram na elaboração desse modelo. Sua principal contribuição científica, sendo uma das principais lacunas encontradas nos trabalhos relacionados, foi a avaliação da necessidade de efetuar a etapa de retriagem, em momento apropriado, a partir de dados de saúde coletados diretamente dos pacientes com o uso de Internet das coisas.

Para o processo de avaliação foi desenvolvido um protótipo cuja finalidade foi aplicá-lo a um cenário desenvolvido para validar o uso do modelo proposto. Com isso, foi possível verificar que o uso de ciência de situação se adéqua com eficiência ao processo de triagem, notificando os enfermeiros o momento em que se deve efetuar a retriagem de um paciente. Esse resultado foi possível ser alcançado com o uso de dispositivos vestíveis que monitoram os sinais vitais do paciente. Sendo assim, também é possível concluir que o uso da Internet das coisas agregou funcionalidades essenciais ao modelo desenvolvido e que, o modelo proposto é viável para utilização em emergências de hospitais.

Existem algumas oportunidades de trabalhos futuros relacionados a esse modelo. Inicialmente pode-se desenvolver as integrações com dispositivos vestíveis reais para o monitoramento de sinais vitais dos pacientes. Além disso, também é possível realizar um estudo de caso com pacientes em um ambiente controlado para analisar como a solução se comportaria em situações reais e, com isso, avaliar atributos como a usabilidade e percepção de valor desse modelo.

Agradecimentos

Os autores gostariam de agradecer a CAPES, ao CNPq e a FAPERGS pelo apoio a essa pesquisa.

Referências

- Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15):2787–2805.
- Blake, M. B. (2015). An internet of things for healthcare. *IEEE Internet Computing*, (4):4–6.
- Borodin, A., Zavyalova, Y., Zaharov, A., and Yamushev, I. (2015). Architectural approach to the multisource health monitoring application design. In *Open Innovations Association (FRUCT), 2015 17TH Conference of*, pages 16–21. IEEE.
- Christ, M., Grossmann, F., Winter, D., Bingisser, R., and Platz, E. (2010). Modern triage in the emergency department. *Deutsches Ärzteblatt international*, 107(50):892.
- da Rocha, C. C. L., da Costa, C. A., and da Rosa Righi, R. (2015). Um modelo para monitoramento de sinais vitais do coração baseado em ciência da situação e computação ubíqua. In *VII Simpósio Brasileiro de Computação Ubíqua e Pervasiva*.

- Gelogo, Y. E. and Kim, H.-K. (2013). Unified ubiquitous healthcare system architecture with collaborative model. *International Journal of Multimedia and Ubiquitous Engineering*, 8(3).
- Hassanalieragh, M., Page, A., Soyata, T., Sharma, G., Aktas, M., Mateos, G., Kantarci, B., and Andreescu, S. (2015). Health monitoring and management using internet-of-things (iot) sensing with cloud-based processing: Opportunities and challenges. In *Services Computing (SCC), 2015 IEEE International Conference on*, pages 285–292. IEEE.
- Jara, A. J., Zamora, M. A., and Skarmeta, A. F. (2012). Knowledge acquisition and management architecture for mobile and personal health environments based on the internet of things. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, pages 1811–1818. IEEE.
- Jentsch, M., Ramirez, L., Wood, L., and Elmasllari, E. (2013). The reconfiguration of triage by introduction of technology. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 55–64. ACM.
- Jobé, J., Ghuysen, A., Gérard, P., Hartstein, G., and D’Orio, V. (2014). Reliability and validity of a new french-language triage algorithm: the elisa scale. *Emergency Medicine Journal*, 31(2):115–120.
- Krafzig, D., Banke, K., and Slama, D. (2005). *Enterprise SOA: service-oriented architecture best practices*. Prentice Hall Professional.
- Mackway-Jones, K., Marsden, J., and Windle, J. (2014). *Emergency triage*. Blackwell Publishing, 3 edition.
- Marilza Pernas, A., Diaz, A., Motz, R., and Palazzo Moreira de Oliveira, J. (2012). Enriching adaptation in e-learning systems through a situation-aware ontology network. *Interactive Technology and Smart Education*, 9(2):60–73.
- Rahmat, M. H., Annamalai, M., Halim, S. A., and Ahmad, R. (2013). Agent-based modelling and simulation of emergency department re-triage. In *Business Engineering and Industrial Applications Colloquium (BEIAC), 2013 IEEE*, pages 219–224. IEEE.
- Satyanarayanan, M. (2011). Mobile computing: the next decade. *ACM SIGMOBILE Mobile Computing and Communications Review*, 15(2):2–10.
- Schilit, B., Adams, N., and Want, R. (1994). Context-aware computing applications. In *First Workshop on Mobile Computing Systems and Applications, 1994. WMCSA 1994*, pages 85–90. IEEE.
- Weiser, M. (1991). The computer for the 21st century. *Scientific american*, 265(3):94–104.
- Westergren, H., Ferm, M., and Häggström, P. (2014). First evaluation of the paediatric version of the swedish rapid emergency triage and treatment system shows good reliability. *Acta Paediatrica*, 103(3):305–308.
- Worster, A., Sardo, A., Eva, K., Fernandes, C., and Upadhye, S. (2007). Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. *Journal of emergency nursing*, 33(4):319–323.

Um Serviço de Representação Ontológica de Contexto Baseada no Padrão de Projeto Estímulo-Sensor-Observação

Ernesto Fonseca Veiga¹, Renato de Freitas Bulcão Neto¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – CEP 74001-970 – Goiânia – GO – Brazil

{ernestofonseca, renato}@inf.ufg.br

Abstract. *Context-aware computing research highlights ontology-based modeling as a dominant technique for context representation, showing the need for evolving in expressiveness and standardization. This paper proposes Hermes Widget, which is a component that provides the ontological context representation service, with support for the Stimulus-Sensor-Observation sensing pattern. This proposal was validated in a monitoring vital signs scenario with contribution of a nursing team. Results highlight the contributions as expressiveness, standardization and interoperability of generated context.*

Resumo. *Pesquisas em Computação Sensível a Contexto destacam a modelagem baseada em ontologias como técnica dominante para representação de contexto, evidenciando a necessidade de uma evolução quanto a expressividade e padronização. Este trabalho propõe Hermes Widget, um componente que oferece o serviço de representação ontológica de contexto, baseada no padrão de sensoramento Estímulo-Sensor-Observação. A proposta foi validada em um cenário de monitoramento de sinais vitais, com contribuição de uma equipe de Enfermagem. Os resultados obtidos destacam as contribuições quanto a padronização, expressividade e interoperabilidade do contexto gerado.*

1. Introdução

O projeto e os serviços realizados por um Sistema de Gerenciamento de Contexto (CMS), de maneira geral, adotam um ciclo de vida ao qual as informações de contexto estão submetidas. Uma abordagem abrangente para o *ciclo de vida do contexto*, apresentada por [Perera et al. 2014], tem início com a etapa de aquisição, que é seguida pelas etapas de modelagem e de raciocínio, finalizando-se com a disseminação do contexto.

Grande parte da pesquisa que envolve este ciclo de vida está voltada para a investigação de abordagens para a *etapa de modelagem* [Bettini et al. 2010]. Algumas das principais necessidades identificadas para esta etapa são a evolução quanto à expressividade e à utilização de padrões para a representação de informação contextual [Sehic et al. 2012]. Neste sentido, as tecnologias da Web Semântica, tais como ontologias e a padronização para descrição de informações e realização de consultas, podem contribuir significativamente para a representação e o gerenciamento do contexto.

A importância e os benefícios de uma representação formal das informações de contexto têm sido destacados por [Perera et al. 2014], que recomendam a modelagem baseada em ontologias para apoiar requisitos como expressividade e interoperabilidade. A padronização e o formalismo oferecidos por esta técnica de modelagem permitem

tanto uma representação quanto o processamento de consultas mais expressivas, além da realização de inferências automáticas de novas informações sobre o contexto.

Neste trabalho apresentamos *Hermes Widget*, um componente de software que realiza a etapa de modelagem do ciclo de vida do contexto, e implementa como principal serviço a representação de contexto baseada em ontologias. Este componente é fundamentado na utilização de padrões da Web Semântica e também no padrão para modelagem de dados de contexto do domínio de sensoriamento, denominado Estímulo-Sensor-Observação, com o objetivo de fornecer um modelo de contexto padronizado e um serviço de representação ontológica reutilizável e interoperável para demais componentes de software e aplicações.

Oferecendo um serviço de representação de contexto independente de domínio, *Hermes Widget* pode ser reutilizado para diferentes cenários de aplicação, sendo ainda parte integrante do projeto do CMS *Hermes* [Veiga et al. 2014], que contempla todas as etapas do ciclo de vida do contexto. Porém, como possui um projeto arquitetural independente dos demais componentes dessa infraestrutura, o componente pode ser integrado a outros sistemas para realização dos serviços relacionados à etapa de modelagem.

O restante deste trabalho está assim estruturado: a Seção 2 contempla os principais fundamentos relacionados à proposta e introduz o CMS *Hermes*; a Seção 3 apresenta o componente *Hermes Widget*, sua arquitetura, requisitos e detalhes de implementação; a Seção 4 discorre sobre os experimentos e a validação do serviço de representação; a Seção 5 apresenta os trabalhos relacionados; e a Seção 6 sintetiza e conclui este trabalho.

2. Fundamentação Teórica

Esta seção apresenta a ontologia SSN e o CMS *Hermes*, introduzindo as informações necessárias para a compreensão deste trabalho.

2.1. Ontologia *Semantic Sensor Network*

Definições semânticas compartilhadas contribuem tanto para a integração do contexto adquirido de diferentes fontes, como para a interoperabilidade deste contexto entre sistemas e aplicações. Neste intuito, a *Semantic Sensor Network* (SSN) é uma ontologia desenvolvida pelo *W3C Incubator Group*, que tem como objetivo central descrever um sistema ou processo de sensoriamento, bem como os demais conceitos e relações deste domínio, conforme o padrão de projeto ontológico Estímulo-Sensor-Observação, ou *Stimulus-Sensor-Observation* (SSO) [Janowicz and Compton 2010].

O padrão de projeto SSO tem como objetivo apoiar a representação semântica de sistemas e dados de sensoriamento, e foi criado com base em princípios ontológicos que o tornam reutilizável para diferentes domínios de aplicação [Janowicz and Compton 2010]. Independente de ontologias de alto nível, introduz um conjunto de classes e relações, sendo as suas principais definições:

- **Estímulos:** mudanças detectáveis no ambiente, seja este físico ou lógico. São o ponto de partida de uma medição, podendo ser direta ou indiretamente relacionadas a propriedades ou características de interesse. Exemplos de estímulos podem ser alterações nos sinais vitais de um paciente.
- **Sensores:** podem ser componentes físicos ou de software que realizam o sensoriamento de um estímulo, por exemplo, um sensor de temperatura corpórea.

- **Observações:** atuam como a relação entre um estímulo de entrada, o sensor que o coleta e a saída deste sensor. Podem também descrever outros parâmetros, tais como espaço (localização) e tempo (instante de uma aferição). Uma observação é o conjunto das informações que descrevem o contexto do sensoreamento.

A Figura 1 apresenta a implementação do padrão SSO na ontologia SSN, com seus principais conceitos e relações.

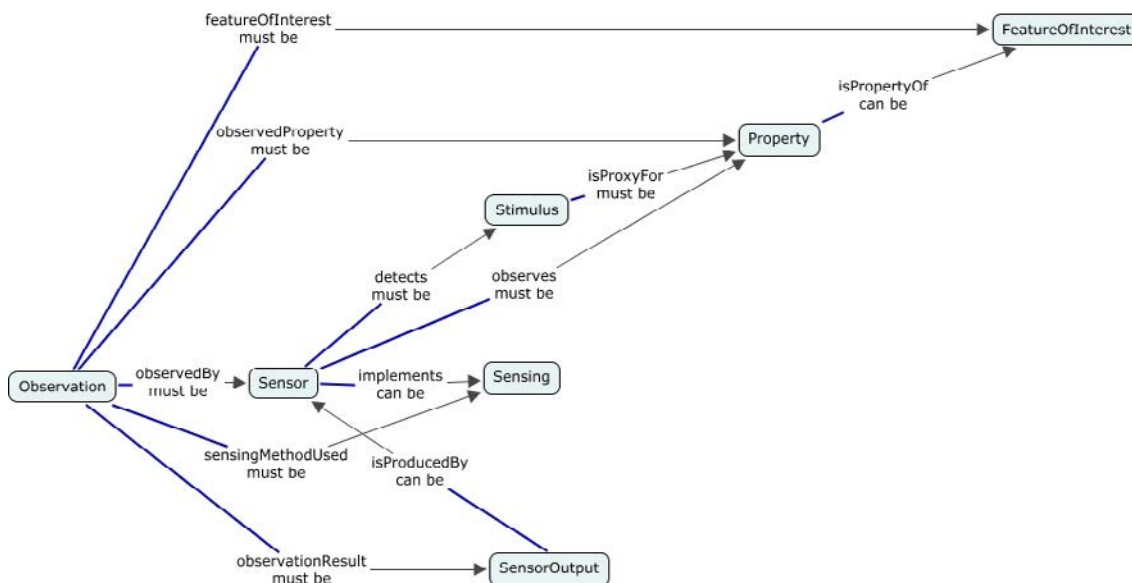


Figura 1. Padrão *Stimulus-Sensor-Observation* [Janowicz and Compton 2010]

A ontologia SSN foi utilizada como base para a padronização do serviço de representação ontológica de contexto, núcleo da proposta deste trabalho.

2.2. Sistema para Gerenciamento de Contexto *Hermes*

Hermes é um CMS baseado em ontologias proposto por [Veiga et al. 2014], que tem como objetivo principal apoiar o desenvolvimento de aplicações sensíveis a contexto. Os principais requisitos para sua construção são: *i*) oferecer serviços de apoio ao gerenciamento do ciclo de vida das informações de contexto; *ii*) infraestrutura baseada em componentes independentes; e *iii*) serviços apoiados na semântica das informações de contexto; destacados por [Bettini et al. 2010, Perera et al. 2014].

A Figura 2(a) apresenta a arquitetura de *Hermes* e seus componentes: *i*) *Hermes Base*, responsável pela comunicação componente/componente e componentes/aplicações na disseminação de contexto [Maranhão et al. 2014]; *ii*) *Hermes Widget*, representação ontológica de contexto coletado de sensores [Veiga et al. 2016]; *iii*) *Hermes Aggregator*, agregador para geração de contexto com maior nível de expressividade; *iv*) *Hermes Interpreter*, interpretação e filtragem semântica de contexto [Maranhão et al. 2014]; e *v*) *Hermes History*, gerenciador de histórico, consultas e acesso a informação de contexto.

No intuito de simplificar a visualização do fluxo de contexto dentro do CMS *Hermes*, a Figura 2(b) apresenta separadamente o esquema de comunicação. O componente *Hermes Base*, responsável pela comunicação em *Hermes*, encapsula o acesso a

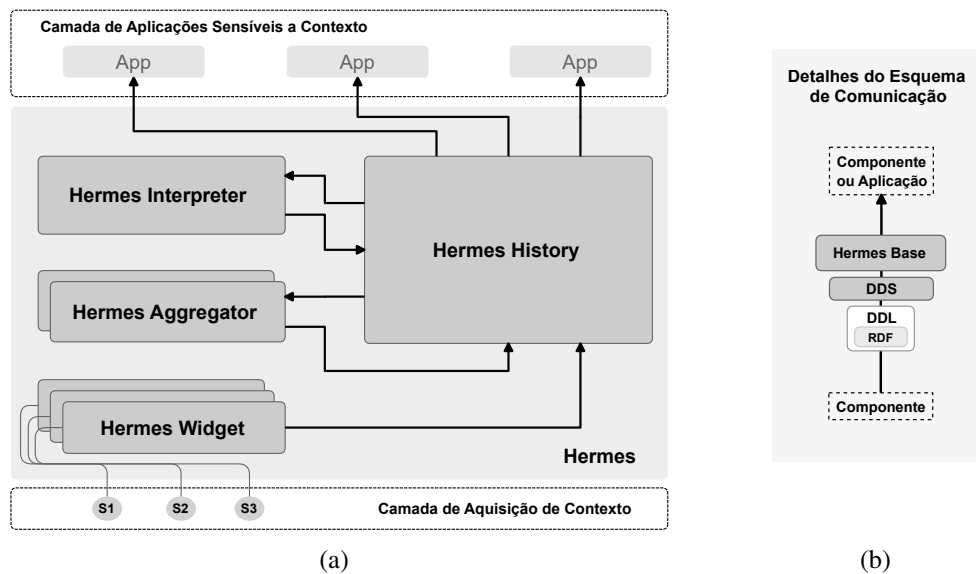


Figura 2. Infraestrutura *Hermes*, adaptada de [Veiga et al. 2016], onde: (a) apresenta o projeto arquitetural de componentes e (b) o esquema de comunicação.

um *middleware* baseado na especificação *Data Distribution Service* (DDS)¹, que permite a comunicação *publish-subscribe* do contexto serializado no formato RDF, padrão para representação de informações na Web Semântica. Este esquema é utilizado na arquitetura de *Hermes* para comunicação entre os componentes e também destes com as aplicações.

3. *Hermes Widget*: Representação Ontológica e Padronização de Contexto

Hermes Widget é o componente responsável por abstrair o serviço de representação ontológica do contexto. Este serviço é fundamental dentro do CMS *Hermes* (ou de qualquer outra infraestrutura de contexto), uma vez que oferece as informações de entrada necessárias à realização dos demais serviços e também às aplicações. A representação ontológica marca em *Hermes* a primeira etapa de transformação dos dados sem processamento, adquiridos de sensores, em contexto de alto nível, que pode ser utilizado por outros componentes e por aplicações sensíveis a contexto.

3.1. Requisitos e Arquitetura

Os principais requisitos de *Hermes Widget* são:

1. **Permitir a especificação de um sensor e do processo de sensoriamento:** o componente deve permitir a descrição de um processo de sensoriamento e de suas relações, tratando termos e eventos deste domínio tais como: sensor, propriedade, entidade de interesse, observação (ou sensoriamento), resultado entre outros.
2. **Representação ontológica das informações de contexto:** a representação realizada por *Hermes Widget* deve representar formalmente as informações de contexto sendo, neste caso, baseada em ontologias.
3. **Padronização do modelo de contexto:** *Hermes Widget* deve permitir a representação padronizada do contexto, independente do domínio de aplicação.

¹DDS: <http://www.omg.org/spec/DDS/>

4. **Disseminação de informações de contexto:** após representar o contexto recebido da camada de aquisição, o *Hermes Widget* deve disponibilizar essas informações para demais componentes do CMS *Hermes* e/ou aplicações interessadas.

A ontologia SSN e o padrão SSO foram definidos para cumprir os requisitos de 1 a 3, oferecendo a base para a construção do modelo de representação de contexto de *Hermes Widget* e também um padrão para esta representação (SSO). A representação baseada em ontologias, cumprindo o requisito 2, atribui formalidade e expressividade ao contexto, permitindo a descrição semântica das informações e a realização de inferências. Este mesmo requisito corrobora com o requisito 3, uma vez que, tanto a representação baseada em ontologias quanto a padronização de um modelo para a descrição do contexto, garantem a interoperabilidade no processo de disseminação.

A Figura 3 apresenta as camadas arquiteturais do componente *Hermes Widget*. Este projeto foi definido com base nos requisitos apresentados, juntamente com os requisitos do próprio CMS *Hermes*, que definem os serviços de configuração, comunicação e notificação de maneira geral para todos os seus componentes.

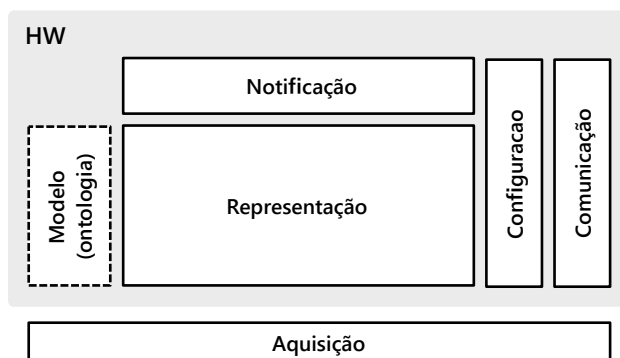


Figura 3. Arquitetura em camadas do componente *Hermes Widget*

A *camada de aquisição*, externa ao *Hermes Widget*, recebe as alterações de contexto detectadas pelos sensores, transmitindo-as à *camada de representação*². A representação é realizada com base em um modelo ontológico, conforme especificado nos requisitos, e o contexto produzido é gerenciado pelas camadas de configuração, comunicação e notificação, que realizam os serviços necessários à disponibilização do contexto representado.

A camada de configuração de um *Hermes Widget* acessa um arquivo de definição *JSON*, que descreve os parâmetros de configuração do componente. Estas definições tratam de maneira específica da criação do tópico que receberá as informações de contexto representadas e que será assinado por aplicações e pelos componentes interessados, que passarão a receber as notificações de contexto desse tópico.

3.2. Serviço de Representação Ontológica

A Figura 4 apresenta os detalhes relacionados ao projeto e implementação do serviço de representação ontológica de contexto. Cada aferição de contexto recebida de um sensor, através da camada de aquisição, é transformada em uma informação de entrada para o serviço de representação, que é composto de duas partes principais:

²*Hermes Widget* e nenhum outro componente de *Hermes* trata a aquisição de dados de sensores físicos.

1. **Gerente de Representação:** é a subcamada em que se realiza a representação ontológica do contexto adquirido. Tem como base a ontologia SSN, que fornece o modelo de contexto para as informações adquiridas.
2. **Gerente de Contexto:** prepara as representações para envio, na forma de instâncias de contexto devidamente serializadas.

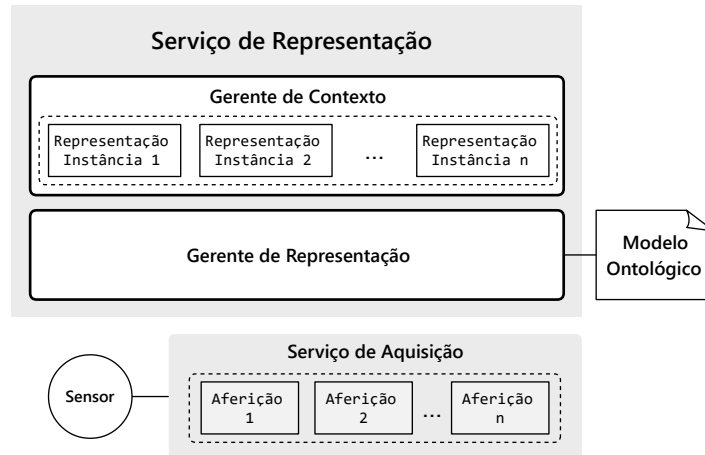


Figura 4. Serviço de representação ontológica de contexto

Na implementação do serviço de representação de contexto foi utilizado o *framework* Jena ³, da Apache, tecnologia *open source* que oferece API's para a manipulação de informação semântica, tais como modelos RDF, ontologias OWL e realização de consultas SPARQL. A ontologia SSN é o modelo utilizado para criação das representações de contexto, tornando o serviço reutilizável para receber entradas de diferentes tipos de sensores e produzir como saída uma representação de contexto padronizada.

4. Cenário de Validação do *Hermes Widget*

Essa seção apresenta a validação de *Hermes Widget* por meio da simulação de um cenário de monitoramento de sinais vitais humanos em Unidade de Terapia Intensiva (UTI).

4.1. Cenário de testes

Para esta validação foram implementados cinco *Hermes Widgets*, sendo um para cada sinal vital, a saber: *pressão arterial*, *temperatura corpórea*, *frequência de pulso*, *frequência respiratória* e *saturação de oxigênio*. O objetivo foi reproduzir o ambiente de monitoramento real de uma UTI.

Não dispondo de sensores físicos para realização dos testes, nem do acesso à uma UTI real para integração com as instâncias do componente *Hermes Widget*, o serviço de aquisição foi realizado na forma de simulação. Essa simulação ocorreu através da leitura dos dados da base pública *PhysioBank* ⁴, que disponibiliza monitoramento de sinais vitais reais coletados de pacientes de UTI. Dessa forma, foram implementados *sensores virtuais* que importam os dados disponíveis na base de acordo com o seu formato nativo e

³Jena: <https://jena.apache.org/>

⁴PhysioBank: <https://physionet.org/mimic2/>

disponibilizam as aferições de acordo com a frequência de leitura especificada, simulando o comportamento de um sensor físico.

A Tabela 1 contém amostras de exemplo dos registros de um paciente identificado por *033n*, coletados da *PhysioBank*. Essa amostra possui frequência de monitoramento definida com intervalo de 20 segundos entre aferições, considerada de alta precisão pela Equipe de Enfermagem do Hospital das Clínicas (UFG), que apoiou este projeto. A primeira coluna apresenta o momento da medição a partir do início do monitoramento, e as demais apresentam os dados aferidos para cada sinal vital, representado pelas suas respectivas siglas:

- *ABPdias*: Pressão sanguínea diastólica
- *ABPsys*: Pressão sanguínea sistólica
- *ABPMean*: Pressão sanguínea média
- *Temp*: Temperatura corpórea
- *Pulse*: Frequência de pulso
- *Resp*: Frequência respiratória
- *SpO2*: Saturação de oxigênio

Tabela 1. Amostra de registro da base MIMIC para o paciente 033n.

Intervalo	<i>ABPMean</i>	<i>ABPsys</i>	<i>ABPdias</i>	<i>Pulse</i>	<i>Resp</i>	<i>SpO2</i>	<i>Temp</i>
sec	mmHg	mmHg	mmHg	bpm	bpm	%	°C
0	81	138	54	53	25	94	37.5
20	78	134	53	53	30	94	37.5
40	79	135	53	52	36	94	37.5
60	78	133	53	52	26	94	37.5

4.2. Validação Funcional da Representação Ontológica

Como descrito anteriormente, o serviço de representação realizado pelo componente *Hermes Widget* utiliza a padronização Estímulo-Sensor-Observação (Figura 1), adotada pela ontologia SSN, no intuito de representar de maneira expressiva e interoperável as informações de um sensoreamento. Para cada informação de contexto aferida é criado um indivíduo do tipo *ssn:Observation*, ao qual serão adicionadas as propriedades:

- *ssn:observes*: que relaciona um sensor *ssn:SensingDevice* a uma propriedade observada *ssn:Property*;
- *ssn:isProducedBy*: relaciona a informação produzida *ssn:SensorOutput* ao sensor *ssn:SensingDevice* que a produz;
- *ssn:observedBy* e *ssn:observationResult*: ligam respectivamente um sensor e a informação produzida por ele, a uma observação *ssn:Observation*.

A Figura 5 apresenta o grafo gerado a partir do modelo de contexto armazenado no repositório que contém a persistência do sensoreamento de duas instâncias de medição de pressão sanguínea ⁵.

Neste exemplo a propriedade do tipo *ssn:Property* descrita é a pressão sanguínea, representada pela instância *ssn:BloodPressure*. Os sensores virtuais que observam esta propriedade são do tipo *ssn:SensingDevice*, e são descritos pelas instâncias *ssn:BloodPressureSensor001* e *ssn:BloodPressureSensor002*. O fato de existirem dois

⁵Por questões de espaço o modelo RDF que deu origem ao grafo não será apresentado. Este grafo representa o modelo na sua integridade.

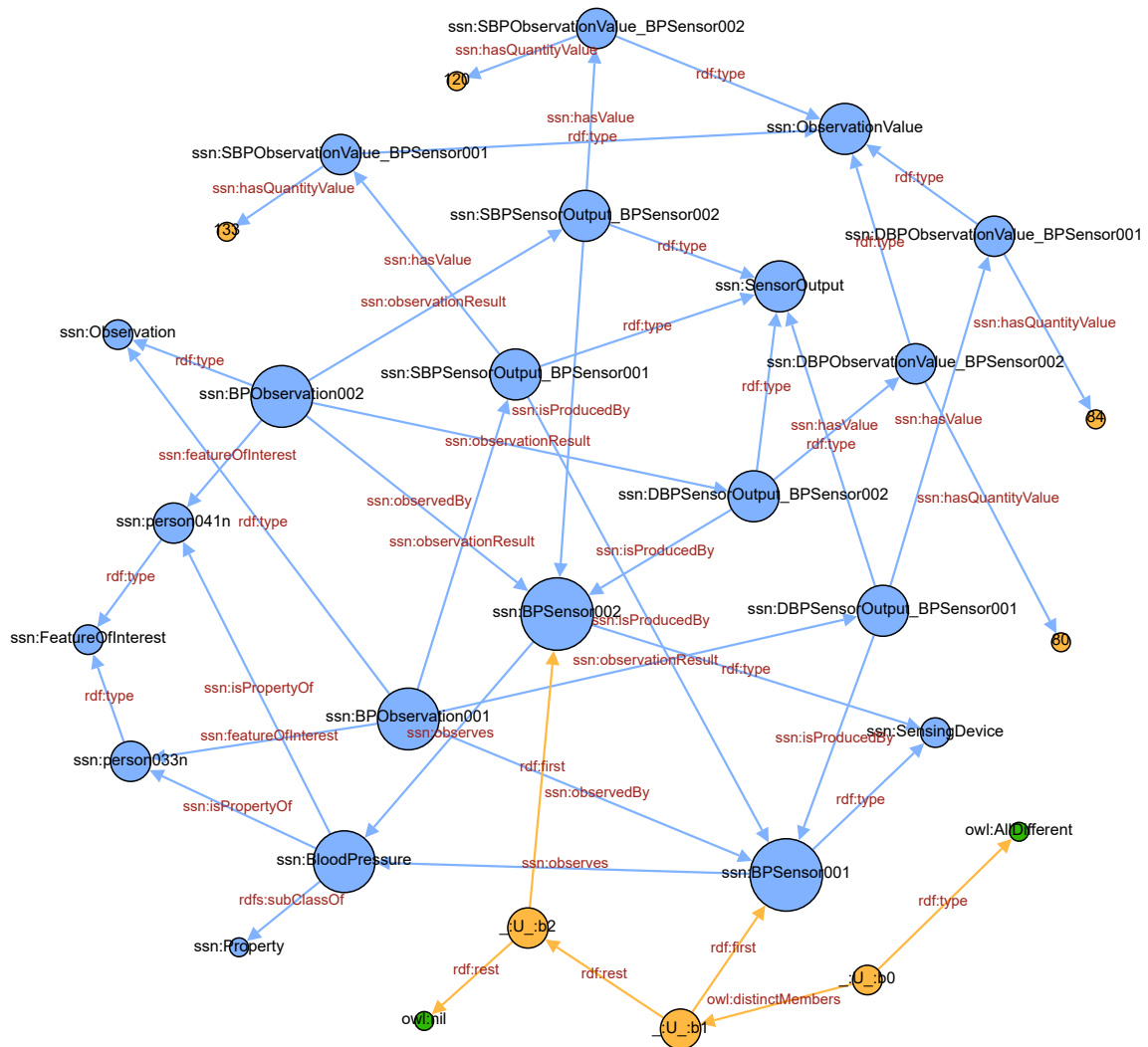


Figura 5. Grafo da representação SSN para duas instâncias de pressão arterial

sensores distintos para o mesmo tipo de sinal vital nesta representação significa que estas foram coletadas de pacientes diferentes.

Os resultados do monitoramento de contexto representado são apresentados por indivíduos do tipo *ssn:SensorOutput*, que são as saídas resultantes das aferições, que incluem os valores observados *ssn:hasValue*, que são representados pelos sensores e possuem um valor observado *ssn:ObservationValue* para cada medição aferida, neste caso (133, 84) e (120, 80). Todas estas propriedades são relacionadas em uma instância da classe *ssn:Observation*, que descreve o monitoramento realizado.

O padrão de projeto Estímulo-Sensor-Observação, implementado pela ontologia SSN, contribui em *Hermes Widget* para a definição de uma *representação de contexto padronizada*, que pode ser utilizada independente do domínio de aplicação. Dessa forma, além de oferecer grande facilidade para o compartilhamento das informações representadas, qualquer componente ou aplicação que utilize o contexto gerado por *Hermes Widget* tratará a informação de maneira análoga, seguindo o padrão que rege a sua estrutura. Portanto, além do formato de representação do contexto, a estrutura semântica também é

padronizada, independente da informação de contexto a ser representada.

Também foram realizados outros experimentos, como uma avaliação da escalabilidade do serviço de representação ontológica de contexto. Como resultado, verificou-se a capacidade do *Hermes Widget* em processar grandes volumes de dados sem impactos no tempo de processamento por representação, como descrito em [Veiga et al. 2016].

5. Trabalhos Relacionados

LoCCAM (*Loosely Coupled Context-Aware Middleware*) [Maia et al. 2013] é um *middleware* de arquitetura orientada a componentes que propõe a redução do acoplamento entre o processo de aquisição de contexto e as aplicações que o utilizam. Os componentes *Context Acquisition Components* (CACs) encapsulam o código de acesso a sensores para fornecer informação de contexto de alto nível.

Segundo o princípio de reusabilidade, CroCo (*Cross-application Context management Service*) [Mitschick et al. 2010] destaca a relevância de um sistema de representação independente de domínio, apresentando um modelo ontológico de propósito geral que trata os principais aspectos do contexto. Porém, CroCo assume que seus provedores já enviem o contexto na forma de triplas RDF, que é recebido pelo serviço *Context Update Service*. CroCo não define um padrão para a representação do contexto.

Em *infinitum* [Jaroucheh et al. 2012], os componentes provedores de contexto (CP's) realizam a representação de informações recebidas, gerando unidades de contexto chamadas *Contextlets*, que são uma parte específica da informação sobre uma determinada entidade. Uma limitação de *infinitum* é a falta de um processo para tornar a representação de contexto flexível em sua arquitetura.

O componente *Hermes Widget* possui uma ideia similar à dos CAC's de LoCCAM e da estrutura de CroCO e de *infinitum* quanto à separação entre aquisição/representação e utilização do contexto. Porém, diferente de LoCCAM, esta proposta utiliza a modelagem baseada em ontologias para oferecer maior expressividade e independência do modelo de contexto em relação a representação baseada em espaço de tuplas, realizada pelos CAC's. Diferendo em relação aos demais trabalhos relacionados, nesta proposta é apresentada a padronização da estrutura de representação ontológica de contexto, neste caso utilizando o padrão para sensoriamento Estímulo-Sensor-Observação.

6. Conclusões

Este artigo apresentou o componente *Hermes Widget*, que recebe dados de sensores e, através do processo de representação ontológica, gera um modelo de contexto altamente expressivo e em um formato de representação com apoio à interoperabilidade.

Como contribuições deste trabalho, destacam-se: *i)* um serviço de representação ontológica de contexto reutilizável e independente de domínio de aplicação; *ii)* modelo de representação de contexto padronizado de acordo com a ontologia SSN (padrão SSO); *iii)* representação de contexto compartilhável e interoperável entre sistemas e aplicações; e *iv)* validação da representação de contexto no cenário de monitoramento de sinais vitais humanos, apoiada por uma base de dados com aferições reais de sinais vitais de pacientes e uma equipe de enfermagem com experiência em UTI.

Destaca-se como trabalho futuro a utilização do componente *Hermes Widget* e do serviço de representação em outros cenários de aplicação além do monitoramento de sinais vitais. Também devem ser implementadas outras aplicações sensíveis a contexto que utilizam as informações representadas em diferentes domínios, no intuito de avaliar extensivamente a reusabilidade e a flexibilidade do componente proposto.

Agradecimentos

O autor Ernesto Fonseca Veiga agradece a CAPES pela bolsa de mestrado concedida a esta pesquisa.

Referências

- Bettini, C., Brdiczka, O., Henriksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive Mob. Comput.*, 6(2):161–180.
- Janowicz, K. and Compton, M. (2010). The stimulus-sensor-observation ontology design pattern and its integration into the semantic sensor network ontology. In *3rd International Conference on Semantic Sensor Networks - Volume 668*, pages 64–78.
- Jaroucheh, Z., Liu, X., and Smith, S. (2012). An approach to domain-based scalable context management architecture in pervasive environments. *Personal Ubiquitous Comput.*, 16(6):741–755.
- Maia, M. E. F., Fonteles, A., Neto, B., Gadelha, R., Viana, W., and Andrade, R. M. C. (2013). Locom - loosely coupled context acquisition middleware. In *28th Annual ACM Symposium on Applied Computing*, pages 534–541.
- Maranhão, G. M., Sene Júnior, I. G., and Bulcão Neto, R. F. (2014). Anatomy of a semantic context interpreter with real-time events notification support. In *20th Brazilian Symposium on Multimedia and the Web*, pages 159–162.
- Mitschick, A., Pietschmann, S., and Meißner, K. (2010). An Ontology-Based, Cross-Application Context Modeling and Management Service. *Int. J. Semant. Web Inf. Syst.*, 6(1):39–54.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *Communications Surveys Tutorials, IEEE*, 16(1):414–454.
- Sehic, S., Li, F., Nastic, S., and Dustdar, S. (2012). A programming model for context-aware applications in large-scale pervasive systems. In *IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 142–149.
- Veiga, E. F., Maranhão, G. M., and Bulcão Neto, R. F. (2016). Development and scalability evaluation of an ontology-based context representation service. *IEEE Latin America Transactions*, 14(3):1380–1387.
- Veiga, E. F., Maranhão, G. M., and Bulcão Neto, R. F. (2014). Apoio ao desenvolvimento de aplicações de tempo real sensíveis a contexto semântico. In *IX Workshop de Teses e Dissertações do XX Simpósio Brasileiro de Sistemas Multimídia e Web*, pages 1–4.

Uma Arquitetura Hierárquica Multinível para Ciência de Situação em Segurança da Informação

Ricardo Borges Almeida¹, Roger da Silva Machado¹,
Diórgenes Y. L. da Rosa¹, Lucas Medeiros Donato²,
Adenauer Corrêa Yamin¹, Ana Marilza Pernas¹

¹Programa de Pós-Graduação em Computação
Universidade Federal de Pelotas (UFPel), Pelotas – RS – Brasil

²De Montfort University – Cyber Security Centre – Leicester, Reino Unido

{rbalmeida, rdsmachado, diorgenes, adenauer, marilza}@inf.ufpel.edu.br,

lucas.donato@myemail.dmu.ac.uk

Abstract. *Information security precautions are inherent in computer systems, but the consequent loss of their mismanagement become larger in high demand connectivity systems, such as the ubiquitous systems. In this perspective, it is proposed in this paper an approach based on scalability, flexibility, heterogeneity and autonomy, aiming to provide situation awareness in ubiquitous environments. The differential of the approach is its multilevel hierarchical architecture designed over the distribution of situational awareness modules in three software components, which allows to provide increased autonomy to the components. The conception strategies and developed usage scenario showed that the approach offers autonomy presenting flexibility and scalability, proving to be opportune in today's distributed and heterogeneous environments.*

Resumo. *Cuidados com a segurança da informação são inerentes aos sistemas computacionais, mas as perdas consequentes de sua má administração se tornam maiores em sistemas com alta demanda de conectividade, como é o caso dos sistemas ubíquos. Nesta perspectiva, é proposta neste trabalho uma abordagem fundamentada na escalabilidade, flexibilidade, heterogeneidade e autonomia, visando fornecer ciência de situação à ambientes ubíquos. O diferencial da abordagem está em sua arquitetura hierárquica multinível projetada sobre a distribuição de módulos de ciência de situação em três componentes de software, o que permite fornecer maior autonomia aos componentes. As estratégias de concepção e o cenário de uso desenvolvido mostraram que a abordagem oferece autonomia apresentando caráter flexível e escalável, se mostrando oportuna para operar nos atuais ambientes distribuídos e heterogêneos.*

1. Introdução

O clássico artigo de Weiser (WEISER, 1991) destaca que o paradigma da Computação Ubíqua (UbiComp) tem como uma de suas premissas prover computação de forma transparente, estando o modelo computacional integrado às demandas do usuário. Nesta perspectiva, a UbiComp visa fornecer o acesso do usuário a seu ambiente computacional em qualquer lugar, todo o tempo, com qualquer dispositivo, de forma não-intrusiva, mantendo o foco do usuário em suas atividades [Lopes et al. 2014].

Uma das características inerentes às aplicações da UbiComp é a ciência de situação. A ciência de situação auxilia as aplicações ubíquas a perceberem modificações em seus contextos de interesse e, quando necessário, reagirem a esta nova situação. A construção de aplicações ubíquas cientes de situação apresenta uma série de desafios: aquisição de informações contextuais a partir de fontes heterogêneas e distribuídas; o processamento dessas informações na busca por situações de interesse; a respectiva atuação que poderá promover a adaptação funcional e não-funcional do ambiente; e o armazenamento e a busca de informações para disseminação aos usuários [Bellavista et al. 2012].

Os princípios da UbiComp e a necessária ciência de situação traz como consequência um aumento considerável na demanda por conectividade destes sistemas. Os desafios são intensificados devido à natureza volátil, espontânea, heterogênea e invisível de como ocorre a comunicação nos sistemas ubíquos [Langheinrich 2010]. Além disso, percebe-se o crescimento natural em tamanho, complexidade e distribuição das atuais infraestruturas computacionais [Onwubiko 2012a].

Tendo em conta estes desafios, o objetivo central deste trabalho é a concepção de uma abordagem denominada *Execution Environment for Highly Distributed Applications - Multilevel Hierarchical Architecture for Situation Awareness* (EXEHDA-MHASA) que visa o fornecimento de ciência de situação por meio de uma arquitetura hierárquica multinível projetada sobre a distribuição de módulos de ciência de situação em três componentes de software, possuindo como principal premissa proporcionar aspectos de escalabilidade, flexibilidade, heterogeneidade, autonomia enquanto estratégia para fornecer segurança a estes sistemas.

O presente artigo está organizado da seguinte forma: a seção 2 apresenta os conceitos relacionados a ciência de situação; a seção 3 discute a concepção da abordagem EXEHDA-MHASA; a seção 4 descreve um cenário de uso; a seção 5 discute os trabalhos relacionados; por fim, a seção 6 apresenta as considerações finais.

2. Ciência de Situação

Uma situação consiste de um conjunto de elementos contextuais de interesse instanciados e relacionados de forma a prover alguma informação válida em um intervalo de tempo específico. Tendo definido o termo situação, a ciência de situação consiste da percepção e compreensão de uma ou mais informações contextuais e a projeção de seus efeitos em um futuro próximo [Onwubiko 2012a]. Percebe-se, então, a existência de três níveis para a obtenção da Ciência de Situação:

- percepção: envolve os processos de detecção, reconhecimento e monitoramento, que levam a consciência de múltiplos elementos situacionais (objetos, eventos, pessoas, sistemas, fatores ambientais) e seus estados atuais (locais, condições, formas, ações);
- compreensão: síntese e correlação dos elementos desconexos identificados no nível de percepção por intermédio de diferentes estratégias de reconhecimento de padrões, interpretação e avaliação. Este nível requer a integração dessas informações para entender como isso vai impactar as metas e objetivos do indivíduo/sistema. Isto é normalmente realizado com o apoio de soluções de processamento de eventos;

- projeção: responsável pela capacidade de antecipação de ocorrências futuras com base na compreensão dos elementos no ambiente atual. A projeção é alcançada por meio do conhecimento da situação, da dinâmica dos elementos e da compreensão da situação, para depois projetar esta informação adiante no tempo e, assim, determinar se elas afetarão os futuros estados do ambiente.

Esta definição de ciência de situação leva a um fluxo onde, na etapa de percepção, os eventos são inicialmente registrados por sensores distribuídos, coletados e, na sequência, contextualizados. Na etapa de compreensão, estes eventos contextualizados são processados buscando identificar possíveis situações de interesse. Finalmente, na projeção as situações detectadas são projetadas adiante no tempo, fornecendo subsídio para a tomada de decisão, podendo realizar atuações diretamente no ambiente monitorado.

3. EXEHDA-MHASA

A abordagem concebida, denominada EXEHDA-MHASA, é caracterizada principalmente pelo provimento de ciência de situação, através de uma arquitetura hierárquica multinível que trata os desafios de escalabilidade, flexibilidade, heterogeneidade e autonomia. A EXEHDA-MHASA foi baseada no *middleware* EXEHDA, beneficiando-se, em especial, da normatização já definida pelo *middleware* quanto aos recursos que compõem o ambiente ubíquo. No EXEHDA, os recursos da infraestrutura física são mapeados para três abstrações básicas, as quais são utilizadas na composição do ambiente ubíquo: a EXEHDAcel; a EXEHDAbase; e o EXEHDA nodo [Lopes et al. 2014].

A EXEHDA-MHASA foi concebida por meio de três componentes de software, sendo o “Collector” proposto com base no EXEHDA nodo e o “SmartLogger” e o “Manager” baseados na EXEHDAbase. Estes três componentes foram desenvolvidos em Python, tendo seus detalhes de concepção descritos nas próximas subseções.

A comunicação entre os componentes estabelece um fluxo de eventos e situações, propondo a formação de uma hierarquia multinível, onde cada nível representa a ciência de situação sobre determinado escopo, unidade ou localização geográfica. Seguindo as premissas do EXEHDA, esta hierarquia pode ser alterada pela agregação ou remoção dinâmica de nodos, permitindo um crescimento vertical e horizontal.

Cada componente foi concebido para ser autônomo, ou seja, em caso de falha de algum componente na hierarquia, os demais continuam operacionais, realizando as detecções de situações de interesse, assim como as atuações pertinentes de acordo com as configurações definidas pelos administradores do sistema. Ainda assim, a arquitetura provê uma estratégia de recuperação de falha (*failover*), onde cada componente pode possuir uma lista de prioridades em suas configurações que consiste de até três endereços para o próximo nível na hierarquia, ou seja, caso a comunicação com um componente superior falhe, o sistema tentará o próximo.

Nessa hierarquia, por padrão, todos os eventos de sensores pré-configurados no componente Collector são encaminhados para o SmartLogger residente na mesma EXEHDAcel para o processamento e armazenamento de eventos. Se o administrador desejar, o sistema oferece a opção de filtragem de eventos, possibilitando que alguns eventos não sejam encaminhados ao SmartLogger. A ideia geral é que a cada nível da hierarquia sejam

repassadas informações/situações com significado cada vez maior e, conseqüentemente, em menor quantidade, diminuindo o tráfego na rede e aprimorando assim a tomada de decisões.

3.1. EXEHDA-MHASA: Collector

Este componente de software foi projetado para ser implantado em um hardware dedicado ou internamente a um dispositivo, como servidores que oferecem serviços de rede (web, e-mail, banco de dados, entre outros). O Collector oferece a possibilidade de realizar localmente a compreensão dos eventos coletados em busca de situações de interesse e, se necessário, executar a decorrente atuação (na etapa de projeção), caracterizando em parte a flexibilidade e a autonomia dos componentes da abordagem proposta.

A etapa de percepção deste componente realiza a identificação e o monitoramento de sensores físicos e lógicos, sendo concebida modularmente para permitir expansão visando fornecer heterogeneidade a proposta, bem como possibilitar a flexibilidade e distribuição da arquitetura, por exemplo, explorando protocolos de comunicação para recebimento de eventos de dispositivos externos.

O conjunto de serviços ativos é controlado pelo seu perfil de execução, o qual deverá ser configurado pelo administrador do sistema buscando refletir as demandas de ciência de situação do ambiente computacional, o que também colabora com a necessidade de heterogeneidade da solução.

Adicionalmente, objetivando aprimorar a capacidade de percepção da abordagem, agregando flexibilidade, heterogeneidade e dinamicidade, foi desenvolvida a capacidade de descoberta automática dos recursos internos ao Collector (interfaces de rede, partições, logs, entre outros), bem como situações a serem avaliadas, com base na especificação de parâmetros nas configurações dos sensores e das situações.

Considerando as necessidades de normalização, contextualização, categorização e priorização dos eventos coletados, e as demandas de processamento de eventos, as etapas de pré-processamento e de compreensão foram projetadas apresentando uma sintaxe alternativa ao tradicional uso de expressões regulares. Em ambos os casos as sintaxes utilizadas visam fornecer uma melhor legibilidade e facilidade na criação e adaptação das expressões existentes, o que novamente colabora com os requisitos de heterogeneidade [McGuire 2007], [EsperTech 2015]. Além disso, as modificações nas regras são carregadas dinamicamente, promovendo a adaptação funcional.

A compreensão dos eventos coletados poderá ser executada em um dispositivo dedicado contribuindo para a escalabilidade da arquitetura concebida. Finalmente, como consequência das situações detectadas na etapa de compreensão, a atuação poderá ser executada de forma distribuída, ou seja, em qualquer dispositivo do ambiente monitorado.

3.2. EXEHDA-MHASA: SmartLogger

O SmartLogger foi projetado para receber eventos de diferentes Collector's e/ou SmartLogger's com o objetivo de fornecer a ciência de situação sobre os dispositivos sob a sua coordenação. Em outras palavras, ele oferece a visão sobre o ambiente monitorado considerando a abrangência da célula (EXEHDAcel) onde ele está inserido, ou ainda, a amplitude de células subordinadas a sua dentro da hierarquia. Este componente foi concebido para ser implantado preferencialmente em um dispositivo dedicado e, além de

oferecer a percepção, compreensão e projeção dos eventos e situações recebidas, ele oferece um repositório para armazenamento dessas informações, permitindo que elas sejam disponibilizadas em uma interface aos administradores do sistema.

Assim como no Collector, as etapas de compreensão e de projeção são opcionais, contribuindo para a flexibilidade e autonomia dos componentes da EXEHDA-MHASA. No SmartLogger, o controle do conjunto de serviços ativos também é realizado com base no seu perfil de execução.

Este componente, por ser baseado na EXEHDAbase, pode ser distribuído entre vários equipamentos, contribuindo para sua escalabilidade. Ainda, no que diz respeito a esta demanda, em caso de sobrecarga deste componente uma possível solução com base na arquitetura é a adição de um novo SmartLogger no mesmo nível arquitetural, realizando a divisão dos dispositivos sob sua coordenação.

Considerando o aprimoramento da visibilidade dos eventos e situações identificadas em cada nível da arquitetura, a compreensão realizada pelo SmartLogger fornece a capacidade de correlacionar eventos de diferentes dispositivos (correlação cruzada) e permite que os dados históricos sejam consultados no repositório local de eventos e situações.

3.3. EXEHDA-MHASA: Manager

O Manager foi concebido no intuito de centralizar a visualização da ciência de situação sobre o ambiente ubíquo como um todo. Assim como o SmartLogger, por ser baseado no EXEHDAbase, ele poderá empregar estratégias de distribuição dos serviços, fornecendo escalabilidade. Esta característica também poderá ser explorada pelos aspectos arquiteturais onde, por exemplo, caso ocorra uma sobrecarga do Manager, podem ser instanciados dois SmartLoggers para divisão da atual carga de responsabilidade exclusiva do Manager, passando os novos SmartLogger's a enviarem os eventos e/ou situações já tratados para o Manager.

A percepção realizada pelo Manager foi projetada para receber eventos de diferentes Collector's e/ou SmartLogger's, com o objetivo de aprimorar o nível de ciência de situação sobre os diversos dispositivos sob a sua coordenação na hierarquia da arquitetura. Igualmente ao SmartLogger, este componente foi concebido para ser implementado preferencialmente em um hardware dedicado.

A compreensão e a projeção também foram projetadas com características similares as disponíveis no SmartLogger, incluindo a correlação cruzada e a atuação distribuída.

Para suportar o armazenamento de eventos e situações no Manager, junto as configurações dos perfis de execução dos demais componentes por meio das *templates*, foi proposto um repositório híbrido de informações contextuais, o qual, em sua implementação foi composto de um modelo não-relacional para eventos e situações e outro relacional herdado do EXEHDA para as demais configurações. Cada modelo deste repositório pode ser formado por vários dispositivos com hardware independente dos demais módulos do Manager, possibilitando a utilização de estratégias de redundância, distribuição e balanceamento de carga.

4. Cenário de Uso

A preocupação com a segurança da informação tem aumentado nos últimos anos, e isto é consequência natural do aumento do número dos incidentes e das perdas financeiras decorrentes. Visando o fornecimento de uma visão integral sobre a segurança das infraestruturas computacionais, Tim Bass (1999) propôs a aplicação dos conceitos de ciência de situação no campo da segurança em redes de computadores. Embora a união destas duas áreas venha sendo estudada há aproximadamente quinze anos, ela ainda constitui um foco de estudo e pesquisa relevante na área de segurança da informação. Por sua vez, é importante registrar que os riscos de segurança tem se potencializado devido as características da UbiComp presentes nas atuais infraestruturas [Onwubiko 2012b].

Um dos desafios de segurança nas atuais infraestruturas distribuídas é a identificação de ataques realizados a partir da mesma fonte à diferentes serviços que podem operar sobre o mesmo dispositivo, sobre dispositivos diferentes, ou ainda, localizados geograficamente distantes. Desta forma, este estudo de caso visa englobar eventos provenientes de diferentes dispositivos, fornecendo assim uma melhor visibilidade sobre a situação de segurança do ambiente. Com isso, será explorada a correlação com vulnerabilidades previamente detectadas e o uso de dados históricos.

O estudo de caso desenvolvido teve como referência o ambiente computacional da Universidade Federal de Pelotas (UFPel) que possui uma infraestrutura computacional com características da UbiComp. Dentre os diversos prédios da UFPel, estão o Campus Anglo (CA) e Campus Capão do Leão (CCL), nos quais ficam alocados os dois datacenters da UFPel. A figura 1 apresenta a infraestrutura concebida para avaliação da EXEHDA-MHASA, onde é exibida a disposição dos componentes Collector, SmartLogger e Manager, destacando o fluxo de comunicação dos eventos e situações. Observa-se também os servidores que possuem o HIDS OSSEC e o Collector executando localmente.

Inicialmente, foi realizado um ataque de força bruta ao serviço de *File Transfer Protocol* (FTP) que executa no servidor denominado *webserver2*. Isso provocou a detecção do ataque pelo OSSEC [OSSEC 2016a] que executa no próprio servidor. Os eventos registrados pelo OSSEC foram coletados, pré-processados e correlacionados no módulo de compreensão pelo Collector implantado no servidor, o que implicou na detecção de uma situação de interesse. Como projeção para esta situação foi realizado o bloqueio temporário do endereço *Internet Protocol* (IP) de origem - identificado na normalização como SRCIP - no *firewall* interno do *webserver2*. Os eventos do OSSEC foram armazenados no SmartLogger, enquanto que a situação foi salva no Manager.

Esta primeira situação detalhada validou a autonomia do Collector, visto o emprego da compreensão e projeção sem utilizar os demais componentes. Além disso, esta situação gerou dados históricos que serão explorados na sequência.

Para validar a correlação cruzada foi realizado um ataque de força bruta em dois servidores FTP que operam no *webserver1* e *webserver2*, a partir de um novo endereço IP. Os eventos registrados pelo OSSEC analisados separadamente não levariam a detecção de uma situação de interesse, pois o mesmo considera que eventos de nível 5 são erros gerados pelo usuário, estando a uma passo do nível 6 que seriam ataques de baixa relevância [OSSEC 2016b].

Dessa forma, os eventos após tratados pelos Collector's de ambos os servidores,

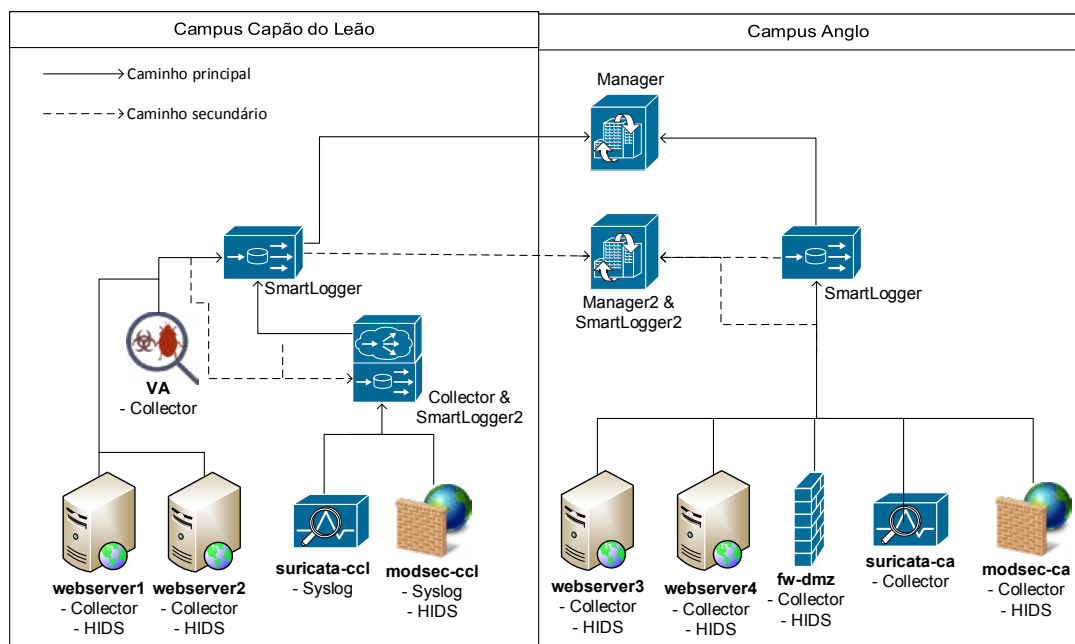


Figura 1. Cenário de uso - fluxo de comunicação dos componentes da EXEHDA-MHASA

são repassados ao SmartLogger, o qual correlaciona os eventos provenientes dos diferentes servidores, identificando uma situação de interesse quando houverem 6 ocorrências de eventos da regra representada pelo identificador 11302 [OSSEC 2016a] do OSSEC, em um intervalo de 120 segundos. Isto provocou a identificação do ataque com maior antecedência, ou seja, antes do atacante realizar as 6 tentativas em cada servidor no intervalo mencionado, provocando uma resposta mais rápida pela atuação na etapa de projeção que realizou o bloqueio no *firewall* do *webserver1* e do *webserver2*. A organização arquitetural no cenário também é importante visto que o SmartLogger foi alocado na mesma célula, neste caso, também mesma infraestrutura física dos servidores mencionados, implicando na independência de comunicação com um possível servidor central geograficamente distante, o que poderia ser necessário em uma arquitetura cliente-servidor.

Logo em seguida, para validar a integração com a análise de vulnerabilidades e uso dos dados históricos, foi iniciado um ataque ao servidor *webserver3* hospedado em outra célula, utilizando o mesmo endereço IP que realizou o primeiro ataque ao *webserver2*. Como consequência, o servidor *modsec-ca* que executa um *Web Application Firewall* operando na frente do *webserver3* como um *proxy* reverso, detectou vários acessos como suspeitos, registrando-os nos logs que foram coletados na etapa de percepção do componente Collector. Estes logs foram pré-processados e repassados para a tarefa de compreensão que identificou uma nova situação. Como projeção, o Collector realizou o bloqueio no *firewall* do servidor.

Os eventos e situações desse ataque foram repassados ao SmartLogger, o qual realiza uma busca no banco de dados de vulnerabilidades previamente identificadas por scripts personalizados que executam no servidor VA. Visto que estes scripts já haviam identificado vulnerabilidades de injeção de *Structured Query Language* (SQL) no refe-

rido servidor, e que as tentativas de ataque registradas são desse mesmo tipo, a situação recebe essas novas informações contextuais (tipo da vulnerabilidade e seus dados). A situação é enviada ao Manager que, ao realizar a compreensão, identifica a existência das informações sobre vulnerabilidades e, como consequência, realiza uma busca na base local (não-relacional) por situações com o mesmo IP de origem, encontrando a situação de ataque ao serviço de FTP registrada na outra célula. Como projeção é realizado o bloqueio do IP em todos os *firewalls* de cada servidor que opera na DMZ do CCL e no servidor fw-dmz do CA.

Com este estudo de caso, destaca-se principalmente, a visão integrada sobre os eventos de segurança gerados por diferentes tecnologias, onde foi possível integrar os ataques realizados nas duas células que estão em infraestruturas geograficamente distantes. Evidencia-se também a flexibilidade da solução, visto a diferença de complexidade das situações, onde foi explorada a união com um sistema de análise de vulnerabilidades. Adicionalmente, é importante ressaltar a exploração de dados históricos das situações previamente identificadas que foram utilizadas para inferir a possibilidade de que o atacante continuaria realizando tentativas de ataques aos servidores. Finalmente, a flexibilidade é também expressa pela diferença das projeções onde foram empregadas ações em um único dispositivo e sobre dispositivos dispersos.

5. Trabalhos Relacionados

A partir da revisão bibliográfica realizada, foi possível identificar dois trabalhos que exploram a concepção ou a implantação de uma arquitetura para soluções *Security Information and Event Management* (SIEM), o que se aproxima da EXEHDA-MHASA.

5.1. SIEM Implementation for Global and Distributed Environments

No trabalho [Anastasov and Davcev 2014], é proposto um modelo e uma arquitetura para implementação de sistemas SIEM que utiliza múltiplos gerenciadores SIEM hierárquicos. O modelo é chamado de “*Hierarchical Managers Model*”. Os autores demonstram como esse modelo e arquitetura poderiam ser criados sobre a HP ArcSight ESM [Hewlett-Packard 2016]. O modelo proposto define que o servidor SIEM central, denominado *Parent Manager*, se comunica com servidores SIEM intermediários (chamados de *Child Managers*). Por sua vez, cada *Child Manager* coleta dados de algumas origens de eventos, normalmente a partir de uma região ou local específico. Os *Child Managers* regionais coletam e armazenam dados, e em seguida, normalizam os eventos antes de passá-los para o servidor SIEM central que realizará a agregação, correlação e relatórios. Dados brutos permanecem nos *Child Managers* locais para fins forenses.

Além de não explorar os conceitos de ciência de situação, o trabalho não apresenta estudos de caso que demonstrem ou quantifiquem as vantagens do modelo proposto. Além disso, não fica formalizada a área de atuação dos *Child Managers*, apenas é informado que o mesmo será responsável por uma implantação SIEM regional, enquanto que na EXEHDA-MHASA, por ser baseado no *middleware* EXEHDA, existe a definição de uma EXEHDA Cel. Neste modelo, também não foram considerados ou expostos possíveis problemas de sobrecarga dos *Child Managers*.

No que diz respeito a arquitetura, apesar deste trabalho apresentar uma abordagem hierárquica, e do componente Manager da HP ArcSight possibilitar o redirecionamento

de eventos para outra implantação da ArcSight, não foi identificado um cenário com mais de três níveis. Além disso, não fica claro o método ou configuração empregada para o repasse dos eventos ou incidentes de um *Child Manager* para o *Parent Manager*, diferentemente da EXEHDA-MHASA que é por meio das regras de compreensão. Finalmente, este trabalho destaca que sua aplicação é indicada para grandes empresas, enquanto que a EXEHDA-MHASA pode ser implantada em pequenas, médias e grandes empresas, no primeiro caso especificamente explorando apenas os componentes Collector e Manager.

5.2. Security Information and Event Management (SIEM) für Klein- und Mittelständische Unternehmen (KMU)

O projeto *Security Information and Event Management (SIEM) für Klein- und Mittelständische Unternehmen (KMU)* (SIMU), disponível em [SIMU 2015], apresenta o desenvolvimento de uma solução semelhante a sistemas SIEM, a qual visa melhorar significativamente a segurança do ambiente tecnológico em redes corporativas sem consideráveis esforços. A arquitetura do SIMU é baseada em componentes que podem ser superficialmente divididos em duas camadas: os coletores e controladores de fluxo; e o motor SIMU [SIMU 2015]. O cenário de avaliação apresentado demonstrou a viabilidade da proposta.

Apesar desse trabalho se assemelhar a EXEHDA-MHASA, realizando também a utilização de soluções sem custo e de código aberto, um dos pontos que deve ser destacado é que apesar de possuir uma arquitetura distribuída por diferentes componentes, a mesma é parcialmente baseada no modelo cliente-servidor, conseqüentemente possui a escalabilidade limitada. A utilização do protocolo IF-MAP, o qual é empregado para comunicação entre clientes e servidores, implica na dependência dos fornecedores empregarem este protocolo. O trabalho não explora a autonomia dos componentes que realizam a coleta, ou seja, em caso de falhas na comunicação, a reposta a possíveis ataques fica comprometida. Por fim, destaca-se que o trabalho não emprega em sua essência os conceitos de ciência de situação.

6. Considerações Finais

Por meio da revisão bibliográfica realizada, foi possível identificar algumas características necessárias para o desenvolvimento de soluções cientes de situação que contemplem os desafios da UbiComp, como a necessidade de flexibilidade, escalabilidade, autonomia e o suporte à heterogeneidade. Estes desafios foram buscados na concepção, prototipação e avaliação da abordagem concebida.

Desta forma, a concepção da EXEHDA-MHASA visou explorar a obtenção de ciência de situação por meio de uma arquitetura hierárquica multinível projetada sobre a distribuição de módulos de ciência de situação em três componentes de software: Collector, SmartLogger e Manager. Após a prototipação da EXEHDA-MHASA e dos testes realizados, conclui-se que a abordagem concebida contribui no tratamento dos desafios da UbiComp, permitindo a detecção distribuída de situações de interesse oferecendo escalabilidade, flexibilidade, heterogeneidade e autonomia.

Como trabalhos futuros pretende-se desenvolver novos testes que busquem quantificar as melhorias oferecidas pela arquitetura concebida. Além disso, espera-se analisar diferentes estratégias que podem ser aplicadas no nível de compreensão, a fim de melhorar

a identificação de situações de interesse, seja para avaliação de eventos onde a incerteza esteja presente, ou aprimorar questões de desempenho, escalabilidade, heterogeneidade, entre outros desafios.

Outra opção de continuação deste trabalho é a avaliação da possibilidade de reconstruir um nodo por meio da comunicação entre os nodos do nível hierárquico inferior, proporcionando maior resiliência à Ciência de Situação.

Referências

- Anastasov, I. and Davcev, D. (2014). Siem implementation for global and distributed environments. In *Computer Applications and Information Systems (WCCAIS), 2014 World Congress on*, pages 1–6.
- Bass, T. (1999). Multisensor data fusion for next generation distributed intrusion detection systems. In *In Proceedings of the IRIS National Symposium on Sensor and Data Fusion*, pages 24–27.
- Bellavista, P., Corradi, A., Fanelli, M., and Foschini, L. (2012). A survey of context data distribution for mobile ubiquitous systems. *ACM Comput. Surv.*, 44(4):24:1–24:45.
- EsperTech (2015). Esper reference version 5.3.0. *EsperTech Inc. - Event Series Intelligence*.
- Hewlett-Packard (2016). Acesso em 26 de maio de 2016. Disponível em: <http://www8.hp.com/us/en/software-solutions/siem-security-information-event-management/index.html>.
- Langheinrich, M. (2010). *Privacy in Ubiquitous Computing*. J. Krumm, ed., CRC Press.
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014). A middleware architecture for dynamic adaptation in ubiquitous computing. *j-jucs*, 20(9):1327–1351.
- McGuire, P. (2007). *Getting Started with Pyparsing*. O’Reilly, first edition.
- Onwubiko, C. (2012a). *Situational Awareness in Computer Network Defense: Principles, Methods and Applications: Principles, Methods and Applications*. Information Science Reference.
- Onwubiko, C. (2012b). *Situational Awareness in Computer Network Defense: Principles, Methods and Applications: Principles, Methods and Applications*. Premier reference source. Information Science Reference.
- OSSEC (2016a). Acesso em 26 de maio de 2016. Disponível em: <http://ossec.github.io>.
- OSSEC (2016b). Acesso em 26 de maio de 2016. Disponível em: ossec-docs.readthedocs.org/en/latest/manual/rules-decoders/rule-levels.html.
- SIMU (2015). Acesso em: 06 dez 2015. SIMU-project. Disponível em: <http://simu-project.de/english/project/index.html>.
- Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3):66–75.

Uma arquitetura para IoT direcionada à ciência do contexto baseada em eventos distribuídos

Rodrigo Souza^{1,4}, João Lopes^{1,4}, Anderson Cardozo³, Tainã Carvalho²,
Patrícia Davet², Alexandre Wolf⁵, Adenauer Yamin^{2,3}, Jorge Barbosa⁵, Cláudio Geyer¹

¹ Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brasil

² Universidade Federal de Pelotas (UFPel) – Pelotas – RS – Brasil

³ Universidade Católica de Pelotas (UCPel) – Pelotas – RS – Brasil

⁴ Instituto Federal Sul-rio-grandense (IFSUL) – Pelotas – RS – Brasil

⁵ Universidade do Vale do Rio dos Sinos (Unisinos) – São Leopoldo – RS – Brasil

{rsslouza, jlblopes, geyer}@inf.ufrgs.br

{ptdavet, trcarvalho}@inf.ufpel.edu.br

{anderson.cardozo, adenauer}@ucpel.edu.br

{awolf, jbarbosa}@unisinos.br

Abstract. *The recent advances in the Internet of Things (IoT) area, which has provided an increasing availability of networked sensors and actuators, has given a new perspective to research in the context awareness in UbiComp. In this sense, the main contribution of this paper is the proposition of COIoT, a distributed architecture for IoT designed with the aim at providing, through rules, the proactive management of the EXEHDA Middleware interactions with the physical environment. To evaluate the functionalities of the proposed architecture we implemented a case study in the agricultural area. The achieved results are promising.*

Resumo. *Os avanços recentes no campo de Internet das Coisas (IoT - Internet of Things), que têm promovido uma disponibilidade cada vez maior de sensores e atuadores conectados em rede, trouxe uma nova perspectiva às pesquisas em ciência de contexto na Computação Ubíqua (UbiComp). Neste sentido, a principal contribuição deste artigo é a proposta do COIoT, uma arquitetura distribuída e orientada a eventos para a IoT construída com o objetivo de prover, através de regras, o gerenciamento proativo das interações do meio físico com Middleware EXEHDA. Para validar as funcionalidades da arquitetura proposta foi implementado um estudo de caso na área da agricultura. Os resultados obtidos foram promissores.*

1. Introdução

Na Computação Ubíqua os sistemas computacionais devem ser capazes de reagir às mudanças de estado das diferentes variáveis contextuais de seu interesse. Essas variáveis contextuais devem ser coletadas em ambientes altamente distribuídos [Knappmeyer et al. 2013]. Por sua vez, os avanços científicos e tecnológicos

recentes no campo de IoT têm viabilizado a utilização de sensores em larga escala, os quais constituem fontes geradoras de informações contextuais para as aplicações ubíquas cientes de contexto [Perera et al. 2014].

Diversos desafios de pesquisa relacionados ao uso da IoT na obtenção de informações contextuais são associados com as diferenças entre os requisitos de alto-nível das aplicações ubíquas e as tarefas de gerenciamento dos dispositivos da IoT, as quais são relacionadas com as características eletrônicas envolvidas [Perera et al. 2014].

A principal contribuição deste artigo consiste em preencher essa lacuna através da proposição do COIoT (Context + IoT), uma arquitetura integrada ao *Middleware EXEHDA* capaz de prover suporte ao tratamento de sensores e atuadores. O EXEHDA (Execution Environment for Highly Distributed Applications) [Lopes et al. 2014a] provê uma arquitetura de software baseada em serviços que visa a criação e gerenciamento de um ambiente ubíquo, bem como a execução de aplicações sobre este ambiente.

O COIoT [Souza et al. 2015] é uma arquitetura baseada em eventos, gerenciada por regras, com processamento distribuído de contexto, capaz de agir proativamente na coleta das informações contextuais do ambiente físico, bem como atuar remotamente sobre o mesmo.

Este artigo está organizado da seguinte forma: A seção 2 apresenta a modelagem do COIoT, detalhando a sua arquitetura e funcionalidades. Na seção 3, é apresentado o protótipo e os testes realizados na área da agricultura. Os trabalhos relacionados são apresentados na seção 4. Por fim, na seção 5, são realizadas as considerações finais deste artigo.

2. COIoT: concepção e modelagem

A abordagem de tratamento de contexto proposta para o COIoT tem suas funcionalidades distribuídas entre dois tipos de servidores: Servidor de Contexto e Servidor de Borda. O Servidor de Borda foi concebido para atuar, principalmente, no gerenciamento das interações com o ambiente físico. Por sua vez, o Servidor de Contexto atua no armazenamento e processamento de informação contextuais [Lopes et al. 2014b].

2.1. Modelo arquitetural proposto

A arquitetura proposta para o COIoT, apresentada na Figura 1, foi concebida com o objetivo de gerenciar diferentes dispositivos da IoT, como sensores e atuadores heterogêneos. Esta arquitetura tem por premissa atuar de forma autônoma, tanto na coleta e processamento das informações contextuais, quanto na atuação sobre o ambiente, uma vez que essas atividades continuam a serem executadas mesmo nos períodos nos quais as aplicações interessadas no seu uso estejam inoperantes.

O processamento do contexto no COIoT se dá de forma distribuída entre os Servidores de Borda e Contexto. O módulo *Motor de Regras* (Servidor de Borda) constitui o primeiro nível de processamento, enquanto o *Processador de Contexto* (Servidor de Contexto) o segundo nível. As regras submetidas ao Motor de Regras devem ser elaboradas de forma a atender, prioritariamente, os eventos críticos, cujo tratamento deve ser realizado no menor tempo possível e com mínimo de falhas. Isso se deve ao fato de que o Servidor de Borda é geralmente alocado fisicamente próximo ao ambiente monitorado,

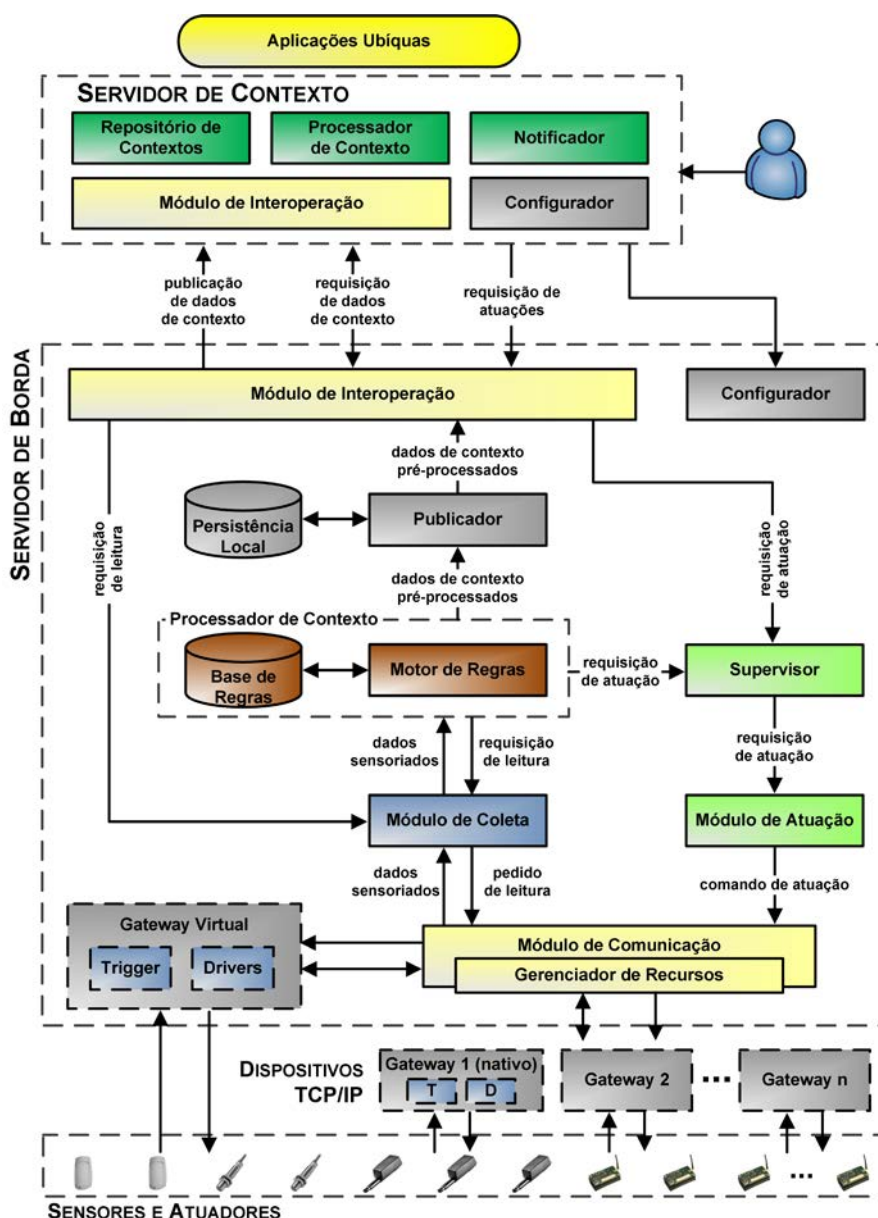


Figura 1. Arquitetura do COIOT

permitindo uma atuação (alertas, ativação/desativação de equipamentos eletromecânicos, etc.) independentemente de uma eventual perda de comunicação com o Servidor de Contexto por decorrência de uma falha na rede. Por outro lado, regras que necessitem incluir o tratamento de informações históricas, acessar dados coletados de outros Servidores de Borda, ou que envolvam outros modelos de processamento de contexto, devem ser processadas no Servidor de Contexto. Ambos os módulos de processamento do contexto foram concebidos tendo como base o modelo de regras tipo ECA (evento-condição-ação) [Terflath 2009], as quais podem ser disparadas a partir de eventos produzidos pelo ambiente. Embora seja utilizado o modelo ECA como mecanismo básico de tratamento do contexto, tanto a condição a ser tratada quanto a ação a ser executada pela regra admitem outros modelos de processamento que podem ser chamados a partir da regra, os quais são decorrência do tipo de domínio de aplicação a ser atendida pelo COIOT.

O módulo *Repositório de Contextos* utiliza um modelo relacional para a representação das informações contextuais, o qual proporciona um registro histórico desses dados. A estrutura do Repositório de Contexto reflete a organização da arquitetura do *Middleware* EXEHDA, contemplando assim o relacionamento entre aplicações, componentes, sensores, ambientes e contextos de interesse. O repositório também armazena a tabela de dados de configuração da arquitetura e as publicações dos sensores existentes no ambiente ubíquo. Estes dados são utilizados pelo módulo Processador de Contexto para disparar as ações apropriadas dependendo das informações contextuais.

Considerando a característica inerentemente distribuída das aplicações ubíquas, os *Módulos de Interoperação* do COIOT foram concebidos para promover a interoperabilidade entre os Servidores de Borda e de Contexto, assim como com outros serviços do *Middleware*. A concepção deste módulo teve como referência o estilo arquitetural REST [Fielding 2000].

O *Notificador* tem a função de gerar notificações a partir dos resultados do processamento do contexto realizado pelo Processador de Contexto. Este módulo utiliza uma estratégia de notificação baseada no modelo *publish/subscribe*, em que recebe subscrições de todos os serviços e/ou aplicações que requerem notificações a respeito das mudanças nos estados dos contextos.

Todas as configurações necessárias para o funcionamento do COIOT são gerenciadas através de uma interface web disponibilizada pelo módulo *Configurador*. Entre as funcionalidades oferecidas tem-se: a configuração de sensores e atuadores (inclusão, remoção e alteração), o gerenciamento de *drivers* de dispositivos, gerenciamento das regras de processamento de contexto, configuração de acesso aos Servidores de Contexto e Borda, entre outros.

O módulo *Publicador* tem a função de disparar as requisições de envio de informações contextuais para as demais camadas do *Middleware*, interoperando com o Servidor de Contexto através do Módulo de Interoperação. As publicações são organizadas em um sistema de fila FIFO e são processadas conforme a disponibilidade da rede. Considerando as possíveis falhas de comunicação entre o Servidor de Borda e o Servidor de Contexto, bem como eventuais atrasos da rede, foi concebido o módulo de Persistência Local cuja função é realizar o armazenamento temporário da fila de informações contextuais até que as mesmas sejam publicadas.

Com o intuito de garantir a interoperabilidade com tecnologias de mercado, e também potencializar a distribuição das iniciativas de coleta e/ou atuação, foram utilizados *gateways* de dois tipos: (i) Gateway proprietário, que tem funcionalidades heterogêneas variando de acordo com seus fabricantes, e o; (ii) Gateway nativo cujas funcionalidades operam de maneira integradas à arquitetura do COIOT. O módulo *Gateway Virtual* age como uma virtualização do Gateway nativo e implementa dois tipos de módulos básicos: *Drivers* e *Triggers*. *Drivers* são módulos arquiteturais responsáveis pelo acesso aos valores das grandezas físicas capturadas pelos sensores, bem como pelas execuções de comandos enviados para os atuadores. Os *Drivers* encapsulam e controlam os sensores e atuadores de uma maneira individualizada, evitando que diferenças operacionais destes dispositivos se propaguem a outros componentes da arquitetura. O *Trigger* gerencia a leitura de sensores através de eventos, tendo sido concebido para lidar com os

dois principais tipos de eventos a serem tratados na IoT: eventos temporais e eventos do ambiente [Perera et al. 2014].

O *Módulo de Comunicação* e o *Gerenciador de Recursos* foram concebidos para gerenciar os aspectos associadas à comunicação entre os *Gateways* e o Servidor de Borda. O Módulo de Comunicação administra as comunicações através de API REST enquanto o Gerenciador de Recursos provê um mecanismo de descoberta que gerencia a entrada e saída de dispositivos na rede, ocorrências típicas da IoT.

O *Módulo de Coleta* tem a função de direcionar as solicitações de coleta aos respectivos *gateways* sob demanda do Motor de Regras, do Servidor de Contexto, ou das aplicações. O módulo *Supervisor* aglutina os comandos de atuação, recebendo os parâmetros de controle e resolvendo eventuais conflitos entre as requisições de oriundas de diferentes fontes. O *Módulo de Atuação* tem uma função semelhante a do Módulo de Coleta, recebendo os comandos de atuação e seus parâmetros operacionais (duração, energia de ativação, etc.) e os encaminhando aos *gateways* correspondentes para processamento adicional.

2.2. Suporte ao tratamento de eventos proposto

Na Internet das Coisas, eventos do ambiente ocorrem quando existe uma alteração importante em algum contexto de interesse, por exemplo, uma temperatura atingindo certo valor ou a identificação da entrada de um usuário em uma sala, entre outros. Esses eventos devem ser interceptados pelo sistema de gerência e notificações devem ser enviadas às aplicações para que as mesmas possam dar o tratamento adequado [Perera et al. 2014].

Ambientes da IoT geram potencialmente uma grande quantidade de eventos que devem ser gerenciados pela arquitetura de suporte. O gerenciamento desses eventos por parte da arquitetura possibilita que os mesmos possam ser tratados no momento em que eles acontecem, permitindo uma reação rápida quando necessário [Razzaque et al. 2016].

Eventos são frequentemente identificados como primitivos (discretos) ou complexos (compostos). Um evento primitivo refere-se a uma ocorrência instantânea, atômica de um acontecimento de interesse em um determinado instante de tempo. Enquanto que um evento complexo (também chamado de evento composto) consiste na combinação de eventos primitivos ocorridos em um determinado intervalo de tempo [Terfloth 2009]. O COIoT provê suporte tanto a eventos primitivos quanto a eventos complexos, os quais podem ser utilizados para disparar regras ECA.

O modelo de tratamento de eventos proposto para o COIoT considera um conjunto de eventos primitivos gerados a partir de: (i) mudanças de estado dos contextos de interesse coletados através de sensores; (ii) ativação/desativação de atuadores; e (iii) alterações na infraestrutura do ambiente computacional. Estes eventos são apresentados na Tabela 1. O suporte a eventos complexos é provido pelo COIoT a partir da composição de eventos através lógicas condicionais tratadas pelas regras ECA e processadas pelos Processadores de Contexto.

3. COIoT: prototipação e testes

Esta seção resume os principais aspectos de prototipação e testes realizados através do projeto AMPLUS (*Automatic Monitoring and Programmable Logging Ubiquitous Sys-*

Tabela 1. Eventos primitivos do COIoT

Evento	Nível de Geração	Descrição
Publication	Gateway/Servidor de Borda	Ocorre quando uma informação contextual é enviada ao Servidor de Borda ou Servidor de Contexto
Actuation	Gateway	Ocorre quando um atuador é disparado
NewDevice	Gateway	Ocorre quando um novo sensor ou atuador é identificado
DeviceDisconnect	Gateway	Ocorre quando um sensor ou atuador é desconectado
NewGateway	Servidor de Borda	Ocorre quando um novo Gateway é inserido
GatewayDisconnect	Servidor de Borda	Ocorre quando um Gateway é desconectado

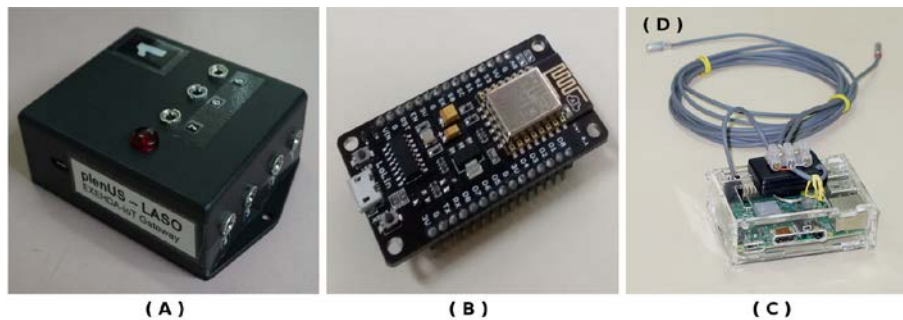


Figura 2. (A) Gateway Nativo; (B) NodeMCU; (C) Raspberry PI; (D) Sensor de Temperatura DS18B20

tem) usados para avaliar as funcionalidades do COIoT. O estudo de caso inclui tarefas relacionadas com o sensoriamento, coleta, processamento e notificação de contexto. Neste estudo de caso foi desenvolvida uma ferramenta para avaliar as principais funcionalidades da arquitetura proposta.

O projeto AMPLUS foi concebido para fornecer serviços móveis e cientes do contexto que permitem o armazenamento dos estados contextuais que caracterizam os equipamentos do Laboratório Didático de Análise de Sementes (LDAS - <http://amplus.ufpel.edu.br/ldas>), a geração de notificações e atuações quando necessário.

3.1. Infraestrutura de hardware

Para avaliar as funcionalidades do COIoT optou-se por utilizar no LDAS um conjunto de dispositivos que consistem de um Gateway nativo, 15 sensores e um atuador. Os sensores selecionados para este estudo de caso são baseados na tecnologia de 1-Wire (<http://www.maximintegrated.com>). Esta tecnologia é caracterizada como uma rede de transmissão de dados com base em dispositivos eletrônicos endereçáveis, e se destaca pela sua versatilidade e a facilidade de implementação. Os sensores de temperatura utilizados podem ser vistos na Figura 2(D). Este sensor é envolto em um invólucro de alumínio para dar maior resistência mecânica e isolar da umidade. O Gateway Nativo (Figura 2 (A)) foi desenvolvido utilizando NodeMCU o qual permite a conexão de até sete dispositivos 1-wire. NodeMCU (<http://nodemcu.com/>) é uma plataforma de código aberto para desenvolvimento de dispositivos IoT (vide Figura 2 (B)). Para explorar a característica reativa da arquitetura, também foi utilizado um atuador (alerta luminoso) com base na tecnologia de 1-Wire. Este atuador é acionado quando é necessária a atenção dos trabalhadores do laboratório com alguns equipamentos.

3.2. Infraestrutura de software: principais características

A maior parte do protótipo do COIoT foi escrita em Python no sistema operacional Raspbian. O hardware utilizado no Servidor de Borda é o Raspberry Pi (<http://www.raspberrypi.org>) (vide Figura 2 (C)). O Servidor de Contexto foi instalado em um hardware com processador Intel E3400-2.6GHz dual-core com 4GB de memória, com o Linux Ubuntu Server como sistema operacional. A tecnologia usada na implementação do Motor de Regras e do Processador de Contexto é o Drools (<http://www.drools.org/>). A leitura dos sensores é feita através de drivers específicos que fazem um tratamento individualizado dos dispositivos de acordo com as características tecnológicas de cada um. O Módulo de Interoperação foi implementado através do Sails.js (<http://sailsjs.org/>), um framework MVC (MVC) para Node.js (<https://nodejs.org>). A API REST desenvolvida fornece recursos para lidar com os sensores e atuadores, bem como para realizar a publicação dos dados coletados. Os dados enviados através das operações REST são estruturados em JSON.

3.3. Infraestrutura de software: soluções desenvolvidas para o LDAS

O suporte do COIoT à operação do cenário de avaliação é provido através de *triggers* de leitura dos sensores e de um conjunto de regras de processamento de contexto. Os *triggers* são utilizados para gerenciar as leituras dos sensores de temperatura das Incubadoras BODs (*Biochemical oxygen demand*) em duas situações: (i) em intervalos de tempo regulares; e (ii) quando o valor está fora da faixa especificada. As Incubadoras BODs são utilizadas no LDAS para realizar testes de germinação de sementes, os quais exigem precisão quanto aos limites específicos para variação da temperatura.

As regras de processamento de contexto utilizadas foram organizadas entre os Servidores de Borda e Servidor de Contexto de modo a atender o cenário proposto. Os critérios utilizados na distribuição das regras foram: (i) minimização no fluxo de dados; e (ii) continuidade do monitoramento mesmo em momentos de perda de comunicação entre os servidores. As regras utilizadas são apresentadas nas Tabelas 2 e 3.

A ferramenta desenvolvida possibilita a seleção do contexto de interesse a ser exibido, que pode ser apresentado na forma de um relatório textual ou através de um modo gráfico. Através da ferramenta o pesquisador do LDAS pode ter acesso à visualização das variações dos valores de temperatura e umidade ocorridas nas Incubadoras BODs durante os períodos de análise, os quais influenciam diretamente nos resultados dos processos de germinação das sementes.

O relatório gráfico desenvolvido permite visualizar simultaneamente as curvas de variação dos valores de vários sensores utilizados no LDAS (vide Figura 3). A seleção dos sensores a serem visualizados é feita a partir de um menu com suporte a múltipla seleção. Também é disponibilizado um recurso de inspeção que permite a comparação dos valores em um determinado instante do tempo. A janela de tempo dos dados que estão sendo visualizados pode ser definida pelo usuário através da mesma interface gráfica que exibe os valores sensorizados.

De forma a promover a proatividade do Projeto AMPLUS com a comunidade de usuários, foram desenvolvidas interfaces para serviços de comunicação: e-mail e SMS para a rede celular. O Servidor de Contexto produz essas mensagens a partir do processamento de regras de contexto de forma autônoma.

Tabela 2. Regras do Servidor de Borda

Nome da Regra	Evento	Condição	Ação
Lê temp. BOD	Publication	Se valor fora do especificado	Aciona alerta luminoso
Publica temp. BOD	Publication	-	Publica temperatura no Servidor de Contexto

Tabela 3. Regras do Servidor de Contexto

Nome da Regra	Evento	Condição	Ação
Lê temp. BOD	Publication	Se valor fora do especificado	Notifica o usuário (e-mail/SMS)
Persiste temp. BOD	Publication	-	Persiste a temperatura no Repositório de Contexto

**Figura 3. Relatório Gráfico**

A rotina dos laboratoristas implica na mobilidade por diferentes ambientes físicos do LDAS. Para resolver esta situação, uma interface foi desenvolvida para alertas visuais, que são ativados sempre que um dispositivo está em um estado que requer atenção. Considerando-se esses alertas, os detalhes podem ser inferidos através da interface da ferramenta de visualização desenvolvida para o Projeto AMPLUS.

A avaliação da arquitetura do COIOT ocorreu através da avaliação de aceitação da ferramenta desenvolvida. O estudo envolveu 10 voluntários, entre professores, alunos e técnicos, com atividades relacionadas ao LDAS. Cada participante utilizou um desktop para acessar a ferramenta. Após a realização de um treinamento básico, os participantes utilizaram a ferramenta de visualização e responderam um questionário de avaliação, considerando a experiência de uso.

O questionário foi construído com base no Modelo de Aceitação de Tecnologia

(TAM), usando uma escala de Likert [Yoon and Kim 2007], que varia entre 1 e 5. Para a aceitação da ferramenta o modelo TAM considera: (i) Facilidade de uso: grau em que o usuário avalia que a ferramenta pode reduzir seu esforço; e (ii) Percepção de utilidade: grau em que o usuário avalia que a ferramenta pode melhorar a sua experiência. A média dos resultados obtidos para facilidade de uso foi de 4,7 enquanto que para a percepção de utilidade foi obtida uma média de 4,4. A análise desses resultados mostrou que a aprovação foi elevada tanto para a facilidade de uso bem como para a percepção de utilidade.

4. Trabalhos relacionados

O estudo da literatura permitiu identificar alguns trabalhos relacionados, entre as quais foram selecionados os seguintes: EcoDiF [Pires et al. 2014], Xively [LogMeIn 2015], Carriots [Carriots 2015] e LinkSmart [Kostelník et al. 2011]. Os aspectos considerados na seleção destes trabalhos relacionados foram: (i) o suporte à heterogeneidade; (ii) suporte ao gerenciamento de eventos; (iii) ciência do contexto; e (iv) interoperabilidade.

Entre os trabalhos relacionados, a ciência de contexto é suportada apenas pelo LinkSmart, mas o suporte oferecido é limitado. Por sua vez, o COIoT oferece um mecanismo para a coleta e processamento distribuído de contexto através de regras, bem como para a atuação sobre o ambiente.

Todos os trabalhos relacionados apresentam estratégias de *trigger* para gerenciar o fluxo de dados transmitidos entre os diferentes dispositivos envolvidos. Um menor fluxo de dados traz benefícios, principalmente, no que tange à escalabilidade e ao consumo de energia. No entanto, Xively e Carriots não oferecem *triggers* associados à coleta. No COIoT, a abordagem de *triggers* foi concebida para permitir a personalização da coleta dos dados através de eventos considerando as características de variabilidade física de cada grandeza monitorada, o que proporciona minimizar o fluxo de dados entre os gateways e o Servidor de Borda.

A manipulação de eventos é suportado por todos os trabalhos selecionados, mas apenas o Carriots utiliza regras neste tratamento. Além disso, o gerenciamento distribuído das regras entre os Servidores de Contexto e Borda é um diferencial em relação aos outros projetos. Esta funcionalidade de processamento de contexto nos trabalhos relacionados, é normalmente restrita a um único dispositivo.

5. Conclusão

Este artigo resume os esforços de pesquisa associados com a concepção do COIoT. COIoT é uma arquitetura para a Internet das Coisas, integrado ao Middleware EXEHDA, que gerencia a coleta e pré-processamento das informações contextuais, oferecendo suporte à atuação no ambiente.

A principal contribuição deste trabalho é concepção de uma arquitetura para IoT direcionada à ciência de contexto. A arquitetura proposta é orientada a eventos, baseada em regras e gerencia a coleta e processamento das informações contextuais, bem como as atuações no ambiente físico de forma distribuída. A estratégia adotada para a COIoT ampliou o escopo de utilização do *Middleware* EXEHDA, permitindo seu uso em diferentes cenários.

Na continuidade da pesquisa os seguintes aspectos devem ser considerados em trabalhos futuros: (i) expandir o uso do COIoT no LDAS, possibilitando o monitoramento de outros equipamentos de laboratório e, conseqüentemente, incorporando outros tipos de sensores e atuadores; e (ii) continuar os procedimentos de integração de COIoT com os diferentes serviços e recursos do *Middleware* EXEHDA.

Referências

- Carriots (2015). Carriots. <https://www.carriots.com/>. Acessado em Maio de 2015.
- Fielding, R. T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. PhD thesis, UNIVERSITY OF CALIFORNIA, IRVINE.
- Knappmeyer, M., Kiani, S. L., Reetz, E. S., Baker, N., and Tonjes, R. (2013). Survey of Context Provisioning Middleware. *IEEE Communications Surveys & Tutorials*, 15(3):1492–1519.
- Kostelník, P., Sarnovský, M., and Furdík, K. (2011). The semantic middleware for networked embedded systems applied in the internet of things and services domain. *Scalable Computing*, 12(3):307–315.
- LogMeIn, I. (2015). Xively. <https://xively.com/>. Acessado em Fevereiro de 2015.
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014a). A Middleware Architecture for Dynamic Adaptation in Ubiquitous Computing. *Journal of Universal Computer Science*, 20(9):1327–1351.
- Lopes, J. L., de Souza, R. S., Pernas, A. M., Yamim, A., and Geyer, C. (2014b). A Distributed Architecture for Supporting Context-Aware Applications in UbiComp. In *IEEE International Conference on Advanced Information Networking and Applications (AINA)*, Victória, Canada.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *IEEE Communications Surveys and Tutorials*, 16(1):414–454.
- Pires, P. F., Cavalcante, E., Barros, T., Delicato, F. C., Batista, T., and Costa, B. (2014). A Platform for Integrating Physical Devices in the Internet of Things. *2014 12th IEEE International Conference on Embedded and Ubiquitous Computing*, pages 234–241.
- Razzaque, M. A., Milojevic-Jevric, M., Palade, A., and Clarke, S. (2016). Middleware for Internet of Things: A Survey. *IEEE INTERNET OF THINGS JOURNAL*, 3(1):70–95.
- Souza, R., Lopes, J., Geyer, C., Garcia, C., Davet, P., and Yamin, A. (2015). Context awareness in UbiComp: An IoT oriented distributed architecture. In *2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, pages 535–538.
- Terfloth, K. (2009). *A Rule-Based Programming Model for Wireless Sensor Networks*. PhD thesis, Freie Universität Berlin.
- Yoon, C. and Kim, S. (2007). Convenience and TAM in a ubiquitous computing environment: The case of wireless LAN. *Electronic Commerce Research and Applications*, 6(1):102–112.

Uma Proposta de Instrumentação Remota Explorando a Ubiquidade da Infraestrutura da IoT

Lucas M. S. Xavier^{1,2}, Huberto Kaiser Filho¹, Patrícia T. Davet¹, Adenauer C. Yamin^{1,2}

¹Centro de Desenvolvimento Tecnológico – Universidade Federal de Pelotas (UFPel)
Rua Gomes Carneiro, 1 – 96.010-610 – Pelotas – RS – Brasil

²Centro Politécnico – Universidade Católica de Pelotas (UCPel)
Rua Gonçalves Chaves, 373 – 96.015-560 – Pelotas – RS – Brasil

{lmdsxavier,hkaiser,ptdavet,adenauer}@inf.ufpel.edu.br

Abstract. *The demands of modern instrumentation systems can be answered by the use of the same technological advances behind the consolidation of IoT and its infrastructures. In both scenarios, applications must deal with challenges such as remote access, information storage and alert generation. In this paper, we present a remote measurement instrument which strives to achieve ubiquitous features through close integration with an IoT platform, exploring the convergence between instrumentation and IoT. Specifically, the instrument, which is called ubiMeter, uses the data acquisition and processing mechanisms of the EXEHDA middleware. Its features were evaluated in two different case studies and the results obtained were judged as promising.*

Resumo. *As tecnologias que vêm permitindo a consolidação das infraestruturas da IoT são as mesmas que respondem às demandas dos sistemas de instrumentação modernos, os quais devem contemplar, entre outros desafios, o armazenamento e o acesso remoto das informações medidas e a geração de alertas. Este trabalho busca explorar essa convergência, apresentando a concepção de um instrumento de medição que, por meio da integração com uma plataforma IoT, assume um caráter ubíquo. Especificamente, foram usados os mecanismos para coleta e tratamento de informações sensorizadas do middleware EXEHDA. O instrumento, denominado ubiMeter, foi submetido a dois cenários de teste. Os resultados obtidos mostraram-se promissores.*

1. Introdução

A Internet das Coisas (*Internet of Things* ou IoT), enquanto paradigma para a evolução da Internet nos próximos anos, prevê que cada vez mais os objetos cotidianos deverão ganhar capacidades de processamento, comunicação e sensoriamento. Essa visão abre espaço para que esses objetos sejam incorporados às redes de computadores e passem a interagir de forma autônoma, promovendo uma ampla integração do mundo físico ao digital [Kortuem et al. 2010].

Dentro dessa perspectiva, as aplicações se tornam capazes de empregar a infraestrutura da IoT para realizar a computação de forma amplamente distribuída e, tanto quanto possível, provendo interfaces homem-máquina proativas, adquirindo um caráter efetivamente ubíquo [Pires et al. 2015].

Em ambientes IoT, existe a necessidade de coletar dados continuamente gerados por um conjunto heterogêneo de sensores e objetos, bem como de disponibilizar essas informações de forma remota para diversas aplicações e usuários [Perera et al. 2013]. Fica caracterizada, assim, uma forte convergência entre os requisitos abarcados pelas plataformas IoT e as demandas dos sistemas de instrumentação modernos, que cada vez mais compartilham dos mesmos desafios tecnológicos e de pesquisa. Essas necessidades se verificam tanto no escopo industrial [Kagermann et al. 2013] quanto no âmbito de laboratórios para ensino e pesquisa [Salzmann et al. 2015].

Assim, a proposta central desse trabalho é o emprego de um *middleware* IoT para o desenvolvimento e prototipação de um instrumento de medição. A abordagem adotada considera que os instrumentos eletrônicos modernos possuem capacidades computacionais crescentes, dedicadas a funcionalidades como autocalibração e visualização das medições. Essas capacidades podem ser estendidas para disponibilizar uma interface de serviços por meio de protocolos típicos de ambientes IoT.

O *middleware* IoT utilizado neste estudo foi o EXEHDA. Este *middleware* provê uma arquitetura de software baseada em serviços que visa criar e gerenciar um ambiente ubíquo, promovendo a execução de aplicações IoT sob este ambiente. Seu foco é permitir que as aplicações adquiram um caráter ciente de contexto, tornando-se reativas e adaptativas em resposta a variações nos seus contextos de interesse. O instrumento desenvolvido foi integrado ao Subsistema de Reconhecimento e Adaptação ao Contexto do EXEHDA, aproveitando os recursos do mesmo para coleta, armazenamento de informações sensorizadas e geração de alertas.

O restante do artigo está estruturado como segue. A Seção 2 apresenta uma visão geral da arquitetura do EXEHDA, com ênfase no Subsistema de Reconhecimento e Adaptação ao Contexto. A concepção do instrumento IoT é discutida na Seção 3. A Seção 4 aborda o desenvolvimento de um protótipo funcional seguindo a arquitetura proposta, bem como os estudos utilizados para avaliação do instrumento. Na Seção 5, são resumidos alguns dos principais trabalhos relacionados. Finalmente, a Seção 6 traz uma série de considerações finais a respeito do trabalho.

2. Middleware EXEHDA

O EXEHDA é um *middleware* voltado para aplicações ubíquas amplamente distribuídas, móveis e cientes de contexto. Para isso, adota uma organização baseada em serviços dinamicamente carregados por um núcleo mínimo conforme a demanda. As principais funcionalidades providas pelo *middleware* estão divididas nos subsistemas de acesso ubíquo, comunicação, execução distribuída e adaptação e reconhecimento do contexto [Lopes et al. 2014b].

O ambiente ubíquo provido pelo EXEHDA é formado por células que agrupam os recursos disponíveis na infraestrutura física. Esses recursos são mapeados para três abstrações básicas [Souza et al. 2015]: (i) EXEHDA nodos, que são dispositivos computacionais responsáveis pela execução das aplicações; (ii) EXEHDA bases, que são o elemento central de uma célula, fornecendo serviços básicos e de referência; e (iii) EXEHDA bordas, que são elementos responsáveis por interoperar os serviços do *middleware* com os dispositivos IoT. A organização celular dos recursos, na visão do *middleware*, é mostrada na Figura 1.

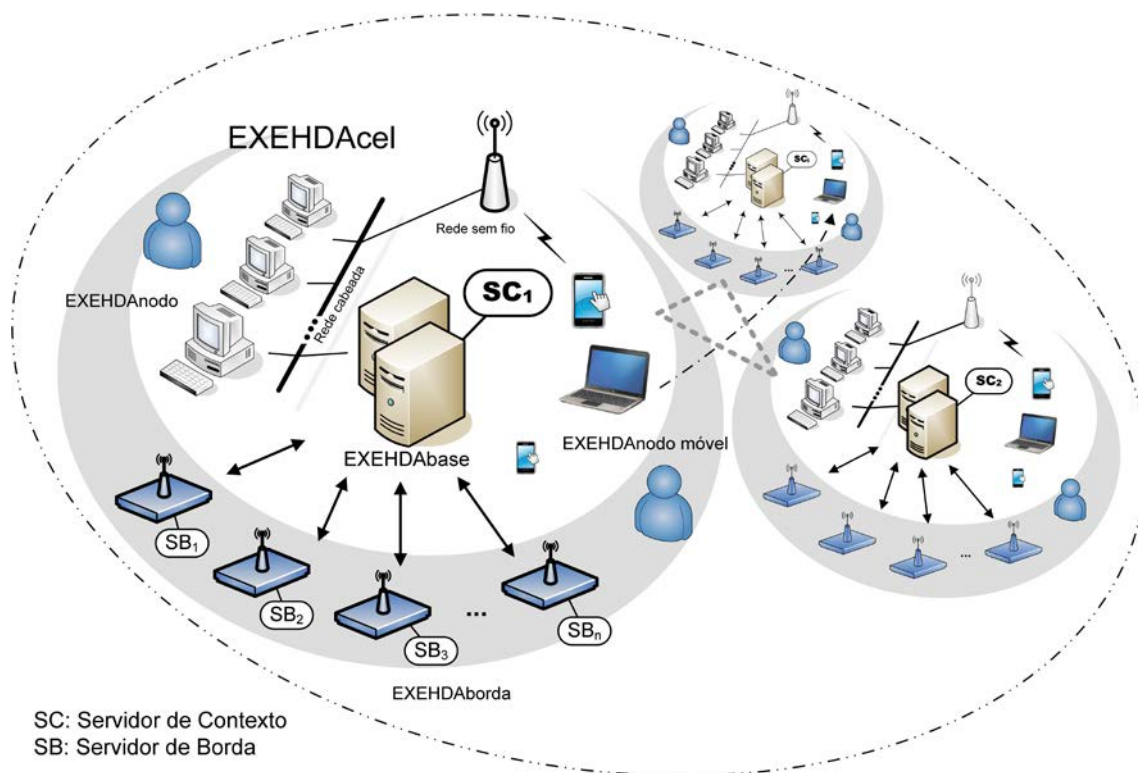


Figura 1. Organização celular do ambiente ubíquo provido pelo EXEHDA.
Fonte: Souza et al. (2015)

Dentre os diversos serviços do EXEHDA, o Subsistema de Reconhecimento e Adaptação ao Contexto (SRAC) é o módulo responsável pela aquisição, pelo armazenamento e pelo processamento de informações contextuais, além do suporte a procedimentos de atuação. Suas responsabilidades estão divididas em dois tipos de servidores [Lopes et al. 2014a]. O Servidor de Borda, executado em um nodo tipo EXEHDA-borda, se destina a gerenciar a interação com o meio físico através de sensores e atuadores, bem como executar o primeiro estágio do processamento do contexto. O Servidor de Contexto, por sua vez, é instanciado em uma EXEHDAbase e atua no armazenamento e no segundo estágio de processamento das informações contextuais, integrando dados históricos e dados provenientes de diferentes Servidores de Borda.

A arquitetura do SRAC prevê, ainda, o uso de *gateways* como dispositivos dedicados à interoperação entre os Servidores de Borda e nodos sensores e atuadores heterogêneos, potencialmente não programáveis. Os *gateways* agem no sentido de interligar esses dispositivos IoT ao SRAC, utilizando, para isso, protocolos como CoAP, MQTT e HTTP sobre a rede TCP/IP. Esses equipamentos têm hardware e capacidades dedicadas para tratar tecnologias específicas, fazendo conversão de protocolos e gerenciamento de dispositivos que não possam ser tratados diretamente por Servidores de Borda.

3. ubiMeter: Visão Geral e Arquitetura

Em ambientes IoT típicos, há um elevado volume de comunicação entre os diversos objetos inteligentes, que se apresentam em grande número e sob formas variadas, diferindo significativamente em suas capacidades computacionais. Nesse cenário, surgem desa-

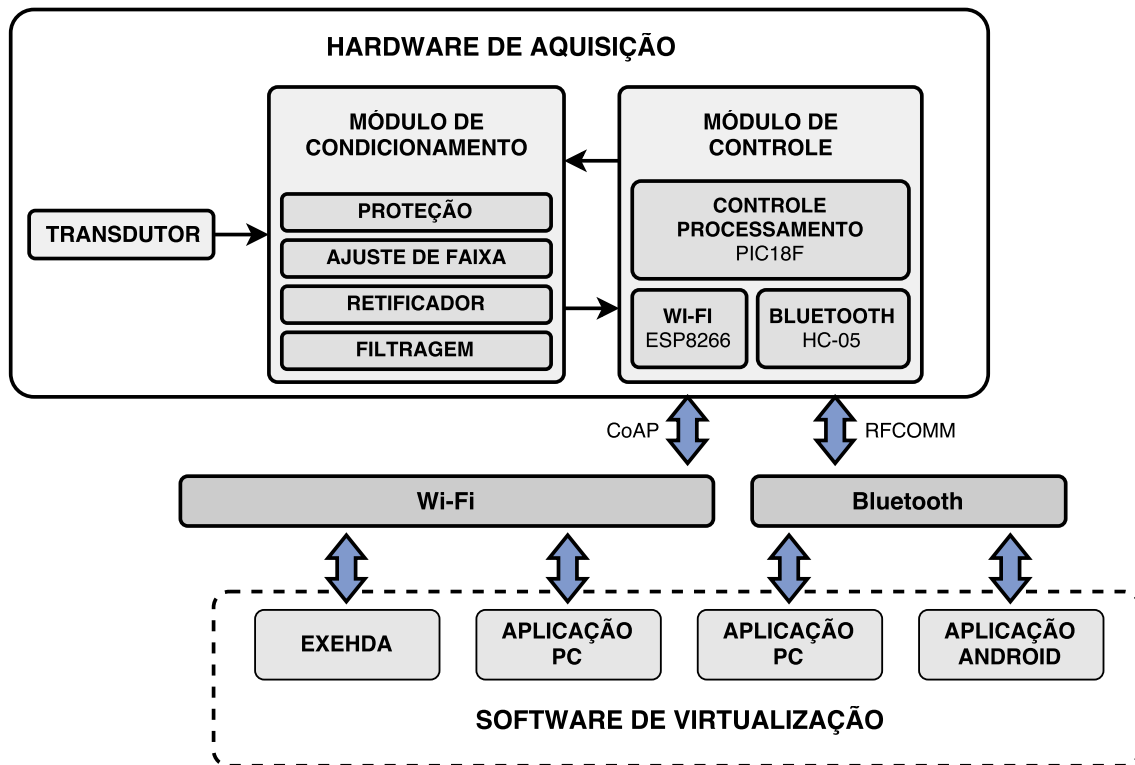


Figura 2. Visão geral da arquitetura do ubiMeter

fios relacionados a aspectos como interoperabilidade, escalabilidade e complexidade de configuração. Assim, o desenvolvimento de novos dispositivos IoT deve priorizar, como premissas (entre outros itens): (i) a operação autônoma, provendo mecanismos para coleta e processamento automatizados dos dados; (ii) o uso de protocolos bem difundidos e padronizados para comunicação com outros objetos inteligentes; e (iii) a simplicidade de configuração, considerando a dinamicidade das infraestruturas de rede.

A solução caracterizada neste artigo propõe a implementação de um instrumento eletrônico que, além de disponibilizar as medições por uma interface de rede, seja capaz de atender às premissas enumeradas. A arquitetura proposta para o ubiMeter pode ser dividida, em um primeiro nível de abstração, em dois grandes blocos: um hardware de aquisição, que provê uma interface de acesso às medições por meio de canais de comunicação sem fio; e uma camada de software, executada em outro dispositivo, que disponibiliza as informações providas para usuários e aplicações. Uma visão geral para a arquitetura do instrumento é mostrada na Figura 2.

O software, nessa perspectiva, pode ser representado tanto por um *middleware* IoT como por uma aplicação específica que seja capaz de interoperar com o hardware de aquisição. A comunicação se dá por meio de uma interface de serviços, disponibilizada pelo hardware de aquisição e implementada sobre os protocolos de comunicação previstos para o instrumento. Os serviços podem ser agrupados em duas classes: funcionalidades de configuração do instrumento e procedimentos para aquisição das grandezas medidas.

O hardware de aquisição, por sua vez, pode ser subdividido em dois blocos. O Módulo de Controle compreende um circuito microprocessado que digitaliza as grande-

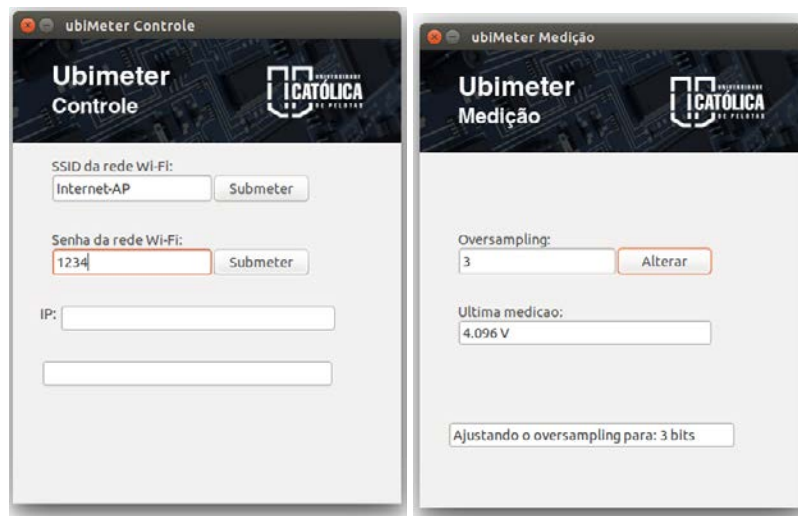


Figura 3. Aplicações *desktop* para visualização dos dados e gerenciamento das configurações

zas medidas e implementa a interface com o software de virtualização. O Módulo de Condicionamento, representado por um circuito analógico, realiza a adequação dos sinais de entrada às características elétricas exigidas pelo microcontrolador do Módulo de Controle.

Quando utilizado de forma integrada ao EXEHDA, o instrumento é modelado como um *gateway*. Nesse caso, o ubiMeter disponibiliza uma interface para configuração e leitura das medições empregando o protocolo CoAP. A comunicação entre o instrumento e o *middleware* se dá de forma assíncrona, com requisições realizadas diretamente pelas aplicações ou disparadas pela avaliação de uma ou mais regras pelo processador de contexto do Servidor de Borda.

4. ubiMeter: Funcionalidades e Avaliação

O protótipo desenvolvido para este trabalho foi um voltímetro para medição de tensões contínuas, baseado na premissa de permitir a operação autônoma ao longo de uma faixa de medição ampla. Para isso, foi utilizado um esquema de seleção automática de faixa de operação, no qual o Módulo de Controle regula a atenuação imposta ao sinal de entrada por meio de um conjunto de relés presente no Módulo de Condicionamento. A opção por um medidor de tensão objetiva simplificar a etapa de validação do ubiMeter, sem prejuízo da generalidade da arquitetura.

O Módulo de Controle teve suas funcionalidades implementadas utilizando um microcontrolador PIC18F4550, um módulo Wi-Fi ESP-01 e um módulo Bluetooth HC-05. O PIC, que possui conversor A/D integrado de 10 bits, assume as funcionalidades de aquisição e controle, que são realizadas em um esquema de multitarefa cooperativo. Os outros dois módulos, por sua vez, realizam a comunicação com o software de virtualização; o módulo Wi-Fi disponibiliza a interface CoAP para coleta de dados; e o módulo Bluetooth permite, além da leitura das medições, a configuração da conectividade Wi-Fi do instrumento. A comunicação Bluetooth é baseada no protocolo RFCOMM, com um protocolo de camada de aplicação próprio.

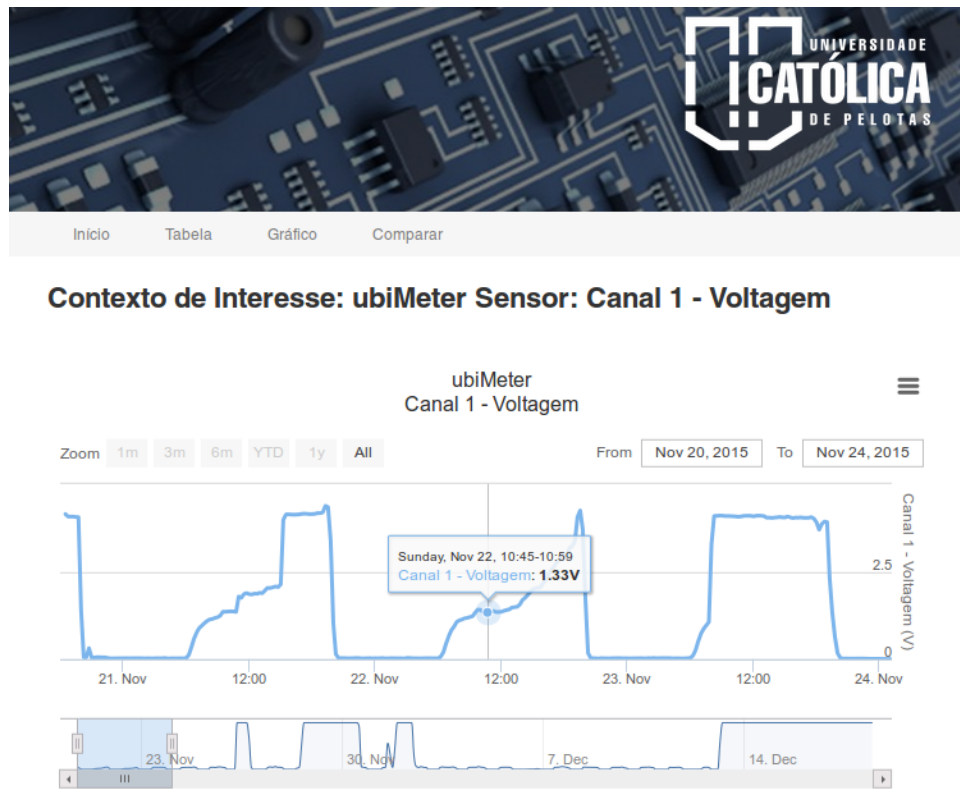


Figura 4. Aplicação web do EXEHDA para visualização das medições obtidas com o ubiMeter

O Módulo de Condicionamento foi implementado como um circuito analógico linear que opera com entrada diferencial de até 350 V e produz como saída um sinal proporcional na faixa de 0V a 5V. Para isso, o circuito possui quatro faixas de operação (com fundo de escala de 2 V, 20 V, 200 V e 400 V), que são selecionadas automaticamente pelo Módulo de Controle por meio de um sistema de *autorange*. O nível de atenuação é controlado por um conjunto de quatro relés do tipo *reed* que alteram a relação entre os resistores de um divisor resistivo, bem como o ganho de um estágio amplificador.

Para demonstrar a interoperação entre o software de virtualização e o hardware de aquisição, foram desenvolvidas duas aplicações gráficas, mostradas na Figura 3, para comunicação entre o ubiMeter e um PC. Uma das aplicações utiliza a interface Bluetooth para oferecer funcionalidades básicas de configuração da conectividade do instrumento. A outra, baseada na interface CoAP, permite uma funcionalidade mínima de visualização das medições e configuração das opções de aquisição e pré-processamento dos dados. Ambas as aplicações foram desenvolvidas em Python e testadas em sistemas baseados em Linux.

No caso de instalação do ubiMeter em um ambiente gerenciado pelo EXEHDA, foi utilizada, para visualização dos dados, uma aplicação web baseada na tecnologia de software trabalhada no grupo de pesquisa [Davet et al. 2015, Kaiser Filho et al. 2016]. A aplicação é mostrada na Figura 4. Entre as principais funcionalidades utilizadas no trabalho, pode-se enumerar a geração de gráficos e tabelas baseados nas medidas realizadas. Outras funcionalidades da aplicação incluem o envio de alertas por SMS ou e-mail,

Tabela 1. Resultados dos testes para a faixa de 20 V

Referência	Média (V)	ubiMeter		Desvio Padrão
		Erro médio	Erro máximo	
3,2 V	3,20	0,01%	1,86%	1,15%
6,4 V	6,42	0,27%	0,95%	0,56%
9,6 V	9,60	0,04%	0,61%	0,40%
12,8 V	12,82	0,14%	0,56%	0,31%
16 V	16,01	0,06%	0,50%	0,23%
19,2 V	19,23	0,14%	0,60%	0,23%

Tabela 2. Resultados dos testes para a faixa de 2 V

Referência	Média (V)	ubiMeter		Desvio Padrão
		Erro médio	Erro máximo	
320 mV	0,322	0,49%	16,74%	12,18%
640 mV	0,638	0,39%	8,73%	5,75%
960 mV	0,962	0,19%	5,61%	4,07%
1,28 V	1,278	0,19%	4,21%	2,85%
1,6 V	1,590	0,63%	3,31%	2,17%
1,92 V	1,926	0,30%	3,13%	1,91%

no caso de medições fora dos parâmetros esperados, e a comparação entre valores de múltiplos sensores.

Para avaliação do trabalho desenvolvido, foram concebidos dois cenários de teste; o primeiro, focado na integração do ubiMeter com o *middleware* EXEHDA; e o segundo, voltado para a verificação de requisitos não funcionais ligados à eletrônica do hardware de aquisição.

No primeiro cenário, o instrumento desenvolvido foi instalado em um ambiente ubíquo gerenciado pelo EXEHDA. Ao longo de três dias, foi monitorada a tensão sobre um fotoresistor localizado em um ponto fixo de uma sala. As medições foram periodicamente coletadas por um Servidor de Borda instanciado em uma Raspberry Pi e armazenadas em um Servidor de Contexto instanciado em uma máquina remota desktop. A Figura 4 mostra dados extraídos do período de testes.

No segundo cenário, o protótipo foi submetido a um conjunto de testes para verificar os erros introduzidos pelo hardware de aquisição. Um conjunto de doze tensões de referência, abrangendo duas das quatro faixas de operação do Módulo de Condicionamento, foi gerado por uma fonte CC ajustável. As tensões foram aferidas com auxílio de um multímetro ET-1002 da Minipa, que possui faixas de operação divididas de forma similar ao protótipo e precisão de $\pm 0,8\%$ mais $0,25\%$ do fundo de escala.

Para cada tensão de referência, foram realizadas trinta medições com o ubiMeter. O Módulo de Aquisição e Comunicação foi configurado para realizar uma sobreamostragem de 64 amostras, estendendo a resolução para 13 bits. Nas tabelas 1 e 2 são mostradas a média das trinta medições, bem como o erro máximo, o erro médio e o desvio padrão.

Com o protótipo operando na faixa de 20 V, nenhuma das medições apresentou erro superior a $0,8\%$ da medição mais $0,25\%$ do fundo de escala (a precisão do próprio

multímetro usado como referência). Assim, a precisão obtida foi similar à de um instrumento típico de mercado.

Na faixa de 2 V, a inclusão de um estágio amplificador no caminho do sinal acenou a flutuação dos valores mensurados, amplificando erros introduzidos nos estágios anteriores do Módulo de Condicionamento. Em função disso, as medições apresentaram grande variabilidade, com desvio padrão superior a 2% e erro absoluto de até 60 mV.

5. Trabalhos Relacionados

Existe uma ampla literatura relacionada ao desenvolvimento de plataformas voltadas para a aquisição de dados produzidos por redes de sensores sem fio, tanto no escopo da IoT como em ambientes ubíquos [Perera et al. 2013, Razzaque et al. 2016]. Contudo, a maioria dos trabalhos não enfatiza as necessidades específicas dos cenários de instrumentação e medição tradicionais, dificultando a comparação direta.

A exceção é o trabalho [Spanò et al. 2015], que propõe uma infraestrutura IoT para aplicações de medição e aquisição de dados distribuídas e executadas de forma concorrente. A plataforma visa, ainda, facilitar a integração de diferentes tipos de sensores e protocolos de comunicação. Assim como no presente trabalho, *gateways* são usados para prover interoperabilidade e realizar a interface com dispositivos de baixo poder computacional. Contudo, a arquitetura proposta não é organizada de forma hierárquica, interpondo um *IoT Server* como única abstração entre as redes de sensores e as aplicações.

Entre os outros trabalhos, que não possuem foco específico em instrumentação, é possível pontuar alguns exemplos de plataformas que poderiam ser usadas para aplicações de medição distribuída.

O projeto LinkSmart [Kostelnik et al. 2011] é um *middleware* para inteligência ambiental com capacidades de ciência de contexto. A plataforma é baseada em uma arquitetura distribuída, orientada a serviços e com componentes estruturados em três camadas: (i) camada de rede, responsável pela comunicação com os dispositivos; (ii) camada de serviço, responsável pelo gerenciamento de eventos, dispositivos e escalonamento de recursos; e (iii) camada semântica. Um de seus diferenciais é a possibilidade de contemplar uma descrição semântica dos dispositivos através do uso de ontologias, o que proporciona interoperabilidade em nível sintático e semântico.

O *middleware* Servilla [Fok et al. 2012] é voltado para redes de sensores heterogêneas. As aplicações são estruturadas na forma de tarefas independentes de plataforma e executadas em máquinas virtuais nos nodos disponíveis. Códigos dependentes de plataforma são abstraídos como serviços, que são dinamicamente vinculados às tarefas em tempo de execução. A plataforma usa uma linguagem própria para descrição de serviços e tarefas.

Xively [LogMeIn Inc. 2016] é uma solução comercial de PaaS que provê serviços de *middleware* para produtos e aplicações IoT. Além de funcionalidades de armazenamento e processamento de dados, a plataforma permite a descoberta de dispositivos por meio de um diretório de serviços, bem como a implantação de aplicações amplamente escaláveis. Embora vários formatos de dados sejam suportados, os dados de entrada não são homogeneizados, e o suporte a interoperabilidade é parcial [Razzaque et al. 2016].

6. Considerações Finais

De forma geral, as tecnologias estratégicas para os sistemas de instrumentação modernos são as mesmas que permitem a materialização do cenário da IoT. Este trabalho busca explorar essa convergência, apropriando-se de uma infraestrutura de IoT para conferir um caráter de ubiquidade a um instrumento de medição remoto. Levando isso em consideração, o desenvolvimento do ubiMeter foi realizado de forma integrada a um *middleware* e contemplando as principais premissas da instrumentação ubíqua.

As funcionalidades do ubiMeter, ao longo de sua avaliação, mostraram resultados promissores nesse sentido, em duas situações distintas: (i) quando do seu uso enquanto instrumento virtual, operado através de uma interface local via Bluetooth; e (ii) quando da sua integração com o EXEHDA, viabilizando o acesso às medições em uma perspectiva distribuída. Pode-se pontuar, por exemplo, a contribuição do *autorange* para a autonomia de operação do instrumento, já que a seleção automática de faixa de medição minimiza a necessidade de intervenção do usuário.

Com o prosseguimento dos esforços de pesquisa, pretende-se incorporar formatos de dados da web semântica aos serviços de medição, instrumentando com metadados as grandezas disponibilizadas. Ainda, entende-se como vantajoso o suporte a protocolos de descoberta de serviços (como o UPnP), o que representaria um avanço ainda maior no sentido de reduzir a complexidade de configuração do instrumento. Finalmente, pode ser interessante a realização de estudos de caso de escopo mais amplo, explorando aspectos como escalabilidade e precisão de forma mais aprofundada.

Agradecimentos

Os autores querem registrar seu agradecimento aos órgãos de fomento CAPES, CNPq e FAPERGS pelo apoio aos trabalhos de pesquisa relacionados ao *middleware* EXEHDA.

Referências

- Davet, P., Kaiser Filho, H., João, L., Xavier, L., Carvalho, T., Fleischmann, A., Yamin, A., Lopes, J., and Souza, R. (2015). Consciência de contexto na IoT: uma arquitetura distribuída e escalável. In *Proceedings of the V SBESC*. CPS.
- Fok, C.-L., Roman, G.-C., and Lu, C. (2012). Servilla: a flexible service provisioning middleware for heterogeneous sensor networks. *Science of Computer Programming*, 77(6):663–684.
- Kagermann, H., Helbig, J., Hellinger, A., and Wahlster, W. (2013). *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Final Report of the Industrie 4.0 Working Group*. Forschungsunion.
- Kaiser Filho, H., João, L., Carvalho, T. a. R., Davet, P., and Yamin, A. (2016). Uma contribuição ao gerenciamento de recursos de sensoriamento e atuação no middleware exehda direcionado à internet das coisas. In *Anais da Escola Regional de Alto Desempenho do Estado do Rio Grande do Sul*, São Leopoldo. SBC.
- Kortuem, G., Kawsar, F., Fitton, D., and Sundramoorthy, V. (2010). Smart objects as building blocks for the internet of things. *Internet Computing, IEEE*, 14(1):44–51.

- Kostelnik, P., Sarnovsk, M., and Furdik, K. (2011). The semantic middleware for networked embedded systems applied in the internet of things and services domain. *Scalable Computing: Practice and Experience*, 12(3):307–316.
- LogMeIn Inc. (2016). Xively.
- Lopes, J., Souza, R., Gadotti, G., Pernas, A., Yamin, A., and Geyer, C. (2014a). An architectural model for situation awareness in ubiquitous computing. *IEEE Latin America Transactions*, 12(6):1113–1119.
- Lopes, J., Souza, R., Pernas, A., Yamin, A., and Geyer, C. (2014b). A distributed architecture for supporting context-aware applications in ubicomp. In *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*, pages 584–590.
- Perera, C., Zaslavsky, A. B., Christen, P., and Georgakopoulos, D. (2013). Context aware computing for the internet of things: A survey. *Communications Surveys & Tutorials, IEEE*, abs/1305.0982.
- Pires, P. F., Delicato, F. C., Batista, T., Barros, T., Cavalcante, E., and Pitanga, M. (2015). Plataformas para a internet das coisas. In *Livro Texto de Minicursos – SBRC 2015*. SBC, Vitória.
- Razzaque, M. A., Milojevic-Jevric, M., Palade, A., and Clarke, S. (2016). Middleware for internet of things: A survey. *IEEE Internet of Things Journal*, 3(1):70–95.
- Salzmann, C., Govaerts, S., Halimi, W., and Gillet, D. (2015). The smart device specification for remote labs. In *REV 2015*, pages 199–208, Bangkok. IAOE.
- Souza, R. S. d., Lopes, J., Gadotti, G., Pernas, A. M., Yamin, A., and Geyer, C. (2015). Gerenciamento proativo de redes de sensores na ubicomp. *Revista Brasileira de Computação Aplicada*, 7(3):53–64.
- Spanò, E., Pascoli, S. D., and Iannaccone, G. (2015). Internet-of-things infrastructure as a platform for distributed measurement applications. In *Instrumentation and Measurement Technology Conference (I2MTC), 2015 IEEE International*, pages 1927–1932.

Uma Proposta de Rede Social Espontânea para Interação Ubíqua em Eventos

Marcelo da Silva dos Santos¹, Cristiano André da Costa¹, Jorge Luis Victoria Barbosa¹

¹Programa de Pós-Graduação em Computação Aplicada
Universidade do Vale do Rio dos Sinos (UNISINOS) – São Leopoldo, RS – Brasil

marcelo@novohamburgo.rs.gov.br, {cac,jbarbosa}@unisinors.br

Abstract. *The popularization of mobile devices and the increasing spread of social networks has fostered the development of applications in order to manage and encourage interactions through their devices. This article presents a model of spontaneous social network to foster the interaction of participants in events such as conferences and congresses. In addition to presenting the model, called Mingle, we show the prototype created for experimentation. The assessment followed two visions, the first focusing on access statistics to the environment and the second checking usability, effectiveness, efficiency and user satisfaction. Results showed good reception Mingle and potential for implementation on a larger scale.*

Resumo. *A popularização dos dispositivos móveis e a crescente difusão das redes sociais fomentou o desenvolvimento de aplicativos com o intuito de gerenciar e incentivar interações através de seus dispositivos. Este artigo apresenta um modelo de rede social espontânea destinada a fomentar a interação dos participantes de eventos, como conferências e congressos. Além de apresentar o modelo, denominado Mingle, mostraremos o protótipo criado para experimentação. A avaliação seguiu duas visões, a primeira focando estatísticas de acesso ao ambiente e a segunda verificando usabilidade, eficácia, eficiência e satisfação dos usuários. Resultados demonstraram boa recepção do Mingle e potencial para implantação em maior escala.*

1. Introdução

A disseminação dos dispositivos móveis, a evolução tecnológica e os avanços científicos têm aproximado cada vez mais a humanidade do conceito de computação ubíqua [Satyanarayanan 2001] [Beach et al. 2008]. Tais dispositivos fazem parte do cotidiano e acompanham constantemente as pessoas em suas vidas diárias, onde os mesmos são capazes de detecção contínua das atividades e informações espaço-temporais do seu usuário, o que auxilia a compreensão dos contextos onde atua. Portanto, a computação ubíqua [Weiser 1991] aliada à popularização dos dispositivos móveis amplia a possibilidade de exploração de formas de interação ubíqua, ou seja, o conceito de computação integrada ao mundo real, a qualquer momento e em qualquer lugar. Neste sentido, os aplicativos baseados em Serviços Baseados em Localização, que utilizam o contexto de localização do usuário para a recomendação de serviços relacionados a esse local específico [Vaughan-Nichols 2009], tem empregado esforços

para fornecer serviços que promovam formas de interação espontânea relacionadas à sua localização, tendo as aplicações no formato de redes sociais empregadas ganhado espaço.

A popularização do uso de redes sociais virtuais em camadas cada vez mais jovens da sociedade vem tornando-se parte do cotidiano [Guedes 2013]. Em termos gerais, uma rede social pode ser caracterizada como uma estrutura composta por pessoas ou organizações, essas ligadas por relacionamentos horizontais e não hierárquicos, onde partilham valores e objetivos comuns [Duarte e Frei 2008]. Uma rede social virtual utiliza sistemas automatizados e a Internet para proporcionar esta interação [Costa 2008]. Este trabalho tem como objetivo apresentar o Mingle, um modelo de rede social espontânea [Laforest et al. 2014] [Costa et al. 2014], onde tal rede será formada por usuários fisicamente presentes em um lugar específico, onde as pessoas interagem de forma espontânea sobre um assunto específico de interesse comum, tais como conferências, exposições, galerias, estádios ou restaurantes.

Nesse âmbito, foi desenvolvido um modelo com o intuito de gerenciar e incentivar interações entre usuários através de seus dispositivos móveis, ofertando serviços baseados em localização, com o estabelecimento de uma rede social espontânea. Este poderia ser aplicado a cenários diversos, em aplicações comerciais ou puramente sociais. A partir do modelo, foi criado um protótipo para um congresso que permitiu avaliar a proposta. O artigo discute a avaliação que considerou aspectos quantitativos e qualitativos do funcionamento do ambiente. A principal contribuição científica do trabalho é permitir a interação entre pessoas de forma ubíqua e espontânea relacionada com contextos atuais. Particularmente, no caso da aplicação o contexto é um congresso e a interação se dá via chat.

Este trabalho está distribuído da seguinte forma: o modelo proposto para a plataforma Mingle é apresentado na Seção 2. A Seção 3 comenta detalhes de implementação e os resultados obtidos e aspectos de interesse desta pesquisa. Na Seção 4 são apresentados mostra de um conjunto de trabalhos relacionados. Finalmente, algumas considerações são feitas na Seção 5.

2. Modelo Proposto

Neste trabalho será demonstrado o modelo utilizado na construção da plataforma Mingle [Costa et. al. 2014], um ambiente para fomentar interações sociais espontâneas em ambiente virtual, proporcionando serviços orientados às pessoas participantes da rede presentes dependente da sua localização específica. Neste trabalho a localização de onde o evento está ocorrendo é denominada célula. Os participantes de uma célula podem interagir uns com os outros acessando a plataforma.

A célula pode ser formada para qualquer contexto, como eventos esportivos, concertos, conferências, e assim por diante. A ideia é que pessoas que frequentam o mesmo evento estão envolvidas na mesma atividade. Desta forma, as interações seriam relevantes durante o tempo de duração das atividades e, depois disso, essas pessoas acabariam deixando de compartilhar o mesmo contexto e a rede se desfazendo. A primeira modelagem do ambiente foi sugerida para o uso em conferências e congressos, onde os serviços foram direcionados para este cenário específico.

A criação de uma nova célula deveria prover, no mínimo, suporte para três requisitos básicos:

- Prover um ambiente de comunicação entre os participantes e palestrantes em congressos e seminários, por meio de salas de comunicação que podem ser criadas pelos próprios participantes;
- Prover um canal para envio de perguntas ao palestrante durante a apresentação, contando com opção para atribuição de pontuação para as perguntas com o intuito de “*rankear*” as perguntas e organizar a ordem de interesse do grupo;
- Prover um canal para troca de links e conteúdos relacionados à discussão.

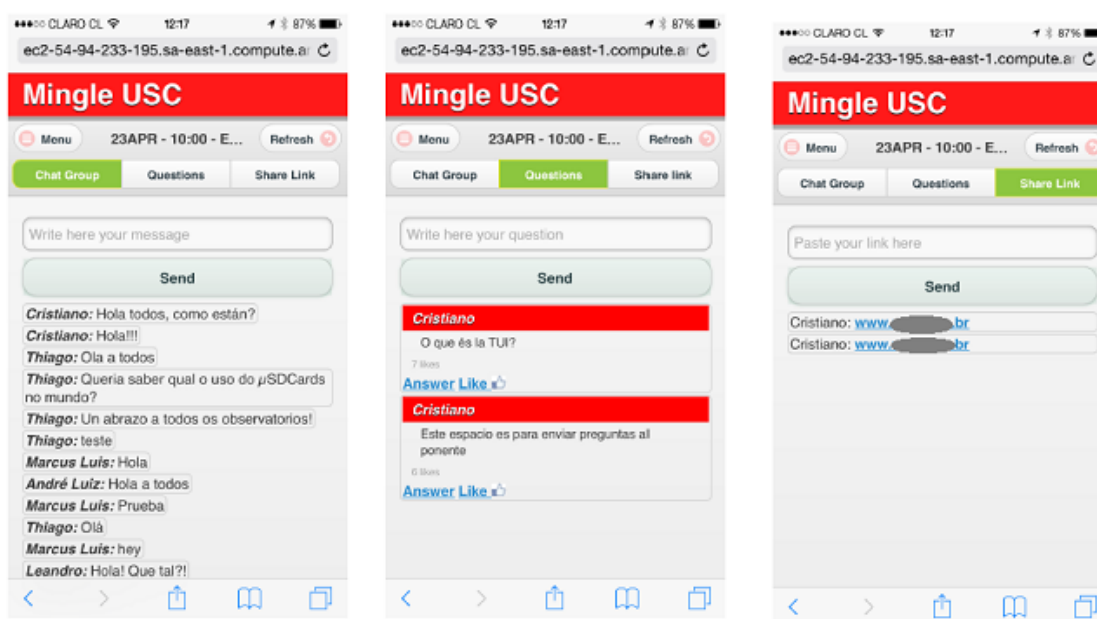


Figura 1. Telas de funcionalidades do Mingle.

Com base nestas informações, foi construído um protótipo inicial com o intuito de validar a proposta. As telas do protótipo com os requisitos acima comentados podem ser vistas na Figura 1. Outros recursos como mapas, agenda, perfil, entre outras, podem ser vistos na Figura 2.

O modelo foi desenvolvido para plataforma web, utilizando o padrão Model-View-Controller (MVC), com módulos comunicando-se via REST através de requisições JSON, permitindo boa comunicação mesmo em situações em que há limitação de recursos e de largura de banda no local. Construído inicialmente como um *WebService*, esta arquitetura permite futuramente a criação de aplicativos nativos para plataformas móveis (Android, IOS), que podem consumi-lo por meio da API desenvolvida, ampliando o número de sensores e informações disponíveis, visto que poderiam ser utilizados os recursos do dispositivo para a coleta de informações de GPS, por exemplo.



Figura 2. Telas de funcionalidades do Mingle.

A Figura 3 apresenta o fluxo interno do Mingle. O servidor recebe solicitações do usuário (Passo 1), disparando uma requisição ao controlador. Inicialmente, obtemos a requisição com os dados provenientes de um objeto JSON. Através da manipulação deste objeto, é capturado o conteúdo JSON do corpo da requisição e preparada para envio ao banco de dados (Passo 2). Depois de avaliado se o pedido foi de uma consulta (agenda, pesquisa...) (Passo 3-a) ou uma entrada no chat da célula (Passo 3-b), ela é enviada para a base de dados respectiva.

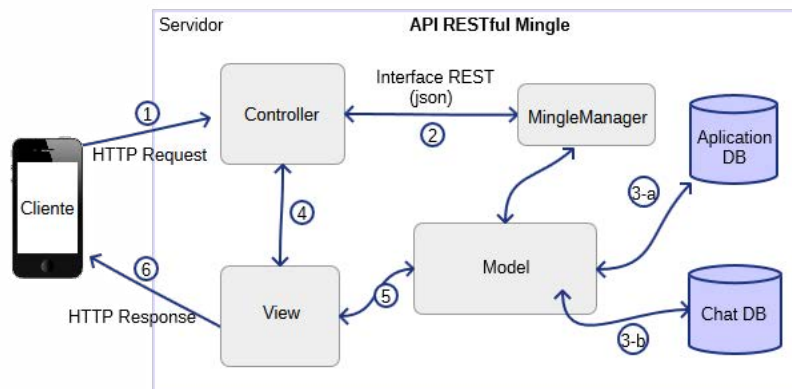


Figura 3. Arquitetura do Mingle e principais interações entre os componentes.

3. Avaliação e Discussão

O experimento foi realizado durante congressos em duas edições. Os dados foram colhidos durante as duas edições do Congresso Internacional dedicado ao TUI (Cartão Universitário Inteligente), evento destinado a análise de tendências, situação e avanço na utilização do TUI no mundo, ambos ocorridos no Chile, a primeira edição compreendida entre os dias 23 e 24 de abril, doravante denominado de Célula 1 e a segunda edição ocorrida no dia 3 de dezembro, doravante denominado de Célula 2, ambos no ano de 2015. Os eventos apresentaram um número expressivo de participantes recorrentes, característica que foi utilizada para avaliação da proposta. Foram colhidos como grupo

amostral aqueles usuários que participaram das duas edições e que preencheram a pesquisa nas duas ocasiões.

A avaliação segue duas abordagens, a primeira focando quantitativamente os números de acesso ao ambiente através das estatísticas geradas pela utilização. A segunda foi uma abordagem qualitativa através de aplicação de questionário junto aos usuários. A primeira etapa de avaliação foi executada por meio do uso de ferramentas digitais para mensuração do conteúdo produzido e volume de interações nos conteúdos propostos. A segunda focou em um nível de realidade que não pode ser quantificado, trabalhando com o universo de significados, das ações, motivos, crenças, aspirações, valores, atitudes e relações humanas, captadas a partir do olhar dos pesquisadores, ou seja, preocupa-se principalmente em compreender e explicar a dinâmica das relações sociais [Minayo 2007]. Para sua avaliação, foi aplicado um questionário direcionado para investigar aspectos como usabilidade, confiabilidade, atendimento as expectativas, pontos negativos e positivos, dentre outras questões.

3.1. Avaliação quantitativa

Para obtenção das métricas de acesso e tráfego no ambiente foram extraídas estatísticas com auxílio da ferramenta Google Analytics (2016). Os dados foram colhidos durante os dias que ocorreram os congressos. Dentre os dados levantados na primeira aferição da ferramenta, no primeiro dia da Célula 1 foi registrado um pico de acessos de 4.255 visualizações de páginas em um único dia, com um tempo médio de 18 minutos e 35 segundos por sessão e que esta média aumentou para 22 minutos e 53 segundos no segundo dia. Nas leituras posteriores notou-se que o número estabilizou na média de 3.100 visualizações durante os quatro dias que se sucederam ao término do evento.

Referente ao comportamento dos usuários seguem números obtidos na primeira e segunda edição do evento:

- Na Célula 1 o ambiente contava com 127 usuários ativos, usuários que apresentavam um mínimo de dois acessos durante o período em que a extração de dados foi efetuada. Em um total de 149 participantes cadastrados temos um percentual de 85,23% de adesão;
- Na Célula 2 o ambiente contava com 84 usuários ativos, usuários que apresentam um mínimo de dois acessos durante o período em que a extração de dados foi efetuada. Em um total de 164 participantes cadastrados temos um percentual de 51,22% de adesão;

Referente ao percentual de tempo de permanência por sessão, aqueles que realizaram autenticação e realizaram interações com o ambiente, seguem números obtidos na primeira e segunda edição do evento:

- Na Célula 1 foi registrado um tempo médio de 20 minutos e 44 segundos por sessão, com uma média de 10 autenticações durante o período avaliado, com tempo médio de 1 minuto e 22 segundos por página visualizada;
- Na Célula 2 foi registrado um tempo médio de 26 minutos e 47 segundos por sessão, com uma média de 12 autenticações durante o período avaliado, com tempo médio de 2 minutos e 56 segundos por página visualizada;

Referente ao número de mensagens trocadas somadas todas as interações com o ambiente, seguem números obtidos na primeira e segunda edição do evento:

- Na Célula 1 foram registradas 106 mensagens trocadas, somando intervenções no chat, troca de links e perguntas/respostas das palestras;
- Na Célula 2 foram registradas 245 mensagens trocadas, somando intervenções no chat, troca de links e perguntas/respostas das palestras;

Os dados demonstraram que a plataforma, em um primeiro momento recebeu um grande número de participantes que efetuaram apenas leituras rápidas as mensagens, sem interação com a ferramenta. Notou-se uma diferença significativa no decorrer no tempo, onde 78% das mensagens enviadas foram do final do primeiro dia até o término do evento. Neste primeiro contato, as mensagens, em sua maioria, eram para testar o ambiente e conforme a confiança aumentou, os conteúdos passaram a ser relevantes e relacionados aos temas propostos durante as palestras.

Comparando as duas Células, na primeira nota-se um número maior de adesões, levando em consideração o número de inscritos no evento comparado ao número de usuários ativos. No entanto, durante a ocorrência da Célula 2 notou-se um número menor de adesões, mas um aumento no tempo de permanência e número de mensagens trocadas.

3.2. Pesquisa Qualitativa

A segunda estratégia de avaliação foi convidar os usuários da plataforma a responderem um formulário de pesquisa com o intuito de conhecer as impressões dos mesmos no uso da ferramenta. O mesmo questionário foi aplicado nas duas Células e com ele buscou-se verificar aspectos tanto da ferramenta, como usabilidade e navegabilidade, como suas expectativas, uso da interface e visão de futuros trabalhos aplicando os recursos apresentados. Para manter equilíbrio no grupo amostral, foi selecionado um grupo de 20 usuários que tenha respondido ambas as edições do evento.

O questionário aplicado para esta avaliação esteve disponível no próprio ambiente em uma das opções do *menu* principal. Essa avaliação seguiu o modelo de aceitação tecnológica (TAM) proposto por Davis (1989) e ampliado por Yoon e Kim (2007), usando escala de Likert (1932). Conforme Davis (1989), dentre os vários fatores que os usuários podem considerar importantes para aceitar ou rejeitar um aplicativo o primeiro deles é a percepção de sua utilidade, ou seja, se ele consegue notar melhora ou ganhos na sua experiência na medida em que ele o utiliza. O segundo fator que pode influenciar é quanto a de facilidade na utilização, a qual é definida como o grau de facilidade com que as pessoas podem empregar uma ferramenta ou objeto a fim de realizar uma tarefa específica e importante. Este critério complementa o primeiro, porque de acordo com pesquisas [Yoon e Kim 2007], fornecer um aplicativo com boa usabilidade não significa que esse será útil para o usuário, ou que os benefícios da aplicação justifiquem uma alta complexidade no seu uso.

Para avaliação, os usuários foram convidados a selecionar uma das opções na escala proposta. Os critérios foram definidos como: Discordo, Discordo parcialmente, Indiferente, Concordo parcialmente e Concordo.

Tabela 1. Resultados da avaliação para o item Percepção de Utilidade

Questão	Concordo	Concordo parcialmente	Indiferente	Discordo parcialmente	Discordo
1. As opções do menu são relevantes.	85,7 % (18)	14,3 % (3)	0 % (0)	0 % (0)	0 % (0)
2. O App facilita a interação com o apresentador.	76,2 % (16)	19 % (4)	0 % (0)	4,8 % (1)	0 % (0)
3. O App estimula a discussão de temas apresentados.	85,7 % (18)	14,3 % (3)	0 % (0)	0 % (0)	0 % (0)
4. O App facilita as interações em uma conferência.	81 % (17)	19 % (4)	0 % (0)	0 % (0)	0 % (0)
5. Gostaria de usar o App em eventos na minha universidade ou empresa, se disponível.	76,2 % (16)	19 % (4)	4,8 % (1)	0 % (0)	0 % (0)

Na Tabela 1, são apresentadas as questões que abordam a percepção de utilidade da ferramenta em relação ao uso na comunicação durante o decorrer das palestras e atividades. Na Tabela 2, podemos ver as questões que abordam a interface da ferramenta avaliando as características de usabilidade. Para embasar a avaliação foram aplicados os conceitos de Nielsen (2014), apresentando a cada questão os parâmetros de avaliação e a definição de cada quesito.

Tabela 2. Resultados da avaliação para o item Facilidade de Utilização

Questão	Concordo	Concordo parcialmente	Indiferente	Discordo parcialmente	Discordo
1. O aplicativo é fácil de entender.	61,9 % (13)	33,3 % (7)	4,8 % (1)	0 % (0)	0 % (0)
2. O aplicativo é fácil de usar.	57,1 % (12)	42,9 % (9)	0 % (0)	0 % (0)	0 % (0)
3. A navegação nas opções é simples.	47,6 % (10)	42,9 % (9)	4,8 % (1)	0 % (0)	4,8 % (1)
4. Com pouco esforço, eu posso interagir em um grupo usando o bate-papo.	95,2 % (20)	4,8 % (1)	0 % (0)	0 % (0)	0 % (0)
5. Com pouco esforço, posso fazer uma pergunta em um grupo.	85,7 % (18)	14,3 % (3)	0 % (0)	0 % (0)	0 % (0)
6. Com pouco esforço, eu posso compartilhar um link.	66,7 % (14)	14,3 % (3)	19 % (4)	0 % (0)	0 % (0)
7. As opções são claras e objetivas.	61,9 % (13)	38,1 % (8)	0 % (0)	0 % (0)	0 % (0)

Também foi apresentada uma questão dissertativa aos avaliadores: "Quais suas sugestões para possamos efetuar melhorias na próxima versão do Mingle?". Comparando as respostas, notou-se que 80% dos participantes acreditam que a proposta incrementou a comunicação entre os usuários durante os eventos, não só aumentando a rede de disseminação de conhecimento e colaboração, mas como ferramenta para discussão de projetos correlacionados. Este número não teve variação significativa entre os eventos.

Outras respostas mostraram impressões positivas também quanto à aparência do ambiente e a aplicação dos recursos para colaboração, pelo canal para trocar experiências e compartilhar informações. Em contraponto, um pequeno número apontou que não vislumbrava a utilidade para auxílio das práticas e que não utilizaria em outros eventos. Outros relataram questões de desempenho na atualização das mensagens.

4. Trabalhos relacionados

Nesta seção são apresentados alguns trabalhos que empregam conceitos de rede social aplicada ao campo da computação móvel e ubíqua. Entre aqueles, foi considerado o mais representativo e também os modelos que tiveram alguma espontaneidade, ou seja, criar redes sociais relacionadas a um contexto específico.

O projeto C3PO [Laforest et al. 2014] propõe um modelo para troca de conteúdo multimídia em redes sociais que denominam Redes Sociais Espontâneas e Efêmeras (SESNs). Eles contam com uma arquitetura *peer-to-peer* distribuída formada espontaneamente por dispositivos móveis de participantes do evento. As SESN podem ser criadas por qualquer usuário e anunciado para outros usuários nas imediações. Estes serviços são descobertos invocando técnicas de redes oportunistas. Confirmando participação, eles serão capazes de comentar as apresentações, compartilhar fotos durante a conferência e os eventos sociais, votar na melhor apresentação ou demonstração, entre outras atividades.

Em [Sancho-Asensio et. al 2014] é apresentado um *framework* orientado à formação de grupos em ambientes de Aprendizagem Colaborativa, efetuando agrupamentos de alunos com base em seus perfis. Como motivação para realizar o agrupamento apresenta o objetivo de fomentar o trabalho colaborativo dos alunos.

Tourist-MSN [Arnaboldi et. al. 2014] projetados para melhorar experiências durante visitas turísticas, permitindo que os indivíduos possam criar, coletar e compartilhar informações úteis, através de mensagens assíncronas ou por chat em tempo real, relacionadas a pontos de visita turística. A aplicação utiliza CAMEO (Context-Aware Middleware for Opportunistic Mobile Social Networks), plataforma para dispositivos móveis capaz de fornecer um conjunto de ferramentas para o desenvolvimento de serviços e aplicações colaborativas (baseado em comunicações oportunistas) com alto impacto social [Arnaboldi et. al. 2011]. Sendo assim, ela fornece um conjunto de funcionalidades comuns a redes sociais móveis com base na coleta eficiente, modelagem e inferência de informações de contexto multidimensional, que visam reconhecer a localização do usuário e fornecer a ele informações pertinentes.

5. Considerações finais

Com a influência da Web sobre todas as facetas da vida contemporânea, vidas virtual e real estão cada vez mais se integrando, criando novas oportunidades para o compartilhamento de informações, desenvolvimento de aplicações que exploram os conceitos de sensibilidade ao contexto e as redes sociais. Neste trabalho, apresentamos o protótipo criado para uma rede social espontânea, denominado Mingle, e realizamos a sua avaliação tendo em mente a ideia de fomentar o compartilhamento de informações e conteúdos que estão relacionados com locais específicos.

O protótipo foi construído como um *webservice* utilizando o padrão MVC, com módulos comunicando-se via RESTful, permitindo boa comunicação mesmo em situações em que há limitação de recursos e de largura de banda no local. Esse recurso é especialmente interessante quando o cliente utilizado se tratar de *smartphones* ou *tablets*, considerando capacidade limitada em termos de poder de processamento e bateria.

Além disso, o cenário desenvolvido permite a criação de aplicativos nativos para plataformas móveis (Android, IOS) consumindo o *webservice* por meio da API desenvolvida, neste caso ampliando o número de sensores e informações disponíveis para calibragem do perfil.

A avaliação discutida nesse artigo mostrou uma perspectiva positiva sobre a proposta, mostrando contentamento e facilidade no uso das ferramentas disponibilizadas aos participantes mesmo sendo tímidas as contribuições à produção de conteúdo, mas, ainda assim, dentro de uma faixa esperada nesse período de adaptação.

Como trabalho futuro, pretendemos expandir o modelo com o gerenciamento automatizado dos contextos [WAGNER, BARBOSA e BARBOSA 2014], permitindo um compartilhamento de informações entre diferentes células, a respeito do usuário e seu histórico. Finalmente, pretendemos aumentar a avaliação da plataforma Mingle em um ambiente real e testar o servidor com cargas geradas a partir desse uso.

Agradecimentos

Os autores gostariam de agradecer ao Santander Universidades pelo apoio à essa pesquisa.

Referências

- Arnaboldi, V.; Conti, M.; Delmastro, F. (2011) "Implementation of CAMEO: a context-aware middleware for opportunistic mobile social networks", IEEE Int. Symposium World of Wireless, Mobile and Multimedia Networks, IEEE Computer Society, pp. 1–3.
- _____. (2014) "Cameo: a novel context-aware middleware for opportunistic mobile social networks", Pervasive and Mobile Computing, vol. 11, Elsevier, pp. 148–167.
- Beach, A., Gartrell, M., Akkala, S., Elston, J. et al. (2008) "WhoThat? evolving an ecosystem for context-aware mobile social networks", IEEE Network, vol. 22, no. 4, pp. 50–55.
- Costa, C., Yamin, A. e Geyer, C. (2008) "Toward a general software infrastructure for ubiquitous computing", IEEE Pervasive Computing, vol. 7, pp. 64–73.
- Costa, C. A., Zaupa, D., Barbosa, J. L. V., Righi, R. D. R., Camargo, J. Daroit, N., Yamin, A. C. (2014) "A spontaneous social network based on mobile devices", Social Network Analysis and Mining, vol. 4, no. 1, pp. 158.
- Davis, F.D. (1989) "Perceived usefulness, perceived ease of use, and user acceptance". MIS Q 319–340
- Duarte, Fábio, Frei, Klaus. (2008) "Redes Urbanas. Redes Urbanas", In: Duarte, Fábio; Quandt, Carlos; Souza, Queila. O Tempo Das Redes, p.156.

- Google Analytics. (2016) “Gráficos de acesso e utilização”. Disponível em: <<https://accounts.google.com/>>.
- Guedes, Taís Morais. (2013) “As Redes Sociais — Facebook e Twitter — e suas influências nos Movimentos Sociais”, Monografia de especialização. Universidade de Brasília. Brasília.
- Laforest, F. et al. (2014) “C3PO: a Spontaneous and Ephemeral Social Networking Framework for a collaborative Creation and Publishing of Multimedia Contents”, In MoWNet 2014, pages 1–6, Rome, Italy, Elsevier.
- Likert, R. (1932) “A technique for the measurement of attitudes”. Arch Psychol 1–55.
- Minayo, M.C.S. (2007) “O desafio do conhecimento: Pesquisa qualitativa em saúde”. 10.ed. São Paulo.
- Nielsen, Jakob. (2014) “Usability 101: Introduction to usability”. Disponível em: <<http://www.useit.com/alertbox/20030825.html>>.
- Sancho-Asensio, A., Sole, X., Montero, J., Navarro, J., Canaleta, X., Vernet, D. (2014) “Support tool for the formation of working groups in collaborative learning environments”, In: Information Systems And Technologies (Cisti), 2014 9th Iberian Conference On.
- Satyanarayanan, M. (2001) “Pervasive Computing: vision and challenges”, IEEE Personal Communications, [S.l.], v. 8, p. 10–17.
- Vaughan-Nichols, S. (2009) “Will mobile computing’s future be location, location, location?”, Computer, 42(2):14 –17.
- Wagner, A.; Barbosa, J.L.V.; Barbosa, D.N.F. (2014) “A Model for Profile Management applied to Ubiquitous Learning environments”. Expert System with Application, v.41, n.4, p. 2023-2034
- Weiser, M. (1991) “The computer for the 21st century”, Scientific American, v. 265, n. 3, p. 94-104.
- Yoon C, Kim S (2007) “Convenience and TAM in a ubiquitous computing environment: the case of wireless LAN”. Electron Commer Res Appl 102–112. doi:10.1016/j.elerap.2006.06.009

13° SBSC - Simpósio Brasileiro de Sistemas Colaborativos

Apresentação

O Simpósio Brasileiro de Sistemas Colaborativos (SBSC) discute o desenvolvimento, a avaliação e o uso de ferramentas para apoiar a colaboração entre pessoas. Alguns exemplos atuais de sistemas colaborativos de sucesso incluem redes sociais, colaboração utilizando dispositivos móveis e ubíquos, sistemas de compartilhamento de arquivos, mundos virtuais, *wikis*, sistemas de gestão de conhecimento, ambientes de desenvolvimento distribuído de software, e ferramentas de comunicação, como videoconferência e *blogs*. As discussões envolvem, sob várias perspectivas, o desenvolvimento e a avaliação de sistemas colaborativos, a investigação de como a colaboração ocorre na prática e a aplicação de sistemas colaborativos em diversos domínios.

Em seu 13º ano, o SBSC se consolida como um importante fórum para pesquisadores, estudantes e profissionais discutirem problemas, técnicas, soluções, inovações e experiências sobre esta área multidisciplinar. Nesta edição, o SBSC aprovou trabalhos fechados e em andamento, em português ou em inglês. O evento também conta com o 7º Workshop de Teses e Dissertações em Sistemas Colaborativos, o WTD-SC 2016.

No SBSC 2016 foram submetidos 50 artigos, dos quais 25 foram aprovados. A taxa de aprovação foi de 50%.

Sabrina Marczak (PUCRS)

Josiane Kroll (PUCRS)

Comitê de Organização

Coordenadoras gerais do evento

Sabrina Marczak (PUCRS)

Josiane Kroll (PUCRS)

Coordenadoras do comitê de programa

Denise Del Re Filippo (UERJ)

Carla Diacui Medeiros Berkenbrock (UDESC)

Coordenadora do Workshop de Teses e Dissertações

Rita Suzana Maciel (UFBA)

Coordenadoras do Comitê Gestor da Comissão Especial de Sistemas Colaborativos

Raquel Oliveira Prates (UFMG) - coordenadora

Thais Helena Castro (UFAM) – vice coordenadora

Comitê de Programa

Adriana Vivacqua (UFRJ)

Alberto Castro (UFAM)

Alberto Raposo (PUC-Rio)

Ana Cristina Garcia (UFF)

Andrea Magdaleno (UNIRIO)

Bruno Freitas Gadelha (UFAM)

Carla Berkenbrock (UDESC)

Carlos Ferraz (UFPE)

Celso Hirata (ITA)

Cesar Tacla (UTFPR)

Claudia Cappelli (UNIRIO)

Claudia Motta (UFRJ)

Cleidson de Souza (ITV e UFPA)

Clever Farias (USP)

Crediné de Menezes (UFRGS)

Cristiano Maciel (UFMT)

Daniel Paiva (UFPE)

Denise Filippo (UERJ)

Edeilson Milhomem Silva (CEULP/ULBRA)

Elaine Oliveira (UFAM)

Flavia Santoro (UNIRIO)

Geraldo Xexéo (UFRJ)

Hugo Fuks (PUC-Rio)

Igor Steinmacher (UTFPR)

Ivan Ricarte (UNICAMP)

Jauvane Oliveira (LNCC)

João Porto de Albuquerque (USP)

Jorge Correia Neto (UFRPE)

José Maria David (UFJF)
Marco Mangan (PUCRS)
Marco Aurelio Gerosa (USP)
Marcos Borges (UFRJ)
Mariano Pimentel (UNIRIO)
Melise Paula (UNIFEI)
Nazareno Andrade (UFMG)
Patrícia Tedesco (UFPE)
Pedro Porfirio Farias (UNIFOR)
Rafael Prikladnicki (PUCRS)
Raquel Oliveira Prates (UFMG)
Renata Araujo (UNIRIO)
Rita Suzana Pitangueira Maciel (UFBA)
Roberta Gomes (UFES)
Roberto Willrich (UFSC)
Rosiane de Freitas (UFAM)
Sabrina Marczak (PUCRS)
Thais Castro (UFAM)
Vaninha Vieira (UFBA)
Wallace Ugulino (PUC-Rio)

Palestras

Technology support for learning social-emotional skills for interaction - the building blocks of CSCW

Geraldine Fitzpatrick (Vienna University of Technology)

In Computer Supported Cooperative Work (CSCW) research, the focus is on enabling people to collaborate and communicate together via advances in mobile and social networking technologies. However, the basic social skills that are employed by people as part of these communications and collaborations are largely assumed and taken for granted. We are interested in how technologies can open up opportunities to support the basics of social interaction, such as connectedness, rapport and empathy, and to support the teaching and learning of social and emotional skills. Drawing on recent advances in Psychology, Social Neuroscience and Prevention Science, as well as ubiquitous computing, I explore the design of new technologies to support the more implicit aspects of social interaction and the learning of social-emotional skills, with a particular concern for interactions in everyday contexts.

Design Science Research: método de pesquisa científica com desenvolvimento de artefatos

Mariano Pimentel (UNIRIO)

Muitas pesquisas na área de Sistemas de Informação são realizadas a partir do desenvolvimento de algum artefato: um sistema computacional, um dispositivo eletrônico, um método ou um modelo. Contudo, os métodos mais tradicionais de pesquisa, como Experimento e Estudo de Caso, não pressupõem o desenvolvimento de um artefato, sendo motivo de várias dúvidas de pós-graduandos(as) nesta área de pesquisa, o que eventualmente resulta em artigos, dissertações e teses com baixa utilidade ou rigor, em que: se faz a avaliação das qualidades do artefato sem com isso apresentar uma contribuição para a base de conhecimento científico (falta de rigor); ou se realiza um Estudo de Caso em cenários não-reais; ou se conduz um Experimento quando não é relevante a verificação da relação entre duas variáveis específicas. É em Design Science Research (DSR) que são encontrados fundamentos que legitimam o desenvolvimento de artefatos como um meio para a produção de conhecimentos científicos do ponto de vista epistemológico e filosófico, principalmente a partir da obra de Hebert Simon (1969) sobre as Ciências do Artificial. Este novo paradigma de fazer ciência tem se popularizado na área de Sistemas de Informação, principalmente a partir da década de 1990, sendo atualmente uma das abordagens de pesquisa mais importantes nesta área. Design Science Research possui semelhanças e diferenças com outras abordagens, podendo até ser confundida com Pesquisa-ação, sendo necessário o(a) pesquisador(a) compreender e diferenciar DSR do ponto de vista ontológico. Do ponto de vista metodológico, vários métodos já foram e ainda estão sendo propostos, não havendo um único método consensual e consagrado sobre como se faz pesquisa em DSR, o que também gera muitas dúvidas nos pesquisadores que adotam esta abordagem. Nesta palestra, irei discutir a relevância de DSR na área de Sistemas Colaborativos do ponto de vista epistemológico-filosófico; sua diferenciação de outras abordagens de pesquisa científica, do ponto de vista ontológico; e a profusão dos métodos de pesquisa em DSR. Serão também discutidos exemplos de pesquisas realizadas em Sistemas Colaborativos seguindo a abordagem DSR.

Trabalhos aceitos no WTD-SC

WTD-SC - 7º Workshop de Teses e Dissertações em Sistemas Colaborativos

O Workshop de Teses e Dissertações em Sistemas Colaborativos (WTD-SC) é um fórum dedicado à apresentação e discussão de trabalhos de mestrado e de doutorado, em Sistemas Colaborativos, desenvolvidos nos programas de pós-graduação no Brasil. O WTD-SC é uma excelente oportunidade para o estudante de pós-graduação apresentar a pesquisa em andamento de sua dissertação de mestrado ou de sua tese de doutorado. Isto permite ao estudante receber comentários e sugestões de pesquisadores experientes na área de Sistemas Colaborativos. São submetidos trabalhos que tenham prazo previsto de conclusão a partir de janeiro de 2017. Os tópicos de interesse são os mesmos dos artigos do SBSC 2016.

Sessão técnica 1- Segurança, privacidade e identidade em sistemas colaborativos

Privacidade para Crianças e Adolescentes em Redes Sociais Online sob a lente da Usabilidade: Um Estudo de Caso no Facebook

Cristiana Silva (PUC Minas)
Glúvia Barbosa (CEFET-MG)
Ismael Santana (CEFET-MG)
Tatiana Silva (Faculdade de Ciências Médicas)
Fernando Mourão (UFSJ)

PryVis: Uma ferramenta para Modelagem de Design de Privacidade

Maria Villela (UFVJM)
Lidia Ferreira (UFMG)
Diego Barros (UFMG)
Raquel Prates (UFMG)
Raquel de Melo-Minardi (UFMG)

Caracterização das Estratégias de Privacidade do Instagram

Geanderson Santos (UFMG)
Glúvia Barbosa (CEFET-MG)
Marcelo Barbosa (PUC Minas)

Uma Análise do Impacto do Anonimato em Comentários de Notícias Online

Julio Reis (UFMG)
Manoel Júnior (UFMG)
Lucas Miranda (UFMG)
Raquel Prates (UFMG)
Fabricio Benevenuto (UFMG)

Sessão técnica 2- Estudos empíricos de avaliação de sistemas colaborativos

Identifying Awareness Requirements in Face-to-Face Collaborative Applications for Users with Autism Spectrum Disorders

Greis Silva (PUC-Rio)
Alberto Raposo (PUC-Rio)

Interoperability Requirement to Enhance Collaboration in Software Product Lines: A Systematic Mapping

Frâncila Neiva (UFJF)
Heleno Campos (UFJF)
José Maria David (UFJF)
Regina Braga (UFJF)
Marco Antônio Araújo (UFJF)
Fernanda Campos (UFJF)
Rita Suzana Pitangueira Maciel (UFBA)

Aplicando o Design Science Research no Desenvolvimento de um Sistema Colaborativo Assistivo

Daniel Silva (UDESC)
Deivid Sartori (UDESC)
Carla Berkenbrock (UDESC)
Celso Hirata (ITA)

Um Estudo Empírico sobre Crowdsourcing: O Brasil na plataforma TopCoder

Carlos Lopes (UFPA)
Ricardo Melo (UFPA)
Renato Ferreira (UFPA)
Muller Miranda (UFPA)
Cleudson de Souza (ITV and UFPA)

Os Programas de Engajamento em Software Livre Atraem Estudantes à Colaboração Voluntária? Um Estudo Empírico do Google Summer of Code

Jefferson Silva (USP)
Igor Steinmacher (UTFPR)
Marco Aurelio Gerosa (IME - USP)

Sessão técnica 3 - Colaboração utilizando dispositivos móveis e ubíquos; novos dispositivos de interação

Uma Ferramenta de Telepresença Imersiva usando Oculus Rift

Philip Dunker (PUC-Rio)
Alberto Raposo (PUC-Rio)

Carona Universitária App: Investigando as Relações de Troca, Segurança e Recompensa no Desenvolvimento de uma Aplicação Móvel Voltada a uma Comunidade

Antônio Henrique Gonçalves Leite (UFMG)
Luiz Paulo Damilton Corrêa (UFMG)
Raquel Prates (UFMG)
Fabricio Benevenuto (UFMG)
Pedro Olmo Vaz de Melo (UFMG)

Processo de Design Colaborativo de Artefatos Tangíveis para Crianças

Isomar Lima (UFAM)
Thais Castro (UFAM)
Adriana Doroteu Dantas (UFAM)
Marcos Paulo Siqueira de Farias (UFAM)

CAPCM - Construtor de arquiteturas pedagógicas para dinâmicas colaborativas com textos multimídia

Renan Almeida (UFES)
Orivaldo Tavares (UFES)
Luiz Reinoso (UFES)

Vox Uma plataforma Crowdsourcing para busca de pessoas desaparecidas e cadastro de moradores de rua

Priscila Knoop Silveira (IFSul)
Rodrigo Ruas Oliveira (UFRGS)
Rodrigo Remor Oliveira (IFSul)

Sessão técnica 4 - Redes Sociais e Colaboração

Análise das Interações Sociais entre os Participantes de um Curso EaD: Uma Revisão Sistemática da Literatura

Francisco S.S Neto (UFAM)
José Francisco Netto (UFAM)
Dhanielly Lima (UFAM)

Caracterização das Estratégias de Qualidade de Uso que Potencializam a Utilização Contínua de Aplicativos Colaborativos de Navegação: Um estudo de caso no Waze

Glívia Barbosa (CEFET-MG)
Geanderson Santos (UFMG)
Lorena Tomagnini (CEFET-MG)
Ismael Santana (CEFET-MG)

Por onde quero ir? Interesses pessoais na escolha de um trajeto urbano

Ana Paula Chaves Steinmacher (UTFPR)
Igor Steinmacher (UTFPR)
Marco Aurelio Gerosa (IME - USP)

PoI: uma Aplicação de Detecção de Pontos de Interesse

Alice Menezes (UFAM)
Thais Almeida (UFAM)
Bernardo Gatto (UFAM)
Eulanda Santos (UFAM)
Eduardo Nakamura (UFAM)
Carlos Mauricio Figueiredo (UEA)

Análise da Ferramenta Blueprint de Valor no Cenário de Ecossistemas de Negócios de Software

Luciana Almeida (UFPA)

Cleidson de Souza (ITV and UFPA)

Um Processo para o Desenvolvimento de Serviços de Colaboração em um Ecossistema de Software Científico

Marcio Arakaki (UFJF)

Guilherme Martins (UFJF)

José Maria David (UFJF)

Regina Braga (UFJF)

Fernanda Campos (UFJF)

Frâncila Neiva (UFRJ)

Sessão técnica 5 - Aprendizagem colaborativa com suporte computacional

Uma Revisão Sistemática da Literatura para Investigação de Estratégias de Ensino Colaborativo

Carla Berkenbrock (UDESC)

Simone Erbs da Costa (FURB)

Avaliação por pares como ferramenta colaborativa na correção de redações: Um experimento com o ambiente educacional MeuTutor

Thyago Tenório (UFAL)

Ig Ibert Bittencourt (UFAL)

Uma Abordagem Colaborativa para Aprendizagem de Programação Orientada a Objetos

Leonardo Davi Machado (UDESC)

Carla Berkenbrock (UDESC)

Ivanete Siple (UDESC)

Celso Hirata (ITA)

Sistema multiagentes para recuperação e análise de informações como suporte a mediação pedagógica em ambientes colaborativos de aprendizagem

Sabrina Panceri (UFES)

Crediné de Menezes (UFRGS)

O uso de Storytelling para apoiar o compartilhamento de conhecimento

Roberta Salim (UFRJ)

Angélica Dias (UFRJ)

Juliana França (UFRJ)

Bruno Lage (UFRJ)

Marcos Borges (UFRJ)

Patrocínio



Organização em cooperação com



Privacidade para Crianças e Adolescentes em Redes Sociais Online sob a lente da Usabilidade: Um Estudo de Caso no Facebook

Cristiana S. Silva¹, Glívia A. R. Barbosa², Ismael S. Silva², Tatiana S. Silva³, Fernando H. Mourão⁴

¹Instituto de Ciências Exatas e Informática – Pontifícia Universidade Católica de Minas Gerais (PUC-MG). Av. Afonso Vaz de Melo, 1.200 - Barreiro de Baixo - Belo Horizonte – MG – Brasil.

²Departamento de Computação – Centro Federal de Educação Tecnológica de Minas Gerais (DECOM/CEFET-MG). Av. Amazonas, 7675 - Nova Gameleira - Belo Horizonte, MG – Brasil.

³Faculdade de Ciências Médicas de Minas Gerais. Alameda Ezequiel Dias, 275 - Belo Horizonte, MG – Brasil.

⁴Departamento de Ciência da Computação - Universidade Federal de São João Del Rey. Praça Frei Orlando, 170, Centro, São João del-Rei, MG - Brasil

cristiana.sant.silva@gmail.com, gliviabarbosa@decom.cefetmg.br,
ismaelsantana@decom.cefet.br, tatabernado@hotmail.com,
fhmourao@ufsj.edu.br

Abstract. *This study aimed to characterize, through a case study on Facebook, such as children and adolescents deal with privacy features on online social networks and whether these resources are adequate for the use of this user's profile. The results indicated that there are usability violations that limit the use of the settings, by that user group that can compromise your privacy and security.*

Resumo. *Este trabalho buscou caracterizar, através de um estudo de caso no Facebook, como crianças e adolescentes lidam com os recursos de privacidade nas redes sociais online e se esses recursos estão adequados ao uso desse perfil de usuários. Os resultados indicaram que existem violações de usabilidade que limitam o uso das configurações, por parte desse grupo de usuário, que podem comprometer sua segurança e privacidade.*

1. Introdução

Projetistas e pesquisadores da Computação e áreas relacionadas (e.g., [Junior et al., 2014],[Rodrigues et al., 2012],[Villela e Prates, 2015]) têm direcionado seus esforços para avaliar e/ou propor mecanismos de configuração de privacidade com o intuito de potencializar a segurança dos usuários dessas redes sociais *online*. Porém, alguns trabalhos têm demonstrado que os modelos de privacidade correntes nem sempre oferecem proteção

adequada para os diferentes perfis de usuários, refletindo em uma disparidade entre a privacidade desejada e a real [Liu et al., 2011][Villela e Prates, 2015].

Conforme relatado por Villela e Prates (2015), essa disparidade ocorre, em tempo de interação do usuário com a rede social *online*, porque nem sempre a interface indica claramente aos vários perfis de usuários as formas e mecanismos para gerenciamento do nível de privacidade de seus dados. Em outras palavras, a usabilidade das configurações de privacidade pode não estar adequada ao uso dos diferentes tipos de usuários dessas redes [Chilana et al., 2012], sobretudo ao grupo formado por crianças e adolescentes [Fitton et al., 2014][Gilutz e Nielsen, 2002][Livingstone et al., 2011][Melo e Baranauskas, 2003].

Isso é um problema, porque apesar da existência de termos que indiquem o uso dessas redes sociais *online* para usuários acima de 13 anos [Boyd et al., 2011], estudos indicam que crianças e adolescentes têm utilizado esse tipo de sistema, na maioria das vezes, sem o acompanhamento dos responsáveis [Livingstone et al., 2011], reforçando a necessidade de também adequar os recursos para configuração de privacidade a esse tipo de usuário.

Dessa forma, torna-se relevante avaliar como a usabilidade dos controles de privacidade em redes sociais *online* reflete na percepção e uso que crianças e adolescentes fazem desses recursos. Isso porque, através dessa investigação será possível alertar sobre as vulnerabilidades que esse grupo de usuários pode estar sujeito, bem como melhorar a usabilidade dos mecanismos de privacidade existentes e/ou auxiliar no desenvolvimento de novas soluções que potencializem a privacidade de crianças e adolescentes nas redes sociais *online* [Fitton et al., 2014][Gilutz e Nielsen, 2002].

Nesse sentido, o objetivo desse trabalho consiste em avaliar e caracterizar como crianças e adolescentes têm lidado com as configurações de privacidade em seus perfis nas redes sociais *online* e se a usabilidade das funcionalidades que permitem tais configurações está adequada a esse perfil de usuários. Para isso foi realizado um estudo de caso no Facebook, considerando crianças e adolescentes do Brasil (i.e., usuários entre 6 a 17 anos, conforme definição do Art. 2º da Lei Nº 8.069 do Estatuto da Criança e do Adolescente (1990)¹ do Brasil).

O Facebook foi escolhido para o estudo de caso por ser a rede social *online* mais utilizada no mundo [Duggan et al., 2014][Duggan e Page, 2015] e porque, embora sua política de uso recomende sua utilização para maiores de 13 anos², dados como os apresentados por Boyd et al. (2011) apontam que crianças abaixo dessa idade fazem uso dessa rede social. Por sua vez, o estudo no Brasil se justifica, inicialmente porque o país é um dos campeões no uso de redes sociais, em especial o Facebook que possui aproximadamente 30% de seus usuários no Brasil na faixa etária entre 6 e 17 anos [comScore, 2015][Consumer Reports News, 2013]. Outra motivação refere-se a estudos que evidenciam os perigos que esse grupo de usuários está sujeito no país (e.g., casos de

¹Estatuto da Criança e do Adolescente do Brasil. Lei Nº 8.069. 13 de Julho de 1990. Available in: http://www.planalto.gov.br/CCIVIL_03/leis/L8069.htm.

²<https://www.facebook.com/help/210644045634222>

pedofilia no Brasil através de redes sociais *online* [O tempo, 2014] e exposição de intimidades por parte de adolescentes com desconhecidos [Balmant, 2012]).

Os principais resultados indicaram como os problemas de usabilidade têm influenciado no grau de conhecimento e uso limitado que crianças e adolescentes fazem das configurações de segurança e privacidade no Facebook. Em suma, esse trabalho apresenta contribuições práticas e científicas para a área de sistemas colaborativos e afins, que não se limitam ao contexto do Brasil. Em termos práticos, alerta os projetistas de interface sobre a importância da usabilidade das configurações de privacidade de redes sociais *online*. De forma complementar, os resultados reportados podem servir de parâmetros para a melhoria e/ou criação de mecanismos de privacidade nesses ambientes colaborativos, com maior usabilidade, para crianças e adolescentes. Já em termos científicos, esse trabalho é relevante porque traz à tona a discussão sobre a importância de se propor e avaliar abordagens que visam o projeto e avaliação de interfaces para configurações de privacidade, em sistemas colaborativos, voltadas para crianças e adolescentes.

2. Trabalhos Relacionados

A privacidade em redes sociais online é uma questão complexa que tem sido explorada em diferentes aspectos. Dentre as possíveis linhas de investigação destacam-se os trabalhos que buscam caracterizar, de forma geral, a percepção e uso das configurações de privacidade nessas redes (e.g., [Albesher e Alhussain, 2013][Liu et al., 2011][Rodrigues et al., 2012]). Por exemplo, em Albesher e Alhussain (2013), os autores investigam a eficiência dos mecanismos de privacidade do Facebook para controlar a interação entre os usuários e discutem sobre a importância de se revisar regularmente esses controles, bem como as políticas de privacidade, de forma que permaneçam sempre adequados ao uso da diversidade de usuários que utilizam essa rede *online*.

De forma complementar, outras pesquisas buscam analisar a privacidade sob a lente de grupos de usuários segmentados como, por exemplo, crianças e adolescentes (e.g., [Boyd et al., 2011][Hinduja e Patchin, 2008][Souza e Dick, 2009]). Em seu trabalho, Boyd et al. (2011) alerta para as consequências de segurança e privacidade que as crianças estão sujeitas ao mentirem sobre a idade em redes sociais *online*, com a permissão dos pais. O estudo indica que muitos pais consentem com essa omissão, mas não acompanham devidamente o uso que suas crianças fazem da rede. Isso as tornam vulneráveis diante de pessoas maliciosas, comprometendo a segurança e privacidade das mesmas.

O trabalho realizado por Hinduja e Patchin (2008), analisou, de forma empírica, quais informações pessoais (e.g., nome completo, foto de perfil, escolaridade e telefone) que os adolescentes expõem no MySpace. Os resultados indicaram que, exceto em relação a foto de perfil, a maioria dos participantes são conscientes e menos de 20% expõem dados como nome completo, escolaridade e telefone. Contudo, alertam para o fato de que mais da metade desses adolescentes utilizam fotos pessoais para compor o perfil e essa informação é uma identificação visual que deve ser controlada.

De forma similar, Souza e Dick (2009) examinam quais informações as crianças compartilham no MySpace e qual o nível de compreensão delas sobre as questões de privacidade. Os autores concluíram que a quantidade de informações pessoais exposta por

esse grupo de usuários está relacionada à, pelo menos, dois fatores: (1) o nível de conhecimento que elas têm sobre privacidade e (2) a adequação dos controles de segurança ao uso desse tipo de usuário. Isso porque os resultados sugeriram que as crianças que entendem melhor o conceito de privacidade e conseguem manipular a interface do MySpace, sem dificuldades, estão menos propensas a divulgar informações pessoais nessa rede *online*. Sendo assim, Hinduja e Patchin (2008) e Souza e Dick (2009) reforçam que análises como estas são necessárias não apenas para crianças e adolescentes, mas para outros perfis de usuários, de forma a alertar para as vulnerabilidades que eles podem estar sujeitos ao expor informações pessoais.

Uma vez que a percepção sobre a privacidade em redes sociais pode ser influenciada também pela cultura, diversos estudos têm procurado entender como pessoas de diferentes países lidam com essa questão (e.g., [Dey et al., 2012][Krasnova e Veltri, 2010][Tsoi e Chen, 2011]). Um exemplo é a pesquisa realizada por Tsoi e Chen (2011), na qual, através do uso de questionários e grupos focais, os autores apresentaram diferenças significativas entre os usuários de Hong Kong e da França em relação aos padrões de uso das redes sociais *online* e às suas respectivas configurações de privacidade.

Embora, diferentes aspectos relacionados à privacidade tenham sido investigados, não foram encontrados trabalhos que fizessem a análise proposta neste estudo, abordando a privacidade para crianças e adolescentes em redes sociais sob a lente da usabilidade. Além disso, o fato do presente artigo apresentar um estudo de caso no Facebook com crianças e adolescentes do Brasil, enraíza discussões significativas não apenas para o país, que é um dos campeões no uso das redes sociais *online*, mas também para toda a comunidade científica e para os profissionais que investigam e projetam aspectos relacionados a usabilidade e privacidade em redes sociais *online*. Isso porque o trabalho traz a tona como as questões de usabilidade podem impactar na percepção e uso que os usuários fazem dos controles de privacidade nessas redes.

3. Metodologia

Para atingir o objetivo proposto foi realizado um estudo de caso em três etapas. Inicialmente, foi realizada uma inspeção de usabilidade com base nas heurísticas propostas por Nielsen (1993). Assim, os recursos de privacidade do Facebook foram avaliados sob a perspectiva de especialistas em Interação Humano Computador (IHC). Essa análise ocorreu com o suporte de um profissional formado em Terapia Ocupacional especialista em psicopedagogia infanto-juvenil, a fim de constatar se essas funcionalidades são adequadas ao entendimento e uso de crianças e adolescentes do Brasil.

Posteriormente, foi realizada uma avaliação com esse grupo de usuários, através do método de entrevista semiestruturada, para detectar como eles utilizam esses recursos e o quão vulnerável eles estão nessa rede social. Finalmente, os resultados foram triangulados para que fosse possível verificar como a usabilidade das configurações de privacidade do Facebook tem impactado na segurança de crianças e adolescentes.

A triangulação refere-se a um procedimento da pesquisa qualitativa realizada através da comparação de dados extraídos por diferentes métodos, cujo intuito é conferir a validade dos resultados obtidos em uma pesquisa científica [Cho e Trent, 2006]. Neste

caso, a validação consiste em buscar diferentes interpretações para a mesma questão de pesquisa [Cho e Trent, 2006]. A seguir os principais resultados do estudo de caso são apresentados.

4. Usabilidade das Configurações de Privacidade do Facebook sob a Perspectiva de Especialistas

Uma das formas de analisar a usabilidade de um sistema é executar a Avaliação Heurística (AH) que é um método qualitativo consolidado para avaliar interfaces [Nielsen, 1994]. Nessa etapa do trabalho, os passos da AH foram executados, com o objetivo de se identificar possíveis problemas de usabilidade nos mecanismos de privacidade do Facebook para crianças e adolescentes do Brasil.

A avaliação foi realizada na primeira quinzena de março de 2015 por dois avaliadores com experiência na aplicação da AH. Os resultados foram validados por um especialista na área de IHC (i.e., profissional da área com mais de sete anos de experiência) e por um Terapeuta Ocupacional, especializado em psicopedagogia e comportamento infanto-juvenil, com mais de cinco anos de experiência na área. A atuação do especialista em psicopedagogia contribuiu para que os especialistas em IHC pudessem justificar porque os problemas de usabilidade identificados impactavam na interação de crianças e adolescentes.

4.1. Análise dos Potenciais Problemas Identificados

Durante a avaliação, foi encontrado um total de 12 problemas nas configurações de privacidade que violaram pelo menos uma heurística de usabilidade. Desses, 83% foram classificados com a gravidade 3 (i.e., problema grande), 8% com gravidade 4 (i.e., catastrófico) e 9% com gravidade 2 (i.e., problema pequeno).

Uma vez que cada problema poderia ser associado à ausência de uma ou mais heurísticas de usabilidade e que uma mesma heurística poderia ser violada em diferentes problemas, para realizar a análise proposta e apresentar os resultados de forma consolidada, foram verificadas quantas heurísticas de usabilidade foram violadas e com que frequência (i.e., incidência) essas violações aconteciam. A Figura 1 apresenta os resultados consolidados.

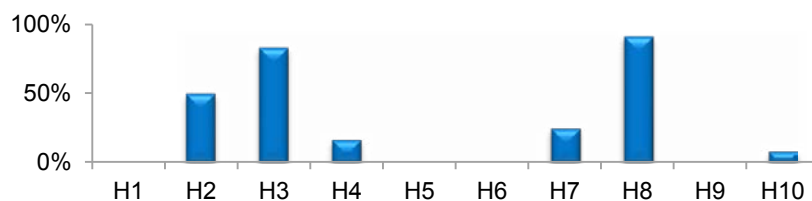


Figura 1. Violação das Heurísticas de Usabilidade nas Configurações de Privacidade do Facebook

Através desses dados, é possível notar que as heurísticas “H3 - Controle e liberdade do usuário”, “H8 - Flexibilidade e eficiência de uso” e “H2 - Correspondência entre o sistema e o mundo real” tiveram grande percentual de violação, ou seja, foram violadas em diferentes problemas. A H8 teve o maior percentual de violação com 92%, H3 teve o

segundo maior percentual com 83% e H2 com terceiro percentual teve 50% do total de incidências nos problemas.

A frequência similar de problemas violando essas três heurísticas pode estar associada à relação existente entre as mesmas. Isso porque, se o projetista não faz uso de elementos em sua interface que façam analogia ao mundo real, dificilmente o usuário consegue interagir com a interface de uma forma eficiente, e ao se deparar com uma interface pouco flexível o usuário se sente desmotivado no uso das funcionalidades. Tais problemas podem comprometer a percepção e uso dos mecanismos de privacidade do Facebook, uma vez que o usuário pode se sentir limitado e, conseqüentemente, desmotivado em aplicar tais configurações em seu perfil [Nielsen, 1994][Nielsen, 1993][Gilutz e Nielsen, 2002].

Levando em consideração o grupo foco desta pesquisa, ao lidar com crianças e adolescentes essas violações se tornam ainda mais grave, pois esse grupo é mais impaciente para lidar com essas dificuldades durante as configurações de privacidade [Fitton et al., 2014][Gilutz e Nielsen, 2002]. Tudo isso diminui as possibilidades desses usuários se protegerem dos perigos que a rede pode oferecer em relação aos outros usuários.

As demais heurísticas, H1 - Visibilidade e *status* do sistema; H4 - Consistência e padrões; H5 - Ajuda aos usuários a reconhecer, diagnosticar e recuperar-se de erros; H6 - Prevenção de erros; e H7 - Reconhecimento ao invés de memorização, apresentaram um percentual de violação inferior a 10%.

5. Apreciação da Privacidade no Facebook sob a Perspectiva de Crianças e Adolescentes

Com o objetivo de investigar a percepção de crianças e adolescentes sobre a privacidade no Facebook e verificar se os problemas de usabilidade levantados na etapa anterior estavam sendo vivenciados por esses usuários, foi realizada uma avaliação por meio de entrevista semiestruturada composta por 30 questões. As questões foram agrupadas de forma que fosse possível identificar: (1) o perfil do participante da pesquisa; (2) detalhes da conta e amigos do Facebook; (3) o conteúdo que é publicado; e (4) conhecer como lidam com a privacidade na rede.

A entrevista ocorreu na segunda quinzena do mês de março de 2015, com crianças e adolescentes brasileiras do estado de Minas Gerais. A mesma foi conduzida por três autores desse trabalho, dois especialistas em IHC e um Terapeuta Ocupacional especializado em psicopedagogia infanto-juvenil. Assim como na avaliação por inspeção (i.e., primeira fase da metodologia desse trabalho), a atuação do especialista em psicopedagogia foi de extrema importância para auxiliar tanto os participantes, quanto os especialistas em IHC na condução e análise dos dados da entrevista, de forma a manter o foco da avaliação na percepção e uso das configurações de privacidade do Facebook por crianças e adolescentes. Ao todo 42 pessoas participaram dessa etapa da pesquisa.

5.1. Perfil dos Participantes

Dos 42 participantes da avaliação, a maioria (75%) foi classificada na faixa etária entre 15 e 16 anos, 9% declararam ter 10 anos, enquanto que 8% têm 11 anos e 8% têm 14 anos. A

divisão de gêneros dos participantes foi 75% do sexo feminino e 25% do sexo masculino. Em relação à formação, constatou-se que a maioria dos participantes (58%) está cursando o ensino médio, 25% concluiu apenas o ensino fundamental e 17% está cursando o nível fundamental. Esses dados indicam que o grupo participante dessa avaliação possui um grau de formação suficiente para compreender e discernir sobre a questão de privacidade no Facebook levantada nesse trabalho.

Os participantes foram questionados sobre as redes sociais online que utilizam atualmente e/ou já utilizaram. A Figura 2 sumariza esses dados e indica o Facebook como a rede social mais utilizada pelos participantes (100%). Em relação ao tempo de utilização do Facebook, constatou-se que todos os participantes possuem conta nessa rede social há pelo menos 2 anos. Quanto à frequência de uso, os dados obtidos mostraram que todos os participantes acessam o Facebook pelo menos duas vezes por semana. Desses, 58% acessam a rede social diariamente, através do uso de dispositivos móveis, por pelo menos 1 hora.

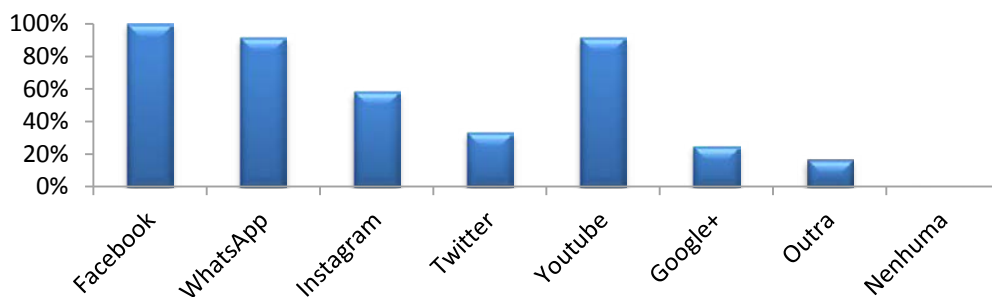


Figura 2. Redes Sociais Online utilizadas

5.2. Contatos e Uso do Facebook

Inicialmente, os participantes foram convidados a responder quantos amigos eles mantêm no Facebook. Os dados obtidos podem ser visualizados na Figura 3, que indicam que a maioria dos participantes dessa pesquisa (67%) possui mais de 500 amigos no Facebook e 33% têm entre 201 a 500 contatos nessa rede social. Ao analisar essa quantidade de amigos distribuída pela idade dos participantes, como demonstrado na Figura 4, é possível perceber, por exemplo, que todos os entrevistados entre 10 e 11 anos possuem de 401 a 500 amigos no Facebook e todos os adolescentes participantes com idade de 14 anos estão conectados a mais de 500 pessoas.

Diante do elevado número de conexões estabelecidas no Facebook, os participantes foram questionados se conheciam pessoalmente todos os amigos que mantêm nessa rede social. 92% afirmaram não conhecer pessoalmente todos os amigos do Facebook. Essa informação chama atenção para o fato de que pessoas mal intencionadas podem se aproveitar dessa situação para cometer crimes, como adquirir dados pessoais para cometer assédios e/ou abusos, ou até mesmo para humilhar ou fazer ameaças (e.g., *cyberbullying*). Essa questão se torna mais evidente quando, desses participantes, 25% afirmam interagir ativamente (e.g., trocam mensagens de texto e fotos), através do Facebook, com esses contatos que eles não conhecem pessoalmente.

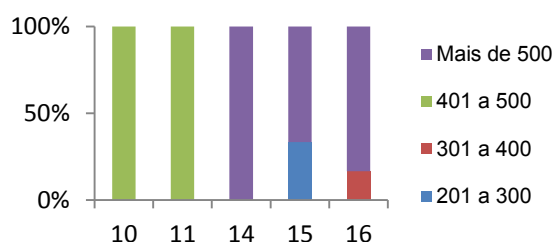
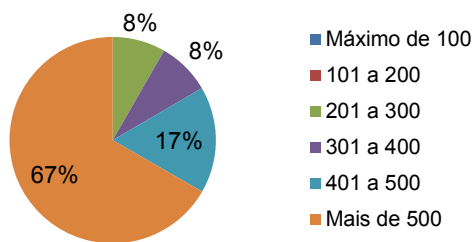


Figura 3. Total de Amigos no Facebook

Figura 4. Idade versus Amigos no Facebook

Os dados apresentados até o momento servem de alerta para a questão de segurança e privacidade dessas crianças e adolescentes na rede. Isso porque um grande número de pessoas pode ter acesso aos dados e informações pessoais desses usuários, que representam um grupo mais frágil e suscetível a ser atraído e enganado por pessoas maliciosas [Boyd et al., 2011][Dey et al., 2012][Dey et al., 2013][Livingstone et al., 2011].

Tal observação reforça a importância de se garantir maior usabilidade para que crianças e adolescentes possam perceber e se sentir motivados a utilizar os mecanismos que permitem configurar privacidade no Facebook, uma vez que esse incentivo pode contribuir para minimizar o risco de exposição desse público.

Os participantes também foram questionados sobre o tipo de conteúdo que publicam no Facebook. Através da Figura 5 é possível constatar que 25% dos entrevistados admitem publicar dados pessoais, como nome, endereço e telefone. Fato que chama atenção para a falta de percepção de algumas crianças e adolescentes para o perigo que podem estar correndo ao expor esses dados, uma vez que os mesmos podem ser utilizados para prejudicá-los.

Além dos dados pessoais, os participantes também admitiram publicar conteúdo que expressam seu estado pessoal (e.g., alegrias, tristezas, conquistas, viagens) e fotos, o que os expõem ainda mais na rede *online*. Ao expor seu estado pessoal para qualquer pessoa, o usuário fica mais propenso a ações de criminosos que podem se aproveitar de um momento de fragilidade para se aproximar. Além disso, o fato de expor fotos pessoais facilita a ação desses criminosos ao visualizar suas vítimas, os lugares que frequentam ou até os bens pessoais que possuem [Boyd et al., 2011][Dey et al., 2012][Dey et al., 2013][Livingstone et al., 2011].

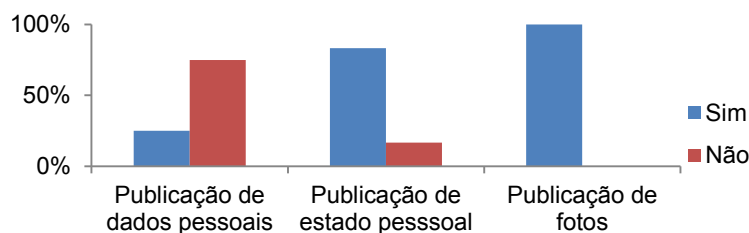


Figura 5. Conteúdos Publicados no Facebook

Esse comportamento se torna mais grave ao analisar a visibilidade dos conteúdos publicados. Verificou-se que embora metade dos participantes limite o acesso às publicações somente a amigos, 42% informaram que a visibilidade de conteúdos publicados

em sua linha de tempo é pública, mesmo o Facebook oferecendo a possibilidade de configurar a visibilidade de cada conteúdo publicado. Além disso, 8% dos participantes disseram não saber qual é visibilidade do conteúdo publicado por eles. Mesmo sendo uma parcela pequena, essa informação é relevante, pois demonstra que ainda existem pessoas, dentro do perfil analisado nesta pesquisa, que não tem consciência sobre a existência do controle de visibilidade de conteúdos publicados. O uso limitado, ou até mesmo o desconhecimento, desse mecanismo de controle pode ser justificado pelo problema de usabilidade exemplificado na avaliação por inspeção (seção 4.1.).

Uma vez analisada a interação com os contatos do Facebook, bem como o tipo e visibilidade do conteúdo publicado pelos participantes, a avaliação com os usuários seguiu de forma a caracterizar: (1) a percepção e uso das configurações de privacidade por crianças e adolescentes e (2) problemas de privacidade já vivenciados por esses usuários. Os resultados são apresentados na próxima subseção.

5.3. Privacidade no Facebook

Inicialmente, os entrevistados foram questionados se estão preocupados com a segurança e privacidade no Facebook e se conhecem a política de privacidade dessa rede social. Dos participantes, 68% afirmaram se preocupar com sua privacidade, porém, 80% disseram não ter conhecimento (i.e., não leram) sobre a política que rege a privacidade no Facebook.

Essa informação pode ajudar a explicar o fato de que menos de 50% dos entrevistados afirmam conhecer e/ou utilizar todas as configurações de privacidade disponíveis no Facebook, conforme demonstrado na Figura 6 que lista os principais controles de privacidade oferecidos pelo Facebook e indica o percentual de participantes dessa pesquisa que conhecem e/ou utilizam cada um desses recursos.

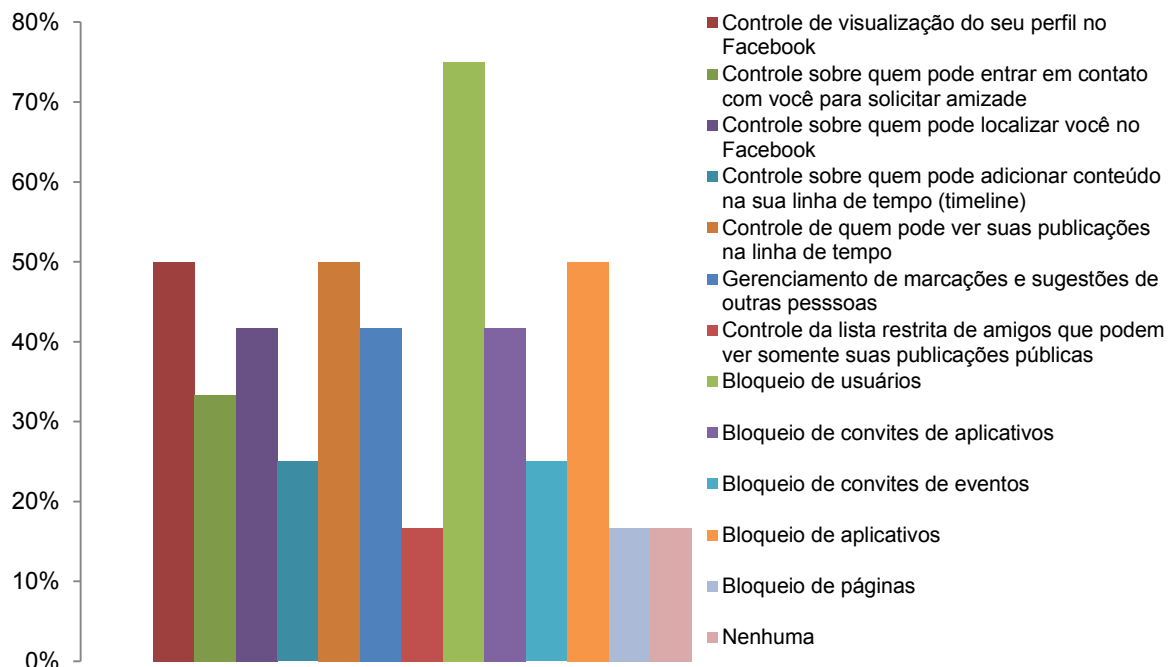


Figura 6. Configurações de Privacidade do Facebook utilizadas

Através desses dados é possível verificar, por exemplo, que a configuração “Controle da lista restrita de amigos que podem ver somente suas publicações e informações públicas”, útil para limitar os conteúdos a um determinado grupo de amigos que não são de total confiança, é conhecida e utilizada por apenas 18% dos entrevistados. Outra configuração que é muito importante é a de “Controle sobre quem pode adicionar conteúdo na sua linha de tempo (timeline)”. Essa também obteve baixo índice de conhecimento, apenas 25%. Essa configuração permite restringir as pessoas que podem adicionar conteúdo na linha de tempo, o que previne o usuário de possíveis transtornos com conteúdos indesejáveis como insultos e ameaças.

Como informado anteriormente, os participantes foram questionados se já vivenciaram problemas de privacidade no Facebook (nesta questão, mais de uma violação poderia ser indicada). Dos entrevistados, como demonstrado na Figura 7, 50% disseram ter sofrido algum tipo de violação de privacidade. Sendo que, 33 % tiveram sua conta invadida e pelos menos 8% admitiram ter sofrido algum tipo de assédio e/ou abuso e problemas com contas falsas no Facebook. Ao contrastar a visibilidade do perfil dos participantes com o tipo de violação de privacidade, é possível concluir, com o auxílio da Figura 8, que as crianças e adolescentes que mantêm um perfil público sofreram diferentes tipos de violações de privacidade.

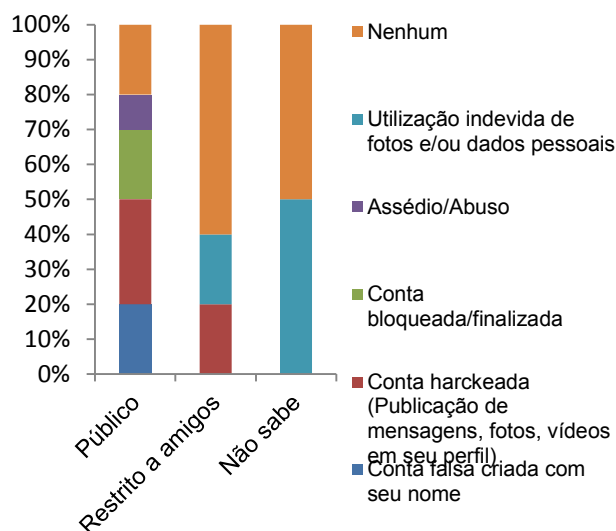
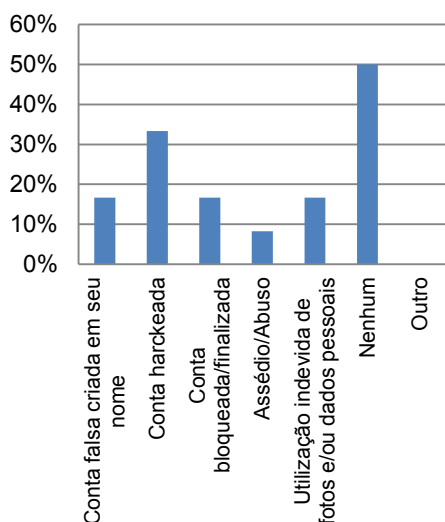


Figura 7. Violações de Privacidade sofridas

Figura 8. Visibilidade versus Violação

Finalizada as análises dessa etapa da metodologia foi possível observar que embora a maioria das crianças e adolescentes, participantes dessa avaliação, afirme preocupar com sua segurança e privacidade no Facebook, menos da metade conhece e/ou utiliza os principais controles disponíveis nessa rede social para se resguardar. Essa falta de conhecimento pode influenciar no uso limitado, ou até mesmo na não utilização, de importantes configurações de privacidade nessa rede social por esse público.

Esse uso limitado é um alerta, porque conforme informado anteriormente, crianças e adolescentes são potencialmente mais vulneráveis nessas redes e o fato de deixarem o perfil público, adicionarem e interagirem com desconhecidos e compartilharem informações

peçoais como dados, fotos e “estados peçoais” aumentam, de forma significativa, as chances de violações de privacidades, como de fato ocorreu com 50% dos participantes.

Os resultados dessa avaliação chamam atenção não somente de pais e responsáveis, em relação aos cuidados que devem ser tomados durante o acesso de seus filhos, mas também dos desenvolvedores de redes sociais *online*, uma vez que a oferta de configurações de privacidade mais claras e adequadas ao uso desse público podem contribuir para aumentar a segurança e privacidade de crianças e adolescentes nessas redes.

6. Triangulação e Discussão dos Resultados

Para verificar como a usabilidade das configurações de privacidade do Facebook tem impactado na segurança de crianças e adolescentes do Brasil, a última etapa dessa pesquisa consistiu em triangular e discutir os resultados obtidos na avaliação de usabilidade por inspeção com os dados coletados durante a entrevista com o público alvo dessa pesquisa.

Os resultados apresentados indicam que, uma possível explicação para que a maioria dos participantes desse estudo de caso desconheça a política de privacidade do Facebook e esteja fazendo um uso limitado dos controles oferecidos por essa rede, é a violação das heurísticas “H2 - Correspondência entre o sistema e o mundo real”, “H4 - Consistência e padronização”, “H7 - Reconhecimento em vez de memorização” e “H10 - Ajuda e Documentação”.

Ao violar H2, H4, H7 e H10, o Facebook não faz uso de elementos de interface compatíveis com a expectativa e experiência dessas crianças e adolescentes, fazendo com que eles encontrem dificuldades em reconhecer e utilizar as opções de configuração de privacidade, bem como em procurar auxílio através dos recursos de ajuda do sistema [Nielsen, 1993][Nielsen, 1994][Gilutz e Nielsen, 2002]. Tais dificuldades podem inviabilizar, ou até mesmo impedir, que as configurações de privacidade sejam aplicadas adequadamente.

Isso porque, uma vez que todas essas heurísticas buscam garantir a facilidade de uso, eficiência, produtividade e satisfação dos usuários, o fato de violá-las contribui para que crianças e adolescentes façam um uso limitado dos mecanismos de privacidade, já que eles não se sentem guiados e motivados em aplicá-las, devido às barreiras (i.e., dificuldades) encontradas na interface disponível para essas configurações.

Logo, essa triangulação evidencia que os problemas de usabilidade refletem no uso que crianças e adolescentes fazem dos recursos de privacidade do Facebook e a pesquisa mostrou que de fato o uso limitado compromete a segurança desses usuários. Isso porque 50% dos entrevistados admitiram ter problemas com violação de privacidade. Sendo assim, a má usabilidade dos recursos para configurar a privacidade aumenta a vulnerabilidade de crianças e adolescentes no Facebook diante de pessoas mal intencionadas.

7. Conclusões e Trabalhos Futuros

Este trabalho buscou apreciar e caracterizar como crianças e adolescentes têm lidado com os controles de privacidade em seus perfis nas redes sociais *online* e se a usabilidade desses controles está adequada a esse perfil de usuários. Para isso foi realizado um estudo de caso

no Facebook, considerando crianças e adolescentes do Brasil, uma vez que esse país é um dos campeões no uso desse sistema que é a rede social mais utilizada no mundo. A análise aqui apresentada é importante porque, embora o uso das redes sociais *online* seja recomendado para maiores de 13 anos, pesquisas indicam que usuários abaixo dessa faixa etária têm utilizado esse tipo de sistema, na maioria das vezes, sem o acompanhamento dos responsáveis, reforçando a necessidade de também adequar os recursos para configuração de privacidade a esse público.

Os principais resultados apresentados e discutidos indicaram como os problemas de usabilidade têm influenciado no grau de conhecimento e uso limitado que crianças e adolescentes do Brasil fazem das configurações de segurança e privacidade no Facebook. Nesse sentido, embora o estudo de caso tenha sido realizado no Facebook, considerando o perfil de crianças e adolescentes do Brasil, os resultados apresentados e discutidos nesse trabalho são relevantes e não se limitam apenas ao contexto desse país. A relevância deste tipo de investigação pode ser sustentada pelo argumento apresentado por Wilson et al. (2009) e Willinger et al. (2010), no qual os autores listam os desafios de pesquisas relacionados a redes sociais *online*.

Segundo Wilson et al. (2009) e Willinger et al. (2010) para que seja possível oferecer soluções que enderecem as necessidades dos diferentes perfis de usuários que utilizam as redes sociais *online*, em todas as situações como, por exemplo, a privacidade, é preciso focar em aspectos dinâmicos e comportamentais durante o uso desse tipo de sistema e, sobretudo, apreciar e caracterizar como os diversos tipos de usuários estão interagindo atualmente, de forma a verificar e endereçar os problemas que podem estar afetando a qualidade dessa interação [Wilson et al., 2009] e [Willinger et al., 2010].

Sendo assim, este trabalho apresenta contribuições práticas e científicas/metodológicas para a área de sistemas colaborativos e afins. Em termos científicos, esse trabalho é relevante porque discute sobre a importância de se propor e avaliar abordagens que visam o projeto e avaliação de interfaces para configurações de privacidade voltadas para crianças e adolescentes, reforçando a relevância de iniciativas como: (1) *User Experience* (UX) para adolescentes [Fitton et al., 2014]; (2) Diretrizes de usabilidade para crianças [Gilutz and Nielsen, 2002]; e (3) *Design* de interface e interação para crianças [Melo and Baranauskas, 2003] no âmbito de sistemas colaborativos. Contudo, vale destacar que a relevância do trabalho não se limita a essa contribuição para as pesquisas na área, isso porque, a metodologia adotada neste estudo de caso pode ser reproduzida na avaliação de outras redes sociais, em relação à usabilidade dos recursos de privacidade, considerando outros perfis de usuários.

Já em termos práticos, o trabalho alerta os projetistas de interface sobre a importância da usabilidade nos controles de privacidade de redes sociais online. De forma complementar, os resultados reportados podem servir de parâmetros para a melhoria e/ou criação de mecanismos de privacidade nesse ambiente virtual, com maior usabilidade, para crianças e adolescentes.

Nesse sentido, como trabalhos futuros propõe-se a criação de um *design* com melhorias e/ou um conjunto de melhores práticas que descrevam os recursos de privacidade

mais adequados para crianças e adolescentes. Além disso, outra direção futura consiste na avaliação dos mecanismos de privacidade em redes sociais *online* sob as perspectivas de outros perfis, como, por exemplo, pessoas com necessidades especiais e idosas.

Referências

- Albesher, A. and Alhussain, T. 2013. Privacy and security issues in social networks: an evaluation of Facebook. In Proceedings of the 2013 International Conference on Information Systems and Design of Communication (ISDOC '13). ACM, New York, NY, USA, 7-10.
- Balmant, O. 2012. Cresce exposição de jovens na internet. O Estado de São Paulo, Caderno Vida, p.27, 02 dez 2012. Estadão Geral
- Boyd, D., Hargittai, E., Schultz, J. and Palfrey, J. 2011. Why Parents Help Their Children Lie to Facebook: Unintended Consequences of the 'Children's Online Privacy Protection Act'. First Monday, 16(11), November 7, 2011.
- Chilana, P., Holsberry, C., Oliveira, F. and Ko, A. 2012. Designing for a billion users: a case study of facebook. In CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12). ACM, New York, NY, USA, 419-432.
- Cho, J. and Trent, A. (2006). Validity in qualitative research revisited. Qualitative Research, 6(3):319--340.
- comScore. 2015. Brazil Digital Future in Focus. Available in: < <https://goo.gl/H51ep1> >. Last access: Dec – 18, 2015.
- Consumer Reports News. 2013. Facebook may let children under age 13 use the site. Available in: <<http://goo.gl/tgqbI7>> Last access: Jun – 07, 2015.
- da Silva, S. R.; e Pereira, R. (2008). Aspectos da interação humano-computador na web social. In Proceedings of the IX IHC 2008, Outubro de 2008, pp.350-553.
- Dey, R., Jelveh, Z. and Ross, R. 2012. Facebook users have become much more private: A large-scale study. In Proceedings of the 4th IEEE International Workshop on Security and Social Networking. IEEE, pp. 346-352.
- Dey, R., Tang, C., Ross, K. W. and Saxena, N. 2012. Estimating age privacy leakage in online social networks. In Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, pages 2836--2840, 2012.
- Dey, R., Ding, Y. and Ross, K. W. 2013. Profiling high-school students with facebook: how online privacy laws can actually increase minors' risk. In Proceedings of the 2013 conference on Internet measurement conference (IMC '13). ACM, New York, NY, USA, 405-416.
- Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A. and Mary Madden, M. 2015. Social Media Update 2014. In Pew Internet and American Life Project. Available in: <<http://www.pewinternet.org/2015/01/09/social-media-update-2014/>>, Last access: Dec – 18, 2015.

- Duggan, M., and Page, D. 2015. Mobile Messaging and Social Media 2015. In Pew Research Center. Available in: <<http://www.pewinternet.org/2015/01/09/social-media-update-2014/>>, Last access: Dec – 18, 2015.
- Fitton, D., Bell, B., Read, J. C., Iversen, O., Little, L. and Horton, M. 2014. Understanding teen UX: building a bridge to the future. In CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14). ACM, New York, NY, USA, 79-82.
- Gill, A. J., Vasalou, A., Papoutsis, C., and Joinson, A. (2011) "Privacy Dictionary: A Linguistic Taxonomy of Privacy for Content Analysis". In Proceedings of CHI 2011, Session: Privacy May 7-12, 2011, Vancouver, BC, Canada.
- Gilutz, S. and Nielsen, J. 2002. Usability of Websites for Children: Children (Ages 3-12) on the Web - 70 Design Guidelines. 3rd Edition. Nielsen Norman Group. Available in: <www.nngroup.com/reports/kids/>. Last access: Dec – 29, 2015.
- Hinduja, S., and Patchin, J. W. 2008. Personal information of adolescents on the Internet: A quantitative content analysis of MySpace. *Journal of Adolescence*, 31(1), 125-146
- Junior, M. P., Xavier, S. and Prates, R. O. 2014. Investigating the use of a Simulator to Support Users in Anticipating Impact of Privacy Settings in Facebook. In Proceedings of the 18th ACM International Conference on Supporting Group Work, ACM, 63-72.
- Krasnova, H., and Veltri, N.F. 2010. "Privacy Calculus on Social Networking Sites: Explorative Evidence from Germany and USA", Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS), IEEE, 2010, pp. 1-10
- Liu, Y., Gummadi, K. P. and Mislove, A. 2011. Analyzing Facebook Privacy Settings: User Expectations vs Reality. In Proceedings of Internet measurement conference, ACM, 61-70.
- Livingstone, S., Ólafsson, K. and Staksrud, E. 2011. Social networking, age and privacy. EU Kids Online, London, UK. Available in: <<http://eprints.lse.ac.uk/35849/>>. Last access: Dec – 29, 2015.
- Melo, A. M. and Baranauskas, M. C. 2003. Design with children: a Semiotic approach. In Proceedings of the Latin American conference on Human-computer interaction (CLIHC '03). ACM, New York, NY, USA, 69-78.
- Nielsen., J. Usability engineering. San Francisco: Morgan Kauffman, 1993.
- Nielsen., J. 1994. Usability inspection methods. In Conference Companion on Human Factors in Computing Systems (CHI '94), Catherine Plaisant (Ed.). ACM, New York, NY, USA, 413-414.
- Nielsen., J. 2000. Why you only need to test with 5 users. 2000. Available in: <<https://goo.gl/qik9ZN>>. Last access: Dec – 29, 2015.
- O TEMPO (Jornal). 2014. Pedófilo se passa por atriz de Chiquititas para abordar crianças. . Available in: <<http://goo.gl/nXGJ4g>>. Last access: Dec – 29, 2015.

- Pereira Junior, M., Xavier, S. I. de R. and Prates, R. O. 2014. Investigating the Use of a Simulator to Support Users in Anticipating Impact of Privacy Settings in Facebook. In Proceedings of the 18th International Conference on Supporting Group Work (GROUP '14). ACM, New York, NY, USA, 63-72.
- Pesce, J. P., Casas, D. L., Rauber, G., and Almeida, V. 2012. Privacy attacks in social media using photo tagging networks: a case study with Facebook. In Proceedings of the 1st Workshop on Privacy and Security in Online Social Media (PSOSM '12). ACM, New York, NY, USA, Article 4 , 8 pages.
- Rodrigues, K. R. H., Canal, M. C., Xavier, R. A. C., Alencar, T. S., and Neris, V. P. A. 2012. Avaliando aspectos de privacidade no Facebook pelas lentes de usabilidade, acessibilidade e fatores emocionais. In Companion Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems (IHC '12).
- Souza, Z. D., and Dick, G. N. 2009. Disclosure of information by children in social networking-Not just a case of "you show me yours and I'll show you mine". *Int. J. Inf. Manag.* 29, 4 (August 2009), 255-261.
- Tsoi, H. K., and Chen, L. 2011. "From Privacy Concern to Uses of Social Network Sites: A Cultural Comparison via User Survey. 2011. In Proceedings of the IEEE Third International Conference on Social Computing (SocialCom), pp.457-464, 9-11 Oct. 2011.
- Villela, M. L. B., and Prates, R. O. 2015. Supporting Designers in Modeling Privacy for Social Network Sites. In Proceedings of 15th Brazilian Symposium on Human Factors in Computing Systems. (IHC'15). November 03-06, 2015.
- Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. N., Zhao, B. Y. 2009. User interactions in social networks and their implications. In Proceedings of the 4th ACM European conference on Computer systems (EuroSys '09). ACM, New York, NY, USA, PP. 205—218 (2009).
- Willinger, W., Rejaie, R., Torkjazi M., Valafar, M., Maggioni, M. 2010. Research on online social networks: time to face the real challenges. *SIGMETRICS Perform. Eval. Rev.* 37, 3, pp. 49--54 (2010).

PryMeVis: Uma ferramenta para Modelagem de Design de Privacidade

Maria Lúcia Bento Villela^{1,2}, Lidia Silva Ferreira¹,
Diego Augusto de Faria Barros¹, Raquel Oliveira Prates¹,
Raquel Cardoso de Melo Minardi¹

¹ DCC – ICEx, Universidade Federal de Minas Gerais, Belo Horizonte – MG, Brasil

²DECOM – FACET, UFVJM, Diamantina – MG, Brasil

{mvillela, lidiaferreira, diego.barros, rprates, raquelcm}@dcc.ufmg.br

Abstract. *The Privacy Design Model (PDM) is a descriptive model that aims at helping designers to model personal information disclosure in Social Network Sites (SNSs). This model takes into account different aspects that influence the level of privacy being offered to users. In this paper we propose PryMeVis - an online tool that allows the creation of visual representations for the models of personal information disclosure in SNSs created using PDM. A preliminary evaluation was conducted and presented, and the results indicate that it can be an important tool to support designers in using PDM to model privacy in SNSs.*

Resumo. *O Modelo de Design de Privacidade (MDP) é um modelo descritivo que tem o propósito de ajudar designers a modelarem o compartilhamento de informações pessoais em Redes Sociais Online (RSOs), considerando diferentes aspectos que influenciam o nível de privacidade a ser oferecido ao usuário. Neste artigo, propomos a PryMeVis - uma ferramenta online que permite a criação de visualizações correspondentes aos modelos de compartilhamento de informações pessoais em RSOs, criados com o uso do MDP. Além disso, apresentamos também uma avaliação preliminar de tal ferramenta, com os resultados indicando que a mesma pode ser importante para apoiar designers no uso do MDP para a modelagem de privacidade em RSOs.*

1. Introdução

A Internet, com mais de três bilhões de usuários¹, tem mudado drasticamente a forma como as pessoas se relacionam socialmente, uma vez que vem sendo cada vez mais utilizada para apoiar a interação social entre seus usuários, através das Redes Sociais Online (RSOs). Entretanto, enquanto o uso desses sistemas cresce, novas preocupações relacionadas à privacidade surgem [Acquisti et al. 2015], uma vez que favorecem o compartilhamento de quantidades crescentes de informações pessoais como parte de sua funcionalidade, se apresentando como uma ameaça à privacidade de seus usuários.

Com isso, estudos que buscam fornecer um melhor entendimento sobre a relação entre privacidade e RSOs têm sido conduzidos. Também foi identificada a necessidade de ferramentas para apoiarem designers na tomada de decisões sobre aspectos de privacidade

¹Dados de Internet Live Stats em março de 2016 (<http://www.internetlivestats.com/internet-users/>)

que devem ser considerados nesses sistemas. Nesse sentido, [Villela and Prates 2015] propôs o Modelo de Design de Privacidade (MDP), que consiste em uma ferramenta epistêmica² para apoiar o designer na elaboração e avaliação do discurso designer-usuário em RSOs, com foco na privacidade relacionada ao compartilhamento de informações pessoais. Como uma ferramenta epistêmica, o MDP não se propõe a fornecer diretamente uma solução para tratar questões de privacidade em RSOs, e sim a aumentar o entendimento do designer sobre o problema e suas implicações, a fim de permitir-lhe gerar soluções alternativas e compará-las entre si.

Existe porém, um número considerável de decisões a serem tomadas durante o design do compartilhamento de informações pessoais em uma RSO usando o MDP. Além disso, uma vez que o designer toma todas as decisões necessárias, pode ser difícil ter uma visão geral dos possíveis estados de privacidade definidos como consequência de suas decisões. A fim de facilitar a visão geral das decisões do designer ao fazerem uso do MDP, bem como dos possíveis estados de privacidade resultantes, [Villela 2016] propôs uma representação visual para o modelo. Assim, o objetivo deste trabalho é apresentar a ferramenta PryMeVis (acrônimo para *PRivacY design model Visualization*), que se propõe a apoiar o uso do MDP, ao fazer uso da sua representação visual para a confecção de modelos de compartilhamento de informações pessoais em RSOs. Além disso, com o objetivo de verificar o valor e a utilidade dessa representação visual, a mesma foi avaliada, juntamente com a versão inicial da ferramenta PryMeVis, com designers de sistemas, que consistem em seus potenciais usuários. Foi verificado, a partir dessas avaliações, que o uso da ferramenta de visualização PryMeVis enriquece a experiência de designers de RSOs, ao projetarem o compartilhamento de informações pessoais nesses sistemas e visualizarem o impacto de suas decisões na privacidade dos usuários.

Na próxima seção, apresentamos os trabalhos relacionados a este estudo. Na seção 3, são apresentados o Modelo de Design de Privacidade (MDP), suas dimensões e sua representação visual. Na seção 4, apresentamos a ferramenta PryMeVis, seguida de uma avaliação preliminar da representação visual do MDP e da primeira versão da ferramenta, na Seção 5. Por fim, na Seção 6, são apresentadas as considerações finais deste trabalho, bem como os trabalhos futuros que podem ser explorados a partir do mesmo.

2. Trabalhos Relacionados

A maior parte das pesquisas sobre privacidade em RSOs consiste em estudos empíricos e tem se dedicado principalmente à compreensão de diferentes aspectos relacionados à privacidade no compartilhamento de informação pessoal nesses ambientes [Stutzman and Hartzog 2012, Bevan et al. 2015]. Por outro lado, poucos estudos têm contribuído com modelos teóricos ou conceituais para tratar privacidade no design de tecnologias online [Belanger and Crossler 2011]. Nesse sentido, algumas ferramentas conceituais têm sido propostas com a finalidade de aumentar a flexibilidade que RSOs oferecem ao seus usuários, a fim de lidarem com aspectos específicos de privacidade [Pang and Zhang 2015, Tierney and Subramanian 2014], como o controle de acesso e o gerenciamento de compartilhamento .

Com o objetivo de apoiar designers em suas decisões sobre aspectos de privaci-

²Ferramenta que apoia o designer em sua reflexão sobre o design que está sendo feito, ajudando-o a obter novo conhecimento ou habilidades [de Souza 2005].

dade a serem considerados no design de RSOs, modelos conceituais são propostos, como o de [Romero et al. 2013] e o de [Epstein et al. 2015]. O primeiro consiste em um modelo que visa apoiar o design de mecanismos que permitem ao indivíduo coordenar sua privacidade enquanto comunica com outro usuário em um canal exclusivo, em RSOs, considerando especificamente sua disponibilidade para comunicação. O segundo consiste em um *framework* voltado para o projeto e avaliação de aspectos relacionados ao compartilhamento de dados capturados automaticamente nesses sistemas, com foco específico nas respostas e reações da audiência à informação compartilhada. Esses modelos se assemelham à proposta do MDP, no sentido de apoiar as decisões do designer sobre como projetar aspectos específicos do sistema, considerando a privacidade dos seus usuários. No entanto, os focos desses modelos, além de serem distintos entre si, são distintos do foco do MDP, que se propõe a apoiar designers a refletirem sobre diferentes aspectos relacionados ao compartilhamento de informações pessoais em RSOs, considerando o impacto de suas decisões nos níveis de privacidade oferecidos aos usuários pelo sistema.

Um ponto que merece ser destacado é que esses modelos conceituais para o apoio do design de privacidade não oferecem quaisquer tipos de ferramentas que ajudem na sua utilização por parte de designers, possibilitando a visualização dos modelos de design criados. As ferramentas de visualização existentes, relacionadas a privacidade em RSOs, são voltadas basicamente para o usuário final, com o intuito de permiti-lo gerenciar a sua privacidade ou aumentar a sua consciência sobre a sua exposição nesses sistemas. Contribuindo para o gerenciamento de privacidade por parte do usuário final, [Gao and Berendt 2013] apresenta uma ferramenta que possibilita aos usuários manusearem de forma eficiente os seus círculos de amizade, ao permiti-los colocar seus contatos em grupos, a fim de gerenciar o que será compartilhado com os mesmos. [Paul et al. 2012], por sua vez, desenvolveram uma interface melhorada para as configurações de privacidade do Facebook, possuindo como principal recurso a aplicação de um código de cores para grupos diferentes de audiências, com o objetivo de melhorar o acesso e a usabilidade oferecida. Visando aumentar a consciência sobre a exposição dos usuários em RSOs, [Mazzia et al. 2012] apresentam uma ferramenta que permite ao usuário entender a visibilidade de seu perfil nesses sistemas. Também dentro desta categoria estão as ferramentas que simulam efeitos de configuração, possibilitando ao usuários prever as consequências de suas escolhas [Pereira Junior et al. 2014, Malandrino et al. 2013]. Nessa mesma linha, [Schlegel et al. 2011] propuseram e avaliaram um mecanismo intuitivo para sumarizar e controlar a exposição de usuários em plataformas móveis. Por fim, [Wang et al. 2015] apresentam uma ferramenta de visualização interativa que ajuda os usuários a compreenderem e explorarem seus próprios traços de personalidade, derivados de seus dados de mídia social, e configurarem suas preferências de privacidade.

Como podemos perceber, embora tais ferramentas consistam em propostas para o tratamento de aspectos específicos de privacidade na interface de RSOs, seus focos são distintos do que estamos propondo nesta pesquisa, um vez que elas são voltadas para o usuário final e não para o designer, no sentido de ajudá-lo a tomar decisões, em tempo de design, sobre como tratar questões de privacidade na interface desses sistemas.

3. Modelo de Privacidade para Design (MDP)

O MDP [Villela and Prates 2015, Villela 2016] é um modelo descritivo que considera o compartilhamento de informações pessoais em RSOs como uma comunicação entre usuá-

rios mediada pelo sistema. Tal comunicação pode ocorrer tanto na forma direta (quando o próprio indivíduo³ compartilha informações sobre ele no sistema), quanto na forma indireta (quando outro usuário compartilha informações sobre o indivíduo). Além da informação sobre o indivíduo que é explicitamente compartilhada, o MDP também considera como informação pessoal seus discursos e atividades dentro do sistema, uma vez que podem levar à inferência de características e atributos pessoais, que impactam a sua privacidade [Kosinski et al. 2013]. Além disso, o MDP considera privacidade no contexto “um-para-muitos” das RSOs, em que as informações pessoais dos seus usuários são compartilhadas em espaços acessíveis a um grupo limitado ou ilimitado de pessoas. Assim, o compartilhamento de informações em interações envolvendo duas ou mais pessoas que se comunicam exclusivamente entre si, em um canal exclusivo, não é tratado pelo MDP.

O MDP é estruturado por meio de *dimensões de privacidade* que descrevem diferentes aspectos relacionados à privacidade no compartilhamento de informações pessoais em RSOs, conforme mostrado na Figura 1. Esses são aspectos sobre os quais o designer deve pensar, e que podem impactar o estado de privacidade dos usuários de RSOs. Para cada uma dessas dimensões, um conjunto de possíveis valores é definido, remetendo a níveis distintos de privacidade para o indivíduo.

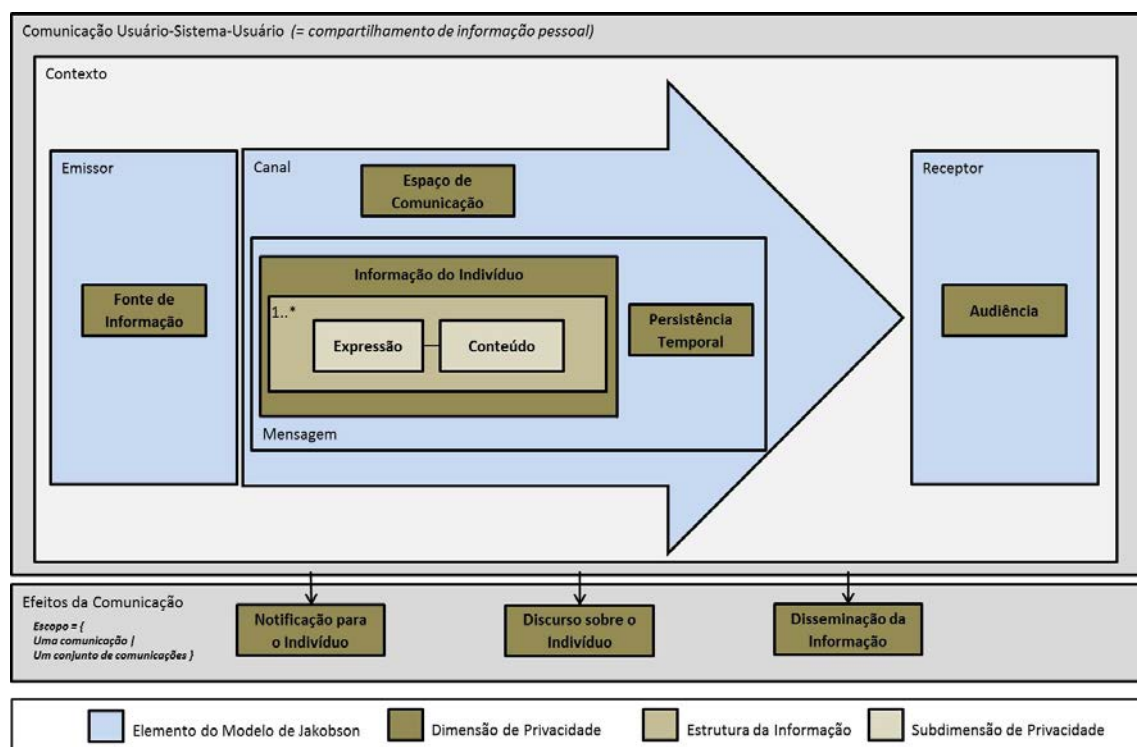


Figura 1. O modelo de design de privacidade (MDP)

O MDP considera os elementos **contexto**, **controle** e **estado de privacidade** relativos à privacidade no compartilhamento de informações pessoais. O **contexto** está relacionado a normas culturais, éticas e morais que governam compartilhamento de informações pessoais em RSOs e funciona como um pano-de-fundo, tendo em vista que as

³Usamos o termo “indivíduo” para nos referirmos ao usuário cuja informação está sendo compartilhada dentro da RSO.

decisões do designer sobre como tratar as dimensões de privacidade, no que se refere à forma como serão atribuídos valores às mesmas, são baseadas no contexto no qual o sistema é projetado. O elemento **controle** está relacionado a quem possui o poder de decisão sobre os valores a serem atribuídos a cada uma das dimensões de privacidade. Isso significa que, em tempo de design, o designer deve decidir, para cada uma das dimensões, a quem deverá ser concedido o controle sobre o valor da mesma: ao *usuário* ou ao *sistema*. O **estado de privacidade** é obtido a partir dos níveis de privacidade relacionados a cada uma das dimensões do MDP. Assim, a combinação de valores atribuídos a cada uma das dimensões de privacidade, em tempo de design ou em tempo de uso, considerando diferentes oportunidades de compartilhamento de informação pessoal, pode levar o indivíduo a atingir diferentes estados de privacidade dentro do sistema. Dessa forma, as decisões do designer em relação ao controle e valor para as dimensões de privacidade determinam os possíveis estados de privacidade que o usuário pode atingir durante o uso do sistema.

3.1. Dimensões de Privacidade

As dimensões de privacidade do MDP foram definidas com base na literatura e em estudos empíricos sobre o compartilhamento de informação e privacidade em RSOs⁴. Essas dimensões são estruturadas em dois níveis: *comunicação usuário-sistema-usuário* e *efeitos da comunicação*, que serão explicados a seguir.

3.1.1. Dimensões de Comunicação Usuário-Sistema-Usuário

A comunicação usuário-sistema-usuário diz respeito à comunicação entre usuários que ocorre através do sistema, com o seu conteúdo consistindo de informações pessoais. As dimensões deste nível modelam os elementos que constituem o compartilhamento de informação pessoal em RSOs, e serão detalhadas a seguir.

Fonte de Informação: Esta dimensão diz respeito a quem pode determinar como, quando e em que extensão a informação pessoal do indivíduo será compartilhada no sistema. Os possíveis valores dessa dimensão são “*indivíduo*”, quando o indivíduo explicitamente compartilha informações sobre ele com outros usuários, ou executa ações que indiretamente podem caracterizá-lo, ao revelar suas opiniões, pontos de vista ou mesmo traços de sua personalidade dentro do sistema; e “*outro usuário*”, sendo que, neste caso, o indivíduo não possui autonomia sobre a sua informação que é compartilhada dentro do sistema, remetendo assim a um nível mais baixo de privacidade.

Espaço de Comunicação: Esta dimensão se refere ao local onde a informação sobre o indivíduo é compartilhada dentro do sistema. O compartilhamento de informações pessoais pode ocorrer em um dos seguintes espaços: “*espaço de perfil do indivíduo*”, que refere-se ao espaço onde são compartilhadas informações mais estáticas, relacionadas a elementos de identidade do indivíduo dentro do sistema; “*espaço de publicação do indivíduo*”, que refere-se ao local onde são compartilhadas informações mais dinâmicas, que refletem situações ou objetos que podem sofrer frequentes atualizações, e sobre as quais outros usuários costumam poder interagir; “*espaço de outro usuário*”, que refere-se a um espaço que pertence a outro usuário do sistema, onde também podem ser compartilhadas

⁴As explicações sobre como foram geradas as dimensões do MDP estão apresentadas em [Vilella and Prates 2015]

informações pessoais do indivíduo; e “*espaço público*”, que diz respeito a um espaço que não pertence especificamente a nenhum usuário e pode ser acessado por todos os usuários e, em alguns casos, também por pessoas fora do sistema.

Informação do Indivíduo: Esta dimensão diz respeito à informação que é compartilhada no sistema e é composta pelas subdimensões **expressão** e **conteúdo**. A subdimensão **expressão** refere-se à forma como a informação é expressa no sistema. O MDP considera que informações podem ser expressas nos formatos: “*predefinido*”, que é o formato das informações que definidas pelo sistema, em tempo de design, cabendo ao usuário apenas decidir se irá compartilhá-las ou não, ao executar determinadas ações dentro do sistema; “*tipado*”, que é o formato das informações que possuem o seu significado definido pelo sistema, em tempo de design, mas são fornecidas pelos usuários, em tempo de uso; e “*livre*”, que é o formato da informação que tem o seu significado definido pelo próprio usuário, em tempo de uso, a partir do seu conteúdo. A subdimensão **conteúdo** refere-se ao nível de pessoalidade da informação sobre o indivíduo que está sendo compartilhada, que diz respeito a quão pessoal ela é [Villela et al. 2015]. Dessa forma, a subdimensão **conteúdo**, pode assumir os seguintes valores no MDP: “*pouco pessoal*”, “*levemente pessoal*”, “*pessoal*”, “*um tanto pessoal*” e “*muito pessoal*”.

Persistência Temporal: Diz respeito ao período de tempo durante o qual a informação sobre o indivíduo fica acessível à sua audiência no sistema. No MDP, a persistência temporal da informação pessoal do indivíduo pode assumir os seguintes valores: “*instantânea*”, quando a informação sobre o indivíduo fica disponível apenas para a audiência que está acessando o sistema no momento em que a mesma é compartilhada; “*limitada*”, quando a informação fica disponível para a audiência por um período limitado (e geralmente curto) de tempo; “*ilimitada em uma direção*”, quando a informação fica visível para a sua audiência por um período ilimitado de tempo, a partir do momento em que é disponibilizada (tempo presente) e prossegue em direção ao passado ou ao futuro; e, por fim, “*permanente*”, quando a informação compartilhada fica sempre acessível à sua audiência, enquanto ela não for excluída pelo usuário.

Audiência: Esta dimensão se refere a quem terá acesso à informação, sobre o indivíduo, compartilhada no sistema. O valor para a dimensão pode ser: “*indivíduo*”, indicando que tem-se um nível máximo de privacidade por um lado, mas um nível mínimo de interação, tendo em vista que ninguém, além do próprio indivíduo, será capaz de acessar a informação; “*selecionada*”, quando o usuário decide quem serão os usuários que farão parte do grupo que irá formar a audiência da informação; “*limitada*”, quando abrange todo o conjunto de usuários do sistema; “*ilimitada*”, quando abrange, além dos usuários do sistema, outras pessoas externas ao mesmo; e “*desconhecida*”, quando o indivíduo ao qual a informação se refere não tem conhecimento de quem poderá ter acesso à mesma.

3.1.2. Dimensões de Efeitos da Comunicação

Os efeitos da comunicação consistem de informações que podem ser geradas a partir da comunicação usuário-sistema-usuário, referente ao compartilhamento de informações pessoais do indivíduo, e que também estão relacionadas a questões de privacidade. A seguir, serão detalhadas as dimensões desse nível.

Notificação para o indivíduo: Esta dimensão diz respeito ao sistema informar adequadamente ao indivíduo quando a informação sobre ele é divulgada ou acessada por outros usuários, e de que forma isso acontece. Os seus valores podem ser: *"completa"*, caso o indivíduo receba, para cada informação pessoal sua que é compartilhada, um completo relatório sobre as interações de outros usuários com a mesma; *"parcial"*, caso o indivíduo seja informado apenas sobre uma parte das interações de outros usuários com sua informação; e *"ausente"*, quando o sistema não fornece ao indivíduo nenhuma informação sobre as interações de outros usuários com sua informação.

Discurso sobre o indivíduo: Dimensão relacionada ao sistema tomar a iniciativa de gerar compartilhamentos de informações do indivíduo. Remetendo a níveis decrescentes de privacidade, os valores para esta dimensão podem ser: *"ausente"*, se o sistema não gera qualquer comunicação a partir de um ou mais compartilhamentos de informações pessoais do indivíduo; *"destaque"*, quando o sistema apresenta informações sobre o indivíduo às quais os usuários já possuem acesso, porém de forma destacada, trazendo-as à atenção dos usuários; e *"novo"*, caso o sistema processe uma ou mais informações pessoais do indivíduo que são compartilhadas, gerando nova informação sobre o mesmo.

Disseminação da Informação: Esta dimensão está relacionada à audiência ser capaz de (re)compartilhar informação pessoal do indivíduo, referindo-se à possibilidade de uma informação pessoal do indivíduo se espalhar através da rede. Tal fato pode gerar problemas de privacidade, devido ao alcance inesperado, por parte do indivíduo, de suas informações [Pereira Junior et al. 2014]. Os valores que essa dimensão pode assumir são: *"ausente"*, quando não é permitida à audiência da informação recompartilhá-la com outras pessoas; *"limitada"*, quando a informação compartilhada sobre o indivíduo pode ser propagada de uma maneira restrita, apenas para uma audiência adicional limitada; e *"ilimitada"*, quando a informação sobre o indivíduo pode ser recompartilhada indistintamente pela audiência, sem nenhuma restrição.

3.2. Representação Visual do MDP

Com o objetivo de auxiliar designers na utilização do MDP, permitindo-os visualizar o impacto de suas decisões no estado de privacidade que o sistema oferece aos seus usuários, foi proposta uma representação visual para o mesmo. Nessa representação visual, cada oportunidade de compartilhamento de informação pessoal no sistema é caracterizada como um "tipo de comunicação" e é representada por um *"honeycomb"*. Esse *"honeycomb"* é formado por hexágonos que representam cada uma das dimensões de privacidade do MDP, cuja disposição remete ao seu posicionamento na estrutura do MDP, como mostrado na Figura 2. No centro do *"honeycomb"* é colocada a descrição referente ao tipo de comunicação que o mesmo representa.

Para cada uma das dimensões de privacidade representadas no *"honeycomb"*, os valores que a mesma pode assumir, em tempo de design, estão associados a tons de uma determinada cor. Na representação em papel, foram utilizados tons de cinza. Assim, quanto mais escuro for o tom de cinza, maior é o nível de privacidade associado ao valor da dimensão. Caso o valor para a dimensão seja definido em tempo de uso, o hexágono correspondente fica sem preenchimento (equivalente à cor branca na representação) e o conjunto de valores que poderão ser atribuídos à mesma são listados em formato textual.

Como o MDP é qualitativo, as suas dimensões possuem diferentes gradações de



Figura 2. Disposição das dimensões de privacidade do MDP na representação visual de um tipo de comunicação

valores. Assim, algumas dimensões vão apresentar possibilidades de valores/cores que remetem a todos os níveis de privacidade, enquanto outras vão apresentar possibilidades de valores que remetem a apenas alguns desses níveis. Nesse caso, é feito um mapeamento valor-cor, de forma que o valor que remete ao nível mais alto de privacidade dentro da dimensão é sempre mapeado na tonalidade mais escura da cor e aquele que remete ao nível mínimo de privacidade é mapeado na tonalidade mais clara. Nesse caso, o valor referente ao nível intermediário de privacidade para a dimensão em questão será mapeado na tonalidade intermediária.

Neste ponto, vale salientar que a dimensão **informação do indivíduo** é representada no “*honeycomb*” apenas através de sua subdimensão **conteúdo**, tendo em vista que são os seus valores que estão associados a níveis de privacidade dentro da dimensão. Assim, na representação visual do MDP, iremos nos referir a tal dimensão como **conteúdo da informação**.

A borda dos hexágonos, na representação visual do MDP, representa o momento em que a definição de valor para a dimensão ocorre, se em tempo de uso ou em tempo de design. Neste último caso, a borda também caracteriza outros aspectos específicos em relação ao escopo do valor da dimensão (nível de sistema ou nível de tipo de comunicação). Além disso, a borda representa também quem tem o controle sobre a dimensão, ou seja, quem pode definir os seus valores, como mostrado na Figura 3.

	Borda	Controle
Dimensão cujo valor é definido em tempo de design (nível de sistema ou de tipo de comunicação)		Sistema – valor fixo no sistema para todos os tipos de comunicação
		Sistema – valor fixo no tipo de comunicação
Dimensão cujo valor é definido em tempo de uso (nível de instância de comunicação)		Sistema
		Indivíduo
		Outro usuário
		Definido em tempo de uso

Figura 3. Bordas representando o momento em que o valor para a dimensão é definido e quem tem controle sobre ela

Considerando os tipos de comunicação referentes a todas as oportunidades de compartilhamento de informação pessoal dentro de um sistema, a visualização fornece ao designer uma ideia geral das diferenças nos níveis de privacidade oferecidos por cada tipo de comunicação, bem como do estado de privacidade que os usuários podem alcançar no sistema. Portanto, se existe um grande número de tipos de comunicação no sistema, isso provavelmente significa uma maior complexidade no sentido de tratar questões de privacidade. Além disso, caso os tipos de comunicação se diferenciem em relação aos seus níveis de privacidade, isso pode ser uma indicação de que designers deveriam deixar claro para os usuários essas diferenças e o contexto em que elas se aplicam.

4. PryMeVis

Com o objetivo de facilitar a criação e avaliação da representação visual e do próprio MDP, desenvolvemos a PryMeVis, uma ferramenta para modelagem e criação de uma representação visual online do MDP⁵. Nessa ferramenta, o designer pode modelar cada tipo de comunicação referente a uma oportunidade de compartilhamento de informação pessoal dentro da RSO, fornecendo valores para os atributos “controle” e “valor” de cada dimensão do MDP. Com base nos valores fornecidos através da interação direta com a interface, PryMeVis representa visualmente os tipos de comunicação, como mostrado na Figura 4. A única diferença entre a representação visual apresentada na seção anterior e a utilizada na PryMeVis é o uso da paleta sequencial⁶ de cores com diferentes intensidades em tons de azul, ao invés dos tons de cinza, que ocorre puramente por questão estética. Também foi utilizado o padrão *fade* (em tom de cinza) para representar o valor “não aplicável” na visualização.



Figura 4. Protótipo da ferramenta de visualização do MDP: versão inicial

⁵Disponível em pensi.dcc.ufmg.br/applications/

⁶Paleta sequencial ColorBrewer: colorbrewer2.org

Inicialmente, foi implementada uma versão online simplificada da ferramenta, fazendo uso da biblioteca Java Script D3.js (que manipula documentos baseados em dados) e do framework Bootstrap⁷. O design foi pensado para ser limpo, responsivo e focado na visualização, priorizando sua disposição na interface e exibindo em uma mesma tela todos os campos que devem ser preenchidos para a construção da visualização, a fim de facilitar a interação e compreensão da modelagem. Isso implicou no fato de o designer poder modelar apenas um tipo de comunicação por vez. Uma limitação da versão inicial da ferramenta é a impossibilidade de armazenar a visualização gerada, tendo o usuário que tirar um *screen shot* da mesma se quiser guardá-la. Entretanto, apesar das limitações, os resultados obtidos com a avaliação PryMeVis (que será apresentada na Seção 5) mostram que a mesma atendeu ao seu principal objetivo, que é servir como um apoio à utilização do MDP e visualização das modelagens geradas com o seu uso.

Após a avaliação da versão inicial da PryMeVis, foram definidos os casos de uso que representam um cenário chave de seu uso por designers, no sentido de aprimorá-la. Nesse cenário, o usuário faz seu cadastro na ferramenta, cria uma modelagem referente a uma determinada RSO e cadastra os tipos de comunicações que representam as oportunidades de compartilhamento de informação pessoal no sistema. Dessa forma, é permitido ao usuário fazer a modelagem da RSO usando o MDP, podendo este acessar posteriormente os dados modelados através da PryMeVis. Foram definidas também funcionalidades desejáveis (que ainda não foram implementadas) a serem incorporadas à sua próxima versão. Um ponto importante é a implementação de um sistema de ajuda nos moldes proposto por [Silveira et al. 2004], que possa ser utilizado pelos usuários para entenderem melhor ou tirarem dúvidas sobre aspectos relativos ao MDP. Com isso, o designer poderia aprender sobre o MDP durante sua interação com a PryMeVis, de uma maneira didática e rápida. Além disso, é de grande interesse pesquisar maneiras alternativas de visualizar, filtrar e comparar as redes modeladas, bem como suas comunicações. Atualmente, a PryMeVis permite que se veja a modelagem de cada tipo de comunicação da RSO separadamente. Porém, a possibilidade de selecionar diferentes tipos de comunicação para visualização em uma mesma tela, por exemplo, pode permitir a comparação dos mesmos no que tange aos níveis de privacidade oferecidos pela RSO aos seus usuários, o que pode ter um valor significativo para o designer.

5. Avaliação da Representação Visual e PryMeVis

Após o MDP ter sido proposto, foi realizada uma avaliação inicial do mesmo junto a designers de sistemas, que consistem em seus potenciais usuários [Villela 2016]. Dentre os objetivos dessa avaliação, pretendia-se coletar indicadores tanto sobre a capacidade da representação visual do MDP, quanto sobre a primeira versão da ferramenta PryMeVis atender a seus propósitos, no sentido de apoiar o uso do MDP. Neste trabalho focamos apenas na parte da avaliação relacionada a estes objetivos, que será descrita a seguir.

5.1. Avaliação Realizada

A avaliação aconteceu em dezembro de 2015 e envolveu seis participantes, designers de sistemas de informação. Além disso, todos eram usuários frequentes de RSOs, acessando esses sistemas mais de uma vez por dia, com exceção de um participante que acessava menos de uma vez por semana. Todos os participantes já haviam alterado suas configurações

⁷Site do D3 (d3js.org) e do Bootstrap (getbootstrap.com)

de privacidade nesses sistemas, ao menos para estabelecer uma configuração padrão para o compartilhamento de suas informações. A avaliação foi realizada em uma única sessão, com duração aproximada de três horas e meia, e consistiu de três etapas: apresentação do MDP, execução das tarefas e realização de grupo focal.

Na apresentação do MDP, foi informado aos participantes o objetivo da avaliação e coletado o consentimento dos mesmos, através da assinatura do termo de consentimento. Os participantes também responderam a um questionário com questões sobre a sua experiência no uso de RSOs e em design de IHC, dentre outros aspectos. A primeira etapa foi encerrada com uma apresentação completa sobre o MDP, com exemplos do seu uso na modelagem de uma RSO existente. Em seguida, na segunda etapa da avaliação, cada participante recebeu material impresso contendo uma visão geral do MDP: sua representação gráfica e uma breve descrição de cada dimensão e seus possíveis valores, bem como as possíveis definições para o controle, e os níveis de privacidade associados ao valores das dimensões. Foi então solicitado aos participantes que executassem quatro tarefas. Dentre elas, as que estavam relacionadas à representação visual eram: (a) usar o protótipo da ferramenta PryMeVis para gerar um modelo consolidado de um tipo de comunicação do Facebook, resultante da discussão em pares realizada a partir de modelos gerados individualmente com o MDP para esse tipo de comunicação; e (b) comparar duas RSOs distintas a partir de seus modelos MDP, buscando identificar diferenças entre as mesmas no que tange à privacidade relacionada ao compartilhamento de informação pessoal - para isso, os participantes receberam as representações tabular e visual equivalentes dessas RSOs⁸. Por fim, na terceira etapa da avaliação, foi realizado com os participantes um grupo focal sobre as tarefas realizadas, quando tiveram oportunidade de falar, dentre outros pontos, de suas opiniões sobre a utilidade da representação visual do MDP, bem como da ferramenta PryMeVis.

O material coletado durante a realização das tarefas consistiu dos modelos gerados pelas duplas de participantes com o uso da ferramenta PryMeVis, do registro por escrito (efetuado pelas mesmas duplas de participantes formadas na tarefa anterior) das diferenças percebidas entre as duas RSOs comparadas, e dos discursos dessas duplas enquanto realizavam as tarefas. Esses materiais foram analisados para a presente avaliação, bem como os discursos de todos os participantes durante a realização do grupo focal.

5.2. Resultados

Conforme mencionado, a avaliação coletou dados sobre o uso da ferramenta PryMeVis e também sobre a representação visual do MDP. Tal representação foi avaliada a partir da tarefa de comparação entre duas RSOs distintas, que foi realizada pelos participantes com base na versão impressa da representação visual dos modelos MDP das mesmas.

Sobre a ferramenta PryMeVis, todos os participantes conseguiram utilizá-la facilmente e entender de maneira imediata como o modelo MDP deveria ser criado dentro da mesma. As dúvidas que surgiram durante o seu uso foram principalmente em relação ao MDP (relacionadas à definição das dimensões e seus valores), e não sobre a PryMeVis em si. Os participantes apontaram algumas inconsistências na forma como os elementos do MDP eram apresentados na interface da ferramenta e fora dela. Além disso, os participantes fizeram sugestões de aspectos a serem incorporados na PryMeVis, sendo alguns deles

⁸Geradas por um especialista no MDP, através da modelagem reversa dos seus tipos de comunicação.

já previstos, embora ainda não implementados na sua versão inicial usada na avaliação. Por exemplo, a inclusão de um sistema de ajuda, com a explicação sobre as dimensões de privacidade e os possíveis valores para seus atributos valor e controle.

Sobre a representação visual, os participantes destacaram que esta os ajudou a entenderem melhor a modelagem que fizeram com o MDP e os seus efeitos, bem como a terem uma visão mais clara das relações entre as dimensões dentro do modelo. Isso foi apontado pelo participante P4: *“olhando a colméia [fazendo referência ao conjunto de hexágonos que representam as dimensões de privacidade do MDP], a gente conseguiu enxergar as relações entre os elementos ‘tá vendo esse aqui: tá falando isso e isso [referindo-se às dimensões e seus valores], tem a ver com esse outro aqui, que tem a ver com aquele outro. Isso aqui [referindo-se ao valor de uma dimensão] talvez não tá batendo”*. P3 também destacou a importância da representação visual no entendimento do MDP: *“O fato de ter uma representação visual, assim, ajudou bastante. Ela organiza um monte de informação de um jeito muito mais fácil. Tá tudo ali...você olha e já sabe”*.

Ao comparar as modelagens MDP das duas RSOs, foi observado que todos os seis participantes, organizados em três duplas, preferiram utilizar a representação visual à tabular. Além disso, alguns deles, aos discutirem com seus pares, expressaram a razão por tal preferência, como pode ser visto na fala de P6, por exemplo, ao discutir com P3: *“Eu acho que essa tabela é irrelevante [...]. O desenho [referindo-se à representação visual] é muito bom [...]. Gostei, resolve o problema”*. P1, durante as discussões no grupo focal, justificou o porquê de ter preferido utilizar a representação visual para comparar as duas RSOs: *“Pela questão das cores, o realce... você vê lá uma cor mais forte, questão da privacidade...na sua cara [...]. Se você pega o texto [se referindo ao texto na tabela], precisa ler. Já na [representação] visual, eu bato o olho e consigo identificar”*.

O uso de tonalidades diferentes de cores para representar os níveis de privacidade remetidos pelos valores das dimensões do MDP foi apontado pelos participantes como sendo de grande importância para o entendimento dos níveis de privacidade oferecidos pelas RSOs modeladas e, logo, para diferenciá-las no que tange à privacidade. A fala de P6 ilustra este ponto: *“Você vê um modelo todo preto, você já sabe que o nível de privacidade é alto, mas quando você vê um modelo todo claro, sabe que o nível de privacidade é baixo [...]. Foi muito útil ter as cores para tentar ver a diferença entre os modelos”*. P1, por sua vez, associou as cores com o baixo nível de controle concedido ao usuário: *“Onde está mais escuro, eu já consigo entender que o usuário não tem tanta liberdade”*. Um problema apontado pelos participantes foi o uso da cor branca (ou ausência de preenchimento do hexágono correspondente à dimensão de privacidade) para indicar que o valor da dimensão será definido em tempo de uso. Segundo eles, o usuário pode incorretamente associar o “branco” ao nível mais baixo de privacidade, devido à sua proximidade com o tom mais claro de cinza, utilizado na representação visual do MDP para denotar tal nível. P6, por exemplo, expôs essa dificuldade da seguinte forma: *“Agora, tem aquela questão das cores [...], por exemplo, branco que significa quando [o valor] não é especificado, mas o branco, teoricamente, seria o nível mais baixo de privacidade. Então isso aí, a questão de significado ficou pobre, né?! Tinha que ser uma outra cor”*.

Houve também sugestões que extrapolam o propósito da ferramenta, como foi o caso de P6, que sugeriu estender o protótipo para fornecer feedback sobre aspectos na RSO modelada que seriam críticos em relação à privacidade do indivíduo, com base nos

valores atribuídos às dimensões do MDP. No entanto, o atendimento dessa sugestão vai além do objetivo da ferramenta, uma vez que requer a inclusão no MDP de um componente semântico que possa identificar situações em que seria relevante alertar o designer sobre o impacto de suas decisões, caracterizadas pela combinação de valores atribuídos às dimensões, no nível de privacidade a ser oferecido pelo sistema aos seus usuários.

Assim, de uma forma geral, percebemos que a representação visual se mostra como um importante suporte ao uso e entendimento do MDP, melhorando a percepção dos designers sobre os efeitos das suas modelagens na forma como a privacidade relacionada ao compartilhamento de informações pessoais é tratada em uma RSO. No entanto, alguns pontos ainda precisam ser melhorados a fim de proporcionar uma representação mais adequada, no caso específico do efeito da atribuição de valores em tempo de uso nos níveis de privacidade proporcionados pelo sistema aos seus usuários. Os resultados também indicam que vale a pena investir para que a versão da PryMeVis aqui apresentada possa evoluir para se tornar uma ferramenta robusta de apoio ao designer no uso do MDP. O fato dos participantes terem levantado pontos que já haviam sido previstos para a ferramenta (como a necessidade de inclusão de um sistema de ajuda) reforça e justifica a importância da implementação desses recursos, como incremento funcional para tornar mais eficiente o seu uso e, conseqüentemente, o uso do MDP.

6. Conclusão

Esse trabalho tem por objetivo apresentar PryMeVis, uma ferramenta online para apoio ao uso do MDP, que consiste em um modelo descritivo para auxiliar designers na elaboração e avaliação do modelo de privacidade de RSOs, relacionada ao compartilhamento de informações pessoais. Tal ferramenta foi criada para facilitar o uso da representação visual do MDP, permitindo que designers criem modelos para o compartilhamento de informações pessoais em RSOs de forma mais eficiente, ao atribuir valores para as dimensões do MDP e visualizar o impacto de suas decisões nos estados de privacidade que podem ser alcançados pelos usuários dentro do sistema.

O propósito da avaliação de um protótipo ainda simplificado da PryMeVis (em relação à sua proposta) foi coletar indicadores iniciais sobre a utilidade da representação visual do MDP, além de informações que pudessem ajudar na definição da direção a ser seguida no desenvolvimento da ferramenta e do investimento requerido para tal. Assim, os resultados mostraram a importância da representação visual como suporte ao uso e entendimento do MDP, uma vez que melhora a percepção dos designers acerca dos efeitos de suas decisões, relativas à modelagem do compartilhamento de informações pessoais, nos níveis de privacidade oferecidos pelo sistema aos seus usuários. A avaliação também nos indicou alguns pontos que precisam ser melhorados em tal representação, especificamente no que tange às dimensões que possuem seus valores definidos em tempo de uso. Além disso, os resultados da avaliação também nos indicaram novos aspectos a serem considerados na ferramenta PryMeVis, como a inclusão de um sistema de ajuda, no sentido de que possa se tornar uma ferramenta robusta a ser oferecida a designers de RSOs para apoio ao aprendizado e uso do MDP.

A refatoração da PryMeVis está em andamento, com o objetivo de gerar uma ferramenta de modelagem mais efetiva no suporte ao uso do MDP. Por se tratar de um ferramenta com maior nível de interação, a arquitetura inicial foi modificada, sendo agora

baseada na abordagem “modelo, visão, controle” (MVC), através do arcabouço Ruby on Rails. Esta decisão visou garantir o baixo acoplamento e alta coesão do sistema, aumentando sua qualidade e tornando-o escalável. A interface e a visualização, assim como na versão inicial, continuaram empregando os padrões abertos de desenvolvimento web (HTML5, CSS3, JavaScript e SVG), além da biblioteca D3.js e Bootstrap. O banco de dados foi modelado possibilitando o armazenamento de mais de uma representação por vez, facilitando o acesso do designer a suas análises.

Os próximos passos consistem em finalizar o desenvolvimento da PryMeVis e disponibilizá-la a designers e interessados na aplicação do MDP. Um sistema de ajuda robusto e que permite ao designer entender a lógica de funcionamento do MDP durante a modelagem está sendo projetado. Além disso, uma segunda avaliação que contemple a interação completa com a ferramenta (não se atendo apenas à modelagem de tipos de comunicação) deve ser realizada, a fim de consolidar as decisões de projeto tomadas.

7. Agradecimentos

Agradecemos aos participantes da avaliação por sua colaboração; às agências CNPq e FAPEMIG, bem como ao InWeb (MCT/CNPq 573871/2008- 6) e MASWeb (FAPEMIG/PRONEX APQ-01400-14) por seu apoio financeiro parcial a esta pesquisa.

Referências

- Acquisti, A., Brandimarte, L., and Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347(6221):509–514.
- Belanger, F. and Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4):1017–1041.
- Bevan, J., Cummings, M., Kubiniec, A., Mogannam, M., Price, M., and Todd, R. (2015). How are important life events disclosed on facebook? relationships with likelihood of sharing and privacy. *Cyberpsychology, behavior and social networking*, 18(1):8–11.
- de Souza, C. S. (2005). *The semiotic engineering of human computer interaction*. MIT Press, Cambridge, MA.
- Epstein, D. A., Jacobson, B. H., Bales, E., McDonald, D. W., and Munson, S. A. (2015). From "nobody cares" to "way to go!": A design framework for social sharing in personal informatics. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*, pages 1622–1636, New York, NY, USA. ACM.
- Gao, B. and Berendt, B. (2013). Circles, posts and privacy in egocentric social networks: An exploratory visualization approach. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '13*, pages 792–796, New York, NY, USA. ACM.
- Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805.
- Malandrino, D., Petta, A., Scarano, V., Serra, L., Spinelli, R., and Krishnamurthy, B. (2013). Privacy awareness about information leakage: Who knows what about me?

- In *Proceedings of the 12th ACM Workshop on Workshop on Privacy in the Electronic Society*, WPES '13, pages 279–284, New York, NY, USA. ACM.
- Mazzia, A., LeFevre, K., and Adar, E. (2012). The pviz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS '12, pages 13:1–13:12, New York, NY, USA. ACM.
- Pang, J. and Zhang, Y. (2015). A new access control scheme for facebook-style social networks. *Computers & Security*, 54:44–59.
- Paul, T., Stopczynski, M., Puscher, D., Volkamer, M., and Strufe, T. (2012). C4ps: Colors for privacy settings. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, pages 585–586, New York, NY, USA. ACM.
- Pereira Junior, M., Xavier, S., and Prates, R. O. (2014). Investigating the use of a simulator to support users in anticipating impact of privacy settings in facebook. In *Proceedings of the 18th ACM International Conference on Supporting Group Work*, pages 63–72. ACM.
- Romero, N. A., Markopoulos, P., and Greenberg, S. (2013). Grounding privacy in mediated communication. *Comput. Supported Coop. Work*, 22(1):1–32.
- Schlegel, R., Kapadia, A., and Lee, A. J. (2011). Eyeing your exposure: Quantifying and controlling information sharing for improved privacy. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 14:1–14:14, New York, NY, USA. ACM.
- Silveira, M. S., Barbosa, S. D. J., and Souza, C. S. d. (2004). Designing online help systems for reflective users. *Journal of the Brazilian Computer Society*, 9(3):25–38.
- Stutzman, F. and Hartzog, W. (2012). Boundary regulation in social media. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 769–778, New York, NY, USA. ACM.
- Tierney, M. and Subramanian, L. (2014). Realizing privacy by definition in social networks. In *Proceedings of 5th Asia-Pacific Workshop on Systems*, pages 1–7.
- Villela, M. L., Xavier, S. I., Prates, R. O., Prates, M. O., Shipman, F., Prates, A. A., and Cardoso, A. A. (2015). Contrasting people's attitudes towards self-disclosure in online social networks and face-to-face settings. In *Proceedings of 21th International Conference on Collaboration and Technology (CRIWG 2015)*, pages 232–247.
- Villela, M. L. B. and Prates, R. O. (2015). Supporting designer in modeling privacy for social network sites. In *Proceedings of XIV Brazilian Symposium on Human Factors in Computer Systems (IHC 2015)*, pages 113–122, Salvador - BA. Sociedade Brasileira de Computação.
- Villela, M. L. B. V. (2016). *Um Modelo de Design de Privacidade para o Compartilhamento de Informacoes Pessoais em Redes Sociais Online*. PhD thesis, Departamento de Ciencia da Computacao, UFMG.
- Wang, Y., Gou, L., Xu, A., Zhou, M. X., Yang, H., and Badenes, H. (2015). Veilme: An interactive visualization tool for privacy configuration of using personality traits. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 817–826, New York, NY, USA. ACM.

Caracterização das Estratégias de Privacidade do Instagram

Geanderson E. dos Santos¹, Marcelo W. Barbosa², Glívia A. R. Barbosa³

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (DCC/UFMG). Av. Antônio Carlos, 6627 - Prédio do ICEX Pampulha, Belo Horizonte, MG – Brasil.

²Instituto de Ciências Exatas e Informática – Pontifícia Universidade Católica de Minas Gerais (PUC-MG). Av. Afonso Vaz de Melo, 1.200 - Barreiro de Baixo - Belo Horizonte – MG - Brasil, CEP 30640-070

³Departamento de Computação – Centro Federal de Educação Tecnológica de Minas Gerais (DECOM/CEFET-MG). Av. Amazonas, 7675 - Nova Gameleira - Belo Horizonte, MG – Brasil.

geanderson@dcc.ufmg.br, mwerneck@pucminas.br,
gliviabarbosa@decom.cefetmg.br

Abstract. *This paper aimed to verify and characterize how Instagram has supported their users' privacy. A qualitative methodology known as Semiotic Inspection Method (SIM) is used to identify and present the privacy strategies proposed by Instagram interface designer. Following, these strategies were compared to the privacy strategies available in the literature in order to verify the equivalence degree of Instagram privacy strategies and the ones available in the literature. Finally, an online questionnaire was sent to Instagram users aiming to verify whether users feel safe using the app based on the privacy strategies. The results showed that the strategies adopted by Instagram designer have contributed to the protection of Instagram's users. Moreover, the results can help in the development and evaluation of privacy settings in other social networks.*

Resumo. *Este trabalho buscou verificar e caracterizar como o Instagram tem suportado a privacidade dos seus usuários. Para isso, o Método de Inspeção Semiótica (MIS) foi aplicado para identificar e apresentar as estratégias de privacidade propostas pelo aplicativo. Posteriormente, essas estratégias foram comparadas com os princípios de privacidade disponíveis na literatura com o intuito de verificar o grau de correspondência entre as estratégias do Instagram e os princípios de privacidade identificados na literatura. Finalmente, um questionário online foi aplicado para usuários do Instagram buscando verificar se esses usuários se sentem seguros ao utilizar esse aplicativo. Os resultados mostraram que as estratégias adotadas pelo Instagram têm contribuído para proteger seus usuários. Além disso, esses resultados contribuem para o desenvolvimento e avaliação dos recursos de privacidade de outras redes sociais.*

1. Introdução

Redes sociais *online* têm se tornado cada vez mais populares, e no Brasil essa utilização não tem sido diferente [comScore 2015]. Essa crescente adoção abre espaço para que essas redes *online* se tornem, cada vez mais, mediadoras da interação entre as pessoas de diferentes regiões, culturas e condições físicas e/ou psicológicas [Villella e Prates 2015]. Contudo, juntamente com todos os benefícios que as redes sociais *online* podem oferecer a seus usuários, surge também a preocupação com a privacidade dos usuários nessas redes [Gill *et al.* 2011][Villella e Prates 2015]. Isso porque, na maioria das vezes, o conteúdo compartilhado (e.g., informações pessoais, fotos, vídeos) refere-se a informações pessoais e verdadeiras que ao serem acessadas sem um controle podem expor seus proprietários de forma que fiquem vulneráveis a pessoas não autorizadas e/ou mal intencionadas [Gill *et al.* 2011].

Com o intuito de potencializar a privacidade dos usuários de redes sociais, projetistas e pesquisadores da área de Interação Humano Computador (IHC) têm direcionado seus esforços para avaliar e propor mecanismos que permitem a configuração de privacidade nessas redes (e.g., [Fong *et al.* 2009][Junior *et al.* 2014][Rodrigues *et al.* 2012][Villella e Prates 2015]). Porém, algumas pesquisas têm demonstrado que os modelos de configuração correntes nem sempre oferecem proteção adequada para os diferentes perfis de usuários, resultando em uma disparidade entre a privacidade desejada e a real [Liu *et al.* 2011][Villella e Prates 2015]. Diante dessa realidade é importante caracterizar quais estratégias de privacidade têm sido utilizadas por softwares sociais para oferecer proteção de seus membros e verificar se essas estratégias contribuem para a privacidade dos mesmos [Young e Quan-Haase 2009].

Motivados pela importância de pesquisas que investigam como as redes sociais *online* têm favorecido a privacidade dos usuários, e considerando o Instagram um fenômeno em termos de popularidade mundial, com cerca de 300 milhões de usuários acessando o serviço mensalmente [Instagram 2015], o objetivo desse trabalho consiste em identificar e caracterizar as estratégias de privacidade do Instagram e verificar se, de fato, essas estratégias potencializam a privacidade dos usuários durante o uso do aplicativo.

Este trabalho foi conduzido a partir de uma metodologia qualitativa que consistiu na aplicação do Método de Inspeção Semiótica (MIS) [de Souza *et al.* 2006] para identificar e apreciar as estratégias de privacidade adotadas pelo projetista na interface do Instagram. Posteriormente, essas estratégias foram comparadas com diretrizes de privacidade disponíveis na literatura a fim de verificar o grau de adequação da proposta do Instagram às mesmas. Finalmente, foi aplicado um questionário aos usuários do aplicativo para verificar se, de fato, os recursos oferecidos têm apoiado no controle de privacidade desses usuários durante o uso do Instagram.

Os resultados dessa pesquisa permitiram constatar que o projetista do Instagram adota estratégias consideradas relevantes para potencializar a privacidade dos usuários. Nesse sentido, esse trabalho apresenta contribuições práticas e científicas para essa linha de pesquisa. Em termos práticos, os resultados reportados, bem como as estratégias de privacidade identificadas neste trabalho, podem servir de parâmetros para a melhoria e/ou

criação de mecanismos de privacidade em redes sociais *online*. Além disso, essa pesquisa alerta os designers de interface sobre a importância de oferecer configurações de privacidade adequadas aos usuários nesse ambiente *online*. Em termos científicos/metodológicos, os resultados do MIS reforçam a aplicabilidade do método, devido à sua fundamentação teórica, para identificar as estratégias de design, comunicadas na interface, que visam potencializar determinadas qualidades de uso, neste caso, a privacidade.

2. Trabalhos Relacionados

A privacidade em redes sociais é um assunto complexo que tem sido investigado por pesquisadores de diferentes áreas sob diferentes perspectivas. Por exemplo, o trabalho realizado por Johnson *et al.* (2012) alerta para o fato de que, nem sempre, configurações de privacidade são capazes de combater a ameaça interna, ou seja, pessoas que fazem parte da própria rede de amigos de um determinado usuário, mas que poderiam usar as informações que eles têm acesso para prejudicar o proprietário do perfil. Em Xavier *et al.* (2014), os autores apreciam a privacidade de usuários em redes sociais através de uma metodologia qualitativa de pesquisa. O estudo mostrou a importância dos usuários serem capazes de denunciar abusos em redes sociais.

Em uma perspectiva diferente, Hull *et al.* (2010) enfatizam como que decisões de projeto de interface podem expor informações de usuários. Eles destacam a implementação do *feed* de notícias e a inserção de aplicativos no Facebook, e como tais mudanças na interface impactaram consideravelmente na privacidade dos membros daquela rede. Outros autores têm explorado quais configurações contribuem para fornecer uma melhor privacidade para membros dessas redes *online* [Young e Quan-Haase 2009][Johnson *et al.* 2012].

Embora os trabalhos apresentados foquem em aspectos de privacidade em redes sociais *online*, não foram encontradas pesquisas que fizessem a análise feita neste estudo, caracterizando as estratégias de privacidade que têm sido adotadas pelas redes sociais *online* através de um estudo de caso no Instagram. Esse tipo de investigação é importante porque seus resultados podem apontar as vantagens e/ou limitações no uso dessas estratégias, de forma a colaborar no desenvolvimento e/ou avaliação das configurações de privacidade dos softwares sociais *online*.

3. Metodologia e Referencial Teórico

Considerando o objetivo deste trabalho, a seguinte questão de pesquisa foi investigada: “*Quais as estratégias de privacidade utilizadas pelo Instagram contribuem para proteger as informações dos usuários?*”. A metodologia adotada para responder essa pergunta consistiu em uma abordagem qualitativa, dividida em quatro etapas.

Inicialmente, foi realizado um levantamento bibliográfico com o objetivo de identificar estratégias de privacidade, presentes na literatura, que poderiam ser utilizadas como referência para verificar a adequação da proposta de privacidade do Instagram. Em seguida, a interface do Instagram foi inspecionada, seguindo o Método de Inspeção Semiótica (MIS) [de Souza e Preece 2004][de Souza *et al.* 2006][de Souza e Leitão 2009]

para identificar e apresentar as estratégias de privacidade comunicadas em sua interface.

A terceira etapa da metodologia buscou comparar as estratégias encontradas na literatura com as identificadas na interface do Instagram. Para finalizar, foi investigado junto aos usuários desse aplicativo se as estratégias identificadas contribuem para a privacidade de seus membros. Para isso foi aplicado um questionário *online* e as respostas foram analisadas de forma a caracterizar a influência das estratégias no uso do Instagram. Para um melhor entendimento dos conceitos e métodos mencionados na metodologia, a seguir o MIS será apresentado, bem como a justificativa pela escolha desse método.

3.1 Método de Inspeção Semiótica

O Método de Inspeção Semiótica (MIS) é fundamentado na Teoria da Engenharia Semiótica – uma teoria explicativa de Interação Humano Computador (IHC). A Engenharia Semiótica (EngSem) nos permite entender os fenômenos envolvidos no design, uso e avaliação de um sistema interativo [de Souza 2004]. Para a EngSem, a interface de um sistema é vista como uma comunicação, unidirecional e indireta, do projetista do sistema aos seus usuários. O objetivo dessa comunicação é transmitir aos usuários a sua visão sobre a quem o sistema se destina; que problemas ele pode resolver e como interagir com ele [de Souza 2004]. O usuário compreende a mensagem à medida que interage com a interface. Essa comunicação é conhecida como metacomunicação e sua mensagem como metamensagem. A metamensagem transmitida do projetista para o usuário é composta por signos – i.e. qualquer coisa que representa algo para alguém [Peirce 1992]. A EngSem classifica os signos em um sistema interativo como: metalinguísticos, estáticos e dinâmicos [de Souza *et al.* 2009][de Souza *et al.* 2010].

No intuito de apreciar a qualidade da comunicação entre o projetista e o usuário, a EngSem foca na avaliação da propriedade de comunicabilidade do sistema, isto é, sua capacidade de transmitir aos usuários as decisões e princípios que guiaram seu design de forma eficiente e eficaz [Prates *et al.* 2000]. Um dos métodos propostos pela EngSem para apreciação da comunicabilidade é o MIS [de Souza *et al.* 2006][de Souza *et al.* 2009][de Souza *et al.* 2010]. Para isso, um avaliador (que tenha conhecimento de IHC e Engenharia Semiótica) inspeciona o sistema com o objetivo de analisar o sistema proposto a partir da metamensagem projetista-usuário [de Souza *et al.* 2006][de Souza *et al.* 2009][de Souza *et al.* 2010].

A avaliação da interação de um sistema utilizando o MIS consiste em cinco etapas [de Souza *et al.* 2006]: (1) inspeção dos signos metalinguísticos; (2) inspeção dos signos estáticos; (3) inspeção dos signos dinâmicos; (4) contraste e comparação entre as mensagens identificadas em cada uma das inspeções e (5) apreciação da qualidade da metacomunicação. Nos passos de 1 a 3, o especialista deve ser capaz de reconstruir a metamensagem proposta pelo projetista a partir dos signos inspecionados. No passo 4, o especialista contrasta as mensagens geradas nos passos anteriores e, a partir desta análise, explora a possibilidade do usuário atribuir significados diferentes a um mesmo objeto (signo). Finalmente, no passo 5, o especialista gera uma versão unificada da metamensagem e registra a sua apreciação final da qualidade do sistema.

Embora originalmente o MIS tenha sido proposto para avaliar a comunicabilidade de um sistema, uma revisão na literatura realizada por [Reis e Prates 2011], revelou que, devido à sua fundamentação teórica, esse método também pode ser utilizado para identificar as estratégias de *design* comunicadas na interface que visam potencializar determinadas qualidades de uso. Neste sentido, trabalhos como o realizado por Coutinho *et al.* (2011) identificam estratégias sonoras para orientação em jogos, através da aplicação do MIS. Já o trabalho realizado por Barbosa *et al.* (2011) utilizou o MIS para identificar estratégias de sociabilidade em redes sociais. Os autores Silva, Marcelle e Oliveira (2014), utilizaram o MIS para identificar estratégias de marketing (MKT) em sites de hotel, e, no trabalho feito por [Silva *et al.* 2015], o MIS é utilizado para identificar estratégias de *gamificação* em aplicativos educacionais. Todas essas evidências justificam a utilização do MIS na identificação de estratégias no *design* de interação.

4. Apreciação da Proposta de Privacidade do Instagram

Projetistas de softwares sociais *online* não podem controlar a interação social que acontece entre os usuários dentro dessas redes sociais. Contudo, decisões de interface podem auxiliar no controle de privacidade dos usuários desses softwares [Strater e Lipford 2008][Hull *et al.* 2010]. Devido ao crescimento abrupto do Instagram e a possibilidade de divulgar informações pessoais através desse aplicativo, este trabalho buscou avaliar a interface do Instagram para identificar e caracterizar sua proposta de privacidade. Para isso foi realizada uma avaliação utilizando o MIS, cujo objetivo foi identificar as decisões do projetista que refletem em estratégias de privacidade descritas na literatura.

A avaliação foi realizada no período de sete dias (entre 05/09/2015 e 11/09/2015) e foi conduzida por dois autores desse trabalho, sendo que ambos já haviam realizado outras avaliações de interface no contexto de softwares sociais *online* utilizando o método em questão [Barbosa *et al.* 2013]. O escopo da avaliação foi limitado à versão em Inglês do aplicativo do Instagram e às seguintes tarefas: (1) pesquisar e seguir usuários, (2) publicar e visualizar conteúdo, (3) interagir socialmente com os seguidores através dos recursos do sistema como *chat* (mensagem) e envio de fotos/vídeos, e (4) configurar aspectos de privacidade relacionados as três tarefas anteriores. A seguir, são apresentados os principais resultados da primeira etapa dessa pesquisa.

5. Proposta de Privacidade do Instagram

A partir da aplicação do MIS, verificou-se que o propósito do projetista em relação ao Instagram é oferecer um espaço para interação e compartilhamento de fotos e vídeos. O sistema dispõe de políticas e termos de uso que, conforme relatado pelo próprio projetista em sua metamensagem, devem ser seguidas porque, “*Ao usar o nosso Serviço, você entende e concorda que estamos fornecendo uma plataforma para você publicar e compartilhar conteúdo, incluindo fotos, comentários e outros materiais ("Conteúdo do usuário"). Isso significa que outros Usuários podem pesquisar por, ver, usar ou compartilhar qualquer Conteúdo do usuário que você disponibilize publicamente através do Serviço, em consonância com os termos e condições desta Política de Privacidade e nossos Termos de uso.*” [Instagram 2015].

Qualquer pessoa que possui e-mail ou uma conta no Facebook, e aceitar a política de privacidade e termos de uso pode criar uma conta no Instagram. Ao criar um perfil, é necessário informar apenas um nome de usuário e senha. A partir da interação com o aplicativo, diversas outras informações poderão ser inseridas no perfil, tais como: Nome, Website, Biografia, Telefone e Sexo. Além disso, é possível disponibilizar uma imagem, cujo objetivo é identificar visualmente o proprietário do perfil, essa imagem também pode ser importada de outras redes sociais (e.g., Facebook e Twitter).

Para que a interação entre os membros aconteça, o projetista do Instagram oferece um recurso de pesquisa e recomendação de pessoas. Para interagir com outros integrantes do Instagram, o proprietário do perfil tem duas opções: (1) seguir outros membros do Instagram ou (2) permitir que outros membros o siga. Devido a questões de privacidade, usuários podem tornar seus perfis privados. Essa opção permite que o proprietário do perfil receba uma requisição de usuários que o queiram seguir, ou seja, ele pode aprovar ou não que outros usuários tenham acesso ao seu perfil (i.e., conceder visualização a suas fotos/vídeos).

Os principais recursos oferecidos para interação e compartilhamento de conteúdo entre o proprietário do perfil e seus seguidores são compartilhamento de fotos/vídeos, mensagens diretas, e envio direto de fotos e vídeos. É importante destacar que o Instagram fornece uma gama de ferramentas para edição do conteúdo antes de ser postado. Membros podem compartilhar conteúdo com outros membros, desde que eles tenham permissão para isso. Além disso, é possível incluir informações como localização, título do conteúdo, ou mesmo marcar seguidores. Outras redes sociais estão vinculadas a opção de compartilhamento, por essa razão, esse conteúdo (i.e., foto ou vídeo), pode aparecer diretamente em outras redes sociais, tais como: Facebook, Twitter, Tumblr e Flickr.

Em relação ao conteúdo multimídia compartilhado (i.e., fotos e vídeos) na tela principal do aplicativo, é permitido que usuários curtam, comentem, ou mesmo enviem uma publicação para outros membros. Além disso, a/o foto/vídeo pode ser compartilhada em outras redes sociais após sua publicação. Caso o usuário deseje desfazer da postagem, ou mesmo queira modificar alguma coisa. É possível editar e deletar publicações.

Conforme mencionado anteriormente, além do compartilhamento de conteúdo entre o proprietário do perfil e seus seguidores, é possível interagir socialmente através de mensagens diretas. O mecanismo de mensagem diretas possibilita um bate-papo entre o dono do perfil e outro membro do Instagram. Além disso, caso seja do interesse dos interlocutores, é possível adicionar outras pessoas à conversa. O conteúdo da conversa é gravado, para posterior visualização e, além disso, é possível anexar fotos ou vídeos à interação. Novamente, o menu de notificações ajuda a manter a privacidade dos usuários do Instagram, uma vez que eles podem impedir que não seguidores os enviem mensagens diretamente.

Com o intuito de garantir uma boa interação entre os usuários que mantem contato através do Instagram, o projetista notifica cada membro sobre as atualizações que ocorrem em seu perfil, por exemplo, quando um contato curte a foto do dono do perfil uma mensagem é enviada ao mesmo, o usuário recebe uma notificação para que tenha a

oportunidade de fornecer um feedback para a pessoa ou o grupo que interagiu com ele (e.g., responder uma mensagem, curtir um conteúdo, aceitar uma solicitação para seguir).

O menu de notificações é uma das principais decisões do projetista, pois essa função é utilizada no controle sobre a exibição de conteúdo compartilhado no perfil de um usuário. Esse recurso é essencial para manter uma boa privacidade dos usuários dessa rede. Entretanto, o menu de notificações não é a única ferramenta que favorece a privacidade no Instagram. Fotos e vídeos indevidamente postados por seguidores e não seguidores podem ser reportados para os administradores do Instagram pelo dono do perfil. Seguidores hostis também pode ser bloqueados pelo dono do perfil, com o intuito de proteger a rede de amigos do proprietário da conta. A Figura 1 ilustra evidências da proposta do projetista para a interface do Instagram.

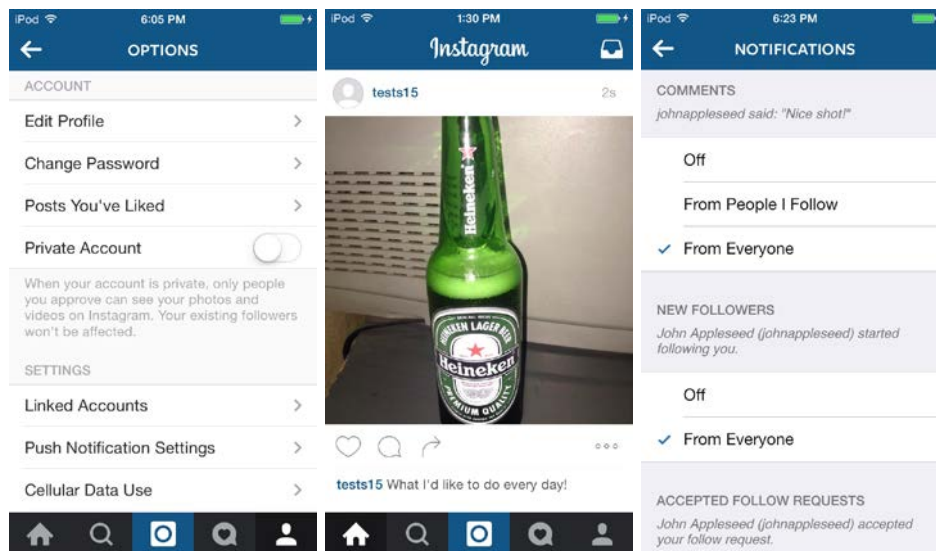


Figura 1 – Proposta do Projetista do Instagram

5.1. Estratégias de Privacidade Identificadas

Uma vez identificada a proposta do projetista em relação a privacidade no Instagram, este trabalho buscou verificar se as decisões comunicadas na interface são compatíveis com as estratégias que, de acordo com a literatura [Young e Quan-Haase 2009][Johnson *et al.* 2012][Xavier *et al.* 2014], podem ser consideradas fatores motivacionais para usuários em ambientes sociais. A Tabela 1 lista as estratégias de privacidades recomendadas na literatura.

Tabela 1 - Estratégias de Privacidade identificadas na Literatura

Estratégias Privacidade		
Cód. Estratégia	Descrição	Referência
E01	Ocultar informações pessoais ou prover falsas informações sobre mim	[Young e Quan-Haase 2009]
E02	Excluir determinadas informações pessoais	[Young e Quan-Haase 2009]

E03	Possibilitar comunicação privada entre usuários	[Young e Quan-Haase 2009]
E04	Permitir a alteração das configurações "default" do sistema	[Young e Quan-Haase 2009]
E05	Remover informações que outros usuários postaram sobre mim	[Young e Quan-Haase 2009]
E06	Desmarcar-se de determinados posts	[Young e Quan-Haase 2009] [Johnson <i>et al.</i> 2012]
E07	Restringir acesso à certas informações pessoais	[Young e Quan-Haase 2009]
E08	Proteger rede de amigos	[Johnson <i>et al.</i> 2012]
E09	Bloquear usuários indesejados	[Young e Quan-Haase 2009]
E10	Denunciar abusos	[Xavier <i>et al.</i> 2014]
E11	Alterar status de um post	[Johnson <i>et al.</i> 2012]
E12	Criar listas customizadas de contatos	[Johnson <i>et al.</i> 2012]
E13	Manter mais de uma conta	[Johnson <i>et al.</i> 2012]

A partir da aplicação do MIS foi possível identificar que o projetista utiliza 11 das 13 estratégias de privacidade já conhecidas na literatura, ou seja, observou-se que o projetista atendeu 85% das estratégias levantadas na literatura para endereçar privacidade. A Tabela 2 apresenta as estratégias de privacidade propostas pela literatura, as decisões do projetista identificadas pelo MIS e exemplos de aplicação destas estratégias no aplicativo. É importante ressaltar que uma estratégia pode remeter a uma ou mais decisões do projetista da interface.

Tabela 2 - Estratégias de Privacidade x Decisão do Projetista

Cod. Estratégia	Cod. Decisão Projetista	Decisão do Projetista	Exemplo de Aplicação no Instagram
E01	DP01	Permitir deixar em branco praticamente qualquer informação sobre o usuário, alteração livre dessas informações e não conferir a veracidade da mesma	Usuários somente precisam informar <i>username</i> e <i>password</i> . Todas as outras informações do perfil podem ficar em branco ou serem alteradas
E02			
E03	DP02	Permitir o envio de mensagens diretas para qualquer usuário	Usuários podem mandar mensagem para qualquer pessoa conectada ao Instagram

E04	DP03	Permitir alteração das configurações padrão do sistema	Usuários podem alterar diversas configurações de privacidade no menu de configurações
E05	DP04	Permitir a remoção de fotos/vídeos postados por seguidores	Usuários podem remover-se de qualquer foto/vídeo postado por outros usuários
E06			
E07	DP05	Permitir perfil privado	Usuários podem configurar o perfil como privado fazendo novos seguidores precisarem de uma aprovação
E08			
E09	DP06	Permitir o bloqueio de seguidores	Seguidores podem ser bloqueados através de uma opção dentro de seus perfis
E10	DP07	Permitir reportar qualquer post no Instagram	Usuários podem reportar para os administradores qualquer foto/vídeo postado na comunidade
E11	DP08	Permitir a edição de uma foto/vídeo	Usuários podem editar uma foto ou vídeo postados
E12	DP00	Instagram falha nessa estratégia	Nulo
E13			

A Figura 4 apresenta como algumas decisões do projetista podem aparecer na interface do aplicativo.

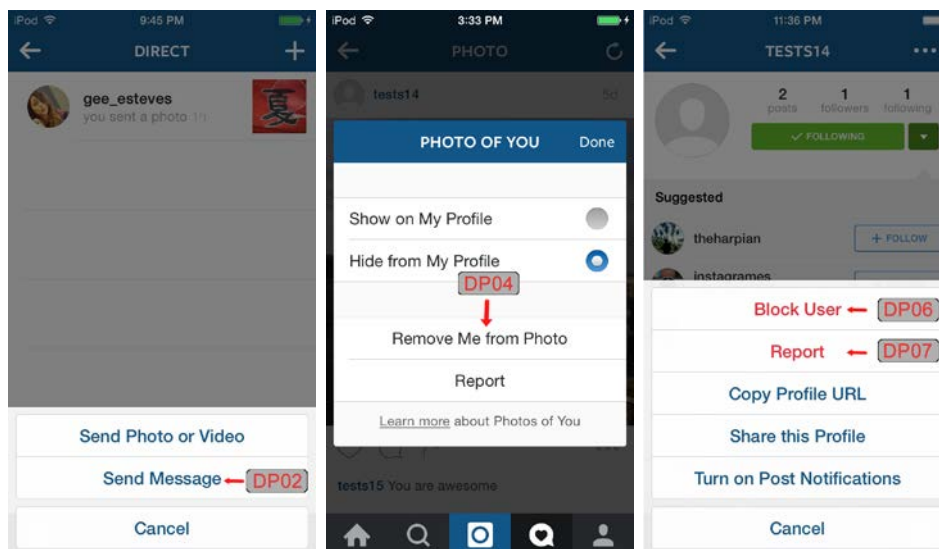


Figura 2 – Decisões do Projetista na interface do Instagram

6. Análise das Configurações de Privacidade do Instagram sob a Perspectiva dos Usuários

Com o objetivo de investigar junto aos usuários do Instagram se eles percebem e adotam as estratégias de privacidade identificadas na análise anterior, foi aplicado um questionário *online* composto por 14 questões de múltipla escolha. As questões foram divididas de tal forma que fosse possível caracterizar: (1) o perfil dos participantes da pesquisa, (2) se os usuários sentem-se seguros quando utilizam o aplicativo, e (3) quais as decisões do projetista são mais relevantes para manter os usuários seguros. O questionário ficou disponível para acesso durante uma semana, entre os dias 6 a 12 de Outubro de 2015. O mesmo foi divulgado por meio de redes sociais, e-mails e fóruns de discussões sobre o Instagram.

Ao todo 62 pessoas participaram da pesquisa. Dos participantes, cerca de 3% foram classificados na faixa etária entre 13 e 18 anos, 77% declararam estar na faixa entre 19 e 30 anos e 19% estão na faixa entre 31 e 55 anos. Embora houvesse as opções que representavam menores de 12 anos e maiores de 56 anos não foram obtidas respostas para esses perfis. Quanto às redes sociais utilizadas, além do Instagram, 97% utilizam o Facebook e o WhatsApp, 79% o Youtube, 71% o Twitter, e 53% o Google+. Outros 10% citaram outras redes sociais, tais como o SnapChat e o LinkedIn.

6.1. Análise das Configurações de Privacidade no Instagram

Inicialmente, os participantes foram questionados se eles se preocupam com questões de privacidade quando utilizam o Instagram, 82% dos participantes afirmaram se preocupar com essa questão durante o uso do aplicativo, enquanto 18% disseram não se importar com privacidade no Instagram.

Nesse sentido, os participantes foram questionados se conheciam as configurações de privacidade do Instagram e se sabem aplicá-las em seus perfis. Conforme demonstrado na Figura 3, embora a maioria (56%) dos participantes afirme conhecer e aplicar os recursos de privacidade em seus perfis, 44% afirmaram não adotar as configurações por razões como, por exemplo, a falta de conhecimento sobre sua existência ou sobre como acessá-las na interface.

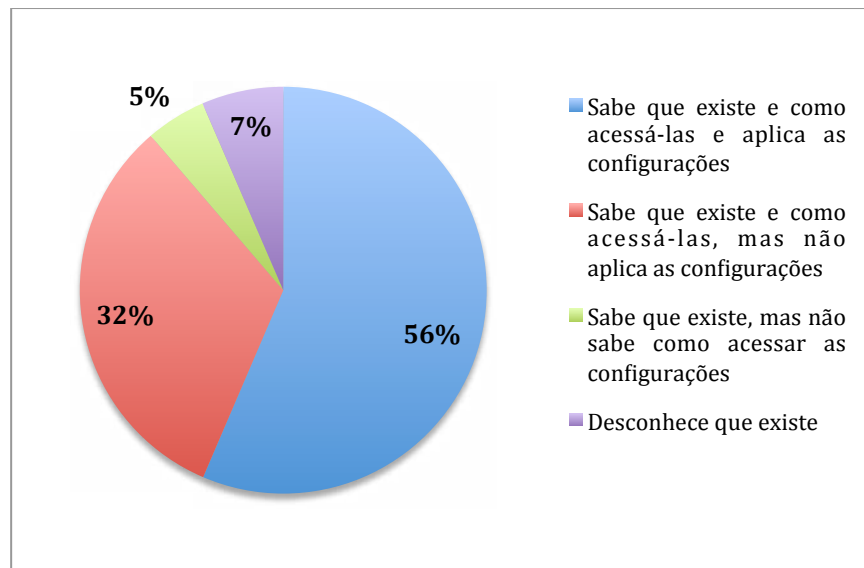


Figura 3 – Grau de conhecimento sobre as Configurações de Privacidade

Fonte: Dados da Pesquisa

Na sequência, os participantes que afirmaram aplicar as configurações de privacidade foram questionados sobre quais recursos eles utilizam. O gráfico da Figura 4 apresenta os resultados obtidos. É possível notar que a principal estratégia de privacidade adotada pelos usuários (42%) refere-se a configurar a visibilidade do perfil para privado. As demais opções são utilizadas por, no máximo, 37% dos participantes. Esse resultado chama atenção para o uso limitado que esses usuários têm feito das possibilidades que o Instagram oferece para resguardar a privacidade no aplicativo.

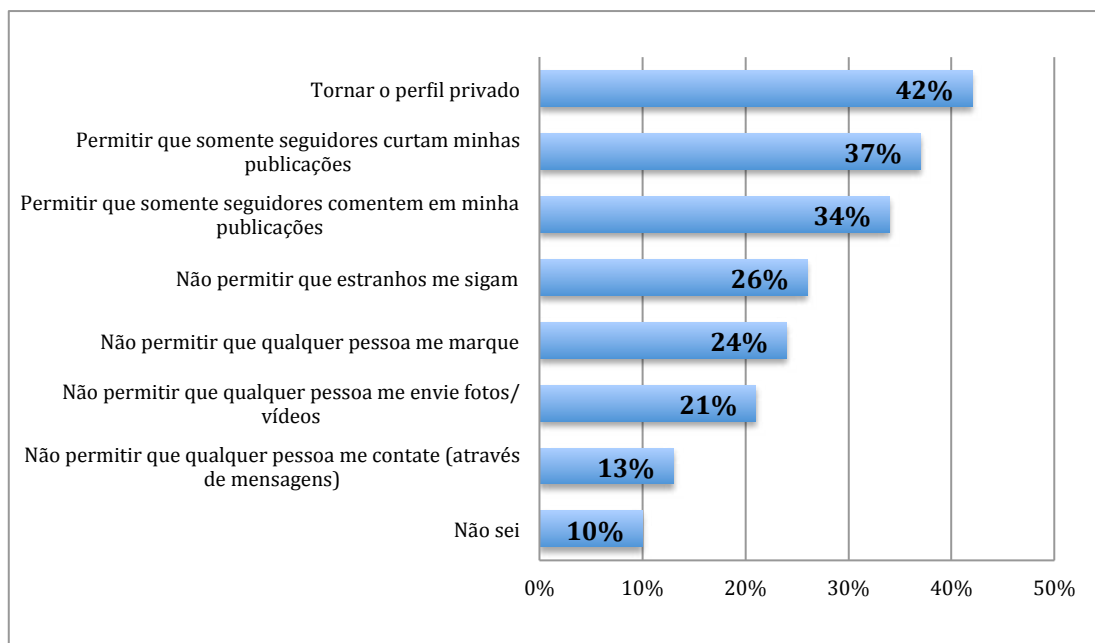


Figura 4 - Configurações de Privacidade do Instagram adotadas pelos Participantes

Fonte: Dados da pesquisa

6.2. Discussão dos Resultados

Como mostrado pela (Tabela 2), o Instagram adota 11 das 13 estratégias propostas pela literatura, ou seja, aproximadamente 85% das estratégias são atendidas. Esse pode ser considerado um fator positivo nesta rede social, porque como mostrado por [Young e Quan-Haase 2009][Jonhson *et al.* 2012], a adoção dessas estratégias pode favorecer para que o ambiente *online* se torne mais seguro. Contudo, o fato do Instagram não atender duas estratégias (E12 – Criar listas customizadas de contatos e E13 – Manter mais de uma conta) pode afetar aspectos de privacidade que deveriam ser considerados no design do aplicativo. Por exemplo, suponha que um usuário X poste uma foto pessoal para seus seguidores. Um usuário mal-intencionado que o segue no Instagram, toma posse daquele arquivo e o usa para prejudicar o usuário X de alguma forma, como por exemplo, postando a foto em um site para adultos denegrindo a imagem do usuário X. Entre as várias formas de evitar tal problema, a implementação de E12 poderia ter incentivado o usuário X a postar essa foto somente para seus amigos mais próximos [Strater e Lipford 2008].

A partir da análise das configurações de privacidade oferecidas pelo aplicativo, com a utilização que os usuários têm feito das configurações, nota-se que o recurso mais utilizado não atinge metade dos usuários (42% tornar o perfil privado). Isso indica que, embora o aplicativo disponibilize diversos recursos para garantir a privacidade de seus usuários, esses recursos têm sido usados de forma limitada, tornando as pessoas vulneráveis na rede. Uma hipótese para essa limitação de uso, é a falta de clareza na interface sobre como aplicar essas configurações nos perfis.

Dessa forma, a triangulação dos resultados indica que as estratégias existentes na literatura para endereçar privacidade podem ser úteis para oferecer recursos que potencializem a segurança dos usuários. No entanto, conforme já alertado por [Xavier *et al.* 2014][Rodrigues *et al.* 2012][Fong *et al.* 2009], e também evidenciado neste trabalho, esses recursos contribuem para privacidade somente quando são bem comunicados na interface, caso contrário, os usuários negligenciarão a utilização dos mesmos, e a possível segurança que tais recursos podem oferecer para membros de redes sociais *online* não serão aproveitadas em sua totalidade.

7. Conclusão

Este trabalho buscou apreciar e caracterizar as estratégias de privacidade adotadas em redes sociais e investigar se os usuários percebem e utilizam os recursos oferecidos. Os principais resultados apresentados e discutidos indicam que apesar do Instagram fazer uso da maioria das estratégias de privacidade, por meio das possibilidades de configurações oferecidas, os recursos disponíveis nem sempre são percebidos e utilizados através da interface.

Embora o estudo tenha sido realizado no Instagram, os resultados apresentados e discutidos são relevantes também para outros softwares sociais e/ou sistemas colaborativos que lidem, sobretudo com informações pessoais de usuários. A relevância deste tipo de investigação pode ser sustentada pelo argumento apresentado por [Willinger *et al.* 2010], onde os autores destacam a importância de apreciar e caracterizar as estratégias de interação, bem como a forma como diversos tipos de usuários estão interagindo, de forma a

verificar e endereçar os problemas que podem afetar na interação dos mesmos com esses sistemas [Willinger *et al.* 2010].

Sendo assim, esse artigo apresenta contribuições técnicas e científicas para as áreas de Sistemas Colaborativos, IHC e afins. Em termos científicos/metodológicos, os resultados do MIS reforçam a aplicabilidade do método, tendo em vista à sua fundamentação teórica, com o objetivo de identificar as estratégias de design, comunicadas na interface de sistemas colaborativos, que visam potencializar determinadas qualidades de uso, neste caso, a privacidade dos usuários. Em termos práticos, os resultados reportados, assim como as estratégias de privacidade identificadas neste trabalho, podem servir de parâmetros para a melhoria e/ou desenvolvimento de mecanismos de privacidade em redes sociais *online*. Além disso, essa pesquisa alerta os projetistas de interface sobre a importância de oferecer configurações de privacidade adequadas aos usuários em ambientes *online*.

Nesse sentido, como trabalhos futuros, poderia ser investigada a possibilidade de identificar signos na interface que permitem classificar as possibilidades de privacidade oferecidas por softwares sociais. Isso ajudaria no projeto e avaliação de outros sistemas/aplicativos que busquem manter seus usuários seguros. Além disso, seria interessante explorar como as estratégias de privacidade podem influenciar na interação do usuário.

Referências

- Barbosa, G. A. R.; Prates, R. O. e Corrê, L. 2011. Análise da sociabilidade de comunidades online para os usuários surdos: Um estudo de caso do orkut. In Proc. of the X IHC & CLIHC '11, (2011), 235-245.
- Barbosa, G. A. R.; Santos, G. E.; and Victor M. de O. Pereira, V. M. O. 2013. Caracterização qualitativa da sociabilidade no Facebook. In Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems (IHC '13). Brazilian Computer Society, Porto Alegre, Brazil, Brazil, 72-81.
- comScore. 2015. The State of Social Media in Brazil and the Metrics that Really Matter. comScore Presentation. Disponível em: <https://goo.gl/l3dg8U>.
- Coutinho, F.; Prates, R. O.; Chaimowicz, L. An analysis of information conveyed through audio in an fps game and its impact on deaf players experience. In: IEEE. Games and Digital Entertainment (SBGAMES), 2011 Brazilian Symposium on. [S.l.], 2011. p. 53–62.
- de Souza, C.S. and Preece, J. 2004. A Framework for Analyzing and Understanding Online Communities. In: Interacting with Computers, Amsterdam, v. 16(3), (2004), 579-610.
- de Souza, C. S., Leitão, C. F., Prates, R. O., and da Silva, E. J. 2006. The semiotic inspection method. In: Proc. of IHC 2006, SBC, (2006), 148-157.
- de Souza, C. S., and Leitão, C. F. Semiotic engineering methods for scientific

- research in HCI. Morgan & Claypool Publishers, Editor: John M. Carroll, 2009.
- de Souza, C. S., Leitão, C. F., Prates, R. O., Bim, S.A., da Silva, E.J. 2010. Can inspection methods generate valid new knowledge in HCI? The case of semiotic
- Fong, P. W., Anwar, M. and Zhao, Z. 2009. A privacy preservation model for facebook-style social network systems. In Computer Security - ESORICS, Springer Berlin Heidelberg, 303-320.
- Gill, A. J., Vasalou, A., Papoutsis, C., and Joinson, A. (2011) "Privacy Dictionary: A Linguistic Taxonomy of Privacy for Content Analysis". In Proceedings of CHI 2011, Session: Privacy May 7-12, 2011, Vancouver, BC, Canada.
- Hull, G., Lipford, H. R. and Latulipe, C. 2011. Contextual gaps: privacy issues on Facebook. *Ethics and Inf. Technol.* 13, 4 (December 2011), 289-302.
- Instagram. 2015. Press News. Disponível em: <https://instagram.com/press/>
- Johnson, M., Egelman, S. and Steven M. Bellovin. 2012. Facebook and privacy: it's complicated. In Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12). ACM, New York, NY, USA, , Article 9 , 15 pages.
- Junior, M. P., Xavier, S. and Prates, R. O. 2014. Investigating the use of a Simulator to Support Users in Anticipating Impact of Privacy Settings in Facebook. In Proceedings of the 18th ACM International Conference on Supporting Group Work, ACM, 63-72.
- Leitão, C. F., de Souza, F. C. S. and Barbosa, C. M. de A. 2007. Face-to-face sociability signs made explicit in CMC. In Proc. of the 11th IFIP TC 13 international conference on Human-computer interaction (INTERACT'07), (2007), 5-18
- Liu, Y., Gummadi, K. P. and Mislove, A. 2011. Analyzing Facebook Privacy Settings: User Expectations vs Reality. In Proceedings of Internet measurement conference, ACM, 61-70.
- Peirce, C.S. The Essential Peirce. Indiana University Press, Bloomington, 1992.
- Prates, R. O., de Souza, C. S., Barbosa, S. D. J. 2000. A method for evaluating the communicability of user interfaces. *interactions.* 7(1), (2000), 31-38.
- Preece, J. 2001(a). Online communities: Usability, Sociability, Theory and Methods. In R. Earnshaw, R. Guedj, A. van Dam and T. Vince (Eds) *Frontiers of Human-Centred Computing, Online Communities and Virtual Environments.*
- Reis, S. de S. and Prates, R. O. 2011. Applicability of the Semiotic Inspection Method: a Systematic Literature Review. In: Proc. of X IHC & V CLIHC'11, (2011), 177-186.

- Rodrigues, K. R. H., Canal, M. C., Xavier, R. A. C., Alencar, T. S., and Neris, V. P. A. 2012. Avaliando aspectos de privacidade no Facebook pelas lentes de usabilidade, acessibilidade e fatores emocionais. In Companion Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems (IHC '12).
- Silva, J. C. R.; Oliveira, E, R; and Barbora, G. A. R. Pereira, V. M. O. 2015. Caracterização de Estratégias de Gamificação em Aplicativos Móveis Educacionais: Um Estudo de Caso do Aplicativo Duolingo. In Proceedings of the 14th Brazilian Symposium on Human Factors in Computing Systems (IHC '15). Brazilian Computer Society, Porto Alegre, Brazil.
- Strater, C. and Lipford, H. R. 2008. Strategies and struggles with privacy in an online social networking community. In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 1 (BCS-HCI '08), Vol. 1. British Computer Society, Swinton, UK, UK, 111-119.
- Villella, M. L. B., and Prates, R. O. 2015. Supporting Designers in Modeling Privacy for Social Network Sites. In Proceedings of 15th Brazilian Symposium on Human Factors in Computing Systems. (IHC'15). November 03-06, 2015.
- Willinger, W., Rejaie, R., Torkjazi M., Valafar, M., Maggioni, M. 2010. Research on online social networks: time to face the real challenges. SIGMETRICS Perform. Eval. Rev. 37, 3, pp. 49--54 (2010)
- Xavier S. I. R., Maria L. B. Villela, Raquel O. Prates, Marcos O. Prates, Antônio A. P. Prates, and Alexandre A. Cardoso. 2014. Migrando das redes sociais offline para as redes sociais online: o que houve com a privacidade?. In Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems (IHC '14). Sociedade Brasileira de Computação, Porto Alegre, Brazil, Brazil, 265-274.
- Young, A. L., and Quan-Haase A. 2009. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In Proceedings of the fourth international conference on Communities and technologies (C&T '09). ACM, New York, NY, USA, 265-274.

Uma Análise do Impacto do Anonimato em Comentários de Notícias Online

Julio Reis, Manoel Miranda, Lucas Bastos, Raquel Prates, Fabrício Benevenuto

¹Universidade Federal de Minas Gerais (UFMG) – Brasil

{julio.reis, manoelrmj, lucasmbastos, rprates, fabricio}@dcc.ufmg.br

Abstract. *In the last years, with the growth of the Internet as a mean of communication, a large number of people has changed the way they consume news, replacing the traditional newspapers and magazines by virtual and online versions and/or web blogs. One of the reasons online news have attracted users is because of the opportunity to comment and express their feelings and opinions about them. In this work, we analyze the content of comments of the online newspaper Reuters news in which we consider the users can comment anonymously, and Facebook in which comments are associated to an identity, often their real world identity. Our goal is to investigate whether users behave differently considering their anonymity or not. More specifically, we intend to verify whether there are significant linguistic differences between comments posted in an anonymous environment (Reuters) and in an environment in which their identity is usually known (Facebook). In addition, we propose to analyze the reactions of users when they are anonymous or not, to headlines with different sentiment associated with them. Our results indicate that linguistic characteristics change according anonymity, and also that users tend to be more negative when they are anonymous.*

Resumo. *Nos últimos anos, com o crescimento da Internet como meio de comunicação, um grande número de pessoas mudou a maneira de consumir notícias, substituindo os tradicionais jornais e revistas impressos por suas versões virtuais online e/ou web blogs. Uma das razões pelas quais as notícias online têm atraído usuários é a oportunidade de comentar e expressar seus sentimentos e opiniões sobre as mesmas. Neste trabalho, analisamos o conteúdo de comentários online postados em notícias do jornal Reuters, no qual consideramos que o usuário posta anonimamente, e do Facebook, onde o usuário está associado a uma identidade, frequentemente a sua identidade no mundo físico. Nosso objetivo é investigar se existe variação no comportamento do usuário considerando o seu anonimato ou não. Mais especificamente, trabalhamos no sentido de verificar se existem diferenças linguísticas significativas entre o comentário postado no ambiente onde é conferido ao usuário certo grau de anonimato (jornal Reuters) e no ambiente onde espera-se que este grau de anonimato não exista (Facebook). Além disso, nos propomos a analisar as reações dos usuários quando eles estão anônimos ou não, para manchetes com diferentes sentimentos associados a elas. Nossos resultados indicam que o conteúdo de comentários postado em notícias pelos usuários no ambiente anônimo e não anônimo, possuem características linguísticas particulares. Além disso, mostramos que o usuário tende a ser mais negativo quando encontra-se anônimo.*

1. Introdução

Nos últimos anos, com o crescimento da Internet como meio de comunicação, um grande número de pessoas mudou a maneira de consumir notícias, substituindo os tradicionais jornais e revistas impressos por suas versões virtuais online e/ou *web blogs* [Mitchelstein and Boczkowski 2009, Allan 2006, Tewksbury 2005]. Atualmente, as notícias online têm sido responsáveis pela captação ou atração de uma quantidade significativa de usuários na Internet [Tatar et al. 2014], possibilitando que o consumo de informação nestes ambientes seja cada vez mais social e interativo [Purcell et al. 2010].

Existem várias características de sistemas online em geral que permitem a participação do usuário, entretanto, a interação por meio de comentários e espaços para debates têm sido amplamente oferecida em jornais online [Domingo et al. 2008], e como consequência vem se tornando a forma mais comum de participação dos leitores neste cenário [Emmer et al. 2011]. Os comentários online em sites de notícias são fóruns particularmente importantes na esfera pública uma vez que possuem a capacidade de proporcionar a discussão de questões coletivamente relevantes para a sociedade, possibilitando a formação de opinião pública [Gerhards and Schäfer 2010]. Vários usuários são atraídos pela oportunidade de comentar e expressar seus sentimentos e opiniões sobre uma notícia, o que nos fornece, por meio desses fóruns de discussão, indicações da opinião, pensamentos e sentimentos de leitores [Kothari et al. 2013].

As notícias online, de forma geral, também têm atraído comentários que se propagam em outros meios de comunicação social, como por exemplo, nas redes sociais. Neste cenário, é comum nos depararmos com compartilhamentos e comentários que relacionam ou fazem referência a uma notícia específica [Kothari et al. 2013]. Inclusive alguns sites de jornais online incluem instalações que permitem o compartilhamento e comentários via redes sociais como Facebook ou Twitter [Diakopoulos and Naaman 2011]. Porém, a natureza dos comentários em redes sociais é consideravelmente diferente daqueles postados em sites de notícias [Kothari et al. 2013].

Fóruns online existem desde o início da Internet, no passado, eles eram muitas vezes dedicados a certas questões ou temas sensíveis e sua população de usuários era relativamente pequena e muitas vezes limitada para usuários tecnicamente sofisticados com preocupações específicas. No Facebook, por sua vez, espera-se que cada usuário envie uma mensagem utilizando uma única identidade online (conta), que corresponde à sua identidade no mundo físico. Em contraste, no ambiente de um jornal online, um usuário pode enviar mensagens públicas usando qualquer nome de conta do utilizador, o que pode não deixar clara a sua real identidade podendo proporcionar assim um certo grau de anonimato para a postagem de mensagens por usuários. Trabalhos recentes evidenciam que o anonimato estimula o discurso negativo [Wang et al. 2014], entretanto, pouco se sabe sobre o impacto do anonimato no comportamento do usuário no contexto de comentários em notícias online. É parte desta lacuna que endereçamos neste trabalho.

Nosso objetivo é investigar se existe variação no comportamento do usuário considerando o anonimato ou não. Parte do nosso esforço depende do uso de análise de sentimentos como uma forma para capturar o sentimento expresso nos comentários de notícias online. Mais especificamente, trabalhamos no sentido de verificar se existem diferenças linguísticas significativas entre o comentário postado no ambiente onde é conferido ao usuário certo grau de anonimato (jornal Reuters) e no ambiente onde espera-se que este grau de anonimato não exista (Facebook). Além disso, nos propomos a analisar

as reações dos usuários quando eles estão anônimos ou não, para manchetes com diferentes sentimentos associados a elas. Por exemplo, será que a polaridade do sentimento dos comentários dos usuários inseridos nestes ambientes (e.g. anônimo – Reuters, não anônimo – Facebook) refletem a polaridade das manchetes (notícias)¹?

Para respondermos a estas questões, analisamos o conteúdo de cerca de 360 mil comentários postados em notícias do jornal Reuters, onde consideramos o anonimato do usuário, e no Facebook, onde o usuário normalmente possui um identificador associado à sua identidade no mundo físico. Nossos resultados evidenciam que o conteúdo de comentários postados em notícias pelos usuários no ambiente anônimo e não anônimo, possuem características linguísticas particulares. Além disso, mostramos que o usuário tende a ser mais negativo quando encontra-se anônimo. Nossas observações têm aplicações para concepção ou manutenção de sistemas que apoiem a interação do usuário neste contexto, uma vez que fornece indícios do comportamento dos usuários nestes cenários. Em última análise, nossos resultados contribuem para a compreensão dos ambientes de comunicação mediada contemporâneos, bem como a sociedade em que vivemos.

O restante desse artigo está organizado da seguinte forma. A Seção 2 descreve os trabalhos relacionados. Em seguida, na Seção 3, descrevemos nossa metodologia experimental elaborada para este trabalho. Depois, na Seção 4, apresentamos e discutimos os resultados. Por fim, a Seção 5 apresenta conclusões e direções para trabalhos futuros.

2. Trabalhos Relacionados

Antes do surgimento da Internet, o efeito do anonimato sobre o comportamento do usuário já atraía diversos pesquisadores. Em 1969 por exemplo, por meio de um experimento com humanos, P. Zimbardo mostrou que as pessoas se tornam mais agressivas ou violentas em ambientes menos limitados por normas sociais [Zimbardo 1969]. Com o advento da Internet e a crescente popularidade de sistemas de computação, foi permitido ao usuário recriar algumas de suas convenções comportamentais sociais do mundo real no ambiente online. Para essa interação com o conteúdo na Web, o usuário normalmente faz uso de um identificador único que o representa dentro do contexto específico [Jøsang and Pope 2005]. Entretanto, grande parte destes cadastros requerem somente informações arbitrárias de perfil (e.g. email, nome), e este aspecto pode ser utilizado para preservar diversos aspectos do usuário (e.g. privacidade), dentre eles o anonimato, cujo impacto continua sendo analisado em diferentes contextos. Em [Moore et al. 2012], por exemplo, o anonimato em posts de um fórum (ou seja, mensagens em fóruns sem nenhum identificador) foi relacionado como um importante critério para se distinguir/identificar ataques na Internet. Já em [Silva et al. 2016] uma rede social anônima é utilizada como cenário para realização de um estudo sistemático que mede em larga escala os principais alvos de discurso de ódio nos meios de comunicação social online.

Para melhor compreensão da diferença entre o conteúdo postado em mídias anônimas e em outras mídias sociais, o trabalho realizado em [Correa et al. 2015] caracterizou mensagens postadas pelos usuários no *Whisper*² e no *Twitter*³. Com o uso do *Amazon Mechanical Turk*⁴, os autores introduziram o conceito de nível de anonimato para de-

¹É válido ressaltar que a polaridade da notícia é obtida por meio da inferência da polaridade da manchete da mesma, considerando sua importância já atestada e discutida em trabalhos anteriores [Reis et al. 2015].

²<https://whisper.sh/>

³<https://twitter.com/>

⁴<https://www.mturk.com/>

terminado texto e a análise revelou que este aspecto, em alguns casos, não é binário. Formalmente, os pesquisadores realizaram um experimento com seres humanos e propuseram uma métrica que representa o teor do anonimato de determinado texto a partir de uma percepção do próprio usuário. Além disso, diferenças linguísticas entre os conteúdos anônimos e não anônimos foram identificadas, e essas características foram utilizadas para treinamento de um classificador que distingue ambos os tipos.

Já o contexto de notícias online especificamente, tem sido bastante explorado em trabalhos recentes (e.g [Reis et al. 2014b, Reis et al. 2014a, Elgesem et al. 2016, Buhl et al. 2016]). Uma pesquisa com 583 jornalistas dos EUA, que fundamentou-se na formação social das teorias tecnológicas, é apresentada em [Nielsen 2014]. O objetivo do referido trabalho era investigar o impacto dessas tecnologias no comportamento do usuário, incluindo cobertura e interação dos usuários no contexto de notícias. Os resultados mostram que os jornalistas não se opõem ao fato de que as plataformas web forneçam mecanismos que permitam o compartilhamento de comentários pelos leitores, entretanto consideram que o anonimato do usuário impulsiona a interação dos leitores entre si.

O impacto do anonimato em comentários foi também explorado em trabalhos recentes [Santana 2014, Cho and Kwon 2015]. Enquanto [Cho and Kwon 2015] correlaciona o anonimato com a qualidade dos comentários, [Santana 2014] compara comentários postados por usuários anônimos e não anônimos investigando o nível de civilidade em ambos. Os resultados fornecem uma evidência empírica de que o anonimato tem efeito nos comentários dos usuários. Já em [Diakopoulos and Naaman 2011] são caracterizados comentários postados no SacBee.com. O estudo apresenta uma discussão sobre as motivações dos usuários para leitura e escrita de comentários em notícias online, além de diferentes perspectivas para a gestão da qualidade dentro do contexto jornalístico.

Neste trabalho damos um foco diferente na direção de analisar o impacto do anonimato no comportamento do usuário. Considerando o contexto de notícias online, investigamos o impacto do anonimato com foco na polaridade expressa nestes comentários. Além disso, inspirados por algumas análises efetuadas para o contexto de redes sociais, investigamos características linguísticas que diferenciam o comportamento dos usuários nestes dois cenários (e.g. anônimo e não anônimo). Nossos resultados fornecem importantes evidências do comportamento humano sobre essas perspectivas com inúmeras aplicações para projetos futuros, tais como sistemas que atuem neste contexto.

3. Metodologia

Para que fosse possível analisar o impacto do anonimato nos comentários postados pelos usuários, inicialmente, coletamos dados de um ambiente anônimo e outro não anônimo. Em seguida, definimos uma estratégia para caracterização do conteúdo nestes dois cenários e inferimos o sentimento de cada um dos registros. A seguir, descrevemos detalhes relacionados à metodologia experimental adotada.

3.1. Base de Dados

Foram utilizados comentários oriundos do jornal Reuters Online⁵, onde o usuário pode postar mensagens utilizando qualquer nome de conta que na maioria dos casos não deixa clara a sua real identidade (ambiente anônimo), e do Facebook⁶, onde o usuário possui

⁵<http://www.reuters.com>

⁶<http://www.facebook.com>

Tabela 1. Sumário de informações dos datasets

Fonte	Período de Coleta	# Notícias	# Comentários
Reuters	Jan/2013-Set/2014	27.051	166.329
Facebook	Mar/2014-Mar/2014	617	195.157

uma identidade, que normalmente corresponde à sua identidade no mundo físico (ambiente não anônimo).

A agência Reuters, em seus vários idiomas de publicação, ficou conhecida por reportar furos jornalísticos estrangeiros. Atualmente com sede em Nova Iorque, a empresa tem expandido seus produtos de negócio e rede de reportagem global para incluir serviços de mídia, financeiros e econômicos e tem abrangência mundial. Já o Facebook, é a rede social mais utilizada no mundo atualmente⁷. As relações entre seus bilhões de usuários têm atraído a atenção de vários pesquisadores no mundo inteiro, como se pode atestar pelo grande número de pesquisas recentes realizadas no Facebook (e.g. [Ellison et al. 2007, Hayes et al. 2016, Marder et al. 2016])⁸.

3.2. Processo de Coleta

A coleta dos dados do jornal Reuters foi realizada com a execução de um *web crawler* cujo propósito era abranger comentários de notícias em destaque no referido jornal, permitindo uma análise mais robusta considerando o objetivo de pesquisa proposto. Acessamos o *archive* deste jornal online e selecionamos o “*Top News*” de publicações para realização de uma coleta de dados retroativa. Para cada uma das notícias recuperadas, foi realizado um *parser* do conteúdo retornado e recuperadas informações relativas às notícias e aos comentários.

Já a coleta de dados no Facebook, foi realizada por meio da API oficial da rede⁹. Foram selecionadas páginas oficiais de jornais conhecidos internacionalmente¹⁰. A partir daí, acessamos os posts destas páginas que compartilharam notícias. Para cada notícia recuperada, coletamos todos os comentários inseridos pelos usuários, além de informações relativas às próprias notícias. A Tabela 1 sumariza informações relacionadas ao conjunto de dados coletado.

3.3. Caracterização

Com o objetivo de identificarmos características próprias de comentários em ambiente anônimo e não anônimo, damos um passo na caracterização da nossa base de dados nestes cenários. Neste trabalho, primeiramente propomos uma caracterização de um conjunto de comentários coletados com foco em características linguísticas deste conteúdo. Nosso propósito é entendermos o uso de atributos linguísticos (e.g. estrutura e gramática das mensagens) em comentários postados nos dois ambientes analisados. Para isso, utilizamos como estratégia para essa caracterização a ferramenta *Linguistic Inquiry and Word*

⁷De acordo com: <http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> - última visita em março de 2016.

⁸Uma rápida pesquisa por artigos publicados a partir de 2012 contendo a palavra “Facebook” retorna mais de 3800 resultados.

⁹<https://developers.facebook.com/>

¹⁰Reuters (<https://www.facebook.com/Reuters>), BBC News (<https://www.facebook.com/bbcnews>), DailyMail (<https://www.facebook.com/DailyMail>) e The NY Times (<https://www.facebook.com/nytimes>)

Tabela 2. Exemplos de comentários rotulados pelo SentiStrength

Força de Polaridade	Comentário
≤ -3	No hope, this people love to kill their own ...very sad
-2	Well, it might be a headache for US sooner or later. These are your puppets sweet US
0	Only what u do for God will last forever
2	The world feels like a better place already.
≥ 3	We will never forget you !!! TOTAL RESPECT !!!

Count (LIWC) [Tausczik and Pennebaker 2010]. O uso dessa ferramenta¹¹ é importante e adequado para nosso estudo pois ele nos permite estimar componentes emocionais, cognitivos, estruturais e gramaticais de um texto fornecido como entrada, baseado na utilização de dicionários contendo palavras e categorias associadas a cada uma.

3.4. Inferência de Sentimento

Depois de definirmos a estratégia para caracterização linguística do conteúdo, propomos um procedimento para inferência do sentimento de cada um dos comentários postados no jornal Reuters e no Facebook. Esforços recentes mostram que não existe um método de análise de sentimentos que seja mais adequado para todos os tipos de dados [Gonçalves et al. 2013]. Assim, ao invés de utilizarmos um método de análise de sentimentos popular, nós brevemente investigamos qual o método seria mais adequado para o nosso conjunto de dados. Para isso selecionamos aleatoriamente do nosso conjunto de dados, sumarizado na Tabela 1, 100 comentários aleatórios com quantidades proporcionais para cada um dos cenários, que foram rotulados manualmente por 3 voluntários do nosso laboratório de pesquisa como positivo, negativo ou neutro. Cada um dos voluntários rotulou as mensagens de forma independente, sem a influência dos demais, permitindo o alcance de um percentual de concordância 92%.

Nós utilizamos este pequeno conjunto de dados rotulado como uma *baseline* para a comparação de desempenho de 8 métodos de análise de sentimento implementados e disponibilizados no sistema *iFeel* [Araújo et al. 2014, Araújo et al. 2016]. O *Sentistrength* [Thelwall et al. 2010] foi o método de análise de sentimentos com o melhor desempenho para o cenário em questão, alcançando uma acurácia de 74% para comentários. Outros métodos apresentaram resultados competitivos com os do *Sentistrength*, entretanto, uma propriedade interessante deste método é que ele mede a força da polaridade do sentimento em textos em uma escala de -5 (muito negativo) a 5 (muito positivo). Acreditamos que isso nos forneça uma perspectiva interessante sobre a intensidade da opinião expressa nos comentários. A Tabela 2 fornece exemplos de comentários, em ambos os cenários, classificados de acordo com a pontuação do *Sentistrength* (optamos por manter a nomenclatura no idioma original, em inglês). Não houve mensagens rotuladas com pontuação 1 e -1 em nosso conjunto de dados.

Uma das limitações deste trabalho está relacionada ao fato de que o conjunto de comentários não representa de forma balanceada o conjunto de notícias. Por exemplo, os comentários coletados no Facebook compreendem mais de um jornal. Entretanto acreditamos que ao analisarmos dados relativos ao mesmo contexto (notícias online), somos capazes de fornecer uma perspectiva geral do impacto do anonimato no comportamento do usuário de forma geral. Para a realização de comparações mais minuciosas, realizamos alguns tratamentos que serão descritos durante a apresentação dos resultados.

¹¹O software está disponível em <http://www.liwc.net/>.

4. Resultados e Discussões

Nesta seção, apresentamos os resultados da análise de dados de comentários de notícias coletados do jornal Reuters e do Facebook.

4.1. Características Linguísticas do Conteúdo

Um dos nossos objetivos neste trabalho era comparar características linguísticas de comentários de notícias postadas pelos usuários no ambiente anônimo (Reuters) e não anônimo (Facebook). Para isso, inicialmente medimos a frequência do número de palavras nos comentários em cada um dos cenários e apresentamos o resultado na Figura 1. Percebemos que para ambos os casos, as curvas obtidas são de cauda longa, entretanto, existem muito mais mensagens curtas (até 100 palavras para Reuters e 50 para Facebook) que longas. Entretanto, as mensagens postadas no Facebook tendem a ser ainda mais curtas do que mensagens postadas diretamente no jornal online.

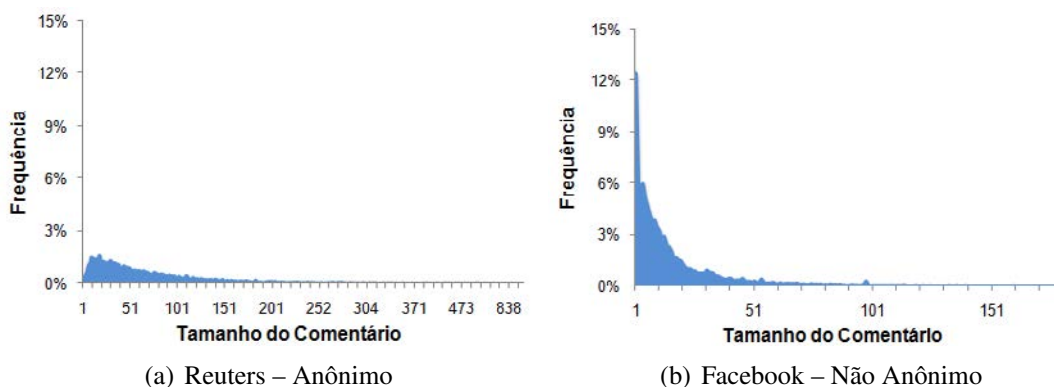


Figura 1. Frequência do número de palavras contidas nos comentários em cada cenário

Em seguida, comparamos e contrastamos amostras de dados destes dois ambientes, conforme detalhamento apresentado em seções anteriores, com a utilização do LIWC [Tausczik and Pennebaker 2010]. Da base de dados apresentada na Tabela 1, nós selecionamos um subconjunto de notícias que possuíam comentários nos dois ambientes: Reuters e Facebook. Para as análises do restante desta subseção, selecionamos uma amostra aleatória de 10 mil comentários inseridos em notícias do jornal Reuters, e 10 mil comentários no Facebook associados a notícias deste subconjunto. As demais análises apresentadas nas subseções seguintes serão efetuadas com todo o conjunto de dados. Em seguida, executamos a ferramenta LIWC e obtivemos o percentual de ocorrência de palavras associadas a cada uma das 85 categorias, para os dois cenários (i.e. anônimo – Reuters, e não anônimo – Facebook).

O LIWC consiste em um conjunto de palavras agrupadas em categorias associadas a quatro classes gerais: *Linguistic Processes* (LP) (e.g., advérbios, pronomes), *Psychological Processes* (PP) (e.g., emoções negativas e positivas), *Personal Concerns* (PC) (e.g, trabalho), e *Spoken Categories* (SC). Além, disso a ferramenta nos fornece taxas de ocorrência de *Punctuation* (PT) (e.g, vírgula, exclamação).

Considerando que o nosso propósito é investigar características intrínsecas ao comentário postado nestes ambientes, focamos nos dicionários ou categorias do LIWC que apresentaram um percentual de variação mais significativo. Por exemplo, para a categoria

Money o percentual de ocorrência para os comentários no Facebook foi de 1,73 enquanto para os dados de comentários no jornal Reuters o percentual foi de 0,78. Neste caso, a razão entre os valores é de 2,21 (1,73/0,78), o que indica que a ocorrência de palavras da categoria *Money* é 2,21 vezes maior em comentários de notícias no Facebook (em relação aos dados de comentários do jornal Reuters). Com base no exemplo apresentado, definimos como significativa a classe cuja razão é menor que 0,5 (ocorrência menor ou igual a metade) ou maior que 2 (ocorrência maior ou igual ao dobro). Acreditamos que isso nos forneça um indício mais preciso de categorias que realmente sejam úteis para a diferenciação dos conteúdos analisados neste trabalho. Os resultados obtidos para as classes significativas do LIWC são apresentados na Tabela 3.

Podemos perceber uma grande expressividade de classes associadas à categoria de preocupações pessoais (*Personal Concerns*) da ferramenta utilizada. O comentário postado em notícias de forma não anônima (Facebook), tem presença significativa de palavras relacionadas ao dicionário de dinheiro do LIWC (*Money*), o que pode ser um indício de que este ambiente impulse discussões mais profissionais e/ou pessoais. Já para o cenário anônimo, percebemos a ocorrência considerável de palavras associadas aos dicionários de religião e morte (*Religion, Death*) do LIWC.

Para ilustrarmos esta situação, mostramos por meio da Figura 2 visualizações de árvores de palavras¹² para os termos “*Money*”, “*Religion is*” e “*Death*”, usando o nosso conjunto de dados analisado. As referidas visualizações nos apresentam frases que se ramificam a partir de um termo raiz definido. As palavras mais frequentes são apresentadas em fonte maior. É possível notarmos por meio das frases formadas, que se mantêm as considerações efetuadas anteriormente. Para a classe “*Money*” (ambiente não anônimo), estão presentes frases como “*money from companies or interest groups aligned against popular opinion and politically correctness.*”, que apresentam teor mais profissional. Já para Reuters frases como “*religion is a human invention. True, but not so concerning the Word God.*” e discussões sobre “*Death penalty*” (classes *Religion* e *Death* respectivamente), reforçam o comportamento mais negativo do usuário para a discussão de temas mais sensíveis quando o mesmo encontra-se anônimo.

Esta análise corrobora resultados apresentados por [Zimbardo 1969] que destacam que o comportamento dos seres humanos é mais negativo em um ambiente menor limitado por normas sociais. Outro aspecto observado está relacionado ao uso intenso de pontuação em comentários não anônimos (Facebook), o que pode indicar que o usuário esteja um pouco mais preocupado com a formalidade neste tipo de ambiente (em contraste com o ambiente anônimo), onde seu perfil e relacionamentos correspondem, em grande parte dos casos, à sua realidade no mundo real.

Tabela 3. Análise das classes significantes do LIWC

Dicion./Classe LIWC	Anôn./Não Anôn. (%)	Anônimo (%)	Não anônimo(%)	Exemplo de Palavras	Categoria LIWC
<i>Money</i>	2,22	1,73	0,78	Audit, cash, owe	<i>Personal Concerns</i>
<i>Religion</i>	0,36	0,43	1,19	Altar, church, mosque	
<i>Death</i>	0,44	0,44	0,99	Bury, coffin, kill	
<i>AllPet</i>	0,48	13,93	28,98	-	<i>Punctuation</i>

¹²<https://www.jasondavies.com/wordtree/>

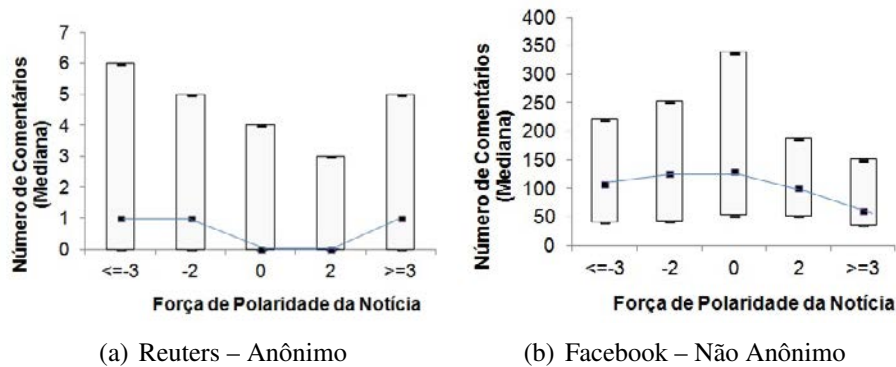


Figura 3. Boxplot para o número de comentários em função do valor da força da polaridade da notícia nos dois cenários

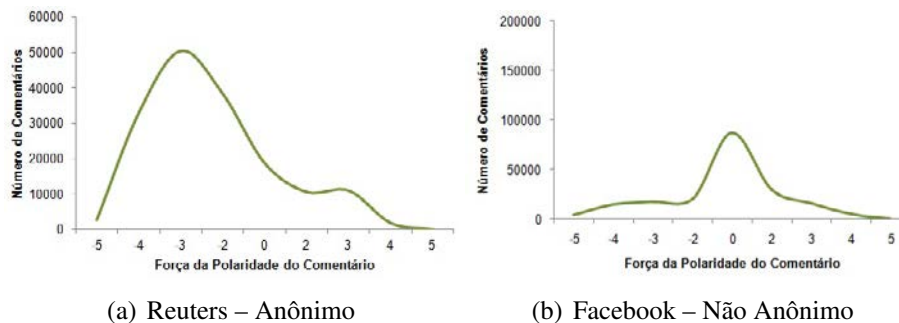


Figura 4. Número de comentários por valor de força de polaridade dos comentários nos dois cenários

histograma de comentários por valor de força da polaridade para cada um dos cenários, conforme representado na Figura 4. Nosso objetivo nesta etapa é entendermos qual a polaridade esperada para comentários inseridos sobre um mesmo grupo de notícias, considerando o usuário nestes dois ambientes (anônimo e não anônimo).

Podemos observar que, para o jornal Reuters em particular, a grande maioria dos comentários possui força de polaridade negativa. Em resumo, 75% dos comentários possuem polaridade negativa, 11% são neutros, e 14% têm força de polaridade positiva. Já para o Facebook, percebemos que essas proporções são distribuídas de forma distinta. Especificamente para este cenário, 74% dos comentários possuem polaridade neutra, 22% são positiva e apenas 2% têm a força de polaridade negativa.

Os resultados identificados, considerando o ambiente onde o usuário encontra-se anônimo (Reuters), corroboram o conhecimento construído por trabalhos realizados anteriormente [D’Costa 2013, Moosa 2014]. Por meio de uma análise mais ampla do ambiente online, os autores afirmam que a seção de comentários é geralmente “casa ou local para hostilidade”. D’Costa (2013) explica que, quando pessoas de diferentes culturas e origens interagem, eles podem ler algo que desafia sua visão de mundo e, em seguida, a lógica e a razão desaparecem. Tudo o que importa é que eles permitam que você saiba o quão errado você, o tópico, a perspectiva ou o assunto são. E de repente, a seção de comentários tornou-se um discurso retórico. De maneira distinta podemos observar que, para o cenário onde o usuário não está anônimo (Facebook), o número de comentários negativos inseridos pelo usuário é bem inferior, o que evidencia que o comportamento do

usuário é variável de acordo com o ambiente e impactado por fatores como anonimato.

A fim de verificar se o valor de força da polaridade do comentário em ambos os cenários é correlacionado com (ou consequência de) o valor da força de polaridade da manchete, elaboramos as Tabelas 4 e 5 que apresentam matrizes de confusão entre a polaridade da manchete (título) e a polaridade do comentário postado naquela referida notícia, para o Reuters (ambiente anônimo) e o Facebook (ambiente não anônimo), respectivamente. Note que para o jornal Reuters, no qual consideramos o usuário anônimo, os comentários com polaridade negativa são publicados de forma independente da polaridade da notícia (manchete). Ou seja, a porcentagem de comentários negativos é significativamente mais elevada para todas as classes de polaridades de notícias. Isso indica que, mesmo que o título esteja resumindo uma história de polaridade positiva, os comentários são de polaridade negativa em mais de 70% dos casos. É curioso notarmos que as notícias neutras são as que recebem o maior percentual de comentários negativos. As notícias de polaridade positiva, por sua vez, receberam mais comentários positivos, ainda que poucos.

Para o Facebook, onde consideramos o não anonimato do usuário, podemos notar uma situação mais uniforme em relação à situação do jornal Reuters em termos de proporção, com grande expressividade de comentários neutros em todas as classes de polaridade das notícias. Tal situação indica que para uma notícia negativa o usuário tende a postar mais comentários negativos (41,12% dos casos), e de forma similar ocorre para as notícias neutras que recebem comentários neutros em mais de 55% dos casos. Já as notícias positivas recebem praticamente a mesma proporção de comentários neutros e positivos (cerca de 39,2%). Finalmente, é importante salientar que, em nosso conjunto de dados, mais de 85% dos usuários postaram em média, em ambos os cenários, apenas um comentário por notícia, o que descarta a possibilidade de que este resultado seja influenciado por usuários que inserem muitos comentários em uma determinada (ou única) notícia.

		<i>Comentários</i>		
		Negativo	Neutro	Positivo
<i>Manchete (Títulos)</i>	Negativa	71,30%	12,71%	16,00%
	Neutra	76,55%	10,46%	12,99%
	Positiva	70,43%	11,37%	18,19%

Tabela 4. Matriz de confusão entre a polaridade da manchete e a polaridade do comentário postado na notícia – Reuters

		<i>Comentários</i>		
		Negativo	Neutro	Positivo
<i>Manchete (Títulos)</i>	Negativa	41,12%	34,53%	24,35%
	Neutra	20,61%	55,07%	24,32%
	Positiva	21,56%	39,25%	39,20%

Tabela 5. Matriz de confusão entre a polaridade da manchete e a polaridade do comentário postado na notícia – Facebook

5. Conclusões e Trabalhos Futuros

Nos dias de hoje as notícias online têm atraído uma quantidade significativa de usuários na Internet. Essas notícias são publicadas de diferentes formas como por exemplo, na

página do próprio jornal ou disseminadas por meio das redes sociais. A interação do usuário nestes ambientes por meio de comentários tem sido cada vez mais presente, e tem possibilitado a discussão sobre diversos assuntos.

Neste trabalho investigamos o impacto do anonimato no comportamento do usuário por meio da análise de mais de 360 mil comentários inseridos em notícias do jornal Reuters e em notícias postadas no Facebook. Endereçamos o ambiente do jornal Reuters com certo grau de anonimato considerando que neste contexto o usuário pode utilizar um identificador qualquer (fraco) que não está associado, necessariamente, à sua identidade no mundo físico. Já no Facebook consideramos que o usuário não está anônimo uma vez que espera-se que seu identificador e círculos de amizade representem sua identidade e relacionamentos reais do mundo *off-line* (identificador forte). Nossos resultados revelam aspectos interessantes sobre o impacto do anonimato no comportamento do usuário considerando o ambiente de notícias online.

Primeiramente, caracterizamos o conteúdo postado nestes dois cenários com objetivo de identificarmos características linguísticas que distingam o comentário postado pelo usuário considerando o fato de que o mesmo esteja anônimo ou não. Para um subconjunto de notícias online comentadas nos dois ambientes percebemos que o comentário postado pelos usuários anônimos possui características linguísticas particulares. Neste estudo descobrimos que comentários postados no Facebook tendem a ser mais curtos, entretanto notícias publicadas neste ambiente são mais comentadas em relação às notícias veiculadas no próprio jornal. Além disso, mostramos presença significativa de palavras relacionadas à classe de preocupações pessoais do dicionário do LIWC, além do diferente uso de pontuações em contraste com comentários postados no ambiente não anônimo. Essas características poderiam ser utilizadas por exemplo, na construção de algoritmos que identifiquem e façam a separação de um conteúdo anônimo de um não anônimo. Esta estratégia já tem sido explorada de forma mais genérica em trabalhos que treinam classificadores para identificação de conteúdo anônimo [Correa et al. 2015], mas poderia ser também explorada para cenários mais específicos.

De forma geral, nossos resultados geram indicadores de que o anonimato pode ser um fator de impacto no comportamento dos usuários quando analisamos comentários postados pelos mesmos no contexto de notícias online. Tal situação corrobora estudos realizados na psicologia social que mostram que o anonimato influencia fortemente o comportamento do usuário – online e *off-line*. Seres humanos se tornam mais agressivos ou violentos (negativos) em um ambiente menos limitado por normas sociais [Zimbardo 1969], além de que este contexto proporciona desinibição do usuário [Pinsonneault and Heppel 1997, Suler 2004]. Alguns trabalhos argumentam que a falta de prestação de contas no ambiente anônimo estimula o discurso negativo [Wang et al. 2014]. Acreditamos que o não anonimato do usuário talvez o reprima quanto a assumir determinadas escolhas.

Neste trabalho damos apenas um primeiro passo no entendimento deste universo, entretanto entendemos que possam haver outros fatores, além do anonimato, que podem influenciar o comportamento dos usuários nestes ambientes, como por exemplo, o fato de que no Facebook o usuário saiba quem é a sua audiência (normalmente sua rede de amigos), podendo assim evitar falar de coisas ruins para não passar uma imagem negativa para seu meio social. Em contrapartida, no jornal Reuters, a audiência esperada é composta por outros leitores que o usuário não necessariamente conhece, e ainda que estes

outros leitores o conheçam não quer dizer que consigam identificá-lo. Ainda assim, esse trabalho contribui revelando várias observações interessantes sobre o impacto do anonimato no comportamento do usuário neste contexto, fornecendo características que podem ser utilizadas para a concepção de novos sistemas, e finalmente, para compreendermos a sociedade em que vivemos.

Como trabalhos futuros, para fortalecimento de nossos indicadores, pretendemos analisar de forma mais aprofundada a relação do anonimato com outros possíveis fatores de influência do comportamento do usuário, inclusive, levando em consideração outros contextos em que notícias são comentadas de forma anônima ou não. Além disso, pretendemos explorar o impacto do anonimato no comportamento dos usuários de forma mais qualitativa. Entendemos, por exemplo, que uma análise dos traços de personalidade dos usuários talvez possa revelar aspectos interessantes que nos permitam compreender de forma mais aprofundada os fatores de impacto do comportamento do usuário diante desta situação. Por fim, acreditamos que a metodologia que apresentamos neste trabalho pode facilitar ou apoiar diferentes esforços de investigação neste domínio impulsionando assim a produção de conhecimento.

Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto FAPEMIG - PRONEX - MASWeb, Modelos, algoritmos e sistemas para a Web (número do processo APQ - 01400-14), e por projetos individuais do CNPq, CAPES e Fapemig. Finalmente, os autores gostariam de agradecer Leandro Araújo pelas inspiradoras discussões no estágio inicial desse trabalho.

Referências

- Allan, S. (2006). *Online News: Journalism And The Internet: Journalism and the Internet*. McGraw-Hill Education. McGraw-Hill Education.
- Araújo, M., Diniz, J. P., Bastos, L., Soares, E., Júnior, M., Ferreira, M., Ribeiro, F., and Benevenuto, F. (2016). ifeel 2.0: A multilingual benchmarking system for sentence-level sentiment analysis. In *Proceedings of the International AAAI Conference on Web-Blogs and Social Media*, Cologne, Germany.
- Araújo, M., Gonçalves, P., Benevenuto, F., and Cha, M. (2014). ifeel: A system that compares and combines sentiment analysis methods. In *WWW (Companion Volume)*. International World Wide Web Conference (WWW'14).
- Buhl, F., Günther, E., and Quandt, T. (2016). Observing the dynamics of the online news ecosystem: News diffusion processes among german news sites. *Journalism Studies*, pages 1–26.
- Cho, D. and Kwon, K. H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. *Computers in Human Behavior*, 51:363–372.
- Correa, D., Silva, L. A., Mondal, M., Benevenuto, F., and Gummadi, K. P. (2015). The many shades of anonymity: Characterizing anonymous social media content. In *Ninth International AAAI Conference on Web and Social Media (ICWSM)*.
- D'Costa, K. (2013). Don't read the comments! (Why do we read the online comments when we know they'll be bad?). *Scientific American*.

- Diakopoulos, N. and Naaman, M. (2011). Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 133–142. ACM.
- Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J. B., and Vujnovic, M. (2008). Participatory journalism practices in the media and beyond: An international comparative study of initiatives in online newspapers. *Journalism practice*, 2(3):326–342.
- Elgesem, D., Feinerer, I., and Steskal, L. (2016). Bloggers' responses to the snowden affair: Combining automated and manual methods in the analysis of news blogging. *Computer Supported Cooperative Work (CSCW)*, pages 1–25.
- Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of facebook "friends": social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12(4):1143–1168.
- Emmer, M., Vowe, G., and Wolling, J. (2011). *Bürger onlin*. Konstanz: UVK.
- Gerhards, J. and Schäfer, M. S. (2010). Is the internet a better public sphere? comparing old and new media in the us and germany. *New media & society*.
- Gonçalves, P., Araújo, M., Benevenuto, F., and Cha, M. (2013). Comparing and combining sentiment analysis methods. In *Proceedings of the 1st ACM Conference on Online Social Networks (COSN'13)*.
- Hayes, M., van Stolk-Cooke, K., and Muench, F. (2016). Corrigendum to understanding facebook use and the psychological effects of use across generations comput. hum. behav. 49 (2015) 507-511. *Computers in Human Behavior*, 56(C):377.
- Jøsang, A. and Pope, S. (2005). User centric identity management. In *AusCERT Asia Pacific Information Technology Security Conference*, page 77. Citeseer.
- Kothari, A., Darwish, K., Mourad, A., and Taei, A. (2013). Detecting comments on news articles in microblogs. In *Proceedings of the 7th International AAAI Conference on Web-Blogs and Social Media*.
- Marder, B., Houghton, D., Joinson, A., Shankar, A., and Bull, E. (2016). Understanding the psychological process of avoidance-based self-regulation on facebook. *Cyberpsychology, Behavior, and Social Networking*.
- Mitchelstein, E. and Boczkowski, P. J. (2009). Between tradition and change: A review of recent research on online news production. *Journalism*, 10(5):562–586.
- Moore, M. J., Nakano, T., Enomoto, A., and Suda, T. (2012). Anonymity and roles associated with aggressive posts in an online forum. *Computers in Human Behavior*, 28(3):861–867.
- Moosa, T. (2014). Comment sections are poison: handle with care or remove them. *The Guardian*.
- Nielsen, C. E. (2014). Coproduction or cohabitation: Are anonymous online comments on newspaper websites shaping news content? *New Media & Society*, 16(3):470–487.
- Pinsonneault, A. and Heppel, N. (1997). Anonymity in group support systems research: A new conceptualization, measure, and contingency framework. *Journal of Management Information Systems*, pages 89–108.

- Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., and Olmstead, K. (2010). Understanding the participatory news consumer, pew internet & american life project. Technical report, Research Report.
- Reis, J., Benevenuto, F., Vaz de Melo, P., Prates, R., Kwak, H., and An, J. (2015). Breaking the news: First impressions matter on online news. In *Proceedings of the 9th International AAAI Conference on Web-Blogs and Social Media*, Oxford, UK.
- Reis, J., Goncalves, P., Miranda, L., Prates, R., and Benevenuto, F. (2014a). A percepção dos usuários sobre a utilização de análise de sentimentos como apoio a seleção de notícias. In *Proceedings of the XIII Brazilian Symposium on Human Factors in Computing Systems*. Sociedade Brasileira de Computação.
- Reis, J., Gonçalves, P., Vaz de Melo, P. O., Prates, R., and Benevenuto, F. (2014b). Magnet news: You choose the polarity of what you read. *Proceedings of the International Conference on Web and Social Media (ICWSM)*.
- Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, 8(1):18–33.
- Silva, L., Modal, M., Correa, D., Benevenuto, F., and Weber, I. (2016). Analyzing the targets of hate in online social media. In *Proceedings of the International AAAI Conference on Web-Blogs and Social Media*, Cologne, Germany.
- Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & behavior*, 7(3):321–326.
- Tatar, A., Antoniadis, P., De Amorim, M. D., and Fdida, S. (2014). From popularity prediction to ranking online news. *Social Network Analysis and Mining*, 4(1):1–12.
- Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.
- Tewksbury, D. (2005). The Seeds of Audience Fragmentation: Specialization in the Use of Online News Sites. *Journal of Broadcasting & Electronic Media*, 49(3):332–348.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558.
- Wang, G., Wang, B., Wang, T., Nika, A., Zheng, H., and Zhao, B. Y. (2014). Whispers in the dark: analysis of an anonymous social network. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 137–150. ACM.
- Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. In *Nebraska symposium on motivation*. University of Nebraska press.

Identifying Awareness Requirements in Face-to-Face Collaborative Applications for Users with Autism Spectrum Disorders

Greis F. Mireya Silva, Alberto Barbosa Raposo

Department of Informatics
Pontifical Catholic University of Rio de Janeiro (PUC-Rio) - RJ - Brazil

{gcalpa, abraposo}@inf.puc-rio.br

***Abstract.** Individuals with Autism Spectrum Disorders (ASD), mainly those with high severity, have difficulties to identify and understand the others' actions and emotions and, therefore, to understand the basic concepts of a collaborative work, even in face-to-face situations. We developed and evaluated a multitouch collaborative game with youths with high severity in ASD. Despite the positive results to encourage the social interaction, we observed that they had difficulties to understand tasks, to identify their partners, and the workspace. Based on these difficulties, we identify a set of awareness requirements to contribute to the design of awareness mechanisms adequate to face-to-face collaborative applications for individuals with ASD.*

1. Introduction

The development of a collaborative application usually considers that users possess the intrinsic characteristics required to easily perform collaboration activities, involving its three dimensions: communication, cooperation and coordination [Ellis, Gibbs and Rein 1991]. The systems attempt to offer awareness support mechanisms so that the collaborative process is performed effectively and without hiccups. Within CSCW context, “awareness” is defined by Dourish and Bellotti (1992) as: “*An understanding of the activities of others, which provides a context for your own activity. This context is used to ensure that individual contributions are relevant to the group’s activity as a whole, and to evaluate individual actions with respect to group goals and progress*”.

A great challenge arises when it comes to design collaborative applications for individuals that present difficulties to understand the most basic concepts of a collaborative activity, including difficulty to interpret what the other person is doing or feeling, even in face-to-face situations. This is the case of individuals with ASD, who also present communication and social interaction deficits, together with restricted and repetitive behavior patterns [APA 2013, WHO 1992]. To strengthen these compromised abilities, it is very important to engage these individuals in collaborative situations, using the appropriate methods to motivate them to identify, learn and develop the activities that encompass a collaborative process. However, current collaborative systems are not designed to take into account these limitations, therefore restricting their use by these individuals.

Studies about face-to-face multitouch collaborative applications for individuals with ASD, mainly for those with low impairment in ASD [Chen 2012, Millen et al.

2011] have used different strategies [Hourcade et al. 2012, Stone et al. 2011] or restrictions regarding the user interface elements, to motivate/force collaboration among users [Battocchi et al. 2009, Bauminger et al. 2007, Gal et al. 2009, Giusti et al. 2011, Piper et al. 2006], presenting positive results by stimulating users' social interaction abilities, and the potential benefits of the technology to encourage users' gestural and verbal communication. However, even the collaborative applications developed for this public, still do not consider notions of awareness for these users that do not know how to collaborate in a typical sense. In this context, there are still many unanswered questions, especially for people with high impairment in ASD¹. For example, what are the needs of individuals with ASD within the context of collaborative work? How can we offer software that allows them to perceive tasks, actions of their peers, their own actions and the workspace? What software characteristics have to be taken into account by developers to offer learning opportunities, regarding the collaboration dimensions, for these users that do not know how to collaborate in a typical sense?

As an initial contribution to this great challenge, we have developed and evaluated a collaborative game for the interaction of two users in a multitouch tabletop, specifically developed for individuals with high severity in ASD. This game includes a set of restrictions called Collaboration Patterns. These patterns restrict, in different ways, user interaction with the interface elements, where they can share resources, exchange information and simultaneously perform activities.

In previous studies [Silva et al. 2014, 2015], we evaluated the benefits of the game in the users with ASD, and we obtained encouraging results, particularly in the user's interest in the game and in the generation of social interaction expressions. However, we observed that users have some difficulties to easily understand the collaborative activities. Therefore, our objective in this work is to identify these difficulties in order to obtain a set of requirements to guide the design of collaborative applications for individuals with ASD.

Our evaluation, in this paper, has focused on the difficulties faced by users during their interaction playing the game, especially those related to the awareness of tasks, game peers, and the context of the group as a whole. Through our analysis of such difficulties, we have verified that it is necessary to not only offer restrictions to motivate collaboration, but also opportunities for collaboration learning. These learning opportunities may be offered by including in the application some awareness mechanisms that are appropriate for users to comprehend the dimensions that involve collaborative work. The results of this evaluation have led to an awareness requirements specification aiming to guide the development of awareness mechanisms adequate to face-to-face collaborative applications designed for individuals with ASD.

In the next section, we present related works about strategies and recommendations for the design of collaborative applications in multitouch tabletops for users with ASD. Then, we present the collaborative game and the collaboration patterns

¹ Individuals with severe or marked deficits in verbal and nonverbal social communication skills cause severe impairments in functioning, very limited or limited initiation of social interactions, and minimal response to social overtures from others (according to the severity levels of the Diagnostic and Statistical Manual of Mental Disorders-DSM-5) [APA, 2013].

included in the game. Next, we detail the game evaluation process, performed with users with ASD, the obtained results, and the set of awareness requirements identified in the evaluation process. Finally, we present conclusions and future work.

2. Related work

Aiming to stimulate collaboration among individuals with ASD, various studies about collaborative multitouch applications have recommended the inclusion of restrictions in the user interface elements. This type of restriction indirectly induces the users to collaborate [Hornecker 2005], because they offer implicit suggestions to act in pre-determined ways or to adopt interaction patterns that indirectly “force” collaboration among users.

Among such studies is the game SIDES [Piper et al. 2006], which requires the collaboration of four users, in its interaction via *enforced turn-taking*, to accomplish the objective of the game. Giusti and collaborators (2011) propose a set of four restrictions called *collaborative patterns*, one of them being, for example, *choosing together pattern*, which requires the touch of two or more users to select one particular object. The authors concluded that the users learned the importance of collaboration as they advanced in the game phases, adequately responding to the collaboration strategies implemented [Weiss et al. 2011].

Similarly, other authors [Battocchi et al. 2009, Bauminger et al. 2007, Gal et al. 2009] have developed two games based on the *Enforced Collaboration Paradigm*, which consists in forcing the simultaneous interaction of two or more users on the same interface element to move it. Using this paradigm, the *Collaborative Puzzle Game* [Battocchi et al. 2009] requires that two users must move together each piece of a puzzle to assemble it; and in the *StoryTable* [Bauminger et al. 2007, Gal et al. 2009] users must create together a common story.

Even studies about applications for individuals with typical development consider that it is important to identify what has to be restricted in a collaborative interface to help people interact problem free. For example, Yuill and Rogers (2012) suggest it is necessary to consider restricting the following aspects: the *degree of awareness* of user actions and intentions, the *degree of control* over the interface, and the *degree of availability of relevant information*.

From the above mentioned related work, we may conclude that interaction restrictions are important to motivate the collaboration among individuals with ASD. The Collaboration Patterns used in the game presented in this work also include a gradual sequence of restrictions to instruct the users on the most basic aspects of collaboration.

Other studies have indicated important characteristics to be considered in the design of collaborative applications for individuals with ASD, which we classify as follows:

Fun applications. Parents of children and adults with ASD have indicated, in an interview performed by Putnam e Chong (2008), the importance of applications to be fun and include learning experiences, in addition to consider each users’ specific characteristics and sensorial integration aspects.

Simple and predictable applications. Understanding simplicity in the sense of “*removing unnecessary information and features so that technology is simpler to understand and intuitive to use*” [Keay-Bright and Howarth 2012]. Simplicity of user interface elements is mandatory, as the intent is to help users to deal with the complexities of the real work. If users are able to feel comfortable from the beginning in their interaction with the application, then it is possible to encourage them to perform more challenging tasks [Hourcade et al. 2012].

Error agnostic applications. Use of applications with no wrong answers neither error messages [Hourcade et al. 2012, Keay-Bright and Howarth 2012], allowing users not to get frustrated and keep motivated throughout the necessary interaction attempts. Keay-Bright and Howarth (2012) also suggest the exclusion of correct answers and instructions, but ASD therapists suggest that this is very important to encourage work progress, so that users perceive that they are in the right path and get even more motivated to continue interaction with the system.

Representation of people in the real world. Such representation may be done through caricatures or photographs, according to the specific characteristics of the users, as some individuals with ASD may see caricatures/avatars as an easier to understand element than common gestural expressions [Moore et al. 2005] and for recognition of emotions [Rosset et al. 2008]. Other individuals may find it easier to identify individuals through photographs or more realistic faces, as photographs may be intuitively more useful when the user is going to interact with the real world [Moore et al. 2005].

Additionally, Zancanaro and collaborators (2010) state that one should notice the following aspects to support the acquisition of social competence by individuals with ASD: consider the specific interests of users and their strong and weak abilities to generate motivation to perform the task; consider up to which point each user is capable of cooperating with others and has perception of them; offer content variety and a feedback with auditory and visual characteristics, according to the user actions level; offer from very simple synchronized actions up to collaborative actions that involve verbal communication or negotiations; and interact accompanied by human meditation.

Moreover, we considered recommendations presented in studies of face-to-face collaborative systems for individuals with typical development, but that we also consider appropriate for individuals with ASD, for its facilitating characteristics. For example, Haller and collaborators (2010) suggest that it is important to provide an environment that contributes to an easy user adaptation and manipulation of interface elements, this way the cognitive work load could be minimized.

However, the recommendations found in the literature are more closely related to interface characteristics to facilitate user interaction, and not specifically related to support awareness. The requirements identified in the present work propose recommendations focused on awareness, which should be also integrated to the gradual sequence of restrictions.

3. Collaborative Game PAR

We developed PAR (acronym in Portuguese for Ask, Help and Receive), a multitouch collaborative tabletop game to encourage social interaction among individuals with high

impairment in ASD. We chose multitouch technology by its advantages on collaborative work. This technology allows high levels of awareness and more fluent interactions [Hornecker et al. 2008].

PAR includes Collaboration Patterns, a set of interaction strategies on elements in a multiuser interface to gradually encourage collaboration among individuals with ASD, mainly those with more level of severity in ASD. Collaboration Patterns were designed based on the constraints proposed by Giusti and collaborators (2011), such as, “Different role pattern”, which assigns different roles for each user, and “Constraints on objects pattern”, which requires two or more users to move an object. The set includes a specific sequence of four Collaboration Patterns. We proposed this sequence following recommendations of experts in ASD. They indicated the collaboration is a complex area for people with ASD with high impairment; therefore, this must be carefully and gradually encouraged. The following sequence of Collaboration Patterns was used in PAR and indicates a gradual reinforcement of collaborative activities:

Passive Sharing Pattern. Each user has a different role to share resources. The tasks are just action and response from one user to another. To ‘share resources’, each user has only to realize his/her own task and know the result of the tasks of the partner, independently of identifying who executed the task and how it was executed.

Active Sharing Pattern introduces the importance of recognizing the role of the partner in collaboration. This is achieved by requiring ‘information exchange’ between users, in addition to sharing resources. Each user receives information from the partner about how to cooperate.

Joint-Performance Pattern uses the same principles as “Active Sharing Pattern” and introduces cooperative ‘simultaneous actions’ to help the other when needed. Users now realize that the participation of both is strictly necessary to achieve a goal, recognizing that they are involved in a collaborative activity.

Unrestricted Interaction Pattern. This pattern does not assign roles or constraints to users, allowing them to collaborate and develop ‘free strategies’ for sharing information, cooperating, and helping each other. This pattern is offered to users after they have interacted with each restricted patterns.

PAR game requires collaboration of two users around a multitouch tabletop, in order to get the sport pieces (shirt, shorts and sneakers) and to dress the soccer players of a team (Figure 1). The sports pieces are randomly distributed on three higher shelves of a warehouse. These pieces need to be taken down so the players can wear them. On the side of each shelf, there is a box. This box serves as a container. One user must put one sport piece into a box that serves as a container and send it to the second user. In the lower part of the warehouse, there is a cart. Second user must receive each sport piece in this cart. When the cart is full with the three sport pieces, it is necessary to take it to the parking lot and give the pieces one by one to the player in a row.

PAR game has three collaborative phases. Each user will have a different role according their place around the multitouch tabletop. One user is located at the upper side of the table (User 1), and the second user is located at the lower side of the table (User 2). At the time of dressing each player, users may move to the right side of the table (Figure 1). The first phase includes the Passive Sharing Pattern (Figure 1.a). The

role of user 1 is to take any sport piece and put it in the box on the shelf. The role of user 2 is to move the cart and take it to the shelf of the descending box to receive the piece sent by User 1. The second phase adds the Active Sharing Pattern (Figure 1.b), because User 1 must have information about the three sport pieces that should be sent. User 2 should ask for any piece by pressing a button. User 1 should then put the piece requested by the partner into the box to send it.

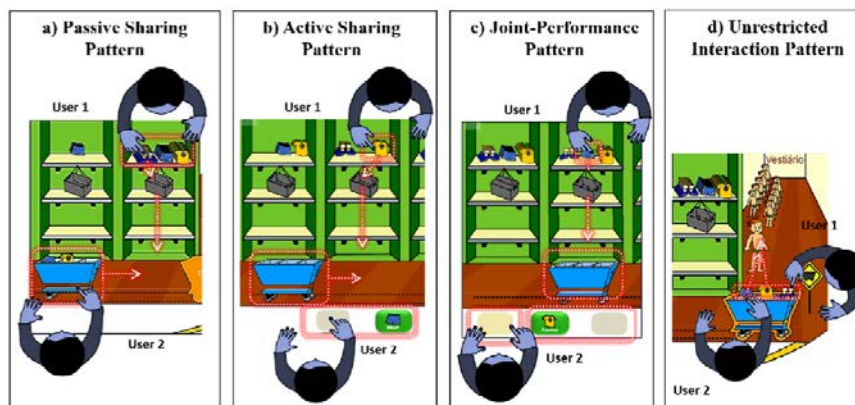


Figure 1. Collaboration Patterns in Collaborative Game PAR.

In addition to the tasks of the previous phases, in the third phase the boxes are already closed at the time that each piece is requested. Therefore, the Joint-Performance Pattern is applied because the box must be opened by both users. User 2 must press a button to open the boxes while User 1 takes and puts the piece requested into the box (Figure 1.c). In the three phases, when the cart is filled with three pieces, User 2 should move the cart to the parking lot. Here, the Unrestricted Interaction Pattern is applied, because each user may take any piece of the car in any order to dress the soccer player (Figure 1.d).

4. Evaluation

In this work, the main research questions of the evaluation include: What difficulties users had in the collaborative process during the game? Are these difficulties related with awareness of tasks, partners or workspace? What are the requirements of users according to their interaction in the game?

4.1. Evaluation Methods

Participants. We evaluated PAR with five youths with high impairment in ASD with ages between 10 and 17 years. These youths have high level of impairment in social interaction and collaboration skills. These users have difficulties in verbal communication, visual contact, and lack initiative to interact and share with others. Some users have behavioral problems and difficulty in respecting rules in a group. We named each user with letters from A to E: A is a 17-year-old boy, B is a 16-year-old boy, C is a 15-year-old girl, D is a 14-year-old boy, and E is a 10-year-old girl. The ethical questions were considered at all stages of the study. Parents of youths signed the respective Term of Free and Clarified Consent to authorize the study and the evaluation process with the youths.

Environment. We performed the evaluation in an institution for people with ASD, where we installed the multitouch tabletop and three cameras located at different angles focusing on the table. This arrangement allowed recording both user-system and user-user interactions.

Procedure. We performed 51 evaluation sessions with ten pairs of users. Each session lasted between 5 and 15 minutes. The evaluation routine was conducted so that each user interacted with the game in the order of phases. During the evaluation sessions, users were accompanied by a therapist who was willing to help and motivate them if needed.

Measures. We analyzed in each phase of the game the users' collaborative behavior related to the understanding of the tasks, the partners' awareness, and understanding of past, current and future activities. Due to the difficulties of users with ASD, it is not possible to anticipate their probable actions or to make a previous categorization of actions. Thus, we made a categorization from both the observations and recordings during the evaluation. For the parts of the game that include the collaboration patterns with restrictions, we classified the users' difficulties related to task awareness, social awareness, and workspace awareness, defined below:

Task awareness refers to the understanding of the tasks and of what the group needs to accomplish their realization [Belkadi et al. 2013], involving task information, such as, what are its objectives, structure, and how to compete it.

Social awareness refers to the knowledge available about the other collaborators within the social context [Belkadi et al. 2013], about the social connections within the group, involving who is the group, its objective, structure, participants.

Workspace awareness refers to the updated knowledge about interactions of collaborators within the shared workspace [Gutwin and Greenberg 1996], involving aspects like: where are the participants working, what are they doing, who is responsible, with which interface elements are they interacting, what are they going to do.

5. Results

As highlighted in [Silva et al. 2014, 2015], the interest of the users increased with each phase in the game. Collaboration Patterns applied in the game generated in the users greater motivation to perform the tasks and to interact with their partners through verbal or gestural expressions, such as have physical contact, see, ask for help, rectify, complain, laugh, commemorate, reject, and encourage. However, after the analysis performed in this work, we also highlighted that during the three phases of the game, in the restricted patterns, users also had some difficulties. We categorized these difficulties into three categories, as detailed in Table 1.

In the first phase of the game, with the Passive Sharing Pattern, users had the larger number of difficulties in the three awareness dimensions. In the second and third phases, the number of difficulties was decreasing in the social and workspace awareness (Figure 2). Users had more difficulties in task awareness with the joint-performance pattern in the third phase of the game, maybe because this phase demanded greater concentration to perform the tasks in the game.

Table 1. Categories of users' difficulties during the game, classified by different kinds of awareness.

Difficulties Related to:	Categories of users' difficulties
Task awareness	<ul style="list-style-type: none"> - Difficulty to understand what is his/her task and how to perform it. - Frustration with an error as result of an action.
Social awareness	<ul style="list-style-type: none"> - User disregards the partner's recommendations and necessities. - User sees the partner as someone who must help him/her doing his/her tasks. - User does the task of the partner when s/he does not perform it as quickly as s/he wants. - User does his/her task and also the partner's task, even if the partner is trying to perform it.
Workspace awareness	<ul style="list-style-type: none"> - User has difficulty to understand the new tasks and s/he continues to perform previous tasks. - User does not make the tasks according to previous tasks of the partner. - User interferes in the task of the partner because s/he thinks that the partner will perform it wrong.

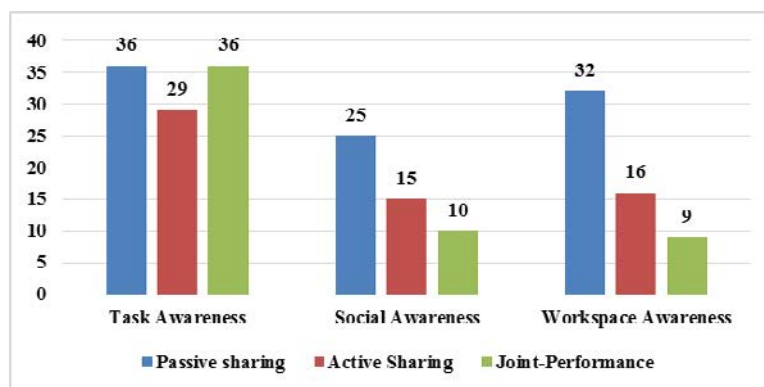


Figure 2. Number of user's difficulties observed in the restricted Collaboration Patterns during all the evaluation sessions.

The number of difficulties related to social awareness is inferior to the number in task and workspace awareness (Figure 2). But that number can be related to the difficulties that some users had even to involve themselves in the collaborative activities. Then, these users perform neither positive nor negative actions related with workspace. Most users (B, C, D, and E) had difficulties related to task awareness, they needed more motivation and orientation to perform the tasks, such as, repeat explanations about the movements required on elements of the game, and indicate who must do the task. Related to social awareness, some users (B, E) had more difficulty to both identify the other as his/her collaboration partner and understand the importance of wishes and interactions of the partner. This difficulty was reduced throughout the phases of the game. Related to workspace awareness, most users (B, C, D, and E) had sometimes difficulty to respect the workspace of other, interfering in their tasks and interactions.

In the first phase of the game (Passive Sharing Pattern), user C had all difficulties categorized in Table 1 at least once. User A had only some difficulties to both follow the wishes of the partner and perform his task according previous tasks of the partner (Figure 3). In the second phase (Active Sharing Pattern), most users had less difficulty (Figure 3) in most of the aspects categorized in Table 1. User E had difficulty to respect the workspace of the partner, but this difficulty decreased in the third phase of the game. Throughout the sequence of Collaboration Patterns, user E learned to respect rules in the collaborative work and waited for the response of the partner. User D had kept a large number of difficulties in the three phases of the game (Figure 3), related to understand what is the task and how to do it. This is the user with most impairment in ASD. He only performed the tasks when the therapist or the partner indicates him, but he does not interest in performing a collaborative work, then it was not possible to observe other difficulties related to social and workspace awareness.

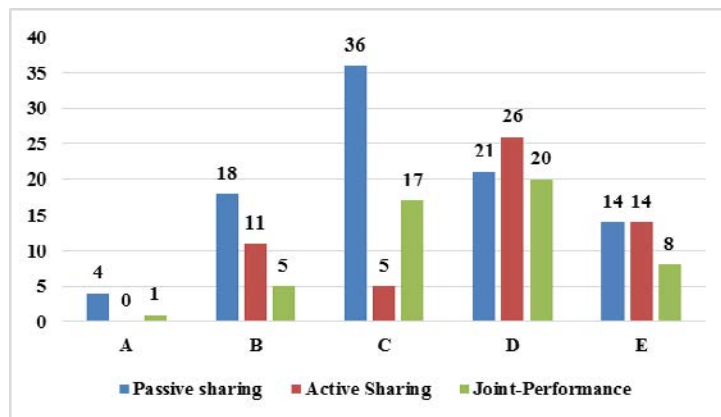


Figure 3. Total number of difficulties of each user in the restricted collaboration patterns during all evaluation sessions.

Since the unrestricted interaction pattern allows a free interaction between users, we did not classify difficulties, but we classified three types of what we called non-collaborative actions, characterized by the disinterest of one or both users (Table 2). These non-collaborative actions were decreasing throughout the phases of the game. In the first phase, four pairs of users performed the three non-collaborative actions while they suit the soccer players. In the second phase, three pairs of users performed only two of these actions. In the third phase, only one pair of users performed once these actions.

Table 2. Types of non-collaborative actions performed by pairs of users during the unrestricted interaction pattern.

Non-Collaborative Actions	Phases of the game		
	1st	2nd	3th
One user dresses the soccer player and the partner does not show any interest in collaborating.	BD CE	BD DE	CD
Users try to take turns and dress the soccer player, but when one of them cannot perform the task, the partner becomes aggressive, discouraging the interaction of the other.	BC	CD	
Both users need constant motivation to try to dress the soccer player.	CD		

PAR evaluation results have led to the identification of users' awareness requirements during their interaction with the game. These requirements are classified in: task awareness, social awareness, and workspace awareness, as detailed in the next subsection.

5.1. Awareness Requirements

The awareness requirements presented here were obtained according to the difficulties of users in their interaction with PAR game and the empirical evaluation results. We also considered some aspects mentioned in other studies (Section 2) together with experts in ASD, to define these requirements.

We organized the awareness requirements obtained in groups of questions (Table 3) that should be responded by developers to support each type of awareness regarding the interaction of users with ASD with face-to-face collaborative applications. These questions are similar to questions for the specification of a system focused on awareness for individuals with typical development [Belkadi et al. 2013], but they comply with difficulties that are particular of individuals with ASD. They do not include all the possible questions for people with typical development, and the proposed questions aim to involve the more specific needs of users with ASD.

Task awareness requirements. To promote knowledge of the actions being performed, and encourage the users identify themselves as responsible for certain tasks, the collaborative applications shall:

- Explain in detail for each user what is(are) his/her task(s) in the system, and everything related to it(them), offering a highly visual and auditory content, avoiding the use of text.
- Use some means of highlighting the task of each user (greater visibility, colors, lights, images, audios, etc.) more than the other elements of the interface, while the user is interacting with that particular task.
- Explain the importance of performing each task and the objective to be accomplished, through images and/or audios. This explanation may be presented at any time, depending on the difficulty that each user may have to understand it.
- Actively display the user name/photo at the screen, hinting that it is his/her turn to perform the task. Constantly motivate the user to perform the task, always encouraging by his/her name and showing elements at the interface that are of his/her interest.
- Show the performance of the task in parallel to the user interaction, in a way that the user "imitates" the actions shown and required by the system.
- Guide the user in each step of the task, indicating which actions to perform, with which elements and types of interactions.
- Show through animations, and/or audios, the explanation about the result of the task. In case the action has been in error, the system shall not show the wrong answers and/or negative messages, but instead an invitation to try again, with the explanation of the actions that must be done to be successful in the new attempt.

Table 3. Questions for awareness support in collaborative applications for individuals with ASD

Type of awareness	Questions
Task awareness	<ul style="list-style-type: none"> - Identification of tasks: What do I know about the task, its structure, content and context? - Objectives: What is the objective of the task? Why should I perform it? - Execution: How do I identify that I am responsible for the execution of the task? What steps should I follow to complete the task? What elements I need to perform on the interface to complete it? - Evaluation: How task execution will be evaluated?
Social awareness	<ul style="list-style-type: none"> - Identification of work performed in group: How do I identify that I belong to a certain group? How and when may I contribute with my group? - Objectives: Why should I interact with other to complete my activity? Why do we work together? - Identification of partners and their tasks: Who are my partners? What are my roles in this group? What are the roles of my partners? How can I identify when my partner(s) need(s) my collaboration?
Workspace awareness	<ul style="list-style-type: none"> - Identification of shared workspace: Where am I working? Where are my partners working? - Knowledge about partners interaction within the shared workspace: Who is responsible for an action or object? How can I help my partners to complete the tasks? - Identification of what is happening now and is going to happen within the workspace: What are the current tasks of my partners? What are they going to do?

Social awareness requirements. To encourage the users to be conscious of their partners and everyone's interactions, the collaborative applications shall:

- Offer a detailed explanation, using audiovisual means, about who is part of the group, highlighting each user as an essential part to the work to be performed.
- Inform about who finished the performance of some task and who must continue to collaborate with the previous task (calling the user by the name), and also show the respective explanation about how to interact and in which elements of the interface.
- Reserve a space to place the photos and names of each user, so that they feel identified.
- Offer intuitive ways, preferably images and sounds, to explain which are the functions of each member of the group and in which moment they should be performed.
- Offer distinct ways for each participant to call for the attention of the other, such that this participant can quickly perceive that his/her partner is trying to communicate.

Workspace awareness requirements. To introduce the users to the knowledge of workspace, the collaborative applications shall:

- Reserve space at the interface for each member of the group, identifying it with distinct characteristics for each one, including his/he photo and name.
- Offer a distinct way to highlight the participation of each user and the place around the interface in which he/she is interacting.
- Invite each participant to cooperate with his/her partner in the required moment, encouraging him/her that his/her help is very important to reach the objective.
- Block and free the elements of the interface in the adequate moments to facilitate users' interaction and encourage them to collaborate.
- Show the actions that are being performed at the workspace and the information about who is doing.

We suggest that each type of awareness requirement be included in collaborative applications sequentially, in order to gradually involve the users in each one of the dimensions that comprise the 3C collaboration model [Fuks et al. 2007]. This has the objective to get users to know each aspect of collaboration, and gradually understanding it as a whole. The 3C collaboration model includes the inter and intra-relationship of three dimensions: communication, which consists of the conversation through the exchange of messages; cooperation, which are interventions performed by participants within the shared workspace; and coordination, which consists in the management of participants, activities and resources [Fuks et al. 2007, 2008].

Initially, we suggest approximating the users to the knowledge about the tasks to be executed in a collaborative environment (task awareness) incorporating cooperation activities. Then, to incorporate some aspects to approximate the knowledge about work partners (social awareness) with communication activities. Finally, to approximate the knowledge about activities performed within the workspace (workspace awareness) incorporating coordination activities.

6. Conclusions

Individuals with ASD present several difficulties in communication and social interaction. They present difficulty to understand even basic concepts of a collaborative activity. Studies about collaborative applications for this public proposed different restrictions and recommendations for applications in order to motivate/force the collaboration among users. However, there are few studies on aspects of awareness for these individuals who do not know how collaborate. Our objective in the present work, as an initial contribution for this challenge, was to identify the difficulties faced by youths with ASD through their interaction in a collaborative game for multitouch tabletop, in order to obtain a set of requirements to guide the design of collaborative applications for individual with ASD, mainly to those with high severity.

We evaluated the collaborative game PAR with youths with ASD between 10 and 17 years, diagnosed with higher level of severity. Despite the encouraging results of the game to engaging users in an attractive experience, gradually encouraging social interaction as mentioned in the previous studies [Silva et al. 2014, 2015], we found, in

this work, that these youths had different difficulties related to the understanding of tasks and how to perform them (task awareness), to the recognition of a partner (social awareness), and to the understanding of the actions and tasks in the workspace (workspace awareness). The results about users' difficulties during their interaction with the game indicated that these users need even more motivation to perform a collaborative work. It is not sufficient to provide easy tasks or restrictive elements to motivate/force the collaboration. It is necessary to offer learning opportunities on how to collaborate, since this area is unknown or difficult for users with ASD. These learning opportunities can be offered by the gradual stimulation of users' awareness through appropriate awareness mechanisms in the interface. It is also important that users gradually understand each dimension of collaboration (cooperation, communication, coordination).

Awareness mechanisms can be developed based on the awareness requirements identified in this work. We believe that these requirements are a promising tool to design appropriate awareness mechanisms to facilitate the users' interaction with the system and with their partners, and to contribute to their learning about a collaborative work. This tool provides a stimulant first step for future research on collaborative systems for individuals with ASD, especially those with high impairment.

As future research, we will develop collaborative applications following these requirements, and then, evaluate the application with users with ASD, in order to identify benefits, limitations, and possible challenges of this approach.

Acknowledgments

Greis Silva thanks FAPERJ for the support granted to her work (process #101.541/2014). Alberto Raposo and Greis Silva thanks FAPERJ for supporting this research (Assistive Technology Program, process #E-26/190.243/2013).

References

- American Psychiatric Association APA. (2013). "Diagnostic and statistical manual of mental disorders, DSM-5®". American Psychiatric Pub. Retrieved March 15, 2015 from <http://goo.gl/aiFD4W>
- Battocchi, A., Pianesi, F., Tomasini, D., Zancanaro, M., Esposito, G., Venuti, P., Ben Sasson, A., Gal, E., and Weiss, P.L.T. (2009). "Collaborative Puzzle Game: a tabletop interactive game for fostering collaboration in children with Autism Spectrum Disorders (ASD)". In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '09), 197-204.
- Bauminger, N., Goren-Bar D., Gal E., Weiss, P.L.T., Yifat, R., Kupersmitt, J., Pianesi, F., Stock, O., and Zancanaro, M. (2007). "Enhancing Social Communication in High-Functioning Children with Autism through a CoLocated Interface". In: Multimedia Signal Processing, 2007. MMSP 2007. IEEE 9th Workshop on, 18-21.
- Belkadi, F., Bonjour, E., Camargo, M., Troussier, N., and Eynard, B. (2013). "A situation model to support awareness in collaborative design". *International Journal of Human-Computer Studies*, 71, 1: 110-129.

- Chen Weiqin. (2012). “Multitouch Tabletop Technology for People with Autism Spectrum Disorder: A Review of the Literature”. In: Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012), 198-207.
- Dourish, P., and Bellotti V. (1992). “Awareness and coordination in shared workspaces”. In: Proceedings of the 1992 ACM conference on Computer-supported cooperative work (CSCW '92), 107-114.
- Ellis, Clarence A., Gibbs, Simon J., and Rein, G. (1991). “Groupware: some issues and experiences. Commun”. ACM , 34, 1: 39-58.
- Fuks, H., Raposo, A., Gerosa, M., Pimentel, M., and Lucena, C. (2007). “The 3C Collaboration Model”. The Encyclopedia of E-Collaboration, Ned Kock (org), 637-644.
- Fuks, H., Raposo, A., Gerosa, M., Pimentel, M., Filippo, D., and Lucena, C. (2008). “Inter-and intra-relationships between communication coordination and cooperation in the scope of the 3C Collaboration Model”. In: Computer Supported Cooperative Work in Design (CSCWD'08), 148-153.
- Gal, E., Bauminger, N., Goren-Bar, D., Pianesi, F., Stock, O., Zancanaro, M., and Weiss, P.L.T. (2009). “Enhancing Social Communication in High-Functioning Children with Autism through a Co-Located Interface”. *AI & Society*, 24, 1: 75-84.
- Giusti, L., Zancanaro, M., Gal, E., and Weiss, P.L.T. (2011). “Dimensions of Collaboration on a Tabletop Interface for Children with Autism Spectrum Disorder”. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'11), 3295-3304.
- Gutwin, Carl and Greenberg, Saul. (1996). “Workspace Awareness for Groupware”. In: Conference Companion on Human Factors in Computing Systems (CHI '96), 208-209. <http://doi.acm.org/10.1145/257089.257284>
- Haller, M., Forlines, C., Koeffel, C., Leitner, J., and Shen, C. (2010). “Tabletop Games: Platforms, Experimental Games and Design Recommendations”. *Art and Technology of Entertainment Computing and Communication*, 271 - 297.
- Hornecker, Eva. (2005). “A design theme for tangible interaction: embodied facilitation”. In: Proceedings of the ninth conference on European Conference on Computer Supported Cooperative Work (ECSCW'05), 23-43.
- Hornecker, Eva, Marshall, P., Dalton, N. S., Rogers, Y. (2008). “Collaboration and Interference: Awareness with Mice or Touch Input”. In: Proceedings of the 2008 ACM Conference on Computer supported cooperative work (CSCW'08), 167-176.
- Hourcade, J. P., Bullock-Rest, N. E., and Hansen, T. E. (2012). “Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders”. *Personal and Ubiquitous Computing*, 16, 2: 157-168.
- Keay-Bright, W., and Howarth, I. (2012). “Is simplicity the key to engagement for children on the autism spectrum?”. *Personal Ubiquitous Computing*. 16, 2: 129-141.

- Millen, L., Hawkins, T., Cobb, S., Zancanaro, M., Glover T., Weiss, P.L.T, and Gal, E. (2011). “Collaborative Technologies for Children with Autism”. In: Proceedings of the International Conference on Interaction Design and Children (IDC’11), 246-249.
- Moore, D., Cheng, Y., McGrath, P., and Powell, N. J. (2005). “Collaborative virtual environment technology for people with autism”. *Focus on Autism and Other Developmental Disabilities*, 20, 4: 231-243.
- Piper, A., O’Brien, E., Morris, M. R., and Winograd, T. (2006). “SIDES: A Cooperative Tabletop Computer Game for Social Skills Development”. In: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW’06), 1-10.
- Putnam, C., and Chong, L. (2008). “Software and technologies designed for people with autism: what do users want?”. In: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '08), 3-10.
- Rosset, B. D., Rondan, C., Da Fonseca, D., Santos, A., Assouline, B., and Deruelle, C. (2008). “Typical Emotion Processing for Cartton but not for Real Faces in Children with Autistic Spectrum Disorders”. *Journal of Autism and Developmental Disorders*, 38, 5: 919-925.
- Silva, G. F. M., Raposo, A., and Suplino, M. (2014). Par: A collaborative game for multitouch tabletop to support social interaction of users with autism. *Procedia Computer Science*, 27, 84-93.
- Silva, G. F. M., Raposo, A., and Suplino, M. (2015). Exploring collaboration patterns in a multitouch game to encourage social interaction and collaboration among users with autism spectrum disorder. *Computer Supported Cooperative Work (CSCW)*, 24(2-3), 149-175.
- Stone, B., Beach, L., Ganci, A., Jones, T., Ribeiro, B., Strouse, E., and Woolley, E. (2011). “Immersive Learning: Developing an Interactive Touch Screen Learning Module for Children with Autism”. In: Proceedings of the 13th International Conference on Engineering and Product Design Education (E&PDE’11), 684-689.
- Weiss, P.L.T., Gal, E., Eden, S., Zancanaro, M., and Telch, F. (2011). “Usability of a multi-touch tabletop surface to enhance social competence training for children with Autism Spectrum Disorder”. In: Proceedings of the Chais conference on instructional technologies research, 71-78.
- World Health Organization. (1992). “The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines”. Geneva. Retrieved March 15, 2015 from <http://goo.gl/kEBJiR>
- Yuill, Nicola and Rogers, Yvonne. (2012). “Mechanisms for collaboration: A Design and Evaluation Framework for Multi-User Interfaces”. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19, 1: 1- 25.
- Zancanaro, M., Gal, E., Parsons, S., Weiss, T., Bauminger, N., and Cobb, S. (2010). “Teaching social competence: in search of design patterns”. In: Proceedings of the 9th International Conference on Interaction Design and Children (IDC '10), 270-273.

Interoperability Requirement to Enhance Collaboration in Software Product Lines: A Systematic Mapping

Frâncila Weidt Neiva¹, Heleno de Souza Campos Junior¹, José Maria N. David¹, Regina Maciel Braga¹, Marco Antônio Pereira Araújo¹, Fernanda Claudia Alves Campos¹, Rita Suzana Pitangueira Maciel²

¹Department of Computer Science, Federal University of Juiz de Fora – Minas Gerais, Martelos – CEP 36036-330, Juiz de Fora-MG – Brazil

²Department of Computer Science, Federal University of Bahia – Bahia, Pituba – CEP: 41810-825, Salvador-BA - Brazil

fran.weidt@gmail.com, camposheleno0@gmail.com, {jose.david, regina.braga, marco.araujo, fernanda.campos}@ufjf.edu.br, ritasuzana@dcc.ufba.br

Abstract. *[Background]: Software Product Lines (SPLs) have been widely discussed in literature as part of solution to build systems solutions in a collaborative way. Providing interoperability support is an essential requirement to achieve effective results in collaborative systems field. Despite the importance of interoperability in SPLs, research related to this field is scarce and scattered in literature. Fulfilling this requirement is not a trivial task and a mapping study about which interoperability solutions have been tackled in SPLs has not been done yet. [Goal]: This paper aims at identifying and classifying research related to interoperability in SPLs. [Method]: Conduct a systematic mapping to identify and classify SPL interoperability solutions. [Results]: The classifications showed that syntactic was the main addressed interoperability level and that architecture and variability issues were the focus of the SPL researches. [Conclusions]: The results have suggested that an expressive number of solutions related to the syntactic interoperability level were proposed. However, some challenges remain opened, such as providing higher interoperability level support, for example, regarding pragmatic, dynamic and conceptual interoperability solutions.*

1. Introduction

The use of software plays a key role in different sectors of society causing an implicit dependency. From this dependency, the need to increase quality and agility, and to reduce costs in software development process has arisen. To address this need some trends have been adopted as globalization of software engineering and use of software architectures based on Software Product Lines (SPLs) [Sugumaran et al. 2006]. The adoption of these trends increases collaboration complexity through the creation of new dependencies between distributed software development teams and organizations. SPLs are now developed and used beyond the boundaries of a single organization [Bosch 2009]. In a collaborative effort, distributed teams and organizations create software products. Regarding significant differences in practices in different teams and organizations, SPL environments are frequently heterogeneous in terms of variability modeling approach, platform, domain vocabulary, and so on. If left untreated this heterogeneity may induce “island solutions” that restricts communication and hinders

collaboration [Dhungana et al 2011]. A software product line is a set of software-intensive systems sharing a common set of features that satisfy specific needs of a particular market segment or mission, and is widely used to build systems solutions in a collaborative way [Clements and Northrop 2001]. These systems are developed from a common set of core assets in a prescribed way. In other words, a software product line comprises systematic reuse of assets to be instantiated or to generate a personalized product. Systematic reuse benefits the agile software development in conjunction with costs reduction, however interoperability challenge increases.

Software developers, usually working globally distributed, use assets built by others developers in a SPL. The selection and combination of different assets that compose a product in a SPL must be interoperable, which is especially difficult when different developers are involved. Interoperability is an important non-functional requirement that is not trivial to be fulfilled. Providing interoperability between SPL assets in order to build the best possible product, frequently demands that all environment be prepared and adapted to enable interoperation. To achieve interoperability requirement is necessary to design an interface which enables communication between assets [Siegmund et al. 2009].

Communication process is usually explored in semiotics, mainly in the study of meaning-making, the philosophical theory of signs and symbols in linguistics field. Semiotics is divided into three dimensions: syntactic, semantics and pragmatics. Syntax acts as a sign, semantics is related with which the sign refers to, and pragmatics is related to the effect of the sign on the interpreter. This effect can be noticed depending on the context where the sign is used [Morris 1938]. These three aspects are important in the communication process, which is fundamental to reach an effective collaboration as revealed in [Ellis 1991].

Using semiotics, we can also explain interoperability in SPLs. SPLs assets need to be combined and to communicate accordingly. In other words, they need to interoperate in order to reach specified goals. Interoperability can be understood as the ability of heterogeneous applications to share procedures and data on the same equipment and/or on distinct platforms [Bernstein 1996]. Interoperability in SPLs is not a trivial requirement. Each SPL has its own particularities and needs. As a result, each SPL requires that different interoperability levels be achieved. For example, suppose a developer working in a SPL that must combine assets that better address his/her expectation. So, he/she must know enough semantic information about assets, as restrictions about its use, which concepts they represent and which nonfunctional requirements they provide. In these situations, to assure that SPL will allow composition not only considering syntactic interoperable assets, but also higher interoperability levels, it is necessary to aggregate additional meaning into a communication protocol. Furthermore, these assets are usually built by globally distributed developers, even developers from different companies. The developer may not have previous information about most of the assets, once they did not develop them or use them before. Despite of this fact, the assets have to fulfill developers' requirements in order to solve their problems properly. In this case, supporting syntactic interoperability is not enough. In order to do this, it is important to consider the context where these assets are inserted, the constraints about their use, the concepts that they represent, and the nonfunctional requirements that they have to fulfill, among others [Neiva et al. 2016]. In this way, a SPL that supports interoperability in different levels is

essential to an effective collaboration among the developers, once it enhances the use and interoperation of assets created by different developers. Without interoperability support, the reuse in SPL is hindered and developers tend to build and use its own assets instead of those already developed by others.

The problem is that despite the importance of the interoperability requirement in SPL domain to enhance collaboration, the proposed solutions related to this research field are scarce and scattered in literature. Besides, to the best of our knowledge there is no study that maps the knowledge in literature about this subject. In this way, this paper aims at systematically mapping the interoperability solutions that have been proposed in software product line literature, mainly in collaborative systems area, identifying trends and research opportunities.

In contrast to the usual process of literature review, systematic reviews and mappings are designed to reduce bias and provide a reliable picture of the current state of the art of a specific research field. Through a predefined protocol, this sort of studies follows precise and strict methodological steps to select and analyze relevant papers.

This paper is organized as follows, Section 2 presents relevant concepts to understand the classifications used in the systematic mapping. Section 3 presents our systematic mapping planning and execution. Section 4 presents the results of our systematic mapping. Section 5 presents the threats to the validity of our study. Finally, Section 6 concludes this paper, with a discussion about our findings.

2. Background

Kitchenham et al. (2004) suggested that we have to consider adopting an evidence-based approach to software engineering. The author argued that evidence-based software engineering emphasizes the need to find and aggregate evidence on a specific topic using secondary studies, such as systematic literature reviews and mapping studies, as the methodological framework for identifying and aggregating evidence [Kitchenham 2010].

Systematic maps and reviews have distinct goals. Systematic reviews aim at identifying the best practices related to specific procedures, technologies, methods or tools based on empirical evidences. This is not a goal for systematic mappings since they do not analyze in details the articles content. Instead, the systematic mappings aim at classifying and conducting a thematic analysis of literature in a specific topic of software engineering [Petersen et al. 2008] [Kitchenham et al. 2011].

This paper contributes to an evidence-based approach to investigate interoperability requirement that is essential to be fulfilled in collaborative systems based on SPLs. This systematic mapping aims to identify studies across the years related to this subject, the most active authors in the area and the main vehicle for publications. In this context, we are interested in classifying the studies according to the interoperability level addressed based on Levels of Conceptual Interoperability Model (LCIM) [Tolk and Muguirra 2003] and according to the software product line element addressed (i.e. architecture, variability, artifact repository, and configuration management). The following two subsections provide the background on the LCIM and the software product line elements, respectively.

2.1 Levels of Conceptual Interoperability Model (LCIM)

There are several models that deal with interoperability in a technical domain [C4ISR 1998], [NATO 2003]. Technical domain is about systems that were already implemented. One of these models is the “Levels of Information Systems’ Interoperability (LISI)” [C4ISR 1998]. LISI model focuses on technology (IT) issues, and are mainly concerned with communication, data exchange and service (application) interoperability.

Other model is the NC3TA Reference Model for Interoperability (NMI). This model is defined and documented within the context of the NATO C3 Technical Architecture (NC3TA) [NATO 2003]. NC3TA establishes a set of standards, architecture and reference models. It aims at evaluate if a system is able to automatic exchange well-defined structural data. The main goal of the LISI and NC3TA models is to establish measures of merit to evaluate the degree of interoperability between two existing systems.

However, to analyze interoperability only considering a technical domain is not enough to achieve an effective collaboration. To cope with this problem, Tolk and Muguira (2003) proposed the “Levels of Conceptual Interoperability Model (LCIM)”. LCIM model is able to determine whether meaningful interoperability between systems is possible in initial phases of the systems development through the alignment of conceptual models. It was firstly developed to be applied in military simulation systems within the U.S. Department of Defense but has been successfully applied in different domains as U.S. Department of Energy solutions for the power-grid [GridWise 2007]. Also, it was featured as a reference model in several contributions, such as [Tolk et al. 2007], [Dobrev et al. 2007], [Zeigler et al. 2007] and [Morris et al. 2004].

LCIM is a framework that deals with conceptual interoperability issues beyond technical interoperability and divides conceptual interoperability into layers. LCIM levels are characterized as follows: Level 0 (No Interoperability) - Stand-alone systems are included in this level; Level 1 (Technical Interoperability) - A communication protocol enables data exchange between systems; Level 2 (Syntactic Interoperability) - A standard structure for information exchange is introduced; Level 3 (Semantic Interoperability) - A common information exchange reference model is used. On this level, the meaning of data is shared in an unambiguous way; Level 4 (Pragmatic Interoperability) - The use of data – or the context of its application – is understood by the participating systems. The context in which the information is exchanged is unambiguously defined; Level 5 (Dynamic Interoperability) - In this level, the systems are able to comprehend the state of changes that occur in the assumptions and constraints that each system is making over time, and is able to take advantage of those changes. Besides, this level ensures that those involved in a message exchange expect that the effects caused by this exchange are the same; Level 6 (Conceptual Interoperability) - This is the highest level and it occurs when the conceptual model is aligned. In other words, the assumptions and constraints generated through the semantic abstraction of the real world are aligned.

2.2 Software Product Line Elements

Interoperability requirement can be related to different elements in a SPL, such as architecture, variability, artifact repositories and configuration management. For

example, software product line architecture usually defines components that have to interact to enable products development. Different architectural designs can be employed to organize these components, such as Common Object Request Broker Architecture (CORBA) and Service Oriented Architecture (SOA) [Barbosa et al. 2011].

A brief description of these concepts, based on [Pohl et al. 2005], follows:

- **Architecture in SPL:** The architecture provides a common, high level structure for all product line applications.
- **Variability in SPL:** Variability defines what can vary in software product lines. In other words, it introduces variation points for the product line.
- **Artifacts Repository:** Artifacts repository comprises a common platform where the artifacts are stored. The artifacts are interrelated by traceability links to ensure the consistent commonality definition and the variability among all artifacts in a software product line [Clements and Northrop 2001].
- **Configuration Management in SPL:** Configuration management aims to manage different versions of the artifacts that are used to develop and maintain software system.

The definitions of the interoperability levels and the different elements involved in SPL presented in this section were used to guide the mapping of interoperability solutions in SPL. Interoperability in SPL enhances the communication between the assets in order to enable products composition in an effective way.

Next section details the systematic mapping conducted in this research.

3. Systematic Mapping

This section presents a systematic mapping of the literature that we conducted to investigate interoperability solutions that have been published in the software product line domain. We organized our systematic mapping based on the main activities proposed by [Kitchenham et al. 2004], such as: planning, conducting and reporting the study.

3.1 Mapping Planning

The planning activity aims to identify the mapping objectives and to define a protocol. The protocol specifies the method to be used in the systematic mapping in order to reduce the researcher bias [Steinmacher et al. 2012]. Moreover, the systematic mapping must be reproducible and the protocol contributes to fulfill this requirement. This section summarizes our protocol.

Research Questions. This systematic mapping aims at answering the questions below:

MQ1) How many studies were published across the years?

MQ2) Who are the most active authors in the area?

MQ3) Which publication vehicles are the main targets for research production?

MQ4) How are the solutions distributed according to its interoperability level?

MQ5) Which software product line artifacts are related to the interoperability solution?

Based on the research questions, we take into account the PICOC method proposed by [Petticrew and Roberts 2008] to define our scope:

- Population (P): Software product lines.
- Intervention (I): Interoperability.
- Comparison (C): No comparison intervention.
- Outcomes (O): Solution.
- Context (C): Software.

To select the papers to be analyzed in order to answer the research questions some criteria were defined.

Inclusion and Exclusion Criteria. The process used to include and exclude a paper was organized in one inclusion criteria (IC) and three exclusion criteria (EC). The inclusion criterion was:

IC1: The study presents an interoperability solution to support software product lines.

The exclusion criteria were:

EC1: The study is not written in English;

EC2: The study is not available on the web; and

EC3: The study is an editorial or summary of keynotes, workshops or tutorials.

After the definition of the research questions and the papers inclusion and exclusion criteria, the following steps define the sources of the research papers and the search string that was performed in these sources.

Sources: Search was done in six electronic databases namely: Science@Direct (<http://www.sciencedirect.com>); El Compendex (<http://www.engineeringvillage.com>); IEEE Digital Library (<http://ieeexplore.ieee.org>); ISI Web of Science (www.isiknowledge.com); Scopus (<http://www.scopus.com>); Inspec (<http://www.theiet.org>) and ACM Digital Library (<http://dl.acm.org>).

The electronic databases were selected based on previous studies [Kitchenham and Charters 2007], [Prikladnicki and Audy 2010] and [Steinmacher et al. 2012].

Search String: The search string was defined based on the PICOC terms and synonyms related to software product line and interoperability, as follows: (*“product line” OR “domain architecture” OR “dssa” OR “product family” OR “application family” OR “core artifacts”*) AND (*“interoperability” OR “mediation” OR “mediator” OR “wrapper” OR “connector”*) AND (*“software”*).

[Silva et al. 2010] and [Siegmond et al. 2009] were known potential interest studies that were used to control if the search string was finding relevant studies.

The mapping was conducted and the details for this step are showed in the following subsection.

3.2 Conducting the Mapping

In this activity we executed the search string on the selected sources and afterwards we performed the study selection process. This activity was done by two researchers in order to avoid selection bias according to personal experience or interest. The resulting process is illustrated in Figure 1 and it comprised the following steps:

The papers titles were read, and those papers that were considered clearly irrelevant to the research question, or appeared more than one time in the results, were excluded. The included papers were analyzed based on their abstracts and keywords, considering the research questions. The papers were fully read and analyzed according to the research questions. As a result, the selected papers had information extracted into a proper form.

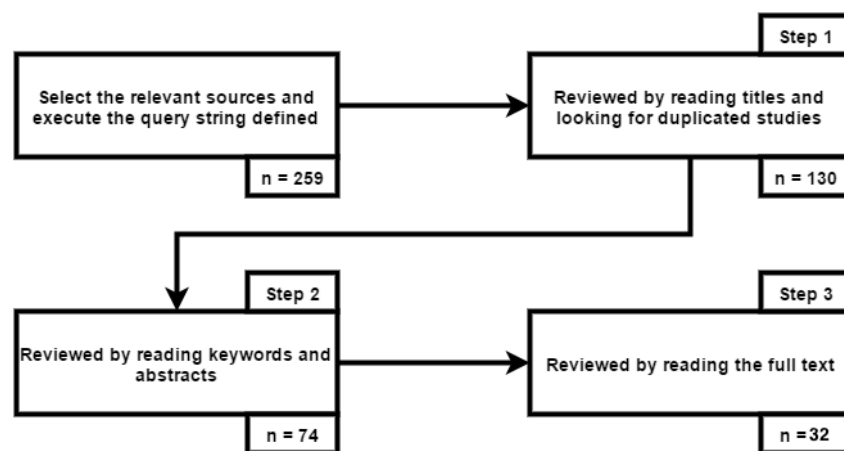


Fig. 1. Paper selection process – based on [Steinmacher et al. 2012].

In the first step, from 259 papers returned by the query, 130 (50,19%) were selected. The next step involved keywords and abstract reading. After the analysis, 74 papers were selected, which represents a reduction of 56,92%. Finally, on the last step, the review was performed by reading the full text. From 74 papers included in the previous step, 32 (43,24%) of them remained.

4. Systematic Mapping Report

The 32 selected papers were analyzed and the information to answer the mapping questions (MQ) was extracted. A summary on what could be observed on each of those questions follows.

Figure 2 represents graphically the answer to MQ1 (How many studies were published over the years?). The distribution of publications over the years is similar but we realized that there was an upward trend in the amount of publications after 2004.

This was probably due to the fact that the first conceptual model to deal with interoperability issues was proposed in 2003 by Tolk and Muguira (2003), motivated by the research agenda of USA Department of Defense Net-Centric Data Strategy. Besides, the Web 2.0 term became widespread in 2004 by O'Reilly Media enterprise increasing the interest in interoperability challenges.

It is important to highlight that the search was performed in 2015 December, so it is possible that papers which were published in this year had not been yet indexed.

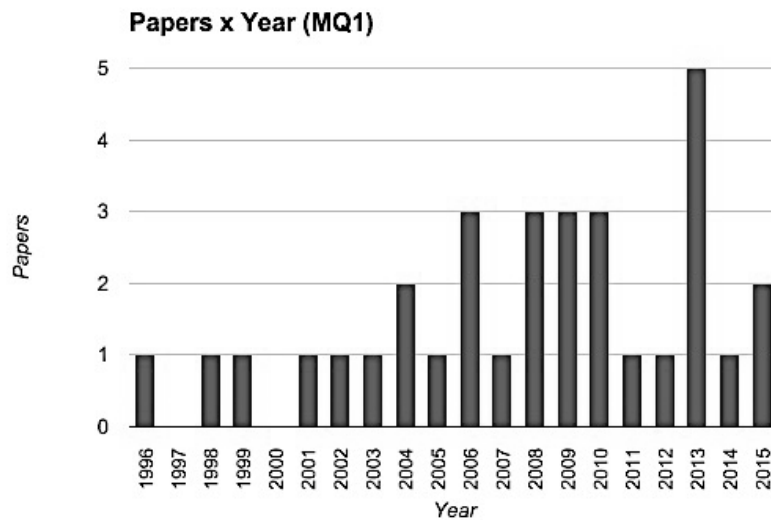


Fig. 2. Number of relevant papers per year.

Considering MQ2 (Who are the most active authors in the area?), we selected the authors that appeared twice or more times on the selected papers. The result is illustrated in Figure 3 and we can observe that the most active researchers have three publications. This fact evidences that interoperability in software product lines has not been widely explored by any group of authors.

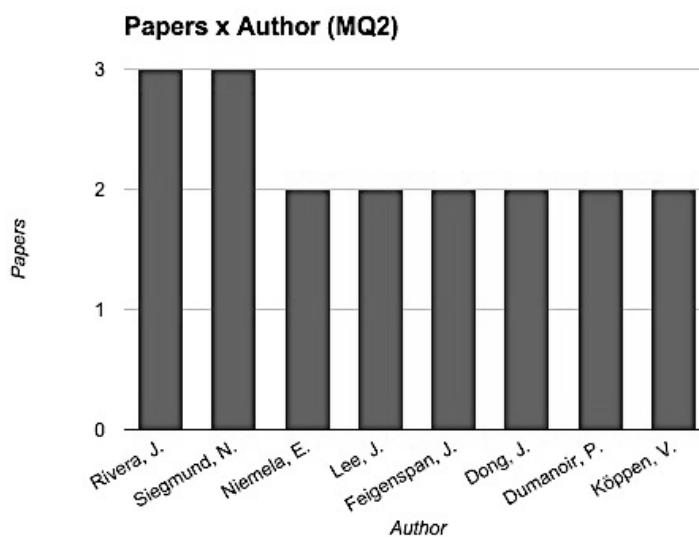


Fig. 3. Number of relevant papers per author.

Regarding MQ3 (Which publication vehicles are the main targets for research production in the area?), Figure 4 presents where the primary studies were published. Most of papers were published in conferences (46,87%), followed by workshops (25%) and journals (21,87%). In Computer Science, the number of Conferences is significantly higher compared with the number of Journals. This fact explains why most papers were published in conferences.

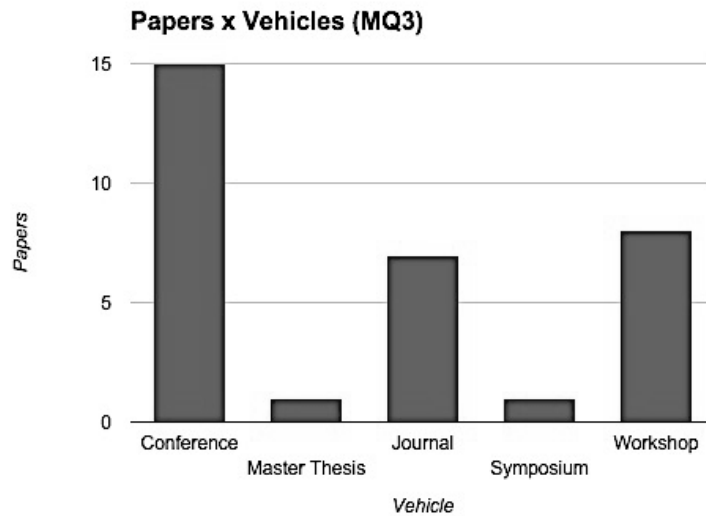


Fig. 4. Number of relevant papers per publication vehicles.

To answer MQ4 (How are the solutions distributed according to their interoperability level?), we used the LCIM to classify the papers focusing on their addressed interoperability level. Figure 5 illustrates the amount of papers and the addressed interoperability levels. First, we can conclude that syntactic was the main target level addressed. Second, we can observe that few interoperability solutions to support SPL in a pragmatic interoperability level were published. The pragmatic interoperability solutions have only been published after 2011. Pragmatic interoperability, as defined in Section 2.1, is the fourth level according to LCIM, thus there are opportunities to investigate how to satisfy higher interoperability levels in SPLs. Furthermore, we have realized that researches which investigate pragmatic interoperability in SPL are scarce in literature.

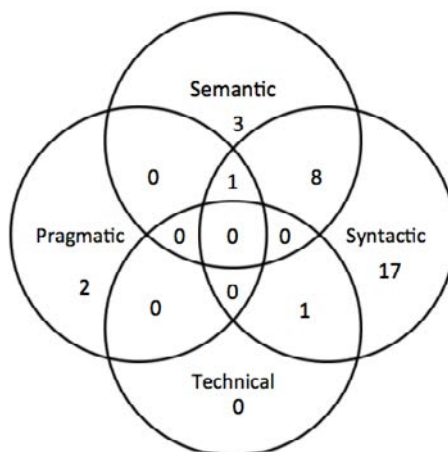


Fig. 5. Venn diagram for interoperability levels.

Finally, to answer MQ5 (Which software product line elements are related to the interoperability solution?), we classified the papers according to the addressed SPL element as defined in section 2.2. Figure 6 illustrates a classification that includes architecture, variability, artifacts repository and configuration management. This Venn diagram allows observing that architecture and variability are the main elements

addressed by interoperability solutions. The proposed solutions used the development of architecture designs as a strategy to enable the SPL to achieve interoperability. Also, once the variability introduces variation points in SPL, it is important that these variation points are aligned with interoperability purposes in order to identify variations that maintains interoperation among assets.

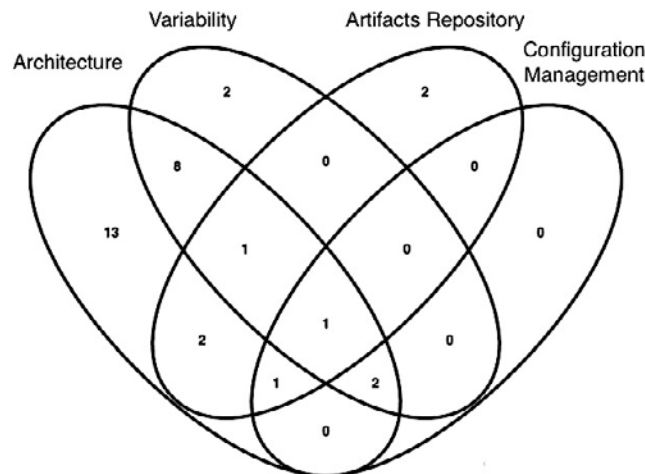


Fig. 6. Venn diagram for SPL elements.

In the next subsection, a summary of MQ4 and MQ5 research questions results is presented. Besides, the classified papers are linked to their references to allow knowing which paper belongs to which classification.

4.1 Summary of the classification according to LCIM and SPL elements

In this subsection we summarize the classification of interoperability level found in the papers according to the LCIM and its relation with the SPL element.

Table 1 provides an overview of the selected papers considering the interoperability level and the SPL elements, allowing to observe which SPL elements are supported by which interoperability level. Due to space limitations, the complete references list to the selected papers could not be presented in this paper. However, they can be found in <https://goo.gl/audSsK>. All external references are marked with a “*”.

Table 1. Distribution of papers according to LCIM and SPL element.

LCIM	SPL elements	References
Technical	Architecture	[Hodson 2007*].
	Variability	[Hodson 2007*].
Syntactic	Architecture	[Abu-matar 2009*] [Carbon et al. 2008*] [Dong 2013a*] [Dong 2013b*] [Dumanoir et al. 2009*] [Ganesan 2010*] [Garcia et al. 2013*] [Goedicke et al. 2002*] [Graubmann 2003*] [Hatzisymeon et al. 2005*] [Hodson 2007*] [Hupf and Davis 2007*] [Kaiser et al. 1996*] [Krishna et al. 2006*] [Losavio et al. 2015*] [Nascimento et al. 2011*] [Niemela and Holappa 1998*] [Niemela 1999*] [Papakonstantinou 2001*] [Razavian and Khosravi 2008*] [Rivera et al. 2008*] [Siegmund et al. 2010*] [Siegmund et al. 2009*] [Smiley et al. 2014*].
	Variability	[Abu-matar 2009*] [Ganesan 2010*] [Goedicke et al. 2002*] [Hartmann et al. 2010*]

Semantic		[Hodson 2007*] [Losavio et al. 2015*] [Nascimento et al. 2011*] [Niemela 1999*] [Razavian and Khosravi 2008*] [Smiley et al. 2014*].
	Artifacts Repository	[Chiang and Lee 2004*] [Dumanoir et al. 2006*] [Hupf and Davis 2007*] [Kaiser et al. 1996*] [Niemela 1999*] [Rivera et al. 2008*] [Smiley et al. 2014*].
	Configuration Management	[Kaiser et al. 1996*] [Nascimento et al. 2011*] [Niemela 1999*].
	Architecture	[Dong 2013a*] [Galindo et al. 2015*] [Graubmann 2003*] [Hatzisymeon et al. 2005*] [Hupf and Davis 2007*] [Lesaint and Papamargaritis 2004*] [Losavio et al. 2015*] [Nascimento et al. 2011*] [Niemela 1999*] [Papakonstantinou 2001*] [Silva et al. 2010*].
	Variability	[Galindo et al. 2015*] [Hartmann et al. 2010*] [Lesaint and Papamargaritis 2004*] [Losavio et al. 2015*] [Nascimento et al. 2011*] [Niemela 1999*].
	Artifacts Repository	[Hupf and Davis 2007*] [Niemela 1999*].
	Configuration Management	[Lesaint and Papamargaritis 2004*] [Nascimento et al. 2011*] [Niemela 1999*].
Pragmatic	Architecture	[Lee 2013*] [Nascimento et al. 2011*].
	Variability	[Lee 2013*] [Nascimento et al. 2011*] [Siegmond et al. 2012*].
	Configuration Management	[Nascimento et al. 2011*].

Table 1 allows us to observe that the number of syntactic solutions is higher than semantic solutions comparing all SPL elements. None SPL elements are supported by dynamic and conceptual interoperability levels. These observations are probably related to the fact that in order to achieve higher interoperability levels the lower levels must be established [Tolk and Muguira 2003]. Naturally, in literature we found more studies discussing lower interoperability levels and the lack of studies addressing high levels can be considered as a good research opportunity.

We also observed during paper analysis that syntactic interoperability solutions were commonly associated with a service-oriented architecture (SOA) [Abu-Matar 2009*] [Dumanoir et al. 2009*] [Losavio et al. 2015*] [Nascimento et al. 2011*] and the exchange of Extensible Markup Language (XML) files [Hatzisymeon et al. 2005*] [Hodson 2007*] [Hupf and Davis 2007*] [Rivera et al. 2008*]. These associations can be explained because (SOA) was proposed as a solution to enable distributed applications development that is an important factor to enhance collaborative systems design [Papazoglou et al. 2007]. XML standard in its turn, is used in Web Services initiatives. Web Services technologies were built to follow SOA concepts. XML is an important technology to provide interoperability across the web.

In semantic interoperability level, interoperation is enriched with semantics. To add semantic to the environment, the use of ontologies is a common strategy [Hupf and Davis 2007*] [Nascimento et al. 2011*] [Silva et al. 2010*]. Ontologies are used to store semantic information and are widely used in semantic web applications as occurred in [Silva et al. 2010*].

In pragmatic interoperability level, collaboration must achieve the expectations of the involved parties. For this purpose, interoperation is enriched with pragmatic information, such as context information. Context information can be modeled through ontologies, in conjunction with semantic information.

SPL architecture and variability were the main target for interoperability solutions, considering all interoperability levels. Providing architecture designs that

considers interoperability issues since its conception is a smart strategy to achieve interoperability in SPL [Batisda and Nieto 2007]. SPL variability must maintain the interoperation among the assets in order to produce well-functioning products.

Interoperability in SPL benefits different domains. In our set of selected papers, we identified interoperability in SPLs applied in e-business [Papakonstantinou and Vassalos 2001*], e-health [Losavio et al. 2015*] [Smiley et al. 2014*], e-learning [Razavian and Khosravi 2008*] and military domain [Rivera et al. 2008*] [Hupf and Davis 2007*] [Dumanoir et al. 2006*] [Dumanoir et al. 2009*], for example. While SPL has been adopted to increase reuse and productivity, the need of collaboration involving scenarios of distributed teams and systems reinforced interoperability as an essential requirement.

There are influences that can threaten the validity of this mapping study and must be identified. The next section presents these threats to the mapping study validity.

5. Threats to Validity

This systematic mapping intended to raise the interoperability solutions to support software product line domain applications. However, it presents some threats to validity and limitations. The results of this mapping study may have been influenced by some uncontrollable limitations. First and foremost, there might be bias regarding to the selected papers during this study. To prevent this bias, two researchers actively reviewed all the papers, agreeing on each selection.

Additionally, errors can be introduced in the protocol definition and the search string may not contain all the relevant keywords, even considering that the protocol was reviewed. As a result, this fact may cause the loss of some valuable studies. To reduce this risk, other researchers reviewed the mapping planning presented in Section 3. The search string was also evaluated using two papers [Silva et al. 2010] and [Siegmund et al. 2009] to control the returned results. We aimed to verify if they would be found by the electronic databases. The two papers appeared in the results generating evidences about the search string correctness.

Furthermore, we did not consider the Springer Link database, so it is possible that relevant studies were not indexed by the selected electronic bases. However, [Kitchenham 2010] conducted a systematic mapping only using the SCOPUS database. The author argued that SCOPUS indexes IEEE, ACM and Elsevier publications, which means that it indexes many of the leading publications. Once SCOPUS was in our set of selected databases, we believe that the selected electronic databases were enough to obtain a big picture of the interoperability research area.

6. Final Remarks

Interoperability is a key requirement to achieve an effective collaboration in collaborative systems. To cope with market needs, collaborative systems can be built using SPLs. In this paper, we presented a systematic mapping to identify and classify the interoperability solutions in software product lines. We aimed to obtain a good coverage of publications related to the targeted area. During the mapping conduction, we started with 259 papers found by the selected electronic sources. These results were filtered, which resulted in 32 papers.

From the obtained results, we identified that after the year 2004 the number of papers published in the interested area has slightly increased, reaching a peak in 2013. We also identified that there is no research group widely exploring the area. The most active researchers in the area have 3 publications.

Analyzing the content of the selected papers, we could notice that most of them (53,12%) only focused on syntactic interoperability level to support SPLs. The semantic interoperability levels were exclusively addressed in 9,37% of the papers and 25% were related to syntactic and semantic level. We can observe that other interoperability levels such as dynamic and conceptual did not appear in the selected papers.

Interoperability enables the communication between the assets to compose a product, maximize reuse and encourage developers to use assets built by others developers. Despite interoperability importance, there is a low number of research in this area. The majority of papers focused on the lowest interoperability levels and addressed architecture and variability in SPLs. Higher interoperability levels represent a long-term research agenda that need to be explored in depth.

Despite this mapping study representing an important contribution itself, it is one piece in a bigger project in collaboration system area. We are making progress in advanced studies to depth understanding strategies to achieve higher interoperability levels (pragmatic, dynamic and conceptual levels). We aim to enhance a meaningful collaboration, regarding collaborative experiments supported by a SPL infrastructure. In these advanced studies, we are considering, for example, organizational factors and the context where the involved companies and developers are inserted.

We already obtained results achieving lower interoperability levels: syntactic, semantic and pragmatic interoperability [Neiva et al. 2015]. We used these levels to enhance collaboration among scientists that collaboratively develop scientific experiments in a software ecosystem that has a SPL in the core of its infrastructure. We obtained positive results as witnessed in [Neiva et al. 2015], however we aim to explorer higher interoperability levels. As revealed by this mapping study, to explore high interoperability levels is a good contribution to literature and offer a lot of research opportunities and questions to be answered. This work can be considered as a necessary step to help researchers that intend to contribute with this research field.

References

- Barbosa, E. A., Batista, T.: Garcia, A. and Silva, E. (2011) “PL-AspectualACME: an Aspect-Oriented Architectural Description Language for Software Product Lines”. In: Proceedings of the 5th European Conference on Software Architecture, Springer-Verlag, pages 139—146.
- Bastida, L. and Nieto, F. J. (2007) “Context-Aware Service Compositions: A Way to Facilitate Interoperability”. In: Enterprise Interoperability II, Springer London, pages 427—430.
- Bernstein P. A. (1996) “Middleware: A Model for Distributed System Services.” In: Communications of the ACM. New York. vol. 39, n. 2, pages 86—98.
- Bosch, J. (2009) “From software product lines to software ecosystems.” In: Software Product Line Conference (SPLC), San Francisco, CA, USA: ACM, Carnegie Mellon University, vol. 446, pages 111—119.

- C4ISR Architecture Working Group. (1998), “Levels of information systems interoperability (LISI)”. March, 30.
- Clements, P.C. and Northrop, L. (2001) *Software Product Lines: Practices and Patterns*. Addison-Wesley.
- Dobrev P., Kalaydjiev O., Angelova G. (2007) “From Conceptual Structures to Semantic Interoperability of Content”. Bookchapter in *Conceptual Structures: Knowledge Architectures for Smart Applications*, Springer Verlag, pages. 192—205.
- Dhungana D, Seichter D, Botterweck G, Rabiser R, Grunbacher P, Benavides D, Galindo JA. (2011) “Configuration of multi product lines by bridging heterogeneous variability modeling approaches”. In: *Software Product Line Conference (SPLC)*, pages 120-129.
- Ellis, C.A., Gibbs, S.J. and Rein, G. (1991) “Groupware: some issues and experiences”. In: *Communications of the ACM*, vol. 34, n. 1, pages 39—58.
- GridWise Architecture Council Interoperability Framework Team. (2007) “Interoperability Context-Setting Framework”, V1.0, July.
- Kitchenham, B. A., Budgen, D. and Brereton, O. P. (2011) “Using Mapping Studies as the Basis for Further Research - A Participant-Observer Case Study”. In: *Information & Software Technology*, vol. 53, n. 6, pages 638—651.
- Kitchenham, B. and Charters, S. (2007) “Guidelines for Performing Systematic Literature Reviews in Software Engineering” Technical Report, Keele University and Durham University Joint Report.
- Kitchenham, B. (2010) “What’s up with software metrics? – A preliminary mapping study”. In: *Journal of systems and software*, vol. 83, n. 1, pages 37—51.
- Kitchenham, B.A., Dyba, T., Jorgensen, M. (2004) “Evidence-Based Software Engineering”. In: *Proceedings of the 26th International Conference on Software Engineering, (ICSE '04)*, IEEE Computer Society, Washington DC, USA, pages 273—281.
- Morris, C. W. (1938), “Foundations of the Theory of Signs”.
- Morris, E., Levine, L., Meyers, C., Place, P., Plakosh, D. (2004) “SOSI: System of Systems Interoperability”, vol. 5, n. 573.
- NATO Allied Data Publication 34 (ADatP-34). (2003) “NATO C3 Technical Architecture (NC3TA)”, Version 4.0.
- Neiva, F.W., David, J.M.N., Braga, R. and Campos, F. (2016) “Towards Pragmatic Interoperability to Support Collaboration: A Systematic Review and Mapping of the Literature”. In: *Information and Software Technology*, vol. 72, pages 137—150.
- Neiva, F.W., David, J.M.N., Braga, R., Campos, F. and Freitas, V. (2015) “PRIME: Pragmatic Interoperability Architecture to Support Collaborative Development of Scientific Workflows”. In: *Proceedings of the 2015 IX Brazilian Symposium on Components, Architectures and Reuse Software (SBCARS)*, pages 50—59.

- Papakonstantinou, Y. and Vassalos, V. (2001) “The Enosys Markets Data Integration Platform: Lessons from the Trenches”. In: Proceedings of the International Conference on Information and Knowledge Management, pages 538—540.
- Papazoglou, M.P., Traverso, P., Dustdar, S. and Leymann, F. (2007) “Service-oriented computing: state of the art and research challenges”. In: Computer, n. 11, pages 38—45.
- Petersen, K., Feldt, R., Shahid, M. and Mattsson, M. (2008) “Systematic Mapping Studies in Software Engineering”. In: Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering (EASE'08), pages 68—77.
- Petticrew, M. and Roberts, H. (2008) “Systematic reviews in the social sciences: A practical guide”. John Wiley and Sons.
- Pohl, K., Bockle, G., Linden, F.V.D. (2005), Software Product Line Engineering, Springer.
- Prikladnicki, R. and Audy, J.L.N. (2010) “Process Models in the Practice of Distributed Software Development: A Systematic Review of the Literature”. In: Information & Software Technology, vol. 52, n. 8, pages 779—791.
- Siegmund, N., Pukall, M., Soffner, M., Koppen, V. and Saake, G. (2009) “Using Software Product Lines for Runtime Interoperability”. In: Proceedings of the Workshop on AOP and Meta-Data for Software Evolution, RAM-SE'09, n. 4.
- Silva, A. P. D., Costa, E., Bittencourt, I. I., Brito, P. H. S., Holanda, O. and Melo, J. (2010) “Ontology-Based Software Product Line for Building Semantic Web Applications”. In: Proceedings of the 2010 Workshop on Knowledge-Oriented Product Line Engineering, KOPLE'10, ACM Special Interest Group on Programming Languages (SIGPLAN), n. 1, pages 1—6.
- Steinmacher, I., Chaves, A.P. and Gerosa, M.A. (2012) “Awareness Support in Distributed Software Development: A Systematic Review and Mapping of the Literature”. In: Computer Supported Cooperative Work (CSCW), vol. 22, n. 2-3, pages 113—158.
- Sugumaran, V., Sooyong, P., and Kyo, C. K. (2006) “Software Product Line Engineering”. In: Communications of the ACM, vol. 49, n. 12, pages 28—32.
- Tolk, A. and Muguira J.A. (2003) “The Levels of Conceptual Interoperability Model”. In: Proceedings of the Fall Simulation Interoperability Workshop. Orlando, Florida: Simulation Interoperability Standards Organization, pages 1—10.
- Tolk, A., Diallo, S.Y., Turnitsa, C.D. (2007) “Applying the Levels of Conceptual Interoperability Model in Support of Integratability, Interoperability, and Composability for System-of-Systems Engineering” In: Journal of Systemics, Cybernetics and Informatics, vol. 5, n. 5, pages 65—74.
- Zeigler, B.P., Hammonds, P.E. (2007), Modeling & Simulation-Based Data Engineering: Introducing Pragmatics into Ontologies for Net-Centric Information Exchange, Academic Press.

Aplicando o Design Science Research no Desenvolvimento de um Sistema Colaborativo Assistivo

Daniel Maniglia Amancio da Silva¹, Deivid Felipe Sartori¹, Celso Massaki Hirata²,
Carla Diacui M. Berkenbrock¹

¹Universidade do Estado de Santa Catarina (UDESC)
Joinville – SC – Brasil

²Instituto Tecnológico de Aeronáutica (ITA)
São José dos Campos – SP – Brasil

danielmanilha@hotmail.com, deividfelipi@gmail.com,
hirata@ita.br, carla.berkenbrock@udesc.br

Resumo. *Os avanços da tecnologia, bem como a crescente preocupação da sociedade com a inclusão social de pessoas com deficiências motivaram o desenvolvimento de ferramentas para apoiar pessoas com atrasos cognitivos. Este trabalho tem o objetivo de definir artefatos de comunicação, em smartphones, para um sistema colaborativo de monitoramento geográfico, possibilitando a comunicação do deficiente intelectual com seus cuidadores em seu deslocamentos cotidianos. O trabalho foi guiado pela metodologia de pesquisa Design Science Research (DSR). A pesquisa foi dividida em três ciclos de descoberta de conhecimento: ciclo de conhecimento do usuário, ciclo de design dos artefatos e ciclo de avaliação da usabilidade. O presente artigo também detalha o levantamento de requisitos, desenvolvimento dos artefatos e avaliação do sistema. Como resultados, destaca-se a geração do conhecimento científico proporcionado pela utilização da metodologia de pesquisa DSR. Além disso, o trabalho visa fortalecer a utilização da abordagem de desenvolvimento centrado no usuário e do design participativo no contexto de desenvolvimento de um sistema colaborativo assistivo.*

Abstract. *Advances in technology and the growing concern of society in relation to the social inclusion of people with disabilities motivated the development of tools to support people with cognitive delays. This work aims to define communication artifacts, used in smartphones, for a collaborative system of geographical monitoring, enabling the communication of intellectual disabled with their caregivers in their daily movements. The work was guided by Design Science Research (DSR). The research was divided into three cycles of knowledge discovery: user knowledge cycle, design cycle of artifacts and usability evaluation cycle. This article also details the requirements gathering, development of artifacts and system evaluation. As a result, there is the generation of scientific knowledge provided by the use of the DSR research methodology. In addition, this work intend to strengthen the use of user centered design and participatory design in the development of a collaborative assistive system.*

1. Introdução

No Brasil, nas últimas décadas, novas políticas públicas têm sido estabelecidas para incentivar o desenvolvimento e uso de Tecnologias Assistivas (TA). Dentre as deficiências legalmente assistidas, pode-se citar a deficiência intelectual.

Deficiência intelectual (DI) é um termo que se usa quando uma pessoa apresenta certas limitações no seu funcionamento mental e no desempenho de tarefas como as de comunicação, cuidado pessoal e de relacionamento social [Almeida 2005].

A vida em comunidade das pessoas com DI tem sido estudada [Fink 1992], [Satcher 2000], [Wahl 1999] e é evidente o aumento das ligações comunitárias, do suporte social e das oportunidades de recuperação pela socialização.

Sistemas computacionais em rede disseminaram-se por todo o sistema social e vem provocando transformações em todos os setores a vida contemporânea [Fuks 2011]. Portanto, sistemas assistivos e colaborativos podem servir como ferramenta de apoio a colaboração e socialização de pessoas com DI.

Porém, existem desafios no desenvolvimento de tecnologias assistivas e sistemas colaborativos. A característica de multidisciplinaridade, as particularidades, limitações e diferentes necessidades dos usuários tornam complexo o desenvolvimento destes sistemas com foco em “usuários especiais”. Para [Druin 2009], desenvolver sistemas é sempre uma tarefa desafiadora e o desafio é ainda maior no desenvolvimento de sistemas para pessoas com necessidades especiais. A combinação de relações de poder, a reduzida habilidade de comunicação e as partes interessadas adicionais, tais como pais, professores ou responsáveis, resulta em uma situação bastante complexa. Já para [Grudin 1994], no desenvolvimento de sistemas, considerar usuários com DI como usuários “vulneráveis” somente evidencia o desconhecimento de métodos adequados de design, a dificuldade na comunicação ou a dificuldade de entender a realidade destes usuários.

A comunicação é um dos pilares dos sistemas colaborativos, uma necessidade para relacionamento social e também uma dificuldade para pessoas com DI. Nesse sentido, o presente artigo detalha a pesquisa em desenvolvimento que busca a geração de conhecimentos sobre a comunicação de pessoas com DI e seus cuidadores. Neste contexto, a Comunicação Aumentativa e Alternativa (CAA) surge como a possibilidade de colaboração envolvendo pessoas com DI. A pesquisa é guiada pela metodologia de pesquisa denominada *Design Science Research* (DSR). O presente artigo detalha os ciclos de pesquisa que possibilitaram a construção dos componentes computacionais para possibilitar a comunicação entre pessoas com DI e seus cuidadores em um sistema de monitoramento geográfico.

O restante do artigo está organizado da seguinte forma: a seção 2 apresenta os trabalhos relacionados, a seção 3 apresenta os conceitos de Design Science Research e descreve a adaptação realizada para a pesquisa descrita neste artigo. Na seção 4 é feita uma discussão sobre a relevância do problema. As seções 5, 6 e 7 descrevem os artefatos de software gerados na pesquisa, a descrição do desenvolvimento da pesquisa e os ciclos de descoberta de conhecimento e a definição do rigor da pesquisa, respectivamente. Por fim, na seção 8 são apresentadas as considerações finais e os trabalhos futuros.

2. Trabalhos Correlatos

Esta seção apresenta alguns trabalhos diretamente relacionados com a pesquisa descrita no presente artigo. São trabalhos que abordam a utilização do DSR em sistemas colaborativos, avaliação de sistemas para pessoas com deficiências e redes sociais em dispositivos móveis.

[Rocha et al. 2015] utilizaram a metodologia *Design Science Research* para definição de artefatos para sistemas colaborativos. A pesquisa demonstra a influencia do tamanho do grupo na participação em bate-papo educacional e produz um modelo matemático para estimar a quantidade máxima de alunos que devem participar do bate papo para manter o nível de participação desejado.

[Pries-Heje et al. 2008] apresentam uma discussão sobre a importância da avaliação. No entanto, a discussão de atividades e métodos de avaliação é limitada e, normalmente, assume uma perspectiva posterior, em que a avaliação ocorre após a construção de um artefato. Tais perspectivas podem assumir que a avaliação é um processo empírico e seus métodos podem ser selecionados da mesma forma como métodos de pesquisa empírica. Os autores ainda, demonstram uma gama mais ampla de estratégias de avaliação, que inclui uma avaliação anterior a construção do artefato. Esta visão adiciona ênfase às estratégias para a avaliação de processos de design, além de projetar produtos, utilizando critérios de qualidade.

[Petrie et al. 2006] apresentam dois estudos de casos que exploram técnicas de avaliação assíncrona em ambientes remotos com participantes com deficiência. No trabalho, são utilizadas as técnicas de questionários e observações. O estudo mostra que os dados quantitativos recolhidos em avaliações remotas assíncronas com usuários com deficiência são comparáveis com os dados coletadas nas avaliações locais.

[Chang et al. 2009] sugerem uma arquitetura geral para redes sociais móveis onde os indivíduos nas redes formam uma comunidade virtual, que convergem e se conectam com o outro, com base na comunicação móvel. Os autores também descrevem uma implementação de redes sociais móveis em colaboração com um programa de emprego para as pessoas com doença mental grave.

[de Araújo et al. 2014] utilizaram o design participativo como ferramenta de apoio para o desenvolvimento de um sistema colaborativo para tratamento multidisciplinar na rede do Sistema Único de Saúde. Eles utilizaram técnicas que possibilitaram que os profissionais da saúde, que irão utilizar o sistema, pudessem opinar durante o processo de construção do mesmo.

Assim como o presente trabalho, [Rocha et al. 2015] e [Pries-Heje et al. 2008] utilizam o DSR como guia para descoberta de conhecimento científico no desenvolvimento de sistemas colaborativos. Sendo que para [Rocha et al. 2015] o artefato resultante é uma fórmula matemática para obter o número ideal de participantes em *chats* e [Pries-Heje et al. 2008] empregam como artefato a avaliação. [de Araújo et al. 2014] utilizam o design participativo para levantamento de requisitos e definição de um sistema colaborativo para a área de saúde. Em [Petrie et al. 2006] é relatada a avaliação, mas no contexto de sistemas onde os usuários são pessoas com deficiências. O contexto do trabalho apresentado por [Chang et al. 2009] tem relação direta com o trabalho relatado neste artigo, pois a pesquisa desenvolvida foca na utilização de redes sociais em dispositivos

móveis.

A multidisciplinaridade da pesquisa descrita neste artigo justifica a correlação com trabalhos desenvolvidos em contextos diferentes. Nesta seção destacamos trabalhos que utilizam a metodologia DSR para sistemas colaborativos para design e avaliação, assim como trabalhos com deficientes e redes sociais móveis. A junção de partes dos escopos dos trabalhos correlatos adicionado ao foco na Comunicação Aumentativa e Alternativa, define o escopo de trabalho da pesquisa descrita no presente artigo.

3. *Design Science Research*

Design Science Research (DSR) é um método de pesquisa centrado na evolução de uma “ciência do projeto” (*design science*) originário da área de engenharia de produção, e que evidencia seu sentido e sua forma de operacionalização. Para [Lacerda et al. 2013], o DSR tem foco na importância da definição das classes de problemas e dos artefatos gerados no âmbito da pesquisa. Para [Aken 2004] as classes de problemas podem consistir em uma organização para a trajetória e o desenvolvimento do conhecimento em uma *design science*. Para [Bax 2015], DSR é uma metateoria que auxilia o pesquisador a criar conhecimento teórico durante os processos de concepção de artefatos, justificando como tais processos podem ser significativos para a comunidade científica.

[Hevner 2007] utiliza o DSR em sistemas de informação. De acordo com o autor, os conhecimentos necessários para realizar uma pesquisa em sistemas de informação envolvem os paradigmas da “ciência do comportamento” e “ciência do design”. A ciência do comportamento aborda a pesquisa pelo desenvolvimento de teorias que explicam fenômenos relacionados com a necessidade de negócio identificada e a ciência de design aborda a pesquisa pelo desenvolvimento e avaliação de artefatos projetados para atender a necessidade de negócio identificada.

No contexto da presente pesquisa, a ciência do comportamento guiará o entendimento dos limites de comunicação e entendimento do software pelos usuários com deficiência intelectual e a ciência do design guiará o desenvolvimento de interfaces adaptadas a esses usuários. Portanto, na presente pesquisa, o fenômeno comportamental investigado é a utilização da tecnologia de comunicação por usuários com deficiência intelectual e o artefato desenvolvido são os componentes e recursos de tela adaptados às necessidades dos usuários.

[Hevner 2007] também sugere um *framework* conceitual, onde define-se que uma pesquisa é realizada em um ambiente produzindo uma base de conhecimento. Os produtos da pesquisa são: teorias e artefatos que devem ser justificados e avaliados para produzirem *feedback* que possa refinar as teorias e os artefatos. Adicionalmente, o autor disponibiliza um conjunto de diretrizes que as pesquisas em sistema de informação devem apresentar. Na Tabela 1 estão as diretrizes aplicadas a presente pesquisa.

4. Relevância do Problema

Pessoas com DI diariamente realizam trajetos, como por exemplo, ida e volta até a padaria, supermercado ou até mesmo para as empresas onde trabalham. Na maioria das vezes, devido a suas deficiências, essas pessoas dependem da supervisão presencial dos cuidadores na realização de seus trajetos.

Tabela 1. Diretrizes do DSR adaptado para a pesquisa

Diretriz	Aplicação da Diretriz
Relevância do Problema	<p>“Para as pessoas sem deficiência a tecnologia torna as coisas mais fáceis. Para as pessoas com deficiência, a tecnologia torna as coisas possíveis” [Radabaugh 1993].</p> <p>Segundo [SNPD 2012] 1,4 por cento da população brasileira tem deficiência mental ou intelectual.</p> <p>A presente pesquisa tem foco na colaboração, provida por tecnologia, para o aumento da autonomia de pessoas com deficiência intelectual</p>
Artefato	O artefato produzido são as telas e os componentes de telas (de <i>smartphones</i>) por onde pessoas com deficiência poderão se comunicar com seus cuidadores
Processo de Busca da Solução	Guiado pelo método de pesquisa <i>Design Science Research</i> utilizou-se conceitos de IHC, mais precisamente do Design Centrado no Usuário e do Design Participativo para pesquisa e desenvolvimento da solução
Rigor da Pesquisa	Para cada ciclo de conhecimento da pesquisa um ou mais conceitos foram utilizados para garantir o rigor da pesquisa. Avaliação dos especialistas, <i>Workshops</i> , observação do uso e avaliação da utilização dos componentes foram utilizados para garantir o rigor da pesquisa durante os ciclos de design
Avaliação	Avaliação de usabilidade, com o envolvimento dos usuários. Utilizando a análise de observação de uso e a análise de utilização de componentes (<i>logs</i>)
Contribuições da Pesquisa	Fortalecimento do uso do <i>Design Science Research</i> na área de Sistemas Colaborativos Análise dos resultados do uso de Design Centrado no Usuário e do Design Participativo no design de Sistemas Colaborativos e Assistivos Registro dos resultados obtidos sobre oportunidades e limites enfrentados pela utilização de dispositivo móvel de comunicação por pessoas com DI
Comunicação da Pesquisa	Visa a comunidade científica interessada no desenvolvimento de Sistemas Colaborativos Assistivos e na Comunicação Aumentativa e Alternativa

Em alguns casos, existem condições de realização destes deslocamentos sem a presença de um cuidador, porém os cuidadores não se sentem seguros em deixar que o DI se desloque sozinho. A insegurança é justificada pela possibilidade da ocorrência de problemas no percurso, os quais a pessoa com DI pode ter dificuldades para resolver sozinha. Por exemplo, pegar um ônibus errado, ter uma crise durante o trajeto, sentir medo ou se sentir perdido.

Para [Glat 2004], o cuidado de pessoas com DI impõe adaptações cotidianas. Tais adaptações podem levar muitas famílias ao isolamento social e conseqüentemente levar a família a reforçar mecanismos de superproteção. [Glat 2004] ainda reforça que, nesses casos, a condição do indivíduo com deficiência pode ser hiperdimensionada em detrimento as suas capacidades e o cuidado pode se tornar exagerado, prejudicando sua independência e autonomia. [Glat 2004] realizou estudos com familiares de pessoas adultas com DI que demonstram a superproteção, evidenciando a necessidade de construção de um espaço ressocializador. A superproteção é evidenciada também por [Rosa and Denari 2012], que encorajam as famílias a oferecer a pessoa com DI maior independência e integração a sociedade.

Neste contexto surgiu a oportunidade de apoiar a autonomia de pessoas com DI por meio do desenvolvimento de um sistema colaborativo de cuidado e localização. Assim, foi idealizado um sistema onde a pessoa com DI e seus cuidadores possam colaborar e dessa forma proporcionar confiança aos cuidadores e um aumento da autonomia de pessoas com DI em seus deslocamentos diários.

Na busca de possibilitar a colaboração dos cuidadores e dos monitorados (com DI), a presente pesquisa teve foco na definição de artefatos de software de comunicação entendíveis para as pessoas com DI. Basicamente, a comunicação é o processo de transferência de informação e implica na transmissão e recepção de mensagens (pensamentos, ideias, desejos e sentimentos) de uma pessoa para a outra [Tomasello and Tomasello 2003]. As capacidades comunicativas são importantes no desenvolvimento e manutenção das relações sociais e pessoas com DI, que frequentemente têm dificuldades de comunicação.

Portanto, no contexto da pesquisa apresentada nesse artigo e na busca do desenvolvimento de uma Comunicação Aumentativa e Alternativa (CAA), o problema de comunicação relatado indicou a oportunidade de desenvolvimento de artefatos de software que possibilitem que pessoas com DI possam colaborar e se comunicar utilizando dispositivos como *smartphones*.

5. Artefatos

Durante o desenvolvimento da pesquisa foi desenvolvido um sistema de monitoramento. Ele possui um “Módulo de Colaboração” utilizado para a iteração da pessoa com DI com o sistema e portanto, este módulo é o foco da pesquisa. O módulo de colaboração possibilitará a troca de mensagens de texto por *chat* entre monitorado com DI e os monitores (cuidadores). O foco da pesquisa é mantido nos artefatos que permitem a comunicação. Portanto, são as telas e os componentes de tela do módulo de colaboração, que possibilitarão a comunicação entre os usuários monitores e monitorados.

Os sistemas colaborativos, em geral, são compostos de ferramentas genéricas, tais como: correio-eletrônico, lista de discussão, fórum, *chat*, entre outros

[Ugulino et al. 2008]. A Figura 1 traz o exemplo de uma tela de *chat* para ilustrar a colaboração de Maria e João que representam monitores que se comunicam por mensagens de texto com José que representa o monitorado com DI que se comunica por símbolos. O sistema colabora informando os eventos do monitoramento (horário de saída, chegada, etc).



Figura 1. Protótipo de tela de *CHAT* do módulo de colaboração

6. Processo de Busca de Solução

O desenvolvimento da presente pesquisa foi dividida em ciclos de design e para melhor representação, cada ciclo foi dividido em três ciclos de conhecimento. O primeiro ciclo de conhecimento é o ciclo intitulado “conhecendo o usuário” e está detalhado na subseção 6.1.

6.1. Ciclo 1 do Design: Conhecendo o Usuário

A Figura 2 demonstra o ciclo de geração de conhecimento, conforme sugere o DSR, para o primeiro ciclo da busca da solução. A entidade “ciência do comportamento”, demonstra o objetivo na busca do conhecimento dos usuários, suas limitações e capacidades. É onde a pesquisa aborda as características dos usuário e como essas características influenciam a utilização do sistema na comunicação.

É utilizada a abordagem de design focado no usuário (UCD, do inglês *User Centred Design*) para fortalecer a estratégia de foco nas necessidades dos usuários. A entidade “ciência de design” demonstra a prototipação das telas de comunicação do sistema. A prototipação, além de demonstrar a funcionalidade de comunicação em desenvolvimento, proporciona subsídios para avaliação. Com a prototipação é possível avaliar se os artefatos de comunicação em desenvolvimento irão atender as necessidades dos usuários.

Finalizando o ciclo, a avaliação de especialistas com base no protótipo identifica sucessos e insucessos dos artefatos de comunicação.

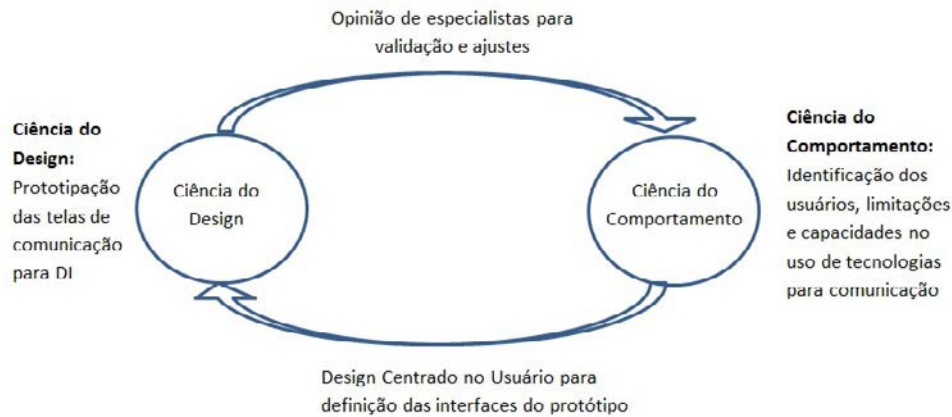


Figura 2. DSR Ciclo 1, adaptado de [Hevner 2007]

O primeiro passo para entendimento dos usuários foi entender a participação de cada usuário na utilização do sistema. Foi realizada uma entrevista aberta para a identificação inicial dos usuários e suas relações com o sistema. Para [Minayo 1993], a entrevista aberta é utilizada quando o pesquisador deseja obter a maior quantidade de informações sobre determinado tema, segundo a visão do entrevistado, e também para obter um maior detalhamento do assunto em questão. Ela é utilizada geralmente na descrição de casos individuais, na compreensão de especificidades culturais para determinados grupos e para comparabilidade de diversos casos.

As entrevistas foram realizadas com profissionais ligados a uma instituição que presta apoio terapêutico a pessoas com DI. Foi realizada uma entrevista com um terapeuta ocupacional e outra com um psicólogo, ambos com contato profissional com os usuários foco do sistema. Diversas foram as contribuições providas da execução das entrevistas, podendo-se destacar a identificação da classificação das pessoas com DI de acordo com sua autonomia. Para identificar o grau de DI de seus pacientes, a instituição envolvida na pesquisa, desenvolveu internamente, uma tabela que os classifica de acordo com a capacidade de autonomia e socialização e relaciona essas características a uma determinada cor, conforme apresentado na Tabela 2.

Após as entrevistas um importante requisito do sistema foi definido: Optou-se por definir que os usuários foco da pesquisa são os usuários que se encaixam nas características da classificação **azul**. Essa definição foi realizada para ir em encontro a um dos objetivos do sistema, que é o aumento da autonomia nos deslocamentos das pessoas com DI, pois nem todos os deficientes intelectuais podem se deslocar sozinhos, como demonstra a Tabela 2.

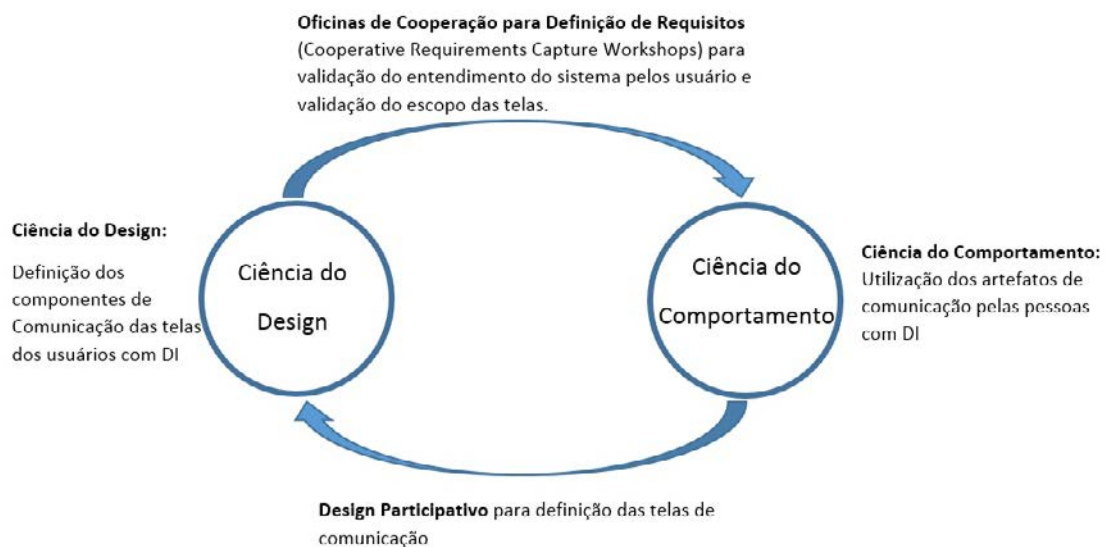
6.2. Ciclo 2 do Design: Definindo os requisitos de usabilidade

A Figura 3 mostra o ciclo de geração de conhecimento para o segundo ciclo do design.

Neste ciclo, a "ciência do comportamento" é definida pela utilização dos artefatos de comunicação pelas pessoas com DI. Em busca de maior assertividade no de-

Tabela 2. Tabela de cores para representação da autonomia das pessoas com DI

Cores	Características
Vermelho	<p>Maior grau de dependência</p> <p>Pouca ou nenhuma autonomia nas tarefas cotidianas</p> <p>Dificuldades severas de locomoção e comunicação</p> <p>Baixa ou nenhuma socialização</p>
Amarelo	<p>Maior autonomia em relação ao “Vermelho”, porém ainda com pouca autonomia</p> <p>Pouca autonomia nas tarefas cotidianas</p> <p>Dificuldades de locomoção e comunicação</p> <p>Baixa socialização</p>
Verde	<p>Maior Autonomia em relação ao “Amarelo”</p> <p>Deficiência intelectual moderada</p> <p>Habilidades de locomoção e comunicação, mesmo apresentando dificuldades</p> <p>Possibilidade de socialização</p>
Azul	<p>Deficiência intelectual leve</p> <p>Realiza pequenas decisões</p> <p>Sociável</p> <p>Podem ser ou não alfabetizados</p> <p>Habilidades de comunicação e locomoção, mesmo apresentando dificuldades</p>

**Figura 3. DSR Ciclo 2, adaptado de [Hevner 2007]**

envolvimento dos artefatos de comunicação, a presente pesquisa utilizou da metodologia de desenvolvimento de sistemas denominada Design Participativo (DP). Para [Muller et al. 1997], o DP envolve um conjunto de teorias, práticas e estudos relacionados com utilizadores (profissionais e trabalhadores) no desenvolvimento de software, hardware ou qualquer atividade relacionada com o computador. A realização de *workshops* foi a prática escolhida do DP para envolvimento dos usuários na definição dos artefatos de comunicação. Já a entidade “ciência de design” demonstra o refinamento das telas

de prototipação do sistema. Os *workshops* possibilitaram a avaliação da usabilidade dos componentes de tela dos protótipos.

Para ser mais específico, foi utilizada a prática de Oficinas de Cooperação para Definição de Requisitos, do inglês *Cooperative Requirements Capture Workshops*. Baseada em uma análise de quem são as partes interessadas em um sistema de computador e quais os problemas de projeto, essa técnica possibilita aos profissionais de software realizarem sete passos para determinar os requisitos de design. Três etapas envolvem o trabalho direto com os usuários e as demais fornecem o contexto dentro do ciclo de vida do desenvolvimento. Os sete passos são respectivamente: análise de partes interessadas, identificação do problema, formular a equipe de entre as quatro classes de partes interessadas (descritas no modelo de participação da técnica), explorar o ambiente do usuário por meio de uma oficina (*User Workshop*), validar o entendimento de ambiente de usuário, identificar o escopo da solução e validar o escopo.

A Engenharia Semiótica foi proposta por [De Souza 2005] para o design de linguagens de interface e tem como base teórica a semiótica, disciplina que estuda signos e linguagens de produção de significado e sentido [da Rocha and Baranauskas 2003]. Para [Barbosa and da Silva 2010] com o intuito de tornar as interfaces amigáveis e naturais, e assim, mais fáceis de serem usadas e menos hostis, são utilizados elementos gráficos para representar comandos, dados ou funções do sistema. [Barbosa and da Silva 2010] ainda definem que em uma interface de tela, um signo é uma mensagem codificada pelo designer para se comunicar com usuário.

O objetivo do *workshop* é trabalhar com os usuários com DI, buscando a identificação do melhor entendimento das telas de comunicação do sistemas voltadas para estes usuário. Ou seja, é utilizar o *workshop* como ferramenta do DP, para definir os artefatos gráficos das telas de comunicação utilizadas pelos usuários com deficiência intelectual, visando a usabilidade e possibilitando assim comunicação.

6.3. Ciclo 3 do Design: Avaliando a Usabilidade

O terceiro ciclo do design busca avaliar se o emprego de técnicas do Design Centrado no Usuário e do Design Participativo, durante as fases de levantamento de requisitos e desenvolvimento da solução, resultaram em uma boa usabilidade do sistema por parte das pessoas com deficiência intelectual.

Um ambiente controlado será preparado para a avaliação da usabilidade. A usabilidade será medida por duas técnicas: A observação do uso e a utilização de componentes. O primeiro se dá pela observação da utilização do sistema pelo usuário e cabe ao observador avaliar o entendimento e capacidade de utilização do usuário. Já a avaliação de utilização dos componentes será a análise dos registros da utilização (*logs*) e das mensagens do *chat*.

A Figura 4 mostra o ciclo de geração de conhecimento para o terceiro e último ciclo do design.

7. Rigor da Pesquisa

De acordo com [Thomas and Hatchuel 2009], para que uma pesquisa seja confiável, ela deve se preocupar não somente com a relevância, mas também com o rigor, que deve estar presente desde a sua condução até a apresentação de seus resultados.



Figura 4. DSR Ciclo 3, adaptado de [Hevner 2007]

O rigor da presente pesquisa está registrada nas Figuras 2, 3 e 4, na opinião de especialistas, oficinas de cooperação e na observação de uso e análise dos componentes, respectivamente.

O primeiro ciclo desta pesquisa resultou na construção de protótipos das telas de comunicação do sistema com base nos conceitos da abordagem UCD. A validação do artefato gerado nesse ciclo foi um método de avaliação analítico chamado Opinião de Especialistas. Para [Prates and Barbosa 2003], métodos de avaliação analíticos são aqueles nos quais avaliadores inspecionam ou examinam aspectos de uma interface de usuário relacionados com a usabilidade.

8. Conclusão e Trabalhos Futuros

A presente pesquisa contempla a realização do design da comunicação em um sistema colaborativo de monitoramento de pessoas com Deficiência Intelectual (DI), produzindo artefatos que possibilitem a comunicação de pessoas com DI com seus cuidadores. Pretende-se assim, proporcionar confiança aos cuidadores e um aumento da autonomia de pessoas com DI em seus deslocamentos diários. A pesquisa foi guiada pela metodologia de pesquisa *Design Science Research* (DSR) e foi dividida em três ciclos de descoberta de conhecimento.

No primeiro ciclo foi gerado um protótipo de baixa fidelidade como artefato da “ciência do Design”. A “ciência do comportamento”, do primeiro ciclo, é definida pela necessidade de conhecimento dos usuários, que no contexto deste estudo, são usuários com capacidades cognitivas reduzidas. O primeiro ciclo se caracterizou pela utilização da abordagem do Design Centrado no Usuário.

O segundo ciclo tem o objetivo de validação e ajustes no protótipo obtido no ciclo 1. Os artefatos do ciclo 2 são os componentes de tela de comunicação, utilizadas pelos usuários com deficiência intelectual e definidos com o apoio destes usuários e baseados

nas abordagens do Design Participativo. Nessa etapa, foram utilizadas como ferramenta as Oficinas de Cooperação para Definição de Requisitos.

O terceiro ciclo da descoberta de conhecimento será executado. A definição do terceiro ciclo está descrita na seção 6.3 do presente artigo, porém os resultados ainda não foram coletados e analisados. O terceiro ciclo usa o resultado da avaliação de usabilidade como o artefato da ciência do design e utiliza um ambiente controlado de utilização para avaliação por observação de uso e utilização de componentes do sistemas (*logs*).

Podemos concluir que o DSR tem potencial para produção de conhecimento no processo de design de sistema colaborativo e assistivo. Destaca-se também a importância das abordagens de Design Centrado no Usuário e do Design participativo como ferramenta para o design da comunicação aumentativa e alternativa no contexto do sistema proposto.

Como trabalhos futuros, a avaliação de usabilidade, em um ambiente controlado de uso, possibilitará a validação das ferramentas utilizadas para a construção do sistema, assim como trará uma nova contribuição científica com a análise dos resultados da avaliação.

Referências

- Aken, J. E. v. (2004). Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules. *Journal of management studies*, 41(2):219–246.
- Almeida, M. S. R. (2005). Caminhos para a inclusão humana. *Porto: Edições Asa*.
- Barbosa, S. D. J. and da Silva, B. S. (2010). *Interação humano-computador*. Elsevier.
- Bax, M. P. (2015). Design science: filosofia da pesquisa em ciência da informação e tecnologia. *Ciência da Informação*, 42(2).
- Chang, Y.-J., Liu, H.-H., and Wang, T.-Y. (2009). Mobile social networks as quality of life technology for people with severe mental illness. *Wireless Communications, IEEE*, 16(3):34–40.
- da Rocha, H. V. and Baranauskas, M. C. C. (2003). *Design e avaliação de interfaces humano-computador*. Unicamp.
- de Araújo, L. P., Berkenbrock, C. D. M., and Mattos, M. M. (2014). Usando pesquisa-ação no desenvolvimento de um sistema colaborativo para tratamento multidisciplinar na rede do sus.
- De Souza, C. S. (2005). *The semiotic engineering of human-computer interaction*. MIT press.
- Druin, A. (2009). *Mobile technology for children: Designing for interaction and learning*. Morgan Kaufmann.
- Fink, P. J. (1992). *Stigma and mental illness*. American Psychiatric Pub.
- Fuks, H. (2011). *Sistemas colaborativos*. Elsevier Brasil.
- Glat, R. (2004). Uma família presente e participativa: o papel da família no desenvolvimento e inclusão social da pessoa com necessidades especiais. In *Anais do 9º Congresso Estadual das APAEs de Minas Gerais*, pages 1–7.

- Grudin, J. (1994). Computer-supported cooperative work: History and focus. *Computer*, (5):19–26.
- Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, 19(2):4.
- Lacerda, D. P., Dresch, A., Proença, A., and ANTUNES JUNIOR, J. (2013). Design science research: método de pesquisa para a engenharia de produção. *Gestão & Produção*, 20(4):741–761.
- Minayo, M. C. d. S. (1993). O desafio do conhecimento científico: pesquisa qualitativa em saúde. *São Paulo/Rio de Janeiro: Hucitec-Abrasco*, pages 01–10.
- Muller, M. J., Haslwanter, J. H., and Dayton, T. (1997). Participatory practices in the software lifecycle. *Handbook of human-computer interaction*, 2:255–297.
- Petrie, H., Hamilton, F., King, N., and Pavan, P. (2006). Remote usability evaluations with disabled people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1133–1141, New York, NY, USA. ACM.
- Prates, R. O. and Barbosa, S. D. J. (2003). Avaliação de interfaces de usuário—conceitos e métodos. In *Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação, Capítulo*, volume 6.
- Pries-Heje, J., Baskerville, R., and Venable, J. (2008). Strategies for design science research evaluation. *ECIS 2008 proceedings*, pages 1–12.
- Radabaugh, M. (1993). Selecting and obtaining assistive technology—ibm national support center for persons with disabilities. *Interagency Project for assistive technology—IPAT. North Dakota: IPAT*.
- Rocha, E. B., Pimentel, M., Diniz, M. C., and Santoro, F. M. (2015). Design science research para o desenvolvimento de um modelo da participação em bate-papo. *iSys-Revista Brasileira de Sistemas de Informação*, 8(1):18–41.
- Rosa, F. D. and Denari, F. E. (2012). Trabalho, educação e família: perspectivas para a pessoa com deficiência intelectual. *Revista Educação Especial*, 26(45):73–90.
- Satcher, D. (2000). Mental health: A report of the surgeon general—executive summary. *Professional Psychology: Research and Practice*, 31(1):5.
- SNPD (2012). Cartilha do censo 2010 - pessoas com deficiências.
- Thomas, H. and Hatchuel, A. (2009). A foundationalist perspective for management research: A european trend and experience. *Management Decision*, 47(9):1458–1475.
- Tomasello, M. and Tomasello, M. (2003). Comunicação linguística e representação simbólica. *M. Tomasello, Origens culturais da aquisição do conhecimento humano*, pages 131–186.
- Ugulino, W., Nunes, R. R., Oliveira, C. L., Pimentel, M., and Santoro, F. M. (2008). Dos processos de colaboração para as ferramentas: A abordagem de desenvolvimento do projeto communicatec. In *Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web*, WebMedia '08, pages 233–240, New York, NY, USA. ACM.
- Wahl, O. F. (1999). Mental health consumers' experience of stigma. *Schizophrenia bulletin*, 25(3):467.

Um Estudo Empírico sobre Crowdsourcing: O Brasil na plataforma TopCoder

Carlos H. Lopes¹, Ricardo Rodrigo M. Melo¹, Renato P. Ferreira¹, Muller G. Miranda¹, Cleidson De Souza^{1,2}

¹Programa de Pós Graduação em Ciência da Computação – Universidade Federal do Pará (UFPA) – Belém – PA – Brasil

²Instituto Tecnológico Vale – Belém – PA – Brasil

{carloshugolopes, ricardorodrigomm, renpina, mulgsm}@gmail.com,
cleidson.desouza@acm.org

Abstract. *This paper describes an analysis of Brazilian software developers in the largest software engineering crowdsourcing platform, TopCoder. We compare Brazilian active coders in this platform with active coders from other countries, especially BRICS, Latin America and Western Europe. Our results can be used to further Brazilian participation in the platform.*

Resumo. *Este artigo descreve uma análise quantitativa dos programadores do Brasil que atuam na maior plataforma crowdsourcing do mundo na área de engenharia de software, a TopCoder. Além disso, considerando apenas os desenvolvedores ativos, uma análise é feita comparando os brasileiros com os programadores de outros países, assim como os brasileiros comparados com os programadores das regiões do BRICS, América Latina e Leste Europeu. Os resultados deste estudo podem ser utilizados para melhorar a participação dos engenheiros de software brasileiros.*

1. Introdução

Segundo Brabham (2008), o termo *crowdsourcing* foi cunhado por Howe e Robinson em Junho de 2006. Esse termo descreve um novo modelo de negócio baseado na *web* que aproveita os conhecimentos e soluções de pessoas geograficamente distribuídas, mas conectadas via Internet. Por exemplo, uma pessoa ou companhia divulga, por meio da Internet, um problema a ser resolvido. Um número vasto de pessoas oferecem soluções para o problema e a recompensa é dada para a solução vencedora. Neste contexto, ambos se beneficiam: tanto a empresa por obter a solução de seu problema ou produto rapidamente e com melhores custos financeiros, quanto o trabalhador que desenvolveu a solução ao ter a possibilidade de ganhar uma recompensa, seja em dinheiro, prêmios, ou até mesmo a própria fama (Malone et al. 2010). A InnoCentive é um exemplo de empresa que oferece prêmios para pesquisadores que apresentam soluções para um problema científico (Zatti 2013) (Malone et al. 2011).

O surgimento do *crowdsourcing* foi proporcionado pelo conceito de inteligência coletiva. Segundo (Alves et al. 2011), a inteligência coletiva é definida como uma inteligência presente em todas as partes do mundo, coordenada em tempo real, e que resulta em competências eficientes continuamente valorizadas. A multidão colabora em

um objetivo comum e pode até mesmo gerar resultados muito melhores do que os feitos por uma só pessoa. Esses conceitos foram potencializados com o advento das novas tecnologias de comunicação, como a Internet. Alguns pesquisadores (e.g., (Malone et al. 2011)) vêem o modelo do *crowdsourcing* como uma nova forma de como o trabalho é ou será feito pelas empresas. Em particular, Malone e colegas afirmam que uma nova era, a era da hiperespecialização, está se iniciando. Ele afirma ainda que muito das riquezas que o mundo está aproveitando atualmente vem do ganho da produtividade dividindo o trabalho em tarefas cada vez menores realizadas por trabalhadores cada vez mais especialistas. Hoje, isso ocorre graças ao aumento do conhecimento do trabalho e das tecnologias de comunicações.

Um exemplo baseado no modelo de *crowdsourcing* é a plataforma TopCoder que é disponibilizada pela empresa Appirio onde os clientes fornecem seus projetos para a TopCoder que separa os mesmos em módulos. Cada módulo é aberto para mais de 600.00 programadores inscritos na plataforma. O objetivo é que eles completem o módulo dentro de um prazo determinado. Os desenvolvedores da melhor solução ganham um prêmio financeiro (Birkinshaw and Goddard 2009).

É objetivo deste trabalho analisar os programadores brasileiros quanto à participação (por meio da quantidade de competições), classificação e linguagens de programação favoritas na plataforma TopCoder. Esta análise tem como intuito verificar como esses desenvolvedores do Brasil se comparam a desenvolvedores de outros países. Desta forma, pretende-se entender o potencial das atividades de desenvolvimento de *software* utilizando *crowdsourcing* no Brasil. Além disso, os resultados desta pesquisa poderiam ser utilizados por outras empresas interessadas em explorar o potencial de atividades colaborativas como o *crowdsourcing* em seus contextos. Vale ressaltar que incentivar a participação brasileira em atividades de *crowdsourcing* pode criar novas oportunidades de renda e mobilidade social (Kittur, 2013).

O restante deste artigo está organizado como segue. A próxima seção apresenta o conceito de *crowdsourcing* e detalha a plataforma TopCoder. A seção 3 descreve a metodologia utilizada neste trabalho, incluindo como os dados foram coletados. As próximas seções, 4 e 5, apresentam os resultados do trabalho bem como a discussão dos mesmos. Finalmente, a seção 6 descreve as conclusões e trabalhos futuros.

2. *Crowdsourcing* e o TOPCODER

2.1. Visão Geral

O termo *crowdsourcing* descreve um novo modelo de negócio baseado na *web* que aproveita as soluções de uma rede distribuída de pessoas por meio de um “convite” aberto à apresentação de propostas. Uma das importâncias do *crowdsourcing* é que ele permite o estabelecimento de espaços de idéias compartilhadas para a produção de inovação. Além disso, o *crowdsourcing* pode promover maneiras de como o trabalho deve ser feito, de maneira rápida, sob controle e baixo custo (Malone et al. 2011).

Um exemplo bastante conhecido de *crowdsourcing* é o *Amazon Mechanical Turk*¹ (AMT) da empresa Amazon. Este serviço utiliza a inteligência humana para realizar tarefas chamadas de *Human Intelligence Tasks*. Este serviço permite que empresas acessem programaticamente uma mão-de-obra diversificada sob demanda. A justificativa por trás desta plataforma é que os humanos conseguem ser mais eficientes do que os computadores em algumas tarefas como por exemplo, identificar objetos em uma imagem ou vídeo, transcrever áudios, remover arquivos duplicados a partir de um diretório, entre outros exemplos. O AMT disponibiliza APIs (*Application Programming Interfaces*) para que os desenvolvedores acessem milhares de trabalhadores remotos. Além disso, as APIs permitem integrar programaticamente os resultados obtidos diretamente em projetos corporativos ou acadêmicos. Por exemplo, Hara e colegas (2013) utilizam o AMT junto com *Google Street View* para que a multidão identifique problemas de acessibilidade em calçadas de cidades dos Estados Unidos.

2.2. A utilização de *Crowdsourcing* por Empresas

Existem vários casos de empresas que adotam o *crowdsourcing*, incluindo a Wikipédia (Huberman et al 2009) e a Innocentive que entrega prêmios pelas soluções desenvolvidas (Zatti 2013). Um dos problemas que a Innocentive disponibilizou para a multidão² foi o vazamento de óleo no golfo do México. Neste caso, 900 soluções foram propostas para o problema. Outro exemplo de uso do serviço da Innocentive foi da empresa farmacêutica Roche que queria encontrar a melhor forma de medir o volume e a qualidade de espécimes clínicos. Em 2008, a empresa patrocinou um concurso recebendo, depois de dois meses, 113 propostas, isto é, muitas soluções em um curto espaço de tempo (Malone et al 2011).

A Netflix, uma empresa de TV por internet dos Estados Unidos com mais de 37 milhões de assinantes em 40 países³, em 2006, criou uma campanha à multidão em geral para encontrar um algoritmo de recomendação mais rápido que o seu. Ela ofereceu US\$1.000.000 para a equipe vencedora (Zatti 2013, Schweissguth 2012). O sucesso foi tão satisfatório que, segundo (Lohr 2009), a empresa pretendia apresentar um novo concurso onde os competidores teriam de dar relevância para dados demográficos e comportamentais com o intuito de modelar os “gostos” dos clientes.

A empresa Dell possui uma plataforma de *crowdsourcing* chamada *IdeaStorm*, onde já conseguiu integrar 280 sugestões e ideias em sua linha de produtos, como a sugestão de montar uma linha de computadores baratos rodando Linux. Seguindo a mesma linha da empresa Dell, o Innovation Jam, definido como um gigantesco brainstorming virtual onde funcionários, clientes e mesmo familiares de funcionários contribuem ativamente, e em tempo real, para a geração de ideias (Schneider et al. 2011).

Além desses, o My Starbucks Idea, é uma a plataforma para colher ideias com uma equipe dedicada em colher informações e envolver a comunidade nos seus

¹ <http://aws.amazon.com/pt/mturk/>

² *Crowd*.

³ <https://signup.netflix.com/MediaCenter>

problemas. Hoje ela é referência em gestão de ideias no mundo todo (Zatti 2013). Tal iniciativa mostra como o *crowdsourcing* busca a determinação das necessidades dos usuários e o que eles querem, a fim de conhecê-los (Schweissguth 2012).

Finalmente, um último exemplo é da Procter & Gamble, uma empresa pioneira em desenvolver soluções de problemas por meio do *crowdsourcing*: ela já realiza esta atividade há mais de 15 anos (Zatti 2013).

2.3. *Crowdsourcing* e Engenharia de Software: A plataforma TopCoder

No contexto de engenharia de *software*, a plataforma TopCoder é um dos exemplos mais famosos de *crowdsourcing*. Esta plataforma administra competições de programação chamadas de SRMs (*Single Round Matches*). Estas competições são realizadas quinzenalmente. Esta plataforma foi criada em Novembro de 2000 e vendida para a empresa Appirio em Setembro de 2013 (Lunden 2013). Ela reúne especialistas do mundo todo em design, desenvolvimento, testes entre outras áreas da engenharia de software. O objetivo do TopCoder é beneficiar tanto os *coders* (programadores de sistemas) e quanto as empresas. Os *coders* tem a oportunidade de serem reconhecidos, de melhorar suas habilidades, demonstrar e receber recompensas por seus conhecimentos. Por outro lado, as empresas buscam boas soluções feitas de maneira eficiente, isto é, rápida e com menor custo.

Com mais de 600.000 coders cadastrados na plataforma, a comunidade TopCoder constrói *softwares* e plataformas para diversos clientes. Isto é acompanhado por testes contínuos e processos de gestão, assim, a plataforma coloca para seus usuários uma nova forma de resolução de problemas que é alcançada através de desenvolvedores espalhados ao redor do mundo.

Na prática, o TopCoder “quebra” os projetos de seus clientes em pedaços e oferece os mesmos para a multidão de desenvolvedores freelances como um ou mais desafios competitivos. Por exemplo, um projeto pode começar com um desafio para resultar na melhor ideia de produto de *software*. Um segundo desafio pode prover um alto nível de detalhamento dos objetivos dos projetos e desafiar desenvolvedores para criar os documentos que melhor traduzem os requisitos dos sistemas⁴. O documento de especificação vencedor pode tornar-se a base do próximo desafio em que outros desenvolvedores competem para o desenho da arquitetura do sistema, especificando as partes necessárias do *software* e as conexões entre eles. Outros desafios são lançados para desenvolver, separadamente, cada pedaço e depois integrá-los. Finalmente, um último desafio pode ser lançado para encontrar e corrigir *bugs* em diversas partes do sistema (Malone et al. 2011).

3. Metodologia

Um programa em Java (*parser*) foi desenvolvido para extrair dados de documentos HTML obtidos a partir da plataforma TopCoder. O programa inicia com a obtenção do

⁴ Existe um fórum na página do TopCoder que permite aos desenvolvedores consultarem os clientes para mais detalhes, e essas perguntas e respostas são visíveis para todos os competidores.

document HTML da página de estatísticas do TopCoder⁵, depois obtém os dados relevantes de cada coder disponíveis no documento como por exemplo: nome, nacionalidade, *rating* (pontuação dentro da plataforma), quantidade de competições, linguagem de programação mais usada e por fim armazena esses dados em um banco de dados, desta forma permitindo consultas e análises. A biblioteca *jsoup*⁶ foi utilizada na construção deste *parser* (Heaton 2002).

Considerando que este trabalho visa analisar, quantitativamente, o desempenho dos *coders* brasileiros no TopCoder, mais especificamente em relação aos países que formam o BRICS, o Leste Europeu e a América Latina, foi necessário realizar consultas nos dados obtidos que retornassem resultados numéricos desses países, tais como: (i) número máximo, mínimo, média e desvio padrão de competições dos *coders*; (ii) quantidade de *coders*; (iii) quantidade de linguagem-padrão escolhida; (iv) e classificação.

Os resultados apresentados na próxima seção são relatados a partir dos *coders* ativos na classificação disponível no site da comunidade⁷. Essa classificação representa os *coders* ativos ordenados, de maneira decrescente, por pontuação. Um *coder* é considerado ativo quando participa de pelo menos uma competição no período de 180 dias. Estes, por sua vez, estão distribuídos em 134 países, incluindo o BRICS (Brasil, Rússia, Índia, China e África do Sul), 15 países da América Latina e 14 do Leste Europeu.

4. Resultados

4.1. Resultados Gerais

A Tabela 1 a seguir apresenta os dez Países mais bem colocados no Topcoder, bem como os melhores e os últimos *coders* colocados por país.

Tabela 1. Classificação dos *coders* por País

Colocação	País	Número de <i>coders</i> ativos	Melhor <i>coder</i> colocado	Último <i>coder</i> Colocado
1°	Rússia	531	1°	8.822°
2°	Japão	653	9°	8.874°
3°	China	1432	7°	8.883°
4°	Ucrânia	232	13°	8.866°
5°	Polónia	204	6°	8.886°
6°	Bielorrússia	92	2°	8.725°
7°	Estados Unidos	579	16°	8.905°

⁵ <http://community.topcoder.com/tc?module=AlgoRank>

⁶ <http://jsoup.org>

⁷ <http://community.topcoder.com/tc>

8º	Croácia	78	69º	8.518º
9º	Iran	94	32º	8.844º
10º	Coreia do Sul	122	70º	8.885º

Em relação ao número de competições, até a escrita deste artigo tinham sido realizadas 603 competições do tipo SRM. O *coder* mais ativo da plataforma, isto é, aquele que participou do maior número de competições é um *coder* dos EUA com 401 competições. Entretanto, o número médio de competições por *coder* ativo é 21,81, o desvio padrão é de, aproximadamente, 37,99 e o mínimo é de 1 competição. Deve-se observar que quantidade de participação em competições não implica em melhor classificação, visto que o *coder* mais ativo ocupa a 1.116ª posição.

Em relação às linguagens de programação, os *coders* mais ativos escolhem como linguagem-padrão: C++, Java, C#, Python e Visual Basic (ver Figura 1).

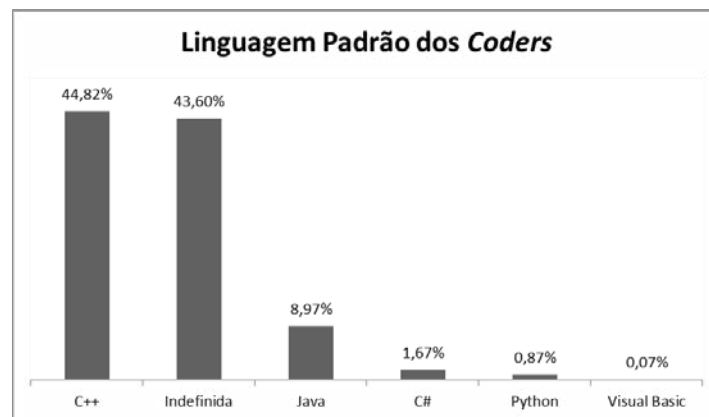


Figura 1. Linguagem de programação padrão escolhida pelos *coders*

4.2. Resultados dos brasileiros

O Brasil está presente no TopCoder e relativamente bem colocado: entre os 13 primeiros Países. Quando da coleta de dados, o Brasil possuía 147 *coders* presentes na classificação geral dos programadores ativos, com o melhor colocado ocupando a 45ª posição (Tabela 2) e o último colocado ocupando a 8825ª posição.

Tabela 2. Classificação dos dez brasileiros mais bem colocados

<i>Coder</i>	Competições	Posição	Linguagens
Brasileiro 1	82	45º	C++
Brasileiro 2	49	151º	C++
Brasileiro 3	92	165º	C++
Brasileiro 4	58	184º	C++
Brasileiro 5	47	292º	C++
Brasileiro 6	291	438º	C++
Brasileiro 7	64	452º	C++
Brasileiro 8	20	549º	C++
Brasileiro 9	38	668º	C++
Brasileiro 10	92	700º	C++

Os brasileiros participaram em média de 25,28 competições, apenas um pouco acima da média mundial. O brasileiro mais participativo esteve presente em 342 competições diferentes. Por outro lado, o *coder* brasileiro que menos participou de competições, esteve presente em apenas 1. Levando em consideração a linguagem de programação mais utilizada pelos brasileiros, o C++ é a linguagem padrão, refletindo a mesma linguagem padrão mundial, conforme a Figura 2.

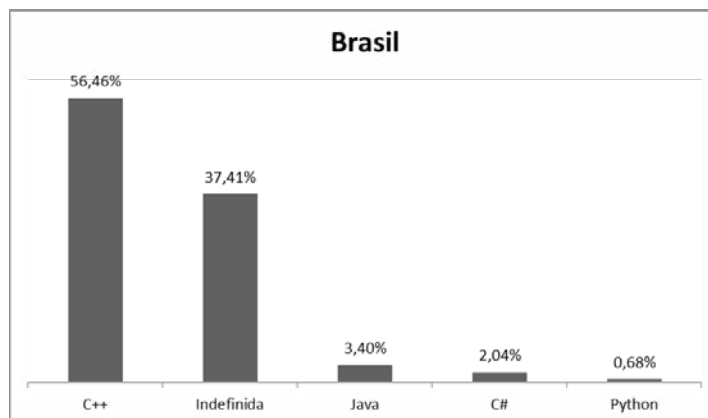


Figura 2. Linguagem de programação padrão escolhida pelos brasileiros

4.3. Brasil em relação a América Latina

Para fins de comparação, tendo os dados obtidos neste trabalho como base, definiu-se que seria traçado, primeiramente, um paralelo entre os programadores brasileiros e os pertencentes à América Latina. Como se pode observar, o Brasil possui o maior número de *coders* ativos, bem como o melhor colocado no *ranking* da TopCoder. Dentre os países da América Latina, o Brasil é o mais bem colocado. O país que mais se aproxima no referido *ranking* é a Argentina, ocupando a 29ª posição, seguida pelo Peru em 33°.

Tabela 3. Classificação dos *coders* por Países da América Latina

Colocação	País	Número de <i>coders</i> ativos	Melhor <i>coder</i> colocado	Último <i>coder</i> colocado
13°	Brasil	147	45°	8825°
29°	Argentina	44	274°	8292°
33°	Peru	86	360°	8898°
38°	Mexico	49	478°	8765°
44°	Colômbia	50	444°	8760°
48°	Bolivia	44	475°	8807°
54°	Venezuela	18	1403°	8871°
58°	Republica Dominicana	12	920°	8074°

Além disso, os dados de número de competições e linguagem de programação foram analisados. No que diz respeito ao número de competições a média entre o Brasil

e a América Latina é a praticamente a mesma por desenvolvedor ranqueado, como pode ser visto na Figura 3.

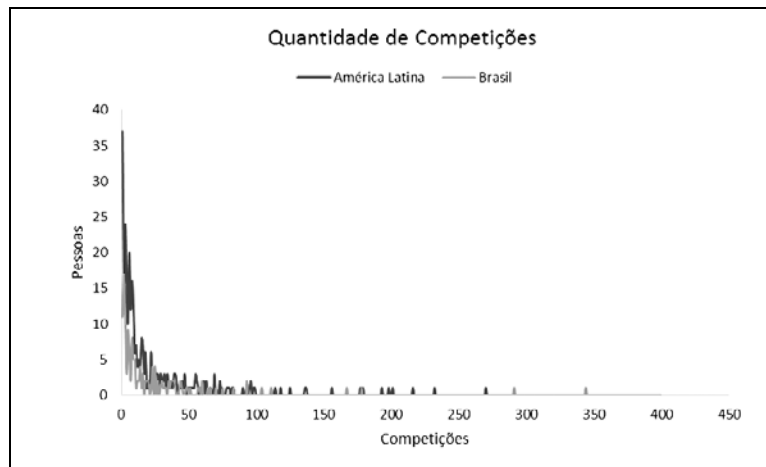


Figura 3. Comparação de quantidades competições dos coders brasileiros em relação a América Latina.

Quanto à linguagem de programação, os brasileiros, bem como o restante dos Países latino-americanos, utilizam mais o C++ como linguagem padrão, seguido pelo Java (Figura 4). Este padrão é similar ao do resto do mundo conforme apresentado na Figura 1 anteriormente.

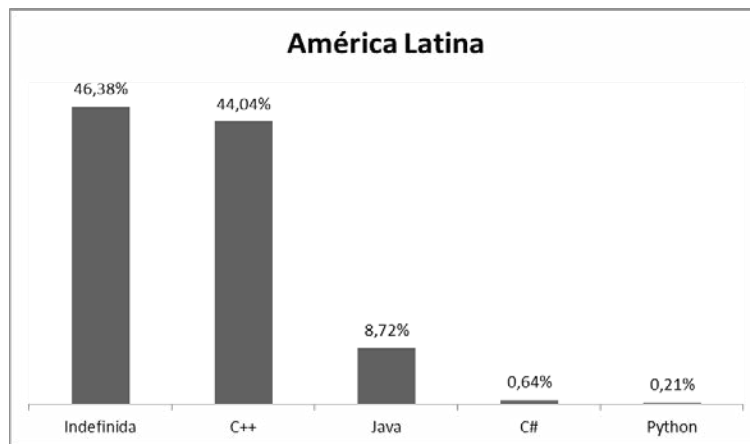


Figura 4. Linguagem de programação padrão escolhida pelos coders da América Latina.

4.4. Brasil em relação ao BRICS

É interessante analisar também como o Brasil está em relação aos países que formam o BRICS, pois os brasileiros fazem parte deste grupo político. A partir dos dados foram analisadas as seguintes características: número de competições, *ranking* e linguagem de programação. Em relação ao número de competições, a média do Brasil é maior comparado a do BRICS que é de 16,74 competições por *coder* ativo. Porém, a Figura 5 mostra que os *coders* pertencentes aos países que formam as siglas BRICS participam de um maior número de competições que os brasileiros. Já o número máximo de

competições, o brasileiro que mais competiu perde para o russo (primeiro do *ranking* geral) que competiu 346 vezes.

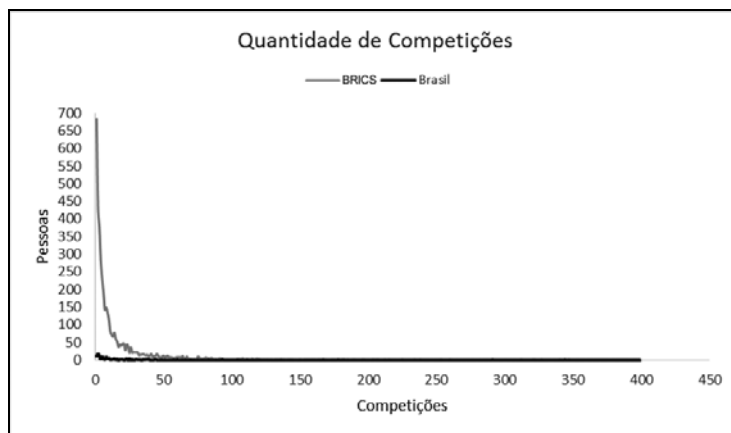


Figura 5. Comparação de quantidades de competições dos coders brasileiros em relação aos outros Países do BRICS.

Conforme a Tabela 4, o Brasil, em relação aos países do BRICS, é o terceiro colocado e está atrás da China e Rússia. É importante destacar que a Rússia é o país que ocupa a primeira posição e a China a terceira posição da classificação geral dos países. A tabela abaixo mostra a classificação dos países do BRICS em relação a todos os competidores ativos do TopCoder.

Tabela 4. Classificação dos coders por Países do BRICS

Colocação	País	Número de <i>coders</i> ativos	Melhor <i>coder</i> colocado	Último <i>coder</i> colocado
1º	Rússia	531	1º	8822º
3º	China	1432	7º	8883º
13º	Brasil	147	45º	8825º
18º	Índia	2264	217º	8906º
36º	África do Sul	11	4º	8767º

Observando a Figura 6, a maioria dos *coders* pertencentes ao BRICS escolhem o C++ como linguagem padrão.

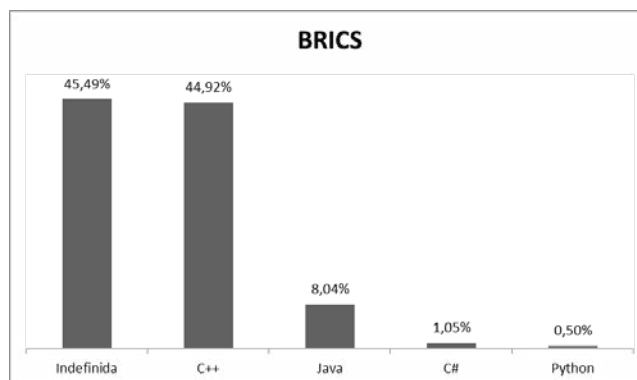


Figura 6. Linguagem de programação padrão escolhida pelos coders do BRICS

4.5. Brasil em relação ao Leste Europeu

Conforme visto anteriormente, o Brasil está atualmente na 13ª posição da classificação geral. Entretanto, em uma análise comparativa apenas com os países do Leste Europeu, o Brasil se encontra na 6ª colocação se comparado apenas aos 14 países presentes no *ranking* do TopCoder e que fazem parte do Leste Europeu: Belarus, Bulgária, República Checa, Croácia, Eslováquia, Hungria, Letônia, Lituânia, Macedônia, Polônia, Romênia, Rússia, Servia e Ucrânia.

Tabela 5. Classificação dos dez *coders* mais bem colocados por Países do Leste Europeu.

Colocação	País	Número de <i>coders</i> ativos	Melhor <i>coder</i> colocado	Último <i>coder</i> colocado
1º	Rússia	531	1º	8822º
4º	Ucrânia	232	14º	8866º
5º	Polônia	204	6º	8886º
6º	Bielorrússia	92	2º	8725º
8º	Croácia	78	69º	8518º
13º	Brasil	147	45º	8825º
14º	Bulgária	69	128º	8897º
17º	Eslováquia	43	60º	7548º
21º	România	106	208º	8538º
28º	Lituânia	12	292º	6518º
30º	República Checa	22	78º	8003º

Outra comparação refere-se a média de competições por desenvolvedor brasileiro que é de 25,285, enquanto que para média do conjunto de Países que compõem o Leste Europeu é de 36,097 competições. Essa disparidade de participação média em competições é comprovada na Figura 7.

A Figura 8 apresenta a escolha das linguagens de programação dos *coders* do Leste Europeu. Novamente, a linguagem C++ apresenta-se como a linguagem mais utilizada.

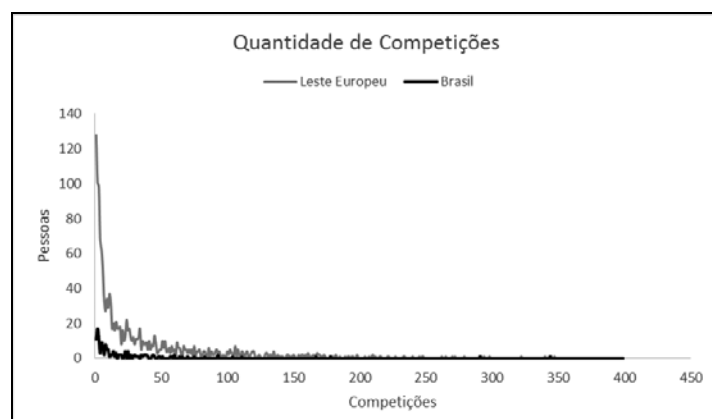


Figura 7. Comparação de quantidades de competições dos *coders* brasileiros em relação aos países do Leste Europeu.

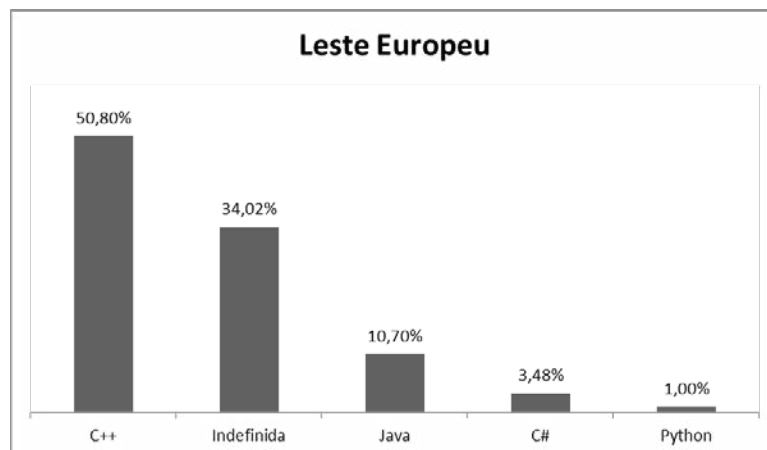


Figura 8. Linguagem de programação padrão escolhida pelos *coders* do Leste Europeu.

5. Discussão

5.1. Visão Geral

A Figura 9 abaixo apresenta o total de pesquisas feitas, em relação ao número total de pesquisas realizadas no *Google* ao longo do tempo para a palavra “*Crowdsourcing*”. Estas são as pesquisas feitas em todo o mundo no período de outubro de 2009 a maio de 2015. O figura abaixo é resultado de uma pesquisa utilizando o *Google Trends*⁸.

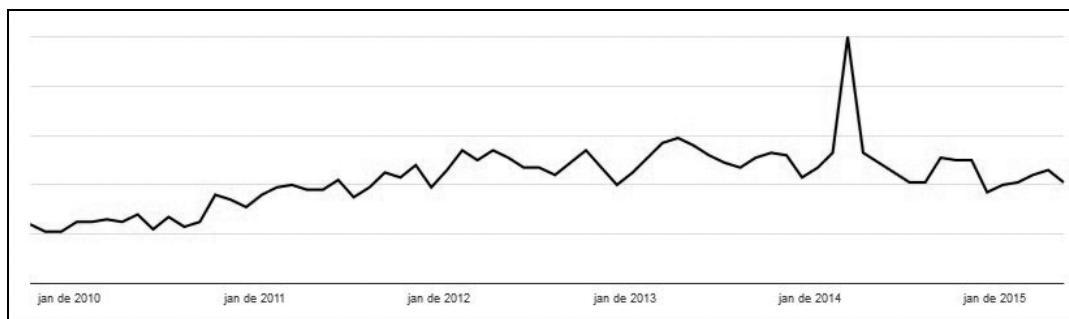


Figura 9. Pesquisa pelo termo *Crowdsourcing* no *Google*

A Figura 9 sugere um crescimento genérico no interesse deste novo modelo de produção. Por ser um tema relativamente recente, nos últimos anos a comunidade científica também tem se esforçado para entender melhor este fenômeno com diversos artigos nas conferências nacionais (SBSC) e internacionais (CSCW e CHI) incluindo (Chan et al, 2016) (Antonio et al, 2014) (Lykourantzou et al, 2016).

Os resultados anteriores sugerem que o Brasil está relativamente bem colocado na plataforma TopCoder, especialmente quando comparado aos outros países da América Latina, mas nem tanto quanto comparados aos outros países do BRICS e os países do Leste Europeu. Entretanto, vale ressaltar que o número de brasileiros ativos no

⁸ <http://www.google.com/trends>

TopCoder ainda é pequeno, com apenas 147 desenvolvedores quando da coleta dos dados utilizados neste artigo.

Assim, um estudo empírico sobre a participação brasileira no TopCoder pode trazer incentivos para a indústria nacional, possibilitando investigar fatores para melhoria da participação no TopCoder como mostram as evidências de outros autores (Li et al, 2013). Os autores acreditam que este trabalho contribuirá para o aumento do conhecimento sobre o *crowdsourcing* no contexto do desenvolvimento de *software* pela comunidade científica brasileira. Isto tem o potencial de estimular outros estudos sobre a participação brasileira neste modelo de negócio. Isto é particularmente importante no atual contexto econômico do Brasil, pois o *crowdsourcing* também cria novas oportunidades de renda e mobilidade social em regiões do mundo onde as economias locais podem ser estagnadas ou sem investimento (Kittur, 2013).

5.2. Trabalhos Relacionados

Diversos aspectos têm sido estudados no contexto de *crowdsourcing*, e ainda mais recentemente, como o *crowdsourcing* pode trazer ganhos para as atividades de desenvolvimento de *software*. Neste contexto, o *crowdsourcing* pode ser definido como “uma realização de determinadas tarefas de desenvolvimento de *software* em nome de uma organização por um grande grupo indefinido de pessoas externas com os conhecimentos técnicos especializados através de um chamado aberto” (Stol, 2014). Além disso, essa abordagem usa uma multidão online para terceirizar tarefas, incluindo requisitos, projeto, codificação, testes, evolução e documentação (Leicht, 2015).

Conforme mencionado anteriormente, a plataforma TopCoder é uma das maiores sendo utilizada no contexto de desenvolvimento de software. Assim, ela tem sido estudada por diferentes pesquisadores. Por exemplo, Douglas (2014) discute algumas características da comunidade TopCoder coletadas através de um *survey* com 2.100 respostas de membros da comunidade. Um dos resultados é que quase a metade dos entrevistados disseram que sua linguagem de desenvolvimento preferida é, não é surpreendentemente, C++ ou Java. Este resultado corrobora com o que foi apresentado neste trabalho.

Outro estudo que analisa a comunidade no TopCoder foi realizado por (Boudreau et al 2014). Estes autores buscaram analisar os fatores que levam à um maior ou menor esforço por parte dos coders. Para tanto, eles realizaram um experimento através de um conjunto de 260 coders escolhidos aleatoriamente para um problema complexo ao longo de 10 dias. Como mencionado antes, a plataforma TopCoder fornece uma infraestrutura pronta para acolher os desafios de programação e tem um sistema de classificação e avaliação de habilidades bem desenvolvidas, que permite a identificação da capacidade de qualquer competidor. Assim, os autores foram capazes de tomar medidas detalhadas e objetivas de esforço, comunicação e performance dos *coders*. Os resultados sugerem que os níveis de esforço são movidos por incentivos financeiros e da interação de outros companheiros de equipe.

De maneira similar a este trabalho, Mattos e colegas (2013) apresentam uma análise de como as empresas brasileiras estão usando as plataformas de colaboração para a inovação aberta. O objetivo do trabalho era identificar as práticas de empresas

quanto a condução de iniciativas de *crowdsourcing*. Os resultados preliminares da pesquisa indicam que os projetos de *crowdsourcing* têm sido bem sucedidos.

Um outro estudo empírico baseado no Topcoder é apresentado por Li e colegas (2013). Com base na literatura existente estes autores propõe 23 fatores de qualidade de software em 2 aspectos: plataforma e projeto. Para tal, eles utilizam um programa rastreador (crawler) para coletar informações de 836 projetos no período de 2003 a 2012. Além disso, eles utilizam um modelo de regressão múltipla para analisar o impacto dos fatores na qualidade de software e sugerem 6 fatores chaves para melhoria da qualidade: pontos de qualidade na plataforma, quantidade de projetos ativos, quantidade dos membros registrados no projeto, pontuação máxima dos desenvolvedores, pontuação do projeto e tamanho da documentação disponível. Além disso, Li e colegas também sugerem 4 aspectos para ajudar na melhoria da qualidade do *software* desenvolvido via *crowdsourcing*: escolher o melhor período para postar o projeto, reduzir a escala do projeto, atrair desenvolvedores mais habilidosos e, finalmente, melhorar a pontuação do projeto. Conforme apresentado, pesquisas em Crowdsourcing vêm crescendo ao longo do tempo, e este trabalho teve como motivação contribuir com a área. Além disso, os dados coletados nesta pesquisa estão sendo armazenados para que se possa em um trabalho futuro analisar o histórico de participação de coders brasileiros, bem como de outros países presentes na plataforma.

6. Conclusão

Essa pesquisa buscou contribuir para o conhecimento sobre o *crowdsourcing* no contexto de desenvolvimento de *software* através de um estudo da plataforma TopCoder. Além disto, este artigo visa estimular estudos adicionais sobre a participação brasileira nesta nova forma de trabalho, além de permitir um maior entendimento sobre o potencial das atividades de desenvolvimento de *software* utilizando *crowdsourcing* no Brasil. Foram coletados dados sobre classificação de coders e países a partir das páginas HTML da própria plataforma, utilizando um *parser* de documentos HTML. Após análises, concluímos que o Brasil está relativamente bem colocado no geral, bem como entre todas as comparações realizadas nesta pesquisa, ou seja, o Brasil em relação à América Latina, ao BRICS e Leste Europeu. Entretanto, vale ressaltar que o número de brasileiros ativos no TopCoder ainda é pequeno, com apenas 147 desenvolvedores. A análise contínua destes dados por um longo espaço de tempo poderá revelar a evolução quanto ao aumento da participação por países, trabalho este que está em andamento.

Vale ressaltar uma ultima relevância deste trabalho para comunidade de sistemas colaborativos: os resultados apresentados buscaram estimular uma contribuição maior em pesquisas empíricas no contexto do crowdsourcing ao invés de focar na construção de plataformas e aplicações. Em outras palavras, entender a participação da multidão (*crowd*), ou pelo menos parte dela, em uma plataforma de crowdsourcing pode servir de ponto de partida para a investigação de fatores que visem o aumento da participação no Topcoder (Li e colegas, 2013). Além disso, plataformas de crowdsourcing também estão relacionadas a comunidades online, logo o entendimento da adoção e retenção destas comunidades é um importante aspecto de pesquisa em Sistemas Colaborativos.

Agradecimentos

Esta pesquisa é financiada pelo CNPq através dos processos 406692/2013-0, 440880/2013-0 e 310468/2014-0.

7. Referências

- Alves, L., et al. (2011) “Um aplicativo baseado em inteligência coletiva para compartilhamento de rotas em redes sociais”. VIII Simpósio Brasileiro de Sistemas Colaborativos.
- Antonio M., et al. (2014) “Filtragem colaborativa de rotas de ônibus usando dados sobre a acessibilidade das vias urbanas”. Anais do XI Simpósio Brasileiro de Sistemas Colaborativos. Curitiba, PR, Brasil.
- Birkinshaw, J. and Goddard, J. (2009) “Qual é seu modelo de gestão?”. HSM Management 75.
- Boudreau, K., et al. (2014) “From Crowds to Collaborators: Initiating Effort & Catalyzing Interactions Among Online Creative Workers”. Social Science Research Network. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384068.
- Brabham, D. (2008) “Crowdsourcing as a model for problem solving”. The International Journal of Research into New Media Technologies. Vol. 14.
- Chan J., et al. (2016) “Improving Crowd Innovation with Expert Facilitation”. Proceedings of the 19th CSCW Conference. San Francisco, CA, USA.
- Douglas, J. (2014) “Getting to Know You!”. <http://www.topcoder.com/blog/results-of-getting-to-know-you-survey/>
- Hara K., et al. (2013) “Combining Crowdsourcing and Google Street View to Identify Street Level Accessibility Problems”. CHI'13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Heaton, J. (2002) “Programming Spiders, bots and Aggregations in Java”. San Francisco, USA.
- Huberman, B., et al. (2009) “Crowdsourcing, attention and productivity”. Journal of Information Science. pp. 758-765.
- Kittur, A., et al. (2013) “The Future of Crowd Work”. In Proceedings of the 2013 conference on Computer supported cooperative work (CSCW 2013). USA, San Antonio, 1301-1318.
- Leicht, N., et al. (2015) “Crowdsourcing in Software Development: A State-of-the-Art Analysis”. 28th Bled eConference. Bled, Slovenia.
- Li, K., et al. (2013) “Analysis of the key factors for software quality in crowd sourcing development: An empirical study on TopCoder.com”. Proceedings of the 2013 IEEE 37th Annual Computer Software and Applications Conference, pp. 812-817.
- Lohr, S. (2009) “Netflix Awards \$1 Million Prize and Starts a New Contest”. <http://bits.blogs.nytimes.com/2009/09/21/netflix-awards-1-million-prize-and-starts-a-new-contest/>.

- Lunden, I. (2013) “Appirio Buys TopCoder To Add More Crowdsourcing, And 500K Developers And Designers, To Its CloudSpokes Network”. <http://techcrunch.com/2013/09/17/appirio-buys-topcoder-to-add-more-crowdsourcing-and-500k-developers-and-designers-to-its-cloudspokes-network/>
- Lykourantzou, I., et al. (2016) “Team Dating: A Self-Organized Team Formation Strategy for Collaborative Crowdsourcing”. Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing System. San Jose, CA, USA.
- Malone, T., et al. (2010) “The collective intelligence genome”. MIT Sloan Management Review. Vol. 41.
- Malone, T., et al. (2011) “The big idea. The age of hyperspecialization”. Harvard Business Review.
- Mattos, C., et al. (2013) “The Use Of Collaborative Platforms For Improved Open Innovation: Brazilian Context”. International Conference on Industrial Engineering and Operations Management. http://www.abepro.org.br/biblioteca/ICIEOM2013_STO_170_979_21786.pdf
- Schneider, D., et al. (2011) “Multidões: a nova onda do CSCW?”. VIII SBSC.
- Schweissguth, S. (2012) “Learning From the Crowdsourcing Efforts at Netflix”. <http://crowdsourc.com>.
- Stol, K. and Fitzgerald, B. (2014) “Two's company, three's a crowd: a case study of crowdsourcing software development”. ICSE.
- Zatti, R. (2013) “Crowdsourcing e as peças de uma nova forma de fazer as coisas”. <http://www.slideshare.net/rafazatti/e-book-crowdsourcing-e-as-peas-de-uma-nova-forma-de-fazer-as-coisas>.

Os Programas de Engajamento em Software Livre Atraem Estudantes à Colaboração Voluntária? Um Estudo Empírico do Google Summer of Code

Jefferson O. Silva^{1,3}, Igor Steinmacher², Marco A. Gerosa¹

¹ Departamento de Computação – Universidade de São Paulo (USP)
Caixa Postal 66.281 – São Paulo-SP – Brazil

² Departamento de Computação – Universidade Tecnológica do Paraná (UTFPR)
Caixa Postal 271 – Campo Mourão-PR – Brazil.

³ Departamento de Computação –
Pontifícia Universidade Católica de São Paulo (PUC-SP)
Caixa Postal 1.670 – São Paulo – Brazil.

silvajo@{ime.usp,pucsp}.br, igorfs@utfpr.edu.br, gerosa@ime.usp.br

Abstract. *Summers of Code, intensive short-term software development programs, may not only foster development but may also inspire its students to engage in OSS. In this study, we conduct a survey with GSoC students' and mentors' to find out either if the students keep contributing after the end of the program, as to understand what makes them enter GSoC. We focus on GSoC because it provides their students with a rare mix of participation rewards including skill development, career leverage, peer recognition, and stipends. We use descriptive statistical for analyzing the students' and mentors' answers. Our findings suggest that the students' main motivation to join GSoC is related to career leverage, using the skills developed in the program to signal their potential to employers. However, we also find that the students' intention to keep contributing to OSS is strong.*

Resumo. *Os programas Summer of Code, programas intensivos de desenvolvimento de software de curta duração, podem não apenas promover o desenvolvimento, mas também inspirar o engajamento de estudantes em projetos de software livre. Neste artigo, realizamos um survey com estudantes e mentores do Google Summer of Code (GSoC) para descobrir se os estudantes continuam a colaborar após o término do programa, bem como entender o que os atrai a participarem do GSoC. Escolhemos o GSoC porque ele provê a seus estudantes uma rara combinação de recompensas de participação incluindo desenvolvimento de habilidades técnicas, alavancagem da carreira, reconhecimento pelos pares e pagamento. Usamos estatística descritiva para analisar as respostas de estudantes e mentores. Nossos resultados sugerem que a principal motivação dos estudantes participarem do GSoC tem relação com alavancagem da carreira, utilizando as habilidades técnicas adquiridas para sinalizar seus talentos a empregadores. Entretanto, também constatamos que a intenção dos estudantes de continuar colaborando com projetos de software livre é forte.*

1. Introdução

Construir e manter comunidades de colaboração aberta sustentáveis é desafiador. Os membros precisam entender como iniciar a comunidade, encorajar o comprometimento, encorajar a contribuição voluntária, além de regular o comportamento de seus participantes [Kraut and Resnick 2012]. O uso de bons princípios geralmente ajuda as comunidades de colaboração abertas a se tornarem bem sucedidas, porém podem não ser suficientes, uma vez que até mesmo comunidades bem estabelecidas, como a Wikipédia, necessitam criar estratégias para atrair um fluxo constante de novatos [Ciampaglia and Taraborelli 2015; Kraut and Resnick 2012].

Estudos anteriores sugerem que programas de engajamento em comunidades de Software Livre (SL) – programas intensivos de desenvolvimento de software de curta duração – não apenas fomentam o desenvolvimento de software, como podem promover a construção das comunidades de SL [Trainer and Chaihirunkarn and Kalyanasundaram and et al. 2014]. Exemplos incluem o Google Summer of Code (GSoC), iRails Girls Summer of Code, Open Summer of Code, Julia Summer of Code¹, Ruby Summer of Code², and Porto Summer of Code³.

Os programas Summer of Code são um tipo de programa de engajamento em SL que pode proporcionar aos estudantes muitas recompensas de participação [Trainer and Chaihirunkarn and Kalyanasundaram and et al. 2014], incluindo parceria com a comunidade, desenvolvimento de habilidades, satisfação pessoal, alavancagem da carreira, reconhecimento pelos pares, pagamento e construção da identidade social. Muitas dessas recompensas são listadas pela literatura de SL como motivação para a colaboração voluntária com projetos de SL [Beecham et al. 2008; Forte and Bruckman 2008; Steinmacher et al. 2014]. Pesquisadores começaram, então, a investigar o potencial dos programas Summer of Code em motivar seus estudantes a contribuírem voluntariamente com SL [Silva et al. 2015; Trainer and Chaihirunkarn and Herbsleb 2014].

No desenvolvimento de SL, as diferentes motivações para participar em uma atividade são tradicionalmente classificadas como intrínsecas ou extrínsecas [Deci et al. 1999]. Pessoas intrinsecamente motivadas extraem sua satisfação da própria realização de uma atividade, enquanto que pessoas extrinsecamente motivadas extraem sua satisfação de contingências externas, como por exemplo salário. Como motivação extrínseca se foca em contingências externas, elas podem ser utilizadas para reforço comportamental [Fang and Neufeld 2009].

Recompensas de participação extrínsecas utilizadas para engajamento podem significativamente comprometer a motivação intrínseca de um indivíduo em condições não triviais [Deci et al. 1999]. A motivação varia em níveis individuais e se mistura com o conhecimento e as habilidades do indivíduo para produzir comportamento. Enquanto a motivação é um estado psicológico, os resultados desse estado é comportamento [Fang and Neufeld 2009].

¹ <http://julialang.org/blog/2015/05/jsoc-cfp/>

² <http://www.rubysummerofcode.org/>

³ <http://blog.portosummerofcode.com/>

Muitos pesquisadores pesquisam os efeitos das recompensas no comportamento dos desenvolvedores. Embora seja verdade que em muitos casos as recompensas ajudam a alinhar os interesses de participantes e organizadores, em alguns outros elas podem induzir um comportamento não desejado. O GSoC é especialmente interessante deste ponto de vista porque ele provê aos estudantes uma ampla mistura de recompensas de participação.

O GSoC levanta questões como: o programa é efetivo no engajamento de desenvolvedores em SL? Pagamento é uma boa estratégia para motivar estudantes para colaboração voluntária em SL? O estudante participaria do programa se não pudesse utilizar essa participação para alavancar a carreira?

O objetivo dessa pesquisa é explorar o papel do GSoC na motivação e engajamento de novos colaboradores em comunidades de SL. Neste contexto, foram definidas duas questões de pesquisa:

QP1. O que motiva os estudantes a entrarem no GSoC?

QP2. O GSoC inspira novatos à colaboração voluntária em projetos de SL?

Para responder nossas questões de pesquisa, conduzimos um survey com estudantes e mentores das edições do GSoC de 2005 a 2015, utilizando técnicas da estatística descritiva para analisar as respostas de estudantes e mentores. Esta pesquisa contribui oferecendo evidência empírica da efetividade dos programas Summer of Code em engajar seus estudantes como novos voluntários em projetos de software, bem como o que leva os estudantes a participarem desses programas.

Este artigo está organizado da seguinte forma: trabalhos relacionados, método de pesquisa, resultados, limitações e ameaças à validade do trabalho e a conclusão.

2. Trabalhos Relacionados

Nesta seção apresentamos mais informações sobre o Google Summer of Code e alguns trabalhos relacionados à motivação e programas de engajamento em SL.

2.1. Google Summer of Code (GSoC)

O GSoC é um programa da Google que remunera aos estudantes para que eles desenvolvam software para diversos projetos SL durante um período de cerca de 3 meses. Atualmente, o valor da remuneração é de US\$ 5.500,00 para o estudante que completar com sucesso todas as fases do programa.

O programa possui 5 metas⁴, sendo que a meta (ii) formou nossa questão de pesquisa (QP2):

- (i). Desenvolver SL para o benefício de todos
- (ii). **Inspirar desenvolvedores iniciantes a começarem a participar no desenvolvimento de SL**
- (iii). Ajudar projetos de SL a identificarem e buscarem novos desenvolvedores e *committers*

⁴http://www.google-melange.com/gsoc/document/show/gsoc_program/google/gsoc2015/help_page#3. What are the goals of this program

- (iv). Prover ao estudante a oportunidade de realizar trabalhos relacionados a seus objetivos acadêmicos (*flip bits, not burgers*)
- (v). Expor o estudante a cenários de desenvolvimento do mundo real

2.2. Motivação de desenvolvedores de SL

Muitos pesquisadores têm investigado o que motiva os desenvolvedores a contribuir voluntariamente com projetos de SL. Beecham et al. [2008] apresentaram uma revisão sistemática sobre motivação na engenharia de software, sugerindo que os engenheiros de software de seu estudo foram provavelmente motivados por uma combinação de três fatores relacionados: suas características, controles internos e moderadores externos. Shapiro e David [2008] encontraram suporte para desenvolvedores motivados por ideologia por meio de surveys. Yu, Jiang, e Chan [2007] identificaram a "obrigação moral" e o "avanço da comunidade virtual" como motivações individuais em uma revisão da literatura. Osterloh e Rota [2004] sugeriram que o comportamento altruísta causado por motivação "pro social" fomentou a colaboração com SL. Benkler [2002] e Osterloh e Rota [2004] sugeriram que a satisfação pessoal e diversão desempenharam um papel importante na motivação para a colaboração voluntária com SL. Tirole e Lerner [2002] propuseram o reconhecimento pelos pares como uma motivação fundamental. Lattemann e Stieglitz [2005] propuseram que os papéis que os colaboradores desempenharam em seu trabalho estão relacionados à motivação, sendo que programadores (ex: ao contrário de gerentes) foram motivados pelo reconhecimento dos pares. Diversos surveys evidenciaram que a aprendizagem em si motivou os indivíduos à colaboração com SL (ex: [Ghosh 2003; Misirli et al. 2014; Oreg and Nov 2008]). Tirole e Lerner [2002] derivaram da teoria econômica que os desenvolvedores de SL foram motivados pela alavancagem da carreira, participando de projetos de SL como uma forma de sinalizar seus talentos a potenciais empregadores. Lakhani e Wolf [2003] examinaram o grau de participação e sua ligação com motivação financeira, concluindo que o nível de contribuição e esforço foi maior para desenvolvedores pagos.

2.3. Programas de engajamento em SL

Os programas de Engajamento em SL – tais como os programas Summer of Code e Hackathons – têm sido pouco estudados até o momento. Trainer, Chaihirunkarn, e Herbsleb [2014] conduziram um estudo de caso de uma biblioteca de bioinformática chamada BioPython para investigar os efeitos do GSoC, concluindo que o GSoC pode proporcionar extensão das funcionalidades, produzir novos conhecimentos e treinar gerações futuras de colaboradores com SL. Trainer et al. [2014] investigaram os efeitos do GSoC e hackathons, concluindo que a maioria dos projetos GSoC resultou em funcionalidades estáveis, e que os hackathons resultaram em protótipos ao invés de ferramentas finalizadas. Silva et al. [2015] estudaram quantitativamente 33 estudantes do GSoC 2014 que trabalharam em projetos da Fundação Apache e não encontraram nenhum desenvolvedor que tenha começado a colaborar com SL após a participação no programa. As investigações do GSoC de Trainer, Chaihirunkarn, e Herbsleb [2014] e Trainer et al. [2014] foram focadas apenas software científicos, e o estudo sobre a efetividade de atração do GSoC 2014 de Silva et al. [2015] focou apenas em desenvolvedores que trabalharam em projetos da Fundação Apache.

3. Método de Pesquisa

Para entender como o GSoC afeta a motivação e engajamento dos estudantes na colaboração voluntária em projetos de SL, conduzimos um survey para aprofundar nosso entendimento sobre: o engajamento dos estudantes em projetos de SL antes e depois de participarem no GSoC; como os estudantes classificam as recompensas de participação normalmente associadas a desenvolvedores de SL; a percepção dos mentores sobre o engajamento de estudantes em projetos de SL antes e depois de participarem no GSoC; e a percepção dos mentores sobre a importância das recompensas de participação dos estudantes. A resposta a estes temas devem nos fornecer evidência empírica sobre o que atrai os estudantes a entrarem no GSoC, bem como prover uma medida do grau de contribuição com projetos de SL após a participação no programa. Escolhemos um survey porque nos permite contatar e coletar informações de um grupo maior de estudantes e mentores. Em adição, o survey fornece padronização aos resultados, o que facilita comparações e o uso de estatística descritiva.

Descrevemos o método em dois passos: coleta dos contatos de estudantes e mentores e seleção dos respondentes (seção 3.1); projeto e administração do questionário (seção 3.2).

3.1. A coleta dos contatos de estudantes e mentores e seleção dos respondentes

Em todas as edições, os organizadores do GSoC publicam no site do programa uma lista com as propostas selecionadas, contendo os nomes dos estudantes aceitos, os nomes dos projetos e os mentores designados. Usando essas informações, procuramos manualmente os e-mails de estudantes e mentores utilizando o mecanismo de busca da Google, uma vez que os organizadores do GSoC não publicam informações de contato. Consideramos que havíamos encontrado o e-mail correto do estudante, quando tínhamos evidências claras que ligavam o nome do estudante ao projeto, como por exemplo quando encontrávamos um blog do estudante contendo o mesmo e-mail que estávamos procurando e publicações a respeito do projeto. Estabelecemos uma amostra de conveniência de 10% do total dos estudantes do GSoC para cada ano, uma vez que a coleta incluía a custosa tarefa de verificação da legitimidade de cada um dos e-mails. Para controlar vieses potenciais, numeramos unicamente todos os participantes das edições de 2005 a 2015 e selecionamos aleatoriamente os estudantes participantes.

3.2. A administração dos questionários e análise dos dados

Por meio de questionário, perguntamos aos estudantes sobre suas contribuições aos projetos de SL antes e depois do GSoC, e algumas questões gerais sobre sua participação no GSoC. Adicionalmente, para avaliar as recompensas de participação, perguntamos aos estudantes se ainda entrariam no GSoC se não houvesse pagamento. As opções das respostas da questão da escala Likert sobre a importância de outras recompensas de participação foram baseadas nos trabalhos de Beecham et al. [2008] e de Misirli et al. [2014].

A Tabela 1 apresenta as questões mais importantes do survey dos estudantes, e cada linha horizontal representa uma quebra de página. O questionário completo está disponível em <https://goo.gl/aQ8QIU>. É importante destacar que o questionário foi aplicado em inglês e a Tabela 1 apresenta a tradução das questões.

Tabela 1. Principais questões do questionário aplicado aos estudantes

<i>Antes do GSoC...</i>	
1.	... você contribuía com projetos de SL além dos seus projetos? (<i>escalas Likert, veja Tabela 2 para os itens</i>)
2.	... você contribuía com o projeto que você escolheu para o programa? (<i>escalas Likert, veja Tabela 3 para os itens</i>)
3.	... você pretendia continuar contribuindo ao projeto (excluindo em outros GSoC)? (<i>escalas Likert, veja Tabela 4 para os itens</i>)
4.	Após o GSoC, você realmente continuou a contribuir (excluindo em outros GSoC)? (<i>escalas Likert, veja Tabela 5 para os itens</i>)
<i>Sobre a época em que você decidiu entrar no GSoC, o quanto você concorda com as seguintes afirmações. Eu teria desistido de entrar no GSoC se ...</i>	
5.	Você participaria do GSoC se não houvesse remuneração financeira da Google? (<i>sim/não</i>)
6.	... não pudesse anunciar a experiência em meu currículo (<i>escalas Likert, veja Figura 1 para os itens</i>)
7.	... não houvesse trabalho tecnicamente desafiador envolvido (<i>escalas Likert, veja Figura 1 para os itens</i>)
8.	... não ganhasse respeito de meus pares em meu trabalho, universidade, comunidade e etc. (<i>escalas Likert, veja Figura 1 para os itens</i>)
9.	... não pudesse continuar a contribuir com projetos de SL (<i>escalas Likert, veja Figura 1 para os itens</i>)
10.	... não houvesse remuneração financeira da Google. (<i>escalas Likert, veja Figura 1 para os itens</i>)
11.	Sexo (<i>masculino/feminino/outro</i>)
12.	Idade
13.	Qual era sua escolaridade?

Conduzimos um piloto do questionário utilizando estudantes brasileiros participantes do GSoC 2015. Depois de alguns poucos ajustes, disparamos e-mails aos estudantes e seus mentores, convidando-os a responder os questionários. Na tentativa de maximizar as respostas, mandamos ainda dois e-mails de *follow up*, lembrando-os de nosso convite. Por fim, utilizamos estatística descritiva para analisar as respostas de estudantes e mentores.

4. Resultados

Nesta seção, apresentamos os resultados do survey, dividindo-os em duas partes. Primeiramente, apresentamos os resultados da análise dos dados dos estudantes e, em seguida, apresentamos os dados relacionados aos mentores. Para cada item das questões, indicamos quantos participantes mencionaram o tema em sobrescrito. Note que essa contagem somente indica o número de respostas, e não necessariamente a importância do tema.

4.1. Survey dos estudantes

Enviamos 1.000 e-mails (10% do total de estudantes) convidando estudantes e recebemos 141 respostas, e nossa amostra contém participantes das edições do GSoC 2005 a 2015.

4.1.1. Perfil dos estudantes

Aproximadamente, um terço dos estudantes está na faixa de 18-20⁽⁴⁵⁾ anos de idade, quase a metade, encontra-se na faixa de 21-25⁽⁶³⁾, um quinto, encontra-se na faixa 26-30⁽²⁶⁾, e apenas alguns poucos estão na faixa 31-40⁽⁷⁾. Em sua primeira participação, os estudantes são majoritariamente formados por graduandos⁽⁵⁸⁾, seguido de graduados⁽⁴¹⁾, e alguns formados no ensino médio⁽²⁴⁾. A minoria entra sem ter o ensino médio completo⁽³⁾, durante a pós-graduação⁽⁶⁾, ou pós-graduado⁽⁷⁾.

Os estudantes tipicamente são desenvolvedores plenos⁽⁶²⁾, com 2 a 5 anos de experiência, ou sêniores⁽⁴¹⁾ com 5 a 10 anos. Identificamos também alguns iniciantes⁽¹⁸⁾ com menos de 2 anos de experiência, e especialistas⁽¹⁹⁾ com mais de 10 anos.

4.1.2. A contribuição de estudantes em projetos de SL antes e depois da primeira participação no GSoC

De forma geral, percebe-se que os estudantes não eram colaboradores regulares de projetos de SL em geral (veja Tabela 2), nem dos projetos para os quais entraram no programa (veja Tabela 3). Isso significa que o GSoC, de fato, *está atraindo potenciais colaboradores de SL*. Podemos ver que o grupo de colaboradores frequentes com projetos de SL (22 estudantes na Tabela 3) é na grande maioria, colaboradores dos projetos para o qual contribuíram durante o GSoC (20 estudantes na Tabela 3 – colaboradores frequentes + *core members*), resultando em 2 colaboradores de outros projetos de SL que não os do GSoC.

Tabela 2. Contribuições do estudante ao SL antes da participação no GSoC

frequência de contribuição	Contagem (%)
nunca	49 (34,8)
raramente	46 (32,6)
ocasionalmente	24 (17,0)
frequentemente	22 (15,6)

Uma possível explicação para este comportamento é que os estudantes começam a colaborar com o projeto como uma maneira de serem aceitos no programa. Existem alguns blogs⁵ na web escritos por estudantes e mentores com dicas sobre como ser aceitos no programa que instruem que os candidatos se envolvam com a comunidade para serem conhecidos.

Tabela 3. Contribuições do estudante para o projeto GSoC antes da primeira participação no programa

frequência de contribuição	contagem (%)
nunca	92 (65,2)
raramente	19 (13,5)
ocasionalmente	10 (7,1)
frequentemente	14 (9,9)
<i>core member</i>	6 (4,3)

De acordo com as respostas ao questionário, a maioria dos estudantes entraram no GSoC com a intenção de continuar contribuindo para seus projetos (veja Tabela 5). Em contrapartida, cerca de um terço dos estudantes entram no programa sem uma visão clara de se vão continuar contribuindo ou não. Quando somamos os alunos indecisos com aqueles que entram no programa decididos a não continuar colaborando futuramente, vemos que o GSoC deve trabalhar a percepção da importância da contribuição de uma quantidade considerável de estudantes (~43%) para que atinja sua meta de inspirar contribuições futuras.

Tabela 4. Intenção em continuar contribuindo com o projeto escolhido antes da primeira participação no GSoC

intenção de contribuição	contagem (%)
definitivamente não	8 (5,7)
não	11 (7,8)
talvez	42 (29,8)
sim	40 (28,4)
definitivamente sim	40 (28,4)

⁵ ex: <http://danielpocock.com/getting-selected-for-google-summer-of-code-2016>

As respostas no questionário sugerem que contribuições tendem a decair no decorrer do tempo após o término do programa, pelo menos para períodos superiores há um ano (veja Tabela 5). Como aplicamos o questionário aos alunos do GSoC 2015 imediatamente após o término do programa, mostramos os dados separadamente porque consideramos que estes estudantes possuem uma perspectiva de contribuição ao longo do tempo diferente dos alunos das edições anteriores. De fato, podemos ver que poucos participantes do GSoC 2015 deixam de contribuir logo após o término do programa, mas vimos que a proporção de não colaboradores e colaboradores-raros aumenta consideravelmente quando consideramos tempo de contribuição superior a um ano. Podemos notar que a proporção dos alunos que tinha a intenção de continuar colaborando com o projeto GSoC (veja Tabela 4) é maior do que aqueles que efetivamente colaboraram (veja Tabela 5). Outro ponto interessante é que alguns dos colaboradores-frequentes se tornam *core-members*.

Tabela 5. Auto avaliação dos estudantes do quanto continuaram a contribuir com o projeto GSoC após o término do programa

contribuição após GSoC?	contagem (%) estudantes 2015	contagem (%) estudantes 2005-2014
não	4 (2,8)	20 (14,2)
raramente	11 (7,8)	19 (13,5)
ocasionalmente	28 (19,9)	18 (12,8)
frequentemente	12 (8,5)	6 (4,3)
<i>core member</i>	7 (5,0)	16 (11,3)

4.1.3. Avaliando as recompensas de participação no GSoC para os estudantes

A maioria dos estudantes entram no GSoC motivados por recompensas de participação extrínsecas, embora exista um número menor, mas significativo de estudantes que busque primariamente as intrínsecas. Da literatura (ex: [Beecham et al. 2008]), nós identificamos que os estudantes entrariam no programa por 5 recompensas de participação: alavancagem da carreira, reconhecimento pelos pares, contribuição com SL, ganho financeiro, incentivos do tipo de trabalho como desafio técnico. Para classificar a importância de cada recompensa de participação na percepção dos estudantes, perguntamos se eles entrariam em um GSoC hipotético que oferecesse todas as 5 recompensas de participação exceto uma. A Figura 1 mostra a contagem das respostas dos estudantes em um formato de respostas Likert (escala de 5 pontos).

Cerca de 44% dos estudantes⁽⁶²⁾ esperavam poder usar sua participação no GSoC para alavancar a carreira (veja Q1 em Figura 1), preferindo desistir de participar do programa caso contrário. Cerca de 35%, não entendem essa recompensa de participação como sendo essencial. No entanto, o uso da participação como diferencial em um currículo levanta dúvidas em cerca de ~20% dos estudantes.

Continuar contribuindo com SL (veja Q2 em Figura 1) é um critério essencial para participação para quase metade (~48%) dos estudantes⁽⁶⁸⁾, o que é um pouco menor do que eles declararam na intenção de continuar contribuindo (~57%) na Tabela 5. Este resultado sugere que a continuidade da colaboração com os projetos GSoC é mais do que um bônus de participação, mas uma experiência da qual os estudantes não desejam abrir mão.

Reconhecimento pelos pares é a recompensa de participação menos essencial na percepção dos estudantes, pelo menos quando comparadas com as outras (veja Q3 em Figura 1). Um ponto interessante é que esta questão teve um grande número de

indecisos⁽³⁹⁾, o que pode indicar que muitos estudantes não tinham refletido a respeito deste tópico anteriormente, ou ainda que a elaboração da questão não tenha sido clara o suficiente.

Remuneração financeira é um critério essencial para participação para cerca de um terço dos estudantes (33%), enquanto que para outros (~37%) não a consideraram essencial (veja Q4 em Figura 1). Contudo, muitos estudantes optaram pela neutralidade neste tópico (~28%). Por outro lado, quando perguntamos se eles participariam se não houvesse pagamento (questão 5 na Tabela 1), ~38% dos estudantes o consideraram como uma recompensa de participação essencial, enquanto ~62% não o consideraram. Uma possível explicação para essa discrepância é que uma das questões possui um formato de resposta Likert enquanto a outra, um formato binário (sim/não). Assim, quando os estudantes puderam expressar suas dúvidas, os dados ficaram discrepantes. Da mesma maneira, podemos ver que os indecisos possuem uma tendência a entender o pagamento como recompensa de participação não essencial, aplicando o raciocínio anterior invertido: quando tiramos a possibilidade de dúvidas, eles o classificam como não essencial.

O desafio técnico é a recompensa em que o maior número de estudantes classificou como essencial para participação (veja Q5 em Figura 1). Da mesma forma, é a recompensa em que os estudantes menos se mostraram indecisos (~14%), e também menos discordaram (~18%).

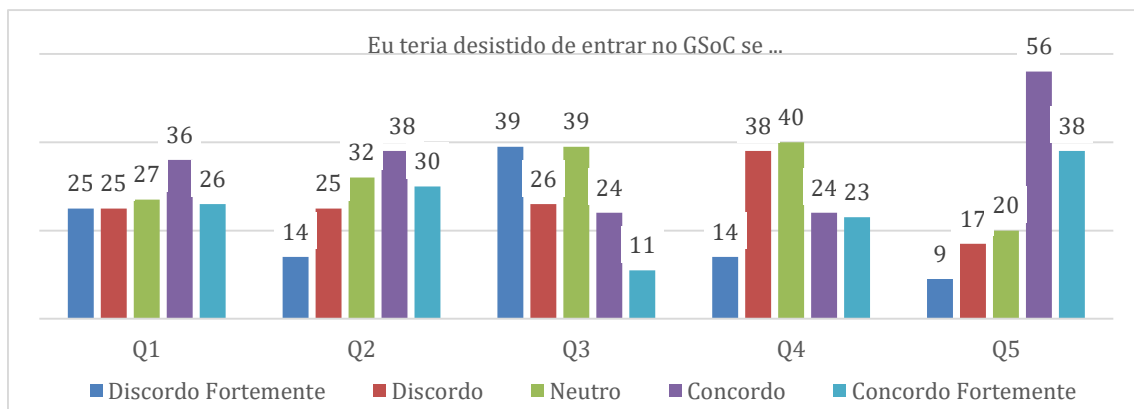


Figura 1. Percepção dos estudantes da importância das recompensas de participação. Q1: ... não pudesse anunciar essa experiência em meu currículo; Q2: ... não pudesse continuar colaborando com SL; Q3: ... não ganhasse respeito de meus pares em meu trabalho, universidade etc.; Q4: ... não houvesse remuneração financeira da Google; Q5: ... não houvesse trabalhos tecnicamente desafiador envolvido

4.2. Survey dos Mentores

Enviamos 730 e-mails convidando mentores e recebemos 53 respostas, abrangendo mentores das edições do GSoC 2005 a 2015. O número de convites enviado aos mentores é menor que o de alunos porque alguns mentores voltam em edições posteriores para orientar outros alunos. O questionário dos mentores pode ser acessado em: <http://goo.gl/forms/paCs3U5ltz>.

4.2.1. A percepção dos mentores sobre as contribuições de seus estudantes com o SL antes e depois do GSoC

Quase metade (~44%) dos mentores entendem que seus estudantes nunca ou raramente contribuem com projetos de SL antes da primeira participação no GSoC (veja Tabela 6). Da mesma forma, os mentores relataram que os estudantes, em sua maioria, nunca ou raramente haviam colaborado com os projetos GSoC antes de entrarem no programa, muito embora exista uma parcela significativa de estudantes com contribuições ocasionais (veja Tabela 7).

Estes dados sugerem o GSoC está atraindo desenvolvedores inexperientes em projetos de SL, o que os torna potenciais futuros colaboradores com projetos de SL futuramente. Este resultado está alinhado com as respostas dos estudantes (veja Tabela 2 e Tabela 3).

Tabela 6. Percepção dos mentores sobre a colaboração dos estudantes com projetos de SL antes de entrarem no GSoC

colaboração com projetos de SL antes do GSoC	contagem (%)
não sei	4 (6,4)
nunca	8 (12,7)
raramente	20 (31,7)
ocasionalmente	19 (30,2)
frequentemente	2 (3,2)

Tabela 7. Percepção dos mentores na colaboração dos estudantes com seus projetos antes de entrarem no GSoC

colaboração com o projeto do mentor antes do GSoC	contagem (%)
não sei	0 (0,0)
nunca	23 (43,4)
raramente	14 (26,4)
ocasionalmente	14 (26,4)
frequentemente	2 (3,8)

Em geral, os mentores percebem um aumento no nível de engajamento dos estudantes em seus projetos após a participação no GSoC (veja Tabela 8), pelo menos no curto prazo, uma vez que a questão não delimitou qualquer período de contribuição. Este raciocínio parece ser razoável, uma vez que já observamos que as contribuições dos estudantes diminuí ao longo do tempo (veja Tabela 5).

Tabela 8. A contribuição dos estudantes aos projetos dos mentores após a participação no GSoC

colaboração com o projeto do mentor após o GSoC	contagem (%)
não sei	1 (1,9)
nunca	8 (15,1)
raramente	16 (30,2)
ocasionalmente	18 (34,0)
frequentemente	10 (18,9)

4.2.2. A percepção dos mentores sobre a importância das recompensas de participação para seus estudantes

De acordo com as respostas da maioria dos mentores⁽⁴¹⁾, a participação dos estudantes no GSoC é fortemente relacionada à alavancagem da carreira dos estudantes (veja M1 na Figura 2). Embora alguns mentores⁽¹⁰⁾ tenham optado em ser neutros sobre a questão, podemos observar que o uso da participação no programa como alavancagem da

carreira é o único ponto em que nenhum mentor discordou. Da mesma forma, foi um dos pontos que menos levantou dúvidas.

A maioria dos mentores⁽³⁴⁾ entende que a colaboração com projetos de SL é um componente forte na decisão dos estudantes a entrarem no GSoC (veja M2 na Figura 2). Este resultado pode ser uma evidência de que o GSoC pode estar servindo como um agente redutor das barreiras para o início da colaboração com projetos de SL, complementando os resultados de Steinmacher, Gerosa, e Redmiles [2015]. Esse resultado está de acordo com as respostas sobre a intenção de continuarem contribuindo após o programa (veja Tabela 4).

Reconhecimento pelos pares é relatada como mais importante pelos os mentores que pelos os estudantes (veja M3 na Figura 2), pelo menos numa comparação simples das escalas Likert da Figura 1 e da Figura 2. Trabalhos anteriores (ex: [Beecham et al. 2008]) afirmam que o reconhecimento pelos pares é um forte motivador para desenvolvedores de projetos de SL, mas não necessariamente para desenvolvedores corporativos, quando comparado a outros motivadores como por exemplo o ganho financeiro. Este resultado faz sentido quando consideramos que o perfil dos mentores (desenvolvedores de SL) difere do perfil de seus estudantes (desenvolvedores corporativos).

O consenso entre os mentores é de que seus alunos são primariamente motivados pelo pagamento (veja M4 na Figura 2), o que contrasta com a avaliação dos estudantes sobre quão essencial o pagamento é para a participação no programa (veja Q4 na Figura 1). Na avaliação dos estudantes, a remuneração financeira é classificada como a quarta recompensa (das cinco listadas) mais essencial para a participação.

A maioria dos mentores⁽³⁷⁾ entende que o desafio técnico de seus projetos é critério essencial para a participação de seus estudantes (veja M5 na Figura 2). Poucos mentores discordam dessa posição⁽³⁾, embora alguns tenham optado por uma posição neutra⁽¹⁰⁾. O desejo dos estudantes em se tornarem tecnicamente melhores faz sentido do ponto de vista do uso da participação como alavancagem da carreira. Neste sentido, as habilidades técnicas adquiridas no programa podem ser utilizadas para se conseguir posições melhores no mercado de trabalho.

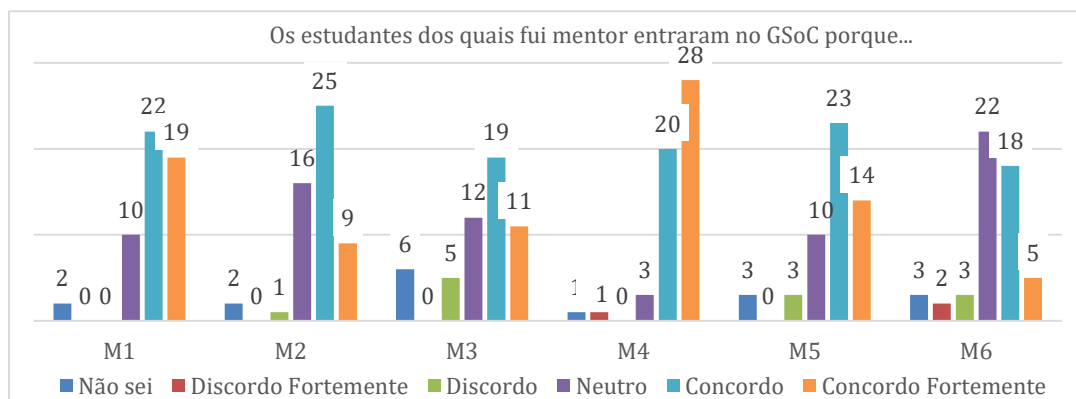


Figura 2. A percepção dos mentores sobre a motivação dos estudantes para entrarem no GSoC. M1: ... queriam anunciar essa experiência em seus currículos; M2: ... queriam colaborar com projetos de SL; M3: ... queriam ganhar respeito de seus pares no trabalho, universidade, comunidade e etc.; M4: ... queriam ser remunerados financeiramente; M5: ... queriam trabalhos tecnicamente desafiadores; M6: ... queriam contribuir com MEU projeto de SL

4.2.3. Por que os estudantes entram (QP1)? e o GSoC, de fato, inspira os novatos a colaborarem com SL (QP2)?

Genericamente, os estudantes entram no programa pelas recompensas de participação, que podem ser extrínsecas ou intrínsecas. Mais precisamente, considerando apenas os fatores motivacionais incluídos nesta pesquisa, os estudantes entram respectivamente (QP1) pelo desafio técnico dos projetos, para colaborar com projetos de SL, por preocupações com a carreira, pela remuneração financeira e pelo reconhecimento pelos pares. Contudo, na visão dos mentores, os estudantes entram, em geral, pela recompensa financeira, pela alavancagem da carreira e pelo desafio técnico respectivamente.

Embora as respostas dos estudantes nos forneçam evidências diretas – uma vez que estão respondendo sobre si mesmos –, devemos considerar que suas respostas podem apresentar vieses de confirmação [Nickerson 1998], tendendo a concordar com as afirmações que apresentam recompensas de participação mais de acordo com os objetivos do GSoC. De fato, as duas recompensas classificadas como menos essenciais à participação e as duas que mais incitaram dúvidas nos estudantes são as que não podem ser mapeadas diretamente aos objetivos do programa: remuneração financeira e reconhecimento pelos pares respectivamente. Nesse caso, a percepção dos mentores pode funcionar como uma visão mais imparcial da realidade.

Outra maneira de avaliar a classificação das recompensas de participação é observar a colaboração dos estudantes com projetos de SL após o término do GSoC. Em geral, numa comparação simples de percentuais, (QP2) o GSoC inspira novatos a colaborarem com projetos de SL, mesmo em períodos superiores há um ano. Podemos observar que a frequência de não-colaboradores e colaboradores-raros antes da primeira participação no GSoC (Tabela 2 e Tabela 3) diminui após a participação no programa (Tabela 4), enquanto as frequências dos colaboradores-frequentes e *core members*, que eram irrelevantes antes do GSoC (Tabela 2 e Tabela 3), aumentam após a participação no programa (Tabela 4), independente do período. Esse resultado sugere uma confirmação da intenção dos estudantes de continuar colaborando com projetos de SL.

Uma questão que surge naturalmente é: se os estudantes (~86%) tem alguma intenção de continuar contribuindo com SL (veja Tabela 4) e eles consideram essa contribuição contínua como essencial, por que então tão poucos estudantes continuam contribuindo (veja Tabela 5)? Afinal, continuar colaborando em projetos tecnicamente desafiadores em SL ofereceria duas das recompensas de participação mais essenciais na visão do estudante.

As respostas no questionário apontam que o perfil típico do estudante é jovem, masculino, cursando a graduação e com pouca ou nenhuma experiência em desenvolvimento de software. Uma interpretação possível do fenômeno, derivada da teoria econômica [Tirole and Lerner 2002], é que os estudantes tipicamente entram no GSoC para alavancar a carreira, utilizando as competências técnicas adquiridas no programa para sinalizarem seus talentos a futuros empregadores. Os mentores também entendem a remuneração financeira como sendo mais essencial do que os estudantes avaliaram. Estendendo a interpretação anterior, os estudantes valorizam o pagamento na mesma medida em que necessitam custear seus estudos e outras despesas básicas, tornando a remuneração financeira mais relevante no curto prazo, mas não o que os alunos necessariamente almejam no longo prazo. Dessa forma, tipicamente a

colaboração voluntária com o projeto GSoC continua até que o estudante se gradue e encontre um trabalho.

5. Limitações e Ameaças à Validade

Este trabalho de pesquisa contribui para a área não muito explorada das motivações dos estudantes para entrarem em programas de engajamento em SL, como o GSoC, e também e sobre os efeitos na atração de novos colaboradores nas comunidades de SL, decorrentes da participação nesses programas. No entanto, apontamos aqui algumas limitações de nosso trabalho.

Em primeiro lugar, os nossos dados podem não ser representativos de toda a população de estudantes, apesar de nossa preocupação em coletar uma amostra que nos permitisse um intervalo de confiança mais adequado. Como se trata de um survey, enviamos convites de participação suficientes tanto para estudantes e mentores, no entanto, nossos resultados ficam dependentes daqueles que nos respondem. Portanto, reconhecemos ser possível chegar a outras conclusões com uma seleção diferente de participantes.

Em segundo lugar, nossa abordagem utilizada na investigação dos fenômenos é apenas quantitativa que, embora facilite a análise dos dados, não nos permite uma investigação mais aprofundada dos contextos de participação, escolhas e intenções reais dos estudantes.

6. Conclusões

Mesmo comunidades de colaboração aberta bem estabelecidas se preocupam em atrair novos colaboradores para substituírem os que as deixam [Ciampaglia and Taraborelli 2015; Kraut and Resnick 2012]. Estudos anteriores sugerem que os programas Summer of Code – programas intensivos de desenvolvimento de software de curta duração – não apenas promovem o desenvolvimento de projetos de software livre (SL), mas também podem promover a construção de suas comunidades [Christopherson et al. 2010; Trainer and Chaihirunkarn and Herbsleb 2014; Trainer and Chaihirunkarn and Kalyanasundaram and et al. 2014; Trainer et al. 2016]. No entanto, ainda não sabemos ao certo se a participação de estudantes em programas de engajamento em SL podem de fato contribuir com novos colaboradores em projetos SL.

Neste trabalho, estudamos o programa de engajamento em SL Google Summer of Code (GSoC), que tem entre seus objetivos inspirar novos colaboradores para atuarem em projetos de SL. Mais precisamente, nós fornecemos evidência empírica investigando duas dimensões do GSoC: o que leva os estudantes a entrarem no programa; e se a participação no programa contribui para o aumento na colaboração com projetos de SL.

Para responder nossa questão de pesquisa, utilizamos um survey para aprofundar nosso entendimento sobre as motivações que levaram os estudantes a ingressar no GSoC. Da mesma forma, também estabelecemos uma classificação das recompensas de participação no GSoC na percepção dos estudantes e mentores. Por fim, utilizamos estatística descritiva e comparações entre as respostas de estudantes e mentores para chegar às nossas conclusões. Nossa amostra abrange estudantes e mentores que participaram das edições do GSoC de 2005 a 2015.

Nossos resultados sugerem que a participação dos estudantes no GSoC na maioria dos casos é fortemente orientada a busca de recompensas de participação extrínsecas. Mais precisamente, os estudantes procuram alavancar a própria carreira participando de um programa global da Google, de tal forma que as competências técnicas adquiridas no programa possam sinalizar seus talentos para futuros empregadores. Contudo, também existe uma componente de motivação intrínseca forte nos estudantes, que os leva, desde antes da primeira participação no programa, a querer continuar colaborando com projetos SL.

Reconhecemos as limitações de nossa amostra e da abordagem exclusivamente quantitativa de nossa pesquisa. No entanto, esta pesquisa ainda está em fase exploratória, e nossos resultados trazem evidências que podem ser investigadas com maior profundidade em pesquisas futuras. Em particular, pode-se incluir uma abordagem qualitativa para aprofundamento do entendimento das motivações dos estudantes, bem como ampliar o tamanho da amostra, tornando-a mais representativa. Outra abordagem é minerar repositórios de software, listas de e-mails, e *issue-trackers* para averiguar a efetividade da participação.

Referências

- Beecham, S., Baddoo, N., Hall, T., Robinson, H. and Sharp, H. (2008). Motivation in Software Engineering: A systematic literature review. *Information and Software Technology*, v. 50, n. 9-10, p. 860–878.
- Benkler, Y. (2002). Coase's Penguin, or, Linux and The Nature of the Firm. . Yale Law School. <https://dash.harvard.edu/handle/1/11363034>, [accessed on Mar 26].
- Christopherson, L., Idaszak, R. and Ahalt, S. (2010). Developing Scientific Software through the Open Community Engagement Process. In *Files.Figshare.Com*.
- Ciampaglia, G. L. and Taraborelli, D. (2015). MoodBar: Increasing New User Retention in Wikipedia through Lightweight Socialization. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*.
- Deci, E. L., Koestner, R. and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, v. 125, n. 6, p. 627–668; discussion 692–700.
- Fang, Y. and Neufeld, D. (2009). Understanding Sustained Participation in Open Source Software Projects. *Journal of Management Information Systems*, v. 25, n. 4, p. 9–50.
- Forte, A. and Bruckman, A. (2008). Why do people write for wikipedia? Incentives to contribute to open-content publishing. In *Proceedings of 41st Annual Hawaii International Conference on System Sciences (HICSS)*.
- Ghosh, R. (2003). Understanding free software developers: Findings from the FLOSS study. *Perspectives on free and open source software*, p. 1–23.
- Kraut, R. E. and Resnick, P. (2012). *Building Successful Online Communities: Evidence-Based Social Design*. The MIT Press.
- Lakhani, K. and Wolf, R. G. (1 sep 2003). Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. *SSRN Electronic Journal*,

- Lattemann, C. and Stieglitz, S. (2005). Framework for Governance in Open Source Communities. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. . IEEE.
- Misirli, A. T., Verner, J., Markkula, J. and Oivo, M. (may 2014). A survey on project factors that motivate Finnish software engineers. In *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*. . IEEE.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, v. 2, n. 2, p. 175–220.
- Oreg, S. and Nov, O. (sep 2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. *Computers in Human Behavior*, v. 24, n. 5, p. 2055–2073.
- Osterloh, M. and Rota, S. G. (1 mar 2004). Open Source Software Development - Just Another Case of Collective Invention? *SSRN Electronic Journal*,
- Shapiro, J. S. and David, P. A. (23 sep 2008). Community-Based Production of Open Source Software: What Do We Know about the Developers Who Participate? *SSRN Electronic Journal*,
- Silva, J. O., Wiese, I., Steinmacher, I. and Gerosa, M. A. (2015). Pagamento Atrai Colaboradores de Longo Prazo ou Prestadores de Serviço ? Uma Investigação Inicial sobre o Google Summer of Code 2014. In *XII Simpósio Brasileiro de Sistemas Colaborativos, SBSC 2015*.
- Steinmacher, I., Gerosa, M. A. and Redmiles, D. F. (2015). Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work & Social Computing*, p. 1379–1392.
- Steinmacher, I., Wiese, I. S., Conte, T., Gerosa, M. A. and Redmiles, D. (2014). The hard life of open source software project newcomers. *Proceedings of the 7th International Workshop on Cooperative and Human Aspects of Software Engineering - CHASE 2014*, p. 72–78.
- Tirole, J. and Lerner, J. (2002). Some Simple Economics of Open Source. *The Journal of Industrial Economics*, v. 50, n. 2, p. 197–234.
- Trainer, E. H., Chaihirunkarn, C. and Herbsleb, J. D. (2014). The Big Effects of Short-term Efforts : Mentorship and Code Integration in Open Source Scientific Software. v. 2, n. 1, p. 1–5.
- Trainer, E. H., Chaihirunkarn, C., Kalyanasundaram, A. and Herbsleb, J. D. (2014). Community code engagements: Summer of code & hackathons for community building in scientific software. In *Proc. International Conference on Supporting Group Work*.
- Trainer, E. H., Kalyanasundaram, A., Chaihirunkarn, C. and Herbsleb, J. D. (2016). How to Hackathon : Socio-technical Tradeoffs in Brief , Intensive Collocation.
- Yu, J., Jiang, Z. and Chan, H. C. (19 apr 2007). Knowledge contribution in problem solving virtual communities. In *Proceedings of the 2007 ACM SIGMIS CPR conference on 2007 computer personnel doctoral consortium and research conference The global information technology workforce - SIGMIS-CPR '07*. . ACM Press.

Uma Ferramenta de Telepresença Imersiva usando Oculus Rift

Philip K. Dunker, Alberto B. Raposo

Departamento de Informática, PUC-Rio
{pdunker,abraposo}@tegraf.puc-rio.br

Abstract. *Telepresence refers to a set of technologies that allows a person to feel as if he is in a place other than his true location. This work presents a tool composed of a head-mounted display (HMD), the Oculus Rift DK1, integrated with a device called remote head, able to film with a stereo camera and to transmit to the Oculus Rift the images in 3D. At the same time, the HMD's gyroscope captures the user's head orientation and sends it to the remote head, which has servo motors able to rotate it in order to allow the user to move the stereo camera without any additional device. The project's goal is to provide the user an experience of immersive telepresence, with a low cost and a simple interface. Some tests with users were performed and indicated the benefit of some features of the tool, such as stereoscopic vision and the remote head rotation.*

Resumo. *Telepresença refere-se a um conjunto de tecnologias que permite uma pessoa se sentir como se estivesse presente em um lugar que não seja a sua verdadeira localização. Neste trabalho apresentamos uma ferramenta composta por um head-mounted display (HMD), o Oculus Rift DK1, integrado com um dispositivo nomeado de cabeça remota (CR), capaz de filmar usando uma câmera estéreo e transmitir para o Oculus Rift as imagens em 3D. Ao mesmo tempo, o giroscópio do HMD captura a orientação da cabeça do usuário e envia para a cabeça remota, que possui servo motores capazes de girá-la a fim de possibilitar o usuário movimentar a câmera estéreo sem nenhum dispositivo adicional. O objetivo é proporcionar uma experiência de telepresença imersiva, com uma interface de baixo custo e simples de usar. Foram realizados testes com usuários com cenários experimentais, que indicaram o benefício de alguns recursos da ferramenta, como a visão estereoscópica e a rotação da cabeça remota.*

1. Introdução

Telepresença refere-se a um conjunto de tecnologias que permite uma pessoa se sentir como se estivesse presente em um lugar que não seja a sua verdadeira localização. Telepresença imersiva é o próximo passo e o objetivo passa por proporcionar a sensação de que o utilizador se encontra completamente imerso num ambiente remoto [Terrile & Noraky, 2012]. Para que isso ocorra, os sentidos do usuário precisam ser estimulados e, quanto mais sentidos forem estimulados e quanto mais fortes forem esses estímulos, maior será a sensação de imersão [Rodrigues, 2015].

Este trabalho tem como objetivo criar uma ferramenta que dê ao usuário uma

experiência de telepresença imersiva. Acreditamos que é possível atingir este objetivo com o uso de equipamentos de baixo custo existentes no mercado, capazes de capturar imagens de um ambiente em 3D, exibi-las para um usuário em 3D, ao mesmo tempo em que este usuário não precisa utilizar nenhum controle físico em suas mãos para manipular as câmeras, pois através de sensores é possível detectar a rotação da cabeça do usuário e girar a câmera da mesma forma. Isso é possível com o uso de um head-mounted display (ou HMD) em conjunto com um dispositivo desenvolvido neste trabalho, chamado de cabeça remota (CR).

Um HMD, de uma maneira geral, é um equipamento de realidade virtual que exhibe ao usuário as informações de um ambiente (geralmente virtual) em 3D. Os equipamentos mais modernos possuem um giroscópio capaz de detectar os movimentos de rotação da cabeça do usuário. A cabeça remota (CR) é um equipamento que utiliza dois servo-motores, uma câmera estéreo e uma placa de prototipagem que possui um micro controlador, chip inteligente que consiste num circuito processador que possui entradas, saídas e uma memória, e comunica-se com um computador via porta USB.

Com esses dois componentes integrados é possível estimular o sentido da visão, o principal sentido para dar o sentimento de imersão ao usuário, com imagens de um ambiente real em 3D (capturadas pela câmera estéreo) e girar a câmera a partir das rotações capturadas pelo giroscópio, de forma natural para o usuário. Dessa forma, o usuário poderá ter um sentimento de presença imersiva no ambiente remoto real.

A ferramenta desenvolvida pode ser usada, por exemplo, em videoconferências, em novas formas de turismo onde o “viajante” pode ver pontos turísticos como se estivesse presente neles sem sair de casa, dentre outras possibilidades de uso. A ferramenta também poderá ser adaptada para ser usada em outras situações, como acoplada a qualquer robô que possua um operador que o controle de longa distância, como: drone, ROV (*Remotely Operated Vehicle*) ou UGV (*Unmanned Ground Vehicle*). No entanto, o presente projeto não foi desenvolvido para uma aplicação específica. O objetivo do trabalho foi o desenvolvimento de um sistema simples e barato, com o qual pudéssemos fazer avaliações sobre a importância dos recursos de estereoscopia e rotação da cabeça remota na telepresença imersiva.

Este artigo está organizado da seguinte forma: na Seção 2 são discutidos alguns conceitos e trabalhos relacionados. Na Seção 3, a solução proposta e os equipamentos são apresentados. Na Seção 4, é demonstrada a arquitetura do sistema e cada um dos componentes é descrito. Uma avaliação da ferramenta feita por usuários é apresentada na Seção 5. Por fim, a conclusão e discussão sobre possíveis trabalhos futuros encerram o artigo.

2. Trabalhos Relacionados

Existe certo desentendimento na literatura a respeito da definição precisa de telepresença e um dos motivos para isso é sua grande quantidade de denominações, que inclui: *synthetic presence*, *virtual presence*, *social presence*, *ego presence* e *telepresence*. Cada um desses termos refere-se a uma experiência que envolve dar a impressão ao usuário de que ele está em um ambiente remoto, virtual ou não, mediado por computador, por vezes em conjunto com outros equipamentos eletrônicos, e cada um dos termos é usado de forma idiossincrática dentro de seu contexto específico.

Hendrix e Barfield observaram que *virtual presence* (sentir-se presente em um ambiente virtual) ou *telepresence* (sentir-se presente em um ambiente remoto real) são fenômenos indistinguíveis psicologicamente. Portanto, parece ser apropriado fazer comparações baseadas apenas na experiência do usuário [Hendrix & Barfield, 1996].

A ideia de telepresença surgiu pela primeira vez a partir de referências anedóticas para a sensação de ser deslocado de uma sala de controle local para uma área remota ao usar uma ferramenta de teleoperação [Minsky, 1980]. Mesmo antes, em 1971, discutia-se a importância da criação de controles “para ajudar o operador a projetar sua presença” em locais de trabalho remotos e fornecendo monitores que “melhoram a identificação do operador com a tarefa” [Johnsen & Corliss, 1971]. A partir desses estudos percebeu-se que a melhor maneira de se criar uma ferramenta de teleoperação era proporcionando ao teleoperador a sensação mais próxima possível de estar em contato direto com o ambiente onde deve ser executada a tarefa.

A dificuldade em compreender o significado de telepresença parte da maneira relativamente descuidada com que o termo é usado. Draper [Draper, 1995] encontrou três definições de telepresença de uso comum: a simples, a cibernética e a experimental. Na definição da simples, telepresença refere-se à habilidade de operar em um ambiente mediado por computador. Por exemplo, Travis, Watson e Atyeo [Travis et al., 1994] escreveram que telepresença “implica em controlar máquinas (usualmente robôs) à distância”.

Na definição cibernética, telepresença é um índice da qualidade da interface humano-computador. Por exemplo, Thomas e Karl Smith [Smith & Smith, 1990] argumentaram que telepresença ocorre quando “as características operacionais da interface humano-computador são compatíveis com as capacidades de desempenho comportamental-fisiológicas e as limitações do ser humano”.

Na definição experimental, telepresença é um estado mental em que o usuário sente fisicamente presente dentro do ambiente mediado por computador [Akin et al., 1983]. Por exemplo, Sheridan [Sheridan, 1992a] descreveu telepresença como uma “sensação de estar fisicamente presente com objetos virtuais no local remoto teleoperado”. Sheridan [Sheridan, 1992b] novamente a descreveu em outro lugar como uma “ilusão” que ocorre quando o usuário “se sente fisicamente presente no local remoto”.

A diferença fundamental entre a telepresença cibernética e experimental é que a definição cibernética se refere à eficiência de displays e controles, enquanto a definição experimental refere-se a algo experimentado pelo usuário humano. A característica diferencial da telepresença experimental é o estado mental, o que inclui a suspensão da consciência do usuário e o afastamento dele para o ambiente mediado por computador. A qualidade da interface humano-computador é frequentemente citada como um fator determinante da intensidade da experiência, mas telepresença experimental é claramente algo além da telepresença cibernética.

No presente trabalho, não entraremos mais a fundo na discussão sobre os diferentes tipos de telepresença. Usaremos a definição mais simples de telepresença de Hendrix e Barfield [Hendrix & Barfield, 1996], mencionada anteriormente: sentir-se presente em um ambiente remoto real. Em particular, estamos interessados no conceito

de telepresença imersiva, que significa prover ao usuário a habilidade de sentir-se presente e interagir com o ambiente remoto através de gestos e movimentos [Rodrigues, 2015].

A telerrobótica, área da robótica responsável pelo controle de robôs semiautônomos à distância, exige que algum tipo de telepresença seja criado, por isso a telepresença e a telerrobótica se misturam. Neste campo de pesquisa existem projetos que usam dispositivos HMDs para o usuário ter a sensação de presença e ser capaz de realizar alguma tarefa robótica com maior facilidade. Isso acontece, principalmente, porque existem ambientes que são inacessíveis para o homem ou porque colocaria a vida de uma ou mais pessoas em risco. Esses sistemas são usados em diversas situações: na exploração espacial como o robô “Robonaut” [Goza, 2004], para fins militares, como em missões de busca e salvamento [Elliott et al., 2012], entre outras situações.

Antes do ano de 2012, o acesso a dispositivos HMDs era mais difícil, já que eram dispositivos caros e com baixa qualidade visual. Em 2012, foi lançada a primeira versão, para desenvolvedores, dos Oculus Rift a um preço acessível e com uma qualidade visual satisfatória. Desde então muitos outros projetos usando esse equipamento começaram a surgir, porém, como ainda estão em fase de desenvolvimento, não há muita informação a respeito. Encontramos alguns projetos similares ao apresentado neste artigo basicamente em vídeos do YouTube ou em sites das instituições desenvolvedoras.

Em 2012, Linda Elliott descreveu experiências feitas com veículos terrestres não-tripulados, usados pelo exército americano em missões de busca e salvamento, onde operadores utilizaram um sistema com visão e áudio estereoscópico e 13 dos 18 operadores preferiram o sistema de telepresença imersivo em oposição aos utilizados hoje em dia com câmeras e monitores comuns [Elliott et al., 2012]. Isso indica o potencial benefício da telepresença imersiva também em operações militares.

A ideia de controlar drones em primeira pessoa também não é nova, existem os chamados drones FPV (*first person view*). Porém, eles possuem câmeras comuns. A Intuitive Aerial, uma empresa da Suécia especializada em drones, está desenvolvendo um projeto objetivando criar drones FPV com câmera estereoscópica (www.intuitiveaerial.com).

Da mesma forma, a manipulação remota de robôs também tem se utilizado de estereoscopia e HMDs. Por exemplo, há um projeto do Robotics Lab (<http://robotics.cs.uml.edu>) da Universidade de Massachusetts em que o objetivo é manipular um robô remotamente do ponto de vista do robô com visão 3D, e assim executar tarefas com precisão¹.

O sistema desenvolvido no presente trabalho relaciona-se com os citados acima porque também tem o objetivo de dar a sensação ao usuário, com um HMD, de que ele se encontra em outro ambiente, onde está a cabeça remota, equipamento criado para gravar e enviar imagens em 3D e interagir girando a câmera. Porém, diferentemente da maioria dos trabalhos acima citados, não visamos inicialmente uma aplicação específica, mas sim o desenvolvimento de um sistema simples e barato que possa ser usado em

¹ www.youtube.com/watch?v=JHlz-Y5qCmY

diferentes aplicações e, no caso específico deste artigo, queremos avaliar o efeito da visão estereoscópica e da rotação remota da cabeça em um cenário típico de videoconferência.

3. Solução Proposta

Para a criação de uma ferramenta de telepresença imersiva, a solução proposta é trazer para o ambiente real a imersão atingida pelos HMDs em ambientes virtuais. Estes equipamentos, além de permitirem a visualização de imagens em 3D, possuem sensores capazes de rastrear a posição e orientação da cabeça do usuário. De posse dessas informações, um software é capaz de transladar e rotacionar uma câmera virtual em um mundo virtual imitando os movimentos reais do usuário, conseqüentemente, este se sente mais imerso do que se a câmera estivesse parada ou se ele a movimentasse com controles comuns. Segundo Sutherland, a variação que ocorre na imagem quando o usuário move sua cabeça deve alternar exatamente do mesmo modo que a imagem de um verdadeiro objeto mudaria para movimentos semelhantes do usuário. Ela é mais importante para a imersão do que a própria apresentação estereoscópica [Sutherland, 1968].

Nossa solução propõe a criação de um equipamento, para ser utilizado juntamente com um HMD, capaz de filmar em 3D e girar esta câmera imitando os movimentos do usuário. A solução proposta é dividida em dois módulos:

- Módulo do operador: responsável por exibir para o operador em 3D imagens do ambiente remoto em tempo real, capturar movimentos de rotação da cabeça do operador e enviá-los para o módulo da cabeça remota.
- Módulo da cabeça remota: responsável pela captura de imagens em 3D e enviá-las para o módulo do operador, ao mesmo tempo em que recebe rotações da cabeça do operador e gira a câmera estéreo com as rotações recebidas.

3.1. Módulo do Operador

O hardware do módulo do operador consiste basicamente no Oculus Rift DK 1, com resolução total da tela de 1280 x 800, mas como é dividida, então a resolução fica 640 x 800 por olho.

Os óculos têm incorporado três sensores que monitoram movimentos da cabeça do utilizador: magnetômetro, acelerômetro e giroscópio. A combinação desses sensores permite rastrear com precisão as rotações da cabeça do utilizador nos eixos X, Y e Z.

O software deste módulo precisa continuamente receber um par de imagens do módulo da cabeça remota, enviá-las para o Rift, ao mesmo tempo que consulta a orientação dele e envia para a cabeça remota.

3.2. Módulo da Cabeça Remota

A cabeça remota (CR) é o módulo mais complexo porque é a junção de três equipamentos eletrônicos. Ela possui os seguintes equipamentos: uma câmera: DUO MLX, da DUO 3D²; uma placa StartUSB for PIC, da MikroElektronika³; e dois mini

² www.duo3d.com

servo motores ES9258, da EMAX⁴.

A câmera DUO MLX é uma câmera binocular que permite filmar em estéreo e é bem compacta, de dimensões 52 x 25.4 x 13.3 milímetros e peso 12,5 gramas. Possui uma boa taxa de quadros por segundo (56 quadros por segundo na sua resolução máxima de 752x480). O único ponto negativo é que as imagens são monocromáticas.

A placa StartUSB for PIC é uma placa com componentes eletrônicos voltada para o desenvolvimento de sistemas embarcados. Ela possui um microcontrolador e os componentes eletrônicos necessários para a comunicação do microcontrolador com um computador via USB.

O servo motor é um dispositivo eletromecânico, que apresenta movimento proporcional a um comando. Sua característica principal é a capacidade de movimentar o seu braço até uma posição e mantê-lo, mesmo quando sofre uma força em outra direção. Em contraste com os motores contínuos que giram indefinidamente, o eixo dos servo motores possui uma faixa de rotação que varia de acordo com o modelo, mas são precisos quanto à sua posição. Foram utilizados dois servo motores do modelo ES9258 da EMAX, que embora tenham bom torque e velocidade para girar a câmera DUO MLX, têm uma faixa de rotação de apenas 60°.

Esses três equipamentos juntos formam a Cabeça Remota (CR) – Figura 1. Com apenas dois servo motores, optamos por girar a cabeça remota apenas nos eixos X e Y, excluindo o giro no eixo Z (movimento de *roll*, conforme detalhe na Figura 1). Foi necessário ainda usar uma fonte externa de 5V que também foi usada de base para prender a placa e os motores. O suporte da câmera é bem simples: um dos motores segura a câmera e o outro motor segura a base do primeiro motor.

4. Implementação

Visando uma implementação modularizada capaz de manipular e testar cada equipamento eletrônico de forma independente, foram implementados cinco componentes. A Figura 2 mostra a integração dos componentes com os módulos.

- CameraDUODll – componente responsável pela comunicação com a câmera DUO MLX, armazenamento das imagens e disponibilização dessas imagens através de uma interface. Nessa DLL ficam todas as funções necessárias para manipular este hardware: a função de inicialização, de início de captura das imagens, de registro da função de call-back, de encerramento do uso do hardware, entre outras.
- Visualizador – componente que recebe as imagens e as reproduz no Oculus Rift. Este componente foi desenvolvido no Unity3D, ferramenta de desenvolvimento de jogos que possui integração com a biblioteca do Oculus Rift, a LibOVR.
- RiftSensor – componente responsável pela captura dos movimentos da cabeça do usuário, ou seja, monitora os sensores do Oculus Rift e é também responsável pelo envio das rotações para o componente CommWithPic.

³ www.mikroe.com

⁴ www.emaxmodel.com

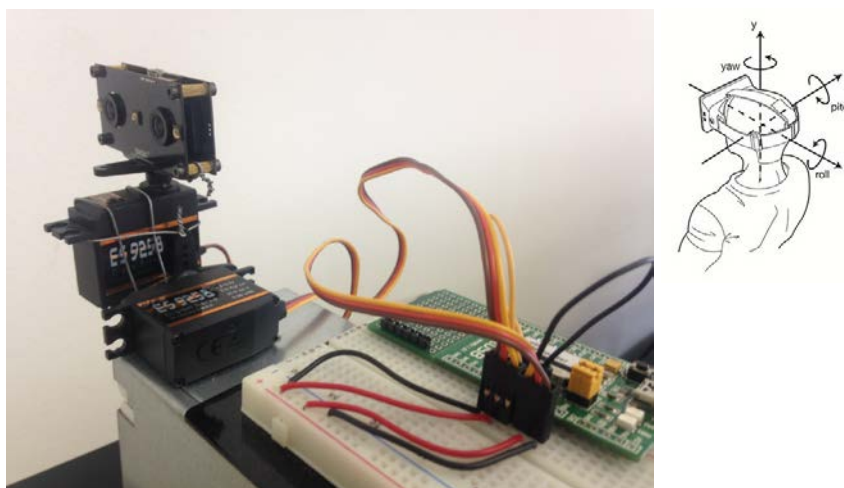


Figura 1. Cabeça remota e eixos de rotação da cabeça do usuário com o HMD.

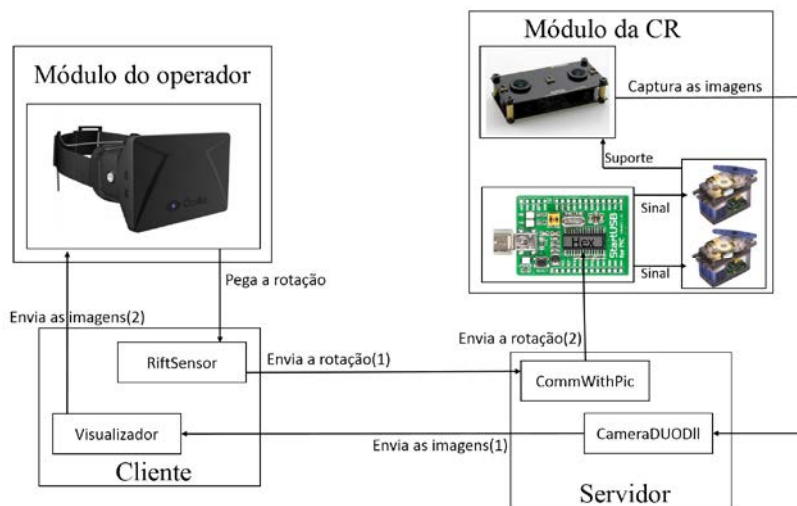


Figura 2. Diagrama dos módulos e componentes.

A comunicação com o microcontrolador é dividida em duas partes, o software executado em um computador e o programa gravado no microcontrolador:

- CommWithPic – componente responsável pelo recebimento e envio das rotações para o microcontrolador. Este componente foi escrito na linguagem C#; ele recebe informações de quanto cada motor deve girar e em qual direção e repassa essas informações para o componente Hex (programa gravado no microcontrolador).
- Hex – componente responsável pelo recebimento das rotações e atuador nos servo motores (programa executado no microcontrolador).

Os componentes CameraDUODII, CommWithPic e Hex fazem parte do servidor e o cliente é composto pelos componentes Visualizador e Rift Sensor. Dessa forma o servidor provê as imagens do ambiente remoto para o cliente.

Duas implementações distribuídas foram testadas: a primeira implementação utilizava os mecanismos do Unity3D e a segunda implementação utilizava a API de

sockets do Windows (Windows Sockets API - WSA). No entanto, nenhuma dessas implementações resultou em um tempo de latência satisfatório para uma atividade em tempo real. Por isso, implementamos o sistema localmente para realização dos testes com usuários. Para que o operador com o Oculus Rift e a Cabeça Remota ficassem em ambientes diferentes foram usados dois cabos extensores do tipo USB e HDMI de 10 metros, que são os cabos necessários para o funcionamento do Oculus Rift, conforme é mostrado na próxima seção.

5. Avaliação com Usuários

A avaliação foi feita baseando-se nas respostas dos usuários por meio da aplicação de dois questionários após a utilização da ferramenta em um cenário mimetizando uma videoconferência que englobava um teste de estereoscopia. Esse experimento teve como objetivo avaliar a sensação de presença dos usuários no ambiente remoto, bem como se a ferramenta de telepresença imersiva traz benefícios ao utilizador quando se trata de videoconferências.

5.1. Cenário de Teste

Doze usuários foram convidados a participar do projeto, voluntariamente, após assinatura do Termo de Consentimento Livre e Esclarecido. O cenário de teste foi composto pelo participante em um ambiente com o HMD enquanto o avaliador encontrava-se em outra sala diante da cabeça remota (CR) e com o modelo físico (MF) usado no teste de estereoscopia (Figura 3).

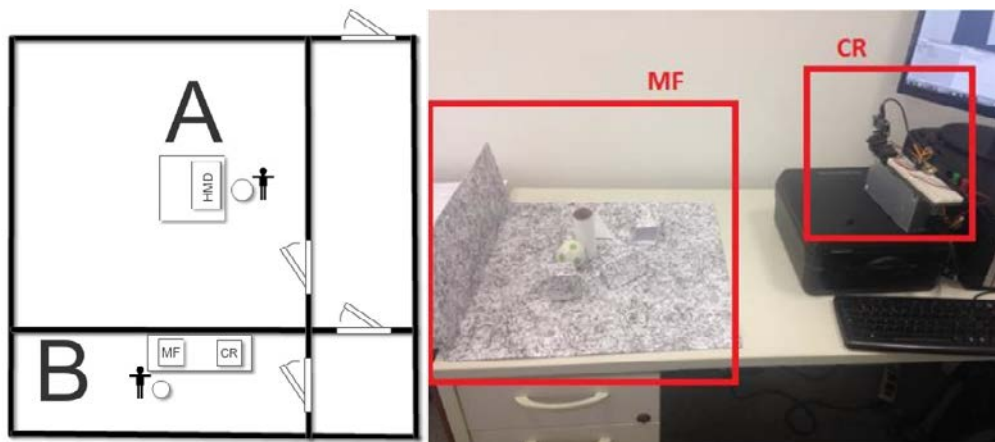


Figura 3. À esquerda: cenário de teste; A: ambiente dos participantes; B: ambiente do avaliador. À direita: cabeça remota (CR) e modelo físico (MF) no ambiente do avaliador.

Para que o ambiente de uma videoconferência fosse inteiramente recriado (com imagem e som) foi necessária a utilização de um sistema para transmissão de áudio, como Skype, ligação via WhatsApp ou ligação telefônica. Dessa forma, foi possível que o usuário interagisse com o avaliador.

O modelo físico utilizado consiste em uma maquete com seis formas geométricas, três delas brancas e três com a mesma textura do fundo (Figura 3). O teste de estereoscopia, por sua vez, apresentou três layouts distintos do modelo físico ao participante que diferiam nas posições dessas formas geométricas. O teste também

contou com três modos de utilização da cabeça remota:

- Com visão estereoscópica e sem rotação das câmeras (3D sRot);
- Sem visão estereoscópica e com rotação das câmeras (2D cRot);
- Com visão estereoscópica e com rotação das câmeras (3D cRot).

A Tabela 1 mostra como ficou definido o modo de utilização da cabeça remota em cada um dos layouts para cada participante.

Tabela 1. Ordem de execução dos testes para cada participante.

Número do participante	Ordem dos layouts (modo da CR)
1,4, 7, 10	A (3D sRot), B (2D cRot), C (3D cRot)
2, 5, 8, 11	A (3D cRot), B (3D sRot), C (2D cRot)
3, 6, 9, 12	A (2D cRot), B (3D cRot), C (3D sRot)

Inicialmente, um tempo do teste foi dispensado para o usuário acomodar-se com o equipamento. Em seguida, o avaliador exibiu um dos layouts com um dos modos de operação da cabeça remota acima citados, iniciando o teste de estereoscopia. O usuário foi questionado em relação à quantidade de formas geométricas e a posição ocupada por elas na maquete, avaliado assim a profundidade dos objetos (do mais perto para o mais longe) enquanto o avaliador anotava as respostas. Este procedimento foi repetido até que os três modos fossem experimentados pelo usuário. Por fim, o usuário respondeu a dois questionários, conforme será descrito a seguir.

5.2. Metodologia de Avaliação

O uso de questionários, uma forma bastante utilizada para medição de presença, permite adquirir dados a respeito da experiência do usuário, tendo em vista que é composto por um conjunto de perguntas direcionadas. Além disso, possui baixo custo, fácil aplicabilidade e não interrompe a experiência do usuário, pois permite a análise e interpretação dos dados em um segundo momento pelo avaliador. Uma desvantagem é que, por serem aplicados posteriormente, dependem da capacidade de memorização dos participantes, que está sujeita ao esquecimento de detalhes vivenciados no momento da experiência.

Um primeiro questionário experimental (Tabela 2) foi desenvolvido especificamente para este cenário de teste. Ele é composto por perguntas que visam obter as impressões do usuário ao utilizar o equipamento e avalia questões subjetivas que variam de acordo com cada participante. Conforto, facilidade de manuseio, sincronia entre rotação da câmera e do HMD e a melhora comparativa entre a visão monocular e binocular foram os parâmetros avaliados por este questionário.

O questionário “*Igroup Presence Questionnaire*” (IPQ) foi o meio utilizado para mensurar o sentimento de presença do usuário no ambiente remoto. Em sua origem, ele foi idealizado para avaliar o senso de presença do usuário em um ambiente virtual. Neste trabalho, o IPQ foi adaptado para avaliar o senso de presença no cenário relatado acima, um ambiente remoto real. Os termos que faziam referências ao ambiente virtual

(“*virtual*”) foram substituídos por referências ao ambiente remoto (“*remote*”).

Tabela 2. Questionário experimental.

Questionário Experimental
Perguntas Iniciais
Você está se sentindo confortável com o HMD?
Você consegue movimentar a câmera da CR com facilidade?
Você consegue perceber algum atraso da imagem comparado ao som?
Perguntas sobre MF (feito 3 vezes, alternado layouts e modos de operação da CR)
Quantos objetos geométricos você vê nesse modelo?
Diga a ordem desses objetos de mais perto para mais longe:
Perguntas finais
Qual a sua impressão com relação a controlar a câmera a partir dos movimentos da sua cabeça?
A velocidade do giro da câmera acompanha a velocidade de movimentação da sua cabeça?
Após mudar da visão mono para a visão estereoscópica, o que você acha da sua noção de profundidade?
Em algum momento da experiência você se sentiu mal? (desorientação, tontura, enjoo, vista cansada ou náusea)

O IPQ foi criado por um grupo de pesquisadores, o Igroup Project Consortium (www.igroup.org/projects/ipq/), com base em questionários anteriores. Numa primeira fase 240 participantes responderam um grande número de itens em uma pesquisa on-line. Estes dados foram analisados por métodos estatísticos e reaplicados em uma segunda fase de pesquisa on-line com mais de 200 participantes. Como resultado, 14 itens foram selecionados para compor o IPQ através de uma escala de sete pontos (-3 a +3), aplicada de forma rápida e simples. Destes, um item tem uma abordagem genérica avaliando a “sensação de estar lá” enquanto os outros 13 medem três fatores (Figura 4): presença espacial, envolvimento e realismo experimentado.

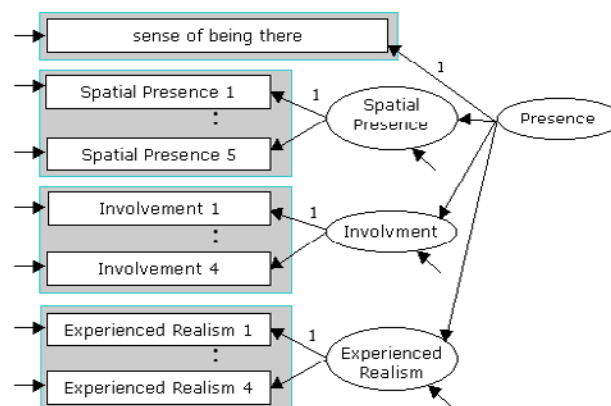


Figura 4. Fatores do IPQ (<http://www.igroup.org/pq/ipq/construction.php>).

Dos 14 itens do IPQ, os últimos 4 (11-14) foram excluídos dos nossos testes porque fazem referência ao realismo experimentado e, por se tratar de ambiente remoto real, a avaliação do realismo não seria coerente.

Devido a sua grande utilização em diversos estudos, este questionário também permite a comparação dos resultados obtidos com os dados coletados por outros pesquisadores, reunidos em um banco de dados.

5.3. Resultado dos Testes

Os 12 participantes dos testes tinham média de idade de 28,5 anos com desvio padrão de 2.5 anos, a maioria (10 deles) do sexo masculino, havendo prevalência na área de exatas. Destes, 75% não possuíam problema de vista e não utilizavam óculos ou lentes de contato. Três participantes não tinham experiência prévia com um HMD.

A análise das respostas da primeira parte do Questionário Experimental, realizada oralmente pelo avaliador durante o período de acomodação do usuário, permitiu inferir que todos os participantes sentiram-se confortáveis ao utilizar o equipamento, sendo que 25% deles classificaram a experiência como muito confortável, 50% como confortável e 25% como normal. Além disso, todos os participantes conseguiram movimentar facilmente a câmera da CR e não perceberam nenhuma latência na transmissão do som em relação à imagem.

A segunda parte do Questionário Experimental questionava sobre a quantidade e a ordem dos objetos presentes no modelo físico. A quantidade era avaliada através da percepção do usuário do número de objetos presentes em seu campo visual, incluindo objetos camuflados com o objetivo de dificultar a visualização dos mesmos. A ordem dos objetos avaliava a noção de profundidade do usuário que deveria informar ao avaliador a sequência dos objetos de mais perto para mais longe.

A Figura 5 representa, de forma genérica, os acertos e os erros obtidos em cada modo de operação. No modo 2D com rotação houve um índice de acerto de 25% (3 usuários). Já no modo 3D sem rotação, aproximadamente, 33,3% dos usuários obtiveram êxito em suas respostas. O maior índice de acertos foi constatado no modo 3D com rotação. Neste, 75% dos participantes responderam corretamente.

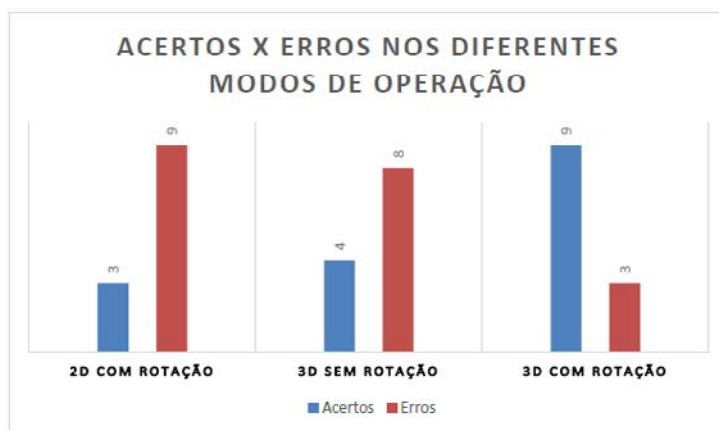


Figura 5. Erros vs. acertos.

A Figura 6 classifica os tipos de erros obtidos, subdividindo-os em erros na ordem dos objetos, erros na quantidade ou em ambos. Dos nove participantes que não responderam corretamente no modo 2D com rotação, três erraram a ordem, quatro não visualizaram algum objeto e dois erraram ambos os quesitos. Já no modo 3D sem rotação, dois erraram na ordem, cinco não visualizaram algum objeto e um errou ambos os quesitos. No modo 3D com rotação, que obteve o menor índice de erros, dois usuários erraram a ordem e um o número de objetos. Neste modo nenhum participante errou os dois quesitos. A partir desses dados podemos perceber que o modo 3D sem

rotação diminuiu o erro na ordem, mas aumentou o erro na não visualização de objetos (quantidade). Isso mostra que o 3D ajuda a compreender a profundidade dos objetos e a rotação da câmera ajuda a visualizar os objetos camuflados, que têm textura igual à do fundo. Portanto, o modo 3D com rotação, por ser uma combinação dos dois fatores, mostrou-se superior aos demais.

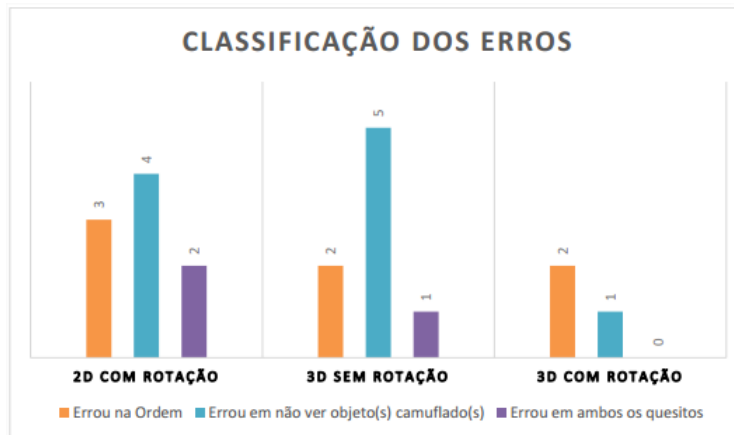


Figura 6. Classificação dos erros.

Após vivenciar os três modos de operação, o usuário encerrou sua experiência com o sistema e foi submetido à terceira parte do Questionário Experimental, em que suas impressões seriam avaliadas. O controle da câmera a partir dos movimentos do usuário foi considerado bastante natural por 42% dos participantes. Este controle foi julgado bom por 50% e médio por 8%. A velocidade do giro da câmera acompanhou a velocidade de movimentação do usuário em 75% dos casos (acompanha = 58%; acompanha muito bem = 17%). Para 25% dos participantes ocorre um atraso entre essas velocidades.

A mudança da visão mono para a estereoscópica foi representativa para todos os usuários. A noção de profundidade melhorou muito em 67% dos casos e para 33% a melhora não foi tão expressiva. Por fim, o usuário foi questionado se em algum momento da experiência sentiu algum tipo de desconforto, tais como: desorientação, tontura, náusea ou vista cansada. Nenhuma dessas sensações ocorreu em 75% dos usuários.

A última etapa do teste com usuários foi o preenchimento do IPQ. Apesar da escala do questionário ter uma faixa de -3 a +3, para fins de análise e estatísticas, os valores são transformados para uma nova faixa de 0 a 6, sendo 0 a pior e 6 a melhor pontuação com exceção dos itens SP2 e INV3 que possuem uma escala invertida. A Tabela 3 mostra a média e o desvio padrão das respostas dos usuários, com os itens SP2 e INV3 já corrigidos.

O item G1 é quesito genérico. Ele sozinho compõe o fator que avalia a sensação de estar no ambiente remoto (G). A média dos itens SP1 ao SP5, que compõem o fator de presença espacial (SP), resultou em 4,17, enquanto a média dos itens INV1 ao INV4, o fator envolvimento (INV), resultou no valor de 3,96 (Figura 7).

Tabela 3. Média e desvio padrão das respostas ao IPQ (RE significa “remote environment”).

IPQ item	“Atalho”	Média	Desvio Padrão
G1	<i>sense of being there</i>	4,25	0,83
SP1	<i>sense of surrounded by RE</i>	4,17	0,69
SP2	<i>felt like just perceiving pictures</i>	4,67	1,31
SP3	<i>sense of not being in RE</i>	5,58	1,61
SP4	<i>sense of acting in RE</i>	4,08	1,38
SP5	<i>sense of being present in RE</i>	4,33	0,85
INV1	<i>awareness of my real (local) env</i>	3,83	1,34
INV2	<i>not aware of my real environment</i>	4,17	1,46
INV3	<i>attention to my real environment</i>	3,92	1,85
INV4	<i>attention captivated by RE</i>	3,92	1,11

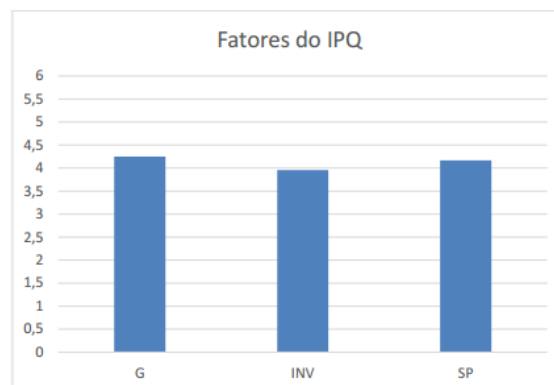


Figura 7. Resultados obtidos nos fatores do IPQ.

Como dito anteriormente, o IPQ permite a análise comparativa entre diversos trabalhos que usaram este questionário. O PQII⁵, estudo realizado pelo Igroup, avaliou os jogos Half-Life e Tomb Raider. Além disso, também foi feita uma comparação entre a média dos 542 dados que compõem o banco de informações do IPQ. A Figura 8 mostra a comparação entre os perfis de presença desses estudos e o obtido no presente trabalho. O perfil de presença do sistema desenvolvido foi superior aos demais em todas as variáveis. Por se tratar de um ambiente remoto real, vale lembrar que o fator realismo foi excluído. Cabe ressaltar, no entanto, que essa comparação é apenas ilustrativa, pois foram experimentos realizados por grupos diferentes, com números de usuários diferentes, com equipamentos e propósitos diferentes. O dado mais relevante é verificar que o sentimento de presença do dispositivo desenvolvido ficou acima da média do banco de informações do IPQ (em preto na Figura 8).

⁵ <http://www.igroup.org/pq/ipq/data.php>

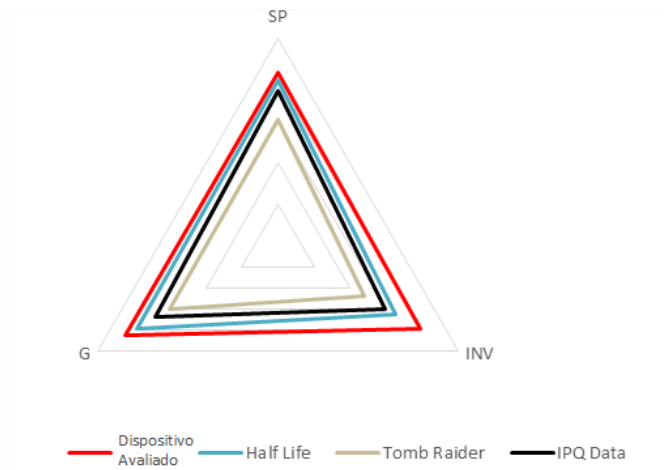


Figura 8. Comparação dos resultados do IPQ com outros trabalhos.

6. Conclusão

Neste trabalho foi proposta e avaliada uma ferramenta de telepresença imersiva. Apesar de algumas limitações no hardware, como a câmera monocromática, faixa de rotação dos servo motores de 60° e ausência de um terceiro servo motor para rotação no eixo Z, o protótipo desenvolvido conseguiu proporcionar ao usuário uma experiência de telepresença imersiva e avaliar a impressão durante seu uso.

A análise dos testes permitiu inferir que o dispositivo consegue reproduzir o ambiente de uma videoconferência e possibilita uma melhora em relação aos dispositivos comuns. Isso ocorre devido à visualização do ambiente em 3D e pela manipulação da câmera através do HMD, sem a necessidade de controles adicionais. Os resultados demonstraram que a ferramenta desenvolvida permite uma boa sensação imersiva. O uso da visão estéreo foi considerado um grande avanço na visualização do ambiente remoto em relação à visão mono, tendo em vista o maior número de acertos nos testes e da percepção, inclusive, de objetos camuflados.

Um dado obtido que pode ter gerado dúvidas foi o fato de durante a primeira fase do questionário experimental todos os participantes terem classificado a experiência como confortável e em um momento posterior, 25% terem experimentado alguma sensação de desconforto. Isso pode ser explicado pela primeira fase do questionário ter sido realizada no momento inicial dos testes, enquanto o usuário experimentava a ferramenta. O desconforto, provavelmente, foi devido ao maior tempo de exposição ao dispositivo. Além disso, alguns participantes mencionaram trepidações durante a movimentação da cabeça remota, o que também pode contribuir para o mal-estar. Essas trepidações podem ter sido causadas pelo suporte de fio de ferro criado para acoplar os motores entre si e à câmera (Figura 1). Esse suporte pode ser melhorado utilizando engrenagens e roldanas que levariam a uma junção mais coesa, evitando a trepidação.

Em face dos dados apresentados o sistema desenvolvido cumpriu as expectativas. Porém, mudanças podem aperfeiçoar o protótipo, bem como destiná-lo a outros fins. As principais melhorias seriam a utilização de hardwares mais modernos como uma câmera estéreo colorida, servo motores de rotação contínua que permitem a rotação em 360° e a utilização de mais um servo motor para possibilitar a rotação da

câmera nos 3 eixos de rotação. Mudanças na implementação do software para um sistema distribuído também são essenciais para a utilização do sistema em longa distância. A utilização do protocolo UDP para envio das imagens seria uma alternativa para o desenvolvimento de um sistema distribuído com pouca latência.

Referências

- Akin, D. L. et al. (1983). "Space Applications of Automation, Robotics and Machine Intelligence Systems (ARAMIS), phase 2". *Volume 1: Telepresence technology base development*.
- Draper, John V. (1995). "Teleoperators for advanced manufacturing: Applications and human factors challenges". *International Journal of Human Factors in Manufacturing*, 5, 1: 53-85.
- Elliott, Linda R. et al. (2012). "Robotic telepresence: Perception, performance, and user experience". Technical Report No. ARL-TR-5928. Army Research Lab Aberdeen Proving Ground MD Human Research and Engineering Directorate.
- Goza, S. M. et al. (2004). "Telepresence control of the NASA/DARPA robonaut on a mobility platform". In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 623-629.
- Hendrix, C.; Barfield, W. (1996). "Presence within virtual environments as a function of visual display parameters". *Presence: Teleoperators & Virtual Environments*, 5, 3: 274-289.
- Johnsen, Edwin G.; Corliss, William R. (1971). "Human factors applications in teleoperator design and operation". New York: Wiley-Interscience.
- Minsky, M. (1980). Telepresence. *Omni*, p. 45-51.
- Rodrigues, F. A. C. (2015). "Immersive Telerobotic Modular Framework using stereoscopic HMDs". 110 f. Dissertação (Mestrado Integrado em Engenharia Informática e Computação) – Faculdade de Engenharia, Universidade do Porto.
- Sheridan, Thomas B. (1992a). "Musings on telepresence and virtual presence". *Presence: Teleoperators & Virtual Environments*, 1,1: 120-126.
- Sheridan, Thomas B. (1992b). Telerobotics, automation, and human supervisory control. MIT press.
- Smith, T. J.; Smith, K. U. (1990). "The human factors of workstation telepresence". In: NASA, Lyndon B. Johnson Space Center, Third Annual Workshop on Space Operations Automation and Robotics (SOAR 1989), 235-250.
- Sutherland, I. E. (1968). "A head-mounted three dimensional display." In: Proceedings of the December 9-11, 1968, fall joint computer conference, part I. ACM.
- Terrile, Richard J., Noraky, James. (2012) "Immersive telepresence as an alternative for human exploration". Aerospace Conference, IEEE.
- Travis, David; Watson, Toby; Atyeo Mike (1994). "Human psychology in virtual environments". *Interacting with virtual environments*: 43-59.

Carona UFMG: Investigando as Relações de Troca, Segurança e Recompensa no Desenvolvimento de uma Aplicação Móvel Voltada a uma Comunidade

Antônio Henrique Gonçalves Leite, Luiz Paulo Damilton Corrêa
Raquel Oliveira Prates, Fabrício Benevenuto, Pedro Olmo Vaz de Melo

¹Departamento de Ciência da Computação (DCC)
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brasil

{antonioh, luiz.correa, rprates, fabricio, olmo}@dcc.ufmg.br

Abstract. *Collaborative consumption applications have been increasingly used, with new proposals for the exchange of goods and services between users. But what are the users motivations to offer and use a particular service when an application is directed to a local community and there is no exchange of money? In this paper, we use the development of a new mobile application for carpooling in a Brazilian university to discuss this issue, based on the results of a survey (n = 287) and considerations of the Brazilian literature about this subject. Thus, we have raised some considerations of reward, exchange and security, and present a new project implementation.*

Resumo. *Aplicações de economia colaborativa têm sido cada vez mais utilizadas, com novas propostas de mercado para a troca de bens e serviços entre usuários. Porém, qual a motivação dos usuários para ofertar e usufruir um determinado serviço quando uma aplicação é voltada a uma comunidade restrita e não há uma troca direta de moeda? Neste trabalho, tomamos como base o desenvolvimento de um novo aplicativo móvel para caronas em uma universidade brasileira para discutirmos essa questão, a partir de resultados de um questionário (n = 287) e considerações da própria literatura brasileira. Com isso, levantamos algumas considerações dos usuários sobre recompensa, troca e segurança, e apresentamos uma nova implementação de sistema.*

1. Introdução

Economia colaborativa ou economia compartilhada (dos termos em inglês “*collaborative consumption*” [Rachel Botsman 2010] e “*sharing economy*” [Hamari et al. 2015]) é um modelo híbrido de mercado que foca no estabelecimento de redes pessoa-para-pessoa de compartilhamento de bens e serviços. Empresas que adotam este modelo não oferecem diretamente um bem ou serviço físico, mas proporcionam uma plataforma que permite que os seus clientes interajam entre si para a troca desses bens e serviços, servindo unicamente como mediadoras entre um **ofertante** e um **demandante**. Recentemente, este modelo tem se tornado popular graças ao surgimento e crescimento de plataformas como a Uber¹, para serviços de transporte, a Airbnb², para hospedagens, e o TaskRabbit³, para

¹<https://www.uber.com/pt/>

²<https://www.airbnb.com.br/>

³<https://www.taskrabbit.com/>

serviços domésticos. Essas aplicações permitem o contato real, fora do sistema on-line, entre pessoas que a princípio não se conhecem, em troca de um serviço que beneficia tanto o ofertante (motorista; anfitrião da hospedagem; prestador do serviço doméstico), que recebe, nestes exemplos, uma quantia em dinheiro, quanto o demandante (passageiro; hóspede; requerente do serviço doméstico), que usufrui do serviço. Neste trabalho, pretendemos investigar o quanto uma mudança de contexto, de uma situação paga para uma voluntária, de uma aplicação geral para uma voltada a uma comunidade específica, podem alterar as funcionalidades de uma solução e gerar uma maior adesão à ferramenta.

No caso específico de uma aplicação de carona, um grande desafio no seu desenvolvimento é definir as dinâmicas de troca entre os seus usuários. Em [Carroll and Bellotti 2015], os autores analisam as diferentes possibilidades de relações de troca que podem existir em sistemas que intermedeiam serviços pessoa-para-pessoa. Um serviço pessoa-para-pessoa pode ser feito a partir de pagamento em dinheiro, mas também pode ser feito de forma voluntária, ou banco de horas, ou novas moedas, como o *Bitcoin* [Nakamoto 2008]. Para cada paradigma, há diferentes dinâmicas de troca e implicações para o design da aplicação. Considerações realizadas em sistemas de transporte pago, como no caso do Uber, nem sempre serão as mesmas para sistemas de caronas, em que há um caráter voluntário. O mesmo vale para projetos que têm em vista comunidades específicas, como é o caso de uma comunidade universitária. Como apontado por [Carroll and Bellotti 2015], as dinâmicas de troca podem variar drasticamente com o contexto.

Assim, neste artigo, discutiremos essas relações de troca, segurança e recompensa a partir de considerações feitas durante o processo de desenvolvimento de um aplicativo de caronas para uma universidade, neste caso, na Universidade Federal de Minas Gerais (UFMG). Para isso, apresentamos o que já foi discutido por alguns trabalhos na literatura (Seção 2); a motivação dos usuários a partir de um questionário (Seção 3); as decisões de design durante o desenvolvimento de um protótipo funcional, com base nessas duas etapas anteriores (Seção 4); uma síntese dessas considerações (Seção 5) e concluímos também divulgando uma versão da aplicação na universidade (Seção 6).

2. Trabalhos Relacionados

Nesta seção apresentamos os trabalhos relacionados à nossa pesquisa organizados em dois temas de interesse: coprodução e outras aplicações de carona.

2.1. Coprodução e relações de troca, segurança e recompensa

Em sistemas de colaboração pessoa-para-pessoa em que não é visada a troca direta de dinheiro ou de bens entre os usuários, novas relações devem ser investigadas. O que motiva um usuário a ofertar determinado serviço? O que ele ganha em troca? Quais as garantias para o demandante que o serviço do ofertante é mesmo seguro? No caso de um aplicativo de caronas, quais as reais vantagens de um motorista em usar determinado aplicativo?

Em [Mauss 1974], o antropólogo Marcel Mauss discute as relações de troca nas sociedades arcaicas. É um dos estudos antropológicos mais importantes sobre reciprocidade e relações de contrato, publicado pela primeira vez em 1924. Em uma série de estudos etnográficos, o ensaio discute que, ao doar um objeto, o doador acaba criando

uma obrigação no receptor de devolver o presente em igual valor. As relações de troca, comuns desde sociedades mais arcaicas, acabam moldando nossa sociedade e relações com outros indivíduos.

Em [Rifkin 2000], por sua vez, novas formas de mercado são discutidas. No livro, o autor se aproxima na discussão do que temos hoje conhecido como Economia Colaborativa, em que as novas tecnologias têm cada vez mais eliminado os conceitos de propriedade, e a base das relações tem sido a troca de experiências e de bens. Em todo caso, há sempre troca.

Em busca de novos paradigmas de relações de troca entre usuários, o conceito de **coprodução** tem sido recentemente explorado na literatura de sistemas colaborativos [Carroll 2013]. Nesses casos, ofertante e demandante compartilham ao mesmo tempo de benefícios na execução de um serviço. O conceito de coprodução originou-se na observação de que a efetiva prestação de serviços sociais às vezes depende da participação ativa de quem recebe esses serviços. [Ostrom 1996] ilustra o exemplo dos crimes nas ruas de Chicago. Segundo suas análises, o crime aumenta quando os policiais fazem a vigia no interior das viaturas, evitando o contato direto com as pessoas das ruas. O policial na rua (ofertante) está em uma posição melhor para coproduzir a segurança pública com a participação ativa das pessoas na própria rua (demandante): a polícia e os moradores passam a se conhecer melhor e assim confiam mais uns nos outros, compartilhando fatos e opiniões que ajudam a entender e solucionar direta ou indiretamente os crimes.

Coprodução não é uma cooperação passiva e é mais do que uma simples participação da comunidade no planejamento de uma ação, mas sim uma ação conjunta, em que todos os envolvidos são beneficiados. Um exemplo institucional é o *Habitat for Humanity*, uma ONG internacional em que a construção de casas é feita através de trabalho voluntário, em que os próprios futuros moradores são mão-de-obra para este trabalho, além daqueles que já receberam auxílio [Carroll 2013].

Tendo em vista esses conceitos, este trabalho pretende investigar as seguintes questões durante o desenvolvimento de um aplicativo de caronas para uma universidade: (1) Existe possibilidade efetiva de coprodução para o contexto investigado (carona, ambiente acadêmico, comunidade fechada)?; (2) Quais as motivações do motorista ao ofertar uma carona, em uma universidade, o que ele espera em troca?; (3) Existem recompensas, além da relação de coprodução, que podem ser consideradas em um aplicativo de carona em uma universidade? Quais?; (4) Quais as preocupações de segurança do motorista (a quais caroneiros ofertar, quais informações prefere conhecer)? e (5) Quais as preocupações de segurança do caroneiro?

2.2. Aplicações de carona em outras universidades

Na própria literatura de sistemas colaborativos no Brasil, alguns processos de desenvolvimento de aplicações de carona já foram discutidos. Em [Vieira et al. 2012], a partir de um questionário divulgado em redes sociais, não direcionado especificamente a uma comunidade, os autores discutiram as motivações de possíveis usuários com relação aos aplicativos de caronas. A partir dos resultados do questionário, eles identificaram que as pessoas estão mais aptas a compartilhar caronas com pessoas mais próximas (amigos e parentes: 90%) e de grupos específicos (mesma empresa, mesma faculdade (80%), mesmo curso (61%)) do que para desconhecidos (6%) ou para pessoas que conhecem

apenas no meio on-line (20%). Para amigos de amigos ou vizinhos, a taxa também é alta, de 55%. Esses dados reforçam a importância da delimitação do uso do aplicativo de caronas desenvolvido a comunidades, uma vez que as pessoas se sentem mais confortáveis em compartilhar bens e serviços com outras que são, de alguma forma, próximas.

Além disso, os autores também investigaram as relações entre os tipos de recompensas e a motivação dos usuários para compartilhar caronas. A maior parte das respostas (47%) foi para a opção “descontos em serviços”, porém, por sua vez, a opção “nenhuma recompensa me motivaria” foi a segunda mais marcada, com 34%. Esse alto índice para a segunda opção pode ter duas explicações: (1) as pessoas não se sentem realmente motivadas de nenhuma forma a usar aplicações de carona, por questões de confiança ou de preferência; ou (2) os usuários poderiam ter outros tipos de motivações para o uso, não buscando necessariamente recompensas diretas e físicas. É com o propósito de investigar as possibilidades da opção 2 que partimos para uma abordagem de questionário direcionada aos alunos da UFMG, conforme discutido na próxima seção. Afinal, quais outras motivações os alunos de uma universidade poderiam ter para usar aplicações de carona, se não puderem receber uma recompensa financeira em troca?

Por fim, em [Macedo et al. 2015] é realizada uma investigação qualitativa com estudantes de uma universidade, que são questionados sobre uma aplicação de caronas para o ambiente acadêmico. A partir de entrevistas com estudantes, os autores chegaram às seguintes considerações, que levamos em conta durante a proposta da aplicação aqui discutida: (1) reputação no sistema e matrícula na universidade são vista com bons olhos para atrair confiança de desconhecidos; (2) motoristas não foram receptivos à ideia de serem mal avaliados por um caroneiro, devido ao caráter voluntário da carona; (3) caroneiros e motoristas preferem, quando houver, avaliações anônimas; (4) participantes preferem não assumir compromissos a longo prazo, na hora de agendar caronas, uma vez que podem haver imprevistos; (5) motoristas e caroneiros não se sentiram à vontade com a ideia de pagamentos, sendo vistos como uma descaracterização do favor; e (6) mulheres se mostram mais vulneráveis na questão de segurança. Todas essas considerações são discutidas nesse artigo em tempo de projeto, na Seção 4, que descreve as decisões tomadas relativas a esses pontos.

3. Questionário on-line: Considerações do contexto da comunidade

Embora pesquisas tenham sido e estejam sendo realizadas em outras universidades brasileiras investigando e desenvolvendo aplicativos de carona (conforme exemplificado na Seção 2.2), decidimos aqui realizar uma investigação direcionada ao ambiente da nossa própria universidade, a Universidade Federal de Minas Gerais (UFMG), durante a fase de desenvolvimento de um novo aplicativo (decisões de projeto estão apresentadas em detalhes na Seção 4). Nesse caso, realizamos um questionário e o distribuímos a membros da comunidade acadêmica, cujos resultados serviram como material para dois direcionamentos: (1) entender as particularidades desta comunidade no qual uma nova aplicação está sendo desenvolvida e que não teriam como ser levantadas em pesquisas anteriores em outras universidades; e (2) investigar a influência de uma comunidade – no caso, a acadêmica – em uma aplicação – no caso, um aplicativo de caronas para dispositivos móveis. Nessa seção, apresentamos os resultados relacionados aos dois direcionamentos, mas mantendo como foco a discussão dos fatores de 1 que influenciam em 2. Ou seja, queremos aqui apresentar o quanto as considerações de uma comunidade podem alterar

ou não as decisões de projeto de uma aplicação, tendo a nossa universidade como estudo de caso.

O questionário teve sua distribuição de forma on-line (a partir do Google Forms⁴) e abordou questões relacionadas aos seguintes temas: uso ou não uso de carro para se locomover na universidade, rotina e motivos; informações sobre caronas ofertadas dentro do campus; informações sobre caronas solicitadas; motivos para dar ou pegar caronas, ou motivos para não dar ou pegar caronas; e uso de aplicações móveis. O questionário foi divulgado entre os meses de dezembro e março de 2016, sendo respondido por 290 pessoas, tendo 287 respostas válidas (3 pessoas não eram membros da UFMG, o questionário possuía uma pergunta que evitava a participação nesses casos).

O questionário foi divulgado em grupos de discussão no Facebook⁵. Na rede social em questão, é possível encontrar grupos direcionados a comunidades e diferentes assuntos dentro dessas comunidades, como é o caso de grupos específicos de alunos da UFMG e grupos direcionados a caronas dentro da própria universidade. No caso, realizamos a divulgação do questionário nos seguintes grupos (em um total de 20): (1) grupos direcionados a caronas dentro da UFMG – nesses grupos, encontraríamos membros que já estão dispostos e abertos à ideia de caronas na universidade (4 grupos; o maior com 6.000 membros e média de 2.500, em março de 2016); (2) grupos que levam o nome da universidade – a fim de divulgar o questionário a um maior número de pessoas (3 grupos; o maior com 14.000 membros e os outros dois com média de 3.000, em março de 2016); e (3) grupos direcionados a diferentes cursos dentro da universidade – com o propósito de obter um conjunto mais heterogêneo de participantes (13 grupos, o maior com 5.000 membros, média de 2.000).

Em relação ao perfil atingido pelo questionário a partir da estratégia de divulgação adotada, 78% foram alunos de graduação, 15,6% alunos de pós-graduação e os demais 6% funcionários ou alunos de cursos técnicos. Sobre a faixa etária, 59,9% têm entre 18 e 23 anos, 31% estão entre 23 e 28 anos, 8,1% os demais (7,1% acima de 28 anos e 1% abaixo de 18). Quanto ao gênero, 51,6% marcaram feminino, 47,4% masculino e 1% preferiu não informar.

Divulgar o questionário nesses grupos fez com que o perfil dos participantes fosse de pessoas mais propensas a oferecer caronas naturalmente e pessoas que, de certa forma, estão interessadas em se entrosar com o público da própria universidade. Dentre os participantes, 108 disseram que vão de carro para a universidade (38% do total), 84 responderam que dão ou já deram carona para alguém dentro do campus. Ou seja, 78% dos motoristas interessados em participar da pesquisa estão abertos à ideia de oferecer caronas. De todos os 287 participantes da pesquisa, 185 (64%) já pegaram carona em algum momento. Dentre os participantes que afirmaram não ter carro ou não ir regularmente de carro para o campus, 68% disseram já ter pego carona. Ou seja, a maioria das pessoas que se sentiram motivadas a preencher o questionário on-line se interessam em oferecer caronas (78% dos motoristas) ou demonstram interesse em pegar caronas (64% dos não

⁴<https://www.google.com/forms/about/>, último acesso em maio de 2016.

⁵Segundo dados de 2014, o Facebook é acessado por 45% dos brasileiros (conforme divulgado em <https://www.facebook.com/business/news/BR-45-da-populacao-brasileira-acessa-o-Facebook-pelo-menos-uma-vez-ao-mes>, último acesso em maio de 2016).

motoristas).

Devido ao direcionamento gerado pela divulgação do questionário e à limitação de alcance via rede social, os números aqui apresentados não constituem uma amostra representativa dos estudantes da universidade como um todo. A UFMG possui atualmente mais de 33 mil alunos de graduação, 14 mil de pós e 1,6 mil de educação profissionalizante⁶. Assim, os participantes do questionário representam apenas 0.6% dos alunos da UFMG, sem levar em consideração professores e funcionários. Ainda assim, este grupo é de interesse uma vez que representam o perfil dos potenciais usuários a quem o aplicativo se destina, isto é: motoristas e não motoristas interessados no assunto e participantes de redes sociais. A seguir, apresentamos os principais tópicos obtidos a partir da análise do questionário.

Propensão para caronas dentro da universidade: Dentre os participantes que vão regularmente de carro para a universidade (81 pessoas, 28% do total de participantes; excluindo aqueles que vão apenas algumas vezes), 81,5% informaram que já deram ou costumam dar caronas para pessoas dentro do próprio campus. Quando questionados sobre a frequência dessas caronas oferecidas, a maioria desses motoristas ou dão caronas na maior parte da semana ou pelo menos uma vez por semana (60% marcou pelo menos uma vez por semana ou mais) ou então dão caronas muito esporadicamente (40% marcou que dá carona pelo menos algumas vezes ao mês ou menos). Ou seja, já existe um movimento por parte de alguns alunos motoristas para a oferta de caronas dentro do campus. E para aqueles que ofertam, essa oferta pode ser frequente e contínua.

Ambiente acadêmico como fator de confiança para oferecer caronas: Foram apresentados aos participantes cinco cenários nos quais eles deveriam responder, em uma escala Likert de 1 (discordo fortemente) a 5 (concordo fortemente), a propensão para oferecer caronas para cada uma das situações apresentadas. Os cenários variavam de situações de maior exposição (cenário 1: dar caronas a desconhecidos em qualquer lugar na cidade) a situações muito restritas ao ambiente acadêmico (cenário 5: dar caronas a conhecidos dentro do campus da universidade). Os cenários estão ilustrados na Figura 1.

A Figura 1 sugere que a disposição para dar caronas aumenta à medida em que o motorista conhece o caroneiro e se encontra em um local de confiança, também já conhecido, como é o caso do campus de uma universidade (confiança em pessoa conhecida > pessoa desconhecida, porém sabe que é membro da universidade > pessoa desconhecida, porém está no ambiente do campus > pessoa desconhecida, fora do campus). O uso de um aplicativo de caronas, por sua vez, melhora um pouco a confiança: o cenário 2 apresenta resultados melhores que o cenário 1. No entanto, o cenário 2 não possui confiança maior que os cenários 3, 4 e 5, em que o caroneiro se encontra no ambiente físico do campus. Tanto no cenário 2 quanto no 3 o número de pessoas que discorda da afirmativa (em vermelho na Figura 1), que não daria carona, é de 20 pessoas (23,8%). Porém, no cenário 3 mais pessoas estariam dispostas a dar caronas que no 2 – em azul, 53 pessoas (63,1%) contra 43 pessoas (51,1%). No cenário 4, em que o vínculo com a comunidade acadêmica é garantido, a disposição para a carona aumenta mais (61 pessoas, 70,6%), ainda que a pessoa seja desconhecida.

⁶Segundo levantamento realizado pela UFMG em novembro de 2015 (https://www.ufmg.br/conheca/nu_index.shtml, último acesso em maio de 2016).

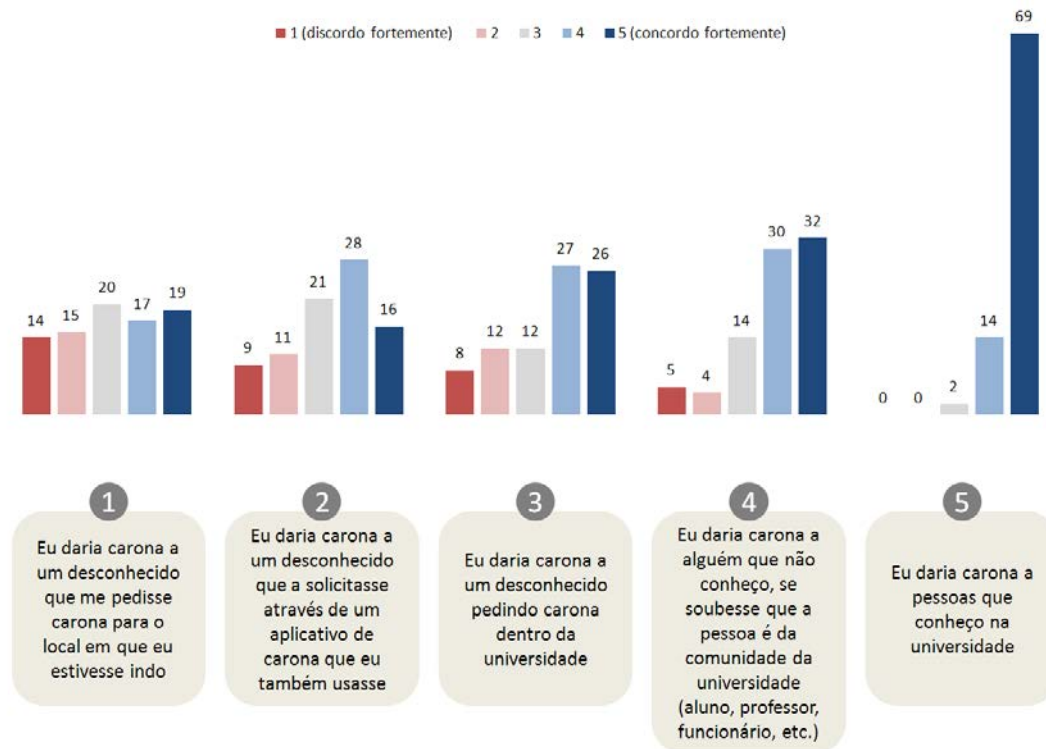


Figura 1. Escala Likert para cenários em situações de carona dentro da universidade: motoristas estão mais dispostos a dar caronas para pessoas de dentro da universidade e conhecidos (cenário 5) do que para desconhecidos em ambientes fora da universidade (cenário 1).

Preocupações com segurança: Das 102 pessoas que afirmaram não pegar caronas, quando questionadas sobre os motivos para isso, 39 marcaram que tinham medo de sofrer alguma violência (38%). Dessas, 33 são mulheres (85%), enquanto apenas 5 são homens (uma pessoa não quis informar o seu gênero). Essa informação reforça o que já foi sugerido de forma qualitativa no trabalho [Macedo et al. 2015], sobre uma preocupação maior por parte das mulheres com a segurança nesse tipo de aplicação. Já das 108 pessoas que informaram que vão de carro para o campus (38% de todos os participantes), 24 disseram nunca ter dado carona a alguém dentro do campus (22% dos motoristas). Dessas 24 pessoas, quando questionadas sobre os motivos para não dar caronas, 12 pessoas (50%) responderam ter medo de sofrer qualquer tipo de violência, e o número entre homens e mulheres nesse caso ficou equilibrado (7 mulheres e 5 homens).

Recompensas através de uma aplicação: Um grande desafio no projeto de sistemas colaborativos de carona é oferecer recompensas para o motorista. Por ser de caráter voluntário, é esperado que tais sistemas sejam mais vantajosos para o caroneiro do que para o motorista, já que este pode não ter muito a ganhar em troca, ao contrário do caroneiro, que ganha a viagem. Assim, investigamos quais as possíveis motivações dos participantes, em especial dos motoristas. Essa investigação foi feita com os participantes que responderam possuir carro e ir regularmente com ele para a universidade ($n = 81$). Os resultados são apresentados na Figura 2.

A recompensa de ajudar um amigo ou conhecido foi a mais votada, sendo marcada por 90,1% dos motoristas. As recompensas de melhorar o trânsito na cidade e contribuir para

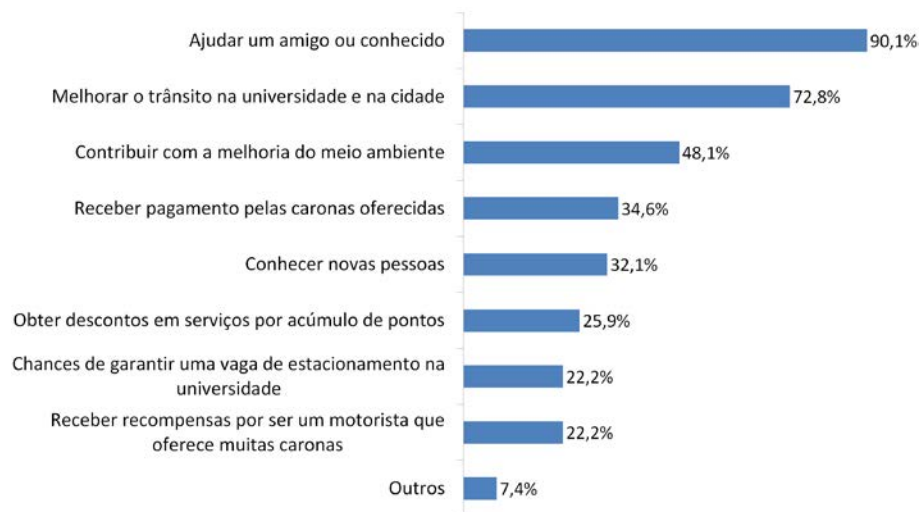


Figura 2. Respostas para a pergunta do questionário “Considerando o uso de um aplicativo de carona para celular, quais das motivações são mais relevantes para você considerar dar carona a alguém?”; essa pergunta foi respondida por 81 motoristas, que podiam marcar mais de uma questão.

o meio ambiente foram a segunda e terceira recompensas mais citadas, respectivamente. A recompensa material mais bem colocada, que consiste em receber um pagamento pelas caronas oferecidas, ficou em quarto lugar, sendo considerada uma motivação relevante por 34,6% dos participantes.

Para aplicações voltadas a comunidades, as recompensas também podem ser de valor local ou mesmo valor social, e não necessariamente valores financeiros. Das respostas apresentadas na Figura 2, a grande maioria dos motoristas tem hoje como principais motivações as recompensas de valor social. No caso da UFMG, uma necessidade que investigamos com os participantes do questionário para uma possível recompensa de valor local foi a questão de estacionamento no campus. Embora “chances de garantir uma vaga de estacionamento” tenha sido a penúltima opção indicada como motivação para usar um aplicativo, parte dos motoristas relataram alguma dificuldade para encontrar vagas de estacionamento no campus. Dos motoristas, 44% relatou dificuldade razoável e 35% difícil ou muito difícil. Pelos dados do questionário, as recompensas de valor social inicialmente são mais privilegiadas por parte dos motoristas da comunidade em questão do que as recompensas de valor local que inicialmente propomos.

4. Carona UFMG: Aplicação móvel e decisões de projeto

Nesta seção discutimos as principais funcionalidades implementadas no aplicativo (app) e as decisões de projeto adotadas. As telas iniciais do aplicativo (login, menu com principais funcionalidades e tela de perfil) estão ilustradas na Figura 4.

Principais funcionalidades: Na aplicação aqui desenvolvida, os motoristas podem divulgar suas rotas com pontos de partida, intermediários e de chegada, além de data e horário (Figura 4.a). Esses roteiros podem ser vistos por outros usuários que estão procurando por transporte, os caroneiros (Figura 4.c). Uma vez identificada uma rota, o caroneiro faz a requisição de carona para o motorista que publicou a rota, que pode ou não aceitar a solicitação. O aplicativo também permite que usuários que não possuem carro divulguem

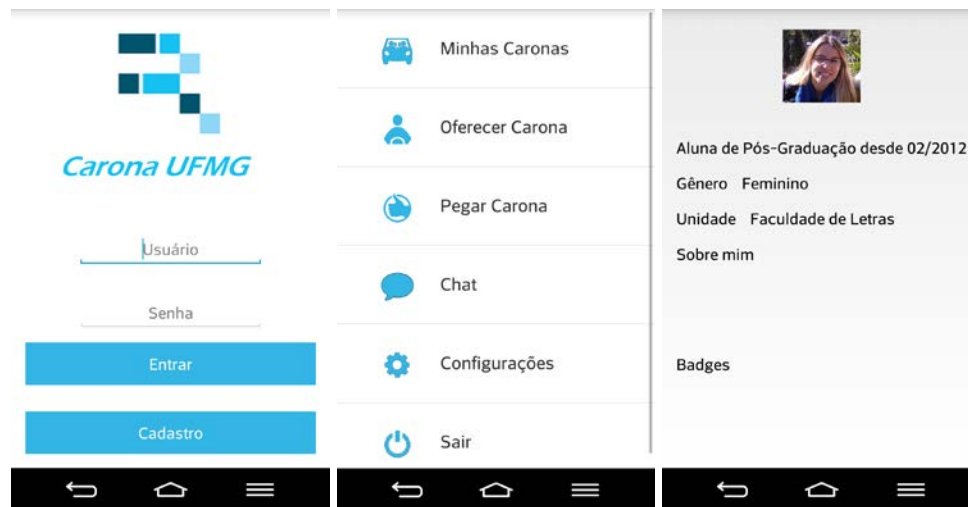


Figura 3. (a) Tela inicial de login (e-mail institucional é exigido); (b) menu de opções do app e (c) tela de perfil de um usuário.

os seus trajetos. Um motorista, ao identificar um trajeto similar ao seu, pode oferecer carona a alguém que precise, ficando a cargo do caroneiro aceitar ou não a carona (Figura 4.b). O aplicativo também conta com um sistema de *bate-papo*, por meio do qual eles podem acertar os detalhes da carona, como combinar o ponto de encontro ideal ou avisar previamente sobre atrasos, por exemplo. Além disso, foi desenvolvido um processo de autenticação para garantir que o usuário é membro da comunidade da UFMG, a fim de aumentar a segurança dentro da rede de caroneiros e a confiança dos usuários ao utilizarem o sistema. Também foi proposto um sistema de reputação, que permite que cada usuário seja avaliado com base em notas dadas por outros usuários e que, quando combinadas, formam uma métrica da reputação do usuário dentro da rede de compartilhamento de caronas. Essa reputação pode ser vista por outros usuários e utilizada como subsídio no momento de decidir se dará ou não carona para o referido usuário. Usuários com a reputação muito abaixo da média também podem ser banidos do sistema, com o intuito de manter a qualidade dos participantes da rede.

Aplicação Android: Uma vez que estamos propondo um sistema colaborativo, a adesão dos usuários é determinante para o sucesso do mesmo. Quanto mais usuários aderirem ao sistema, maior será o número de rotas disponíveis e de pessoas procurando por carona, aumentando, assim, o número de caronas agendadas [Teal 1987][Kaufman 2002]. Para a implementação de uma primeira versão do aplicativo, foi escolhida a plataforma Android, marcada como sendo utilizada por 73,2% das pessoas que preencheram o questionário, podendo ter mais chances, portanto, de adesão.

Arquitetura: O sistema proposto é baseado no modelo cliente-servidor. O objetivo desta divisão é separar a arquitetura e responsabilidades em dois ambientes. Foi implementado um *WebService* com o propósito de receber as requisições e intermediar as interações entre os usuários do aplicativo, cumprindo, assim, o papel de servidor. Este *WebService* foi baseado no modelo REST - *Representational State Transfer*, que possui as características de ser *stateless*, *cacheable* e ter uma interface uniforme. Essa arquitetura permite que a base de dados de caronas possa ser consumida por outros sistemas, facilitando o surgimento de novas soluções que possam tirar proveito desses dados.

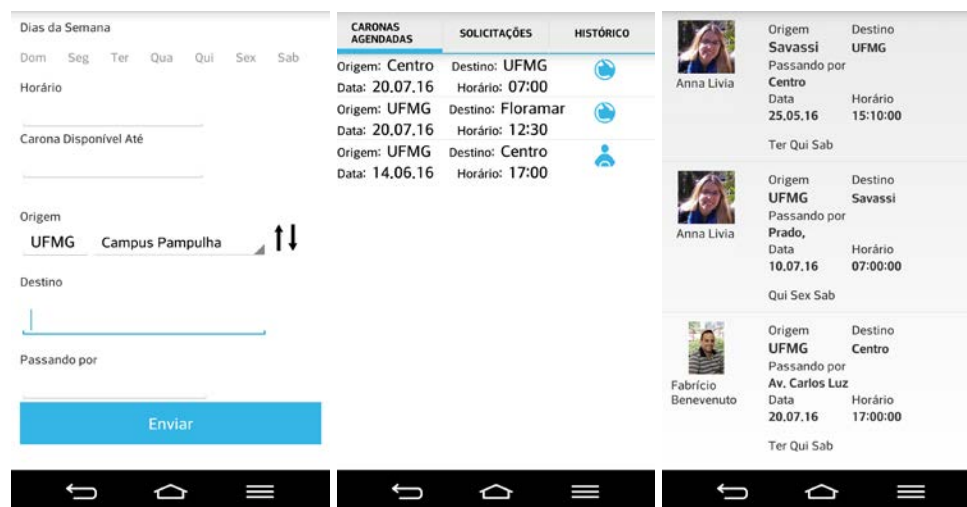


Figura 4. (a) Tela para o motorista cadastrar uma carona; (b) Tela de caronas agendadas, cadastradas tanto por caroneiros, quanto por motoristas e (c) Lista de opções de caronas para serem requeridas, em que é possível visualizar a foto de perfil e outras informações do usuário.

Autenticação: A solução proposta consiste na utilização do e-mail institucional da universidade no processo de cadastro de usuários. Sempre que um usuário se inscreve no sistema de caronas, ele é requisitado a informar, obrigatoriamente, um e-mail do domínio da universidade. O sistema, em seguida, envia para o endereço de e-mail informado um link único que deve ser acessado pelo usuário a fim de validar o e-mail utilizado no momento do cadastro. Enquanto não é feito esse processo de validação, a conta do usuário permanece inativa e o usuário não consegue fazer login no app. O que garante que apenas membros da comunidade acadêmica consigam acessar o sistema.

Cadastro de Rotas: Como o sistema de caronas tem o intuito de atender apenas à comunidade universitária, o aplicativo só permite o cadastro de rotas que tenham a UFMG como ponto inicial ou final. O aplicativo também oferece uma listagem dos prédios da instituição. O sistema só permite que os pontos de origem e destino escolhidos fora do campus da universidade sejam indicados a partir de uma listagem pré-definida de bairros das cidades da Região Metropolitana de Belo Horizonte. Isso faz com que não seja possível que endereços específicos sejam incluídos no sistema e fiquem disponíveis para todos os outros usuários, a fim de manter a privacidade dos mesmos. Uma vez solicitada uma carona, o caroneiro e o motorista podem utilizar o bate-papo a fim de combinarem os locais específicos de início e fim da carona.

Caronas totais x caronas parciais: Para que uma carona possa ser combinada com sucesso, não é necessário que os dois passageiros tenham a mesma origem e o mesmo destino. Um pessoa pode, por exemplo, passar pelo bairro de outra a caminho da sua própria casa. Nesse caso é possível que seja dada uma carona parcial, na qual o caroneiro não acompanha o motorista por todo o seu trajeto. Por esse motivo, o aplicativo permite que, ao cadastrar uma rota, o usuário também cadastre pontos intermediários pelos quais passa no seu trajeto de/para a universidade.

Caronas únicas x Caronas recorrentes: Um dado motorista pode percorrer uma rota uma única vez ou então percorrê-la com certa periodicidade. Da mesma forma, uma

pessoa que procura por caronas pode precisar percorrer o trajeto uma única vez ou com certa frequência. Por esse razão, as rotas podem ser cadastradas pelo usuário como únicas ou recorrentes. As caronas únicas possuem datas definidas e são exibidas no sistema do momento do seu cadastro até a data da carona. Já as recorrentes são exibidas desde sua inclusão no sistema até uma data final escolhida pelo usuário no momento do cadastro.

Informações de perfil: O perfil de cada pessoa apresenta dados pessoais tais como nome, foto, data de ingresso e vínculo com a universidade. As informações apresentadas são preenchidas pelo próprio usuário no momento do cadastro no sistema. Também é disponibilizado um campo que pode ser preenchido com texto livre.

Reputação: Após uma carona, a pessoa que ofereceu carona pode avaliar, com uma nota de 1 a 5 estrelas, o passageiro e vice-versa. A partir dessas avaliações, cada usuário do sistema terá a sua própria nota de reputação, calculada levando em conta as suas avaliações. A literatura mostra que os usuários desse tipo de sistema têm aversão à ideia de ter avaliações negativas publicadas. Essa aversão é ainda maior quando se trata dos motoristas uma vez que estes veem o ato de oferecerem carona como um favor. Na literatura também é citado que as avaliações devem ser, preferencialmente, anônimas [Macedo et al. 2015]. Por esse motivo, propomos um sistema de reputação que pretende utilizar as avaliações de forma anônima para o cálculo das reputações e ao mesmo tempo preservar os usuários com notas ruins. A solução proposta na aplicação faz uso, portanto, do conceito de *gamefication* chamado de *badges* ou conquistas para identificar os usuários que se destacam de maneira positiva em relação aos outros. Um usuário ganha uma *badge* sempre que este conseguir completar uma realização notável dentro do sistema. Essas conquistas são apresentadas nos perfis de cada usuário que as conquistou e servem de subsídios para outros usuários decidirem se devem ou não dar caronas. Esse recurso também serve como forma de incentivo para os usuários, que podem se sentir reconhecidos ao recebê-las e motivados a conquistar outras [Grant and Betts 2013]. Dessa forma, os usuários nunca têm acesso diretamente às notas que receberam em suas avaliações, mas conseguem ter um feedback do seu desempenho por meio das *badges*. Já os usuários que se destacarem negativamente dentro do sistema, acumulando avaliações ruins, nunca serão expostos mas podem ser banidos do sistema por parte dos administradores, o que objetiva manter certo nível de qualidade dos usuários da plataforma.

Bate-papo: O sistema também conta com um serviço de bate-papo para que motoristas e passageiros possam conversar livremente. Este foi criado para permitir que os usuários possam combinar de forma detalhada a carona, como escolher um ponto de encontro exato ou avisar sobre um possível atraso. Os usuários também podem usar o bate-papo para a troca de informações pessoais como telefones e endereços. O bate-papo conta, ainda, com uma funcionalidade que permite que um usuário abra o mapa da cidade, marque um ponto e envie para o destinatário, que por sua vez consegue abrir o mapa e ver qual ponto foi marcado pelo remetente. O bate-papo entre dois usuários só é possível depois que um tiver enviado uma solicitação de carona para o outro. Uma vez que a carona foi dada, esse meio de comunicação entre eles deixa de existir.

Tabela 1. Diferenças entre a aplicação de caronas desenvolvida na universidade em relação a aplicações tradicionais: vantagens e considerações de projeto

	Carona UFMG	Outros apps tradicionais de carona (fora do contexto de uma comunidade; exs.: BlaBlaCar, Zumpy)
Autenticação	É feita via e-mail institucional; somente membros da comunidade acadêmica tem acesso. Evita o surgimento de usuários <i>fakes</i> .	Feita por meio de cadastro próprio ou por meio de logins via redes sociais; aberta a qualquer usuário
Banco de Dados	É armazenado e fica disponível para o uso da própria instituição; os dados podem ser utilizados para entender e melhorar a mobilidade na instituição	É armazenado por uma empresa e somente esta tem acesso às informações
Integração	Permite a integração com outros sistemas da instituição	Pode ter integração com outras empresas (como Facebook e convênios com demais empresas)
Customização	Customizável para cada instituição; é possível realizar a alteração das informações contidas no perfil, nome das unidades e domínios dos e-mails utilizados para a autenticação, entre outros	Não são customizáveis, domínio de uma empresa
Recompensas	A fim de aumentar a adesão ao sistema, permite a utilização de recompensas locais (como vagas na universidade)	Recompensas como desconto em gasolinas, ingressos de cinema, etc.; recompensas financeiras

5. Decisões de Projeto: Motivações e Contrastes

Das funcionalidades da aplicação apresentadas na seção anterior, muitas se diferem das aplicações tradicionais de caronas (como a BlaBlaCar⁷ e o Zumpy⁸), que são abertas a qualquer público. Na Tabela 1, apresentamos as principais diferenças ao se comparar esses tipos de aplicativos com o aqui apresentado, direcionado a uma comunidade. Essa tabela pretende ressaltar algumas das vantagens que podem surgir ao se ter uma aplicação direcionada a um contexto e as decisões de projeto aqui definidas.

Em relação ao que foi discutido nas seções anteriores, resumimos nas Tabelas 2 e 3 os principais resultados, com o propósito de fazer uma breve apresentação do que foi considerado no projeto da aplicação e ressaltar pontos importantes a serem considerados em aplicações voltadas a comunidades.

6. Considerações Finais

Nesse artigo, apresentamos o processo de desenvolvimento de um sistema para o compartilhamento de caronas dentro de uma comunidade acadêmica. Uma pesquisa exploratória foi realizada a fim de identificar o estado atual da utilização de caronas na UFMG. Foi identificada a existência de uma cultura de compartilhamento de caronas e um potencial de expansão da utilização destas pelos membros da comunidade. A preocupação com a segurança pessoal e a dificuldade em encontrar outras pessoas que façam o mesmo trajeto foram citadas como as principais barreiras para que ocorra essa expansão. O artigo

⁷<https://www.blablacar.com.br/>

⁸<http://www.zumpy.com.br/>

Tabela 2. Considerações sobre aplicações de carona encontradas em outras publicações no SBSC e levantadas no questionário on-line aqui aplicado com membros da comunidade acadêmica

Considerações	Fonte	Impacto na especificação do sistema
Coprodução: em uma aplicação semelhante a aplicações de economia colaborativa, porém voltadas a comunidades e de caráter voluntário, o conceito de coprodução pode ser interessante de ser explorado (no caso de caronas: encontrar formas para que motoristas e caroneiros sejam mútua e instantaneamente beneficiados)	Conceito de coprodução explorado em [Carroll, 2013]	Motoristas ganham conquistas (<i>badges</i>) podendo ganhar vantagens e vagas na universidade, caroneiros ganham a carona. Coprodução não está sendo diretamente considerada, ambos não ganham vantagens instantaneamente, porém ao longo de um tempo. Pelo questionário, recompensas de valor social são as mais importantes, portanto pode haver coprodução nesse sentido: motorista ganha a companhia de um amigo, no caso, o caroneiro.
Usuários estão propensos a interagir com pessoas conhecidas mais do que desconhecidas	[Vieira et al., 2012]; Questionário	No caso, a aplicação ainda não prioriza automaticamente conhecidos de desconhecidos, ou amigos próximos.
Usuários se sentem mais seguros em comunidades e com as pessoas dessas comunidades locais	[Macedo et al., 2015]; Questionário	A aplicação exige e-mail institucional no processo de cadastro. O que garante que apenas membros da universidade tenham acesso ao sistema. Isso também evita o surgimento de usuários <i>fakes</i> .
Recompensa social, em uma comunidade, pode ser mais bem vista do que qualquer outra	Questionário	O próprio uso de um aplicativo voltado a uma comunidade específica viabiliza essa consideração. Ranking de top usuários também promove esse reconhecimento. Esse ranking já pode ser feito a partir do banco de dados do aplicativo e é a partir deste que espera-se distribuir as vagas reservadas.
Recompensa local também pode ser um atrativo a ser considerado	[Vieira et al., 2012]; Questionário	A partir do engajamento de cada usuário no sistema e das avaliações feitas por estes depois de cada carona, é gerado um ranking de usuários. Esse ranking pode ser utilizado para dar recompensas aos usuários mais bem classificados. No caso específico da UFMG, propomos que algumas vagas de estacionamento sejam reservadas a estes motoristas.
Motoristas não gostam de ser mal avaliados, pois veem a oferta de carona como um favor sendo feito	[Macedo et al., 2015]	Apesar de o sistema permitir que os usuários sejam avaliados entre si, as notas não são exibidas na aplicação. Essas avaliações são refletidas no aplicativo por meio de conquistas (<i>badges</i>). Sempre que um usuário conseguir um determinado número de notas máximas nas avaliações, por exemplo, ganha uma badge. As <i>badges</i> de um determinado usuário são exibidas em seu perfil público, o que serve de subsídio para que outras pessoas decidam pegar ou dar carona para aquele usuário. Essas <i>badges</i> só são concedidas para boas conquistas, reconhecendo e incentivando o bom comportamento dos usuários e não expondo os usuários com médias ruins nas avaliações.

Tabela 3. Considerações sobre aplicações de carona encontradas em outras publicações no SBSC e levantadas no questionário on-line aqui aplicado com membros da comunidade acadêmica (continuação)

Considerações	Fonte	Impacto na especificação do sistema
Preferência para agendamentos de caronas em prazos curtos	[Macedo et al., 2015]	O sistema lista as caronas disponíveis considerando o prazo restante para o horário da carona. Caronas para daqui a uma hora são exibidas antes de caronas para daqui a cinco horas na listagem de caronas disponíveis, por exemplo. Caronas para hoje são exibidas antes de caronas para amanhã.
Permitir informar horários contínuos e frequentes	Questionário	Motoristas e caroneiros podem dispor de horários contínuos (vários horários durante uma semana ou no mês), sem precisar informar o mesmo horário sempre.

também descreveu alguns detalhes da arquitetura do sistema e suas principais funcionalidades. As decisões de projeto foram discutidas e baseadas em trabalhos descritos na literatura e em um questionário realizado com membros da universidade. Embora aqui tenhamos apresentado algumas decisões do desenvolvimento do aplicativo, uma versão já se encontra pronta e disponível para a comunidade acadêmica da UFMG por meio da Google Play, plataforma de distribuição de aplicativos da plataforma Android. O aplicativo pode ser encontrado no link: <http://www.carona.dcc.ufmg.br/>.

Na discussão sobre o conceito de coprodução e recompensas, em que ambos (motorista e caroneiro) possam ser beneficiados por uma aplicação, identificamos que há possibilidades de benefícios para ambos. Porém, para o motorista os benefícios sociais apresentaram-se mais relevantes do que os financeiros e até mesmo os locais. Pode ser que pela cultura de caronas já existente na universidade, os motoristas estejam mais dispostos a oferecer caronas não visando recompensas, mas sim ajudar os outros membros da comunidade. Novas formas de recompensas locais ainda devem ser discutidas na comunidade, como foi o caso aqui da garantia de vagas de estacionamento, apontadas como alvo, devido a dificuldade de encontrar vagas atualmente no campus.

As preocupações de segurança também foram aqui investigadas e medidas foram tomadas durante o desenvolvimento da aplicação apresentada, como a exigência de um e-mail institucional para o acesso ao sistema e a necessidade parcial de algumas informações, como a foto de perfil, gênero e informações acadêmicas (curso, unidade acadêmica, vínculo com a universidade). Ainda que não obrigatórias (um usuário pode não postar sua foto de perfil, por exemplo), essas informações são fortemente sugeridas pela interface do sistema.

Como próximos passos, pretendemos realizar entrevistas direcionadas a motoristas e caroneiros sobre algumas decisões de projeto discutidas aqui, como a adesão do sistema de *badges* e avaliação de outros usuários. Também pretendemos realizar um estudo do comportamento dos usuários a partir dos *logs* de utilização do sistema e a disponibilização desses *logs* anonimizados para a utilização da comunidade científica. Novas funcionalidades serão incluídas na ferramenta como recomendação de caronas baseada nos históricos de utilização de cada usuário e um sistema anti-fraudes, a fim de garantir que apenas os usuários que realmente deram caronas recebam os incentivos ofe-

recidos aos usuários melhor avaliados pelo sistema.

7. Agradecimentos

Agradecemos aos participantes do questionário por sua colaboração; e às agências CNPq, CAPES e FAPEMIG, bem como ao InWeb (MCT/CNPq 573871/2008- 6) e MASWeb (FAPEMIG/PRONEX APQ-01400-14) por seu apoio financeiro parcial a esta pesquisa.

Referências

- Carroll, J. M. (2013). Co-production scenarios for mobile time banking. In *End-User Development - 4th International Symposium, IS-EUD 2013, Copenhagen, Denmark, June 10-13, 2013. Proceedings*, pages 137–152.
- Carroll, J. M. and Bellotti, V. (2015). Creating value together: The emerging design space of peer-to-peer currency and exchange. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15*, pages 1500–1510, New York, NY, USA. ACM.
- Grant, S. and Betts, B. (2013). Encouraging user behaviour with achievements: An empirical study. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, pages 65–68.
- Hamari, J., Sjöklint, M., and Ukkonen, A. (2015). The sharing economy: Why people participate in collaborative consumption. *American Society for Information Science and Technology. Journal*. Epub ahead of print. Published online: 2 June 2015.
- Kaufman, S. (2002). Why people (dont) carpool and change for the better: A social capital framework for investigating environmental behaviour. In *2nd National Conference of Sustainable Campuses, RMIT*.
- Macedo, R. d. F., Gomes, A. L., Andrade, N., and Maia, D. C. (2015). Caronas à brasileira: uma investigação centrada no usuário sobre o projeto de sistemas de caronas em uma universidade brasileira. In *Simpósio Brasileiro de Sistemas Colaborativos - SBSC 2015*.
- Mauss, M. (1974). *Ensaio sobre a Dádiva*. EPU, São Paulo.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. <http://bitcoin.org/bitcoin.pdf>.
- Ostrom, E. (1996). Crossing the great divide: co-production, synergy, and development. *World Development*, pages 1073–1087.
- Rachel Botsman, R. R. (2010). *What's mine is yours: the rise of collaborative consumption*. HarperCollins Publisher.
- Rifkin, J. (2000). *The Age of Access: The New Culture of Hypercapitalism Where All of Life Is a Paid-For Experience*. Putnam Publishing Group, The.
- Teal, R. (1987). Carpooling: Who, how and why. *Transportation Research*, 21A:203–214.
- Vieira, V., Fialho, A., Martinez, V., Brito, J., Brito, L., and Duran, A. (2012). An exploratory study on the use of collaborative riding based on gamification as a support to public transportation. In *SBSC*, pages 84–93. IEEE Computer Society.

Processo de Design Colaborativo de Artefatos Tangíveis para Crianças

Isomar Lima, Thais Castro, Adriana Dantas, Marcos Siqueira

Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Av. Rodrigo Otávio, 3000 – 69077-000 – Manaus – AM – Brazil

{isomar.slima,dricadoroteu}@gmail.com, {thais,mpsf}@icomp.ufam.edu

Abstract. *This paper presents a collaborative design process, based on interaction design for tangible learning artefacts, with a description of a design case of an artefact to help on middle schools' maths learning, the Tangible Tangram. On this experience, we made appropriation of techniques based on the approach Human Centred Design (HCD) where we worked with two levels of users: the maths teacher and the student. Student's profile (end user) is investigated considering which tasks he performs and where the interaction is more critical. In this way, the intermediate prototypes are evaluated in order to enhance the interaction (metacommunication) between students with the artefact and among students aiming at solving problems.*

Resumo. *Este artigo apresenta um processo de design colaborativo, baseado no design de interação para artefatos tangíveis para aprendizagem, contendo a descrição de um caso de design de um artefato para auxiliar na aprendizagem da matemática, o Tangram Tangível. Nessa experiência, nos apropriamos de técnicas da abordagem de Design Centrado em Humanos (HCD), trabalhando com dois níveis de usuário: o professor de matemática e o aluno. O perfil do aluno é investigado considerando quais atividades ele faz e onde a interação é mais crítica. Sendo assim, os protótipos intermediários são avaliados proporcionando, um maior potencial de interação entre os alunos com o objeto e entre os alunos entre si.*

1. Introdução

Design de interação é um fator indispensável ao se desenvolver *software* interativos ou produtos [Oliveira, 2013] como: celular, videogame, impressora, guichê eletrônico, dentre outros. Define aspectos essenciais dos produtos interativos e questões durante os processos de design. Seu objetivo é redirecionar essas questões, trazendo um olhar crítico sobre avaliação utilizando as perspectivas de uso e comunicação para dentro do processo de *design*. Isto possivelmente é o motivo pelo qual tem sido objeto de estudo de diversas pesquisas relatadas na literatura.

O foco da pesquisa e do estudo em *Design de Interação* são as relações humanas estabelecidas através dos artefatos interativos, que funcionam também como meios de comunicação interpessoal, além de aspectos centrais referentes ao uso real e como são utilizados para a avaliação de produtos interativos [Barbosa; Silva, 2010]. É importante, então, que durante todo o ciclo do projeto de *software* ou artefato tangível, se observe o comportamento e a intenção dos diversos níveis de usuários e da própria equipe de *design*, confrontando e revendo os objetivos.

Este artigo visa apresentar uma maneira de projetar artefatos tangíveis que o nosso grupo de pesquisa já vem utilizando ao longo de cinco anos. Essa maneira de projetar é inspirada no conceito de “*Human Centred Design*” (HCD), no qual os usuários dos artefatos são considerados como humanos, com desejos, mudanças de humor, variação na criatividade e atenção e etc. [Bevan, 1999]. Os métodos utilizados em HCD envolvem todas as pessoas da equipe na utilização do artefato, e não apenas o usuário final. Desses métodos, o que se aproxima mais com nossa maneira colaborativa de projetar artefatos tangíveis é o Design Participativo (DP) [Rosemberg et al, 2008].

Sendo assim, o processo que utilizamos é fortemente baseado no design participativo, mas em escala menor, com menos pessoas envolvidas. Desenvolvemos essa adaptação porque nossos projetos, em sua maioria, são voltados a crianças, com ou sem deficiência. Os projetos para crianças com deficiência ou inclusivos possuem especificidades próprias e constituem um subgrupo da maneira colaborativa de projetar artefatos para crianças.

Dessa forma, a adaptação do DP para o projeto de artefatos para crianças é discutida neste artigo com um cenário real de design de um artefato tangível, o Tangram, que consiste em um jogo com sete peças: um quadrado, um paralelogramo e cinco triângulos, formando um quebra-cabeça. A configuração geométrica de suas peças permite centenas de composições, tornando-o um material didático criativo no contexto de objetos concretos para auxiliar na aprendizagem de matemática.

É importante salientar que neste artigo utilizamos a palavra design no contexto de Interação Humano Computador (IHC), referindo-se ao processo inteiro de concepção e construção de algum artefato interativo, seja ele software ou hardware. Lawson [Lawson, 2006] define design como um processo em três etapas: análise da situação, síntese de uma solução (intervenção) e avaliação da nova situação.

Para que o design colaborativo descrito seja mais bem detalhado, esse trabalho está estruturado da seguinte forma: a Seção 2 apresenta uma descrição sobre a utilização de objetos concretos na Matemática e seu potencial como um objeto de aprendizagem tangível; a Seção 3 descreve a metodologia utilizada; a Seção 4 descreve a adaptação do DP para o processo de design colaborativo de artefatos para crianças utilizado; a discussão sobre os desdobramentos e implicações da proposta são discutidas na Seção 5.

2. Objetos Concretos utilizados na Matemática

Objetos concretos são utilizados no processo ensino-aprendizagem de matemática há muito tempo, embora haja novas propostas de utilização em sala de aula, conforme relatado nas pesquisas em [Guimarães e Rodrigues, 2015] e [Botas e Moreira, 2013]. Esses objetos visam diminuir a dificuldade dos alunos com operações e conceitos mais abstratos como frações, geometria plana, proporção e sistema de numeração decimal.

Algumas propostas como a de Guimarães e Rodrigues [2015] discutem as definições de objetos manipuláveis digitais a partir de materiais concretos já amplamente utilizados. Outros autores como [Einsenberg, 2003] discutem o conceito de artefatos tangíveis, em que esses artefatos representam uma oportunidade de criar sistemas físicos de modelagem computacionalmente aumentados, os quais contribuem tanto para os recursos digitais (editáveis) quanto no aspecto físico do modelo tangível. Entende-se por artefatos tangíveis, nesse contexto, interfaces ou objetos de interação do

usuário, onde o mesmo manipula objetos físicos, para que haja modificações no meio digital.

Ao manipular artefatos tangíveis a criança se beneficia de algumas vantagens: relacionam novos conhecimentos com os que já sabiam, fazem e testam hipóteses, pensam onde aplicar o que estão aprendendo, expressam-se por meio de várias linguagens, aprendem novos métodos, novos conceitos, aprendem a serem críticos sobre os limites de aplicação dos novos conhecimentos. Para que isso ocorra, é necessário que esses objetos apresentem potenciais para boas interações, como uma boa aparência, e estejam inseridos em um ambiente computacional interativo, onde ferramentas para comunicação e cooperação sejam oferecidas para as crianças terem acesso ao registro de sua interação e às trilhas de aprendizagem que percorrem.

Para exemplificar, Zuckerman et al [2005] aponta os modelos de sistemas interativos direcionados para área da matemática com a utilização de artefatos tangíveis e objetos físicos (manipuláveis) que acompanham os efeitos do movimento da tela de computador. Portanto, no contexto deste trabalho, utilizamos um processo próprio de design colaborativo de um objeto manipulável amplamente utilizado em escolas, o tangram, o qual é descrito abaixo juntamente com o seu objetivo.

2.1 Tangram

O Tangram é um antigo jogo chinês com sete peças: um quadrado, um paralelogramo e cinco triângulos (dois grandes, dois pequenos e um médio), formando um quebra-cabeças. A configuração geométrica de suas peças permite centenas de composições, tornando-o um material didático criativo. É útil para desenvolver o raciocínio lógico e geométrico (habilidades de visualização, percepção espacial e análise das figuras) e exercitar as relações espaciais e as estratégias de resolução de problemas [Diniz, 2002].

Nossa proposta de intervenção para o tangram é uma solução envolvendo a divisão em camadas, formando um artefato multicamadas. Uma camada é a base, onde as peças do tangram são unidas. Nela há o processamento das informações obtidas pela manipulação das peças. Outra camada é das peças manipuláveis pelas crianças onde, ao serem manipuladas, as mesmas fornecem feedback ao professor e aos próprios alunos, positivos ou negativos, informando se foram agrupadas corretamente. Por último, há a camada do *software* representada pelo App que se comunica com as peças manipuláveis por meio de *bluetooth*, gerando *logs* com o registro das manipulações dos alunos ao jogarem, fornecendo assim trilhas de aprendizagem que, posteriormente, podem ser utilizadas pelos professores para aferir a aprendizagem de cada aluno.

O Tangram Tangível pode ser utilizado em atividades individuais ou em pequenos grupos. Essas atividades são, basicamente, de dois tipos: encaixar todas as peças corretamente na moldura quadrada e montar as peças de forma a reproduzir figuras pré-definidas. Mas o professor pode incrementar essas atividades em conjunto com as funcionalidades do App, pedindo o cálculo da medida de algum lado da figura montada, a identificação de determinadas figuras escondidas nas demais, etc. A Figura 1 ilustra algumas construções básicas do Tangram, em especial quatro maneiras diferentes de montar uma casa com as peças.



Figura 1 - Modelos de Casa no Tangram. Fonte: Diniz, 2002.

No jogo Tangram é possível aprender conceitos geométricos de frações, lado correspondente, ângulo correspondente, formatos semelhantes, área e perímetro de figuras planas, proporcionalidade entre os lados dos triângulos, ângulos congruentes, polígonos congruentes e semelhantes formados pelas peças do jogo [Genova, 2005]. Esses conceitos fazem parte do conteúdo de geometria trabalhados tanto nas séries iniciais quanto nas finais do ensino fundamental, em profundidades diferentes.

3. Metodologia

Visando obter um olhar sobre os usuários, considerando seu próprio contexto, adotamos o design centrado no ser humano (HCD)[Maguire, 2001], onde se considera as mudanças de desejos, intenções e atenção inerentes a cada ser humano. Das diversas perspectivas adotadas em IHC ao longo dos anos, esta nos possibilitou um contato permanente com os usuários e *stakeholders*, como algumas crianças, professores de matemática e designers com perfis diferentes. Portanto, nossa equipe de design é composta por um professor de matemática do ensino fundamental, um aluno de graduação com experiência em desenvolvimento com arduino, duas crianças para os testes formativos e a professora orientadora e o próprio autor como designer principal e gerente do projeto. O professor de matemática, que mantém contato constante com os alunos do 6º. ano do Ensino Fundamental, testa algumas hipóteses de cenários de interação com o tangram tradicional com esses alunos, fornecendo *feedback* ao restante da equipe de design.

Como metodologia de projeto e desenvolvimento (design), adotamos o Design Participativo e fizemos pequenas adaptações em processos específicos. A adaptação principal se deu principalmente pela dificuldade de manter crianças como participantes da equipe de design. Como se trata de um objeto de aprendizagem, se as crianças fizessem parte do design estariam contaminadas pelo estilo de abordagem para geometria e acabariam se adaptando ao tangram tangível e não o contrário, pois em geral, elas são muito boas em perceber padrões em jogos.

Em termos de pesquisa, sua natureza é aplicada, uma vez que buscamos desenvolver o artefato tangível para utilizá-lo como auxiliar na aprendizagem de

ambiente e um organismo que, com um conjunto de estímulos, fazem o organismo interagir com o objeto. No caso do Tangram Tangível essas relações possibilitam ao usuário identificar as funcionalidade do Tangram sem a necessidade prévia de explicação já que por ser um jogo geométrico facilita a intuição. Uma vez definidas as entradas na etapa anterior, procede-se a avaliação formativa ainda nesses estágios iniciais, com especial atenção para as *affordances* e diferenças de *meaning making* ou processo de aquisição de sentido para os designers, professores e alunos.

4. Criação de protótipos interativos de alta fidelidade com *affordances* apropriadas – nessa fase os protótipos já estão com o hardware e *software* definidos para possibilitar uma melhor interação entre os elementos do artefato e entre as crianças.

5. Avaliação dos protótipos de alta fidelidade – procede-se a avaliação até que um teste de aceitabilidade forneça resultados positivos.

No processo descrito na Figura 2 observamos as etapas de desenvolvimento de protótipos tangíveis utilizadas por pesquisas na área (descritas acima) e adaptamos ao nosso contexto real de aplicação, que é o desenvolvimento de artefatos tangíveis para crianças caracterizados como objetos educacionais tangíveis no domínio da matemática.

Na faixa principal do processo que definimos, as tarefas envolvendo formação de equipe, prospecção e revisão da literatura correspondem à fase 1; as tarefas de escolha do artefato físico existente e planejamento de funções correspondem à fase 2; definimos como avaliação formativa a tarefa de selecionar e aplicar métodos de avaliação da interação conhecidos, tanto para protótipos de baixa quanto de alta fidelidade, correspondendo de forma incremental às fases 3 e 5; e consulta aos logs existentes em uma base de dados em conjunto com o design do protótipo, correspondendo à fase 4. Esse processo só termina com a publicação do protótipo aceito, após um conjunto de testes com crianças.

4. Desenvolvimento do Protótipo

Seguindo o processo de design descrito na Figura 2 para o caso de design Tangram Tangível, houve uma articulação entre diferentes perfis para compor o time de desenvolvimento. Isso envolveu a participação de um professor de matemática da segunda etapa do ensino fundamental da rede pública de ensino local, duas crianças que, apesar de não testarmos diretamente os protótipos iniciais com elas, forneceram feedback prático sobre o uso do tangram utilizado em sala de aula, e os membros da equipe responsáveis pela parte técnica de desenvolvimento dos protótipos. Essa última subequipe, representada na figura 2 como *Designer*, é composta por um aluno de graduação em Ciência da Computação cujo papel interno é de desenvolvedor do App, a professora orientadora que atua como gerente e dois alunos de mestrado, um atuando na concepção e implementação das ideias definidas e outro na concepção das avaliações formativas.

Os instrumentos iniciais de entrevista e observação foram aplicados com o professor e as duas crianças participantes da equipe para definição das *affordances* e funcionalidades desejáveis para o protótipo. Na revisão da literatura foram encontrados trabalhos com desenvolvimento de protótipo para outros domínios, os quais utilizamos como uma comparação para as técnicas definidas e orientaram a própria escolha do artefato de aprendizagem da matemática. Orientando essa escolha também foi realizado o workshop de ideias com todos os envolvidos, mas de forma prática, com atividades de

composição de figuras no tangram. O planejamento das funcionalidades do tangram, além de receber os subsídios das tarefas anteriores, consultou alguns conjuntos de *guidelines* para desenvolvimento de artefatos tangíveis, os quais apresentaremos na subseção a seguir.

Para o primeiro protótipo funcional, definido após o *workshop* de ideias, definimos os materiais a serem utilizados, que foram: Arduino Uno, *Jumper wires*, *USB Cable*, LED, Tecido inteligente, Cartolina, Emborrachado, Papelão Branco, Régua, Lápis, Pincel e Fitas. Nesse protótipo inicial projetamos um quadrado grande com os desenhos das peças do tangram em forma da figura de uma casa, que serviria como base para testar as funcionalidades com as luzes de LED e sensores via tecido condutivo. Realizamos uma avaliação formativa com teste exploratório por todos os integrantes da equipe do projeto. Como o protótipo atendeu às expectativas e as funcionalidades atendidas, passamos para o outro protótipo, dessa vez um que pudesse também atender às *affordances*.

Os materiais para o segundo protótipo foram quase todos os mesmos, somente acrescentamos o *shield Bluetooth* e o sensor *hall*. Os componentes utilizados, então, foram: Arduino Uno, Sensor *Hall*, *Shield Bluetooth*, *Jumper wires*, *USB Cable*, LED, Tecido inteligente, Cartolina, Emborrachado, Papelão Branco, Régua, Lápis, Pincel e Fitas. Foram feitas as marcações para cada sensor hall nas 7 peças que compõem o Tangram e toda a etapa de acabamento das peças. Cada peça recebeu um Arduino Uno e um *Shield Bluetooth* para fazer comunicação com o Arduino servidor que por vez comunica-se com o dispositivo Android. Isso é ilustrado pela Figura 3, que mostra as marcações e pela Figura 4 que mostra a composição da peça.

As Figuras 5 e 6 ilustram as peças do tangram já com os Arduinos e sensores embutidos. Os materiais escolhidos são leves, o que torna a *affordance* melhor. A base do tangram, onde ele é montado foi recortada como uma primeira tentativa de explorar e combinar as diferentes possibilidades de manipulação. Segundo os resultados de nossa avaliação formativa para essa etapa, ainda realizada com a própria equipe do projeto, definimos que no próximo protótipo, já a ser manipulado por quarenta crianças em um contexto real de uso, a base será fixa, detectando as peças por mapeamento de região ocupada por cada uma.



Figura 3. Marcações para os sensores hall.

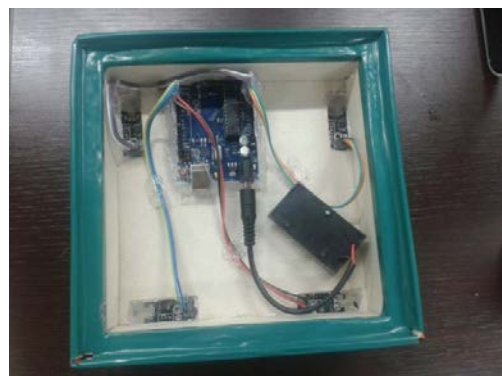


Figure 4. Peça quadrada com sensores hall e Arduino.

Consideramos também os *guidelines* existentes na literatura para desenvolvimento de múltiplos artefatos tangíveis [Zuckerman et al., 2016], [Heijboer, Marigo, e Van Den Hoven, 2008], [Kameas et al., 2005] e [Chu, 2015]. Desses trabalhos, o guideline mais completo e significativo foi o descrito por [Antle, Alissa N., e Alyssa F. Wise, 2013], sintetizado na Tabela 1.



Figura 5. Peças triângulos e quadrado com sensores hall, Módulo NRF24L01 e Arduino.



Figura 6. Peças Tangram finalizadas com sensores hall, Módulo NRF24L01 e Arduino.

Tabela 1 - Guidelines para Interfaces Tangíveis para Usuários.

Fonte: por [Antle, Alissa N., e Alyssa F. Wise, 2013].

		Physical Objects	Digital Objects	Actions on Objects	Informational Relations	Learning Activities
Perspective	Guidelines					
<i>Info processing</i>	1. Distributing information across modalities can increase effective working memory capacity	X	X		X	
	2. Making mappings between the form and behavior of physical and/or digital objects and real-world entities coherent can reduce extraneous cognitive load				X	
<i>Constructivist</i>	3. Creating authentic tasks and using personal objects Can support learners in forming individually.	X				X
	4. Using spatial, physical, temporal or relational properties can slow down interaction and trigger reflection	X	X	X	X	X
<i>Embodied</i>	5. Distributing parts of metal operations to actions on physical and/or digital objects can simplify and support mental skills			X	X	
	6. Leveraging image schemas in input actions can improve usability and system learnability			X		
	7. Using conceptual metaphor(s) based on image schemas to structure interaction mappings may bootstrap learning of abstract concepts				X	

<i>Distributed</i>	8. Designing objects that allow spatial reconfiguration can enable mutual adaptation of ideas	X	X			
	9. Using concrete representations can support interpretation of symbolic representations of abstract concepts	X	X		X	
<i>Collaborative</i>	10. Creating configurations in which participants can monitor each other's activity and gaze can support the development of shared understandings	X	X			
	11. Distributing roles, information and controls across the TUI learning environment can promote negotiation and collaboration	X				X
	12. Creating constrained or codependent access points schemes can compel learners to negotiate with each other			X		

As *guidelines* para objetos físicos e digitais, correspondentes às categorias processamento de informação, construtivista e distribuída estão contempladas em nosso processo de design colaborativo de artefatos tangíveis para crianças (Figura 2). As que correspondem a incorporadas e colaborativas são explicadas abaixo:

- (i) Partes distribuídas de operações mentais para ação em objetos podem simplificar e fornecer suporte às habilidades mentais – isso é atingido na especificação das atividades a serem realizadas com o protótipo, algumas das quais já definidas para a avaliação em situação real de uso.
- (ii) Nivelamento de esquemas de imagem em ações de entrada podem melhorar a usabilidade e aprendizagem do sistema – as imagens da operação desejada é mostrada no App para que a criança reflita sobre sua composição e a materialize no objeto físico. A imagem que vê no App é a que vai formar com as peças.
- (iii) Uso de propriedades físicas, espaciais, temporais e relacionais podem diminuir a interação e aumentar a reflexão – esse é um equilíbrio que procuramos obter na especificação das tarefas. Uma será de construção de imagem, outra de observação de uma figura construída para reflexão e outra envolvendo dedução de medidas. Tudo isso verificado automaticamente pelo artefato Tangram Tangível.
- (iv) Criação de configurações que possibilitem que as pessoas visualizem suas atividades e as dos demais – para isso, desenvolvemos uma funcionalidade para registro dos logs e utilizamos uma ferramenta de visualização gráfica como *plugin*.
- (v) Distribuição de papéis, informação e controle através dos artefatos nos ambientes de aprendizagem podem promover negociação e colaboração – atividades em grupo proporcional isso. O protótipo (App e físico) possibilita a materialização dessas estratégias fornecendo registro individual e em grupo, de acordo com a atividade e, posteriormente identificação das trilhas de aprendizagem no registro dos logs (erros e acertos), vídeos e anotações.
- (vi) Criação de um esquema de pontos de acesso restrito e co-dependente podem estimular os aprendizes a negociar entre si – temos somente um artefato físico

tangram. Portanto, nas atividades em grupo, as crianças terão que negociar para gerarem as soluções.

4.1. Criação do aplicativo para o Tangram Tangível

Para esta tarefa, um dos pontos mais importantes foi o planejamento do conteúdo da aplicação que teve como foco conhecer e entender o público-alvo e o seu nível de instrução.

Após se ter um entendimento completo de quem é o público-alvo (conhecimento proveniente das primeiras tarefas do processo da Figura 2) e como ele pensa e reage, deu-se início ao design do App. Esse processo começa com a decisão sobre o que se quer comunicar e os meios que se pretende usar para transmiti-lo. A Figura 7 mostra algumas telas do App desenvolvido.

Durante o design do App levou-se em conta não apenas ilustrações de alta qualidade, mas também a escolha apropriada de estilos, para se obter uma comunicação eficaz já que imagens, gráficos e ilustrações refletem a emoção mais rapidamente do que palavras. No entanto, deve-se tomar muito cuidado com este tipo de recurso, já que podem transmitir uma mensagem diferente da que se pretende e também para que não se torne algo cansativo para o usuário [Wolfgram, 1994].

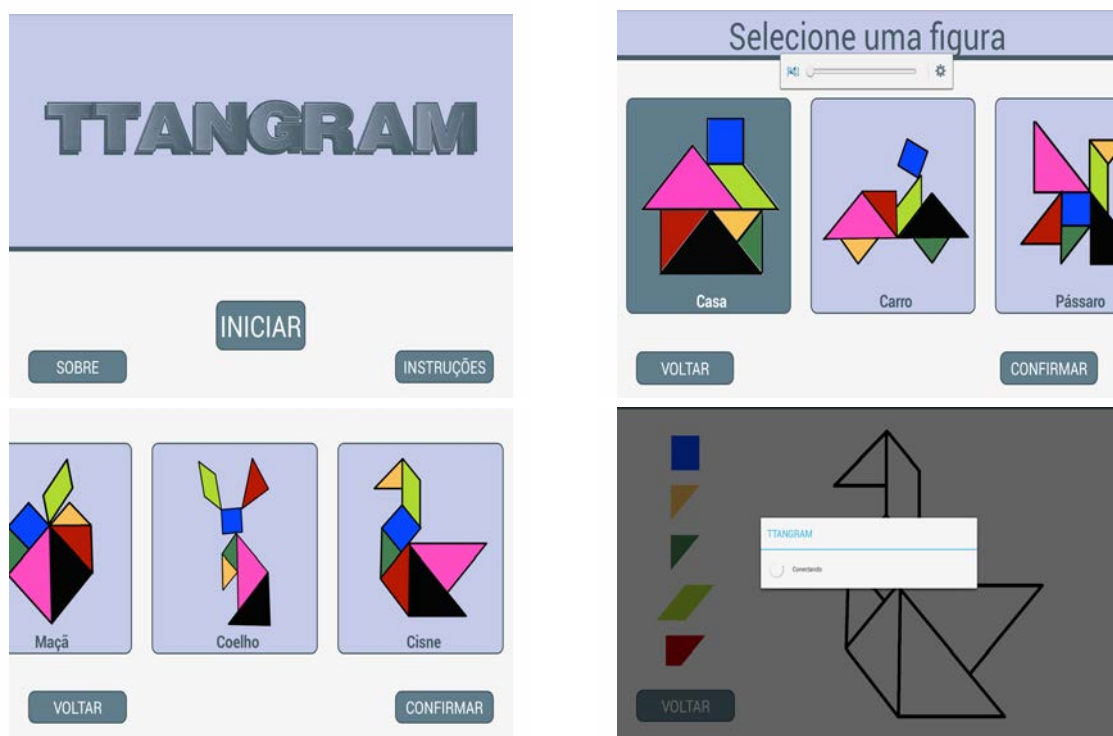


Figure 7 - Telas do App

Ao longo do design do App, além do planejamento do conteúdo, foram considerados os aspectos técnicos tais como: objetivos pedagógicos que podem ser classificados em tutoriais, aplicativos, programação, aplicativos, exercícios e prática,

multimídia e Internet, simulação, modelagem e jogos [Valente, 1999]. O *software* aqui descrito classifica-se como aplicativo e jogo. Outro aspecto focado foi a interação, pois este aspecto em Apps traz à tona um novo nível de controle implícito. O uso apropriado da interatividade resulta na efetiva lembrança da mensagem por parte do usuário [Wolfgram. 1994], pois as pessoas lembram mais daquilo com que interagem.

5. Especificação para objetos de aprendizagem tangíveis

Com base em recomendações e *guidelines* já mencionados nas seções acima, definimos algumas especificações para criação de artefatos tangíveis utilizados como objetos de aprendizagem:

- **Possuir dimensões e/ou o peso aceitáveis:** não podem ser nem muito pesados nem muito leves para que atendem ao requisito de *affordance*.
- **Fornecer ideias multissensoriais para contemplar diferentes estilos de aprendizagem e necessidades diversas:** essas ideias ou representações devem permitir que outros sentidos das crianças sejam empregados nas atividades, ajudando-as a raciocinar. O objetivo é oferecer oportunidades de interação entre as crianças, favorecendo assim a exploração com mais liberdade das figuras geométricas, seus perímetros, áreas e ângulos.
- **O produto deve ser resistente, de forma que as crianças não tenham receio de quebrar na hora de brincar (manipular):** as bases para interfaces tangíveis devem ser feitas de materiais firmes tais como madeira, plástico ou papel madeira. Suas extremidades devem possuir cantos arredondados usados para definir uma transição plana entre duas faces, normalmente definida a partir de uma aresta do modelo sólido ou de faces. A parte eletrônica do objeto deve ficar embutida para que não haja contato físico da criança.
- **Ser manipulável (móvel e independente):** o objeto tangível deve ser independente de energia com uso de baterias permitindo a exploração, construção e divisão de formas geométricas. A interface tangível deve continuar sendo usada para que as crianças possam explorar as especificidades para a qual a mesma foi desenvolvida mesmo com os recursos eletrônicos desativados.
- **Feedback:** Deve possuir uma opção de ajuda e possuir *feedback* positivo e negativo. Segundo Kekalainen (2005): Os métodos de aquisição do *feedback* dos usuários podem ser divididos em duas categorias:
 - a) **Implícito:** realizado pelo sistema, baseando-se em informações adquiridas indiretamente dos usuários, como histórico de navegação, cliques, entre outros;
 - b) **Explícito:** realizado pelos usuários, baseando-se em informações diretas do usuário, como a indicação de pertinência de um resultado, por exemplo, nos sistemas que possuem algum mecanismo de *feedback* explícito, a indicação da relevância dos resultados pode ser binária, quando apenas é dito se o resultado é relevante ou não, ou em escala, quando existem alguns níveis (mais que dois) de relevância possíveis a serem escolhidos. Essa escala pode ser efetuada usando números (de 1 à 5, por exemplo), letras ou com descrições (“irrelevante”, “pouco relevante”, “relevante”, “muito relevante”, por exemplo).
- **Estímulo ao raciocínio:** Provendo reflexão do pensamento lógico na resolução dos problemas abordados pelo conteúdo e o objetivo principal da interface.

- **Independência do tutor:** A interface deve proporcionar aos alunos (crianças) uma certa independência do tutor ou professor, para que a mesma consiga realizar algumas atividades de forma independente (sem ajuda externa), oferecendo segurança à criança durante a execução da atividade, motivando-a a explorar suas funcionalidades.
- **Ligação entre teoria-prática:** interfaces e artefatos tangíveis para crianças devem ser intuitivas e lúdicas, a exemplo das atividades de montagem comuns em brincadeiras infantis e atividades escolares.
- **Permitir uso colaborativo:** artefatos tangíveis devem oferecer aos seus usuários diferentes formas de interação, facilitando o controle, a coordenação, a colaboração e a comunicação entre as partes envolvidas. Dessa forma, as soluções são criadas com negociação e colaboração, desde o início da atividade.
- **Flexibilidade para criação de atividades:** os artefatos de hardware e software tangíveis devem possibilitar a mudança e implementação de novas atividades pelos professores e tutores sem maiores traumas de uma maneira muito simples.
- **Adaptação do protótipo (material) aos conteúdos:** as configurações devem ser adaptáveis aos conteúdos propostos pelos professores.

Interface compreensível: uma interface tangível deve ser fácil de usar. Na interface do Tangram Tangível o professor escolhe o nível de dificuldade e a figura que a criança irá montar e o tipo de atividade ou pode recorrer ao menu de ajuda em caso de dúvidas e consultar maneiras de configurar novas propostas de uso.

As especificações acima estão em fase de formalização para que possam ser reutilizadas por outros pesquisadores e professores que precisem definir um objeto de aprendizagem tangível para a aprendizagem de matemática. Conforme mencionado anteriormente, as especificações acima foram definidas com base nas guidelines descritas em [Zuckerman et al., 2016], [Heijboer, Marigo, e Van Den Hoven, 2008], [Kameas et al., 2005] e [Chu, 2015], utilizando como princípio a avaliação formativa que, segundo [Araújo e Soares, 2014] possibilita verificar problemas de interação e de interpretação de interface durante o processo de design de um artefato.

Atualmente o protótipo está sendo preparado para a avaliação em uma situação real de uso, em uma gincana de matemática a ser realizada com cerca de quarenta crianças do 6º ano do ensino fundamental de uma escola da rede pública de ensino local. Durante essa gincana, os *logs* serão armazenados para compor a base inicial de logs relacionados a atividades, já ficando disponíveis para visualização. Após a gincana, será realizada uma análise quantitativa dos resultados, considerando os itens que definem uma boa interação com objetos tangíveis e identificação das trilhas de aprendizagem individuais e em grupo.

6. Conclusão

Interfaces tangíveis agregam valor a objetos computacionais (*software*), possibilitando que haja mais interação entre os usuários e o registro das interações dos usuários com os artefatos, possibilitando abordagens diferenciadas para as necessidades específicas dos usuários.

Conforme exemplificado pelas referências neste trabalho, existem várias interfaces tangíveis propostas para a área educacional. Sabe-se que *design* de interfaces interativas não é mais exclusivamente gráfico, mas também envolve forma e materiais.

Assim, características como ergonomia, peso, textura e outros aspectos ligados a materiais físicos compõem novas dimensões para projeto e avaliação de interfaces [Cardador, 2013]. Neste trabalho, baseados em nossas experiências anteriores, nas recomendações, *guidelines* e abordagens encontradas na literatura, propomos um processo de design colaborativo de artefatos tangíveis para crianças. Descrevemos seu funcionamento através da prototipação de um caso de design, o Tangram Tangível, o qual é voltado para auxiliar na aprendizagem da matemática, abordando especificamente o conteúdo de geometria plana.

Juntamente com a prototipação do Tangram Tangível, neste artigo definimos uma especificação para objetos de aprendizagem tangíveis no contexto de aprendizagem de matemática na educação básica. Essa especificação está em processo de formalização para ser disponibilizada para a configuração de novos artefatos para o mesmo propósito e ser usada também para configurar novas atividades em um sistema de gerenciamento de objetos educacionais tangíveis.

A preparação de um bom material e a forma como são organizados na sala induz o aluno à manipulação imediata dos objetos (alta *affordance*), sem a necessidade de leituras longas e cansativas. Entretanto, ao utilizar esse tipo de tecnologia em uma sala de aula real, o professor deve proceder um planejamento minucioso, o que requer tempo e habilidade do professor na utilização dos artefatos.

7. Agradecimentos

Esse trabalho recebeu apoio da FAPEAM – Projeto “*Intelligent Framework to Support Collaboration*”, Ed. 016/2013, Res. 035/2013. Agradecemos aos alunos de graduação e pós-graduação, professor e crianças que participam da equipe de design. Agradecemos também a CAPES, através do projeto CARECO - Sistemas de Recomendação para Manutenção Colaborativa de Software, aprovado sob o edital n. 071/2013, o qual possibilitou algumas definições deste artigo, via trocas de ideias.

Referências

- Antle, Alissa N. and Wise, Alyssa F. "Getting down to details: Using theories of cognition and learning to inform tangible user interface design." *Interacting with Computers* 25.1 (2013): 1-20.
- Araújo, Eduardo Cruz, and Luiz Fernando Gomes Soares. "Designing iDTV Applications through Interactive Storyboards." *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*. ACM, 2014.
- Bakker, Saskia; Antle, Alissa N. and Van Den Hoven, Elise. "Embodied metaphors in tangible interaction design." *Personal and Ubiquitous Computing* 16.4 (2012): 433-449.
- Barbosa, J, D, S., Silva, S, B. *Interação Humano Computador*. Editora Campus, Elsevier Editora, Rio de Janeiro – RJ- Brasil, 2010.
- Bevan, N., Curson, I. *Planning and Implementing User-Centred Design*. Teddington, Middlesex, TW11 0DU, UK, CHI 99 15-20 MAY 1999.
- Chu, Jean Ho. "Designing Tangible Interfaces to Support Expression and Sensemaking in Interactive Narratives." *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 2015.

- Cardador, D.; Ugulino, W.; veja, K.; Filippo, D.; Raposo, A.; Fuks, H. Uma abordagem sistemática de prototipação colaborativa para a criação de tangíveis. In: *Simpósio Brasileiro De Sistemas Colaborativos* (SBSC'13), 5., 2013, Manaus, Brasil. Anais... Manaus: SBC, 2013.
- Diniz, I, M.; et. Al. A matemática das sete peças do tangram. CAEM-IME/USP, 2002.
- Eisenberg, M. Mindstuff: *Educational Technology beyond the computer*. Convergence, Colorado-USA, 2003.
- Genova, A. C. Brincando Com Tangram e Origami. Editora Global-São Paulo – 2005, Edição: 4 , Vol. 18.
- Heijboer, Marigo, and Elise Van Den Hoven. "Keeping up appearances: interpretation of tangible artifact design." *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. ACM, 2008.
- Kameas, Achilles, Irene Mavrommati, and Panos Markopoulos. "Computing in tangible: using artifacts as components of ambient intelligence environments." *Ambient Intelligence* (2005): 121-142.
- Kekalainen, J. (2005). *Binary and graded relevance in IR evaluations - comparison of the effects on ranking of IR systems*. In: *Information Processing and Management: an International Journal*, v.41 n.5, p.1019-1033. *International Journal*, v.41 n.5, p.1019-1033, 2005.
- Kitchenham, B. Dyba, T. Jorgensen, M. *Evidence Based Software Engineering*, In Proceedings of 26th International Conference on Software Engineering (ICSE'04), pp 273-281, 2004.
- Lawson, Bryan. *How designers think: The design process demystified*. Routledge, 2006.
- Maguire, Martin. "Methods to support human-centred design." *International journal of human-computer studies* 55.4 (2001): 587-634.
- McGrenere, Joanna, and Wayne Ho. "Affordances: Clarifying and evolving a concept." *Graphics Interface*. Vol. 2000. 2000.
- Oliveira, G. C. C. vamos fazer design de interação? ilustração e diagramação: estúdio grampo, 2013. <http://designdeinteracao.com.br/wp/wp-content/uploads/2013/10/vfdi.pdf>.
- Rosemberg, C., Schilling, A., Bastos, C., Araripe, R. Prototipação de Software e Design Participativo: uma Experiência do Atlântico. 21-24 Outubro | Porto Alegre – RS, Brasil, 2008.
- Valente, J. A. (Org) – O computador na sociedade do conhecimento. Campinas, SP: Unicamp/NIED, 1999.
- Wolfgram, E. D. Criando Em Multimídia. Editora Campus, Edição 87, Rio de Janeiro- RJ, Dezembro de 1994.
- Zuckerman, O. S. A., Resnick, M. *Extending tangible interfaces for Education: Digital Montessori-inspired manipulatives*. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 05), pages 859-868, Portland, USA. ACM Press. 2005.

Zuckerman, O., Gal, T., Keren-Capelovitch, T., Karsovsky, T., Gal-Oz, A., & Weiss, P. L. T. (2016, February). DataSpoon: Overcoming Design Challenges in Tangible and Embedded Assistive Technologies. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 30-37). ACM.

CAPCOM - Construtor de Arquiteturas Pedagógicas para Dinâmicas Colaborativas com Multimídia

Renan F. Almeida¹, Orivaldo de L. Tavares¹, Luiz F. Reinoso¹

¹Departamento de Informática – Universidade Federal do Espírito Santo (UFES)
Vitória – ES – Brasil

ralmeida@ifes.edu.br, tavares@inf.ufes.br, luizfreinoso@gmail.com

Abstract. *This article presents a platform to set up pedagogical architectures to promote collaborative dynamics that use multimedia created in field activities. Called CAPCOM, this new platform offers digital resources to support the development of collaborative educational activities with multimedia records made in the field. This research deals specifically with the digital resources to support collaborative dynamics to be performed after the field activity. With this platform for building educational architecture for collaborative dynamics with multimedia texts, the teacher creates according to the desired educational goal activities.*

Resumo. *Esse artigo apresenta uma plataforma para configurar arquiteturas pedagógicas para promover dinâmicas colaborativas que usem multimídias criados em atividades de campo. Denominada CAPCOM, essa nova plataforma oferece recursos digitais para apoiar o desenvolvimento de atividades pedagógicas colaborativas com os registros multimídias feitos em campo. Este trabalho trata especificamente dos recursos digitais para suporte à dinâmicas colaborativas a serem realizadas após a atividade de campo. Com essa plataforma para construção de arquiteturas pedagógicas para dinâmicas colaborativas com multimídias, o professor cria atividades de acordo com o objetivo pedagógico pretendido.*

1. Introdução

As atividades desenvolvidas em espaços abertos podem apresentar diversas oportunidades de aprendizado, podendo auxiliar os participantes a visualizar ou compreender as relações que existem entre os conceitos teóricos, adquiridos em sala de aula, com os conhecimentos práticos adquiridos pela experiência, conhecido como conhecimento empírico.

Considerada uma estratégia de ensino, tirar os alunos da escola para passeios, visitas técnicas, excursões ou visitas à exposições, pode fomentar para o aprendiz a oportunidade de moldar sua aprendizagem com novos conhecimentos, ampliando assim o interesse e fortalecendo sua aprendizagem. O professor assume o papel de facilitador da aprendizagem, nessa estratégia, ao propor atividades pedagógicas motivadoras e muitas vezes inéditas para os aprendizes.

Aulas de campo devem se articular entre formação teórica e conhecimentos práticos. Elas têm como objetivo ampliar o aproveitamento dos conhecimentos adquiridos, associando das abstrações mentais aos aspectos físicos e naturais dos lugares explorados nas atividades desenvolvidas em espaços abertos.

[Seniciato and Cavassan 2004] mostram, em um estudo feito com 97 alunos da sexta série do ensino médio, que a realização de aula de campo sobre botânica, após as aulas teóricas, foi notado ganho de aprendizagem nos conceitos estudados, chegando a um aumento de 40% no conhecimento dos alunos sobre os assuntos das aulas.

Em [Oliveira and Correia 2013] foi descrita um experiência similar onde alunos do terceiro ano do ensino médio participaram de uma atividade, em um espaço aberto, cujo o objetivo era investigar o papel da aula de campo como mecanismo facilitador no processo de ensino-aprendizagem sobre o ecossistema recifal, importante ambiente do litoral de Alagoas. Segundo [Oliveira and Correia 2013]:

“A partir dos resultados obtidos nos questionários aplicados nas duas fases, observou-se que os estudantes tiveram uma mudança considerável na aquisição de conhecimentos quando comparados os dados antes e após a aula de campo.”

Por se tratar de uma atividade essencialmente colaborativa, aulas de campo apresentam diversas oportunidades de aprendizado e grandes desafios associados a essa estratégia de ensino, dentre os desafios dessa atividade estão: gerenciamento dos conhecimentos adquiridos, fixação desses conhecimentos, socialização das informações registradas, gerenciamento das atividades conduzidas, delimitação do espaço a ser explorado, monitoramento do percurso feito pelos participantes, entre outros elementos.

Usando tecnologias que permitem explorar e registrar dados do mundo físico e processar esses dados, as aulas tornam-se substancialmente mais dinâmicas e motivantes, de modo a permitir experiências reais e virtuais com os dados produzidos durante uma aula de campo, proporcionando a aprendizagem colaborativa.

Segundo [Chagas 2002], desenvolver atividades colaborativas é a “condição necessária para que as redes de aprendizagem se constituam e se mantenham, em prol de concretizar a construção de conhecimento pelos seus intervenientes”.

Segundo [Vieira and Tavares 2015] a flexibilidade dos ambientes virtuais colaborativos ainda não está condizente com a complexidade resultante das variações de cenário para cooperação. [Menezes et al. 2008] evidenciam, para sustentar as práticas pedagógicas, a dificuldade dos ambientes virtuais colaborativos para atender a implantação de novas propostas pedagógicas.

Este trabalho apresenta os resultados obtidos com a associação e implementação dos conceitos de Arquitetura Pedagógica (AP) [Carvalho et al. 2005] e os conceitos que fundamentam o MOrFEU (Multi Organizador Flexível de Espaços Virtuais) [Menezes et al. 2008] cujo principal objetivo é libertar o professor da complexidade da tecnologia digital necessária para sustentar as prática pedagógicas.

Segundo [Reinoso and Tavares 2015], uma Arquitetura Pedagógica é constituída de: 1. objetivo pedagógico (o que aprender); 2. atividade(s) pedagógica(s) (o que fazer para aprender); 3 metodologia (como fazer cada atividade); 4. recursos digitais (recursos disponíveis para apoiar o desenvolvimento de cada atividade). A pesquisa desenvolvida por [Vieira et al. 2011] verificou a necessidade de requisitos de software para *groupwa-*

res, classificados como Learning Management Systems (LMS), mais aderentes às necessidades de cada *groupwares* e que favoreça a criação de APs, com maior articulação e coordenação das atividades colaborativas.

Como resultado da combinação dos conceitos supracitados, este artigo propõe um Construtor de Arquiteturas Pedagógicas para Dinâmicas Colaborativas com Multimídia, onde o professor possa descrever e criar atividades pedagógicas, de modo flexível, para que o aprendiz possa revisitar os registros feitos em um aula de campo, para realizar colaborativamente operações sobre eles, tais como classificá-los segundo o objeto ou o assunto do registro, ou compará-los com registros feito por colegas, de modo a fazer uma metareflexão sobre os eventos vivenciados na atividade em campo.

Para modelar a plataforma CAPCOM foram incorporados os conceitos criados para o projeto MOrFEu tais como Veículo de Comunicação (VCom), Unidade de Produção Intelectual (UPI) e Template para VCom, de modo a permitir que as Arquiteturas Pedagógicas, criadas pelo professor, possam ser instrumentadas com recursos digitais configurados sob medida para elas.

Esse trabalho expande a proposta apresentada em [Pereira et al. 2015], e apresenta o projeto e implementação de um espaço virtual flexível e colaborativo para criação e uso de APCs (arquiteturas pedagógicas colaborativas) com registros multimídia de atividades de campo.

2. Ferramentas usadas pela plataforma CAPCOM

A Google oferece serviços de processamento de imagem com georreferência, necessários para aplicações como as que podem ser geradas pela plataforma CAPCOM. No desenvolvimento do CAPCOM, foram usadas duas soluções do Google: Google Street View e o Google Maps.

O Google Street View é constituído do Google Maps, um recurso para editoração e criação de mapas geográficos e o Google Earth, uma ferramenta para navegação em mapas do globo terrestre. Apesar de integradas funcionam de forma independente, utilizando como base as posições de latitude e longitude. O recurso disponibiliza vista panorâmica de 360° na horizontal e 290° na vertical, permite que os usuários visualizem partes de algum lugar no mundo, que esteja mapeado na plataforma, ao nível do chão. A API do Google Street View é utilizada para visualizar e editar mídias do tipo foto esférica.

O Google Maps é um serviço gratuito para visualizar e interagir com mapas e imagens de satélite da terra. Com as implementações desse recurso é possível traçar rotas nos mapas, fazer marcações, visualizar estruturas geográficas etc. A API do Google Maps é utilizada para visualizar o contexto onde as mídias foram produzidas e visualizar mídias do tipo ponto no mapa.

Para o desenvolvimento da aplicação web foi utilizado a API do mindmeister, HTML, CSS, Java Script e PHP, a framework Yii, que é repleta de recursos digitais para desenvolvimento orientado a objetos e padrão MVC (Model-View-Controller).

3. CAPCOM e as Práticas Pedagógicas

No contexto contemporâneo, as teorias relacionadas ao aprendizado são normalmente referidas como teorias de aprendizagem construtivista, com ênfase na aprendizagem ativa

e autorregulada [Lei et al. 2009]. Tal abordagem salienta a proposta dos experimentos nos quais o conhecimento não é transmitido, mas construído nas atividades ou interações sociais.

As teorias construtivistas de aprendizagem, têm um impacto significativo sobre as práticas pedagógicas, especialmente na promoção da aprendizagem ativa. Nesse contexto as teorias construtivistas solicitam diferentes ambientes de aprendizagem para suportar os elementos que fundamentam essa teoria.

Segundo [Lei et al. 2009] os ambientes construtivistas são complexos, realistas e significativos. Em um ambiente de aprendizagem complexo, o aprendiz é desafiado a construir novos conhecimentos que possam ser aplicados, de preferência no dia a dia, em situações concretas. Tal definição tem o potencial de tornar a aprendizagem uma experiência significativa pessoal, elemento esse que contribui para motivação dos aprendizes no processo de aprendizagem. A motivação, por sua vez, é um elemento essencial para alcançar um nível mais profundo de aprendizagem.

Fundamentado na teoria do construtivismo, o CAPCOM adota o paradigma AP (arquitetura pedagógica), que apoia-se na pedagogia da incerteza, pressupondo que o conhecimento nasce do movimento da dúvida, da incerteza, da busca de novas alternativas, do debate e da troca [Carvalho et al. 2005]. Buscando se diferenciar do ensino convencional de sala de aula e da materialização de apostilas de cursos para a educação à distância, uma AP abrange processos para a criação de novidades, descobertas e experimentações, centrada no ponto de vista do aprendiz, o protagonista da aprendizagem, durante a autoria de soluções individuais ou coletivas [Menezes et al. 2008].

As APs são estruturas de aprendizagem, que combinam abordagem pedagógica, inteligência artificial, educação a distância e concepção de tempo e espaço, e que são inspiradas em princípios das ideias construtivistas de Piaget e da pedagogia da pergunta de Freire, estabelecendo a Pedagogia da incerteza que implica em: (i) soluções de problemas reais, (ii) transformação de informações em conhecimento, (iii) autoria, expressão e interlocução, (iv) investigação e criação de novidades e (v) autonomia e cooperação [Carvalho et al. 2005].

Seguindo as teorias construtivistas e o conceito de Arquitetura Pedagógica, a CAPCOM foi modelada com o propósito de apoiar a construção colaborativa do conhecimento, seguindo a orientação das APs Colaborativas criadas pelo professor. Tal abordagem é implementada nas atividades com foco no tratamento colaborativo de multimídia, proporcionando a consolidação do conhecimento adquirido nas atividades de campo.

A forma como o sistema foi idealizado permite atender as necessidades de reflexão sobre uma aula de campo e seu respectivo objetivo pedagógico, atendendo tanto atividades conduzidas com crianças até atividades desenvolvidas com adultos, em aulas de campo relacionadas ao ensino de matemática, ciência (classificação, estruturação de conhecimento etc), engenharia, ciências biológicas, geografia, computação, por exemplo.

4. Conceitos do MOrFEU usados na plataforma CAPCOM

O projeto MOrFEU (Multi-Organizador Flexível de Espaços Virtuais) tem como objetivo libertar o professor da complexidade das tecnologias e ferramentas necessárias para sustentar as práticas pedagógicas propostas pelo professor. Segundo [Menezes et al. 2008] o

MOrFEU se pauta nos seguintes aspectos: plasticidade, ergonomia, redução da repetição de trabalho e redução da sobre carga cognitiva. Tais características são muito relevantes para a CAPCOM, dado que essa plataforma foi pensado para fomentar novas oportunidades de aprendizado em um contexto pós aula da campo.

Neste trabalho são usados três conceitos do MOrFEU: VCom (Veículo de Comunicação); UPI (Unidade de Produção Intelectual) e Template para VCom. Esses conceitos são usados na modelagem do CAPCOM, de modo a permitir ao professor flexibilidade para configurar os recursos digitais usados nas atividades colaborativas de consolidação do conhecimento desenvolvido na aula de campo, por meio da realização de operações cognitivas (colaborativas) sobre os textos multimídia elaborados no campo, tais como: classificação, criação de textos multimídia mais elaborados, etc.

Um VCom é um canal para troca de comunicações entre pessoas que nele assumem papéis, as vezes simultaneamente, de criadores e consumidores de conteúdo [Pereira 2015]. Alguns exemplos que materializam o conceito de VComs são: aplicativos de publicação de mensagem, blogs da internet, redes sociais digitais, revista, relatório de uma aula de campo, mapas mentais etc.

Segundo [Vieira 2011] “O conceito de UPI (unidade de produção intelectual) permite que todo texto (uma foto, um texto escrito, um vídeo, um áudio etc) seja registrado com autoria, data, versão etc, de modo independente de onde ela esteja publicada.” Uma UPI é a unidade base de toda informação produzida, unidade essa que preserva a autoria e a característica de texto que essa unidade tem. Uma UPI pode assumir zero ou muitas associações com VComs. Como exemplo, em um VCom que se constitua em um relato de uma aula de campo, uma UPI pode ser um de seus componentes, em outro VCom que se constitua em um fórum, a mesma UPI pode ser uma das postagens e, em um terceiro VCom que se constitua em um mapa mental, a mesma UPI pode ser um dos elementos constituintes desse mapa. Tal flexibilidade pode ser configurada por uma ferramenta para construção de VCom, onde os critérios para a recepção e publicação de UPIs podem ser ajustados de acordo com a necessidade.

O conceito de Template para VCom permite configurar diferentes formas visuais para um mesmo conjunto de UPIs associados a um VCom. Segundo [Vieira 2011], “Cada VCom pode estar associado a diferentes templates, o que permite que cada usuário possa definir como deseja visualizá-lo.” Como exemplo, um mapa mental, tipo de VCom, pode assumir vários arranjos, ao mesmo tempo.

Em um arranjo de mapa mental, todos os conceitos podem estar ligados de forma caótica e, em outro arranjo, as ligações entre os conceitos podem seguir uma orientação circular, no entorno do conceito central. Nesses exemplos, a UPI e o VCom continuam com a mesma estrutura, o que varia é apenas o aspecto visual, o Template do VCom.

5. Arquitetura Pedagógica Colaborativa

Atividades colaborativas são definidas como “[...]uma proposta pedagógica na qual estudantes ajudam-se no processo de aprendizagem, atuando como parceiros entre si e com o professor, com o objetivo de adquirir conhecimento sobre um dado objeto[...]” [Leite et al. 2005]

Para [Dillenbourg et al. 1995] aprender de forma colaborativa remete a uma

situação na qual duas ou mais pessoas buscam aprender algo juntas, essa proposta é considerada muito relevante para apoiar o processo de aprendizado.

Aprendizagem de forma colaborativa requer um ambiente diferenciado, já que os professores e alunos desempenham novos papéis nessa abordagem. Os alunos se tornam elementos ativos e responsáveis por sua aprendizagem, nesse contexto o professor deixa de ser o único detentor do conhecimento e passa a promover as propostas pedagógicas, onde os alunos podem progredir com seus próprios esforços [Castro and Menezes 2011].

Usando a estrutura de AP definida por [Reinoso and Tavares 2015], para criar uma prática pedagógica colaborativa, é necessário definir os seguintes elementos: o objetivo pedagógico (o que aprender), atividades (o que fazer), método (como fazer) e recursos digitais (com que ferramentas).

O modelo de colaboração usado como referência pelo professor para criar cada AP colaborativa é o modelo 3C - comunicação, coordenação e cooperação [Pimentel et al. 2006].

A coordenação é desempenhada especialmente pelo professor que organiza os recursos, os participantes das tarefas e acompanha o desenvolvimento das atividades.

A comunicação é desempenhada por todos, mas os alunos, em especial, dependem da comunicação para colaborar na construção das atividades, para argumentação e negociação de compromissos.

A cooperação é desempenhada especialmente pelos alunos, seja nas atividades em grupo ou individuais, visto que mesmo nas atividades individuais existe a cooperação. Durante a cooperação os participantes podem processar e criar novos objetos.

Seguem três exemplos de Arquiteturas Pedagógicas criadas na plataforma CAP-COM (Construtor de AP para dinâmicas colaborativas com textos multimídia):

AP1: Arquitetura Pedagógica Cadeia Alimentar.

Contexto: Após uma aula de campo que acontece em um zoológico, o professor decide criar atividades pedagógicas na qual os alunos possam trabalhar com as mídias produzidas e registradas durante a aula, em um servidor (computador central), para aprender sobre a cadeia alimentar de um grupo de animais que dividem um mesmo ecossistema.

Objetivo pedagógico: – Aprender sobre os níveis tróficos de uma cadeia alimentar, dado um conjunto de seres vivos de um ecossistema.

Atividades: – Associar mídias produzidas em uma aula de campo, usando mapas mentais para registrar a associação.

Método: – O professor seleciona um conjunto de mídias que precisa ser explorado pelos alunos. O professor forma grupos de três alunos e solicita que os alunos colaborem para associar todas as mídias produzidos em uma aula de campo.

Recurso digital: – São usados recursos para edição de mapas mentais, suporte a grupos de colaboração e edição de UPIs.

Colaboração: - No recurso digital “edição de mapas mentais” os alunos trabalham com a fase de cooperação para construir o mapa mental. No recurso “suporte a grupos de colaboração” os alunos e o professor trabalham a fase de coordenação e

comunicação para viabilizar os resultados a serem produzidos. No recurso “edição de UPIs” os alunos trabalham a fase de cooperação para compor as novas UPIs.

AP2: Arquitetura Pedagógica Relato de Grupo.

Contexto: Após uma aula de campo, que acontece em uma área industrial, o professor decide criar atividades pedagógicas colaborativas para os alunos relatarem sobre os conhecimentos adquiridos, usando as mídias produzidas e registradas durante a aula.

Objetivo pedagógico: – Criar relatórios colaborativos com as UPIs da aula em campo.

Atividades: – Combinar logicamente as UPIs (multimídia) e produzir um relatório das atividades desenvolvidas durante a aula de campo

Método: – O professor define que os alunos de um grupo podem acessar apenas os textos (UPIs) que produziram; forma grupos de cinco alunos e desafia o grupo a produzir um relatório das atividades conduzidas durante a aula de campo.

Recurso digital: – São utilizados recursos para edição de UPIs, suporte à colaboração e negociação para uso de UPIs.

Colaboração: - Com o recurso “edição de UPIs” os alunos trabalham a fase de cooperação para compor as novas UPIs. Com o recurso “suporte a grupos de colaboração” os alunos e o professor trabalham a fase de coordenação e comunicação para viabilizar os resultados a serem produzidos. Com o recurso digital “negociação de UPIs”, os alunos e o professor trabalham a fase da comunicação, coordenação e cooperação em conjunto, para viabilizar o uso de uma mídia, de autoria de um aluno, por parte de outro aluno do mesmo grupo.

AP3: Arquitetura Pedagógica Fórum de Vídeos.

Contexto: Após uma aula de campo, onde um grupo de alunos de jornalismo entrevista pessoas que estão assistindo a uma peça de teatro, o professor decide criar atividades pedagógicas na qual os alunos possam apresentar e discutir sobre o processo de confecção dos vídeos registrados em cada entrevista durante a aula.

Objetivo pedagógico: – Aprender a registrar e usar vídeos Atividades – Combinar os vídeos de forma sequencial e registrar os vídeos

Método: – O professor define que a atividade será em grupo; define que só UPIs do tipo vídeo podem ser utilizadas; forma grupos de cinco alunos e desafia os alunos a se ajudarem para registrar seus vídeos de forma sequencial.

Recurso digital: – São usados recursos para edição de UPIs, suporte à colaboração, negociação para uso de UPIs e registro de novas UPIs

Colaboração: - Com o recurso digital “edição de UPIs” os alunos trabalham a fase de cooperação para compor as novas UPIs. Com o recurso digital “suporte a grupos de colaboração” os alunos e o professor trabalham a fase de coordenação e comunicação para viabilizar os resultados a serem produzidos. Com o recurso digital “negociação de UPIs”, os alunos e o professor trabalham a fase da comunicação, coordenação e cooperação em conjunto, para viabilizar o uso de uma mídia, de autoria de um aluno, por parte de outro aluno de qualquer um dos grupos. Com o recurso digital “registro de UPIs” os alunos

trabalham a fase de cooperação para enriquecer as possibilidades de vídeos que podem ser usados na execução da atividade.

6. O Papel do Professor

Durante as atividades, o professor deve produzir conflitos cognitivos entre os seus alunos, incentivando-os a pensar, a analisar quaisquer conteúdos, por meio do raciocínio lógico, da análise e da comparação [Giordani et al. 2006].

As interações entre os aprendizes podem ser afetadas por aspectos culturais diferentes como: estilos de comunicação; atitudes diante de conflitos; abordagens para executar uma tarefa; estilos de tomadas de decisões, atitudes diante de novas descobertas e abordagens sobre o conhecimento [Leite et al. 2005]. Por tanto a solução precisa atender as particularidade dos aprendizes, visando facilitar a interação.

Criando AP colaborativas, com o uso do CAPCOM, o professor ou mediador das atividades pedagógicas pode incentivar o trabalho colaborativo entre os alunos.

Como exemplo de uso do CAPCOM, o professor pode trabalhar o estudo sobre felinos, no contexto de uma aula de campo dentro de um zoológico, onde os alunos produzem textos multimídia com registros das observações que fizeram, e em um momento posterior à aula de campo, o professor pode propor o uso de uma AP (arquitetura pedagógica), com um Recurso Digital onde os alunos possam colaborar para construir relatos e consolidar conhecimentos sobre os felinos, a partir dos registros feitos por toda a turma.

Em outro exemplo, o professor pode criar uma AP, com a CAPCOM, para usar as multimídia produzidas pelos alunos em uma atividade de campo. Essa AP dispõe de Recurso Digital para os alunos construírem, de modo colaborativo, um relato descritivo das atividades executadas, incluindo mídias do tipo foto, foto esférica, vídeo e áudio.

As mídias do relato podem fazer associações entre os conceitos apreendidos durante a visita à atividade e os mídias (UPIs) produzidos, de modo a criar uma classificação (taxonomia) dessas mídias.

[Ocepek et al. 2013] defende que alunos e professores, com diferentes estilos de aprendizagem, tendem a preferir determinadas mídias em detrimento a outras, de acordo com o estilo de aprendizagem. Nesse sentido o professor precisa incentivar os alunos a registrarem as informações de modo colaborativo, podendo os registros produzidos atenderem aos estilos de aprendizagem de cada aluno.

Antes da atividade de campo e após essa atividade, o professor pode promover o debates sobre o propósito pedagógico das atividades, apresentando informações que possam ser verificadas durante a atividade, estimulando nos alunos o aprendizado colaborativo e o aprendizado por associação.

7. A plataforma CAPCOM

A plataforma CAPCOM usa UPIs criadas em atividades de campo, por uma turma, para a criação de arquiteturas pedagógicas que propõem atividades pedagógicas de consolidação do conhecimento adquirido na atividade em campo. Essas arquiteturas pedagógicas são usadas pelos aprendizes em momento posterior às atividades de campo.

Segundo [Vieira and Tavares 2015] novas práticas de ensino-aprendizagem sustentadas pelas arquiteturas pedagógicas sugerem plataformas tecnológicas sintonizadas às necessidades de interação sob protocolos distintos. Os ambientes colaborativos, principalmente os ambientes virtuais de aprendizagem, devem apresentar ao professor, enquanto coordenador das atividades, recursos funcionais capazes de tornar possível a implementação das suas estratégias de ensino e aprendizagem.

Aulas de campo são atividades que permitem reflexão sobre os conhecimentos articulados em estudos teóricos, as atividades realizadas em campo geram dados que são organizados em relatórios. A plataforma AM [Pereira et al. 2015] permite o registro de UPIs (textos) durante a realização das atividades de campo, mas carece de suporte para a reflexão sobre esses textos elaborados no campo, em momento posterior à aula de campo.

Essa pesquisa propõe a plataforma CAPCOM com recursos digitais que suportem atividades de reflexão sobre as UPIs criadas em campo, de modo a permitir novas oportunidades de aprendizagem. Entre as atividades citadas estão: consultar, manipular e combinar, em diferentes modos definidos pelo professor, os registros (UPIs) produzidos em aulas de campo que tenham sido feitos com o uso dos recursos da plataforma AM. Sendo assim, a plataforma CAPCOM permite a criação de arquiteturas pedagógicas com atividades que são complementares e posteriores às atividades realizadas com a plataforma AM.

A plataforma AM [Pereira et al. 2015] fornece recursos digitais para suporte às atividades pedagógicas em aulas de campo. Com ela, um educador pode delimitar a área onde a atividade de campo será realizada, definir subáreas onde informações diferenciadas podem ser acessadas, definir regras de comportamento para os alunos em cada uma das subáreas, acompanhar o trajeto de cada aluno da turma e etc.

Cada aluno, por sua vez, com o uso de um dispositivo móvel, pode criar registros multimídia locativos (UPILocs) ao longo do trajeto que realizar, acessar as informações disponíveis em cada subárea que estiver visitando, interagir com colegas e etc.

A plataforma CAPCOM - Construtor de Arquiteturas Pedagógicas para Dinâmicas Colaborativas com Multimídia dispõe de uma Biblioteca de Recursos Digitais. A Figura 1 apresenta os componentes da plataforma CAPCOM. Também apresenta os atores importantes no cenário de uso dessa plataforma: o professor que usa a CAPCOM para criar as APs, configurar os recursos digitais e visitar os textos (UPIs), produzidos pelos alunos; os alunos que usam os recursos digitais para desenvolverem colaborativamente as atividades propostas em cada AP e criarem novos textos multimídia (UPIs) e os visitantes que usam os recursos digitais para visualizarem os textos (UPIs) produzidos.

Para criar uma Arquitetura Pedagógica (AP) na plataforma CAPCOM, o professor descreve o objetivo pedagógico, a atividade pedagógica e a metodologia, em texto escrito. Depois de cadastrar essas informações, o professor seleciona e configura os recursos disponíveis que vão compor os Recursos Digitais da AP, podendo selecionar mapa mental, composição de mídias, grupos de usuários, classificação de UPI etc.

O professor usa o CAPCOM para construir arquiteturas pedagógicas colaborativas. Os aprendizes executam cada AP (arquitetura pedagógica), usando cada recurso digital alocado pelo professor na AP, para desenvolverem as atividades pedagógicas propostas.

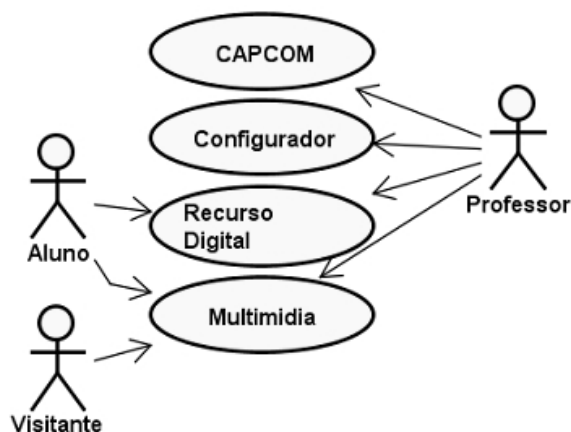


Figura 1. Diagrama de casos de uso para a CAPCOM.

A CAPCOM permite que o professor configure os Recursos Digitais para cada AP criada por ele, atendendo as necessidades específicas de uma aula de campo.

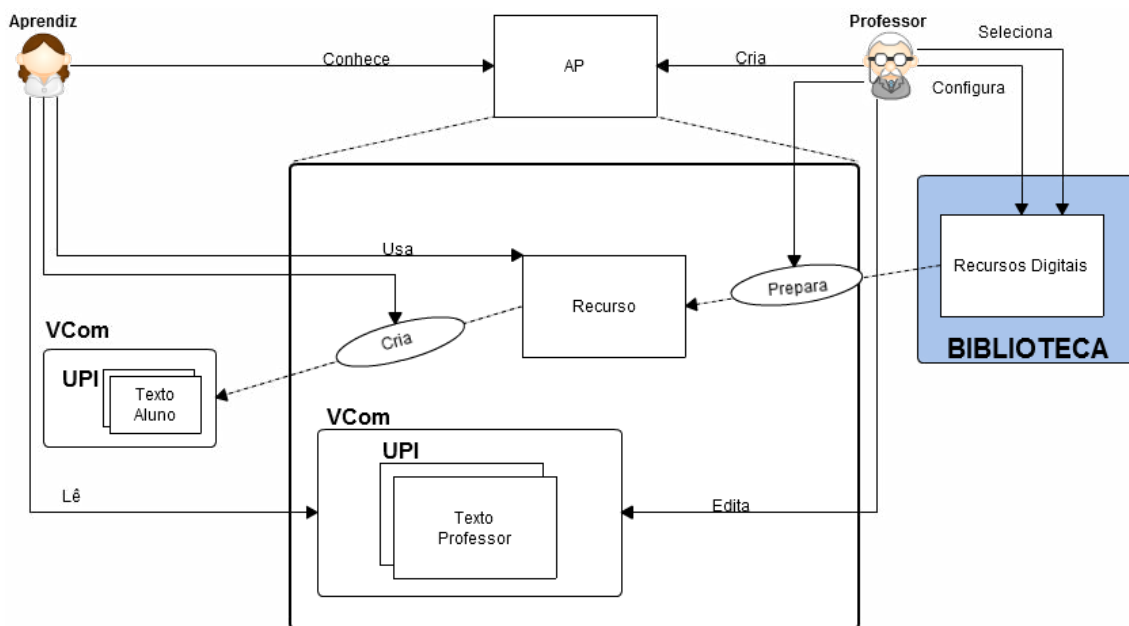


Figura 2. Arquitetura e funcionamento da plataforma CAPCOM.

A Figura 2 apresenta a arquitetura da plataforma CAPCOM. A CAPCOM permite que o professor configure cada Recurso Digital de uma AP e define as regras e a forma da atividade - VCom. Já os alunos podem processar os registros produzidos na aula de campo - UPI - na atividade proposta pelo professor, de modo colaborativo, e modificar o aspecto visual do conteúdo produzido na atividade - Template de VCom - também colaborativamente.

A CAPCOM implementa diversos editores que permitem a manipulação de UPIs para compor novas UPIs (UPIs compostas) ou a criação de VComs (VCLocs) ou templates, de modo a permitir novas formas de organizar e apresentar as UPIs criadas na aula de campo. Tudo isso em atividades realizadas após a aula de campo.

Tabela 1. Distribuição e uso dos recursos digitais da biblioteca do CAPCOM.

Recursos	Arquiteturas Pedagógicas		
	AP1	AP2	AP3
Edição de mapas mentais	Sim	Sim	Sim
Suporte a grupos de colaboração	Sim	Sim	Sim
Edição de UPIs	Sim	Não	Sim
Negociação de UPIs	Não	Sim	Não
Registro de UPIs	Não	Não	Sim

Baseadas nas APs apresentadas na seção 5, foram levantados os seguintes recursos digitais que compõem a biblioteca da CAPCOM:

1. Edição de mapas mentais
2. Suporte a grupos de colaboração
3. Edição de UPIs
4. Negociação de UPIs
5. Registro de UPIs

A Tabela 1 apresenta a distribuição e uso dos recursos da biblioteca da plataforma CAPCOM para criação e uso das APs.

Usando as UPIs de uma aula de campo, o professor pode definir:

- regras sobre o uso delas (um aluno pode usar apenas suas UPIs ou pode usar UPIs de colegas);
- grupos de usuários para colaborar no uso dessas UPIs (textos em múltiplas mídias, como fotos, textos escritos, vídeos, áudios etc) e na criação de novas UPIs compostas;
- a atividade como individual e habilitar a colaboração apenas em outra fase da atividade;
- se a AP terá como foco a composição de novas UPIs ou a classificação das UPIs criadas.

Os recursos digitais configurados pelo professor para cada arquitetura pedagógica, permitem ao aluno articular as UPIs que foram criadas durante a aula de campo, de modo a produzir novos registros, a partir da associação dessas UPIs de diversas formas.

O resultado da manipulação das UPIs pelo aluno fica disponível em outra parte da CAPCOM, onde o professor e o aluno podem ter acesso. A Figura 3 apresenta um exemplo de UPI criada por recurso digital da plataforma CAPCOM, onde é incorporado um menu para interligar a UPI da figura com outras UPIs.

8. Trabalhos Correlatos

A Tabela 2 apresenta uma comparação da plataforma CAPCOM com outros sistemas que também têm o objetivo de ajudar o educador a criar atividades associadas a sua prática pedagógica colaborativa.

Na coluna “Projetado para” é apresentado o(s) modo(s) de uso do sistema: “Móvel” significa que o sistema foi desenvolvido para usar recursos de dispositivos



Figura 3. Um exemplo de UPI gerada em uma AP.

móveis e “Web” indica que o sistema foi construído para usar recursos da Web. A coluna “Principal Uso” apresenta o propósito principal de construção do sistema. A coluna “Colaboração” define se o sistema suporta a colaboração durante a utilização, dentre as diversas possibilidades oferecidas. A coluna “Atividade” indica se o sistema permite a criação de atividades pedagógicas.

Tabela 2. Comparação entre sistemas digitais para atividades pedagógicas.

Sistema	Projetado para	Principal uso	Colaboração	Atividades
MOrFEu	Web	Educacional	Sim	Sim
Aris Games	Móvel	Jogos	Não	Sim
Youubi	Móvel	Educacional	Sim	Sim
Plataforma CAPCOM	Móvel e Web	Educacional	Sim	Sim

A plataforma Youubi, assim como a plataforma CAPCOM, também é projetada para dispositivos móveis, tem suporte à locatividade, tem suporte a colaboração durante as atividades e é voltada para Aplicações Educacionais. Entretanto, na plataforma Youubi as atividades pedagógicas são pré-definidas, o professor não pode ajustar a atividade às necessidades pedagógicas da turma. Já o CAPCOM permite que o professor crie as atividades de acordo com sua proposta pedagógica.

9. Considerações Finais

De modo complementar à proposta do AM, este trabalho apresentou a plataforma CAPCOM (Construtor de Arquiteturas Pedagógicas Com textos Multimídia), que permite a criação e uso de arquiteturas pedagógicas (APs) para a realização de atividades de aprendizagem colaborativa, com o uso de textos multimídia (UPIs), produzidos durante atividades de campo.

O uso das APs proporciona novas oportunidades de aprendizado, aproveitando o potencial pedagógico dos registros (UPIs) produzidos e armazenados nas atividades de campo, com o uso da plataforma AM (Aprendizagem em Movimento) [Pereira et al. 2015].

A plataforma CAPCOM foi criada para completar o ciclo pedagógico de uma aula de campo que é constituído de: planejamento da aula de campo, a aula propriamente

ditada e as atividades de reflexão sobre os dados coletados em campo. A plataforma CAPCOM fornece recursos para criar, configurar e usar arquiteturas pedagógicas que tratem de atividades posteriores à aula de campo.

O artigo apresentou três exemplos de arquiteturas pedagógicas criadas pela plataforma CAPCOM e contextualizou o uso de cada uma dessas APs, bem como, descreveu alguns aspectos da dinâmica colaborativa entre os alunos realizadas durante o uso de cada AP.

Em continuidade à pesquisa descrita aqui, pretende-se ampliar o uso dos recursos digitais da CAPCOM em dispositivos móveis com iOS e Windows Phone. Também pretende-se ampliar as possibilidades de textos que a plataforma pode trabalhar, incluindo vídeos esféricos e outras mídias ainda pouco exploradas, mas que têm grande potencial para o propósito pedagógico.

Referências

- Carvalho, M. J. S., Nevado, R. A. d., and Menezes, C. d. (2005). Arquiteturas pedagógicas para educação a distância: concepções e suporte telemático. *Anais do XVI Simpósio Brasileiro de Informática na Educação, Juiz de Fora-MG. Brasil*, pages 351–360.
- Castro, A. and Menezes, C. (2011). Aprendizagem colaborativa com suporte computacional. *Sistemas Colaborativos*, pages 135–153.
- Chagas, I. (2002). Trabalho colaborativo condição necessária para a sustentabilidade das redes de aprendizagem. *Redes de aprendizagem. Redes de conhecimento. Lisboa: Conselho Nacional de Educação*, pages 71–82.
- Dillenbourg, P., Baker, M. J., Blaye, A., and O'Malley, C. (1995). The evolution of research on collaborative learning. *Learning in Humans and Machine: Towards an interdisciplinary learning science.*, pages 189–211.
- Giordani, A. C. C., Audino, D. F., and Cassol, R. (2006). Inserção do google earth no ensino de geografia. *IN: Jornada da Educação, UNIFRA, Centro Universitário Franciscano, Santa Maria/RS*, pages 1–8.
- Lei, P.-L., Kao, G. Y.-M., Lin, S. S., and Sun, C.-T. (2009). Impacts of geographical knowledge, spatial ability and environmental cognition on image searches supported by gis software. *Computers in Human Behavior*, 25(6):1270–1279.
- Leite, C. L. K., PASSOS, M. d. A., Torres, P. L., and Alcântara, P. R. (2005). A aprendizagem colaborativa na educação a distância on-line. In *CONGRESSO INTERNACIONAL DE EDUCAÇÃO A DISTÂNCIA*, volume 12, pages 1–10.
- Menezes, C. S., de Nevado, R. A., de Castro Jr, A. N., and Santos, L. N. (2008). Morfeu—multi-organizador flexível de espaços virtuais para apoiar a inovação pedagógica em ead. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 1, pages 451–460.
- Ocepek, U., Bosnić, Z., Šerbec, I. N., and Rugelj, J. (2013). Exploring the relation between learning style models and preferred multimedia types. *Computers & Education*, 69:343–355.

- Oliveira, A. and Correia, M. D. (2013). Aula de campo como mecanismo facilitador do ensino-aprendizagem sobre os ecossistemas recifais em alagoas. *Alexandria Revista de Educação em Ciência e Tecnologia*, 6(2):163–190.
- Pereira, B., Almeida, R., Tavares, O., and de Menezes, C. (2015). Am: Uma plataforma digital para criação e uso de veículos de comunicação locativos para atividades pedagógicas. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, pages 131–139.
- Pereira, B. C. (2015). Aprendizagem em movimento - uma plataforma para a criação e uso de veículos de comunicação na realização de arquiteturas pedagógicas. Master's thesis, Programa de pós graduação em em Informática da UFES, Vitória–ES.
- Pimentel, M., Gerosa, M. A., Filippo, D., Raposo, A., Fuks, H., and Lucena, C. J. P. (2006). Modelo 3c de colaboração para o desenvolvimento de sistemas colaborativos. *Anais do III Simpósio Brasileiro de Sistemas Colaborativos*, pages 58–67.
- Reinoso, L. and Tavares, O. (2015). Mvlibras: ambiente digital para comunidades de aprendizagem com recursos inclusivos para surdos. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, pages 772–781.
- Seniciato, T. and Cavassan, O. (2004). Aulas de campo em ambientes naturais e aprendizagem em ciências—um estudo com alunos do ensino fundamental field classes in natural environment and science learning—a study with students from elementary school. *Ciência & Educação*, 10(1):133–147.
- Vieira, R. R. M. (2011). Uma arquitetura de software para o morfeu: Apoiando a realização de arquiteturas pedagógicas em espaços virtuais colaborativos. Master's thesis, Dissertação (Mestrado em Informática)—UFES, Vitória–ES, Agosto.
- Vieira, R. R. M., dos Santos, O. L., do Prado Rafalski, J., Bada, E. M., da Silva, H. F. A., and de Menezes, C. S. (2011). Coordenação nas atividades colaborativas em ambientes de aprendizagem—uma avaliação na implementação de arquiteturas pedagógicas. *RENOTE*, 9(1).
- Vieira, R. R. M. and Tavares, O. d. L. (2015). Modelo para descrição de espaços colaborativos: Apoiando a realização de arquiteturas pedagógicas. *Sistemas Colaborativos*, pages 189–196.

Uma tecnologia social baseada em *Crowdsourcing* para busca de pessoas desaparecidas e cadastro de moradores de rua

Priscila Francielle Knoop Silveira¹, Rodrigo Remor Oliveira¹, Rodrigo Ruas Oliveira²

¹IFSul – Campus Sapucaia do Sul, ²UFRGS

priscilaknoop@outlook.com, remor@sapucaia.ifsul.edu.br,
ruasoliveira@inf.ufrgs.br

Resumo. Este artigo descreve Vox, uma plataforma que amplia a divulgação de desaparecidos, moradores de rua e abrigados, a fim de auxiliar na busca e na reinserção destas pessoas na sociedade. Vox é uma plataforma colaborativa que oferece um banco autossustentável para busca de pessoas. Vox permite a visualização de fotos, cadastro e criação de perfis com atributos físicos detalhados, inserção de descrições acerca das circunstâncias de avistamento e desaparecimento, e mapeamento dos lugares em que a pessoa foi supostamente vista. A plataforma visa a tornar-se um banco de dados nacional e uma referência na busca por desaparecidos, moradores de rua e abrigados.

Abstract. This paper describes a platform--namely Vox-- that increases the dissemination of information regarding missing and homeless people, in an attempt to aid their reintegration into society. Vox is a collaborative platform that provides a self-sustainable database for searching people. Vox allows visualization of photos, registration, the creation profiles with detailed physical attributes, the description of the circumstances in which a person disappeared or was spotted, and the mapping of the places in which the person was allegedly spotted. The project aims at becoming a national database and a reference in the search for missing and homeless people.

1. Introdução

Cerca de 200 mil pessoas desaparecem anualmente no Brasil [Lima, 2011]. As causas são diversas: idosos acometidos por perdas de memória; adolescentes que fogem de casa ou de abrigos e outras instituições públicas; adultos que optam por deixar suas casas, empregos e rotinas; dependentes químicos; além de vítimas fatais de acidentes, calamidades ou crimes [Ferreira, 2013]. Além disso, em uma quantidade significativa dos casos, os desaparecidos não têm capacidade para voltar para casa por conta própria. Em uma pesquisa realizada por Garcia (2011), em 302 famílias, foi observado que aproximadamente 10% dos desaparecidos apresentavam algum tipo de deficiência física ou intelectual. Esse fato é alarmante pois dependendo da deficiência, o desaparecido pode não ser capaz de fornecer informações pessoais ou as de um contato próximo. Nessas situações e em muitas outras, tais como dependência química e idosos com perdas de memória, é bastante comum que os desaparecidos acabem à margem da sociedade, inseridos em populações em situação de rua ou encaminhados a abrigos, aumentando ainda mais a população em situação de vulnerabilidade social presente no país [Mattos

and Ferreira, 2005]. Em um relatório feito pela Delegacia de Desaparecidos de Florianópolis em 2013, a guarda municipal constatou que em uma busca inicial dentre 250 moradores de rua, 40 deles foram identificados como pessoas desaparecidas que estavam sendo procuradas por suas famílias em Florianópolis e pelo restante do estado [Souza, 2013].

Para piorar a situação, a sociedade brasileira é extremamente ineficiente ao tentar lidar com o problema dos desaparecidos. Casos de desaparecimentos são investigados por delegacias, órgãos de administrações públicas e instituições não governamentais, como, por exemplo, a “SOS Desaparecidos”. Entretanto, muitos casos acabam arquivados em delegacias, seja pela crença de que tais pessoas deixaram suas casas sem intenção de retornar, seja pela falta de interligação entre as diversas entidades. A título de comparação, os Estados Unidos possuem o sistema de alerta *America’s Missing: Broadcast Emergency Response*, resultado de um trabalho conjunto entre sociedade civil, meios de comunicação e agências governamentais norte-americanas para os casos de desaparecimento infanto-juvenil. Tal projeto mostrou-se extremamente ágil para os casos de divulgação e busca de crianças desaparecidas [Garcia, 2011].

Portanto, diante da falta de interligação entre instituições e da quantidade imensa de pessoas que desaparecem todos os anos, fica evidente a falta de algo que possibilite a comunicação entre esses diferentes setores. Surge assim o seguinte questionamento: como seria possível correlacionar as informações das pessoas desaparecidas com as dos moradores de rua de uma maneira rápida e viável?

O presente artigo apresenta uma plataforma colaborativa a fim de facilitar a divulgação de desaparecidos, moradores de rua e abrigados, e também auxiliar na busca de desaparecidos e na reinserção de moradores de rua à sociedade. A plataforma, denominada Vox, possui os três seguintes objetivos específicos.

- **Colaboração.** Vox permite que os próprios usuários do sistema insiram informações sobre as pessoas, estejam elas desaparecidas, em situação de rua, ou abrigadas. Com isso, Vox pretende que a própria sociedade coopere com a procura e reinserção de pessoas à mesma. Além disso, Vox oferece dois possíveis tipos de cadastro: cadastro de pessoas (com descrição precisa ou aproximada da mesma) e cadastro de avistamentos (com indicação de local e descrição da circunstância).
- **Busca rápida e facilitada.** Vox possui um banco de dados com opções de busca de desaparecidos e correlação automática de informações acessível a qualquer pessoa com um navegador ou celular. Dessa forma, Vox pretende agilizar a reinserção das pessoas na sociedade e diminuir a invisibilidade social da população de moradores de rua
- **Comunicação.** Vox auxilia na comunicação entre diversas instituições, sejam elas governamentais ou não, e incentiva o uso de plataformas *crowdsourcing* sociais. Em outras palavras, Vox visa a incentivar o conceito de plataformas autoalimentadas pela comunidade e aplicadas à mesma na resolução de um problema social.

Primeiramente, o projeto pretende permitir que sejam cadastradas diversas pessoas, estejam elas desaparecidas, em situação de rua, ou abrigadas. Com isso, possibilitar que a própria sociedade coopere com a procura e reinserção de pessoas à sociedade, seja através de cadastro de pessoas, assim como cadastro de avistamentos.

Pretende também tornar a busca de desaparecidos e a correlação de informações mais rápida e prática, diminuindo a invisibilidade social da população em situação de rua, e incentivar sua reinserção à sociedade. Auxiliar na comunicação entre diversas instituições, sejam elas governamentais ou não, e incentivar o uso de plataformas *Crowdsourcing*, ou seja, a utilização do conceito em que as informações da plataforma são fornecidas pela própria comunidade e incentivar o uso de tecnologias sociais, isto é, a utilização de plataformas aplicadas para a resolução de um problema social.

2. Trabalhos relacionados

Entre os sistemas relacionados encontrados destacam-se diversos sites que fazem a divulgação de desaparecidos, e alguns aplicativos com o mesmo fim. Porém, até o momento não há qualquer site que divulgue fotos de moradores de rua ou abrigados, e tampouco existem aplicativos voltados para essa população. Na Tabela 1 podemos ver uma comparação entre sistemas similares ao proposto pela plataforma, considerando aspectos tais como: se permite a divulgação de pessoas desaparecidas; se correlaciona as informações das pessoas desaparecidas com as dos moradores de rua; se o sistema é aberto para qualquer usuário fazer cadastros (colaborativo); se possui um banco de dados nacional para inserção de qualquer pessoa; e se permite que um usuário informe um avistamento de uma pessoa desaparecida.

Tabela 1. Tabela comparativa de sistemas relacionados

	Divulgação de desaparecidos	Correlação das informações	Sistema colaborativo	Banco de dados nacional	Cadastro de avistamentos
PLID	✓	✗	✗	✗	✗
ReDESAP	✓	✗	✓	✗	✗
Sinesp Cidadão	✓	✗	✗	✗	✗
Find People	✓	✗	✓	✓	✗

O Programa de Localização e Identificação de Desaparecidos (PLID), desenvolvido pelo Núcleo de Apuração Criminal (NAC) do Ministério Público do Estado do Rio de Janeiro, trata de um banco de dados de desaparecidos, que procura auxiliar na busca e identificação destas pessoas. Ele cruza informações de diversos órgãos, como o Detran, o Disque Denúncia, o Departamento de Polícia Rodoviária Federal, e a Polícia Civil e Militar do Rio de Janeiro, entre outros [MP-RJ, 2015]. Distancia-se do projeto aqui abordado por ser voltado para o Rio de Janeiro e abordar apenas o tema de desaparecidos, não aderindo a instituições que acolhem moradores de rua.

O aplicativo Sinesp Cidadão, disponibilizado pela Secretaria Nacional de Segurança Pública do Ministério da Justiça, possui a funcionalidade de acesso a informações sobre pessoas desaparecidas. As informações utilizadas no aplicativo pertencem ao banco de dados de delegacias especializadas de alguns estados, sendo que três estados aderiram à aplicação até o momento [Ministério da Justiça, 2015]. Distancia-se do projeto por não considerar a relação com moradores de rua, e também por ser dependente das delegacias de desaparecidos. Isso faz com que o usuário comum não possa cadastrar desaparecidos de forma independente, e torna o aplicativo obsoleto em estados que não aderiram ao projeto.

A Rede Nacional de Identificação e Localização de Crianças e Adolescentes Desaparecidos (ReDESAP), é um site voltado também para a divulgação de desaparecimentos, focando em crianças e adolescentes. É um site-cadastro, que objetiva uma base única de dados de crianças e adolescentes desaparecidos [Secretaria de Direitos Humanos, 2015]. O site diferencia-se do projeto por abranger apenas crianças e adolescentes desaparecidos, ou seja, não considera adultos e a relação de desaparecidos com a população em situação de rua.

O *Find People* é um aplicativo produzido pela *startup IlhaSoft*, que visa auxiliar na busca de desaparecidos através de um banco nacional. Qualquer pessoa pode cadastrar um desaparecido, e o aplicativo traz um mapa com pessoas cadastradas perto do usuário [Campi, 2013]. O aplicativo permite ao usuário adicionar novos desaparecidos, mas também se distancia do projeto por não considerar a relação com moradores de rua. Apesar disso, a função de disponibilizar um mapa com pessoas cadastradas perto do usuário torna o aplicativo mais interativo que os demais.

Dentre os trabalhos apresentados, é possível verificar que nenhum possibilita a correlação de informações sobre desaparecidos com as de moradores de rua e abrigados. Também se verifica a necessidade de permitir que a população geral tenha um papel mais ativo na busca de pessoas. Isso seria viabilizado ao possibilitar que qualquer usuário cadastre avistamentos de pessoas que estão desaparecidas. Acredita-se que pelos motivos apresentados, nenhum dos trabalhos relacionados tornou-se um banco nacional para busca de pessoas. Diferentemente desses trabalhos, o *Vox* atende à todas as características analisadas.

3. A plataforma Vox

Esta seção descreve a plataforma *Vox*. Inicia-se com um resumo da pesquisa de campo efetuada para amparar o planejamento da plataforma. A seguir, apresenta-se uma visão geral da plataforma. Após, detalha-se cada um dos componentes da arquitetura.

3.1 Pesquisa de Campo

Para alcançar os objetivos da plataforma *Vox*, foram realizadas pesquisas de aprofundamento sobre este problema social para que a plataforma se adapte da melhor maneira possível ao o público alvo. Além da pesquisa dos trabalhos similares apresentados, foi realizada uma visita ao abrigo Bom Pastor, em Novo Hamburgo, RS, com o intuito de validar a proposta do projeto e encontrar possíveis pontos que necessitem de modificações.

A visita efetuada confirmou a necessidade de inserir informações físicas detalhadas das pessoas desaparecidas, para que ao adicionar um novo morador, seja possível verificar as pessoas com características similares, e possivelmente conseguir uma identificação. Além disso, também se se incluiu outros aspectos na busca (além do nome), tais como região.

3.2 Visão Geral

A plataforma usa o conceito de *Crowdsourcing*, que se encaixa no problema em questão, fazendo assim, a plataforma um projeto colaborativo. Sendo uma plataforma *Crowdsourcing*, o projeto propõe a cooperação e a comunicação direta entre familiares

de desaparecidos, moradores de rua, ONG's e diversas instituições, sejam elas governamentais ou não.

Como pode ser visto na Figura 1, a arquitetura da plataforma Vox consiste em basicamente três componentes: o *Vox mobile* que representa um aplicativo desenvolvido para *Android*, o *Vox web*, que representa a aplicação Web, e o *Vox server* que representa o servidor que centraliza as informações oriundas do *Vox mobile* e *Vox web*.

A arquitetura foi planejada para suportar várias conexões ao mesmo tempo. A plataforma emprega *Restfull Web Services* através do framework *Jersey*, proporcionando rapidez no consumo e envio de dados entre os clientes e o *Vox server*. Além disso, o *Restfull Web Services* traz várias vantagens para a plataforma, tais como a reutilização de software, aumento da produtividade, interoperabilidade e escalabilidade. O banco de dados utilizado foi o MySQL, uma plataforma gratuita comumente utilizada no mercado. Para a persistência de dados no banco foi utilizado o framework *Hibernate*, juntamente com a especificação JPA (*Java Persistence API*), proporcionando assim maior facilidade ao trabalhar com bancos de dados relacionais e orientação a objetos. As próximas subseções descrevem em detalhes cada componente do modelo.

As requisições efetuadas pelos clientes são trafegadas utilizando o protocolo HTTP (*Hypertext Transfer Protocol*) e utilizando JSON (*JavaScript Object Notation*) para uma formatação leve de troca de dados, proporcionando rapidez à plataforma.

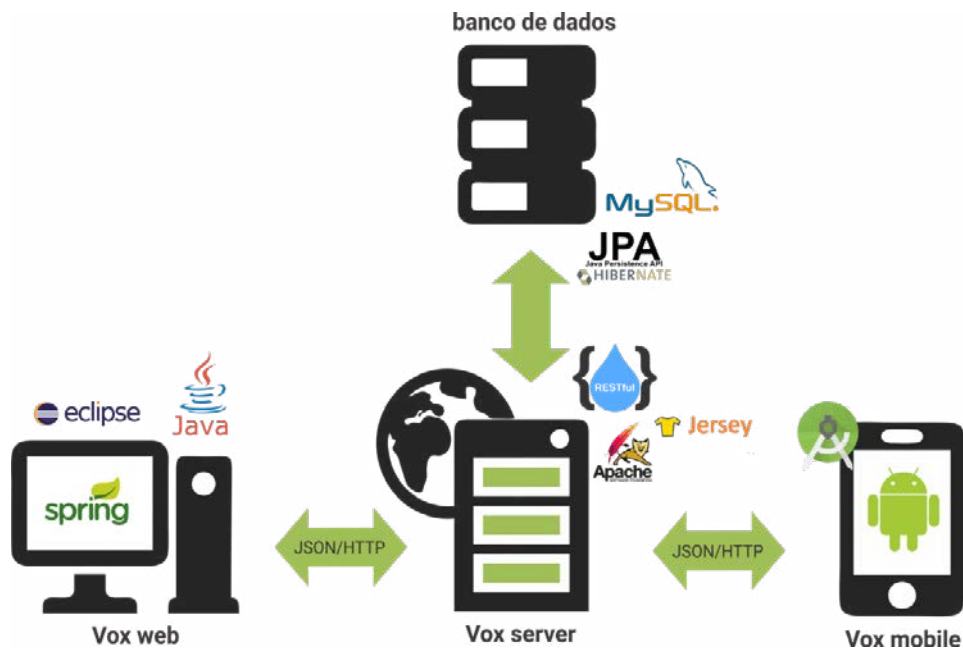


Figura 1. Arquitetura do sistema

3.3 Vox server

Para possibilitar a persistência de dados da aplicação foram utilizadas diversas ferramentas. Com a utilização dessas tecnologias foi possível que as classes demarcadas com a anotação *@Entity* fossem facilmente persistidas no banco, e que os relacionamentos entre classes também fossem facilmente definidos. É nessa camada que se define que cada objeto da classe *Usuario* possui um relacionamento com a classe *PessoaProcurada*, e que cada novo cadastro de pessoas deve informar qual usuário foi responsável pela ação.

Com a finalidade de modelar as informações que serão armazenadas no banco de dados, foi criado um modelo lógico representado as tabelas necessárias ao sistema, seus atributos e seus relacionamentos. O modelo está ilustrado na Figura 2. No diagrama entidade relacionamento é possível observar que cada usuário (*Usuario*) pode possuir diversas informações de contato (*InfoContato*), e que um usuário pode cadastrar diversos desaparecidos, moradores de rua e abrigados (*PessoaProcurada*), assim como registros de avistamentos (*Avistamentos*) para uma determinada pessoa. Também é possível observar que um registro de avistamento obrigatoriamente deve possuir circunstâncias detalhadas (*Circunstancia*) e uma localização (*Localizacao*).

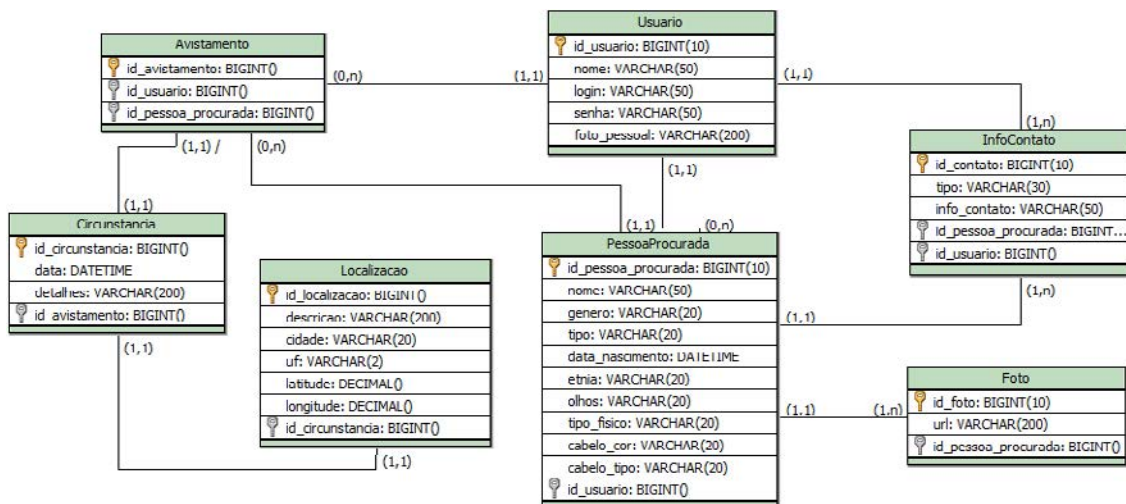


Figura 2. Modelo lógico do banco de dados

A camada de persistência está presente no *Vox server* componente que faz com que seja possível realizar a comunicação com o banco de dados. Esse componente está dividido em classes *DAO's*, *Pojos* e *Services*. A seguir será explicado mais sobre a camada de *Services* desse componente.

O padrão de persistência de dados DAO (*Data Access Object*) permite separar as regras de negócio das regras de acesso ao banco de dados. Todas as funcionalidades de um banco de dados, tais como inserir, buscar, atualizar e deletar objetos são manuseadas nessa parte da aplicação. Nela também estão todas as instruções, métodos, consultas e funções que possibilitam que a plataforma *Vox* insira novas pessoas e consiga acessá-las dependendo da demanda.

A plataforma realiza a comunicação entre os componentes através de um *Web Services* baseados em *REST*. Este por sua vez, utiliza o protocolo *HTTP* como protocolo de aplicação. Essa camada da aplicação proporciona uma comunicação rápida entre serviços e clientes, através de requisições que são transmitidas e recebidas através do formato *JSON*. Os recursos são dispostos em um servidor *REST* através de diversos links que levam o usuário à diversos estados da aplicação.

Para criar um *Web Service Restful* foi usada a especificação *JAX-RS (Java API for RESTful Web Services)*, utilizada juntamente com a implementação *Jersey*. O *JAX-RS* possui um foco em *POJOS (Plain Old Java Object)*, oferecendo um conjunto de anotações que podem ser utilizadas. O *HTTP* é assumido como protocolo de aplicação, permitindo a utilização dos diferentes métodos de requisição utilizados no protocolo, tais como *GET*, *POST*, *PUT* e *DELETE*. A API também permite definir diversos tipos de

conteúdo que os métodos produzem e consomem, tais como XML ou JSON. Para utilizar facilmente esses tipos de conteúdo, a aplicação também utiliza a JAXB (*Java Architecture for XML Binding*), que fornece uma API, as ferramentas e um framework para mapear objetos Java e documentos XML ou JSON.

No *Vox server* há diversos recursos que utilizam diferentes métodos de requisições e representam diferentes ações na aplicação. Por exemplo, no cadastro de novas pessoas procuradas do sistema *Vox*, é utilizado um método POST para que quando o *Vox mobile* acesse a URI “/pessoaProcurada” o servidor insira uma nova pessoa no banco de dados. Todas as requisições utilizam o formato de dados JSON, reconhecido por sua simplicidade e rapidez na troca de dados.

Sendo assim, além dessas ferramentas facilitarem o desenvolvimento da plataforma, também possibilitam uma facilidade de compreensão e manutenção do código. Um Web Service também independe de linguagens de programação, possibilitando uma interoperabilidade na aplicação e uma reusabilidade, pois seus recursos podem ser utilizados em diferentes aplicações.

3.4 Vox mobile

O componente *Vox mobile* é o componente apresentado ao usuário final, que permite acesso às funcionalidades da plataforma. Entre elas, o cadastro de contas, o cadastro de pessoas desaparecidas, moradores de rua e abrigados e também o cadastro de novos avistamentos dessas pessoas. Também a visualização de perfis, mapas com avistamentos recentes e a pesquisa por certas características.

Para que o *Vox mobile* faça o consumo do *Vox server* está sendo utilizada a classe *RestTemplate* do Spring framework, responsável pelo consumo do *Vox server*. O uso dessa classe simplifica o consumo do *Vox server*, sendo necessário informar apenas a URL do serviço requisitado, o tipo de retorno, o método de requisição do recurso e parâmetros quando existentes. Por exemplo, quando um usuário do *Vox mobile* entra no perfil de uma pessoa procurada, a aplicação requisita o método *buscarPessoaPorId()*. Para consumir esse método do *Vox server*, é necessário utilizar o método *getForObject* (URL, *PessoaProcurada.class*, id). Ao consumir esse método, também é necessário que o *Vox mobile* esteja preparado para receber as informações retornadas pelo *Vox server*. Para receber e enviar as informações no formato *Json* está sendo utilizada a biblioteca *Google Gson* em algumas partes do código para a serialização e deserialização de objetos. A biblioteca oferece métodos simples para a conversão de objetos (*toJson()* e *fromJson()*), de objetos Java para representação *Json*, e vice e versa. A biblioteca permite, inclusive, a conversão de objetos complexos.

Para melhorar a experiência do usuário final, o *Vox mobile* utiliza o Material Design, um guia para design visual definido pela Google que pretende simplificar e unificar o layout dos aplicativos. O padrão inclui o uso de paletas de cores, ícones, animações, tipografias e hierarquias definidas pelo Material Design, objetivando que os aplicativos se tornem mais simples, claros e elegantes. O padrão também define recursos animados no uso de sombras e perspectiva, baseado na ideia de que um elemento flutua pela aplicação com sombra realista e profundidade próximas às propriedades de um material na vida real. O uso do padrão melhora a experiência do usuário, pois proporciona uma interface elegante e simples. Além disso, proporciona a sensação de familiaridade ao usuário pela proximidade do layout presente nas próprias ferramentas da Google.

Além disso, a aplicação utiliza a API do *Google Places* para que o usuário possa adicionar novos avistamentos através de um campo *autoComplete* para endereços. Ao usuário digitar o endereço de um avistamento, ele recebe uma lista de possíveis endereços. Ao clicar em um item, o *Vox mobile* recebe a descrição do local, a cidade, estado e coordenadas, fazendo com que se consiga diversos dados com poucos cliques do usuário. O *Vox mobile* utiliza também a API do *Google Maps* para permitir a visualização de mapas nos perfis de pessoas procuradas e para a colocação de pontos que representam as últimas localizações dessas pessoas.

3.5 Vox web

A plataforma Vox possui uma aplicação Web para fins de divulgação do aplicativo *Android* e suas funcionalidades. Também possui informações sobre a plataforma e seus objetivos.

O framework utilizado neste componente é o Spring MVC, que segue o padrão MVC (*Model View Controller*), que é um modelo de arquitetura de software que separa a representação da informação e a interação do usuário. O modelo possui três camadas: *Model*, *View* e *Controller*. A camada de *Models* consiste no encapsulamento dos dados da aplicação; A camada de *Views* é responsável por solicitar informações das *Models* necessárias para gerar representações de saída; A camada de *Controllers* é responsável por receber as requisições do usuário e executar determinadas ações.

Sempre que um usuário acessa a URL da página principal do *Vox web*, essa ação gera uma requisição para a *HomeController*, que recebe uma entrada e decide como processá-la. No caso dessa ação necessitar de acesso as regras de negócio da aplicação, a *Controller* requisita uma *Model* para tratar dessas regras. Ao fim disso, a *HomeController* requisita uma *View* para apresentar a saída do processo, enviando um retorno ao usuário na forma de um documento HTML referente a página inicial da aplicação.

As ideias principais do MVC são a reusabilidade de códigos e a separação dos diferentes conceitos utilizados, pois permite a separação da lógica de negócio, apresentação e navegação. A Figura 3 a seguir exemplifica uma requisição feita por um usuário, a comunicação da *Controller* com a *Model* para busca de informações, e a resposta retornada ao usuário através de uma *View*.

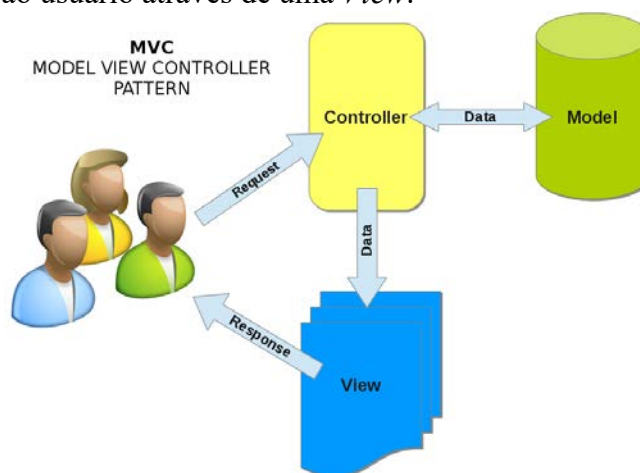


Figura 3. Estrutura de uma requisição utilizando MVC

4. Aspectos de avaliação

Na etapa atual de avaliação do projeto, pretende-se apresentar o sistema em abrigos, albergues e instituições que possam estar interessadas na sua utilização. O abrigo e albergue Bom Pastor, em NH, demonstrou interesse no projeto, e pretende-se apresentar o protótipo desenvolvido a outros locais que possam se beneficiar do uso da ferramenta.

Após a etapa de implementação do projeto, ocorrerá uma etapa de avaliação do protótipo. Pretende-se avaliar o protótipo através de questionários e entrevistas, considerando alguns critérios específicos. O protótipo desenvolvido será avaliado utilizando o *Technology Acceptance Model*, mais conhecido como modelo de aceitação de tecnologia (TAM). O TAM está fundamentado basicamente em dois construtos: a utilidade percebida e a facilidade de uso percebida, sendo que ambos medem completamente os efeitos das variáveis externas, como características do sistema, processo de desenvolvimento, treinamento, na intenção de uso (Silva, Pimentel, Soares; 2012). Sendo assim, os dados coletados em questionários e entrevistas serão avaliados utilizando esses dois critérios, ou seja, no quanto o usuário acredita que a ferramenta é útil e de fácil usabilidade.

5. Discussão

O Vox proporciona o cadastro de contas de usuários, que podem representar indivíduos ou uma organização. Para um usuário cadastrar uma conta é necessário que sejam informados alguns campos básicos, como e-mail, login e senha. Posteriormente um usuário pode alterar dados e cadastrar novas informações de contato. Usuário já cadastrados podem logar na plataforma informando o login e senha, e também podem fazer logout. Na Figura 4 está ilustrada a tela de login de usuários, que permite que o usuário escolha entre efetuar login na plataforma caso já seja cadastrado, ou cadastrar-se.

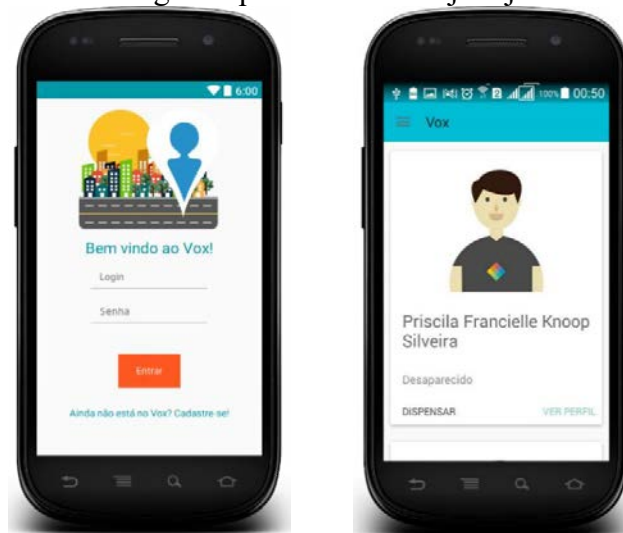


Figura 4. Tela de login e tela principal de divulgação de pessoas

A plataforma proporciona a divulgação de pessoas desaparecidas, moradores de rua e abrigados de acordo a localização do usuário logado. Além disso, na tela principal da aplicação são divulgadas pessoas a partir do critério de ordem de avistamento, ou seja, são exibidos no topo da lista pessoas cadastradas recentemente, ou que foram avistadas recentemente. Nessa tela o usuário possui a opção de “descartar” uma pessoa, isto é,

passar o perfil e visualizar o seguinte, ou de visualizar suas informações na tela de perfil. Tela ilustrada na Figura 4.

A plataforma permite a visualização de perfis detalhados, com informações físicas e avistamentos mais recentes de pessoas desaparecidas, moradores de rua e abrigados. Permite também que qualquer usuário cadastre um desses perfis, informando características físicas detalhadas da pessoa a ser adicionada e o último avistamento dela. (Figura 5)

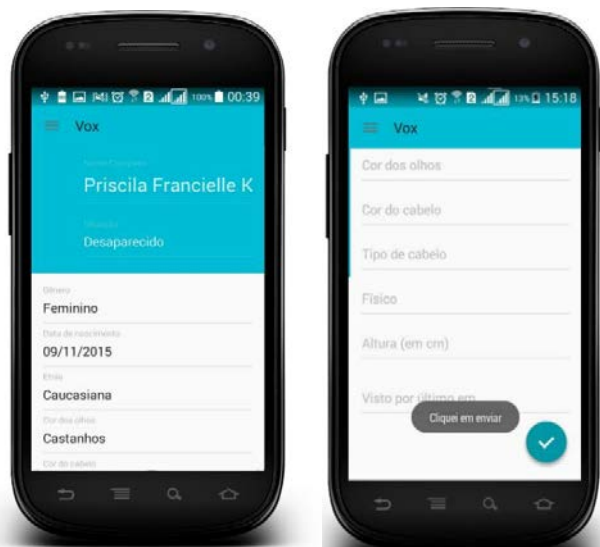


Figura 5. Tela de perfil e cadastro de pessoas

A plataforma permite que qualquer usuário cadastre um avistamento de uma pessoa desaparecida, morador de rua ou abrigado. Para isso, basta informar a data em que a pessoa foi supostamente vista, em que localidade ela estava e fornecer detalhes sobre as circunstâncias do avistamento. Cada usuário também possui um tipo de *feed*, em que é possível visualizar todas as pessoas que cadastrou, juntamente com a última vez em que elas foram vistas. Telas ilustradas na Figura 6.

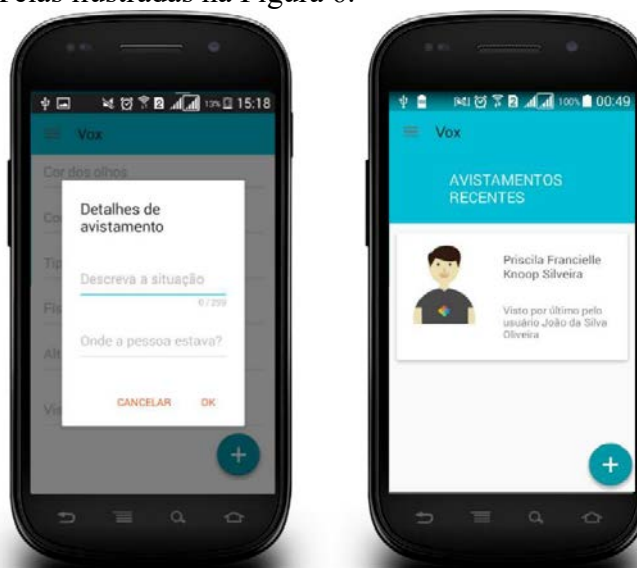


Figura 6. Telas de cadastro de avistamentos e avistamentos recentes

6. Conclusões e Trabalhos Futuros

Com a presente pesquisa foi possível identificar que o tema do é pouco abordado e que não existem aplicações voltadas para o problema específico da relação entre pessoas desaparecidas e moradores de rua. Assim, este trabalho propôs uma plataforma para a diminuição de um problema relevante na sociedade por meio de uma tecnologia social.

Foram desenvolvidas funcionalidades visando facilitar a busca por desaparecidos e também diminuir a invisibilidade social da população de rua. A criação da plataforma colaborativa Vox possibilita uma interação de diversos setores, e também permite que a população civil tenha um papel mais ativo no processo de busca de pessoas. A fim de cumprir com os objetivos propostos, foram implementadas funcionalidades que possibilitam diminuir do impacto do problema apresentado através de um banco de dados nacional para busca de pessoas.

Dentre as tecnologias utilizadas, foi possível identificar que com a utilização do Web Service REST foi possível alcançar uma rapidez na troca de dados, essencial para aplicações móveis. Através do *Framework Hibernate* a persistência de dados foi desenvolvida de forma simples e prática. Quanto ao layout da plataforma, a utilização do padrão de layout Material Design proporcionou simplicidade e elegância à aplicação.

Atualmente, o projeto encontra-se em fase de avaliação e posteriormente irá para fase de divulgação e implementação em instituições de todos os estados brasileiros, a fim de acelerar o processo de criação de um banco nacional para busca de pessoas.

Entende-se que há dificuldades enfrentadas quando o proposto é uma plataforma colaborativa. Primeiramente, por possibilitar a inserção de dados pela população, há o problema de confiabilidade nos dados inseridos, que podem muito bem ser inseridos com a intenção de brincadeira ou de dificultar a utilização da aplicação. Além disso, sistemas colaborativos e tecnologias sociais não são imensamente utilizados, pois o fato de ajudar a aplicação e a sociedade não são motivações o suficiente para que um usuário utilize plataformas desse tipo.

Como trabalho futuro pretende-se adicionar funcionalidades que divulguem a plataforma, a tornem mais confiável e que também a tornem mais atrativa para o público em geral. Para isso, seria interessante desenvolver um sistema de avaliações de usuários, para que seja possível avaliar um usuário ou uma informação como confiável ou não, alertando para possíveis dados falsos; a inserção de selos, ou seja, pequenas premiações dentro da aplicação para usuários que sejam ativos na aplicação e de alguma corroboram para o banco de dados de busca, como forma de motivação; e também realizar uma campanha de divulgação do projeto para ONG's e instituições que trabalhem com moradores de rua, abrigados e também desaparecidos ao redor do país, para evitar que a aplicação se torne obsoleta em outros locais. Outro trabalho futuro seria integrar a aplicação com redes sociais, para que assim seja possível que cada perfil possua a opção de compartilhamento (compartilhar pessoas desaparecidas e avistamentos, por exemplo). Além de atrair mais usuários, isso facilitaria a divulgação de pessoas desaparecidas, moradores de rua e abrigados. Conseqüentemente, a adição de mais usuários aumentaria a quantidade de casos de desaparecimentos resolvidos, pois a divulgação não estará limitada à usuários da plataforma.

Referências

- Campi, M. (2013) “App brasileiro ajuda a encontrar pessoas desaparecidas”, <http://exame.abril.com.br/tecnologia/noticias/app-brasileiro-ajuda-a-encontrar-pessoas-desaparecidas?page=1>, agosto. 2015.
- Ferreira, L. (2013) “De problema de família a problema social: notas etnográficas sobre o desaparecimento de pessoas no Brasil contemporâneo”, <http://aa.revues.org/426>, julho. 2015.
- Garcia, C. (2011) “O que é o desaparecimento de crianças e adolescentes? ”, http://www.moreirajr.com.br/revistas.asp?fase=r003&id_materia=4546, junho. 2015.
- Lima, V. (2011) “Desaparecidos: o descaso do Estado”, [http://www2.camara.leg.br/camaranoticias/radio/materias/REPORTAGEM-ESPECIAL/403230-DESAPARECIDOS--O-DESCASO-DO-ESTADO-\(05'46%22\).html](http://www2.camara.leg.br/camaranoticias/radio/materias/REPORTAGEM-ESPECIAL/403230-DESAPARECIDOS--O-DESCASO-DO-ESTADO-(05'46%22).html), junho. 2015.
- Mattos, R. e Ferreira, R. (2005) “O idoso em situação de rua: Sísifo revisitado”, <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-166X2005000100004&lng=en&nrm=iso>, julho. 2015.
- Ministério da Justiça. (2015) “Aplicativo Sinesp Cidadão ganha módulo para localizar pessoas desaparecidas”, <http://www.justica.gov.br/noticias/aplicativo-sinesp-cidadao-ganha-modulo-para-localizar-pessoas-desaparecidas>, julho. 2015.
- Ministério Público do Estado do Rio de Janeiro. (2015) “Programa de Localização e Identificação de Desaparecidos”, <http://plid.mprj.mp.br/apresentacao>, julho. 2015.
- Secretaria de Direitos Humanos. (2015) “Rede Nacional de Identificação e Localização de Crianças e Adolescentes Desaparecidos”, <http://www.desaparecidos.gov.br/index.php/redesap>, julho. 2015.
- Silva, P. e Pimentel, V. e Soares, J. (2012) “A utilização do computador na educação: aplicando o Technology Acceptance Model (TAM).”, <http://www.ies.ufpb.br/ojs/index.php/biblio/article/viewFile/14208/8113>, julho. 2015.
- Souza, C. (2013) “Moradores de rua estão na lista da delegacia de desaparecidos de Florianópolis”, <http://ndonline.com.br/florianopolis/noticias/128969-moradores-de-rua-estao-na-lista-da-delegacia-de-desaparecidos-de-florianopolis.html>, junho. 2015.

Análise das Interações Sociais entre os Participantes de um Curso EaD: Uma Revisão Sistemática da Literatura

Francisco S. Sousa Neto, José Francisco M. Netto, Dhanielly P. R. de Lima

Programa de Pós-Graduação em Informática (PPGI) – Universidade Federal do Amazonas (UFAM) – Manaus – AM – Brasil

{francisconeto, jnetto, dhanielly}@icompufam.edu.br

Abstract. *Nowadays the advance of information technology growing facility of your use, has propitiated in several advance in more areas of knowledge, for example in Distance Education. In this form, this paper presents a Systematic Literature Review (SLR) with proposal to identify and analyze the methods that existing that further the education, with finality the monitoring of students interactions in Learning Management System (LMS). Results of SLR showing the search realized in this ambit use or define methods of Distance Education that being explored and can guide to support future research.*

Resumo. *Atualmente, o avanço da tecnologia da informação e a crescente facilidade de sua utilização têm propiciado diversos avanços nas mais diversas áreas do conhecimento como, por exemplo, na Educação a Distância (EaD). Desta forma, este artigo apresenta uma Revisão Sistemática da Literatura (RSL) com o propósito de identificar e analisar os métodos existentes que promovem a educação, tendo como finalidade o acompanhamento das interações dos estudantes nos Ambientes Virtuais de Aprendizagem (AVA). Os resultados da RSL mostram que as pesquisas realizadas neste âmbito utilizam ou definem métodos para EaD que ainda estão sendo explorados e podem nortear e apoiar pesquisas futuras.*

1. Introdução

Atualmente, a Educação a Distância (EaD) vem ganhando espaço no campo educacional como um novo processo de aprender e ensinar. Esta modalidade de ensino-aprendizagem se apresenta como uma metodologia inovadora para suprir algumas lacunas na educação tradicional, devido aos diversos contextos sociais em que as pessoas vivem.

De acordo com Santos e Oliveira (2011), essa modalidade de ensino rompe a relação face a face entre professores e alunos e, o processo de ensino e aprendizagem ocorrem em ambientes que transcendem o espaço da sala de aula tradicional, realizando-se em outros espaços e tempos que diferem dos marcados pelas escolas convencionais, atendendo as demandas cada vez mais crescentes de segmentos diferenciados da sociedade.

A EaD propõe diversas características específicas no processo de acompanhamento do desempenho dos alunos, dentre elas, destacamos algumas características do Modelo 3C de Colaboração, que envolve a comunicação, coordenação e cooperação. Segundo Pimentel *et al.* (2006), a comunicação se realiza por meio da

troca de mensagens; a coordenação se realiza por meio do gerenciamento de pessoas, atividades e recursos; e a cooperação se realiza por meio de operações num espaço compartilhado para a execução das tarefas.

Podemos citar como exemplo do modelo 3C aplicado ao Ambiente Virtual de Aprendizagem (AVA), a ferramenta de comunicação: bate-papo, cujo objetivo é possibilitar a troca de mensagens entre os membros de um grupo, além disso, ela também contém elementos de coordenação e cooperação que são necessários para organizar e documentar a comunicação (FUKS *et al.* 2007).

Além do modelo 3C outra característica importante no processo educacional é a interação, que pode ser percebida pelo ciclo que engloba o contexto colaborativo, coordenativo e cooperativo por meio da interatividade dos alunos e professores no AVA.

Nessa perspectiva, alguns trabalhos têm um foco especial no aprendizado mais eficiente ou mais atraente através das interações aluno-professor, aluno-aluno. Para fundamentar essa pesquisa buscamos respaldo em elementos teóricos e metodológicos advindos dos trabalhos de Campana (2008); Lima *et al.* (2014); Soller (2001); Jaques (1997); Boff (2008); Bakharia & Dawson (2011); Bremgartner e Netto, (2011); Mora *et al.*, (2014); Lima e Netto (2015) entre outros.

Contudo, a EaD apresenta-nos algumas peculiaridades no âmbito educacional, com isso, nota-se que esta modalidade de ensinar e aprender introduz uma nova concepção educacional. Nesse sentido, este novo processo de ensinar, rompe a idealização presencial no processo de ensino-aprendizagem. Consequentemente, a Educação a Distância surge como um novo ato pedagógico que não concentra responsabilidade no professor, como era percebido no modelo de ensino presencial.

Com base nessas informações e nos trabalhos abordados anteriormente, esta Revisão Sistemática da Literatura (RSL) tem como objetivo principal investigar, identificar e analisar o estado da arte atual, no qual se encontram as pesquisas que contemplam o monitoramento e o acompanhamento das interações dos alunos nos Ambientes Virtuais de Aprendizagem. O objetivo é que os resultados adquiridos direcionem para desenvolvimento de trabalhos futuros. Nesse sentido, os trabalhos mais importantes foram categorizados e analisados, especificando quais técnicas e métricas foram e estão sendo utilizadas nas pesquisas que referenciam o ensino e aprendizagem nos AVAs.

Para relatar esta RSL, além da seção de introdução, o artigo está estruturado da seguinte maneira: nas seções 2 e 3 é apresentada a fundamentação teórica. A Seção 4 discorre sobre o protocolo utilizado na RSL, identificando as questões de pesquisa. Nas seções 5 e 6 são apresentados os resultados advindos do processo empregado na RSL, assim fazendo menção as características dos trabalhos. Na Seção 7 é apresentada uma avaliação crítica dos trabalhos apontando as características dos mesmos. Por fim, a Seção 8 conclui o trabalho, apontando as contribuições, limitações e trabalhos futuros.

2. A Importância das Interações no Processo do Ensino-Aprendizagem

A análise das interações no processo do Ensino a Distância surge para ampliar a prática educacional pedagógica, e para que isso se torne uma realidade é necessário promover

situações em que os alunos se defrontem com as diversas maneiras de aprender. Por isso, percebe-se que a interação social contribui no processo do conhecimento como construção do saber, interativo, interpessoal e participativo. Além disso, segundo Siebra *et al.* (2006) a análise de interação é importante para descobrir mais sobre o aluno, como por exemplo, suas dificuldades ou dúvidas.

A interação que ocorre na EaD deve ser vista como um fenômeno pedagógico e não simplesmente como uma mera questão de distância geográfica. A distância física conduz uma lacuna na comunicação, um espaço psicológico de concepções errôneas potenciais entre instrutores e alunos, que precisa ser suplantado por técnicas especiais de ensino. A superação da distância se dá por meio de procedimentos diferenciadores na elaboração da instrução e na facilitação da interação (MOORE, 2007).

Segundo Messa (2010), o processo de ensino-aprendizagem no âmbito da EaD contemporânea, em especial nos AVAs, está relacionada à utilização do material didático e das Tecnologias da Informação e Comunicação (TICs) disponíveis à mediação do tutor e à interação entre os alunos.

Além disso, os cursos nessa modalidade de ensino promovem a interação social entre o aluno e o professor. Desta forma, percebe-se que grandes benefícios surgem com o uso da Internet e das mídias digitais. Entretanto é notório alguns desafios para este método de aprendizagem, no qual é possível perceber a importância do acompanhamento do aluno para o professor. De acordo com Bremgartner e Netto (2011), há possibilidade de interação entre alunos e professores, compartilhando o conhecimento, devido os avanços das Tecnologias Digitais de Informação e Comunicação (TDICs) e da procura por plataformas educacionais.

De acordo com Silva e Navarro (2012), a dinâmica ensino-aprendizagem que envolve a interação professor-aluno deve abranger todos os aspectos, englobando assim as suas condições de vida, sua relação com a escola, a percepção e a compreensão do conhecimento sistematizado a ser estudado. Há diversas maneiras de ensinar e aprender e estão fortemente relacionadas nas ações efetuadas pelos mediadores em um AVA. Todavia, a interação entre os participantes promove a transformação no processo de ensino-aprendizagem.

Para Machado e Kazuko (2009), a EaD é complexa, visto que não demanda apenas em mensurar a quantidade de alunos que atingem ou mesmo a quantidade de interação que se tem, mas implica em uma reflexão, em seus parâmetros pedagógicos, sobre suas intencionalidades, já que poderão servir tanto ao ajustamento quanto para desenvolvimento educacional social.

Portanto, a aprendizagem não se dá pela simples transmissão de algo que está fora, mas sim, depende do desequilíbrio cognitivo, o qual é provocado num processo de interação, da ação do sujeito sobre o objeto do conhecimento, dos esquemas de significação que possibilitam o estabelecimento de relações com o novo equilíbrio (PINTO *et al.* 2002).

3. Intervenções Pedagógicas como Processo Colaborativo e Avaliativo na Educação a Distância

As intervenções pedagógicas vem se tornando práticas educadoras que ajudam no processo de ensino-aprendizagem. Dessa forma, é possível por meio delas realizar uma análise do desenvolvimento ou o regresso do estudante. Todavia, podemos considerar que a interação é um ato comunicativo, que envolve seres humanos cujo objetivo é a comunicação interpretada como um ato interativo.

De acordo com Fuks *et al.* (2005), a comunicação envolve a negociação de compromissos e o conhecimento. Através da coordenação o grupo lida com conflitos e se organiza para evitar que os esforços de comunicação e de cooperação sejam perdidos. A Educação a Distância incrementada por recursos que permitem a comunicação e a realização de atividades promovem a construção de uma cultura informatizada e um saber cooperativo e colaborativo, por ações que privilegiam a interação (LIMA e SAUER, 2006).

Atualmente, inúmeros trabalhos estão sendo realizados na área da Informática Educacional, sendo notório o desenvolvimento de protótipos para acompanhar as interações nos fóruns de discussão, mensagens, chats, wikis etc. Além disso, constata-se, igualmente, que há poucas pesquisas na área da Educação visando as intervenções pedagógicas na EaD. A EaD segundo Nova e Alves (2003), surge como uma modalidade de ensino-aprendizagem que possibilita a mediação, por intermédio dos suportes tecnológicos, digitais e de rede, seja inserida em sistemas de ensino presenciais, mistos ou completamente realizada através da distância física.

De acordo com Gerosa *et al.* (2003), alguns fatores citados anteriormente ajudam o docente a coordenar os aprendizes, podendo intervir a tempo quando a discussão começar a tomar rumos não desejados e podem avaliar a evolução da turma e do curso. Estes fatores também podem ser utilizados para guiar os aprendizes durante sua atuação nos fóruns e para a construção de filtros e mecanismos de coordenação que evitem antecipadamente os erros.

Contudo, a interação social corresponde a uma parte cada vez mais significativa da aprendizagem, pois cada vez mais os estudantes se sentem motivados a aprender na interação com seus pares. Essa característica do estudante clama, então, por práticas pedagógicas baseadas na colaboração.

A colaboração observada no ambiente virtual manifesta-se como um fenômeno que ocorre tanto entre os alunos quanto entre aluno e professor. O desejo de colaborar, entre os alunos, evidencia-se muito claramente entre os iniciantes que, repetidas vezes, manifestam seu interesse em estabelecer trocas e parcerias com seus colegas. A fala do aluno, dirigindo-se por e-mail aos seus colegas, representa esse interesse de troca de experiências na área de conhecimento do curso (Mülbert *et al.* 2006).

4. Protocolo da Revisão Sistemática

Segundo os critérios de Kitchenham e Charters (2007), que descrevem os passos necessários para a condução de uma RSL, buscamos encontrar pesquisas atuais e relevantes que contribuam para a realização de futuras pesquisas, que possam direcionar o ensino e aprendizagem no contexto da EaD. Além disso, foram analisadas publicações

científicas, com o propósito de caracterizar e identificar técnicas utilizadas para desenvolver sistemas que auxiliem no ensino e na aprendizagem em AVAs.

A RSL é uma técnica de pesquisa baseada em evidências da literatura científica, conduzida formalmente, seguindo as fases de um protocolo bem definido. Tais fases incluem a realização de atividades de planejamento, de execução e de resultados a fim de responder às questões de pesquisa definidas na Seção 4. As buscas dos artigos foram realizadas por meio da submissão de *strings* às bases de dados indexadas em máquinas de busca digitais.

4.1 Objetivo

Esta RSL tem como objetivo principal seguir o protocolo definido pelo GQM (*Goal-Question-Metric*) desenvolvido por Basili *et al.* (1994), no qual define alguns métodos a serem coletadas a fim de atender alguns objetivos. Os objetivos desta RSL estão destacados no Quadro 1.

Quadro 1. Objetivo conforme paradigma GQM

ANALISAR	Publicações científicas.
COM O PROPÓSITO DE	Caracterizar e analisar.
EM RELAÇÃO A	Realizar um levantamento do estado da arte existente nas diversas literaturas, que tenham como finalidade analisar as interações sociais entre os participantes de um curso EaD por intermédio de uma abordagem multiagente.
DO PONTO DE VISTA DOS	Dos pesquisadores.
NO CONTEXTO	Acadêmico.

4.2. Questões da Pesquisa

A RSL apresentada neste artigo tem como finalidade responder as questões de pesquisa que são apresentadas nas Subseções 4.2.1 e 4.2.2, com o propósito de atrair estudos similares a temática abordada.

4.2.1 Questão Principal da Pesquisa

Com o propósito identificar os trabalhos que analisam as interações dos participantes de um AVA com o objetivo de auxiliar os professores a coordenarem as suas ações; essa RSL apresenta como questão principal de pesquisa:

Quais os trabalhos que analisam as interações sociais entre os estudantes em um Ambiente Virtual de Aprendizagem e quais os dados coletados e critérios utilizados por esses trabalhos nas análises?

4.2.2 Questões Secundárias da Pesquisa

A fim de responder a questão principal de pesquisa e de identificar quais as características, técnicas e critérios utilizados nos trabalhos, as questões secundárias da pesquisa são:

- Quais foram os requisitos considerados para analisar as interações entre os participantes?

- Quais as técnicas e os métodos utilizados na coleta e análise das interações sociais entre os estudantes?

4.3 Estratégias de Busca

As buscas foram realizadas nas bibliotecas digitais: SCOPUS, IEEE Xplore e ACM *Digital Library*. Já as buscas manuais foram realizadas em eventos e conferências que atendem os requisitos para esta pesquisa, entre eles destacamos: Simpósio Brasileiro de Informática na Educação (SBIE), Simpósio Brasileiro de Sistemas Colaborativos (SBSC), Conferência Internacional sobre Informática na Educação (TISE) e Workshop de Informática na Escola (WIE).

4.4 Idiomas

Os idiomas escolhidos para realizar esta busca nas bibliotecas digitais e nas buscas manuais foram o Inglês e o Português, sendo que o Inglês foi utilizado nas bibliotecas por ser o idioma adotado na maioria das conferências e periódicos existentes atualmente que estão relacionados com o tema desta pesquisa. Para as buscas manuais, o idioma utilizado foi o idioma adotado pela própria conferência/evento, que foram citados na Seção 4.3.

4.5 Expressão de Busca

A *string* geral, submetida às bibliotecas digitais é apresentada a seguir, no formato de expressão lógica e utiliza os operadores *OR* e *AND*. Ela foi devidamente ajustada para cada mecanismo de busca e a sua estrutura foi elaborada de acordo com as características e especificidades da *string*. Desta forma a *string* de busca utilizada foi:

((*{virtual learning environment}* OR *{learning management system}* OR *{E-learning}* OR *{moodle}* OR *{interactions relations groups}* OR *{interaction analysis groups}* OR *{social network analysis}* OR *{Timeline}*) AND (*{multi-agents systems}* OR *{intelligent agent}*)).

4.6 Critérios de Inclusão e Exclusão

Com o propósito de eliminar estudos considerados desnecessários para as questões de pesquisa, as seguintes restrições foram adotadas: artigos que não estavam escritos na língua inglesa ou portuguesa, artigos duplicados, artigos sem a versão completa e artigos que não foram publicados em periódicos ou anais de conferências.

Com o intuito de facilitar a busca e a seleção desses artigos, foram definidos alguns critérios de inclusão e exclusão, os critérios adotados encontram-se no Quadro 2:

Quadro 2: Critérios de Inclusão e Exclusão

Critérios de Inclusão (CI)	Critérios de Exclusão (CE)
CI1: Artigos disponíveis na íntegra para download. CI2: Artigos de conferências ou periódicos. CI3: Podem ser selecionadas publicações que envolvam Sistemas Multiagente no processo de ensino e aprendizagem em ambientes virtuais.	CE1: Publicações que não atendem os critérios de inclusão. CE2: Artigos que não apresentam uma ferramenta ou aplicação, ou seja, artigos teóricos e de mapeamentos sistemáticos são desconsiderados. CE3: Artigos duplicados ou pagos.

Ao realizar as buscas nas bibliotecas mencionadas a seguir, obtivemos um resultado total de 269 artigos, conforme apresenta a Tabela 1.

Tabela 1: Resultados Retornados através da String de Busca

Bibliotecas Digitais	String de busca	Pago	1º Filtro	2º Filtro	3º Filtro
IEEE	170	-	33	12	3
SCOPUS	51	6	8	4	2
ACM	39	11	5	2	1
TOTAL	269	17	46	18	6

Ainda segundo a Tabela 1, é possível constatar que na primeira análise avaliativa foram identificados 17 artigos pagos. Após a exclusão dos artigos pagos, realizou-se uma leitura do título e do *abstract* (resumo) dos 252 artigos restantes, sendo que desses 252 artigos, restaram somente 46 para a próxima fase de análise. Para análise do 1º filtro, realizamos uma leitura da introdução e conclusão dos 46 artigos. Para o 2º filtro haviam 18 trabalhos, onde foram realizados os procedimentos do 1º filtro mais a leitura parcial nas seções identificadas como relevantes para pesquisa. Todavia, para o 3º filtro restaram 6 artigos que foram lidos na íntegra a fim de identificar as informações relevantes para o desenvolvimento da pesquisa.

5. Extração de Informações

Para a extração das informações, nos baseamos em fornecer um conjunto de respostas inerentes para cada questão de pesquisa que foram definidas na Seção 4.2. Este procedimento de extração dos dados assegura a aplicação utilizada para todos os artigos selecionados e conseqüentemente facilita na classificação dos mesmos, obtendo desta forma um resultado satisfatório e relevante nesse processo.

As informações extraídas dos artigos pré-selecionados para o 2º filtro são detalhadas no Quadro 3 onde serão apresentadas dados analisados pelos autores.

Quadro 3: Dados Analisados dos Artigos e Referências

Dados Analisados	Referências
E-mail, Mensagens, Interação e Comunicação	Lima <i>et al.</i> (2014); Jaques, (1997); Bakharia e Dawson, (2011), Nolker e Zhou, (2006); Soller, (2001); Blot, (2014); Dráždilová <i>et al.</i> (2008), Stepanyan, (2014), Gerosa <i>et al.</i> (2003), Fuks <i>et al.</i> (2005).
Fórum de Discussão e Tarefas	Lima <i>et al.</i> (2014), Bremgartner e Netto, (2011); Jaques, (1997), Soller, (2001); Gerosa <i>et al.</i> (2003), Fuks <i>et al.</i> (2005).
Colaboração, Gerenciamento e Tutoria	Soller, (2001); Boff, (2008); Dráždilová <i>et al.</i> (2008), (Bentivoglio, 2009), Bremgartner e Netto, (2011); Stepanyan <i>et al.</i> (2014); Gerosa <i>et al.</i> (2003), Fuks <i>et al.</i> (2005).
Acompanhamento e Formação de Grupos	Lima <i>et al.</i> (2014); Daradoumis <i>et al.</i> (2006); Stepanyan <i>et al.</i> (2014); Dráždilová <i>et al.</i> (2008).

6. Trabalhos Analisados

Nesta seção são apresentados os trabalhos resultantes da pesquisa. Primeiramente foram pré-selecionados 18 artigos no 2º filtro, onde foram aplicados os critérios de inclusão e exclusão, resultando 6 artigos a serem analisados no 3º filtro. Para atingir o objetivo dessa RSL, apresentamos no Quadro 4 elementos que embasam as respostas relacionadas às questões de pesquisa abordadas anteriormente, onde foram identificadas características e métodos aplicados nestes trabalhos. Nesta perspectiva analítica dos trabalhos, foi possível analisar as diversas abordagens acerca das interações sociais nos ambientes virtuais de aprendizagem.

Quadro 4: Características e Métodos de Visualização

Autores	Características			
	Sistemas Multiagente	Interações Sociais	Plataforma	Visualização
Lima <i>et al.</i> (2014)	Sim	Aluno/Aluno Professor/Aluno	Moodle	Sociogramas/ Grafos/ Tabelas
Jaques, (1997)	Sim	Aluno/Aluno	Moodle	-
Rocha e Romani, (2001)	-	Professor/Aluno	Moodle	Gráficos/Grafos
Bakharia e Dawson, (2011)	-	Professor/Aluno	Moodle	Gráficos Sociais/ Sociogramas
Dráždilová <i>et al.</i> (2008)	-	Aluno/Aluno	Moodle	Gráficos
Soller, (2001)	Sim	Aluno/Aluno	Moodle	-

O trabalho de Lima *et al.* (2014), apresenta uma estratégia baseada numa arquitetura multiagente, tendo como objetivo analisar as interações entre os estudantes que ocorrem nos fóruns e nas mensagens de um AVA, os resultados dessas análises são

estruturados e apresentados ao professor sob a forma de sociogramas e de tabelas, além disso, o professor recebe alertas via e-mail dos alunos com baixa ou nenhuma interação.

Jaques (1997), apresenta um experimento utilizando uma arquitetura multiagente para realizar o monitoramento das ferramentas de comunicação em um ambiente de Ensino a Distância. Este sistema analisa as discussões que se encontram em andamento nas ferramentas de lista de discussão, chat e newsgroup, disponibilizando ao professor informações estatísticas (percentual de participação dos alunos e número de mensagens trocadas), e identificando possíveis associações nas interações, tais como, tópicos e subtópicos que interessam ao aluno e aos grupos de alunos que interagem mais entre si.

O trabalho de Rocha e Romani (2001), apresenta a ferramenta InterMap (*Interaction Map*), que provê um modelo de interface gráfica interativa. Com o auxílio de técnicas de visualização da informação o trabalho representa graficamente os dados das ferramentas de comunicação em ambientes de Educação a Distância na Web.

Bakharia e Dawson (2011), apresentam uma ferramenta que tem como objetivo analisar a interação no fórum entre os estudantes de um curso on-line e assim proporcionar uma representação visual de fácil interpretação da organização do grupo social nos fóruns.

Drázdilová *et al.* (2008), analisa através de gráficos em redes sociais os dados do sistema de e-Learning. Os dados analisados contém registros sobre as atividades dos alunos em um sistema de AVA. Identificamos também neste trabalho métodos de agrupamento e, conseqüentemente, foram utilizados métodos de análise de rede social. As redes sociais analisadas neste trabalho são representadas por grupos de estudantes que possuem contatos semelhantes e interagem em círculos sociais.

O trabalho de Soller (2001), proporciona uma melhor compreensão da interação do grupo e orienta a melhor forma de apoiar o processo de aprendizagem colaborativa. Todavia, propõe um quadro de princípios para o estudo e análise da interação de grupo que é construído através da combinação de diferentes aspectos e questões de colaboração e aprendizagem.

Os estudos foram selecionados de acordo com as características e dados individuais de cada um. Os resultados obtidos com as análises destes trabalhos envolviam diferentes indivíduos, como gestores, estudantes, mediadores e tutores; podendo ser destinados a um grupo, por exemplo, uma ferramenta que auxiliasse na intervenção pedagógica do mediador, com a finalidade de melhorar a aprendizagem, utilizando técnicas de visualização e assim otimizar o trabalho do professor/mediador.

Nos trabalhos que foram selecionados, identificamos que os dados analisados foram o conteúdo do estudante, as interações sociais, o feedback ao estudante, o envio de notificações (SMS, e-mails, no AVA, etc.), diante dessa verificação percebemos que ainda há muito que ser pesquisado nesta área, enfatizando as intervenções pedagógicas em tempo real utilizando técnicas e métodos de visualização.

Nesta pesquisa foram identificados um número baixo de trabalhos que descrevem a intervenção pedagógica, no entanto, muitos trabalhos não indicam alguma possibilidade de intervir pedagogicamente, uma vez que as intervenções pedagógicas mediadas pelos professores contribuem no processo de ensino-aprendizagem.

7. Discussões e Desafios

O desenvolvimento de pesquisas que envolve o processo educativo na modalidade EaD, vem fomentando análises no campo de investigação, colaboração e articulação entre os pesquisadores envolvidos com esta nova tendência educacional. Todavia, esses estudos abordam discussão de como ampliar a interação social entre os estudantes, uma vez que esse tipo de comunicação nesses ambientes torna-se dinâmico em virtude do desenvolvimento desses ambientes de aprendizagem, onde a socialização virtual é monitorada.

Neste contexto, o trabalho de Bentivoglio (2009), analisa as discussões que se encontram em andamento nas ferramentas de lista de discussão, *chat e newsgroup*, disponibilizando ao professor informações estatísticas (percentual de participação dos alunos e número de mensagens trocadas), e identificando possíveis associações nas interações, tais como, tópicos e subtópicos que interessam ao professor e grupos de alunos que interagem mais entre si, etc. Entretanto, a capacidade de visualizar a estrutura gráfica social e a evolução da comunidade é crucial para o sucesso e serve como um indicador antecipado do sucesso de um design de atividade, bem como as informações sobre a participação dos alunos e seu desempenho educacional.

Os artigos analisados não discutem explicitamente sobre o interesse de intervir pedagogicamente nas interações sociais entre os participantes em um curso EaD. Por exemplo, as tecnologias poderiam promover interações sociais educativas, as quais são geralmente mensuradas pela quantidade de alunos participativos de um fórum interativo no ambiente virtual de aprendizagem em tempo real. Entretanto, nos ambientes virtuais de aprendizagem podemos tecnologicamente acompanhar de maneira estratégica as interações entre os estudantes.

Considerando os diferentes níveis de interação social, Machado e Kazuko (2009), apresentam um modelo de aprendizagem colaborativa projetada para ajudar um sistema de aprendizagem colaborativa inteligente a identificar áreas problemáticas e a interação de grupo-alvo. O modelo descreve possíveis indicadores de aprendizagem colaborativa eficaz, e para cada indicador, recomenda estratégias para melhorar a interação entre pares. Este modelo de aprendizagem colaborativa dirigiu o projeto e desenvolvimento de duas ferramentas que automatizam a codificação, e auxiliam a análise da conversa de aprendizagem colaborativa e de atividade.

8. Conclusões

Esta pesquisa apresentou uma Revisão Sistemática da Literatura, observando o estado da arte em relação às interações sociais entre os estudantes em um Ambiente Virtual de Aprendizagem. Todavia, foram obtidos por intermédio desta RSL, resultados satisfatórios para responder a seguinte pergunta: como auxiliar o professor a identificar os diversos estágios das interações sociais entre os alunos no decorrer do curso em um Ambiente Virtual de Aprendizagem, possibilitando intervenções pedagógicas?

Desta forma, foi realizada uma coleta de artigos que dissertavam sobre as interações sociais dos estudantes. Entretanto, esta análise foi direcionada em pontos-chaves, a fim de obter uma colaboração maior de alunos e professores nos experimentos

das interações sociais nos Ambientes Virtuais de Aprendizagem, com o objetivo de identificar essas interações.

Nesta perspectiva, é necessário que a pesquisa na área de Educação a Distância promova a evolução da aprendizagem conceituando características que colaboram com o desenvolvimento educacional, socializando os alunos neste processo de ensino. Todavia, as ferramentas tecnológicas são extremamente importantes para o acompanhamento das interações sociais que desenvolvem a socialização em ambientes virtuais. Com isso, os artigos apresentados na RSL nos ajudaram a compreender os diversos contextos em que a pesquisa na área de computação é influente nesta modalidade de ensino.

Contudo, percebe-se que existe uma carência de estudos sobre as interações sociais entre os participantes em ambientes virtuais. Sendo assim, para trabalhos futuros pretende-se desenvolver um sistema multiagente que realize o acompanhamento das interações sociais entre cursistas de um curso EaD e forneça essas informações por intermédio de Data Visualization ao professor, a fim de lhe auxiliar na coordenação do curso on-line.

Por fim, esta RSL contribuiu na identificação dos principais elementos que podem ser utilizados no desenvolvimento desse sistema e a sua validação será realizada por intermédio de avaliações e testes.

Agradecimentos

Agradecemos o apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) através do Edital N° 003/2015.

Referências

- Bakharia, Aneasha, & Shane Dawson. (2011). SNAPP: A Bird's-Eye View of Temporal Participant Interaction. International Conference on Learning Analytics and Knowledge LAK '11, ACM, Banff, Alberta.
- Basili, Victor, Gianluigi Caldiera, & H. Dieter Rombach. (1994). Goal Question Metric. Paradigm Encyclopedia of Software Engineering, Kansas.
- Bentivoglio, C. A. (2009, November). Time Analysis of Forum Evolution as Support Tool for E-Moderating. In Intelligent Systems Design and Applications, 2009. ISDA'09. Ninth International Conference on (pp. 1131-1135). IEEE.
- Blot, Guillaume, Pierre Saurel, & Francis Rousseaux. (2014). Pattern Discovery in E-Learning Courses: A Timebased Approach. International Conference on Control, Decision and Information Technologies, CoDIT, Washington, DC 20057, Estados Unidos.
- Boff, Elisa (2008). Colaboração em Ambientes Inteligentes de Aprendizagem Mediada por um Agente Social Probabilístico, Tese de Doutorado, UFRGS, Porto Alegre.
- Bremgartner, V., & Netto, J. F. M. (2011). Auxílio Personalizado a Estudantes em Ambientes Virtuais de Aprendizagem Utilizando Agentes e Competências. In Anais do Simpósio Brasileiro de Informática na Educação (Vol. 1, No. 1).

- Campana, V. (2008). Agentes para Apoiar o Acompanhamento das Atividades em Ambientes Virtuais de Aprendizagem. Anais Do Simpósio Brasileiro de Informática na Educação XIX (SBIE), Fortaleza.
- Daradoumis, T., Martínez-Monés, A., & Xhafa, F. (2006). A Layered Framework for Evaluating On-line Collaborative Learning Interactions. *International Journal of Human-Computer Studies*, Elsevier.
- Dráždilová, P., Martinovic, J., Slaninová, K., & Snášel, V. (2008). Analysis of Relations in eLearning. In 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society Washington, DC, USA.
- Fuks, H., Raposo, A., Gerosa, M. A., Pimentel, M., Filippo, D., & Lucena, C. J. P. (2007). Inter- e Intra-relações entre Comunicação, Coordenação e Cooperação. Anais do IV Simpósio Brasileiro de Sistemas Colaborativos (SBSC), Rio de Janeiro-RJ, ISBN 987-85-7669-116-7.
- Fuks, H., Gerosa, M. A., Filippo, D., & Lucena, C. J. P. (2005). Informações Estatísticas e Visuais para a Mediação de Fóruns Educacionais, *Revista Brasileira de Informática na Educação (RBIE)*.
- Gerosa, M.A., Pimentel, M.G., Fuks, H. & Lucena, C.J.P. (2003), Coordenação de Fóruns Educacionais: Encadeamento e Categorização de Mensagens, XIV Simpósio Brasileiro de Informática na Educação, Rio de Janeiro-RJ.
- Jaques, P. A. (1997). Um Experimento com Agentes de Software para Monitorar a Colaboração em Aulas Virtuais. *Revista Brasileira de Informática na Educação*, Comissão Especial de Informatica na Educação , Rio Grande do Sul.
- Kitchenham, B., & Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering Version 2.3*. Engineering, Durham, UK.
- Lima, D., & Netto, J. F. (2015). Análise dos Resultados de um Sistema Multiagente que Identifica e Caracteriza as Relações Sociais dos Alunos de um Ambiente Virtual de Aprendizagem. In Anais do Simpósio Brasileiro de Informática na Educação (Vol. 26, No. 1, p. 1274).
- Lima, D., Gadelha, B., Netto, J. F., & Bremgartner, V. (2014). MoodleGroups: um Aplicativo para Identificar as Relações Sociais entre Alunos em Ambientes Virtuais de Aprendizagem. *Simpósio Brasileiro de Sistemas Colaborativos*. October (Vol. 6, p. 9).
- Lima, I. G., Sauer, L. Z., Soares, E. M. S. (2004). Discutindo Alternativas para Ambientes de Aprendizagem de Matemática para Cursos de Engenharia. In: *World Congress On Engineering And Technology Education*, Anais. *Engineering Education in the Changing Society*, Guarujá.
- Machado, S., & Kazuko, T. (2009). Mediação Pedagógica Em Ambientes Virtuais De Aprendizagem : a Perspectiva Dos Alunos. *Congresso Nacional de Educação*, EDUCERE.
- Messa, W. C. (2010). Utilização De Ambientes Virtuais De Aprendizagem – AVAs: A Busca por Uma Aprendizagem Significativa. *Revista Brasileira de Aprendizagem*

Aberta e a Distância, (Ambientes Virtuais), pag:1–49.

- Mora, N., Caballé, S., Daradoumis, T., Gañán, D., & Barolli, L. (2014). Providing Cognitive and Social Networking Assessment to Virtualized Collaborative Learning in Engineering Courses. In *Intelligent Networking and Collaborative Systems (INCoS), 2014 International Conference on* (pag. 463-468). IEEE.
- Moore, M. G., Kearsley, G., Galman, R., & Mello, A. (2007). *Educação a distância: uma visão integrada*. Cengage Learning.
- Mülbert, A. L., Girondi, A., Pereira, A. T. C., & Nakayama, M. K. (2013). Affection and Conflict in Learning Management System Interactions. *RIED. Revista Iberoamericana de Educación a Distancia*.
- Nolker, R., & Zhou, L. (2006). Mapping Discussion Roles: From the Classroom to the Online Discussion Board. In *Association for Information Systems - 12th Americas Conference On Information Systems, AMCIS 2006* (Vol. 7, pp. 3961–3964).
- Nova, C., Alves, L. (2003), *Educação a Distância: Limites e Possibilidades, Educação a Distância: Uma Nova Concepção de Aprendizado e Interatividade*. São Paulo: Futura, 2003. p. 43-50.
- Pimentel, M., Gerosa, M. A., Filippo, D., Raposo, A., Fuks, H., & Lucena, C. J. P. (2006). Modelo 3C de Colaboração para o Desenvolvimento de Sistemas Colaborativos. *Anais do III Simpósio Brasileiro de Sistemas Colaborativos (SBSC), Natal-RN*. ISBN 85-7669-097-7.
- Pinto, S. C. C., Schlemmer, E., dos Santos, C. T., Pérez, C. C., & Rheinheimer, L. R. (2002). *AVA: Um Ambiente Virtual Baseado Em Comunidades*. XIII Simpósio Brasileiro de Informática na educação–SBIE2002. São Leopoldo-RS.
- Romani, L. A. S., & ROCHA, H. V. D. (2001). InterMap: Visualizando a Interação em Ambientes de Educação a Distância Baseados na Web. *Encuentro Internacional De Informática En La Educación Superior*, 1.
- Santos, M., & Oliveira, M. (2011). *Interação e Comunicação em Educação a Distância, Interação e Comunicação em Comunidades de Aprendizagem*, Santa Cruz-RN.
- Siebra, S. A., Salgado, A. C., & Tedesco, P. A. (2006). A Process for User Interaction Analysis in Collaborative Environments. *Anais do III Simpósio Brasileiro de Sistemas Colaborativos (SBSC), Natal-RN*. ISBN 85-7669-097-7.
- Silva, O. G., & Navarro, E. C. (2012). A Relação Professor-Aluno No Processo Ensino-Aprendizagem. *Revista Eletrônica Interdisciplinar*.
- Soller, A. (2001) Supporting Social Interaction in an Intelligent Collaborative Learning System. *International Journal of Artificial Intelligence in Education (IJAIED)*, Pittsburgh pp.40-62
- Stepanyan, K., Mather, R., & Dalrymple, R. (2014). Culture, Role And Group Work: A Social Network Analysis Perspective On An Online Collaborative Course. *British Journal of Educational*.

Caracterização das Estratégias de Qualidade de Uso que Potencializam a Utilização Contínua de Aplicativos Colaborativos de Navegação: Um estudo de caso no Waze

Lorena Tomagnini¹, Glívia A. R. Barbosa¹, Geanderson E. dos Santos², Ismael S. Silva¹

¹Departamento de Computação – Centro Federal de Educação Tecnológica de Minas Gerais (DECOM/CEFET-MG). Av. Amazonas, 7675 - Nova Gameleira - Belo Horizonte, MG – Brasil.

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (DCC/UFMG). Av. Antônio Carlos, 6627 - Prédio do ICEx Pampulha, Belo Horizonte, MG – Brasil.

lorenagft2013@gmail.com, gliviabarbosa@decom.cefetmg.br,
geanderson@dcc.ufmg.br, ismaelsantana@decom.cefetmg.br

Abstract. *This study aimed to verify and characterize, through a case study in Waze, the use of quality strategies adopted by collaborative applications of navigation and the contribution these strategies in the continued use of this type of application. The results showed that the designer takes strategies considered relevant to promote a collaborative environment that maintains the user to interact continuously.*

Resumo. *Este trabalho buscou verificar e caracterizar, através de um estudo de caso no Waze, as estratégias de qualidades de uso adotadas por aplicativos colaborativos de navegação e a contribuição das mesmas para o uso contínuo desse tipo de aplicação. Os resultados mostraram que o projetista adota estratégias consideradas relevantes para promover um ambiente colaborativo que mantém o usuário motivado a interagir continuamente.*

1. Introdução

A utilização de aplicativos móveis em diferentes contextos colaborativos (e.g., aplicativos sociais, educacionais e aplicativos de localização/navegação) vem crescendo e esse crescimento abre espaço para que diversos aplicativos da mesma categoria sejam disponibilizados para os usuários (e.g., aplicativos de comunicação: WhatsApp, Telegram e Hangouts) [de Oliveira *et al.* 2012][Google Play, 2016].

Diante dessa realidade, surge o desafio de motivar os usuários no uso constante desses aplicativos de modo que eles se mantenham fidelizados e engajados na utilização dos mesmos [de Oliveira *et al.* 2012]. Pensando neste aspecto, projetistas de interface e pesquisadores da Computação (e.g., [Barbosa *et al.* 2013][Silva *et al.* 2015]) têm investigado a aplicabilidade de adotar qualidades de uso (e.g., a colaboração, sociabilidade e gamificação) nos modelos de interface e interação de aplicativos colaborativos com o

intuito de incentivar o uso contínuo desse tipo de aplicação.

Uma categoria de aplicativos colaborativos que vem ganhando força, por seu uso constante, é a de aplicativos de navegação. Nesses aplicativos, o usuário pode solicitar o fornecimento de rotas, os trajetos com menor tempo de deslocamento, a visualização de mapas em tempo real, entre outras funcionalidades [Tarouco, 2013][Google Play, 2016]. Dentro deste domínio, o aplicativo Waze se destaca com mais de 50 milhões de usuários ao redor do mundo [Waze, 2016], sendo que, no Brasil, esse número ultrapassa os 6 milhões de usuários, fazendo do país o 2º no ranking de utilização do Waze [O Globo, 2013][Tarouco, 2013][Google Play, 2016].

Motivados pelo desafio de manter os usuários engajados na utilização de aplicativos móveis colaborativos, considerando a grande adoção dos aplicativos de navegação e sua importância no uso cotidiano e, pelo fato do Waze ser um fenômeno em termos de adoção mundial, o objetivo deste trabalho consiste em analisar e caracterizar como o aplicativo Waze tem motivado seu uso contínuo, de forma a: (1) identificar e demonstrar as estratégias comunicadas na interface que podem atuar como fatores motivacionais para este uso contínuo e (2) apresentar quais estratégias de fato contribuem para o engajamento dos usuários, considerando a perspectiva dos mesmos.

Os resultados indicaram que o Waze utiliza estratégias, tais como, usabilidade, comunicabilidade, acessibilidade, colaboração, sociabilidade e gamificação, que são consideradas relevantes para potencializar a experiência satisfatória dos usuários e, sob a perspectiva dos usuários participantes dessa pesquisa, essas estratégias contribuem para que eles se mantenham engajados no uso do aplicativo. Sendo assim, esse trabalho apresenta contribuições práticas e científicas para o projeto e avaliação de interfaces de aplicativos móveis colaborativos.

Isso porque, além de caracterizar a proposta motivacional para uso contínuo do Waze, as estratégias identificadas neste trabalho podem ser utilizadas na melhoria e/ou no desenvolvimento de soluções que visam potencializar o uso contínuo de outros aplicativos móveis colaborativos, tanto para o contexto de aplicativos de navegação, quanto para outros domínios. Em termos científicos/metodológicos, os resultados do MIS reforçaram a aplicabilidade do método, devido à sua fundamentação teórica, para identificar as estratégias de *design* comunicadas na interface, que visam potencializar determinadas qualidades de uso, neste caso, as estratégias motivacionais para uso contínuo de softwares/aplicativos.

2. Trabalhos Relacionados

Na literatura, foram identificados alguns trabalhos que buscaram avaliar sistemas e/ou aplicativos colaborativos com o intuito de identificar estratégias de qualidades de uso (e.g., usabilidade, acessibilidade, comunicabilidade, colaboração, sociabilidade e gamificação) que guiaram seu *design* para potencializar o sucesso e a motivação na adoção do sistema.

Seguindo essa linha, Coutinho *et al.* (2011) aplicaram o MIS para identificar as estratégias sonoras utilizadas em jogos para orientar seus jogadores durante a interação, através de um estudo de caso no jogo *FPS Half Life 2*. Posteriormente, os autores

verificaram o impacto dessas estratégias na acessibilidade de jogadores surdos. A pesquisa revelou que a adoção de recursos de áudio, como elemento estratégico para orientar jogadores durante a interação, pode comprometer a acessibilidade e o desempenho de jogadores surdos, quando não há opções alternativas para esse público, uma vez que esse tipo de usuário não poderá fazer uso dessa informação sonora durante a utilização do jogo.

O trabalho realizado por Barbosa *et al.* (2013) buscou analisar e caracterizar como o Facebook apoia a sociabilidade de seus membros. Para isso, o MIS foi utilizado com o objetivo de identificar as estratégias adotadas pelo projetista dessa rede social para prover sociabilidade aos seus usuários. Os resultados mostraram que o Facebook adota estratégias consideradas relevantes para promover uma interação social de qualidade em redes sociais *online*.

Já o trabalho realizado por Silva *et al.* (2014) investigou, através do MIS, a eficácia das estratégias de *marketing* na interface de sites de hotel, através de um estudo de caso no *website* do apart-hotel Promenade Champagnat. Finalmente, a pesquisa conduzida por Silva *et al.* (2015) buscou caracterizar como as estratégias de gamificação estavam sendo adotadas como fator motivacional em aplicativos educacionais. Os autores apresentaram um estudo de caso no aplicativo Duolingo, cujo foco é o ensino de idiomas, e através da aplicação do MIS demonstrou os benefícios de adotar a gamificação em aplicativos educacionais.

Os trabalhos apresentados nesta seção investigaram os benefícios de adotar qualidades de uso como fator motivacional para melhorar a experiência do usuário em sistemas/aplicativos colaborativos. Além disso, todos os autores argumentaram e enfatizaram a importância desse tipo de investigação para a melhoria e/ou desenvolvimento das diferentes categorias de sistemas/aplicativos. Contudo, a maioria dos trabalhos focou na caracterização da adoção de uma qualidade de uso específica e nenhum deles analisou a utilização dessas estratégias no contexto de aplicativos colaborativos de navegação.

Nesse sentido, o presente trabalho se difere dos demais, uma vez que apresenta uma caracterização de todas as estratégias de qualidade de uso que visam potencializar a utilização contínua de aplicativos colaborativos de navegação. Essa investigação é relevante, uma vez que esse tipo de aplicativo tem se tornado cada vez mais popular. Além disso, a identificação das estratégias adotadas, sob a perspectiva de diferentes qualidades de uso, pode contribuir para que projetistas de interface e pesquisadores compreendam melhor a aplicabilidade de cada qualidade de uso adotada como estratégia motivacional nesse contexto.

3. Metodologia e Referencial Teórico

Considerando o objetivo deste trabalho, a seguinte questão de pesquisa foi investigada: “*Quais estratégias de qualidade de uso são utilizadas pelo Waze que contribuem para motivar o uso desse aplicativo?*”. A metodologia adotada para responder essa questão consistiu em uma abordagem qualitativa, dividida em 3 etapas.

Inicialmente, foi realizado um levantamento bibliográfico com o objetivo de identificar diretrizes de qualidades de uso, presentes na literatura, que poderiam ser

utilizadas como fator motivacional para uso contínuo de sistemas/aplicativos colaborativos. Em seguida, a interface do Waze foi inspecionada, seguindo os passos do Método de Inspeção Semiótica (MIS) [de Souza *et al.* 2006] para identificar e apresentar quais decisões foram tomadas pelo projetista do Waze para motivar seus usuários, e se essas decisões são compatíveis com as estratégias levantadas na literatura.

A terceira etapa buscou investigar junto aos usuários do Waze se as estratégias adotadas pelo aplicativo contribuem como fator motivacional para uso contínuo dessa aplicação. Para isso foi aplicado um questionário *online* e as respostas foram analisadas de forma a caracterizar a influência das qualidades de uso na utilização do Waze. A seguir, será descrito o conceito de qualidades de uso adotado neste trabalho. Posteriormente o MIS será apresentado, bem como a justificativa pela escolha desse método.

3.1. Qualidade de Uso como fator motivacional para uso de sistemas/aplicativos

A Experiência do Usuário (do inglês *User Experience (UX)*) está relacionada a como um indivíduo se sente enquanto usuário de um produto, sistema ou serviço, e ao prazer e à satisfação evocados no contato com esse artefato [Norman e Nielsen, 2013]. No âmbito de soluções tecnológicas a UX é potencializada por propriedades que caracterizam a qualidade de uso dos sistemas interativos [Norman e Nielsen, 2013].

Segundo Prates *et al.* (2007), a incorporação dessas propriedades pode variar de acordo com o objetivo e/ou domínio do sistema. Contudo, de forma geral, independente do sistema, para que o apoio computacional de um software e/ou aplicação seja satisfatório para o usuário, é desejável que o projetista comunique as suas intenções e concepções através da interface projetada (i.e., comunicabilidade), remova os obstáculos que impedem o usuário de acessar e interagir com o mesmo (i.e., acessibilidade) e torne o uso adequado aos usuários alvo da aplicação (i.e., usabilidade). Em outras palavras, o sistema deve prover, simultaneamente, as respectivas propriedades de qualidade de uso: comunicabilidade, acessibilidade e usabilidade [Prates *et al.* 2007].

Além dessas, que são consideradas essenciais para todos os tipos de sistemas/aplicativos, outras propriedades têm sido incorporadas como estratégias para motivar o uso contínuo dos diferentes domínios de sistemas/aplicativos. Por exemplo, a colaboração e a sociabilidade que têm sido potencializadas em sistemas colaborativos, e a *gamificação*, que tem sido utilizada em diferentes domínios de softwares para manter os usuários engajados [Pereira *et al.* 2010][de Oliveira *et al.* 2012][Zichermann e Cunningham, 2011].

Segundo Preece (2001) e Prates *et al.* (2007) a colaboração e a sociabilidade, referem-se às regras sociais, como privacidade, liberdade de expressão, confiança e outros aspectos que surgem da interação entre pessoas. Assim, sistemas colaborativos (e.g., redes sociais *online*, editores de conteúdo, blogs) com uma boa colaboração e sociabilidade, são aqueles que têm políticas colaborativas e sociais estabelecidas, que são compreensíveis por seus membros e, além disso, aceitáveis e suficientemente práticas para que possam ser implementadas. A gamificação, por sua vez, é caracterizada pela utilização dos princípios de *design* de jogos, como interação e colaboração, em um contexto que não trata especificamente de um jogo (e.g., profissional, social, educacional) para envolver usuários e auxiliar na resolução de desafios [Zichermann e Cunningham, 2011].

Todas essas propriedades visam maximizar a melhoria da UX, e por isso podem atuar como fatores motivacionais para que os usuários adotem e utilizem continuamente um aplicativo colaborativo. Uma forma de caracterizar se essas propriedades foram implementadas no projeto de interface e interação é através da avaliação da interface, que pode ser conduzida através do Método de Inspeção Semiótica, o MIS, que será apresentado na próxima seção.

3.2. Método de Inspeção Semiótica

O Método de Inspeção Semiótica (MIS) é baseado na teoria da Engenharia Semiótica (EngSem), uma teoria explicativa de Interação Humano Computador (IHC), que permite entender os fenômenos envolvidos no *design*, uso e avaliação de um sistema interativo [de Souza *et al.* 2006].

A avaliação utilizando o MIS consiste em cinco etapas [de Souza *et al.* 2006]: (1) inspeção dos signos metalinguísticos; (2) inspeção dos signos estáticos; (3) inspeção dos signos dinâmicos; (4) contraste e comparação entre as mensagens identificadas em cada uma das inspeções e (5) apreciação da qualidade da metacomunicação. Nos passos de 1 a 3, o especialista deve ser capaz de reconstruir a metamensagem proposta pelo projetista a partir dos signos inspecionados. No passo 4, o especialista contrasta as mensagens geradas nos passos anteriores e, a partir desta análise, explora a possibilidade do usuário atribuir significados diferentes a um mesmo objeto (signo). Finalmente, no passo 5, o especialista gera uma versão unificada da metamensagem e registra a sua apreciação final da qualidade do sistema.

Embora originalmente o MIS tenha sido proposto para avaliar a comunicabilidade de um sistema, uma revisão na literatura realizada por Reis e Prates (2011), revelou que, devido à sua fundamentação teórica, esse método também pode ser utilizado para identificar as estratégias de *design* comunicadas na interface que visam potencializar determinadas qualidades de uso.

Neste sentido, trabalhos como o realizado por Coutinho *et al.* (2011) identificaram estratégias sonoras para orientação em jogos, através da aplicação do MIS. Já o trabalho realizado por Barbosa *et al.* (2011) utilizou o MIS para identificar estratégias de sociabilidade em redes sociais. Silva *et al.* (2014), utilizaram o MIS para identificar estratégias de *marketing* em sites de hotel, e Silva *et al.* (2015) aplicou o MIS para identificar estratégias de gamificação em aplicativos educacionais.

Demonstrada a aplicabilidade do MIS nesses contextos, a adoção desse método para avaliar o aplicativo Waze permitirá, por meio da apreciação da metamensagem do designer, identificar estratégias de qualidade de uso comunicadas na interface desse aplicativo para manter os usuários motivados em seu uso contínuo. A seguir, apresentamos os principais resultados da primeira etapa dessa pesquisa.

4. Apreciação da Proposta de Projeto do Waze para Motivar seu Uso Contínuo

Nesta seção, é apresentada a análise da interface do Waze, realizada utilizando o MIS, indicando a proposta do projetista para potencializar (i.e., motivar) o uso contínuo desse aplicativo, bem como as estratégias de qualidade de uso adotadas por ele. A avaliação foi realizada no período de três semanas, durante o mês de setembro de 2015, e foi conduzida

por dois autores desse trabalho, dos quais um já possui mais de 05 anos de experiência na aplicação do MIS no contexto de avaliação de sistemas colaborativos. O escopo da avaliação contemplou os principais cenários de uso do Waze, como: (1) traçar a rota para um destino; (2) inserir alertas; (3) troca de mensagens entre os contatos e (4) editar o perfil. A seguir, os principais resultados dessa etapa são apresentados.

4.1. Proposta do Waze para motivar seu uso contínuo

A partir da análise do MIS verificou-se que a intenção do projetista em relação ao Waze é disponibilizar um aplicativo cujo objetivo é fornecer a melhor rota para um determinado destino, através da contribuição de uma comunidade composta por usuários em trânsito (e.g., motoristas e passageiros). Além disso, o Waze visa proporcionar uma experiência mais prazerosa e dinâmica, através da colaboração entre os usuários que compartilham informações de trânsito e tráfego durante sua utilização (e.g., alertas de congestionamentos, acidentes, pontos de interdição, radares, pontos de fiscalização, preços em postos de gasolina, atualização de dados no mapa, entre outras).

Para atingir seu objetivo principal, o Waze estimula a colaboração entre os usuários e, conseqüentemente, sua utilização contínua fazendo uso de estratégias como: (1) a criação de uma rede de contatos para interação; (2) bate papo entre os contatos; (3) o estabelecimento de um *ranking* da interação, (4) a oferta de níveis de evolução de permissão e (4) a oferta de recompensas por colaboração. Em relação ao *ranking*, o usuário avança nas posições na medida em que acumula pontos por utilizar as funcionalidades do aplicativo. O Waze permite que seus colaboradores compartilhem suas posições no *ranking* de modo que possam estabelecer competições.

Em relação aos níveis de evolução de permissão, os mesmos são divididos em Waze bebê, adulto, guerreiro, cavaleiro e rei, que também são alcançados em função do acúmulo de pontos por utilização do aplicativo. Na medida em que evolui, o usuário recebe determinadas vantagens, como por exemplo, a possibilidade de editar um mapa. Os pontos são obtidos a partir da criação de alertas, da distância percorrida utilizando o Waze e/ou da atualização de mapas (e.g., inserindo nome nas ruas, número às edificações). De forma complementar, o Waze oferece recompensas, que são conquistadas a partir do momento que o usuário atende um conjunto de critérios, como, por exemplo, adicionar o primeiro amigo.

Outra forma de incentivar a interação entre os usuários, proposta pelo projetista do Waze, consiste na possibilidade de integrar o uso do aplicativo com as redes sociais, como Facebook e Twitter, de modo que seja possível: (1) localizar contatos dessas redes para adicioná-los no Waze, (2) compartilhar o percurso, o local atual e/ou o horário estimado de chegada ao destino. Além disso, o Waze oferece aos usuários a opção de enviar mensagens aos contatos. Essas mensagens podem ser públicas (i.e., para todos os contatos), privadas (i.e., para um contato ou um grupo específico) e/ou um “*Beep Beep*”, que consiste em um tipo de mensagem onde um usuário chama atenção de outro. A Figura 1 ilustra evidências da proposta do projetista para a interface do Waze.



Figura 1 – Proposta do projetista do Waze

4.2. Estratégias para Motivar o Uso Contínuo do Waze

Uma vez identificada a proposta do projetista do Waze para motivar seu uso contínuo, este trabalho buscou verificar se as decisões comunicadas na interface desse aplicativo são compatíveis com as estratégias que, de acordo com a literatura [Nielsen, 1994][de Souza e Preece, 2004][Pontes, 2008][Furtado *et al.* 2010][Werbach e Hunter, 2012] [Knoll, 2014][W3C, 2016], podem ser consideradas diretrizes de qualidades de uso que contribuem para manter os usuários motivados durante a interação em ambientes colaborativos. A Tabela 1 lista 43 estratégias levantadas na literatura, subdivididas em 6 qualidades de uso.

Tabela 1 – Estratégias do projeto de interface e interação para motivar uso contínuo de sistemas colaborativos identificadas na literatura

Qualidade de Uso	Cód.	Estratégia (Diretrizes)*	Autor(es)
Sociabilidade	S1	Propósitos bem definidos	[de Souza e Preece, 2004]
	S2	Considerar as diferenças de cada usuário	[de Souza e Preece, 2004]
	S3	Políticas bem definidas	[de Souza e Preece, 2004]
	S4	Permitir que as pessoas desenvolvam suas próprias identidades <i>online</i>	[de Souza e Preece, 2004]
	S5	Encorajar a empatia e confiança entre os membros	[de Souza e Preece, 2004]
	S6	Encorajar a reciprocidade	[de Souza e Preece, 2004]
	S7	Manter discussões em evidência	[de Souza e Preece, 2004]
	S8	Diversificar possibilidades de comunicação	[de Souza e Preece, 2004]
	S9	Garantir proteção e privacidade	[de Souza e Preece, 2004]
Colaboração	C1	Participação do usuário como colaborador	[Furtado <i>et al.</i> 2010]
	C2	Compreensão dos objetivos	[Furtado <i>et al.</i> 2010]
	C3	Incentivar a participação coletiva	[Furtado <i>et al.</i> 2010]
	C4	Acompanhamento do progresso individual	[Furtado <i>et al.</i> 2010]
	C5	Quantificação da contribuição do usuário no objetivo	[Furtado <i>et al.</i> 2010]
	C6	Configuração de características colaborativas	[Furtado <i>et al.</i> 2010]
Gamificação	G1	Múltiplos caminhos	[Werbach e Hunter, 2012]
	G2	Orientar o usuário na interação	[Werbach e Hunter, 2012]
	G3	Níveis de evolução	[Werbach e Hunter, 2012]
	G4	Erro como aprendizado	[Werbach e Hunter, 2012]
	G5	Competição e colaboração	[Werbach e Hunter, 2012]

	G6	Diversão	[Werbach e Hunter, 2012]
	G7	Incorporar a narrativa ao jogo	[Werbach e Hunter, 2012]
	G8	Definir metas	[Werbach e Hunter, 2012]
	G9	Estabelecer <i>rankings</i>	[Werbach e Hunter, 2012]
	G10	Reconhecimento de padrões	[Werbach e Hunter, 2012]
	G11	Permitir coleções	[Werbach e Hunter, 2012]
	G12	Surpresa e prazer inesperado	[Werbach e Hunter, 2012]
	G13	Presentes / Bonificação	[Werbach e Hunter, 2012]
	G14	Reconhecimento de uma realização	[Werbach e Hunter, 2012]
	G15	Informar <i>status</i> de progresso	[Werbach e Hunter, 2012]
Usabilidade	U1	Visibilidade do sistema	[Nielsen, 1994][Knoll, 2014]
	U2	Minimização da carga de memória	[Nielsen, 1994][Knoll, 2014]
	U3	Facilidade de acesso	[Nielsen, 1994][Knoll, 2014]
	U4	Ajuda e documentação	[Nielsen, 1994][Knoll, 2014]
	U5	Bom aproveitamento do espaço da tela.	[Nielsen, 1994][Knoll, 2014]
	U6	Interação incentivadora	[Nielsen, 1994][Knoll, 2014]
	U7	Configuração persistente	[Nielsen, 1994][Knoll, 2014]
Comunicabilidade	C1	Explicitar caminhos de interação	[Pontes, 2008]
Comunicabilidade, Acessibilidade e Usabilidade	M1	Contextualização ao usuário e Adequação às Funcionalidades	[Pontes, 2008][W3C, 2016][Nielsen, 1994][Knoll, 2014]
	M2	Tratamento de erros e auxílio ao usuário.	[Pontes, 2008][W3C, 2016][Nielsen, 1994][Knoll, 2014]
Comunicabilidade e Usabilidade	M3	<i>Feedback</i> claro, imediato e informativo.	[Pontes, 2008][Nielsen, 1994][Knoll, 2014]
Acessibilidade e Usabilidade	M4	Acessibilidade diversa e facilitada	[W3C, 2016][Nielsen, 1994][Knoll, 2014]
	M5	Previsibilidade e consistência	[W3C, 2016][Nielsen, 1994][Knoll, 2014]

Posteriormente, conforme demonstrado na Tabela 2, foi realizado um mapeamento entre as decisões do projetista do Waze e as estratégias motivacionais identificadas a partir da literatura. É importante ressaltar que uma estratégia pode refletir em uma ou mais decisões do projetista da interface. Além disso, na Tabela 2 foi utilizada a coluna “*Cod. Estratégias*” para referenciar as diretrizes de qualidade de uso apresentadas na Tabela 1.

Tabela 2 - Estratégias motivacionais versus Decisões do projetista do Waze

Cod. Estratégias	Cod. e Decisões do Projetista		Exemplo de aplicação no Waze	Qualidade de Uso
C1	D1	Oferecer diferentes tipos de alerta	O Waze disponibiliza a possibilidade de gerar alertas sobre: congestionamentos, acidentes, pontos de interdição, radares, preços em postos de gasolina, atualizações no mapa, entre outros. Estes alertas permitem aos usuários colaborar entre si no alcance do objetivo final, que é a determinação da melhor rota.	Colaboração
C3				Colaboração
G5				Gamificação
S5				Sociabilidade
S6				Sociabilidade
S8				Sociabilidade

C3	D2	Oferecer pontos por atividades executadas.	Durante a interação com as funcionalidades do Waze (e.g., solicitar rota, informar alerta, editar mapa, adicionar um contato) o usuário recebe pontos que permite posicioná-lo em um <i>ranking</i> e o habilita para receber recompensas.	Colaboração
G5				Gamificação
G14				Gamificação
G12	D3	Oferecer recompensas por uma conquista.	O Waze oferece recompensas a partir do momento que o usuário atende um conjunto de critérios e alcança uma conquista, como, por exemplo, adicionar o primeiro amigo.	Gamificação
G13				Gamificação
G14				Gamificação
G6				Gamificação
G3	D4	Oferecer níveis de evolução de permissão	Durante a interação com as funcionalidades do Waze, o usuário acumula pontos e pode evoluir seu nível de permissão no aplicativo (e.g., Waze bebê, adulto, guerreiro, cavaleiro e rei). Quanto mais evoluído, maior as possibilidades de interação com o aplicativo e intervenção na definição de rotas por parte do usuário.	Gamificação
G5				Gamificação
C4				Colaboração
U6				Usabilidade
G1	D5	Oferecer além da digitação, a opção de comando de voz.	O Waze disponibiliza a opção de realizar a busca por pontos de interesse, através digitação e comando de voz.	Gamificação
C1				Comunicabilidade
M4				Acessibilidade e usabilidade
C3	D6	Oferece a opção de pesquisar/adicionar/interagir com amigos.	Durante a interação com as funcionalidades do Waze, o usuário pode interagir com times (i.e., grupos de contatos), adicionar e pesquisar por contatos, e se comunicar com eles através de mensagens públicas ou privadas.	Colaboração
G6				Gamificação
S5				Sociabilidade
S6				Sociabilidade
S7				Sociabilidade
S8				Sociabilidade
U2	D7	Armazenar preferências do usuário	O Waze armazena automaticamente uma lista de destinos recentemente percorridos pelo usuário. Estes e outros destinos podem ser salvos como favoritos. Isso facilita a busca e recuperação de destinos favoritos.	Usabilidade
U3				Usabilidade
U7				Usabilidade
U4	D8	Oferecer opção de ajuda	O aplicativo disponibiliza ao usuário a opção de Ajuda para especificar problemas mais comuns e como resolvê-los. Além disso, o Waze oferece caminhos múltiplos para acesso a Ajuda.	Usabilidade
G1				Gamificação
G5	D9	Estabelecer um <i>ranking</i> de interação	O Waze exibe um <i>ranking</i> de pontuação, com a classificação dos usuários. O <i>ranking</i> pode ser compartilhado de modo que os usuários estabeleçam competições entre si.	Gamificação
G9				Gamificação
G15				Gamificação
C4				Colaboração
S2	D10	Oferecer configurações de acordo com o perfil do usuário	O usuário pode configurar o idioma padrão, unidades de medida (milhas ou quilômetros), alterar o <i>design</i> de seu carro de acordo com sua preferência e alterar o esquema de cores do mapa.	Sociabilidade
S4				Sociabilidade
C6	D11	Oferecer configurações de interação.	O usuário pode ativar ou desativar bate-papos públicos, mensagens privadas, <i>pop-up</i> de alertas e conexões com redes sociais.	Colaboração
S9				Sociabilidade
M2	D12	Recalcular rotas quando o usuário desvia o caminho.	O aplicativo calcula, automaticamente, uma nova rota para o usuário, quando o mesmo se desvia da rota especificada.	Comunicabilidade, acessibilidade e usabilidade.
G4				Gamificação

A Figura 2 evidencia alguns exemplos de decisões do projetista do Waze que são referenciadas através da coluna “Cód. e Decisões do Projetista”, da Tabela 2.

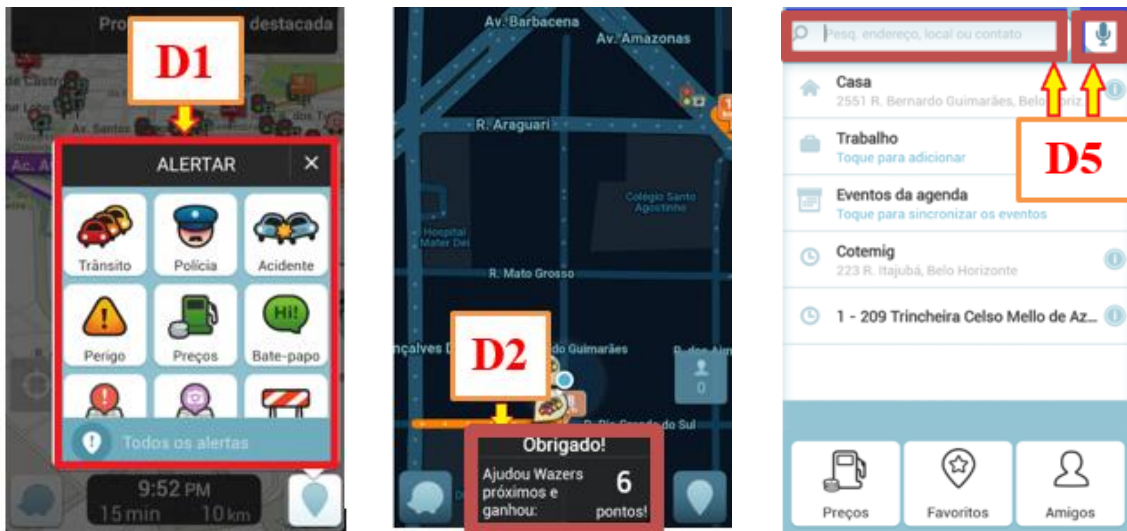


Figura 2 – Interface do Waze com evidências das decisões do projetista

Ao comparar a proposta do Waze, identificada através da avaliação com o MIS, com as estratégias motivacionais identificadas na literatura, foi possível constatar que as decisões do projetista do Waze estão alinhadas com as diretrizes de qualidades de uso que segundo Nielsen (1994), de Souza e Preece (2004), Pontes (2008), Furtado *et al.* (2010), Werbach, e Hunter (2012), Knoll (2014), W3C (2016), podem favorecer a utilização contínua do aplicativo e potencializar a experiência satisfatória dos usuários.

Para um melhor entendimento desses resultados, verificou-se que o projetista adotou 67% das estratégias identificadas na literatura. Além disso, não foi observado o uso de novas estratégias, ou seja, todas as decisões do projetista, observadas através do MIS, são compatíveis com pelo menos uma diretriz de qualidade de uso encontrada na literatura. Esse trabalho também buscou verificar a incidência de uma determinada qualidade de uso no projeto de interface e interação do Waze. Dessa forma, o gráfico da Figura 3 indica a frequência em que uma mesma qualidade de uso foi representada, através da implementação de suas diretrizes, no Waze.

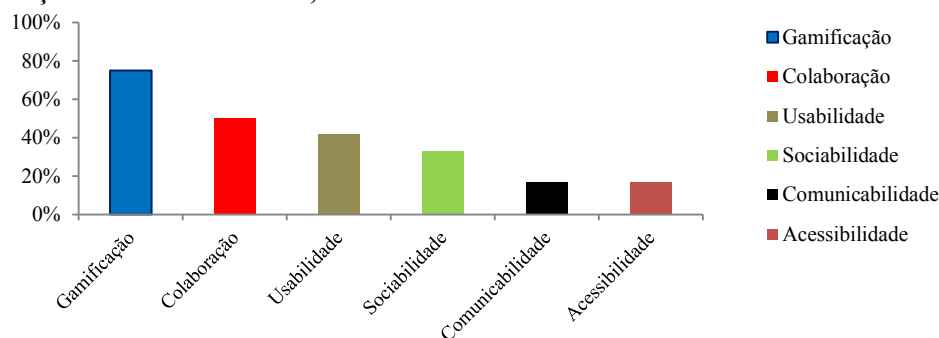


Figura 3 - Incidência das qualidades de uso nas decisões do projetista do Waze

A partir dos dados obtidos foi possível observar que o projetista do Waze adotou

estratégias que contemplam todas as qualidades de uso que visam potencializar a experiência satisfatória do usuário. Porém, houve uma predominância das seguintes propriedades: gamificação, colaboração, usabilidade e sociabilidade que, além da qualidade da interação, buscam motivar o uso prazeroso e contínuo do aplicativo.

5. Análise da Motivação de uso do Waze sob a perspectiva dos Usuários

Com o objetivo de investigar junto aos usuários do Waze se as estratégias de colaboração, sociabilidade, gamificação, usabilidade, acessibilidade e comunicabilidade adotadas pelo projetista do Waze contribuem como fator motivacional para uso desse aplicativo, foi aplicado um questionário *online* com 13 questões de múltipla escolha para os usuários do Waze. As questões foram definidas e agrupadas de modo a identificar: (1) o perfil dos participantes da pesquisa, (2) se eles se sentem motivados ao utilizar o Waze e (3) quais os fatores que mais contribuem para essa motivação. A aplicação do questionário ocorreu no período de 2 semanas, durante a segunda quinzena do mês de setembro de 2015. O mesmo foi divulgado através de redes sociais, e-mails e fóruns de discussões sobre o aplicativo.

Ao todo 185 pessoas participaram da pesquisa, respondendo ao questionário. Em relação ao perfil dos participantes 10% foram classificados na faixa etária entre 13 e 20 anos, a maioria (51%) declarou estar na faixa entre 21 e 28 anos, 20% na faixa de 29 a 36 anos, 14% na faixa de 37 a 56 anos e 5% acima de 56 anos. É importante destacar que a participação de usuários menores de 18 anos é válida, uma vez que o aplicativo pode ser utilizado não apenas por motoristas, mas também por passageiros (i.e., qualquer usuário em trânsito). Quanto ao tipo de aplicativos que utilizam, onde mais de uma categoria poderia ser marcada, 100% dos participantes afirmam utilizar aplicativos de navegação, 78% declararam utilizar redes sociais, 31% utilizam aplicativos de jogos, 97% fazem uso de aplicativos de mensagens instantâneas, 66% utilizam aplicativos de serviços (e.g., banco, taxi) e 5% fazem uso de aplicativos de outras categorias.

5.1. Análise das Motivação dos Usuários

Inicialmente, os participantes foram convidados a responder sobre as motivações iniciais que os levavam a escolher o Waze como aplicativo de navegação (i.e., o que motivou a fazer *download* do aplicativo). Nessa questão, mais de uma opção poderia ser marcada. O resultado obtido mostra que a maioria dos participantes (57%) escolheu usar o Waze pelo fato do aplicativo ser gratuito e 37% começaram a utilizá-lo por recomendações de amigos.

Além disso, os participantes também foram questionados sobre a frequência semanal de uso do Waze, e após fazerem o *download* do aplicativo, se os mesmos sentiam-se motivados durante sua utilização. Os resultados demonstraram, que mais da metade dos entrevistados, 59%, utilizam o Waze pelo menos cinco vezes por semana, e que 72% dos participantes se sentem motivados a fazer uso contínuo do aplicativo. A Figura 4 apresenta os fatores que mais motivam os usuários dessa pesquisa a utilizarem o Waze continuamente. Esses fatores correspondem às decisões do projeto de interface e interação do Waze listadas na Tabela 2. Nessa questão, mais de um fator poderia ser indicado, porém, não era obrigatória.

Através da Figura 4 é possível constatar que, embora os participantes pudessem optar por não responder essa questão, ou ainda marcar a opção “Nenhum dos Fatores”, as

decisões do projetista do Waze, que refletem em estratégias de qualidades de uso para potencializar a experiência do usuário e manter o uso contínuo da aplicação, contribuem como fatores motivacionais para a utilização do aplicativo.



Figura 4 – Fatores motivacionais para uso do Waze. Os fatores (i.e., as estratégias motivacionais) estão agrupados por qualidade de uso conforme a legenda da figura.

Conforme observado no gráfico da Figura 4, as estratégias de usabilidade e colaboração, identificadas na interface do Waze, se destacam em relação aos fatores que mais motivam os usuários na utilização contínua do aplicativo. Entretanto, é importante ressaltar que, em alguns casos, essas qualidades de uso apareceram associadas a outras, como, por exemplo, a estratégia mais referenciada pelos participantes: “*Apresenta o trajeto de uma rota de forma, clara, precisa e objetiva*”, engloba características de comunicabilidade e usabilidade.

Em seguida, os entrevistados foram questionados, explicitamente, se características como colaboração, sociabilidade e gamificação contribuem como fatores motivacionais

para a utilização contínua do Waze. Os resultados indicaram que para 87% dos participantes a colaboração contribui para que eles se mantenham engajados no uso do aplicativo. Isso reforça o argumento de que a colaboração é uma qualidade de uso que favorece o engajamento dos usuários no uso de aplicativos de navegação. Além disso, 67% afirmaram que as características de sociabilidade também influenciam positivamente no uso contínuo do Waze. Já a utilização de dinâmica de jogos (i.e., a gamificação) foi apontada por 42% dos participantes como fator motivacional. De forma complementar, no questionário foram feitas perguntas que buscaram verificar a satisfação do usuário em relação ao aplicativo. Os resultados indicaram que 94% dos participantes pretendem continuar utilizando o Waze como aplicativo de navegação.

6. Discussão dos Resultados

O Waze é um dos principais aplicativos colaborativos de navegação, com mais de 50 milhões de usuários e classificado, de acordo com Google Play (2016), com uma média de 4,6 estrelas (em uma escala máxima de 5 estrelas). Diante dessa aceitação, esse trabalho buscou caracterizar as estratégias de qualidade de uso adotadas pelo projetista desse aplicativo que contribuem como fator motivacional para uso contínuo do Waze.

Por meio da análise do MIS, foi possível verificar que o projetista utiliza elementos conhecidos no âmbito das qualidades de uso. Tais elementos podem ser vistos como estratégias que o Waze adota para motivar os usuários e que podem contribuir para mantê-los engajados na utilização contínua do aplicativo. Os resultados obtidos, por meio da pesquisa com os usuários, sustentam o argumento de que as estratégias identificadas na interface do Waze exercem influência motivacional para uso contínuo do aplicativo.

Nesse sentido, esta pesquisa evidencia a importância de adotar estratégias de qualidade de uso como, comunicabilidade, acessibilidade, usabilidade, colaboração, sociabilidade e gamificação, das quais, sobressaíram a usabilidade e a colaboração, no projeto de interface e modelo de interação de aplicativos colaborativos de navegação. Tal adoção contribui, assim, para que essas propriedades se consolidem como importantes estratégias para potencializar o engajamento e a experiência satisfatória dos usuários no uso desse tipo de aplicação.

7. Conclusões e Trabalhos Futuros

Este trabalho buscou apreciar e caracterizar as estratégias de qualidades de uso adotadas em aplicativos colaborativos de navegação e a contribuição das mesmas para o uso contínuo desse tipo de aplicação. Para isso foi realizado um estudo de caso no Waze e os resultados apresentados permitiram constatar que o projetista desse aplicativo adota estratégias consideradas relevantes para promover um ambiente colaborativo que mantém o usuário motivado a interagir continuamente.

Em termos de contribuições, apesar de apresentar um estudo de caso no Waze, a apreciação realizada se faz relevante tanto em termos técnicos, quanto científicos/metodológicos para a área. Isso porque, em termos científicos/metodológicos, os resultados do MIS reforçaram a aplicabilidade do método, devido à sua fundamentação teórica, para identificar as estratégias de *design* comunicadas na interface, que visam

potencializar determinadas qualidades de uso, neste caso, as estratégias motivacionais para uso contínuo de softwares/aplicativos. Já em termos técnicos, as estratégias apresentadas neste trabalho podem ser utilizadas na melhoria e/ou no desenvolvimento de soluções que visam potencializar o uso contínuo de outros aplicativos móveis colaborativos, tanto para o contexto de aplicativos de navegação, quanto para outros domínios.

Como trabalhos futuros, um ponto interessante a ser investigado seria a possibilidade de identificar componentes na interface (i.e., signos), que permitem classificar as estratégias motivacionais oferecidas pelos aplicativos. Isso ajudaria no projeto e avaliação de outros aplicativos/sistemas que buscam manter seus usuários motivados em sua utilização contínua.

Referências

- Barbosa, G. A. R.; Santos, G. E.; and Victor M. de O. Pereira, V. M. O. 2013. Caracterização qualitativa da sociabilidade no Facebook. In Proc. IHC '13, p. 72-81.
- Britto, T.C., Martins, L.C.G., Penteado, B.E. e Anacleto, J.C. 2011. Técnicas de Prototipação para Smartphones no Apoio à Avaliação de Interfaces com o Usuário. In Proc. IHC&CLIHC'11, p. 39- 42.
- Brown, Q., Bonsignore, E., Hatley, L., Druin, A., Walsh, G., Foss, E., Brewer, R., Hammer, J. e Golub, E. 2010. Clear Panels: A Technique to Design Mobile Application Interactivity. In Proc. DIS'10, p. 360-363.
- Coutinho, F.; Prates, R. O.; Chaimowicz, L. 2011. An analysis of information conveyed through audio in an fps game and its impact on deaf players experience. In Proc. of SBGAMES'11. p. 53–62.
- de Oliveira, D. H. D., de Miranda, L. C., de Miranda, E. E. C., da Silva, L. F. 2012. Prototipação de interfaces de aplicativos para dispositivos móveis: estado da arte e desafios de IHC. In Proc. of IHC '12, p. 315-324.
- de Souza, C. S., Leitão, C. F., Prates, R. O., and da Silva, E. J. 2006. The semiotic inspection method. In: Proc. of IHC'06, p. 148-157.
- de Souza, C. S., Leitão, C. F., Prates, R. O., Bim, S.A., da Silva, E.J. 2010. Can inspection methods generate valid new knowledge in HCI? The case of semiotic inspection. In: International Journal of Human-Computer Studies, 68(1-2), p. 22-40.
- de Souza, C.S. and Preece, J. 2004. A Framework for Analyzing and Understanding Online Communities. In: Interacting with Computers, v. 16(3), p. 579-610.
- Forlizzi, J. and Battarbee, K. 2004. Understanding experience in interactive systems. In Proc. of DIS '04, p. 261-268.
- Furtado, V. et al. 2010. Apoio ao projeto de sites colaborativos: uma abordagem fundamentada na associação de requisitos de colaboração e de interação com pré-padrões de interface. In Proc. of IHC'10.
- Google Play. 2016. Disponível em: <https://play.google.com/store/apps>.

- Knoll, R. C. Desenvolvimento de heurísticas de usabilidade para tablets. Caderno de Estudos Tecnológicos, FATEC, Bauru, Julho 2014. V.2, n.1, p. 93-109.
- Nielsen, J. Usability engineering. Elsevier, 1994.
- Norman, D. and Nielsen, J. 2013. The Definition of User Experience. Disponível em: <<https://goo.gl/Jx1bZj>>. Último acesso em: 29 de março de 2016.
- O Globo. 2013. Waze já tem 6 milhões de usuários no Brasil, que é o 2º mercado do app. Disponível em: <<http://goo.gl/xImxvq>>. Último acesso em: 25 de Março de 2016.
- Pereira, R.; Baranauskas, M. C. C. & da Silva, S. R. P. 2010. Softwares sociais: uma visão orientada a valores. In Proc. of IHC '10, p. 149-158.
- Pontes, P. E. 2008. Especificação de requisitos para comunicabilidade em websites na engenharia semiótica. 108 p. Faculdade de Informática do Centro Universitário Ritter dos Reis, Porto Alegre, 2008.
- Prates, R. O. and Barbosa, S. D. J. 2007. Introdução à Teoria e Prática da Interação Humano Computador fundamentada na Engenharia Semiótica. In: T. Kowaltowski e K. K. Breitman (Org.). Sociedade Brasileira de Computação. SBC. Rio de Janeiro.
- Preece, J. 2001. Online communities: Usability, Sociability, Theory and Methods. In R. Earnshaw, R. Guedj, A. van Dam and T. Vince (Eds) Frontiers of Human-Centred Computing, Online Communities and Virtual Environments.
- Reis, S. de S. and Prates, R. O. 2011. Applicability of the Semiotic Inspection Method: a Systematic Literature Review. In: Proc. of X IHC & V CLIHC'11, p. 177-186.
- Silva, M. R., Coutinho, F. R. S. and Oliveira, A. P. G. 2014. Estratégias de comunicação de marketing para o website corporativo do Promenade Champagnat: usando o MIS para o turismo. In Proc. of IHC'14, p. 333-336.
- Silva, J. C. R.; Oliveira, E, R; and Barbora, G. A. R. 2015. Caracterização de Estratégias de Gamificação em Aplicativos Móveis Educacionais: Um Estudo de Caso do Aplicativo Duolingo. In Proc. of IHC '15.
- Tarouco, F. 2013. A metrópole comunicacional e a popularização dos apps para dispositivos móveis. Seminário Internacional de Pesquisa em Comunicação. Epistemologia e desafios da pesquisa no campo da comunicação, 2013.
- W3C. 2016. Designing for inclusion. 2012. Disponível em: <<http://goo.gl/dJX65Y>>. Último acesso em: 25 de Março de 2016
- Werbach, K.; Hunter, D. For The Win: How Game Thinking Can Revolutionize Your Business. Filadélfia, Pensilvânia: Wharton Digital Press. 148 p., 2012.
- Waze. 2016. Disponível em: <<https://www.waze.com/>>. Último acesso em: 25 de Março de 2016
- Zichermann, G.; Cunningham, C. 2011. Gamification by Design - Implementing Game Mechanics in Web and Mobile Apps. (1st ed.). O'Reilly Media, Inc.

Por onde quero ir? Interesses pessoais na escolha de um trajeto urbano

Ana Paula Chaves¹, Igor Steinmacher¹, Marco Aurélio Gerosa²

¹Departamento Acadêmico de Computação (DACOM)
Universidade Tecnológica Federal do Paraná – Câmpus Campo Mourão (UTFPR-CM)
Caixa Postal: 271 – 87301-899 – Campo Mourão – PR – Brasil

²Instituto de Matemática e Estatística – Universidade de São Paulo (IME-USP)
Rua do Matão, 1010 – 05508-090 – São Paulo – SP – Brasil

{anachaves,igorfs}@utfpr.edu.br, gerosa@ime.usp.br

Abstract. *When moving throughout the city, people may have different interests. Previous researches have considered user preferences to recommend paths and destinations. However, these preferences are generally related to time, distance and cost. In this paper, we assume that other personal and social interests can influence the path choice and should be considered in customized recommendation systems. Thus, this paper presents an exploratory study aiming to analyze what people consider as their interest while choosing the path. We conducted interviews and qualitatively analyzed the data by using open and axial coding. Our results evidenced that people have different interests in different contexts and aspects that influence the choice range from the landscape to the opinion of acquaintances about the path.*

Resumo. *Quando se deslocam pelas cidades, as pessoas podem ter diferentes interesses. Pesquisas tem se preocupado em considerar as preferências dos usuários para fazer recomendações de trajetos e destinos. Entretanto, essas preferências são, em geral, relacionadas a tempo, distância e custo do trajeto. Este artigo considera que outros interesses pessoais e sociais podem ter impacto na escolha do trajeto e que, portanto, deveriam ser considerados em sistemas de recomendação personalizada. Para isso, o artigo apresenta uma pesquisa exploratória que analisa o que as pessoas consideram como seu interesse no momento da escolha do trajeto. Foram realizadas entrevistas e os dados qualitativos coletados foram analisados utilizando codificação aberta e axial. Os resultados evidenciam que as pessoas têm diferentes interesses em contextos diferentes e que os aspectos que podem influenciar a escolha vão desde a paisagem até a opinião de pessoas conhecidas sobre o trajeto.*

1. Introdução

Sistemas georreferenciados fazem parte da vida das pessoas. A popularização da tecnologia de GPSs (*Global Positioning Systems*), o uso de mapas digitais e sistemas de recomendação reduziu o esforço necessário para escolher seus trajetos [Ballatore and Bertolotto 2015]. A ampla utilização de aplicativos como Waze¹ e Google Maps²

¹ <https://www.waze.com>

² <https://maps.google.com>

mostra que as pessoas têm interesse em buscar por melhores trajetos, tanto no seu dia-a-dia quanto em situações não rotineiras.

Entretanto, pessoas têm diferentes interesses quando trafegam pelas cidades, o que traz a preocupação em gerar mapas diferentes para cada pessoa, com base no perfil do usuário [Ballatore and Bertolotto 2015][Campigotto et al. 2016]. Recentemente, pesquisadores têm se preocupado em considerar fatores como restrições do local (origem, destino, pontos de interesse etc.), restrições de tempo e preferências do usuário para oferecer recomendações de trajeto personalizadas [Chen et al. 2015]. Esses trabalhos, em geral, consideram como interesse elementos variados, relacionados às condições de trânsito e das vias, por exemplo, distância, tempo, custo, tráfego, consumo de combustível, entre outros [Campigotto et al. 2016].

Pesquisas voltadas para recomendação em redes sociais baseadas em localização (LBSN – *Location-Based Social Networks*) tendem a considerar informações do perfil social do usuário (por exemplo, utilizando algoritmos de filtragem colaborativa) e seu contexto estático (idade, gênero, profissão, histórico de locais visitados). Bao et al. (2012) utilizam opiniões de “*experts locais*” sobre pontos de interesse. Contudo, esses trabalhos ainda têm pouco enfoque no contexto dinâmico, que incluem localização atual do usuário, humor, a atividade que está realizando no momento (por exemplo, em férias ou no trabalho) e condições ambientais (como condições climáticas e relevo) [Bao et al. 2015]. Além disso, pessoas podem possuir outros tipos de preferências que influenciem na escolha do trajeto, que dizem respeito a características individuais e sociais, como paisagens que gostam de ver, lugares que gostam de frequentar, pontos de interesse que gostariam de conhecer, pessoas com quem gostariam de interagir, entre outros. Esses fatores têm sido recentemente explorados para recomendação de lugares para visita (especialmente voltados para turistas) [Chen et al. 2015; Kurashima et al. 2012; Pires et al. 2015], mas ainda são desconsiderados pelos algoritmos de recomendação de rotas. Ademais, as pesquisas não discutem sobre quais interesses poderiam influenciar a escolha menos óbvia de um trajeto, por exemplo, um trajeto mais longo ou mais demorado.

Diante deste cenário, esta pesquisa se concentra em identificar evidências empíricas de quais tipos de preferências afetam a escolha do trajeto e em que contexto elas são relevantes, respondendo à questão “*O que as pessoas consideram interessante para a escolha do trajeto durante o deslocamento urbano?*”. Para isso, o objetivo deste artigo é apresentar os resultados de uma entrevista realizada com sete sujeitos sobre seu perfil de deslocamento em trajetos rotineiros e não rotineiros. Os tipos de deslocamento foram divididos em rotineiros e não rotineiros por partir do pressuposto de que pessoas em atividades do seu dia-a-dia têm interesses diferentes na escolha do trajeto em relação a pessoas em atividades esporádicas. No contexto desta pesquisa, por **perfil de deslocamento** entendem-se as razões que levam os indivíduos a se locomoverem e a forma como escolhem seus trajetos. Para os **trajetos rotineiros**, foram consideradas quaisquer atividades que o indivíduo faz constantemente no seu dia-a-dia, ainda que não ocorram com frequência ou horários pré-determinados. Por exemplo, ir ao mercado é considerado como uma atividade rotineira, apesar de não acontecer diariamente e não ter um dia ou horário pré-definido. Para os **trajetos não rotineiros**, foram consideradas atividades imprevistas e/ou esporádicas, realizadas fora do contexto diário. Por exemplo, atividades realizadas durante o período de férias ou fins de semana, viagens esporádicas a trabalho ou turismo, entre outras. Os dados qualitativos coletados nas

entrevistas foram analisados utilizando técnicas de codificação aberta e axial [Corbin and Strauss 2014]. Os resultados destacam as diferenças de perfil de deslocamento para um mesmo indivíduo, quando em situações rotineiras ou não rotineiras, e que as características que podem influenciar vão desde as paisagens existentes no caminho até a opinião de conhecidos sobre o trajeto. Como contribuição, espera-se que este resultado forneça subsídios para novas pesquisas que objetivem a criação de modelos, métodos e processos para o desenvolvimento de sistemas de recomendação personalizada, que considerem a influência dos interesses pessoais e sociais para a escolha do trajeto.

2. Trabalhos relacionados

No deslocamento urbano, distância de um trajeto e o tempo de chegada até o destino são aspectos importantes para a escolha de um trajeto [Yang et al. 2015]. Entretanto, pesquisadores já se atentaram para o fato de que pessoas possuem diferentes preferências que influenciam essa escolha [Ballatore and Bertolotto 2015]. Yang et al (2015) afirma que o serviço de roteamento deveria apoiar mais que um critério e capturar as incertezas, incorporando as preferências e o contexto dos usuários.

De acordo com Adomavicius e Tuzhilin (2011), as abordagens para usar informações contextuais para recomendação podem ser divididas em dois grupos: pela pesquisa solicitada pelo usuário do trajeto ou pelas suas preferências informadas ou inferidas. Os contextos usualmente analisados são direcionados a horários locais, condições climáticas ou localização [Adomavicius and Tuzhilin 2011]. Os critérios elegidos como preferências em geral são relacionados à distância do trajeto, tempo de viagem, consumo de combustível, custos com pedágios, quantidade de semáforos [Yang et al. 2015]. De maneira geral, os trabalhos sugerem que o usuário tenha autonomia para decidir quais desses elementos tem mais impacto na sua escolha do trajeto para, então, recomendar a rota mais próxima da necessidade do usuário. Especificamente para congestionamentos, os trabalhos são destinados a sugerir alterações no trajeto, com base na análise dinâmica do fluxo de trânsito [Pires et al. 2015].

Sistemas para planejamento de viagens utilizam preferências por lugares e popularidade de pontos de interesse, obtidos a partir de fontes como redes sociais baseadas em localização (*Location-Based Social Networks*), georreferenciamento de fotos e histórico de viagens [Chen et al. 2015; Kurashima et al. 2012], e até mesmo a opinião de outros indivíduos sobre os pontos de interesses [Bao et al. 2012; Su et al. 2014]. Outros trabalhos sugerem o uso de sensores como RFID para monitorar o comportamento dos turistas e sugerir otimização de trajetos [Tsai and Lai 2013; Tsai and Chung 2012].

Entretanto, na literatura, há pouco conhecimento sobre as razões que levam as pessoas a elegerem um ou outro trajeto, a despeito do tempo, distância e custo. Abou-Zeid et al. (2008) sugerem que pessoas precisam de opções para que reflitam sobre a felicidade do trajeto que realizam. Existem poucas evidências que dizem quais interesses pessoais influenciam na escolha menos óbvia de um trajeto.

3. Método de Pesquisa

Foram realizadas entrevistas com indivíduos a respeito do seu padrão de deslocamento em situações rotineiras e não rotineiras. Foram entrevistados homens e mulheres, com idades entre 18 e 62 anos e com diferentes profissões (agente de viagens, engenheiro,

estudante, servidor de instituição pública, professor, aposentado). Além disso, são oriundas de cidades pequenas, médias e grandes, nos estados do Paraná e São Paulo. A escolha dos entrevistados foi realizada por conveniência, buscando por pessoas que se adequassem ao perfil desejado, garantindo que todas as características mencionadas acima estariam igualmente representadas. As entrevistas foram realizadas online, por voz, gravadas³ e transcritas.

Inicialmente, os entrevistados foram informados sobre a questão de pesquisa “*o que você considera interessante para a escolha do trajeto durante o deslocamento urbano?*”. Em seguida, foi apresentado o conceito esperado de **trajetos rotineiros**. As pessoas foram convidadas a imaginar o seu contexto diário para responder às questões. Após a primeira rodada de perguntas, o mesmo processo foi utilizado para **trajetos não rotineiros**. As questões respondidas em ambas as rodadas foram:

- Por que você se desloca?
- Como você se desloca?
- Como você escolhe o trajeto?
- O que te leva/levaria a fazer trajetos diferentes?
- Como você escolhe o seu modo de viagem?
- Você conhece/procura conhecer os arredores do seu destino?
- O que te leva/levaria a conhecer?

A seguir, serão apresentados os procedimentos de utilizados para a análise qualitativa das entrevistas.

3.1. Análise qualitativa

Após a transcrição das entrevistas, foram executadas as técnicas de codificação aberta e axial propostas por Corbin e Strauss (2014) para o método *Grounded Theory*. A codificação aberta consiste em atribuir um código substantivo ao dado, compará-lo com outros dados e designá-lo em categorias e/ou subcategorias. A codificação axial consiste na formação e desenvolvimento de um conceito, auxiliando o pesquisador a realizar a integração das categorias geradas a partir da codificação aberta, elaborando conexões entre as categorias e as sub-categorias [Cassiani et al. 1996].

Corbin e Strauss (2014) propuseram também a codificação seletiva, que é a busca pelo fenômeno central que define todas as categorias para formar a teoria fundamentada nos próprios dados. Entretanto, para o propósito desta pesquisa, as fases aberta e axial foram suficientes para responder a questão de pesquisa.

As entrevistas foram analisadas em duas etapas. Cada etapa envolveu as respostas de uma rodada de perguntas (trajetos rotineiros e não rotineiros). Os códigos foram criados a partir dos textos das entrevistas, baseados na percepção dos pesquisadores. Categorias foram utilizadas para agrupar códigos referentes a um mesmo significado conceitual. Durante a análise, conforme aumentava a interação com os dados, as categorias e subcategorias evoluíam a fim de aumentar a clareza dos conceitos. Os resultados dessa análise são apresentados na Seção 4 deste artigo.

³ As gravações foram previamente autorizadas pelos respondentes.

A etapa de codificação axial foi realizada alternadamente à codificação aberta, em um processo iterativo. Entretanto, ao final da codificação aberta, todas as relações entre categorias foram revisadas e refinadas.

4. Resultados

Para facilitar a apresentação dos conceitos, esta seção está dividida em: **Trajetos não rotineiros** e **Trajetos rotineiros**. Os **Trajetos não rotineiros** são divididos em **Trajetos a trabalho** e **Trajetos a lazer**. Essa subdivisão emergiu da análise, visto que pessoas em trajetos não rotineiros possuem perfis de deslocamento diferentes de acordo com o objetivo da atividade (trabalho ou lazer). Isso pode ser claramente observado na citação “É melhor pegar um exemplo específico, né, porque muda. Se for a trabalho é uma coisa, se for em férias é totalmente diferente” [E1]. Essa diferença se manifesta nas diversas categorias, desde o modo de deslocamento (meio de transporte) até a flexibilidade em conhecer os arredores do destino.

As demais categorias discutidas a seguir são: Motivo de deslocamento; Meio de transporte e Critério para escolha do meio de transporte; Características importantes do trajeto; Critério para escolha do caminho e Flexibilidade do trajeto; e Flexibilidade para conhecer os arredores.

4.1. Trajetos não rotineiros

As Tabelas 1 e 2 apresentam as categorias, subcategorias e códigos identificados para Trajetos não rotineiros a trabalho e para Trajetos não rotineiros a lazer, respectivamente. A análise dos dados sobre trajetos não rotineiros mostrou que o principal motivo desse tipo de deslocamento são as **Atividades de lazer**, **Atividades profissionais** e **Atividades pessoais** que acontecem esporadicamente. Para cada um desses contextos, o perfil de deslocamento e as características importantes do trajeto foram divergentes, o que destaca a relevância da atividade do indivíduo como fator de influência para a definição de seus interesses no trajeto.

Tabela 1. Categorias, subcategorias e códigos identificados para Trajetos não rotineiros a trabalho.

Categoria	Subcategoria	Códigos
Motivo do deslocamento	Atividades profissionais	Trabalho, Estudo, Compromissos,
	Atividades pessoais	Alimentação, Necessidade/obrigação, Comércio,
Meio de transporte	-	Táxi, A pé, Carro pessoal, Avião, Bicicleta
Critério para escolha do meio de transporte	-	Custo, Saúde, Distância, tempo
Flexibilidade do trajeto	Sempre o mesmo caminho	Falta de opção de trajeto
	Caminhos diferentes	Compromissos diferentes, Conhecer lugares novos
Características importantes do trajeto	Características da via	Trajetos fáceis, Trajetos não cansativos, Distância, Fluxo de Trânsito, Segurança, Condições de rodovias
	Características do local	Comércio, Atrações turísticas, Características/cultura locais, Coisas diferentes, Conhecer lugares novos,
Critérios para escolha do caminho	Segurança	Caminho mais movimentado
	Informações de terceiros	Seguir um guia, Procurar em um mapa, Perguntar a conhecidos
	Clima	Clima favorável
	-	Pessoas idosas, Tempo
	Características da via	Caminhos mais simples, Fluxo do trânsito, Distância
Flexibilidade para conhecer os arredores	Características do local	Conhecer lugares novos
	Oportunidade	Alimentação, Explorar os arredores do hotel
	Tempo	Tempo
	-	Desinteresse por locais familiares

Tabela 2: Categorias, subcategorias e códigos identificados para Trajetos não rotineiros a lazer.

Categoria	Subcategoria	Códigos
Motivo do deslocamento	Atividades turísticas	Tirar foto, Conhecer lugares novos, Observação das paisagens/locais, Conhecer lugares bonitos, Sair de casa, Conhecer lugares novos
	Atividades sociais	Visitar bares, Visitar conhecidos
	Atividades pessoais	Tomar sol
	Atividades de lazer	Lazer, Passeio
Meio de transporte	-	Carro pessoal, Carro alugado, A pé, Ônibus, Avião
Critério para a escolha do meio de transporte	-	Distância
Flexibilidade do trajeto	Caminhos diferentes	Inconvenientes no caminho, Ruas esburacadas, Pedágio
Características importantes no trajeto	Características do local	Atrações turísticas, Características/cultura locais, Coisas interessantes, Coisas diferentes, Culinária local, Paisagens, Lugares bonitos, Conhecer lugares novos
	Características das vias	Fluxo de trânsito, Segurança, Distância,
Critérios para escolha do caminho	Características do local	Visitar atrações turísticas, Paisagens, Curiosidade
	Informações de terceiros	Perguntar a conhecidos, Consultar sites de viagens, Procurar em um mapa
	Características das vias	Caminhos mais ágeis, Sinalização, Caminhos mais simples, Fluxo do trânsito, Facilidade do trajeto, Inconvenientes no caminho, Distância, Tempo
	Oportunidade	Oportunidade, Aleatória
Flexibilidade para conhecer os arredores	-	Visitar atrações turística, Tempo, Experimentar atividades novas, Conhecer lugares novos, Conhecer coisas interessantes, Curiosidade, Coisas diferentes

Quando direcionado às **Características importantes do trajeto**, a análise mostrou que as pessoas em **Trajetos a lazer** consideraram importante **Visitar atrações turísticas, Paisagens, Conhecer lugares novos**, conhecer **Características/cultura local** e conceitos relacionados. Importante notar que pessoas em **Trajetos a trabalho** também apontaram esporadicamente algumas dessas características como sendo importantes, mas poucos entrevistados relacionaram esses fatores aos **Critérios para a escolha do caminho**. Para escolher o caminho, as pessoas em **Trajetos a trabalho** salientaram fatores como o **Fluxo do trânsito**, da **Distância** do trajeto e de serem **Caminhos mais simples** ou **Caminhos mais movimentados**, “*para evitar riscos*” [E1]. Essa diferença pode ser observada nos trechos a seguir:

Trajetos a lazer:

“A passeio são as atrações turísticas, né. Cada dia você faz um caminho pra chegar em um lugar diferente.” [E1]

“Por exemplo, lá em [cidade turística de litoral], as vezes que eu andei por lá pra ir conhecer os lugares, eu procurava sempre ir por onde tinha mais lugares interessantes pra eu conhecer né. Tanto a praia quanto os pontos turísticos de lá.” [E7]

“Ah, eu procuro paisagens, né. Imagens bonitas, pode ser tanto em cidade como floresta, né, você ver uma colina, uma paisagem de campo, uma paisagem de cidade, alguns prédios. Cada um tem seu tipo de beleza.” [E5]

Trajetos a trabalho:

“Podem ter vários fatores aí influenciando, né. A questão da distância, do quão otimizado você pode fazer o trajeto. Como que é a questão do fluxo [do trânsito]. Se for uma coisa que você tem que ir em horários não tão comuns ou pelo menos um pouco mais restritos, é ver a questão de segurança também. [...]” [E6]

A despeito da preocupação com os fatores convencionais do trajeto, quando questionados sobre a **Flexibilidade do trajeto** e os **Critérios para a escolha do trajeto**, embora pessoas em **Trajeto a trabalho** tenham destacado a importância de fazer **Sempre o mesmo caminho** e afirmado escolher **Caminhos mais simples** ou **Caminhos mais movimentados**, por razões de **Segurança** e **Tempo**, elas também apontaram para o interesse de conhecer os arredores, como pode ser observado nos trechos abaixo:

“No trabalho, [procuro conhecer] sim, perto do hotel, geralmente, que é onde a gente tem oportunidade.” [E1].

“Ali próximo ao hotel, eu procuro [conhecer]. Eu não abro muito o raio não.” [E6]

“Geralmente não dá tempo, mas gostaria sim, de conhecer tudo.” [E2]

Nesses casos, a **Oportunidade** foi mencionada como uma característica importante. Além de **Explorar os arredores do hotel**, pessoas que fazem **Trajeto a trabalho** apontaram a **Alimentação** como um fator que colabora com a **Flexibilidade para conhecer os arredores**. Por exemplo, a entrevistada [E1] afirmou que quando está a trabalho, aproveita para conhecer os arredores quando sai para comer, *“porque a única coisa que a gente faz além de trabalhar é comer” [E1].*

Outra característica interessante dos **Trajeto não rotineiros** foi o **Des/interesse por locais familiares**. Por familiares, entende-se locais não rotineiros, porém que o indivíduo tem a sensação de ser um local conhecido. Por exemplo, bairros ou ruas desconhecidas na sua própria cidade. Alguns entrevistados afirmaram que quando precisam realizar **Atividades pessoais** esporádica, não observam o trajeto, ainda que seja a primeira vez que passa pelo local. Isso pode ser visto no trecho a seguir:

“Então daí também depende do local onde eu estou [...]. Por exemplo, em [cidade em que moro] mesmo, quando eu preciso ir em um lugar onde eu não vou com tanta frequência, eu quase nem reparo nas coisas assim em volta [...]. Agora quando eu fui pra [cidade turística][por motivo profissional] eu já reparava tudo, muito mais, né. Eu olhava tudo.” [E7]

Entretanto, o entrevistado E3 afirmou ser *“bem observador”* e reconhecer *“pelo menos as paisagens”* dos locais por onde transita. Essa contradição mostra que o perfil individual é relevante.

Nos **Trajeto a lazer**, o tempo deixa de ser um aspecto restritivo. A inversão da relação com o tempo faz com que as pessoas estejam mais dispostas a observar. As pessoas se mostraram flexíveis a fazer **Caminhos diferentes** (nenhum entrevistado mencionou fazer sempre o mesmo caminho), utilizando como **Crítério para escolha do caminho** conceitos como **Visitar atrações turísticas**, **Curiosidade**, **Sinalização** e até mesmo uma escolha **Aleatória**, como a citação abaixo:

“[Escolho o caminho] Aleatoriamente. Nem que seja o mais longo. Porque daí eu vou acabar vendo outras coisas.” [E4]

Nesse tipo de trajeto, as pessoas disseram que procuram **Conhecer bem os arredores**, têm pré-disposição para **Experimentar atividades novas** e fazem caminhos diferentes com o objetivo de **Conhecer lugares novos**. Em locais históricos ou

turísticos, os caminhos mais interessantes são os que possibilitam **Visitar as atrações turísticas**. Esses conceitos podem ser observados nas citações a seguir:

“Mas assim, na questão de estar lá em [cidade turística], se fosse um caminho mais distante, eu procuraria fazer trajetos diferentes, pra conhecer lugares diferentes.” [E7]

“Se tiver já mais uma flexibilidade maior de tempo ou de dias, e também talvez se tiver alguma outra motivação, tipo, alguém comentar alguma coisa assim 'ah, se você mudar o caminho ou se for por esse outro caminho tem alguma coisa interessante pra se conhecer ou pra se ver', isso de repente pode me motivar a fazer algum trajeto alternativo.” [E6]

“Então, é mais pra conhecer mais lugar, pra ver se identifica com algum... alguma paisagem. Procuo encontrar um lugar que me faça sentir bem, né. Isso que me atrai.” [E5]

É importante observar pelas citações que a novidade é um conceito relevante para a definição de interesse. De maneira geral, o interesse ou preferência de um indivíduo é definido pela escolha de algo que ele já conhece e lhe agrada. Entretanto, quando estão em **Trajetos a lazer**, as pessoas demonstraram fazer escolhas a despeito dos seus interesses. Isso foi especialmente observado nos contextos em que o entrevistado descreveu o Trajeto não rotineiro como uma **Oportunidade** (por exemplo, quando estavam em um local distante de seu local de moradia), o que fica evidente no trecho a seguir:

“[Faço caminhos diferentes] Pra ver se tem alguma coisa legal! Porque daí eu não sei se eu vou passar lá de novo! Vai que tem alguma coisa legal!” [E4]

Em decorrência disso, as **Atividades de lazer e Turísticas** mencionadas foram as mais variadas, como simplesmente **Sair de casa, Experimentar a culinária local, Visitar bares, Restaurantes, Lojas, Atrações turísticas**, entre outros. As citações a seguir são exemplos dessas atividades:

“Quando eu estou de férias: ah, daí [eu me desloco] pra beber, pra tomar sol, pra passear, pra tirar foto, pra.... risos” [E1]

“Agora quando você está passeando, você às vezes... sair de casa é o objetivo. Então você não vai com pressa, você está andando. [...] Você não está nem aí com o resto, você está ali só mesmo curtindo o lugar né.” [E3]

Quando se trata de viagens, a diferença de perfil de deslocamento de acordo com o tipo do trajeto ocorre até mesmo na escolha do meio de transporte. Primeiramente, é fato que a escolha do meio de transporte da viagem pode justificar a utilização de meios de transporte alternativos no destino. Por exemplo, pessoas que viajam de avião, recorrem a táxis e ônibus no destino, o que não acontece com pessoas que viajam com carro próprio. No caso dos **Trajetos a trabalho** o **Táxi** foi mencionado como meio de transporte. Já as pessoas em **Trajetos a lazer** disseram utilizar **Carro alugado, Ônibus** e até mesmo caminhar **A pé**. Foi evidenciado também a pré-disposição dos indivíduos de utilizar meios de transporte mais lentos quando estão em **Trajetos de lazer**.

“E geralmente é de carro. Às vezes a pé se for próximo, mas senão, é de carro. Olha quando você viaja de avião, aí vai de ônibus, né [...]” [E3]

Além disso, a questão do **Custo** apareceu uma única vez, relacionados a **Trajetos a trabalho**, quando o entrevistado afirmou utilizar o **Táxi** porque o custo com transporte era pago pelo empregador.

A **Segurança** também foi mencionada por pessoas em **Trajetos não rotineiros** como uma característica relevante para a escolha do caminho:

“A trabalho eu faço sempre o mesmo [trajeto] e procuro o mais movimentado, por não conhecer, pra evitar riscos.” [E1]

“Eu procuro usar o caminho que eu tenho certeza que dá certo, né.” [E5]

Finalmente, a análise destacou que especificamente para **Trajetos não rotineiros**, tanto para atividades de lazer quanto para trabalho ou atividades pessoais esporádicas, as pessoas usualmente buscam informações sobre o trajeto antes de escolher. As pessoas disseram **Perguntar a conhecidos** ou **Procurar em um mapa** o melhor caminho a seguir. Esses conceitos são exemplificados nos trechos abaixo:

“Se você tem alguém que você conhece e que já foi, você vai perguntar quais são os lugares legais. A gente usa muito o [site de viagens e destinos] como foco de informação, muito mesmo, pra descobrir tudo, onde abastecer, onde visitar, onde comer, onde se hospedar, enfim, tudo.” [E1]

Especificamente com relação ao mapa, um dos entrevistados ressaltou o aspecto negativo da característica estática dos mapas, como pode ser observado a seguir:

“Distância do trajeto né. Porque você olha no mapa, né, parece que é tudo seguro. Mas às vezes você tá num caminho, daí você não acha seguro. Mas quando eu procuro pelo mapa é pela distância percorrida e não pela segurança da via.” [E5]

4.2. Trajetos rotineiros

As categorias, subcategorias e códigos identificados para Trajetos rotineiros podem ser encontrados na Tabela 3. Quando analisando as respostas relacionadas a **Trajetos rotineiros**, pode-se concluir que por rotina entende-se, primariamente, **Atividades profissionais** (trabalho e estudo). Exceção a essa regra foi a entrevistada E2 cuja profissão é organizar excursões turísticas. Embora ela considere as viagens como seu trabalho, é profissional autônoma e organiza excursões sob demanda. Sendo assim, quando questionada sobre sua rotina, ela afirmou *“Me desloco muito pouco. Saio para igreja, banco, visitar família.” [E2]*, o que demonstra que ela desconsiderou suas viagens como parte de sua rotina. Além disso, as pessoas consideraram como rotineiras atividades pessoais (religiosas, sociais, domésticas e físicas) e alimentação. Outras atividades rotineiras mencionadas estão relacionadas à **Atividades domésticas** (como fazer compras), **Atividades religiosas** e à **Atividades físicas** (como natação, pilates e academia).

Quando o foco é direcionado à questão de pesquisa *“O que as pessoas consideram interessante para a escolha do trajeto durante o deslocamento urbano?”* a análise mostrou que o conceito chave é, de fato, o **Tempo**. Todas as **Características importantes do trajeto** possuíam alguma relação com a redução do tempo do trajeto.

Tabela 3: Categorias, subcategorias e códigos identificados para Trajetos rotineiros

Categorias	Subcategorias	Códigos
Motivo do deslocamento	Atividades pessoais	Alimentação, Visitar família, Médico
	Atividades domésticas	Banco, Compras/Comércio
	Atividades profissionais	Trabalho, Estudo
	Atividades religiosas	Igreja
	Atividades físicas	Natação, Pilates, Corrida, Ciclismo, Academia
Meio de transporte	-	Carro, A pé, Moto, Bicicleta, Ônibus
Critério para escolha do meio de transporte	-	Tempo, Volumes para carga, Praticidade,
Flexibilidade do trajeto	Sempre o mesmo caminho	Falta de opção de trajeto, Hábito,
	Caminhos diferentes	Otimização da rotina
Características importantes do trajeto	Alimentação	Alimentação saudável
	Características das vias	Facilidade de acesso a bairro, Exclusividade de veículos leves, Fluxo do trânsito, Sincronismo dos semáforos, Prioridade em determinada direção das vias, Sinalização, Bom para caminhada
	Segurança	Segurança
Critério para escolha do caminho	Características das vias	Quantidade de semáforo, Velocidade da via, Fluxo do trânsito
	Características do local	Relevo
	Tempo	Tempo, Otimização da rotina, Distância
	-	Aleatório, Disposição física
Flexibilidade para conhecer os arredores	Caminhos diferentes	Pontos de referência, Alimentação, Otimização da rotina, Observação de paisagens/locais, Buscar por utilidades do dia-a-dia, Curiosidade, Explorar ruas novas, Desinteresse em locais familiares

Ademais, alguns **Crítérios para a escolha do trajeto** também circundam a questão do **Tempo**. O tempo foi mencionado explicitamente como um critério, por exemplo, no excerto “*É, eu acho que o princípio que norteia o caminho é o tempo que você leva pra sair de um lugar e chegar no outro. Isso é o mais importante. O maior objetivo é chegar mais rápido no lugar que você planejou.*” [E3].

Em outros casos, o **Tempo** é associado a outros critérios, como **Fluxo do trânsito** e o **Relevo**. Por exemplo, o entrevistado E5 afirmou fazer um trajeto mais longo rotineiramente, porque utiliza bicicleta e o relevo da cidade faz o caminho mais longo ser mais rápido.

“Então, [eu uso] bicicleta, [...] na maioria das vezes eu faço o mesmo caminho, não mudo muito não. Porque tem o relevo da cidade, né [...] Eu escolho uma [rua] que parece mais plana pra eu tentar ir em cima da bicicleta, pra ir mais rápido. [...] Se sáisse minha moto do consórcio eu com certeza faria outro caminho. [...] Porque daí eu ia mais direto, né? Seria o menor caminho pra ir.” [E5]

O trecho anterior mostra a maior relevância do **Tempo** com relação à **Distância** do percurso. Em outras características, isso também foi observado. O **Tempo** norteia a escolha do meio de transporte, a escolha do caminho, a flexibilidade para fazer trajetos diferentes. Por exemplo, a entrevistada E6 afirmou fazer sempre o mesmo caminho, porque “*agora a coisa já está mais ou menos esquematizada.*” [E6]. Entretanto, ela afirma que esporadicamente tenta caminhos alternativos, para “*explorar alguma rua nova que foi aberta ali no bairro*”. Quando questionada sobre os interesses que a levam a explorar, a entrevistada retomou a questão do tempo:

“Por exemplo, como é que está o fluxo de trânsito, a sinalização. Porque às vezes esses caminhos alternativos podem se tornar mais extensos, mas devido à redução no fluxo de trânsito, a gente pode gastar menos tempo.” [E6]

Alguns entrevistados mencionaram a **Falta de opções de trajeto** em situações rotineiras. Essa característica está supostamente relacionada ao perfil dos entrevistados, que ou utiliza rodovia para chegar ao trabalho, ou reside muito próximo dos seus destinos rotineiros. Nestes casos, as pessoas afirmaram conhecer bem os arredores do destino e não prestar atenção nas características do trajeto.

“É, eu não lembro nome de rua, mas eu conheço tudo.” [E5]

Por outro lado, a análise mostrou que as pessoas em **Trajetos rotineiros** têm menos interesse em observar o trajeto, o que deve supostamente ocorrer pelo **Desinteresse em locais familiares**. Nenhum dos entrevistados mencionou, como Características importantes do trajeto, conceitos relacionados às **Características do local**. Vale observar que todas as pessoas que disseram conhecer todas as localidades ao redor dos destinos rotineiros são moradoras de cidades pequenas (até 100 mil habitantes) e vivem na mesma cidade há muitos anos. Essa característica de perfil pode ter sido determinante para este fator. Os demais entrevistados afirmaram que as **Características importantes do trajeto** e a **Flexibilidade para conhecer os arredores** são relacionadas a **Pontos de referência** ou **Buscar utilidades do dia-a-dia**, como mercado, posto de gasolina, pet shop, restaurantes ou igrejas.

“[Conheço] só as principais referências, não exploro o lugar.” [E1]

“Olha, o que me chama a atenção são coisas do meu dia-a-dia. Posto de gasolina, mercado, farmácia, pet shop, às vezes a gente está... tem dias que a gente está com mais pressa e não repara muito nas coisas que estão na rua, mas às vezes você está passando por aquela rua pela milésima vez, aí você fala assim ‘nossa, ali tem uma igreja!’. Nunca tinha visto que tinha uma igreja ali. A gente acaba descobrindo as coisas assim, com o passar do tempo, né. Mas, às vezes tem outro prédio assim que a gente não presta muita atenção.” [E3]

A imersão das pessoas nas questões do trabalho e responsabilidades, bloqueando a atenção a outros aspectos interessantes do trajeto, foi curiosamente observado na entrevista com a agente de viagens. Como foi anteriormente discutido, ela não considerou suas excursões como trajetos rotineiros, apesar de ser seu trabalho. Entretanto, quando questionada sobre as **Características importantes dos trajetos**, ela não mencionou as **Atrações turísticas** dos locais que visita. Pelo contrário, ela demonstrou preocupação com a **Segurança** e disse que a o mais importante no caminho é ser **Bom para caminhada**, como pode ser observado nas citações abaixo:

“Como sempre tenho pessoas idosas sempre peço um caminho que não seja difícil e cansativo. Seriam os melhores caminhos. [...] Como sempre tenho um guia que acompanha em viagens vou onde eles me levam. Geralmente não conheço o caminho. Mesmo sendo a mesma viagem duas ou três vezes, sigo eles. Meu trabalho. Não viajo sozinha nunca.” [E2]

Outros conceitos apontados pelos respondentes foram a necessidade de **Compras/Comércio** e a **Alimentação**. Especificamente sobre a **Alimentação**, opções diferentes de podem fazer com que as pessoas optem por outro trajeto. A citação a seguir aponta para essa questão:

“Mais [importante é] restaurante. Acho que alimentação de um modo geral, das coisas mais simples. Por exemplo, coisas que faltam aqui em

[cidade que mora], opções de alimentação saudável. Se tivesse um lugar assim eu procuraria.” [E1]

5. Discussão

A análise das entrevistas forneceu diversos indícios de aspectos que podem ser considerados interessantes para pessoas que estão se deslocando pelas cidades. O primeiro resultado importante foi a confirmação da diferença entre os perfis de deslocamento de pessoas em situações rotineiras daquelas em situações não rotineiras. Mais do que isso, a análise mostrou diferenças significativas no interesse de pessoas em trajetos não rotineiros a trabalho e a lazer. Abou-Zeid et al. (2008) sugere que quando pessoas estão em sua rotina, elas não consideram completamente sua “felicidade com a viagem”, o que pode explicar as diferenças no perfil de deslocamento. As pessoas que estão a lazer, ou que estão em lugares diferentes, ficam mais suscetíveis às influências que as características do trajeto podem oferecer, como as diferentes paisagens, estabelecimentos e atividades locais.

Para pessoas em trajetos não rotineiros, um aspecto importante para a recomendação de trajetos personalizados é explorar a **Oportunidade**. Especialmente para os trajetos a trabalho, como o tempo de deslocamento e os horários são fatores restritivos, as pessoas disseram conhecer os lugares em que tem oportunidade (próximo ao hotel, próximo a restaurantes). Sistemas de recomendação de trajetos poderiam utilizar recomendações de pontos de interesses próximos para sugerir um trajeto alternativo que explore as características locais (pontos turísticos, paisagens urbanas/rurais, cultura). Neste caso, embora o interesse dos indivíduos precise ser considerado, a análise mostrou que as pessoas têm curiosidade de conhecer “coisas novas” ou “diferentes” que caracterizem o local que visitam. Portanto, essa recomendação deveria ser realizada a despeito dos interesses do indivíduo. A recomendação não deveria impedir que o usuário experimente algo novo [Ballatore and Bertolotto 2015]. Uma maneira de selecionar os locais de interesse próximos poderia ser através do histórico de visitas e similaridade de perfis (utilizando algoritmos de filtragem colaborativa), considerando suas restrições de tempo, segurança e disponibilidade de meio de transporte.

Para pessoas em trajetos a lazer, a questão do tempo é menos importante, o que pode ser destacado pelo fato de as pessoas às vezes optarem por meios de transporte mais lentos do que os que usam em sua rotina. Sistemas de Recomendação de trajetos personalizados poderiam melhorar a experiência do usuário durante o seu passeio criando trajetos semelhantes aos propostos acima (para trajetos não rotineiros de trabalho), porém com maior flexibilidade de tempo e opções de meio de transporte (por exemplo, poderia sugerir uma caminhada por alguma rua específica). As questões de segurança continuam sendo importantes. As recomendações poderiam incluir contextos como a quantidade de dias que o usuário tem disponível no lugar, o tempo que as pessoas costumam demorar em cada ponto de interesse (que pode ser obtido com histórico de visita), considerar a chegada em restaurantes próximo dos horários de refeições, além de sugerir trajetos cíclicos, que retornem ao ponto de partida no final do percurso, melhorando o aproveitamento da viagem.

Outro conceito importante destacado pela análise foi o fato de as pessoas habitualmente **Perguntarem a conhecidos** sobre os locais não rotineiros. Técnicas de

crowdsourcing poderiam ser utilizadas para apoiar a recomendação. Utilizar contatos em redes sociais para identificar pessoas conhecidas e sugerir pessoas que já visitaram o destino também são opções que podem ser exploradas.

Com relação aos trajetos rotineiros, os resultados mostram algo esperado: o impacto do tempo nas escolhas do trajeto. Ferramentas comerciais de recomendação de rotas, em geral, focam neste aspecto. Entretanto, os resultados mostram também a dificuldade em observar outros aspectos do trajeto. A pesquisa apresentada em [Abou-Zeid et al. 2008] mostra que uma mudança temporária no comportamento deve ser efetiva para induzir uma modificação comportamental ou minimamente afetar a escolha das pessoas. Por exemplo, a entrevistada E1, apesar de ter apontado a alimentação próxima ao destino como uma forma de otimização da rotina, disse que procuraria por lugares que oferecessem uma “*opção de alimentação saudável*”. Talvez oferecer uma opção de trajeto com alimentação saudável possa ser a “*janela da mudança*” que a entrevistada precisa para reconsiderar sua escolha de caminho. Neste caso, é importante que sejam realizadas novas pesquisas direcionadas especificamente para pessoas em trajetos rotineiros, sugerindo pequenas variações em seus trajetos e verificando seu comportamento com relação às mudanças. Uma abordagem poderia utilizar os interesses das pessoas extraídos de redes sociais para sugerir novos cenários que façam com que a pessoa reflita sobre seu trajeto e avaliar se essa reflexão pode gerar novas escolhas que não sejam exatamente influenciadas apenas pelo tempo do trajeto.

6. Limitações

Como passo inicial desta pesquisa, foram entrevistados sete sujeitos. Com apenas essas entrevistas não é possível afirmar que se alcançou a saturação teórica, conforme pregado pelas *Grounded Theory*. Para os Trajetos não rotineiros, é necessário que se defina um número maior de perfis possíveis, já que a diferença entre perfis pode impactar no surgimento de novos conceitos. Entretanto, para os Trajetos rotineiros, apesar das diferenças de perfis dos entrevistados, a pesquisa chegou muito próxima da saturação teórica. Isso se deve ao instrumento de pesquisa utilizado. Conforme discutido na Seção 6, as entrevistas mostraram que, quando convidadas a raciocinar sobre seus interesses no trajeto, as pessoas em Trajetos rotineiros têm dificuldades em observar aspectos que não estejam relacionados ao Tempo. Sendo assim, utilizando o instrumento de pesquisa apresentado neste artigo, algumas poucas entrevistas seriam necessárias para chegar à saturação teórica. Todavia, é necessário criar novos instrumentos que auxiliem pessoas em trajetos rotineiros a considerar outros aspectos, mais relacionados à satisfação pessoal, conforto e felicidade no trajeto. Mais entrevistas estão planejadas a fim de explorar este e outros dos pontos levantados durante a discussão.

Além disso, pode-se dizer que existe um potencial viés de seleção dos participantes, já que os perfis analisados podem não representar todos os possíveis perfis de pessoas. Isso se justifica pelo caráter exploratório preliminar desta pesquisa. Como para esta etapa da pesquisa seria inviável ter uma amostra que represente toda a população e, visando aliviar o problema, definiu-se um escopo específico e características que foram consideradas inicialmente relevantes. Dentro dos perfis definidos, foram convidadas pessoas que se encaixassem nas características.

Outra potencial ameaça à validade está relacionada à subjetividade da classificação dos dados. A pesquisa utilizou procedimentos de *Grounded Theory* que, para mitigar essa ameaça (comum a métodos qualitativos, com caráter interpretativistas), prega que toda a análise seja fundamentada nos dados coletados. Além disso, houve discussão sobre o processo de análise com um especialista em *Grounded Theory* durante a realização das etapas do método.

7. Conclusão

Neste trabalho foi investigado “*O que as pessoas consideram interessante para a escolha do trajeto durante o deslocamento urbano?*”. Para responder à questão, foram realizadas entrevistas e os dados qualitativos coletados foram analisados usando codificação aberta e axial. Os resultados, resumidos nas Tabelas 1, 2 e 3 da Seção 4, mostraram que quando não influenciadas pelo tempo, as pessoas observam paisagens, pontos de interesse, tem disposição de experimentar coisas novas (culinária, atividades) e que, além de consultar mapas, as opiniões de outras pessoas são importantes fontes de informação no momento de escolher um caminho ou destino desconhecido. Além disso, foi possível observar que as pessoas em sua rotina têm dificuldade em observar aspectos que não sejam relacionados ao tempo de deslocamento.

Este trabalho consiste em uma pesquisa preliminar de caráter exploratório, que visa fornecer subsídios para as próximas etapas desta pesquisa. Novas pesquisas precisam ser realizadas para confirmar os resultados obtidos. Como potencial contribuição deste trabalho tem-se que, para sugerir trajetos personalizados, as informações sobre os interesses individuais e sociais são aspectos que podem ser obtidos por meio de informações disponíveis em redes sociais, em sugestões de contatos e de outras pessoas com similaridades de perfil. É necessário estudar o uso de algoritmos de filtragem colaborativa e informações de contexto para relacionar essas informações aos aspectos dinâmicos, como atividade que o usuário está realizando, meio de transporte, entre outros. Com isso, espera-se que seja possível utilizar os interesses pessoais e sociais dos indivíduos na recomendação de trajetos, melhorando sua experiência no deslocamento urbano.

Agradecimento

A autora Ana Paula Chaves agradece à Fundação Araucária e à Secretaria de Estado da Ciência, Tecnologia e Ensino Superior do Paraná (SETI), pelo apoio financeiro.

Referências

- Abou-Zeid, M., Ben-Akiva, M. and Bierlaire, M. (2008). “Happiness and travel behavior modification”. In European Transport Conference.
- Adomavicius, G. and Tuzhilin, A. (2011). “Context-aware recommender systems”. *Recommender systems handbook*. Springer. p. 217–253.
- Ballatore, A. and Bertolotto, M. (nov 2015). “Personalizing Maps”. *Communications of ACM*, v. 58, n. 12, p. 68–74.
- Bao, J., Zheng, Y. and Mokbel, M. F. (2012). “Location-based and Preference-aware Recommendation Using Sparse Geo-social Networking Data”. In Proceedings of the

- 20th International Conference on Advances in Geographic Information Systems. , SIGSPATIAL '12. ACM.
- Bao, J., Zheng, Y., Wilkie, D. and Mokbel, M. (2015). “Recommendations in location-based social networks: a survey”. *GeoInformatica*, v. 19, n. 3, p. 525–565.
- Campigotto, P., Rudloff, C., Leodolter, M. and Bauer, D. (2016). “Personalized and situation-aware multimodal route recommendations: the FAVOUR algorithm”. *CoRR*, v. abs/1602.09076.
- Cassiani, S. H. D. B., Caliri, M. H. L. and Pelá, N. T. R. (1996). A teoria fundamentada nos dados como abordagem da pesquisa interpretativa. *Revista Latino-Americana de Enfermagem*, v. 4, p. 75–88.
- Chen, C., Zhang, D., Guo, B., et al. (2015). “TripPlanner: Personalized Trip Planning Leveraging Heterogeneous Crowdsourced Digital Footprints”. *IEEE Transactions on Intelligent Transportation Systems*, v. 16, n. 3, p. 1259–1273.
- Corbin, J. and Strauss, A. (2014). “Basics of qualitative research: Techniques and procedures for developing grounded theory”. Sage publications.
- Kurashima, T., Iwata, T., Irie, G. and Fujimura, K. (2012). “Travel route recommendation using geotagged photos”. *Knowledge and Information Systems*, v. 37, n. 1, p. 37–60.
- Pires, C. S., Aguiar, M. S. De and Ferreira, P. R. (2015). “Mobile ACORoute-Route Recommendation Based on Communication by Pheromones”. *Polibits*, n. 51, p. 27–32.
- Su, H., Zheng, K., Huang, J., et al. (2014). “Crowdplanner: A crowd-based route recommendation system”. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*. . IEEE.
- Tsai, C. Y. and Lai, B. H. (2013). “A Customized Visiting Route Service under RFID Environment”. In *Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual*.
- Tsai, C.-Y. and Chung, S.-H. (2012). “A personalized route recommendation service for theme parks using {RFID} information and tourist behavior”. *Decision Support Systems*, v. 52, n. 2, p. 514–527.
- Yang, B., Guo, C., Ma, Y. and Jensen, C. S. (2015). “Toward Personalized, Context-aware Routing”. *The VLDB Journal*, v. 24, n. 2, p. 297–318.

PoI: uma Aplicação de Detecção de Pontos de Interesse

Alice A. F. Menezes^{1,3}, Thais G. Almeida¹, Bernardo B. Gatto¹,
Eulanda M. dos Santos¹, Carlos Maurício S. Figueiredo², Eduardo F. Nakamura¹

¹Universidade Federal do Amazonas (UFAM) – Manaus, AM – Brasil

²Universidade do Estado do Amazonas (UEA) – Manaus, AM – Brasil

³Fundação Paulo Feitoza (FPF Tech) – Manaus, AM – Brasil

{alice.menezes, tga, bernardo, emsantos, nakamura}@icompu.fam.br,
cfigueiredo@uea.edu.br

Abstract. *With the emergence of social networks such as Twitter, Instagram and Foursquare, there was also the introduction of the concept of Participatory Sensor Networks (PSNs). Through this type of sensing, it is possible to collect general information of a particular location, using data provided by users and their mobile devices. Based on a data set of 333,286 check-ins, we present an application that uses Foursquare data to provide the Points of Interest (PoI) in Brazilian capitals, in other words, the most frequented venues, based on the number of check-ins. The aim is to provide the necessary information for tourism and advertising campaigns.*

Resumo. *Com o surgimento de redes sociais, como o Twitter, o Instagram e o Foursquare, houve também a introdução do conceito de Redes de Sensores Participativas (RSPs). Através deste tipo de sensoriamento, é possível coletar informações gerais de uma determinada localidade, utilizando dados disponibilizados por usuários e seus dispositivos móveis. Baseado em um conjunto de dados de 333.286 check-ins, apresentamos uma aplicação que, através de dados do Foursquare, fornece os Pontos de Interesse em capitais brasileiras, ou seja, os locais mais frequentados, baseado no número de check-ins do local. O objetivo é prover as informações necessárias para o turismo e para campanhas publicitárias.*

1. Introdução

Com a evolução dos celulares, hoje chamados de *smartphones*, tornou-se possível a detecção de dados relacionados aos usuários, como as suas opções e sua localização [Silva et al. 2014]. A introdução de sensores como GPS, acelerômetro, microfone, câmera e giroscópio nestes dispositivos [Lane et al. 2010], permitiu o sensoriamento de diversas áreas, em sua maioria urbanas. Esta rede de sensores, composta por todos os integrantes de uma sociedade que possuem dispositivos móveis, é denominada Rede de Sensores Participativa [Silva et al. 2013a].

Nas Redes de Sensores Participativas (RSPs), os usuários são entidades móveis autônomas capazes de sensoriar todo o ambiente no qual estão inseridos [Silva et al. 2014]. Assim, é possível adquirir informações relacionadas a uma determinada localidade através de dados de redes sociais e de dispositivos móveis utilizados por estes usuários.

A Figura 1 apresenta uma visão geral das RSPs, mostrando que os usuários agem como nós sensores, assim como as redes sociais agem como nós *sink* (sensores centralizadores de dados). Desta forma, a coleta de dados através de redes sociais permite identificar condições diversas como o trânsito, o clima, os eventos naturais, os aspectos econômicos, entre outros fatores.

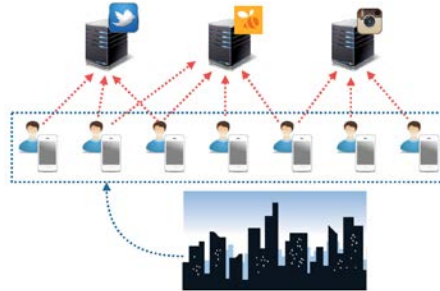


Figura 1. Visão geral das Redes de Sensores Participativas.

Outro aspecto importante das Redes de Sensores Participativas, é que os dados podem ser coletados para análise assim que o usuário os disponibiliza nas redes sociais. Isto permite a criação de soluções que provêm informações em tempo hábil para a tomada de decisões.

As principais contribuições deste trabalho são: um modelo de detecção de Pontos de Interesse, através dos dados da RSP do Foursquare, e um modelo de inferência de idade de usuários em redes sociais através de imagens. Vale ressaltar que o segundo modelo, é utilizado para prover informações auxiliares para o primeiro modelo.

Como resultado, apresentamos a aplicação PoI, cujo público-alvo é composto por turistas, empresários e órgãos governamentais que desejem identificar pontos turísticos e os locais mais frequentados por um público específico, através de dados como idade, sexo e categoria do local frequentado, tornando possível a realização de campanhas e divulgação de produtos.

Para a aplicação PoI, os Pontos de Interesse são os locais mais frequentados de uma região. Neste trabalho, fazemos a medição dos locais mais frequentados através da quantidade de *check-ins* do Foursquare que estes locais possuem em um determinado período, supondo que quanto maior é o número de *check-ins*, mais frequentado é o local.

As próximas seções deste trabalho estão divididas da seguinte forma: na seção 2 apresentamos os trabalhos relacionados, na seção 3 apresentamos o método proposto e a arquitetura da aplicação. Na seção 4, apresentamos a base de dados utilizada, os experimentos realizados e o resultados obtidos. Por fim, apresentamos as conclusões e as direções futuras para este trabalho.

2. Trabalhos Relacionados

Nesta seção é apresentado um resumo das soluções recentes para Pontos de Interesse e para soluções de aprendizagem de máquina para características visuais.

2.1. Pontos de Interesse

No trabalho proposto por [Lee and Sumiya 2010], os autores apresentam um modelo de pesquisa para detecção de eventos locais, utilizando experiências coletivas de uma

população através de dados compartilhados no Twitter.

O modelo proposto consiste em 3 etapas:

1. Coleta de dados geo-localizados através de um sistema de monitoramento de redes sociais para *microblogging*;
2. Identificação de regiões de interesse (RoIs) dos usuários do Twitter, através de medidas de regularidades do comportamento populacional;
3. Detecção de eventos geo-sociais, através da comparação entre medidas de regularidades.

Os autores utilizaram o algoritmo de aprendizagem de máquina K-means para agrupar os *tweets* geo-localizados obtidos através do sistema de monitoramento, e assim, identificar as RoIs. Vale a pena ressaltar que diferentemente deste trabalho, onde os autores definiram RoIs como sendo regiões compostas por aglomerações de usuários, o trabalho proposto neste artigo assume que Pontos de Interesse consistem em localidades (e.g., restaurantes, parques, teatros) onde um número expressivo de usuários realizou *check-in*.

Como experimento, os autores fizeram uma lista de 15 eventos festivos que iriam ocorrer no Japão, para verificar se o modelo proposto conseguiria determinar a existência desses eventos. Como resultado, foram coletados 9 eventos festivos (60% de acurácia). Porém, através do modelo, também foi possível detectar a ocorrência de eventos naturais que ocorreram no mesmo período.

Já no trabalho proposto por [Silva et al. 2013b], os autores utilizaram a RSP do Instagram, visando mostrar os desafios e as oportunidades que emergem do sensoriamento participativo realizados através dos usuários desta rede.

Para os experimentos, foram coletadas 2,3 milhões de fotos do Instagram utilizando o Twitter, o que mostra a abrangência da rede. Entre 30 de junho e 31 de julho de 2012, foram coletados 2.272.556 *tweets* contendo fotos georeferenciadas, postadas por 482.629 usuários. Cada *tweet* é composto de coordenadas (latitude e longitude) e o horário de compartilhamento da foto. Através dos dados coletados, foi possível construir uma aplicação que identifica os Regiões de Interesse em uma localidade.

O método utiliza um algoritmo de agrupamento para identificar Regiões de Interesse a partir das fotos do Instagram. Com isto, a popularidade de uma região baseia-se no número de diferentes usuários que compartilharam uma foto na mesma. Vale ressaltar que o nosso trabalho, diferentemente do trabalho descrito anteriormente, utiliza dados provenientes do Foursquare para identificar Pontos de Interesse em uma localidade, baseando-se apenas na quantidade de *check-ins* do local.

Em outro trabalho, a abordagem apresentada por [Frias-Martinez et al. 2012], utiliza dados geo-localizados que levam em consideração a privacidade do usuário. O método proposto, identifica automaticamente as atividades mais comuns entre os usuários de uma área específica, através da padronização dos dados provenientes do Twitter.

O método também identifica Pontos de Interesse automaticamente, sendo que estes possuem uma grande quantidade de *tweets* relacionados a eles. Para estudo de caso, foram utilizados dados da cidade de Nova Iorque em um período de 49 dias.

2.2. Características Visuais

Em [Jou et al. 2015], é proposto um modelo onde conjuntos de conceitos visuais (imagens onde o seu contexto é conhecido) são utilizados para enriquecer inferências envolvendo análise de polaridade. Mais detalhadamente, é argumentado que cada cultura e idioma são únicos. Logo, explorando a singularidade de cada cultura e idioma é possível entender os efeitos e relações existentes entre os sentimentos e as emoções semânticas através de informações contidas em multimídias sociais. Este trabalho explora os conceitos propostos por [Borth et al. 2013], onde estruturas semânticas chamadas pares adjetivo-substantivo são aplicados em um *framework* de análise de polaridade. Entretanto, em [Borth et al. 2013], esses conceitos são aplicados em um contexto multilíngue, onde apenas análise semântica é realizada. Em [Jou et al. 2015], o elemento visual é adicionado, e é formalizado um *framework* dependente de idioma para a descoberta automática de construções adjetivo-substantivo.

O processo de construção do *framework* multilíngue de sentimentos visuais proposto começa com a mineração de imagens e meta-dados com base em palavras-chaves. *Tag's* de imagens (T1, ..., T5) são rotuladas com partes do discurso (análise sintática, por exemplo. artigo, numeral, verbo) e adjetivos e substantivos são usados para formar combinações de pares de candidatos substantivo-adjetivo (PSA), enquanto outros são ignorados. Finalmente, os candidatos PSA's são filtrados com base em vários critérios (incluindo características semânticas de cada linguagem e eliminação de sentimentos neutros), que ajudam a remover pares incorretos, formando um *framework* final com a diversidade e abrangência necessárias para sua generalização.

Localização geográfica através de imagens é um problema de visão computacional onde, através de uma imagem (em geral, imagens de monumentos ou estátuas populares) deve-se inferir a localização geográfica de onde esta imagem foi obtida. Problemas de iluminação, mal posicionamento da câmera, oclusão entre prédios, árvores e outros monumentos, tornam esse problema extremamente desafiador. Além disso, entender quais características em uma imagem devem ser exploradas para uma maior representação (quais características são mais importantes naquele monumento) e discriminação (quais características diferenciam aquele monumento dos outros monumentos) são problemas ainda em aberto em visão computacional. O *framework* proposto por [Jin Kim et al. 2015] é uma solução para o problema de reconhecer um lugar representado por uma imagem de consulta usando um grande banco de dados de imagens rotuladas. Essa abordagem faz uso de extração de características visuais e predição através de um banco de classificadores baseados em *Support Vector Machines* (Máquina de Vetores Suporte).

Para incentivar a comunidade científica a buscar soluções para esse problema, é proposto um *benchmark*, denominado MediaEval Multimedia Benchmark [Larson et al. 2015]. Nesse *benchmark*, é apresentada uma tarefa que exige que os participantes desenvolvam algoritmos que predizem automaticamente a localização geográfica de multimídia sociais. No trabalho de [Ballan et al. 2010], os desafios de se explorar multimídias sociais para interferência de localização geográfica são abordados. Em muitos casos, multimídias sociais tais como fotos e vídeos, possuem meta-informações (títulos, descrições, comentários) que, quando analisadas de forma adequada, podem indicar de maneira muito precisa a localização geográfica da imagem ou vídeo.

As multimídias sociais também estão associadas com etiquetas de contribuições

(*likes*, *dislikes*) dos usuários e comentários que podem ser explorados através de análise de sentimento. Muitas vezes, os itens de multimídia enviados por usuários estão conectados em uma rede social, neste caso, informações como conexões sociais e pontos de vista (opiniões) também podem estar disponíveis. No caso de vídeos, o sinal multimídia envolve padrões temporais que podem ser exploradas através de algoritmos de visão computacional. Mídias que envolvem vídeos possuem, em geral, um canal visual e um canal de áudio. Informação de áudio possui conteúdo falado e/ou ambiental, onde o contexto possui grande correlação com a localização geográfica. No trabalho de [Choi et al. 2012] é proposta uma solução para dado um banco de dados de imagens/vídeos de treinamento com suas respectivas coordenadas geográficas e dados textuais, estimar a localização geográfica de um vídeo (contendo meta-informações, tais como mensagens e pontos de vista).

3. Método Proposto

Nesta seção é descrito o método proposto, que contém mineração de dados em redes sociais, extração de características visuais a partir da foto do perfil do usuário e uma aplicação *Web* para o processamento e exibição de informações, relacionadas aos Pontos de Interesse de uma localidade.

3.1. Mineração de Dados e Características Visuais

O método proposto neste trabalho, consiste na coleta de dados de usuários e *check-ins* do Foursquare, através do Twitter. Posteriormente, através da API do Foursquare¹, são extraídas informações dos locais onde os *check-ins* foram realizados, além de informações do usuário que realizou o *check-in*. Neste sentido, como os dados relacionados a idade do usuário não podem ser obtidos da rede social, foram coletadas imagens do perfil destes usuários, para inferir a idade dos mesmos através de técnicas de aprendizagem de máquina.

O Twitter² é uma rede social para *microblogging*, na qual os usuários podem enviar e receber atualizações através de mensagens de até 140 caracteres, chamadas de *tweets*. Esta rede social, lançada em Julho de 2006, possui cerca de 320 milhões de usuários ativos mensalmente³. Atualmente, o Twitter possui duas APIs de coleta de dados, sendo elas:

- **Streaming API:** coleta de dados em tempo real, sem limite de *tweets* na coleta;
- **Search API:** coleta de dados históricos de até uma semana, com um limite de 2.700 *tweets* a cada 15 minutos.

Estas APIs, utilizam *queries* de busca para a coleta de dados. Para a solução proposta, a *query* utilizada foi "swarmapp.com", pela qual é possível obter *links* de *check-ins* do Foursquare.

O Foursquare⁴ é uma rede social baseada em localização, cujo principal serviço é o *check-in*, um evento que é sensoriado quando a hora e o local de um usuário são

¹<https://developer.foursquare.com/>

²<https://www.twitter.com>

³<https://about.twitter.com/pt/company>

⁴<https://www.foursquare.com>

coletados [Silva et al. 2014]. Os locais cadastrados no Foursquare são associados a uma das 10 categorias atualmente existentes na rede social, das quais podem ser observados alguns exemplos na Tabela 1.

Tabela 1. Exemplos de locais existentes nas categorias do Foursquare.

Categoria do Foursquare	Exemplos de locais
<i>Arts and Entertainment</i>	Cinemas, museus e casinos.
<i>College and University</i>	Escolas, laboratórios universitários e centros de estudo.
<i>Food</i>	Restaurantes, cafeterias e padarias.
<i>Nightlife Spot</i>	Bares, clubes de <i>rock</i> e boates.
<i>Outdoors and Recreation</i>	Parques, academias e praias.
<i>Professional and Other Places</i>	Fábricas, auditórios e centros médicos.
<i>Shop and Service</i>	Lojas, salões de beleza e supermercados.
<i>Travel and Transport</i>	Aeroportos, estradas e hotéis.

Através do Foursquare, também é possível extrair informações de um usuário e associá-lo ao local no qual o *check-in* foi realizado. Para a solução proposta, extraímos o sexo e a foto de perfil do usuário, para a inferência da idade do mesmo. Neste caso, aplicamos técnicas de processamento de imagens e de visão computacional, para que os padrões existentes nas imagens pudessem ser extraídos e analisados.

Como as imagens podem encontrar-se parcialmente corrompidas por ruídos, oclusão, má qualidade de câmeras e iluminação não adequada, o processamento de imagens é aplicado para realçar características visuais desejadas, tais como contornos, texturas, formas ou cores. Esse processo de aperfeiçoamento produz imagens mais adequadas para análise. Por exemplo, é possível remover o ruído, nitidez e brilho de uma imagem, tornando possível a identificação das principais características.

3.2. Arquitetura da Solução

A arquitetura proposta neste trabalho é composta por um servidor *web*, uma aplicação *web* e um gerenciador de banco de dados, como ilustrado na Figura 2. Desta forma, as requisições dos usuários são enviadas ao servidor para que este se comunique com as camadas da aplicação. Logo em seguida, a aplicação solicita dados do gerenciador de banco de dados. Este, por sua vez, retorna os dados solicitados para a aplicação, a fim de que a mesma direcione o resultado para o servidor. Por fim, o servidor envia uma resposta para o *browser* dos usuários.

3.2.1. Servidor *web*

O servidor *web* utilizado foi o **Apache HTTP Server** devido ao fato de seu código ser aberto, apresentar uma ótima performance e confiabilidade, possuir uma vasta documentação e uma comunidade extremamente ativa.

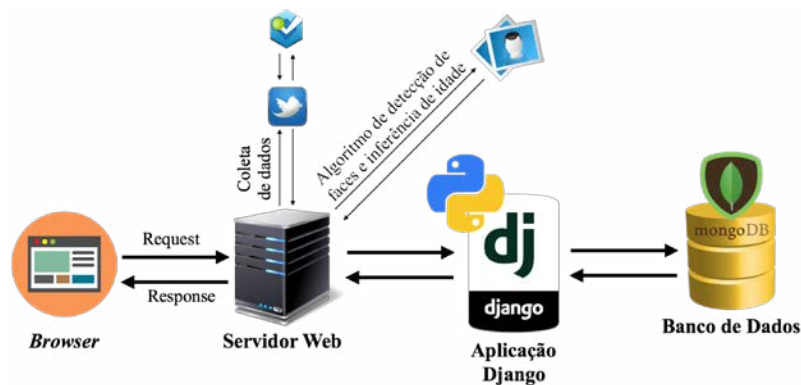


Figura 2. Visão geral da arquitetura da aplicação.

3.2.2. Aplicação web

Dentre as inúmeras opções de linguagens e ferramentas para desenvolvimento *web* no mercado, a aplicação proposta neste artigo utiliza um *web framework* de alto nível, escrito na linguagem de programação **Python**, conhecido como **Django**.

O *framework* Django proporciona um desenvolvimento *web* ágil, sem que os elementos de *design* e performance sejam negligenciados. Isto é possível, devido a inúmeras ferramentas e padrões de projetos que o compõe. As principais ferramentas do Django são [Santana and Galesi 2010]:

- **ORM - Object-Relational Mapper (Mapeador objeto-relacional)**: por meio desta ferramenta é possível abstrair a comunicação entre a camada de modelos de dados (classes de domínio) e o banco de dados. Ou seja, ela permite o mapeamento automático de modelos de dados escritos utilizando o paradigma Orientado a Objetos para o paradigma relacional;
- **URL Dispatcher (Despachador de URL)**: por meio desta ferramenta é possível gerenciar qual função será executada de acordo com a URL requisitada. Além disso, o URL Dispatcher provê mecanismos que facilitam a internacionalização do site e permite apresentar URLs simples e amigáveis ao usuário;
- **Sistema de templates**: fornece uma linguagem própria para geração de *templates*. No contexto de aplicações desenvolvidas em Django, um *template* consiste em uma *string* que, quando combinada com elementos de páginas *web* (*tags* HTML, por exemplo), gera uma saída consistente. Deste modo, é possível garantir que blocos de código e de representação não se misturem;
- **Sistema de administração**: consiste em uma interface de administração que possibilita gerenciar os objetos das classes de domínio do *site*. Desta forma, não é necessário implementar operações **CRUD** (*Create, Read, Update, Delete*) para cada uma das classes do domínio.

A aplicação *web* proposta neste trabalho apresenta como uma de suas principais características o fato de ser responsiva. Isto é, o *layout* da aplicação é capaz de se adaptar a diversos tipos de dispositivos, como pode ser observado na Figura 3. Para desenvolver o caráter responsivo, foram utilizadas, em conjunto com o Django, as seguintes tecnologias: **HTML5** e **Bootstrap**.

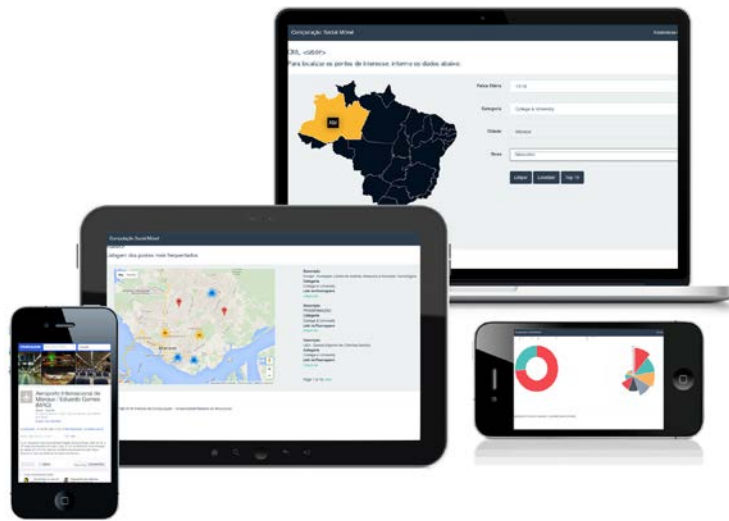


Figura 3. Dispositivos que podem utilizar a aplicação devido a sua responsividade.

O HTML (*Hypertext Language Markup*) é uma linguagem utilizada para publicação de conteúdo em navegadores *web*. Desde 2004, a sua quinta especificação (HTML5) vem sendo desenvolvida. Como novidades, o HTML5 apresenta: novas *tags* (delimitadores de conteúdo) que auxiliam no processo de divisão da página (cabeçalhos, seções, rodápes, etc) de forma intuitiva, um recurso chamado *canvas* que possibilita a renderização de desenhos em tempo de execução, uma API que permite a visualização e manipulação de mapas, entre outros recursos [Poyatos 2014]. Já o Bootstrap é um *framework front-end* amplamente utilizado pela indústria. Em sua estrutura interna, o *framework* possui arquivos CSS (*Cascading Style Sheets*) e Javascript que, quando combinados, permitem construir interfaces com poderosos recursos visuais [Rodrigues 2015].

3.2.3. MongoDB

Devido ao fato de bancos de dados relacionais apresentarem problemas de escalabilidade à medida que o volume de dados armazenados aumenta rapidamente, um novo padrão de banco de dados foi desenvolvido com o objetivo de representar dados sem utilizar inúmeras restrições entre eles [Simões 2014]. Este novo padrão é conhecido como **noSQL** (*Not Only SQL*) [Medeiros 2014].

O noSQL utiliza como modelo de dados o **Modelo Relacional Não Normalizado** (MRNN). Desta forma, os dados são persistidos em estruturas desnormalizadas, o que possibilita recuperar dados sem a utilização de operações de junções (JOINS), uma vez que as estruturas de armazenamento já se encontram aglutinadas. Dentre os principais sistemas gerenciadores de banco de dados que utilizam noSQL está o **MongoDB** [Membrey et al. 2011].

O MongoDB é um Sistema Gerenciador de Banco de Dados (SGBD) baseado em documentos, ou seja, os dados são armazenados em documentos ao invés de serem armazenados em tabelas. Estes documentos são estruturados utilizando uma forma binária do JSON (*Javascript Object Notation*) conhecida como BSON (*Binary Json*). Isto

ocorre principalmente porque o BSON suporta mais formatos de dados e a performance de indexação destes dados é muito maior [Foundation 2016, Banker 2011].

A aplicação proposta nesse artigo opera com um grande volume de dados extraídos de redes sociais e, por ser destinada a *web*, necessita de um alto grau de escalabilidade. Além disso, é necessário que as estruturas que armazenam os dados sejam extremamente flexíveis, como por exemplo, o tempo de processamento gasto para incluir mais um campo na estrutura deve ser mínimo. A partir das necessidades, foi escolhido o MongoDB como Sistema Gerenciador de Banco de Dados.

4. Experimentos e Resultados

Nesta seção, apresentamos os experimentos e resultados relacionados a detecção de pontos de interesse e a inferência de idade dos usuários, utilizada para auxiliar na aquisição de informações necessárias para a aplicação proposta.

4.1. Detecção de Pontos de Interesse

Os dados da aplicação são provenientes do Foursquare e foram coletados entre 14/12/2015 e 29/12/2015. A coleta dos dados foi realizada através do Twitter, utilizando a *query* de busca "swarmapp.com". Para este período foram coletados 333.286 *tweets* provenientes de capitais brasileiras, contendo URLs do Swarm⁵, aplicação responsável por registrar os *check-ins* do Foursquare.

Após a coleta dos dados, foi realizada a limpeza dos mesmos. Esta etapa se fez necessária devido ao fato de que alguns dados possuíam anomalias, o que invalidava a sua utilização pelos modelos presentes na solução deste trabalho. Dentre as anomalias, podemos destacar: *check-ins* cujas URLs foram coletadas, mas durante a extração de dados viu-se que os mesmos foram excluídos, e *check-ins* cujas contas dos usuários foram excluídas ou tornaram-se privadas.

Com a limpeza e a extração dos dados, verificamos que o Rio de Janeiro era a capital brasileira com a maior quantidade de *check-ins*, seguida por Porto Alegre e São Paulo (Figura 4). Além disso, verificamos que as 5 categorias do Foursquare com mais *check-ins* nesse período eram *Shop and Service*, *Food*, *Outdoors & Recreation*, *Professional & Other Places* e *Residence*, nesta ordem. Isto pode ser justificado pelo fato de que os dados são referentes a uma época festiva, o Natal.

Através dos *check-ins* também é possível recuperar informações a respeito do usuário que realizou o mesmo. Assim, verificamos que, neste período, cerca de 75% dos *check-ins* foram realizados por homens. Também coletamos as fotos dos perfis dos usuários a fim de que fossem utilizadas para a inferência da idade dos mesmos, conforme é descrito na subseção 4.2. Após a inferência de idade, tornou-se possível listar os locais mais frequentados de acordo com a faixa etária dos usuários.

Por fim, verificamos que cerca de 50% dos *check-ins* ocorreram no horário comercial (08:00 às 17:00), enquanto 42% ocorreram durante a noite e a madrugada.

Assumindo que os locais mais frequentados de uma região são os locais com o maior número de *check-ins*, identificamos através da aplicação PoI os locais mais frequentados nas capitais brasileiras e suas adjacências, de acordo com cada categoria do

⁵<https://www.swarmapp.com/>

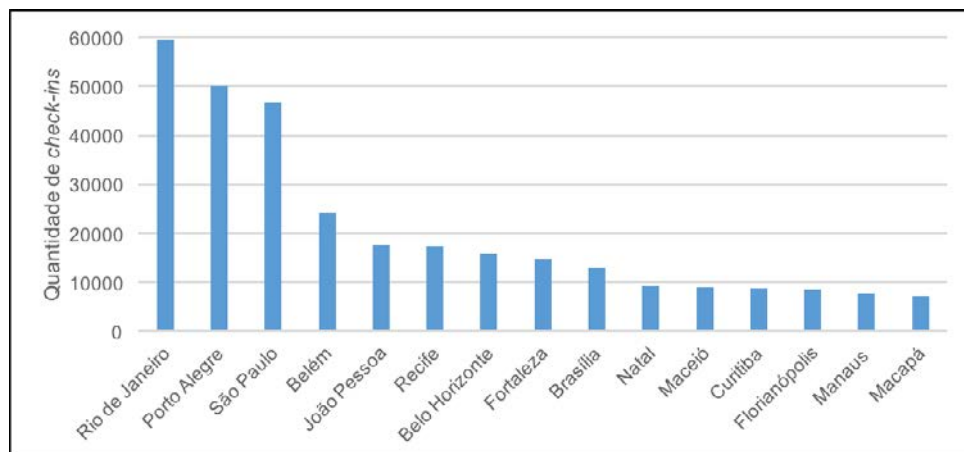


Figura 4. Quantidade de *check-ins* do Foursquare em 15 capitais brasileiras, extraídos no período de 14/12/2015 a 29/12/2015.

Foursquare. Como exemplo, apresentamos a Figura 5, na qual é possível identificar Pontos de Interesse em diferentes categorias na cidade de São Paulo e outras regiões próximas, nos dias 20/12/2015 e 21/12/2015. Para este exemplo, apresentamos a Tabela 2, na qual é possível verificar os 3 locais mais frequentados em 4 categorias, ou seja, os 3 principais Pontos de Interesse destas categorias.

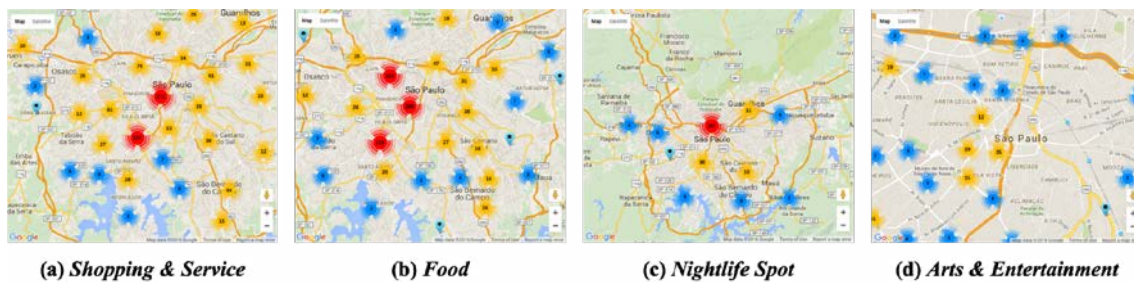


Figura 5. Pontos de Interesse em diferentes categorias na cidade de São Paulo e regiões próximas, nos dias 20/12/2015 e 21/12/2015.

Além das categorias do Foursquare, também é possível determinar Pontos de Interesse utilizando os dados dos usuários. Um exemplo disso, seria verificar os locais mais frequentados por homens ou mulheres em uma região.

Com os exemplos apresentados nesta subseção, verificamos que a aplicação PoI, através das redes sociais, pode prover informações necessárias para diversos fins, dentre os quais podemos destacar:

- Verificação de pontos de interesse de acordo com parâmetros previamente estabelecidos (categoria do local, idade e sexo do usuário, etc.). Isto pode auxiliar em campanhas publicitárias que queiram atingir um determinado público-alvo, por exemplo;
- Listagem dos locais mais frequentados em um região para o auxílio de turistas.

4.2. Inferência de Idade dos Usuários

Estimar a idade de pessoas através de imagens é um problema desafiador, pois envolve fatores que, em geral, geram estruturas não-lineares que são difíceis de serem modeladas

Tabela 2. Lista dos 3 locais mais frequentados nos dias 20/12/2015 e 21/12/2015, divididos por categoria referentes à cidade de São Paulo.

Categoria do Foursquare	Quantidade de <i>check-ins</i>	Lugares mais frequentados
<i>Shop & Service</i>	1331	Shopping Metrô Boulevard Tatuapé, Linda Unha, Havaianas.
<i>Food</i>	1249	Cris Doces e Salgados, Central de Massas, Lanches Escadão.
<i>Nightlife Spot</i>	386	Ilhabela Irish Pub, Barcearia, Coconut Brasil.
<i>Arts & Entertainment</i>	363	Estúdio Iquiririm, Cinemark, Cinépolis.

e classificadas por algoritmos de aprendizagem de máquina modernos. Na estimativa de idade através de imagens, utilizam-se técnicas de extração de características que são necessárias para discriminar e representar o processo de envelhecimento e aparência das faces em imagens.

A tarefa de estimativa de idade por imagens herda as características dos problemas de reconhecimento de objetos através de imagens, tais como o processo de detecção da face, a localização das características faciais, a formulação vetor de características e classificação. Dependendo da aplicação, o resultado da classificação pode ser uma idade exata ou uma faixa etária.

Outro fator importante, que pertence ao problema estimativa da idade, é o intervalo de idades consideradas. Este parâmetro, é um aspecto importante do problema, com diferentes características de envelhecimento que aparecem em diferentes grupos etários. Portanto, um modelo treinado para lidar com uma determinada faixa etária pode não ser aplicável a outras faixas etárias. Em redes sociais, essas imagens encontram-se em ambiente ainda mais desafiador, onde oclusão, mal posicionamento da câmera e iluminação podem tornar a extração de características das imagens difícil.

Para resolver os desafios listados acima, desenvolvemos um método baseado em classificação de conjuntos de imagens. Nesse método, a estimativa de idade de cada pessoa é obtida por meio de várias instâncias de padrões de um mesmo indivíduo, ao invés de apenas uma instância.

Com o rápido desenvolvimento de tecnologias de processamento e armazenamento de vídeos, sequências de imagens de um mesmo indivíduo são facilmente obtidas em redes sociais. Conseqüentemente, a tarefa de estimativa de idade pode ser aprimorada se for utilizado um conjunto de imagens para a extração de características. Por exemplo, cada amostra é um conjunto de imagens faciais com diferentes poses, iluminações e expressões. Utilizando um conjunto de imagens, a estimativa de idade pode alcançar maior precisão, pois conjuntos de imagens geralmente contêm mais informações do que uma única imagem. Ainda que esse conjunto de imagens obtido das redes sociais de um mesmo indivíduo tenha baixa qualidade, pode-se explorar a relação temporal entre essas

imagens, aprimorando a precisão do classificador.

Considerando que um conjunto de imagens de um mesmo indivíduo foi coletada, têm-se os seguintes passos para estimar a idade deste; 1) Detecção da face do indivíduo na imagem, 2) Extração de características da face, tais como região dos olhos, nariz e lábios, 3) Seleção de características, 4) Classificação do conjunto de imagens.

Para detecção de faces, utilizamos o método desenvolvido por [Viola and Jones 2004], que extrai características baseados no Haar-like e utiliza o classificador Ada-boost em cascata. Para a extração de características utilizou-se os seguintes métodos: Histograma de Gradientes Orientados (HOG), LBP e Gabor. Estes métodos possuem características complementares, tais como textura, cor e forma. A seleção de características é feita por meio da Análise dos Componentes Principais [Jolliffe 2002] e um classificador baseado nos vizinhos mais próximos [Cunningham and Delany 2007], que determina a faixa etária a qual pertence o conjunto de imagens dado.

O principal problema em se criar uma base de dados para a tarefa de estimar a idade através de imagens, é a ausência do ground-truth (informação sobre a idade real das pessoas nas imagens). No entanto, é possível inferir visualmente um rótulo para cada imagem manualmente. Para tanto, é necessário que um grupo de pessoas determinem a idade aproximada de cada imagem da base de dados. Nos experimentos apresentados a seguir, fizemos uso de 10 pessoas que avaliaram cada uma das imagens da base de dados. O valor final de cada rótulo foi obtido por meio da moda do conjunto de avaliações de cada imagem (Tabela 3).

Tabela 3. Base de dados obtida através de imagens de redes sociais.

Faixa etária	Moda	Variância	Homens	Mulheres	Total
12-17	14	2.9	8	11	19
18-23	20	2.7	11	29	40
24-28	25	1.8	9	12	21
29-35	32	3.1	7	10	17
36-41	38	2.5	7	6	13
42+	43	3.4	3	7	10

A base de dados confeccionada, é composta por 120 conjuntos de fotos de perfis de usuários. Cada conjunto possui entre 7 e 12 imagens, que foram extraídas de redes sociais, tais como o Foursquare. As imagens variam em relação ao fundo (*background*), resolução, orientação e iluminação. Para efeitos de simplificação, consideramos que todas as imagens em um mesmo conjunto são pertencentes a um único indivíduo, e que foram obtidas em um intervalo de tempo razoavelmente próximo. Quando há mais do que uma face na imagem, supomos que a maior imagem pertence ao proprietário do perfil.

O objetivo desse experimento, é determinar se é possível utilizar algoritmos de aprendizagem supervisionada para estimar a idade das pessoas em redes sociais. A fim de

normalizar as imagens, utilizamos o algoritmo proposto por [Viola and Jones 2004] para retirar apenas as imagens das faces de cada foto, sendo que todas as imagens de faces foram redimensionadas para 40x40 pixels. Em seguida utilizamos os seguintes descritores de características: Histograma de Gradientes Orientados (HOG), LBP e Gabor. A dimensão de cada vetor de características foi reduzida, utilizando-se a Análise dos Componentes Principais. Os parâmetros empregados encontram-se na Tabela 4.

Tabela 4. Principais parâmetros utilizados no experimento.

	Dimensão do vetor D1	Dimensão do vetor D2
HOG (Vh)	900	37
LBP (Vl)	1024	43
Gabor (Vg)	1092	51

Para a execução dos testes, a Análise dos Componentes Principais foi fixada para extrair 95% da energia acumulada em cada um dos vetores de características. Para classificação, fez-se uso do vetor concatenado $V_f = V_h || V_l || V_g$, onde $A = B || C$ significa que A é formado pela concatenação de B com C. O esquema de validação empregado, é baseado no *leave-one-out*, onde um conjunto é separado para teste e o restante é utilizado para treino.

A matriz de confusão apresentada na Tabela 5, contém as informações sobre a taxa de classificação dos conjuntos em relação a quantidade de classes. A acurácia total do método pode ser definida então por 79.05% ($(73.68\% + 87.5\% + 85.71\% + 70.58\% + 76.92\% + 80\%)/6$). Levando em consideração que as imagens utilizadas para formar a base de dados foram obtidas em ambiente não controlado, o classificador possui uma boa acurácia. Além disso, a base de dados empregada no experimento não está balanceada em relação ao número de conjuntos por classe, por esse motivo tem-se a tendência de se classificar um conjunto na classe que possui o maior número de conjuntos.

Tabela 5. Matriz de confusão.

		Classes de Teste					
		12-17	18-23	24-28	29-35	36-41	42+
Classes de Treino	12-17	73.68	-	-	-	7.69	-
	18-23	15.78	87.5	14.28	17.64	7.69	10
	24-28	5.26	7.5	85.71	5.88	7.69	10
	29-35	5.26	2.5	-	70.58	-	-
	36-41	-	-	-	-	76.92	-
	42+	-	2.5	-	5.88	-	80

5. Conclusão e Trabalhos Futuros

Neste trabalho, apresentamos uma aplicação *web*, cujo objetivo é auxiliar empresários, turistas e órgãos governamentais na análise de um perímetro urbano. Desta forma, os mesmos podem identificar locais propícios para o turismo, a realização de campanhas e propagandas, e a divulgação de produtos.

Como exemplo, podemos citar as campanhas de saúde, que buscam atingir um público-alvo específico de acordo com os locais frequentados e a faixa etária. Com a aplicação PoI, o tipo de local pode ser identificado através das categorias do Foursquare, e a idade pode ser inferida pela foto dos perfis dos usuários que frequentam os Pontos de Interesse.

Dentre os desafios deste trabalho, podemos destacar a proveniência de dados de redes sociais, pois algumas cidades estudadas possuíam menos de 1.000 *check-ins* no período de uma semana. Além disso, as imagens obtidas para a inferência de idade foram obtidas de um ambiente não controlado.

Como trabalho futuro, planejamos aprimorar o método de inferência da idade dos usuários, por meio de redes sociais. Para isto, pretendemos não só utilizar apenas as imagens, mas também outros dados que podem ser coletados de perfis, como por exemplo, o sexo, a forma de escrita (culto ou coloquial), a rotina, os locais mais frequentados e os círculos de amizade.

6. Referências

- Ballan, L., Bertini, M., Del Bimbo, A., Meoni, M., and Serra, G. (2010). Tag suggestion and localization in user-generated videos based on social knowledge. In *Proceedings of second ACM SIGMM workshop on Social media*, pages 3–8. ACM.
- Banker, K. (2011). *MongoDB in action*. Manning Publications Co.
- Borth, D., Ji, R., Chen, T., Breuel, T., and Chang, S.-F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM.
- Choi, J., Friedland, G., Ekambaram, V., and Ramchandran, K. (2012). Multimodal location estimation of consumer media: Dealing with sparse training data. In *Multimedia and Expo (ICME), 2012 IEEE International Conference on*, pages 43–48. IEEE.
- Cunningham, P. and Delany, S. J. (2007). k-nearest neighbour classifiers. *Multiple Classifier Systems*, pages 1–17.
- Frias-Martinez, V., Soto, V., Hohwald, H., and Frias-Martinez, E. (2012). Characterizing urban landscapes using geolocated tweets. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 239–248. IEEE.
- Foundation, D. S. (2016). Django documentation. url = <https://docs.djangoproject.com/en/1.9/>. [Acessado em Janeiro/2016].
- Jin Kim, H., Dunn, E., and Frahm, J.-M. (2015). Predicting good features for image geo-localization using per-bundle vlad. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1170–1178.

- Jolliffe, I. (2002). *Principal component analysis*. Wiley Online Library.
- Jou, B., Chen, T., Pappas, N., Redi, M., Topkara, M., and Chang, S.-F. (2015). Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, pages 159–168. ACM.
- Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., and Campbell, A. T. (2010). A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150.
- Larson, M., Kelm, P., Rae, A., Hauff, C., Thomee, B., Trevisiol, M., Choi, J., Van Laere, O., Schockaert, S., Jones, G. J., et al. (2015). The benchmark as a research catalyst: Charting the progress of geo-prediction for social multimedia. In *Multimodal Location Estimation of Videos and Images*, pages 5–40. Springer.
- Lee, R. and Sumiya, K. (2010). Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks*, pages 1–10. ACM.
- Medeiros, H. (2014). Introdução ao mongodb. url = <http://www.devmedia.com.br/introducao-ao-mongodb/30792/#ixzz3yRHHTBfC>. [Acessado em Janeiro/2016].
- Membrey, P., Plugge, E., and Hawkins, D. (2011). *The definitive guide to MongoDB: the noSQL database for cloud and desktop computing*. Apress.
- Poyatos, H. (2014). Programando em html5. url = <http://www.devmedia.com.br/programando-em-html5/31040/#ixzz3ySFtOR3r>. [Acessado em Janeiro/2016].
- Rodrigues, J. (2015). Como criar uma aplicação responsiva com bootstrap. url = <http://www.devmedia.com.br/como-criar-uma-aplicacao-responsiva-com-bootstrap/33185>. [Acessado em Janeiro/2016].
- Santana, O. and Galesi, T. (2010). Python e django: desenvolvimento ágil de aplicações web. *São Paulo: Novatec*.
- Silva, T. H., de Melo, P. O. V., Viana, A. C., Almeida, J. M., Salles, J., and Loureiro, A. A. (2013a). Traffic condition is more than colored lines on a map: Characterization of waze alerts. In *Social Informatics*, pages 309–318. Springer.
- Silva, T. H., Vaz de Melo, P., Almeida, J. M., and Loureiro, A. A. (2013b). Uma fotografia do instagram: Caracterização e aplicação. *XXXII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*.
- Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J., and Loureiro, A. A. (2014). Revealing the city that we cannot see. *ACM Transactions on Internet Technology (TOIT)*, 14(4):26.
- Simões, A. (2014). Comparando o nosql ao modelo relacional. url = <http://www.devmedia.com.br/comparando-o-nosql-ao-modelo-relacional/30917/#ixzz3yRM5amNk>. [Acessado em Janeiro/2016].
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.

Análise e Design de Ecossistemas de Negócios usando a Abordagem *Blueprint* de Valor: Um Estudo Empírico

Luciana A. Almeida¹, Cleidson De Souza^{1,2}

¹Programa de Pós-Graduação em Ciência da Computação – UFPA – Belém, PA – Brasil

²Instituto Tecnológico Vale – Belém, PA – Brasil

lu.abdon.si@gmail.com

cleidson.desouza@acm.org

Abstract. *A startup is a company created in the context of extreme uncertainty and risk. Despite this, a rich set of startups can bring economic benefits to the geographic region where they are located. Thus, there is strong interest aimed at assisting in the creation and success of these startups. However, an important aspect of these startups, the business ecosystem in which they operate and the risks management associated with this ecosystem, is often overlooked. One exception is the value blueprint approach. Thus, this work describes an empirical study aimed at the application and analysis of the value blueprint in a qualitative study with startups.*

Resumo. *Uma startup é uma empresa em fase inicial criada em condições de extrema incerteza e riscos. Um rico conjunto de startups pode trazer benefícios econômicos para a região geográfica onde elas estão localizadas. Assim, existem diversos estudos visando auxiliar na criação e no sucesso destas startups. Entretanto, estas pesquisas não enfatizam um aspecto importante destas startups: o ecossistema de negócios onde elas estão inseridas e a gestão dos riscos associados a este ecossistema. Uma exceção são as pesquisas que utilizam a abordagem conceitual blueprint de valor. Assim, este trabalho descreve um estudo empírico com o objetivo de analisar como a blueprint de valor pode auxiliar um conjunto de startups.*

1. Introdução

O crescimento acelerado de acesso a Internet entre 1995 e 2000, durante a primeira bolha da Internet, apoiado pela convergência digital e o aumento da largura de banda (Leon 2009), possibilitaram o desenvolvimento de modelos de negócios para oferecer produtos/serviços diferenciados através de um meio digital. Nesse período surge, no mercado de tecnologia, o termo *startup*, definido como “*uma organização formada para encontrar um modelo de negócio repetível, escalável e rentável*” (Ries, 2012). Ou seja, empresas recém-criadas que visam transformar ideias em produtos. Estas empresas estão inseridas em um contexto de riscos e incertezas e contam com poucos recursos financeiros. Isto quer dizer que as principais técnicas e ferramentas de planejamento elaboradas durante o século XX já não acompanham as necessidades de negócio atuais.

De acordo com uma pesquisa de Shikhar Ghosh (Gage 2012) existem evidências de que três quartos das empresas apoiadas por capital de risco não retornam o capital dos investidores nos EUA. O *National Venture Capital* estima que 25% a 30% dessas empresas falham e mais de 95% não vêem o retorno planejado sobre o investimento

(Gage 2012). Assim, profissionais de várias áreas – administração, TI, etc – buscam desenvolver ferramentas, metodologias e abordagens que se adequem ao desenvolvimento e gestão de produtos. Alguns exemplos são: *design thinking* (Mueller e Thoring 2012); *customer development* (Blank 2012); *Business Model Canvas* (BMC) (Osterwalder e Pigneur 2010) e *lean startup* (Ries 2012).

Para complicar ainda mais este quadro, as *startups* emergem em um cenário de negócios diferenciado: as cadeias tradicionais de valor já não representam a realidade das relações entre as partes do negócio. O termo que melhor representa este cenário é denominado de ecossistemas de negócios. Moore (1993) define ecossistemas de negócios como “*uma estrutura dinâmica de organizações interligadas que dependem umas das outras para a sobrevivência mútua.*” Isto quer dizer que muitas organizações de software já não funcionam como unidades independentes que podem oferecer produtos isolados. Ao contrário, existe uma grande dependência entre produtos e empresas, isto é, as empresas precisam colaborar para que todas tenham sucesso (Taveira 2012).

Adner (2012) argumenta que “*com frequência cada vez maior, o sucesso depende não só de sua capacidade de cumprir promessas, mas também do quanto uma série de parceiros – alguns visíveis, outros não – conseguem cumprir o que prometem.*” Isto significa que, grande parte dos gestores focam em gerenciar os riscos de execução de seus projetos. No entanto, estes gestores não gerenciam as dependências que existem em um projeto, isto é, as dependências entre as diferentes empresas de um ecossistema de negócios (Adner 2012). Visando endereçar este problema, Adner (2012) propõe a abordagem *blueprint de valor* para auxiliar a identificação desses riscos. Esta também auxilia na visualização do ecossistema e de todos os elementos envolvidos no mesmo. Em outras palavras, o *blueprint* de valor é uma abordagem que torna explícito o ecossistema e suas dependências.

Assim, este trabalho descreve um estudo empírico com o objetivo de avaliar a *blueprint de valor*. Este estudo foi conduzido com um conjunto de *startups* de tecnologia da informação incubadas em uma agência de tecnologia e inovação. Utilizando uma abordagem de pesquisa qualitativa alguns integrantes de *startups* foram entrevistados e os relatos da utilização da *blueprint de valor* foram analisados. Os resultados descrevem as vantagens, desvantagens e sugestões de melhoria da *blueprint* de valor.

O restante deste artigo está organizado da seguinte forma. A Seção 2, a seguir, descreve os principais conceitos necessários para o entendimento deste trabalho incluindo, ecossistemas de negócios e riscos de negócios. A Seção 3, apresenta a análise e o *design* de ecossistemas utilizando a *blueprint* de valor. A Seção 4, mostra a metodologia utilizada nesta pesquisa. Depois disso, a próxima Seção explica os resultados do trabalho. Então, a Seção 6, apresenta uma discussão sobre esses resultados. Finalmente, a Seção 7, traz as considerações e trabalhos futuros.

2. Ecossistemas e Riscos de Negócios

2.1. Ecossistemas de Negócios

O termo Ecossistemas de Negócios é inspirado no conceito de ecossistemas biológicos. O conceito de ecossistema foi formulado para especificar o comportamento cooperativo

dos organismos da natureza (Tansley 1935). A diferença entre um sistema e um ecossistema, é que um ecossistema se desenvolve gradualmente até atingir um equilíbrio dinâmico. Um ecossistema evoluído é, portanto, relativamente autônomo (Ikenami e Salerno 2015). Entretanto, esta estabilidade pode ser abalada por desequilíbrios internos ou turbulências externas causadas inclusive por outros ecossistemas. Alguns desses ecossistemas podem sequer conseguir se organizarem de forma estável e morrerem ainda incipientes (Tansley 1935).

Nesse contexto, todos os membros de um ecossistema de negócios – clientes, organizações, sociedade, ambiente e agentes reguladores – convivem de maneira coordenada ao longo do tempo. Valor é criado por meio de processos complexos e dinâmicos, que envolvem diversas relações entre os atores do ecossistema, nos quais todos devam capturar valor (Jesus 2014). Para ser mais específico, os atores interagem competindo e colaborando, principalmente para desenvolver suas próprias capacidades e, potencialmente, alcançar a governança do ecossistema (Jesus 2014).

Em vários setores da economia, uma importante mudança operada por organizações vem acontecendo. Gradualmente, estas tem deixado de usar as cadeias de suprimentos e passado a adotar parcerias e esquemas de colaboração para oferecer “soluções” melhor elaboradas (Adner 2012).

2.2. Riscos da Co-inovação e da Cadeia de Adoção

Muitos projetos concentram suas forças exclusivamente em nível de execução, por que gestores competentes sabem que alcançar o sucesso requer foco obsessivo nas capacidades internas da organização. Entretanto, muitas vezes estes gestores são surpreendidos pelos “pontos cegos”, isto é, pelos riscos da co-inovação e/ou da cadeia de adoção (Adner, 2012). Os *riscos da co-inovação* consistem das tecnologias desenvolvidas externamente à empresa e que devem existir para que o produto da empresa tenha sucesso. Estes riscos independem da natureza da inovação. Ou seja, o risco pode ser tecnológico (ex: *uma bateria melhor para carros elétricos*), processual (*um novo processo que garanta qualidade*), organizacional (*uma abordagem de vendas integrada que combine ofertas de várias divisões*), entre outras opções. Já os *riscos da cadeia de adoção* referem-se aos intermediários que se encontram no caminho entre um produto e o cliente final do mesmo. Isto inclui a necessidade dos parceiros em adotar a proposta de valor, fazendo com que os clientes finais possam ter a oportunidade de conhecer o produto. Por exemplo, um lojista precisa vender o produto em suas lojas (físicas ou virtuais) para que um cliente tenha a opção de comprá-lo. Enfrentar riscos como o da cadeia de adoção pode ser decisivo no sucesso ou fracasso de novos, pois quando as cadeias de adoção são frágeis os resultados são incertos.

Os riscos de co-inovação e cadeia de adoção demonstram a preocupação com as dependências de um negócio, ou seja, os riscos fundamentais a serem considerados em estratégias de ecossistemas (Adner, 2012). Essas estratégias se baseiam não apenas na cooperação entre pessoas, mas também na cooperação entre empresas dos mais variados setores. Por exemplo, para uma indústria de automóveis, não basta produzir um carro mais confiável, veloz e eficiente: é preciso também oferecer sistemas de entretenimento e computação de bordo de última geração. Empresas de marketing devem oferecer companhias de publicidade e organizar comunidades de usuários; as telecomunicações precisam não só prover o serviço de voz, mas toda experiência de mídia, e assim por

diante. Para ajudar a configurar este cenário Adner propõe uma abordagem, chamada *blueprint* de valor, que será discutida na próxima Seção.

3. Análise e *Design* de Ecossistemas usando a *Blueprint* de Valor

Para o processo da gestão de riscos tanto no ambiente interno da empresa quanto externo, é preciso primeiro identificar os relacionamentos atuais da organização, ou fora dela, que precisam mudar para que o produto tenha sucesso. Apenas depois desta identificação será possível iniciar a definição de estratégias para gerenciar os riscos entre as dependências (Adner 2012). Nesse sentido, a *blueprint* de valor, tem a finalidade de ajudar a identificar estes riscos e desenhar o ecossistema de negócios correspondente, o que Adner (2012) denomina de Mínimo Ecossistema Viável (MVE). O MVE é a menor configuração de elementos que podem ser reunidos em um ecossistema e ainda criar valor comercial. A *blueprint* de valor pode ser utilizada em empresas em contextos diversos. Esta pesquisa analisa sua utilização em *startups*.

O *blueprint* mostra os componentes de um ecossistema, não se concentrando apenas na identificação de componentes tradicionais de negócios (Adner 2012). Para identificar estes componentes 5 aspectos devem ser identificados: identifique o cliente final; identifique seu próprio produto; identifique seus fornecedores; identifique seus intermediários (Quem está entre os mentores do produto e o cliente final? Quem suprirá o produto após disponibilizado?; e por quem o produto deve passar para chegar ao cliente final? Além da identificação desses componentes, o *blueprint* de valor informa aos gestores a situação atual do ecossistema por meio de uma metáfora simples, com base no sinal de trânsito contínuo representado por luzes verde-amarelo-vermelho, para indicar o alinhamento de cada componente. Isto é, para os riscos de co-inovação, *verde* significa que as partes associadas estão prontas; *amarelo* significa que elas ainda não estão prontas, mas que existe um plano (que pode demorar um pouco, mas que será conseguido eventualmente); e *vermelho* que significa que essas partes não estão prontas e que não existe nenhum plano claro definido para elas. Para os riscos de adoção, *verde* significa que seus parceiros estão ansiosos para participar; *amarelo* significa que eles estão neutros, mas abertos a entrar no ecossistema; e *vermelho* significa que eles têm razões claras para manter a sua condição anterior e assim *não* participar do ecossistema tal como ele está desenhado.

A utilização do *blueprint* de valor permite comparar estratégias diferentes. Por exemplo, a seguir, o *blueprint* de valor é usado para explicar por que a *Amazon* e a *Sony* tiveram resultados tão diferentes no desenvolvimento do mercado de *e-readers*. A Figura 1 apresenta o *blueprint* de valor do *Sony Reader* no seu lançamento (Adner 2012). Os elementos representam: (i) E-Ink fornecedora da tela; (ii) *Sony* DRM para gerenciar o conteúdo entre outros componentes; (iii) o leitor digital da *Sony*; (iv) intermediários, os varejistas; e (v) clientes finais o grande público leitor. Além disso, um leitor digital precisa disponibilizar algo para ser lido, completando a proposta de valor aos clientes. Neste ponto entram os complementadores desse ecossistema. Ou seja, a *Sony* precisava convencer: (vi) os autores, editoras e seu próprio varejista, o *Sony Connect.com* a comprar a ideia. Porém as editoras eram problemáticas tanto em termos de co-inovação quanto nas frentes de adoção, pois aspectos econômicos e jurídicos desta nova tecnologia deveriam ser esclarecidos principalmente quanto aos direitos autorais (Adner 2012). A falta de conteúdo adequado significou que este leitor

não ia avançar no mercado. E por fim (vii) a conectividade, na qual o usuário precisava conectar o leitor ao PC para transferir o livro comprado para o leitor.

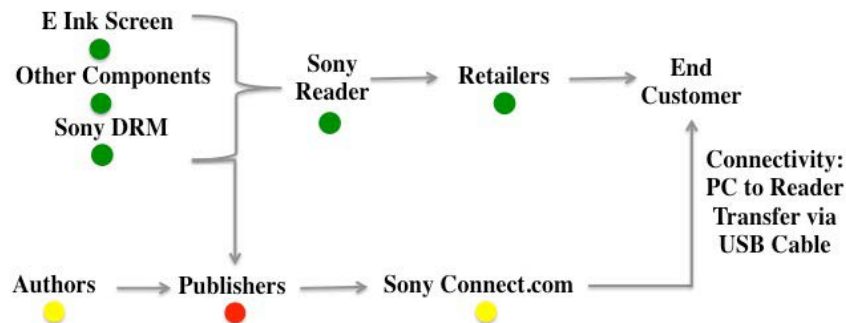


Figura 1. Exemplo de um *blueprint* de valor (Sony Reader).
Adaptado de: Adner, R. (2012).

Este exemplo ilustra como o *blueprint* de valor permite a identificação de incertezas no caminho do conteúdo até o dispositivo, independente da excelência do hardware. Estas incertezas no ecossistema sugerem a um novo conceito utilizado no contexto da *blueprint* de valor: a reconfiguração do ecossistema. A reconfiguração de um ecossistema consiste na alteração das interações entre os elementos no sistema, ou seja, em atuar para mudar o risco dessas relações. O *blueprint* de valor é usado como ponto de partida tal que, a partir de qualquer elemento do mesmo, é possível alterar o arranjo de atividades, atores e links do mesmo. Isto é feito fazendo cinco perguntas fundamentais: O que pode ser separado? O que pode ser combinado? O que pode ser mudado? O que pode ser adicionado? O que pode ser retirado? (Adner 2012). A partir destas perguntas é possível pensar em uma reorganização dos elementos do ecossistema presentes na *blueprint* de valor, e assim, mitigar os riscos do ecossistema (fazendo com que as cores se tornem “verdes”) para o lançamento de um produto. A possibilidade de reconfigurar um ecossistema ajuda a refletir sobre os pontos necessários para a criação de um ecossistema de valor para um produto.

A Figura 2 apresenta o exemplo de *blueprint* de valor para o lançamento do *Kindle* da *Amazon*. Os elementos representam: (i) complementadores iniciais, como a conectividade sem fio no dispositivo desenvolvido, o que permitiria a utilização da rede da *Amazon* para baixar o livro instantaneamente, (ii) o uso de uma arquitetura fechada e proprietária para prevenir que os livros sejam copiados, impressos, ou colocados disponíveis de forma ilegal e que por sua vez reduziriam os riscos de adoção pelos editores e autores (com luzes amarelas). Além disso, (iii) o *Kindle* oferecia inicialmente 90 mil títulos em sua biblioteca, além de oferecer aos seus clientes (iv) *Amazon.com* preços de livros com valores bem reduzidos em comparação com os títulos impressos, e com isso, a *Amazon* lucra sobre a venda de seus dispositivos (Adner 2012).



Figura 2. Exemplo de um *blueprint* de valor (Kindle da Amazon).
Adaptado de: Adner, R. (2012).

Em outras palavras, a Amazon apostou em um ecossistema melhor estruturado oferecendo aos usuários uma proposta de valor mais completa. Os gestores desenharam uma estratégia extremamente segura para as editoras (quanto ao conteúdo) e forneceram uma experiência completa e confortável para os usuários finais.

4. Metodologia

4.1. Visão Geral

Este trabalho descreve uma pesquisa empírica. Estudos desta natureza auxiliam a entender como e por que as coisas funcionam e nos permitem usar esta compreensão para alterar nosso mundo materialmente em várias áreas (Silva e Menezes 2001). Nesse contexto, existe um conjunto de estratégias para coletar informações a respeito da realidade estudada, que compreendem a metodologia. Em vista das metodologias disponíveis, a selecionada para a condução deste trabalho é a metodologia qualitativa (Silva e Menezes 2001) que permite compreender e avaliar a utilização de uma ferramenta ou abordagem por um grupo de profissionais.

Nesse sentido, a pergunta de pesquisa elaborada para nortear este estudo é a seguinte: “Como a abordagem *blueprint* de valor pode contribuir com as *startups* de software no contexto da gestão de riscos de ecossistemas de negócios? ”. Para ajudar a responder a essa questão optou-se pela coleta de dados com base em entrevistas semi-estruturadas individuais e coletivas (Silva e Menezes 2001). Estas possibilitam maior liberdade no momento da entrevista, e com isso pode-se explorar outras perguntas/respostas, dando liberdade também aos entrevistados. As entrevistas foram realizadas no mês de setembro de 2015.

Após a realização da coleta de dados, as transcrições das entrevistas e a análise de dados foram feitas. A análise de dados foi realizada com base na técnica de codificação e categorização (aberta e axial) proveniente da Teoria Fundamentada em Dados (TFD) (Strauss e Corbin 2008).

4.2. Contexto

O contexto do estudo compreende *startups* de tecnologia e de software. Esse contexto foi definido devido ao interesse em avaliar a abordagem em um cenário real. Para iniciar o processo de coleta de dados com as *startups* foi realizada uma pesquisa em busca de incubadoras, aceleradoras, fábricas de software, dentre outros que permitiriam a realização da pesquisa. Após um contato inicial, obteve-se uma resposta positiva da incubadora A¹. Assim, depois de diversas reuniões com a direção da incubadora, a primeira atividade realizada na incubadora foi a apresentação de uma palestra sobre a abordagem *blueprint* de valor. Esta foi ministrada em agosto de 2015 e teve a participação de onze *startups* incubadas. As *startups* tiveram liberdade para decidir participar, ou não, deste estudo. Assim, das *startups* presentes na palestra três tiveram interesse em participar. Estas *startups* serão mencionadas neste trabalho apenas como *startup* – 1, *startup* – 2 e *startup* – 3 por motivo de confidencialidade.

¹ Por motivos de confidencialidade, o nome das *startups* e dos participantes de todo o processo desta pesquisa serão anônimos.

A segunda atividade realizada foi um mini-workshop com as startups. Este workshop foi dividido em dois momentos. Primeiro, a apresentação dos conceitos sobre a *blueprint* de valor de forma detalhada. Nesse momento, explicou-se o passo-a-passo para a construção de um *blueprint* de valor, por meio de vários exemplos. E, segundo, foi sugerido aos empresários que elaborassem o *blueprint* inicial de seus produtos. O primeiro contato dos empresários com a *blueprint* foi mediado pela primeira autora, o qual se propôs a auxiliar na elaboração do *blueprint* de valor e a tirar dúvidas sobre a abordagem de um modo geral.

Ao final do mini-workshop foram entregues aos empresários um documento contendo alguns exemplos de *blueprints* e as atividades a serem realizadas por eles como: a versão inicial do *blueprint* e as descrições dos possíveis riscos (co-inovação e cadeia de adoção) encontrados. Os empresários tiveram duas semanas para realizarem estas atividades.

Novas reuniões de acompanhamento foram marcadas durante o mês de setembro, para que dúvidas adicionais pudessem ser respondidas e para que houvessem discussões acerca dos *blueprints* elaborados pelas startups.

Finalmente, entrevistas com os empresários das startups foram realizadas para buscar respostas para a questão de pesquisa, que retrata a contribuição da *blueprint* de valor no cenário de ecossistemas de negócios de software. A seguir são apresentados os processos de coleta e análises dos dados.

4.3 Processo de Coleta e Análise dos Dados

Na etapa de coleta de dados foram realizadas 3 entrevistas, sendo 2 informantes na “startup – 1”, 1 informante da “startup – 2” e 1 informante da “startup – 3”. Estas entrevistas foram gravadas para posterior análise. Além disso, informações foram coletadas nos vários momentos de interação entre pesquisador e participantes da pesquisa antes das entrevistas, o que colaborou para o desenvolvimento da entrevista semi-estruturada. Em média foram realizados 4 encontros com cada empresário ao longo de 2 meses de coleta de dados, sendo que cada reunião em média durou 3 horas e as entrevistas tiveram em média duração de 90 min. Estas reuniões foram realizadas para discutir os *blueprints* elaborados, bem como, para entender como esta abordagem estava sendo entendida e utilizada pelos empresários.

As entrevistas foram realizadas com o apoio de um guia de entrevistas. Este guia é composto por perguntas sobre as vantagens, desvantagens e sugestões de melhorias para a *blueprint*. Porém, para ajudar a alcançar o objetivo da pesquisa de forma ampla outros questionamentos foram incluídos no guia como: “utilizam outras ferramentas ou abordagens de projeto? ”, “qual o estágio do produto? ”, “a abordagem ajuda na comunicação?” Se sim “em que situação? Como? ”, e por fim, “Quando elaborar o *blueprint* de valor? ”.

As entrevistas foram transcritas e analisadas com o auxílio da ferramenta MAXQDA 11 (Verbi 2015), sendo analisada pelo método de categorização/codificação da Teoria Fundamentada em Dados (Strauss e Corbin 2008). A Figura 3 mostra as categorias resultantes codificação.

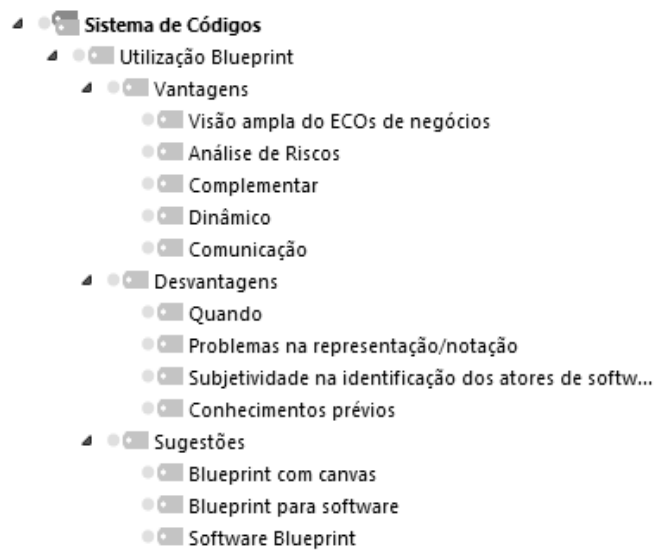


Figura 3. Lista de códigos e categorias.

Na Figura 3 todas as categorias e subcategorias contemplam trechos de entrevistas os quais ajudam a responder à questão de pesquisa deste trabalho. Neste instante são encontradas as categorias interessantes para a obtenção dos resultados desta pesquisa. As categorias mais relevantes são relacionadas entre si. Essas categorias são: (i) “vantagens” que são relatos com detalhes para avaliar as contribuições da abordagem no contexto estudado, (ii) “desvantagens” com relatos sobre pontos que podem ser considerados problemáticos, e, finalmente, (iii) “sugestões” apresentadas pelas *startups* para a abordagem. Cada uma destas categorias é discutida em detalhes na próxima Seção.

5. Resultados

Esta pesquisa buscou analisar a utilização da abordagem *blueprint* de valor no cenário de ecossistemas de negócios de software. Isto foi possível ao analisar as opiniões sobre as vantagens, desvantagens e sugestões para a utilização desta abordagem. Todos estes aspectos emergiram no estudo conduzido conforme a metodologia descrita na seção anterior.

Os resultados da análise permitem perceber, por exemplo, que a abordagem traz diversas *vantagens* aos empreendedores. Por exemplo, ela oferece uma visão ampla acerca dos elementos que compõem seus ecossistemas. Além disso, a análise de riscos sobre esses elementos pode ser utilizada de forma complementar a outras abordagens. Finalmente, ela é uma abordagem dinâmica e facilita a comunicação entre os profissionais da *startup*. Entretanto, a abordagem também apresenta *desvantagens*, entre elas o fato de ser subjetiva e o fato de não ser claro o melhor momento para sua adoção. Em decorrência de alguns pontos específicos entre as vantagens e as desvantagens os entrevistados deram algumas *sugestões* baseadas em seus produtos e negócios. Cada um destes resultados – vantagens, desvantagens e sugestões – serão apresentados a seguir. Ressaltando que as sugestões serão mencionadas juntamente as seções de vantagens e desvantagens pois surgem em decorrência de pontos específicos destas. Seguindo a tradição qualitativa, cada um destes resultados é apresentado acompanhado de trechos dos dados coletados nas entrevistas.

5.1 Vantagens

Dentre as vantagens mencionadas pelos entrevistados as principais são: o *blueprint* permite uma visão ampla do ecossistema de negócios e análise de riscos de co-inovação e da cadeia de adoção, facilita a comunicação, dinamismo e finalmente, ele é complementar a outras modelos e abordagens.

Quanto a visão ampla dos ecossistemas de negócios, os entrevistados apontam as vantagens em se ter conhecimento acerca dos componentes técnicos, de projeto e de negócios de seus ecossistemas, e principalmente em se identificar esses componentes. Isso compreende desde fornecedores, distribuidores, parceiros-chave até os clientes finais e ainda todos os complementadores de um ecossistema. Os trechos de entrevistas a seguir ilustram esta vantagem:

“(...) o principal é conseguir ver de uma forma geral o teu ecossistema né? Consegui ver no meu caso desde os fornecedores até o cliente final...” Startup – 2.

A *startup 2* afirma que o *blueprint* conseguiu ajudar a definir os papéis existentes em seus negócios permitindo a descoberta das dependências e do melhor caminho a percorrer para expandir o negócio.

“Mas no que [o blueprint] me ajudou foi [em] saber a relação entre o fornecedor e a gente. Não especificamente no fornecimento, quem vai ser [o] fornecedor [ou] quem não vai, mas na relação que a gente tem com fornecedor, que poder a gente tem, que poder a gente não tem. Acho que me ajudou a perceber isso.” Startup – 2.

A vantagem da análise de riscos agrupa relatos sobre duas diferentes vantagens: a primeira abrange as vantagens em conhecer os possíveis riscos adicionais existentes nos ecossistemas; e, a segunda expõe o quanto é importante tentar de alguma forma mitigar esses riscos e contornar a situação. De fato, os resultados sugerem a importância da reconfiguração do ecossistema (ver seção 3), isto é, nas estratégias para mitigar todos os gargalos problemáticos para o lançamento de um produto.

Os empresários tiveram a oportunidade de elaborar seus *blueprints* de valor e a partir deste realizar a análise dos riscos (ver seção 4). Dessa forma, eles descreveram a seguinte vantagem:

“(...) outra vantagem é que eu consigo analisar muito bem a partir desse ecossistema montado, os riscos de cada ponto em cada item do meu sistema, em cada nó, eu vou chamar de nó ... do ecossistema eu consigo ver quais são os meus principais riscos, as principais vantagens, os prós e os contras (...)” Startup – 2.

A *startup – 2* descreve que após a elaboração do *blueprint* “a gente consegue a partir daí elaborar uma estratégia de ação, a partir daí a gente já usa outras coisas alguma pesquisa que a gente fez de mercado com os clientes e fica mais fácil de ver [os riscos], neste gráfico do que num ‘textoção’ [o entrevistado refere-se ao texto do plano de negócios, que também menciona análise de riscos externos].”

Prosseguindo com a análise das vantagens da *blueprint*, tem-se que: avaliar as configurações alternativas para um ecossistema, é reconfigurá-lo (ver seção 3). Devido a essa possibilidade, os entrevistados classificaram a abordagem como dinâmica. Com a

elaboração do *blueprint* de valor é possível entender, acompanhar a evolução e fazer alterações no mapeamento que irão refletir na configuração real do ecossistema. Como relata a startup – 2.

(...) eu achei muito legal muito legal mesmo, porque é visual, eu gosto de ver coisas gráficas, eu acho que todo mundo né? A gente olha uma figura e já entendeu, não precisa estar em um texto explicando, a gente já entendeu aquilo ali e eu acho que o blueprint é assim, igual o canvas, ajuda bastante e é bem dinâmico eu gostei disso.”

O entrevistado da *startup* 2 reforça também o caráter visual da abordagem que é útil, pois desta forma é possível que os poucos integrantes da *startup* possam ter conhecimento de um ou mais setores dentro do negócio.

A abordagem por não se tratar de uma ferramenta computacional, a elaboração e alteração do mapeamento é feita de forma manual. Em vista disso, os entrevistados reportam uma sugestão que é a elaboração de um software para auxiliar no *design* de diagramas *blueprint*, tornando a criação do *blueprint* mais interativa.

Outra vantagem destacada é que o *blueprint* facilita a comunicação entre os profissionais da *startup*. Nesse sentido a startup 3 relata:

“(...) a abordagem é interessante, porque, tu tens que te debruçar numa realidade e ela provoca um debate, isso que é legal né? Então, tu viste? Nós começamos a conversar e as horas se passaram e ela provoca o debate e tu começa a ver realmente a realidade da empresa, começa a olhar a empresa, o ambiente, o ecossistema da empresa. Acho que o grande benefício é esse, provocar o debate em cima do negócio”. Startup – 3.

A *startup* 2 relatou uma situação positiva com relação a comunicação dentro de sua equipe, como relata a seguir:

“...por exemplo naquela primeira análise que a gente fez eu desci lá expliquei para os funcionários e eles já conseguiram perceber um monte de coisas que não percebiam, então porque a gente não corta aqui e ‘faz assim, faz assado’, então ficou bem legal pra trabalhar. ” Startup – 2.

Por fim, a última vantagem identificada é o fato da abordagem *blueprint* complementar outras abordagens e modelos já utilizadas pelas *startups*. Esses modelos e abordagens para gestão de projetos são: plano de negócios (Rosa 2013) e o *Business Model Canvas* (BMC²) (Osterwalder e Pigneur 2010). O plano de negócios é um dos requisitos obrigatórios para que estas *startups* pudessem candidatar-se a uma vaga na incubadora. Pode-se constatar que as *startups* consideram o plano de negócios necessário, porém caracterizam-no como uma abordagem textual pouco dinâmica, pois sentem a necessidade de consultá-lo sempre que necessário, mas encontram grande dificuldade para analisar certos pontos em meio a uma “*prosa gigante*” como o entrevistado da startup 1 se refere ao seu plano de negócios. A seguir um relato da startup 3 sobre isto:

² Osterwalder e Pigneur (2010) apresentam o *Business Model Canvas* (BMC) que é um modelo composto por 9 blocos, os quais ajudam a empreendedores criarem seus modelos de negócios, por meio de conceitos sobre como uma organização cria, entrega e captura valor.

“O plano de negócios ele é uma coisa meio trabalhosa e acaba se tornando pouco dinâmica, né? É o ‘documentozão’ e a gente tem que analisar muitas coisas. O canvas não, é um negócio que a gente tá vendo visualmente, que a gente pode alterar rapidamente e o blueprint a mesma coisa.” Startup – 3.

O modelo mencionado pelos entrevistados é o BMC. Esta modelo é muito utilizada e bastante elogiada principalmente pela agilidade de visualização que proporciona aos entrevistados sobre seus modelos de negócios. As *startups* mencionam que ao fazerem seus *blueprints* perceberam que os riscos externos não faziam parte do *canvas*, que eles já haviam elaborado, e por isso relatam que a elaboração do *blueprint* de valor dos seus ecossistemas após a elaboração de seus *canvas* os ajudou a perceberem os pontos cegos de seus negócios com maior eficiência. A startup 1 relata:

“...a gente adiciona alguém ou tira alguém? Vai atualizando [o] blueprint como a gente atualiza o canvas, entendeu? Porque como eu te falei, tem muita coisa que tá no blueprint que não está no canvas por isso que eu acho que [o blueprint] complementa o canvas.” Startup – 1.

Em meio às entrevistas, a *startup – 2* faz uma sugestão, que é uma fusão entre a abordagem *blueprint* de valor e o BMC, como pode ser visto nos trechos de entrevistas a seguir:

“(...) o que eu acho que seria muito interessante mesmo era fazer uma espécie de fusão do canvas com o blueprint aí eu acho que ficaria um negócio bem legal mesmo, bem dinâmico, por que se juntar o canvas com o blueprint eu acho que sai uma coisa sensacional, porque o blueprint consegue avaliar os riscos daquele modelo lá do Canvas se aquilo é válido ou não.”. Startup – 2.

5.2 Desvantagens

Esta seção analisa os relatos sobre as desvantagens da abordagem tem-se: subjetividade quanto a identificar os atores de software e quando elaborar o mapeamento.

A desvantagem quanto a subjetividade está relacionada a dificuldade de correlacionar os elementos apresentados no *blueprint* genérico, ao contexto de software. Um trecho de entrevista, a seguir, mostra a visão da *startup 1* sobre esta questão.

“Olha uma dificuldade que eu tive com blueprint foi quando eu fui fazer foi justamente àquela questão, porque depende muito de cada produto, de cada negócio (...) seria o que pegar pra colocar [no] blueprint (...)” Startup – 1.

A desvantagem, exposta acima, acerca da dificuldade em identificar os atores para software, gerou mais uma sugestão, os entrevistados propõem a criação de um *blueprint* específico para software, ou seja, um modelo pré-definido. Já que sentem dificuldade em identificar quais componentes devem considerar e como deve ser desenhada a estrutura de relacionamentos em seus ecossistemas.

Quanto a desvantagem sobre quando elaborar o *blueprint* ou qual o melhor momento para dar início a elaboração em negócios de software. Observa-se que as *startups* entrevistadas encontram-se em diferentes fases de elaboração de seus produtos, portanto, exprimem opiniões com base nesses estágios e em suas breves experiências de negócios. Nesse sentido, as startups 2 e 3 relatam que:

“Eu acho que depende, porque se fosse mais no início a gente não teria ainda amadurecido muito bem quem seriam, por exemplo, os nossos fornecedores e a

gente não teria uma ideia realmente clara do que eu quero, nosso contexto, eu acho que no meu caso eu conheci essa abordagem no momento perfeito para minha empresa que é o momento que eu tenho muito bem definido quem são meus clientes, quem é o meu público alvo, como eu atinjo meus clientes quem são os clientes dos meus clientes, porque na verdade eu só vendo se o usuário final tiver necessidade de comprar, o meu cliente só vai me comprar se o cliente dele quiser comprar também, eu tenho muito claro toda a minha cadeia, agora [depois de usar a blueprint], desde o meu fornecedor até o desenvolvimento e montagem.” Startup – 2.

E por fim, a *startup – 3* já tem o seu produto lançado no mercado e relata que nesta fase é interessante estruturar o *blueprint*, mas afirma que dependendo do tamanho da empresa, ou seja, uma empresa maior, o *blueprint* com certeza deveria ser elaborado antes.

“Eu não sei se no lançamento de um produto isso funcionaria (...) Dessa forma, o que acontece, para mim funcionou. Hoje digamos se não tivesse o blueprint eu teria menos uma abordagem que ajudou. E ainda tem mais, o meu produto é diferente, se eu sentar no computador agora e quiser mudar todo ele eu mudo né? Eu não mudo a alma, mas eu mudo o conteúdo, o design, a forma de passar e tudo mais. Eu tenho essa possibilidade, mas eu não sei, em uma grande empresa acho que isso deveria ser feito antes, pois requer muito dinheiro de investimento e tem que ter muito cuidado, por isso com certeza antes seria bem melhor.” Startup – 3.

A próxima seção aborda a discussão entre os resultados deste trabalho, qual o impacto positivo para o contexto de riscos em ecossistemas de negócios, e também uma comparação com estudos relacionados aos temas abordados neste trabalho.

6. Discussão

Os resultados da seção anterior permitem constatar que estruturar o ambiente de negócios incluindo as relações entre as empresas, bem como definir os papéis desempenhados por cada uma delas pode auxiliar a identificar e diminuir os riscos desse ambiente, além de auxiliar a amadurecer a proposta de valor de um produto. Isto é, apenas gerenciar os riscos técnicos e de projetos, já não é suficiente para as empresas se manterem competitivas no mercado atual. Elas precisam gerenciar as relações com outras empresas de seu ecossistema. Isto é particularmente importante para a comunidade de Sistemas Colaborativos brasileira, pois a mesma tem historicamente focado em colaborações entre pessoas de uma forma direta (através de estudos empíricos) ou indireta (através de ferramentas de apoio a colaboração). Uma possível exceção para isto são os estudos de emergências e crises e estudos de processos de negócios que enfatizam colaboração entre diferentes organizações, embora geralmente focando nas pessoas e nos papéis que eles desempenham. Neste estudo, o foco é apenas na colaboração – através das dependências e dos riscos – entre organizações, sem lidar com pessoas ou papéis. Assim, este artigo visa enfatizar a colaboração em um nível de abstração diferente de estudos anteriores.

Esta seção também discute brevemente os resultados anteriores no contexto de outras metodologias existentes na literatura de *startups* que visam nortear e diminuir os

riscos de negócios das mesmas. Por exemplo, o *customer development*³ (Blank e Dorf 2012) surgiu com a missão de tornar o processo de abertura de *startups* menos arriscado. Ao se comparar o *customer development* e a abordagem *blueprint* de valor nota-se que os conceitos são complementares. Ambas focam em aprendizado e não mais apenas em execução (Blank e Dorf 2012). Entretanto, a *blueprint* possui conceitos adicionais como a possibilidade de se projetar e visualizar o MVE do negócio (seção 3). Este MVE pode basear-se nos testes de hipóteses e experimentações do *customer development* e nas estratégias de reconfiguração de ecossistemas da *blueprint* - “*as cinco alavancas para a reconfiguração de ecossistemas*” (Adner 2012). Ou seja, utilizando as abordagens em conjunto, além do foco no MVP (Mínimo Produto Viável) é possível também pensar em um MVE. Isto sugere uma outra vantagem da abordagem *blueprint* de valor que é a permitir uma visão sobre estratégia de negócios.

Existem duas outras abordagens frequentemente utilizadas em *startups*, a saber: o plano de negócios e o *business model canvas*, ou simplesmente BMC (Osterwalder e Pigneur 2010). O *blueprint* se diferencia do plano de negócios principalmente por considerar a existência de um caminho a ser percorrido pelo produto até o cliente final. O plano de negócios aborda de forma geral muitos aspectos de negócios, que podem gerar deduções de negócios precipitadas sobre as condições do mercado. Além disso, conforme apresentado na seção de resultados, ele é um documento grande que não acompanha a agilidade de empresas de software. Em contraponto o *blueprint* de valor é dinâmico e proporciona agilidade ao *design* de ecossistemas de negócios e ainda aborda a existência de componentes adicionais e a gerência de riscos.

Em relação ao BMC pode-se perceber que o mesmo aborda questões como o desenvolvimento de uma boa compreensão do ambiente da sua organização, ajudando a desenvolver modelos de negócios mais competitivos e mais fortes. Isto é, com o BMC percebe-se a preocupação em descrever a lógica de como uma organização cria, entrega e captura valor. Ele analisa também o diferencial frente a seus concorrentes, mas não aborda a visão sobre ecossistemas, e também não foi elaborado para gerenciar riscos, como a abordagem *blueprint*.

Finalmente, é importante comentar sobre as desvantagens indicadas da abordagem, bem como as sugestões. Devido a questões de espaço, não se pode discutir todas. A respeito da desvantagem sobre a subjetividade da abordagem. Pode-se analisar o ponto de vista a seguir. Adner (2012) apresenta uma visão sobre o ecossistema, embasada na existência de fornecedores de componentes e de fornecedores de complementos. Com base nisso, o *blueprint* genérico é composto por componentes comumente encontrados em ecossistemas de negócios. No entanto, dependendo do produto de software, os elementos de negócios podem distanciar-se da visão de fornecedores, distribuidores, varejistas, entre outros papéis comumente encontrados em ecossistemas de negócios. Nesse contexto, os entrevistados sugerem a criação de um *blueprint* específico para software, ou seja, um modelo pré-definido. Osterwalder e Pigneur (2010) chamam estes modelos de negócios com características similares de padrões. Isto é, ao se analisar o *canvas* pode-se perceber que algumas empresas possuem o mesmo padrão de modelo de negócios. De maneira análoga, o *blueprint* para

³ A metodologia *customer development*, descreve que a melhor maneira de achar um modelo de negócios escalável, repetível e lucrativo é executar o seguinte fluxo: definir hipóteses; criar experimentos; testar; e obter *insights* e voltar ao primeiro passo (Blank e Dorf, 2012).

software, pode apresentar padrões diferentes para cada tipo de produto de software. Isto é algo que os autores pretendem investigar no futuro. Em resumo, acredita-se que uma análise de padrões é importante para nortear e ajudar a pensar nas estratégias e no ecossistema propriamente dito.

7. Considerações Finais e Trabalhos Futuros

O contexto atual de negócios tem se tornado diferente de anos anteriores a partir da integração de várias empresas, isto é, atualmente, uma empresa pode possuir diversas relações comerciais com outras. *Apple* e *Samsung*, por exemplo, ao mesmo tempo que são parceiras em um aspecto (micro-processadores) são concorrentes em outro (telefones celulares). Assim, entender as relações entre as empresas torna-se essencial para o sucesso de uma organização, especialmente *startups*, que estão naturalmente inseridas em um cenário de riscos e incertezas. A abordagem *blueprint* de valor (Adner 2012) difere de propostas por enfatizar a análise de riscos em atores de ecossistemas de negócio. Além disso, ela sugere estratégias para reorganizar o ecossistema visando aumentar as chances de sucesso de uma empresa, e especialmente *startups*.

Este trabalho descreveu um estudo empírico com 3 *startups* da área de tecnologia de informação que utilizaram o *blueprint* para entender o seu ecossistema, bem como, refletir em como ele precisaria ser reorganizado para diminuir os riscos destas empresas. A análise da utilização do *blueprint* por estas empresas mostrou que o mesmo traz diversas vantagens, embora ele ainda tenha algumas desvantagens.

De um modo geral, o trabalho foca na colaboração entre empresas, ao invés de pessoas, um aspecto pouco explorado na comunidade brasileira e internacional de Sistemas Colaborativos. Os autores esperam que este artigo sirva de motivação para outros autores realizarem pesquisa neste aspecto.

As limitações deste trabalho correspondem: ao número de participantes da pesquisa, embora tenha-se constatado contribuições e algumas desvantagens, este comportamento não pode ser generalizado; outro ponto, é que se gostaria de ter acompanhado a utilização da abordagem pelos entrevistados durante um período de tempo maior. Porém, isto não foi possível por questões de prazos e disponibilidade dos participantes.

Agradecimentos

Os autores gostariam de agradecer ao CNPq pelo financiamento desta pesquisa através dos processos 485070/2013-8 e 310468/2014-0, e em especial a Chamada 59/2013 MCTI/CT-Info/CNPq com processo 440880/2013-0. A primeira autora também gostaria de agradecer a CAPES pela bolsa de mestrado.

Referências

- Adner, R. (2012) “The Wide Lens – What Successful Innovators See That Others.” Miss Penguin Group (USA) Inc., 375 Hudson Street, New York, USA.
- Blank, S. e Dorf, B. (2012) “The Startup Owner’s Manual.” 571 páginas. ISBN-10:0984999302.
- Gage, Deborah. (2012) “The Venture Capital Secret: 3 out of 4 Start-ups Fail.” The Wall Street Journal. Small business. 2012.

- Ikenami, Rodrigo Kazuo e salermo, Mario Sergio. (2015) Ecosistema de negócio: estudo exploratório acerca da delimitação de suas fronteiras. XXXV Encontro Nacional de Engenharia de Produção. Perspectivas Globais para a Engenharia de Produção Fortaleza, CE, Brasil, 13 a 16 de outubro de 2015.
- Jesus, Leandro. (2014) “Das cadeias de valor tradicionais para os Ecosistemas. ” Conteúdo BPM. Disponível em: < <http://www.bpmglobaltrends.com.br/das-cadeias-de-valor-tradicionais-para-ecossistemas/> >. Acessado em janeiro 2016.
- Leon, André. (2009) “Convergência Digital mídias Integradas. ” ComCiência, Campinas, n.110, 2009.
- Maurya, A. (2012) “Running Lean”. 2 nd ed., Califórnia: O’Reilly, 2012.
- Mueller, Roland M.; thoring, Katja. (2012) “Design Thinking vs. Lean Startup: A Comparison of Two User-Driven Innovation Strategies.” In: International Design Management Research Conference, p. 151-160, 2012, Boston.
- Moore, J.F. (1993) “Predators and Prey, A New Ecology of Competition.” Harvard Business Review, May-June.
- Osterwalder, A. e Pigneur, Y. (2010) “*Business Model Generation.*” 1st ed., New Jersey: John Wiley & Sons, Inc., 2010.
- Ries, E. (2012) A Startup Enxuta. 1a ed. São Paulo: Leya, 2012.
- Rosa, Claudio. 2013. Como elaborar um plano de negócios. SEBRAE 2013.
- Silva, E. e Menezes, E. (2001) “Metodologia da Pesquisa e Elaboração de Dissertação. ” Florianópolis– 3. ed. rev. atual. – Florianópolis: Laboratório de Ensino a Distância da UFSC. 121p. (2001).
http://www.scielo.br/pdf/prod/v22n1/aop_t6_0010_0155.pdf.
- Strauss, A. e Corbin, J. (2008) “Pesquisa Qualitativa: Técnicas e Procedimentos para o Desenvolvimento de Teoria Fundamentada. ” Artmed.
- Tansley, A. G. (1935) “The Use and Abuse of Vegetational Concepts and Terms.” Ecology. Ecology Society of America, v. 16, ed. 3, p.284-307, jul, 1935.
- Taveira, A. (2012) “Análise dos aspectos que afetam o surgimento e desenvolvimento de um Ecosistema de Software. ” 62 f. Trabalho de Conclusão de Curso (Graduação) - Ciência da Computação, Centro de Informática, UFPE, Recife.
- VERBI Software. (2015) “MaxQDA – The art of text analysis. 2015.” Disponível em: <http://www.maxqda.com>.

Um Processo para o Desenvolvimento de Serviços de Colaboração em um Ecossistema de Software Científico

Marcio G. Arakaki¹, Guilherme G. Martins¹, José Maria N. David¹, Regina Braga¹, Fernanda Campos¹, Frâncila W. Neiva²

¹Instituto de Ciências Exatas – Departamento de Ciência da Computação –
Universidade Federal de Juiz de Fora (UFJF)
Caixa Postal 36.036-330 – Juiz de Fora – MG – Brasil

²Programa de Pós-Graduação em Informática –
Universidade Federal do Rio de Janeiro (UFRJ)
Caixa Postal 68.530 – Rio de Janeiro – RJ – Brasil

{marcio.arakaki, guilherme}@ice.ufjf.br, {jose.david, regina.braga,
fernanda.campos}@ufjf.edu.br, fran.weidt@gmail.com

Abstract. *Software ecosystem capability to evolve by employees, enables the creation of an environment that can meet specific collaboration needs of each team. However, developing new services to an ecosystem can be a difficult task. This paper proposes a process for the development of collaboration services to a Scientific Software Ecosystem, E-SECO. To evaluate the process, a service was created and extended in accordance with the steps in the process. Thus, it was possible to obtain evidence regarding the validity of the process.*

Resumo. *A capacidade de ecossistemas de software de evoluírem por colaboradores possibilita a criação de um ambiente que possa atender às necessidades de colaboração específicas de cada equipe. Entretanto, desenvolver novos serviços para um ecossistema pode ser uma tarefa difícil. Este trabalho propõe um processo para o desenvolvimento de serviços de colaboração para um Ecossistema de Software Científico, o E-SECO. Para avaliar o processo, um serviço foi criado e, posteriormente, estendido de acordo com as etapas apresentadas pelo processo. Com isso, foi possível obter evidências que apontam para a validade do processo.*

1. Introdução

Experimentação científica é uma atividade colaborativa e complexa que muitas vezes exige um conhecimento diversificado por parte dos cientistas. Essa natureza da experimentação científica e o avanço da tecnologia possibilitaram o crescimento da *e-Science* que, segundo Mattoso *et al.*, (2008), pode ser definida como uma ciência em larga escala cujo objetivo é apoiar os cientistas no desenvolvimento de experimentos colaborativos e distribuídos por meio de uma infraestrutura computacional.

O E-SECO [Freitas *et al.*, 2015] é uma abordagem que utiliza os conceitos de Ecossistemas de Software Científico (ECOSC) a fim de auxiliar cientistas durante as etapas de um experimento científico [Belloum *et al.*, 2011]. Freitas *et al.*, (2015) definem um ECOSC como um Ecossistema de Software que relaciona fornecedores de software científico, institutos de pesquisa, cientistas, órgãos de fomento, instituições

financiadoras, e as partes interessadas nos resultados da pesquisa. Pela própria natureza de Ecossistemas de Software, o E-SECO é uma plataforma aberta, o que possibilita a sua evolução por meio da adição de serviços à infraestrutura e de melhorias sugeridas por desenvolvedores externos à equipe mantenedora. Com isso, é possível desenvolver os serviços necessários para oferecer suporte à colaboração entre os cientistas. Com o objetivo de promover as atividades neste contexto, os serviços de colaboração devem apoiar a comunicação, a coordenação, a cooperação e, por fim, a percepção, entre outros elementos que interferem diretamente no modelo de colaboração [Fuks *et al.*, 2003].

A forma como cientistas colaboram varia de acordo com o contexto do experimento, o ambiente de trabalho e as pessoas envolvidas. Por isso, durante o desenvolvimento, os aspectos específicos do contexto em que os serviços de colaboração serão inseridos devem ser considerados, de forma a atender os requisitos funcionais e não funcionais de cada instância do experimento. Mesmo contendo características específicas, estes serviços possuem elementos comuns e variáveis quanto aos conceitos de colaboração. Nesse cenário, o uso de Linhas de Produto de Software (LPS) possibilita reduzir o retrabalho e diminuir o esforço do desenvolvimento de novos serviços [Gurp *et al.*, 2001].

Ao desenvolver serviços de colaboração para o E-SECO, é necessário considerar as singularidades da infraestrutura do ecossistema, de modo a garantir que seus aspectos de qualidade, tais como flexibilidade, extensibilidade e escalabilidade [Freitas *et al.*, 2015], sejam mantidos. O problema é que, uma vez que o E-SECO é uma iniciativa *open-source* em que toda comunidade pode contribuir para sua evolução, fazer com que esta comunidade considere, durante o desenvolvimento de serviços de colaboração, às singularidades da infraestrutura do ecossistema se torna um trabalho difícil e complexo. A falta de uma iniciativa que apoie a comunidade durante o desenvolvimento dificulta a manutenção de um ecossistema de software aderente aos requisitos não funcionais e, ao mesmo tempo, prejudica a criação de serviços de qualidade que poderiam ser reutilizados.

Para enfrentar tal problema, este trabalho contribui com um processo de referência para o desenvolvimento de serviços de colaboração no contexto de um ecossistema de software científico, em específico, considerando a infraestrutura do E-SECO [Freitas *et al.*, 2015]. No contexto de LPS, processos também são utilizados para apoiar o desenvolvimento de software [Gimenes and Travassos, 2002] [Kang *et al.*, 2002]. Porém, no melhor do nosso conhecimento, não encontramos propostas de processos de desenvolvimento aplicados ao desenvolvimento de serviços de colaboração em ecossistemas de software científico.

O restante deste artigo no está organizado da seguinte forma: a seção 2 apresenta alguns trabalhos relacionados ao processo proposto, a seção 3 descreve as etapas do processo, a seção 4 descreve o funcionamento do processo enquanto a seção 5 apresenta sua avaliação, por fim, a seção 6 apresenta as considerações finais do trabalho.

2. Trabalhos Relacionados

Existem alguns trabalhos na literatura que apoiam a colaboração no contexto de *e-Science*. Zhang et al., (2014) apresentam Confucius, uma ferramenta de suporte à concepção colaborativa de *workflows* científicos por meio do apoio de uma ontologia de

proveniência. Confucius foi desenvolvido a partir do sistema Taverna¹, um sistema de gerenciamento de *workflow* de código aberto. Confucius se baseia no rastreamento das interações humanas durante o processo de concepção de um *workflow*, gravando-as com o apoio da ontologia de proveniência. Segundo os autores, essa ontologia é centrada nos conceitos de *workflow*, que compreendem, por exemplo, a organização de tarefas, a ligação entre os dados, e a definição de pré-requisitos. Em suma, esta ontologia armazena informações sobre as atividades, usuários envolvidos com a concepção, nível de acesso de cada usuário a cada atividade do *workflow*, e as provê ao sistema para apoiar a coordenação do processo. Apesar de Confucius fornecer apoio à colaboração, não oferece, por exemplo, serviços de comunicação para apoiar os cientistas em experimentos geograficamente distribuídos apoiados por uma infraestrutura que utiliza a abordagem de ecossistemas de software.

Miranda *et al.*, (2014) apresentam CollabCumulus, uma abordagem que auxilia a análise e a consulta de dados de proveniência produzidos pela execução de *workflows* científicos em ambientes distribuídos. Seu objetivo é abstrair para os cientistas a estrutura e a linguagem de consulta ao banco de dados que os apoia nos experimentos. Para isso, CollabCumulus apresenta um portal, que funciona como uma rede social, que se comunica com a base de dados de proveniência, de modo a possibilitar que um cientista realize e compartilhe suas análises e opine sobre as análises de outros cientistas. Assim, como o trabalho anterior, este trabalho não fornece apoio à colaboração de forma distribuída em um contexto aberto, além disso, a abordagem não apresenta uma proposta para apoiar a criação de novos serviços de colaboração ou evoluir os serviços existentes.

Gadelha *et al.*, (2009) apresentam uma LPS para a derivação de serviços de colaboração. Essa LPS foi construída a partir da análise de algumas aplicações colaborativas de modo a representar as semelhanças e diferenças nelas encontradas. Esta abordagem se baseia no Modelo de Colaboração 3C [Fuks *et al.*, 2003] para orientar o desenvolvedor a selecionar os conceitos de colaboração desejados no serviço. Assim como a proposta Gadelha *et al.*, (2009), a pesquisa deste artigo se baseia no Modelo 3C para organizar os conceitos de colaboração, entretanto, estes conceitos são utilizados por meio de uma Ontologia de Colaboração [Martins *et al.*, 2015]. Além disso, Gadelha *et al.*, (2009) apresentam uma única LPS que deriva como produto uma aplicação de colaboração que contém vários serviços, por exemplo, uma aplicação que contém Lista de Discussões e Troca de Mensagens, enquanto este trabalho propõe a utilização de uma LPS no contexto de uma infraestrutura (aberta) que utiliza a abordagem de ecossistemas de software. Além disso, neste contexto, as pesquisas previamente apresentadas não oferecem um processo para a construção de serviços de colaboração considerando os requisitos que a infraestrutura necessita.

Diferente dos trabalhos encontrados na literatura, esta pesquisa contribui para os avanços das pesquisas em sistemas colaborativos no contexto de *e-Science* através de uma proposta de processo para o desenvolvimento de serviços de colaboração (criação e/ou extensão) aplicado ao domínio de ECOSC. A proposta de processo é apresentada a seguir.

¹ <http://www.taverna.org.uk/>

3. Processo para o Desenvolvimento de Serviços de Colaboração

Para desenvolver serviços de colaboração, no contexto do E-SECO, os desenvolvedores devem seguir o processo de desenvolvimento proposto de acordo com a infraestrutura do ecossistema. Este processo tem por objetivo padronizar a criação e a evolução de componentes de colaboração, de modo a apoiar a variabilidade, especificidade e interoperabilidade entre os serviços de colaboração desenvolvidos.

Ao se desenvolver um serviço de colaboração, o mesmo será integrado à infraestrutura do ecossistema e poderá ser utilizado por qualquer instância do E-SECO, de acordo com suas necessidades. Para isso, o processo de desenvolvimento envolve os conceitos de Linhas de Produto de Software (LPS), utilizando um repositório *online* como núcleo de artefatos, para apoiar a produção dos serviços de colaboração que são utilizados na instância.

A adoção de processos traz benefícios [Gimenes and Travassos, 2002], entre eles a documentação e o compartilhamento de boas práticas. No E-SECO, as boas práticas representadas por meio do processo proposto visam garantir serviços reusáveis e compatíveis, o contínuo atendimento aos requisitos não funcionais demandados por um ECOSC. Adicionalmente, pode diminuir a complexidade do desenvolvimento de serviços, uma vez que os desenvolvedores agora possuem um processo bem estabelecido para guiá-los.

Dessa forma, acredita-se que, sem um processo para auxiliar o desenvolvimento de serviços de colaboração, estes poderão ser construídos sem uma adequação aos processos já existentes na infraestrutura do E-SECO [Freitas *et al.*, 2015]. Por exemplo, serviços devem estar alinhados aos processos de suporte à interoperabilidade, de busca e recuperação de componentes, entre outros. A seguir, um processo para o desenvolvimento de serviços de colaboração será apresentado. Tal processo é inspirado em processos presentes na literatura ([Gimenes and Travassos, 2002] [Kang *et al.*, 2002]) e aborda elementos do contexto do E-SECO ([Costa *et al.*, 2015] [Freitas *et al.*, 2015] [Neiva *et al.*, 2015]).

3.1. Desenvolvimento de Serviços de Colaboração

A Figura 1 ilustra o processo proposto para o desenvolvimento de serviços de colaboração. Este processo é composto por seis atividades principais.

Etapa 1: Planejamento Inicial. Atividade inicial do processo por meio da qual é gerada uma especificação inicial sobre as características desejadas para um serviço de colaboração (Figura 1 - 1). Como resultado, tem-se uma especificação inicial de um serviço de colaboração.

Etapa 2: Busca. Utilizando as informações da etapa anterior, realiza-se uma busca na plataforma do ecossistema a fim de encontrar uma LPS que atenda total ou parcialmente à especificação criada (Figura 1 - 2). Como resultado da busca podem ocorrer três casos. No primeiro caso, nenhuma LPS que derive um serviço de colaboração, que atenda às necessidades desejadas, foi encontrada (Figura 1 - 3A), com isso, o serviço deverá ser desenvolvido desde o início (Processo de Criação). No segundo caso, encontra-se uma LPS que atende parcialmente às necessidades desejadas (Figura 1 - 3B). Com isso, a LPS será desenvolvida a partir dos artefatos existentes de modo a evolui-la (Processo de Extensão). Por fim, no terceiro caso, encontra-se uma

LPS que atende completamente às necessidades desejadas (Figura 1 – 3C). Com isso, não é necessária nenhuma alteração na LPS, uma vez que já existe a possibilidade de derivação de um serviço com as características desejadas. Como resultado desta etapa, tem-se o conjunto de artefatos pertencentes à LPS, caso alguma seja encontrada. É importante destacar que a atividade de busca é responsável por evitar retrabalho, ou seja, o desenvolvimento de serviços de colaboração redundantes. Contudo, não faz parte do escopo deste trabalho definir uma estratégia de busca e recuperação de serviços.

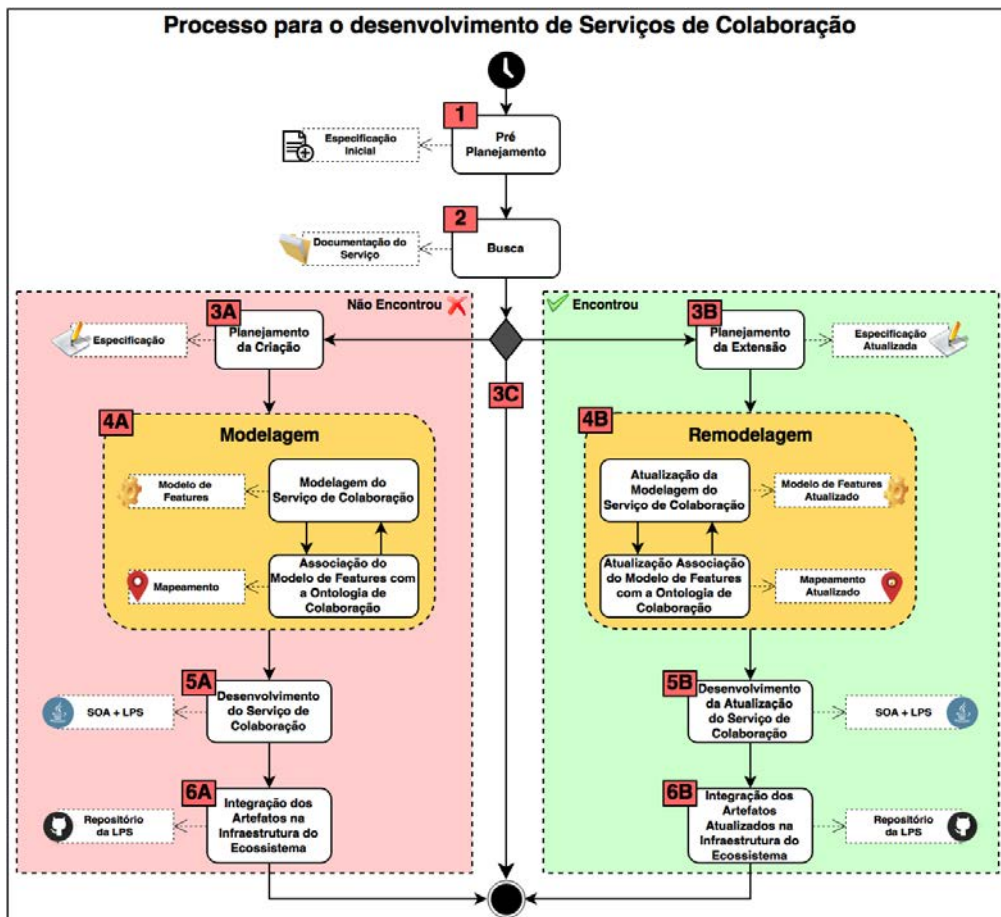


Figura 1. Processo para o Desenvolvimento de Serviços de Colaboração

As etapas posteriores do processo possuem semelhanças e diferenças entre a criação e a extensão de uma LPS que derive um serviço de colaboração. Por isso, serão divididas em dois casos, o primeiro referente à criação e o segundo à extensão.

Etapa 3: **Planejamento**. No caso da criação (Figura 1 - 3A), para desenvolver um novo serviço de colaboração deve-se fazer o seu planejamento, tendo como base a especificação inicial, de modo a definir as características do serviço. Além disso, deve-se analisar sua necessidade e sua contribuição para o ecossistema. Ao final desta etapa, tem-se um conjunto de documentos referentes à especificação do serviço de colaboração, por exemplo, descrição das funcionalidades e *features*, descrição de casos de uso, documento de requisitos e qualquer outro documento que a equipe desenvolvedora julgar necessário. O caso da extensão (Figura 1 - 3B), assim como no caso anterior, se refere ao planejamento do serviço de colaboração, a diferença é que o trabalho se concentra na evolução da documentação existente, e, caso seja necessário,

poderão ser adicionados novos documentos. Ao final desta etapa, tem-se um conjunto atualizado de documentos referentes à especificação do serviço de colaboração.

Etapa 4: **Modelagem**. Uma das propostas do E-SECO é que seus serviços sejam desenvolvidos de acordo com uma LPS. Para representar as variabilidades e semelhanças dos serviços presentes na LPS, utiliza-se o Modelo de *Features* [Kang *et al.*, 2002]. Como apresentado em Costa *et al.*, (2015), realiza-se o mapeamento do Modelo de *Features* com a Ontologia de Colaboração de modo a agregar mais semântica ao Modelo de *Features*. A Ontologia de Colaboração apresenta os conceitos que envolvem os serviços de colaboração em um ecossistema, e o relacionamento entre esses conceitos, possibilitando a verificação de consistência dos serviços gerados e a derivação de novos conhecimentos por meio do mecanismo de inferência [Martins *et al.*, 2015]. Como pode ser observado na Figura 1 (itens 4A e 4B), a modelagem é uma atividade cíclica que varia entre a criação do Modelo de *Features*, e seu mapeamento com a ontologia. Isso se deve ao fato de que a ontologia relaciona os conceitos de colaboração, e estes podem influenciar a criação e modificação do Modelo de *Features*, apresentando, por exemplo, conceitos que deveriam ser contemplados pelo modelo que, em um primeiro momento, não haviam sido incluídos. No caso da criação de serviços (Figura 1 - 4A), um novo Modelo de *Features* e um novo arquivo de mapeamento entre ele e a Ontologia de Colaboração devem ser criados segundo o planejamento da etapa anterior. Ao final desta etapa, tem-se um arquivo XML com o Modelo de *Features* do serviço de colaboração, e um arquivo XML com o mapeamento entre o Modelo de *Features* e a Ontologia de Colaboração. No caso da extensão (Figura 1 - 4B), tanto o Modelo de *Features* quanto o arquivo de mapeamento devem ser alterados segundo a evolução planejada na etapa anterior. Ao final desta etapa, tem-se o arquivo XML do Modelo de *Features* e o arquivo XML do mapeamento atualizados.

Etapa 5: **Desenvolvimento**. Essa atividade representa o desenvolvimento do serviço de colaboração. Para seguir o padrão de desenvolvimento do E-SECO, os serviços devem ser desenvolvidos seguindo os conceitos de (i) uma LPS, o que possibilita maior variabilidade do serviço de colaboração, além de facilitar a evolução do serviço por meio do reuso, e (ii) os conceitos de Arquitetura Orientada a Serviço (SOA do inglês *Service Oriented Architecture*), de modo a facilitar a composição e a interoperabilidade entre os serviços do ecossistema [Neiva *et al.*, 2015]. É importante destacar que cada serviço de colaboração possui sua própria LPS, na qual os componentes são *features* e os produtos são serviços que apresentam um subconjunto dessas. No caso da criação (Figura 1 - 5A), o desenvolvimento deve ser realizado desde o início de acordo com o planejamento e a modelagem realizadas nas etapas anteriores. Ao final desta etapa, tem-se a LPS desenvolvida e apta a derivar serviços de colaboração. No caso da extensão (Figura 1 - 5B), o desenvolvimento da evolução deve ser realizado a partir da LPS existente de acordo com o planejamento e a modelagem realizadas anteriormente. Ao final desta etapa, tem-se a LPS atualizada e apta a derivar novos serviços de colaboração. É importante destacar que, nesta etapa, tanto na criação quanto na extensão, é desejável que se realizem testes à LPS de modo a verificar se o que foi desenvolvido está consistente com o que foi especificado. Estes testes devem ser documentados e anexados ao conjunto de artefatos da LPS do serviço de colaboração.

Etapa 6: **Integração**. Nesta última atividade, o serviço desenvolvido será integrado à Infraestrutura do ecossistema, juntamente com os artefatos gerados nas atividades anteriores. Neste trabalho, a integração ocorre através de um repositório no

GitHub², onde todos os serviços desenvolvidos, e os artefatos relacionados a eles, são armazenados e disponibilizados para serem utilizados por qualquer instância do E-SECO que o deseje. No caso da criação (Figura 1 - 6A), cria-se uma nova área no repositório para armazenar os artefatos gerados em todas as etapas anteriores. Ao final desta etapa, tem-se a LPS integrada à infraestrutura do E-SECO. No caso da extensão (Figura 1 - 6B), os novos artefatos, e os que foram modificados, são armazenados na área já existente do serviço estendido. Ao final desta etapa, tem-se a LPS integrada à infraestrutura do E-SECO. Em ambos os casos, os artefatos submetidos à integração com a infraestrutura do E-SECO serão avaliados e, caso sejam aprovados, serão anexados à infraestrutura.

Após a integração da LPS do serviço de colaboração à infraestrutura do ecossistema, ele estará apto a derivar serviços que poderão ser utilizados por qualquer instância do E-SECO. Como parte do trabalho, o processo proposto foi avaliado de duas formas, uma avaliando a criação e outra a extensão de serviços. Para a criação, o processo foi utilizado para a criação de um novo serviço de colaboração para o E-SECO. A extensão foi avaliada por meio de um estudo experimental onde um serviço de colaboração presente no ecossistema foi estendido. As avaliações são descritas a seguir.

4. Processo de Criação de Serviço em Ação

Com o objetivo de avaliar a viabilidade do processo quanto ao caso de criação de um novo serviço de colaboração para o E-SECO, esta seção apresenta a criação de um serviço de Lista de Usuários.

Seguindo o processo, a primeira etapa é o planejamento inicial, na qual cria-se uma especificação inicial da Lista de Usuários contendo uma breve descrição do serviço e das funcionalidades desejadas, neste caso, uma Lista de Usuários que apresente as seguintes informações sobre seus participantes: Nome, Função e *Status*.

Na segunda etapa do processo, realiza-se uma busca no repositório por uma LPS que consiga derivar um serviço de Lista de Usuário compatível com as informações descritas na etapa anterior. Como não há uma LPS que atente total, ou parcialmente, à especificação inicial, a busca não retorna nenhum conjunto de artefatos, o que indica que o caso de criação de um novo serviço deverá ser utilizado. É importante destacar que o processo prevê a etapa de busca, entretanto, não existe uma estratégia para tal. A princípio, a busca é realizada verificando a descrição de todos os serviços presentes no repositório manualmente. A especificação de uma estratégia de busca foge ao escopo deste trabalho.

Na terceira etapa do processo, realiza-se o planejamento do novo serviço de colaboração. Nesta atividade, foram descritos os requisitos desejados para a Lista de Usuários, são eles: (1) todos os usuários cadastrados devem ser listados; (2) deve-se informar qual a função ocupada por um usuário; (3) deve-se informar se um usuário está *online* ou *offline*, ou seja, seu status.

Na quarta etapa, cria-se o Modelo de *Features* do serviço e o mapeamento do modelo com a Ontologia de Colaboração. O modelo foi criado de acordo com o planejamento da etapa anterior, como ilustrado na Figura 2. Para criar o Modelo de

² https://github.com/pgcc/plscience-ecos/tree/master/web/files/Collaboration_Services

Features utilizou-se a ferramenta FeatureIDE³. Após a criação do modelo, realiza-se o seu mapeamento com a ontologia. Para analisar a Ontologia de Colaboração de modo a auxiliar a criação do arquivo de mapeamento, utilizou-se a ferramenta Protégè⁴.

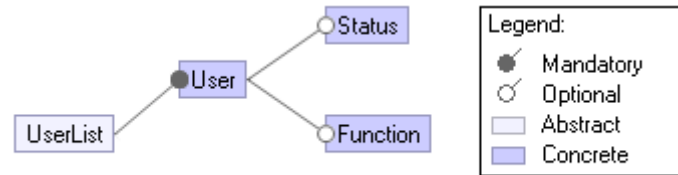


Figura 2. Modelo de *Features* do Serviço de Colaboração Lista de Usuários

Após a modelagem, na quinta etapa do processo, a Lista de Usuários foi desenvolvida baseando-se nos artefatos gerados anteriormente, principalmente do Modelo de Features, pois ele reflete a LPS da Lista de Usuários implementada. A lista foi desenvolvida na forma de um serviço como prevê o processo, e suas *features* foram implementadas como métodos do serviço, de forma a facilitar sua reutilização. Após desenvolver o serviço, criou-se um arquivo WSDL com sua descrição.

Na última etapa, todos os artefatos (incluindo o serviço) foram integrados ao ecossistema. Para isso, criou-se um local com nome “*User List*” no repositório de LPS no GitHub, onde foram adicionados todos os artefatos gerados nas etapas do processo, incluindo o código e descrição do serviço. Com isso, o E-SECO foi evoluído e a Lista de Usuários desenvolvida poderá ser reutilizada, assim como seus artefatos. A utilização por uma instância do E-SECO do serviço de Lista de Usuários derivado da LPS desenvolvida através do processo está destacada na Figura 3.

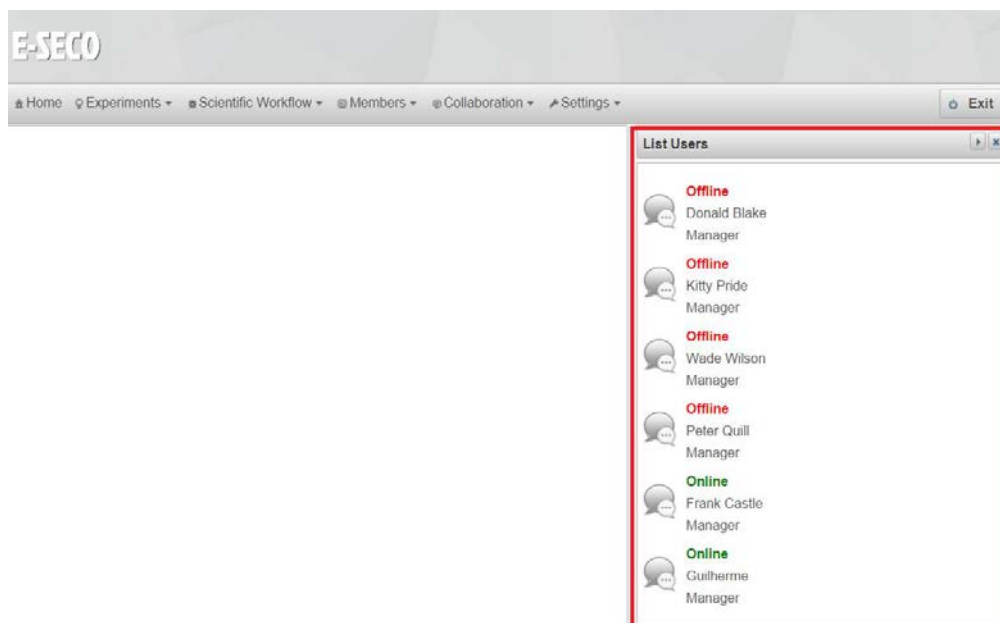


Figura 3. Serviço de Lista de Usuários Utilizado em uma Instância do E-SECO

Como pode ser observado na descrição da utilização do processo, ele se mostrou útil para padronizar a criação de serviços de colaboração, assim como para organizar os

³ http://www.witi.cs.uni-magdeburg.de/iti_db/research/featureide/

⁴ <http://protege.stanford.edu/>

artefatos gerados em cada etapa de modo a facilitar sua utilização, reutilização e evolução. Dessa maneira, as evidências geradas apontam para a viabilidade do processo. A seguir a avaliação do caso de extensão do serviço de colaboração será apresentado.

5. Avaliação do Processo: Caso de Extensão

Para avaliar o processo proposto quando ao caso de extensão, realizou-se um experimento formalmente definido de modo que os participantes aplicassem o processo para evoluir um serviço de colaboração do E-SECO. Primeiramente foi definido o escopo do experimento utilizando o GQM [Basili *et al.*, 2001]: “**Analisar o processo proposto para evoluir a infraestrutura do ecossistema com o propósito de avaliar seu suporte ao desenvolvimento de serviços de colaboração com respeito a sua usabilidade, compreensibilidade e completude do ponto de vista de desenvolvedores no contexto do desenvolvimento de serviços para o E-SECO.**”.

A partir do escopo foram definidas quatro hipóteses alternativas. Espera-se, através da avaliação destas, apresentar evidências sobre a utilidade, compreensibilidade, usabilidade e completude do processo. São elas:

- H0: O processo proposto auxilia na extensão de serviços de colaboração do E-SECO.
- H1: O processo apresenta todas as informações necessárias para a execução do mesmo.
- H2: Todas as informações presentes no processo são de fácil compreensão.
- H3: O processo pode ser utilizado sem a necessidade treinamento prévio.

O experimento consiste na aplicação do caso de extensão do processo por um participante voluntário com o intuito de evoluir a infraestrutura do E-SECO. Em específico, foi dado ao participante a tarefa de evoluir a LPS do serviço de Lista de Usuário existente no ecossistema, com isso, o participante deveria seguir as etapas apresentadas no processo, de modo a gerar os artefatos neles descritos. Essa evolução consistia na adição da possibilidade de acrescentar os usuários das Lista de Usuários em grupos. Para simular um contexto mais realista e avaliar a usabilidade do processo, não foi dado um treinamento prévio para o uso do processo.

A seleção do participante foi criteriosa, isso porque o processo foi criado visando apoiar a desenvolvedores que já possuem um certo conhecimento prévio, fator que teve de ser levado em conta na seleção de indivíduos, de forma a não prejudicar os resultados do estudo. O processo é direcionado a desenvolvedores com bom conhecimento em desenvolvimento com Java *Web*, SOA e LPS, além de conhecimento sobre edição e compreensão de Modelos de *Features*, e conhecimento básico em ontologias. Para caracterizar os participantes, utilizou-se um questionário⁵, onde foi selecionado apenas o participante que possuía o conhecimento previsto para o experimento.

A coleta de dados foi feita por meio de observação direta e por um questionário de avaliação⁶. No questionário encontravam-se perguntas fechadas referentes a cada atividade e duas perguntas abertas referentes ao processo como um todo. A análise dos dados coletados é apresentada na seção 5.2.

⁵ <https://drive.google.com/open?id=0B1uZYT7TKziBTWYxbmtlalhkNEk>

⁶ <https://drive.google.com/open?id=0B1uZYT7TKziBZWFaRmkxZ3ROQ3M>

5.1. Execução do Experimento

Este estudo considerou que as informações contidas na descrição do objetivo do experimento podem ser consideradas equivalentes às informações necessárias na primeira etapa do processo, a etapa do Planejamento Inicial. Portanto, na primeira atividade do estudo, o participante deveria buscar no repositório um serviço que estivesse de acordo com a descrição da tarefa. Esperava-se que o participante encontrasse o serviço de lista de usuários já disponível no ecossistema.

Como esperado, o participante conseguiu encontrar e recuperar as informações referentes a Lista de Usuários existente no ecossistema. Durante a execução da atividade foi possível constatar que o participante encontrou o serviço no repositório pelo seu nome (*User_List*), e baseado no Modelo de *Features* presente no repositório, ele identificou que o serviço se aproximava bastante do que ele precisava. Como o participante encontrou a LPS do serviço de Lista de Usuários disponível no ecossistema, ele seguiu o caso de extensão do processo proposto. É importante ressaltar que o serviço encontrado é o serviço que foi desenvolvido na seção anterior.

De posse dos artefatos referente à Lista de Usuários recuperada do repositório, a primeira ação do participante foi observar o Modelo de *Features* do serviço, posteriormente, ele analisou as descrições das *features* presentes no modelo e as comparou com os requisitos definidos para o serviço. Seguindo a descrição da etapa três do processo, o participante deveria criar o planejamento da extensão, para isso, ele estendeu as descrições existentes de modo a acrescentar o novo requisito de criação de grupos. O participante não sentiu a necessidade de criar nenhum documento adicional e avançou para a etapa seguinte.

Após o planejamento, o participante começou a quarta etapa do processo, a etapa de modelagem. Nessa atividade, o participante deveria atualizar Modelo de *Features* e o arquivo de mapeamento entre o modelo e a Ontologia de Colaboração. Para atualizar o modelo, ele utilizou o FeatureIDE e para verificar a ontologia para auxiliar a atualização do arquivo de mapeamento, ele utilizou o Protégè. Durante a atualização do modelo, o participante utilizou parte do seu tempo para analisar o modelo existente. Acredita-se que esse comportamento foi influenciado pelo fato de que, o processo enfatiza a consistência dos artefatos já existentes. Diante disso, é necessário observar as relações da nova *feature* com as já presentes no modelo. A Figura 4 apresenta o Modelo de *Features* atualizado.

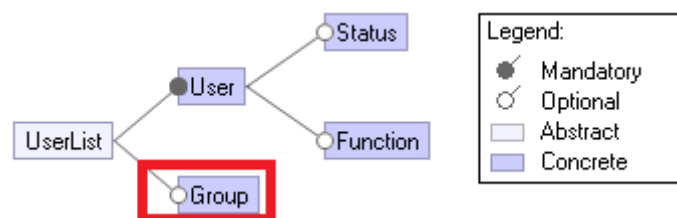


Figura 4. Modelo de *Features* do Serviço de Colaboração Lista de Usuários Atualizado

Na etapa de desenvolvimento, o participante deveria implementar a *feature* de grupo para a Lista de Usuários. Neste momento, ele deveria atentar-se aos princípios de SOA, de forma que a Lista de Usuários continuasse funcionando como um serviço após a implementação. Durante esta etapa, o participante estudou o código existente a fim de

entender seu funcionamento, para posteriormente, desenvolver a *feature* de grupo. Essa atividade tomou muito tempo, uma vez que não há nenhum elemento no processo que auxilie o desenvolvedor a nível de código, ou seja, essa etapa depende fortemente dos conhecimentos e capacidades do participante. O participante conseguiu desenvolver a nova *feature* na forma de um novo método do serviço de Lista de Usuário e, em seguida, gerou a nova descrição (arquivo WSDL) para o serviço. Para testar a nova funcionalidade do serviço de Lista de Usuários, um serviço contendo a *feature* Grupo foi derivada da LPS e adicionada a uma instância local do E-SECO (Figura 5).

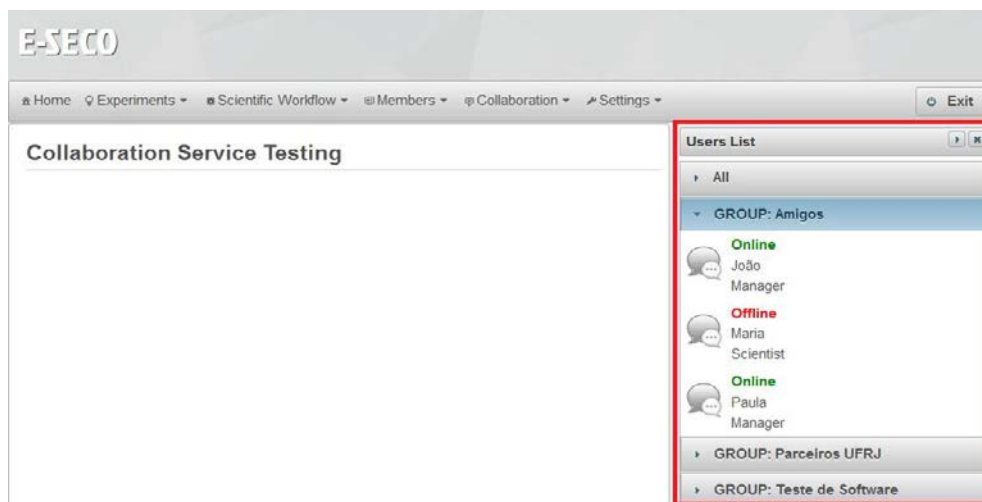


Figura 5. Serviço de Lista de Usuários com a *Feature* Grupo Utilizado em uma Instância do E-SECO

A última etapa do processo é a integração, nela o participante teria que atualizar o repositório da LPS da Lista de Usuários com os artefatos gerados nas etapas anteriores, disponibilizando assim o serviço no ecossistema. Para isso, o participante conferiu e reuniu todos os artefatos gerados durante a estudo, localizou a área do repositório no GitHub, referente a Lista de Usuários, e realizou o *upload* dos artefatos. Nessa atividade o participante não apresentou nenhuma dificuldade. Após a integração dos artefatos, o participante respondeu ao questionário. A seguir, a análise do resultado será apresentada.

5.2. Análise do Resultado

Para analisar os dados obtidos durante a realização do experimento, as respostas dos questionários foram fornecidas por meio de uma escala Likert de cinco níveis, contendo um nível intermediário como proposto por [Laitenberger and Dreyer, 1998]. Para cada resposta presente nas questões fechadas, foi atribuída uma pontuação: 1 para “discordo fortemente”, 2 para “discordo”, 3 para “não consigo e/ou não desejo responder”, 4 para “concordo” e 5 para “concordo fortemente”. Para cada etapa executada pelo participante, são apresentadas três questões, uma referente a utilidade do processo, uma sobre sua compreensibilidade e uma sobre sua completude. A Figura 6 apresenta um resumo dos resultados de cada etapa. É importante lembrar que a primeira etapa, referente ao planejamento inicial, não foi executada pelo participante, como discutido anteriormente.

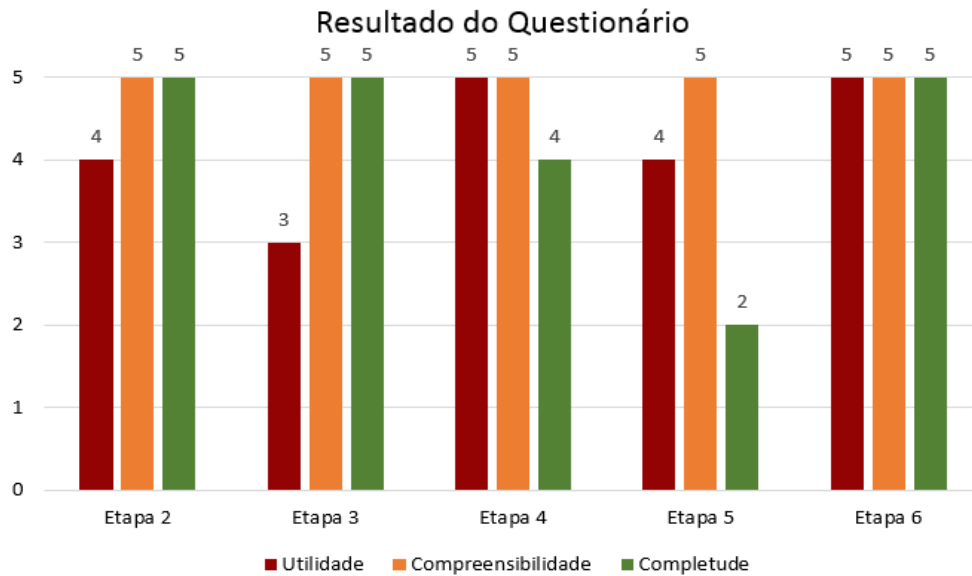


Figura 6. Resultado das Etapas do Questionário

Como pode ser observado na Figura 6, os resultados da etapa 2 indicam que as informações contidas no processo foram úteis para a realização da busca, foram de fácil compreensão e estavam completas. Nesta etapa, a observação direta indicou que as informações fornecidas pelo estudo foram suficientes para o participante realizar a busca, além disso, a complexidade da busca foi baixa, uma vez que existiam poucos serviços no repositório. Os resultados da etapa 3 indicam que as informações para a atualização do planejamento contidas no processo foram de fácil compreensão e estavam completas, contudo, o participante não concordou nem discordou que a descrição desta etapa foi útil. Nesta etapa, a observação direta indicou que o participante considerou esta etapa simples, pois se tratava de uma evolução simples de uma LPS já existente, e por não necessitar de nenhum documento adicional. Os resultados da etapa 4 indicam que as informações para a atualização do Modelo de *Features* e do mapeamento foram úteis, estavam compreensíveis e completas. Nesta etapa, a observação direta indicou que a descrição dos artefatos, a descrição dos cuidados que deveriam ser tomados e a indicação de ferramentas para executar as atividades foram muito importantes para o participante. Os resultados da etapa 5, onde o desenvolvimento é realizado, se mostrou útil e compreensível, entretanto, o participante discordou que a descrição desta etapa estava completa. Nesta etapa, a observação direta indicou que, apesar de descrever adequadamente os artefatos envolvidos, o processo não oferece suporte ao desenvolvedor a um nível de código, com isso, a atividade de implementação foi considerada bastante trabalhosa. Por fim, os resultados da etapa 6, responsável pela integração dos artefatos atualizados ao repositório da LPS, se mostrou útil, compreensível e completo. Nesta etapa, a observação direta indicou que o participante já estava familiarizado com a utilização do GitHub, o que facilitou esta atividade.

Além das questões fechadas, o participante respondeu a duas questões abertas. A primeira pergunta foi: “Como você descreve o suporte oferecido pelo processo em relação a realização do experimento?”. Nesta questão, o participante indicou que o processo foi de grande auxílio durante o experimento, enfatizando que provavelmente não teria conseguido completar a tarefa sem o auxílio do mesmo. A segunda pergunta

foi: “Você sentiu falta de um treinamento para sobre a utilização do processo?”. Nesta questão, o participante indicou que não houve necessidade de um treinamento prévio sobre o processo.

As informações analisadas apresentam evidências que indicam que as hipóteses levantadas podem ser confirmadas, apontando para a viabilidade do processo quanto ao caso de extensão. Com isso, pode-se dizer que o processo auxiliou a extensão da LPS do serviço de Lista de Usuários e que atendeu aos requisitos de usabilidade, compreensibilidade e completude, o que responde às questões de pesquisa apresentadas.

5.3. Ameaças à Validade

É preciso considerar as ameaças à validade presente neste experimento. Quanto à validade externa, (i) o processo está limitado ao contexto do E-SECO, mesmo dentro do contexto de *e-Science*, ecossistemas podem apresentar muitas diferenças, como, por exemplo, não utilizarem uma LPS; além disso, (ii) o processo se limita à criação e extensão de serviços de colaboração. Quanto à validade de resultados, o estudo não foi realizado com uma quantidade suficiente de participantes para se obter uma análise estatisticamente relevante, o que pretende-se realizar como um trabalho futuro.

6. Considerações Finais

Com advento da *e-Science*, a experimentação científica alcançou novas escalas, as equipes de cientistas passaram a estar geograficamente distribuídas, novas tecnologias foram inseridas e, com isso, a necessidade de sistemas que apoiassem a realização de experimentos e, principalmente, a colaboração nesse cenário. Pensando nisso, foi desenvolvido o E-SECO, um ecossistema de software científico (ECOSC) que, por meio de uma combinação de serviços, oferece mecanismos de apoio a colaboração e à realização de experimentos científicos colaborativos.

Para evoluir o E-SECO e criar serviços de colaboração que atendam às necessidades das equipes de cientistas, é necessário considerar os princípios e tecnologias existentes no ecossistema. Diante disso, esse trabalho propôs um processo de desenvolvimento de serviços de colaboração para o E-SECO. Esse processo considera os conceitos de LPS de forma a possibilitar a variabilidade e a reutilização dos serviços. Duas atividades são descritas no processo: (i) a criação de novos serviços de colaboração e a (ii) extensão de serviços de colaboração já existentes. A primeira auxilia o desenvolvedor a criar novos serviços, enquanto a segunda auxilia o desenvolvedor a estender um serviço existente. Em ambos os casos, ao final do processo o novo serviço (ou extensão do serviço) é disponibilizado no ecossistema por meio do repositório de LPS.

Para avaliar o processo proposto foi desenvolvido uma Lista de Usuários seguindo o caso de criação do processo, e esse mesmo serviço foi estendido em um estudo experimental seguindo o caso de extensão do processo. Os dados coletados e analisados durante o estudo experimental apresentaram evidências de que o processo proposto atende aos requisitos de usabilidade, compreensibilidade e completude desejados. Entretanto, o estudo carece de resultados com relevância estatística, o que faz necessária a realização de mais experimentos e em uma escala maior. Espera-se também que o processo possa ser refinado conforme mais estudos sejam realizados.

Como trabalhos futuros em relação ao processo cabe citar: (i) realizar um estudo experimental sobre a criação de um novo serviço de colaboração, e não apenas sobre a extensão; (ii) avaliar a possibilidade de sua aplicação em outros domínios; (iii) criação de mecanismos de busca e gerência dos serviços desenvolvidos e disponibilizados no ecossistema, o que implica na definição de uma especificação inicial formal; (iv) evoluir o processo para que ele considere a composição de serviços de colaboração a partir dos já existentes. Cabe citar também trabalhos referentes aos serviços de colaboração desenvolvidos, como: (i) avaliar como os serviços de colaboração desenvolvidos utilizando o processo proposto, oferecem suporte a colaboração; (ii) avaliar a capacidade de interoperabilidade entre os serviços desenvolvidos utilizando o processo.

Agradecimentos

Aos participantes do Núcleo de Pesquisa em Engenharia do Conhecimento (NEnC). Este projeto é parcialmente financiado pela CAPES, CNPq e FAPEMIG. Guilherme G. Martins é apoiado pela CAPES.

Referências

- Basili, V. R., Caldiera, G. and Rombach, H. D. (1994) “The goal question metric approach”, Software Architecture. Working IEEE/IFIP Conference.
- Belloum, A., Inda, M. A., Vasunin, D., Korkhov, V., Zhao, Z., Rauwerda, H., Breit, T. M., Bubak, M. and Hertzberger, L. O. (2011) “Collaborative e-Science Experiments and Scientific Workflows”. *Internet Computing, IEEE*, v.15, n.4, pages 39-47.
- Costa, G. C. B., Braga, R., David, J. M. N. and Campos, F. (2015) “A scientific software product line for the bioinformatics domain”. *Journal of Biomedical Informatics*, v.56, n.C, pages 239-264.
- Freitas, V., David, J. M. N., Braga, R. and Campos, F. (2015) “Uma Arquitetura para Ecossistema de Software Científico”, In: IX Workshop em Desenvolvimento Distribuído de Software, Ecossistemas de Software e Sistemas-de-Sistemas (WDES 2015), CBSOft, Belo Horizonte, MG, pages 41-48.
- Fuks, H., Raposo, A. B. and Gerosa, M. A. (2003) “Do Modelo de Colaboração 3C à Engenharia de Groupware” In: WEBMIDIA 2003, Simpósio Brasileiro de Sistemas Multimídia e Web, Trilha especial de Trabalho Colaborativo Assistido por Computador, Salvador, BA, pages 445-452.
- Gadelha, B., Nunes, I., Fuks, H. and Lucena, C. J. P. (2009) “An approach for developing groupware product lines based on the 3C collaboration model”, In: 15th international conference on Groupware, Berlin, Germany, pages 328-343.
- Gimenes, I. M. S. and Travassos, G. H. (2002), “O Enfoque de Linha de Produto para Desenvolvimento de Software”, In: XXI Jornada de Atualização em Informática (JAI), Evento Integrante do XXII Congresso da SBC, pages 1-31.
- Gurp, J. van, Bosch, J. and Svahnberg, M. (2001) “On the notion of variability in software product lines”, In: Software Architecture, 2001. Proceedings. Working IEEE/IFIP Conference, pages 45-54.
- Kang, K. C., Lee J. and Donohoe, P. (2002) Feature-Oriented Product Line Engineering. In *IEEE Software*, v. 19, n. 4, pages 58-65.

- Laitenberger, O. and Dreyer, H.M. (1998). “Evaluating the usefulness and the ease of use of a web-based inspection data collection tool”. In: Software Metrics Symposium, pages 122-132.
- Martins, G. G., Neiva, F. W., David, J. M. N., Braga, R. M. and Campos, F. C. A. (2015) “Colaboração no Ciclo de Vida de um Experimento Científico no Contexto de um Ecossistema de Software”, In: XII Simpósio Brasileiro de Sistemas Colaborativos, Salvador, BA, pages 114-121.
- Mattoso, M., Werner, C., Travassos, G. H., Braganholo, V. and Murta, L. (2008) “Gerenciando experimentos científicos em larga escala”, In: Seminário Integrado de Software e Hardware, Belém. SEMISH, Congresso da SBC, pages 121-135.
- Miranda, G. C. V., Souza, J. A. B. V., Braganholo, V. and Oliveira, D. (2014) “CollabCumulus: Uma Ferramenta de Apoio à Análise Colaborativa de Proveniência em Workflows Científicos”, In: XI Simpósio Brasileiro de Sistemas Colaborativos, Curitiba, PR, pages 94-101.
- Neiva, F. W., David, J. M. N., Braga, R. M., Campos, F. C. A. and Freitas, V. (2015) “PRIME: Pragmatic Interoperability Architecture to Support Collaborative Development of Scientific Workflows”, In: IX Brazilian Symposium on Components, Architectures and Reuse Software, Belo Horizonte, MG, pages 50-59.
- Zhang, J., Kuc, D. and Lu, S. (2014) Confucius: A Tool Supporting Collaborative Scientific Workflow Composition. In *IEEE Transactions on Services Computing*, v. 7, n. 1. pages 2-17.

Uma Revisão Sistemática da Literatura para Investigação de Estratégias de Ensino Colaborativo

Simone Erbs da Costa¹, Edilson Hipólito da Silva², Leandro Otávio Córdova Vieira¹,
Carla Diacui M. Berkenbrock¹

¹Universidade do Estado de Santa Catarina (UDESC)
Joinville – SC – Brasil

²Instituto Federal de Santa Catarina (IFSC)
Jaraguá do Sul – SC – Brasil

si.gen@terra.com.br, edilsonhipolito@gmail.com,

leandro@whap.com.br, carla.berkenbrock@udesc.br

Abstract. *This paper presents a systematic review investigating the state of the art of collaborative learning between teachers. The searches were conducted in the digital libraries of the ACM and IEEE Xplore in the selected period of 2009-2015 and were reviewed 591 papers. Here we highlight the results obtained and the methodologies used in the evaluations. The results demonstrate that the strategies in collaborative learning among teachers are still not considered aspects of educational cooperation, restricting the collaboration factors geared to the needs of students or the construction of teaching materials, which are managed individually and not so collaborative. We also try to check trends, strategies and limitations related to collaborative learning.*

Resumo. *Este artigo apresenta uma revisão sistemática que investiga o estado da arte do ensino colaborativo entre professores. As buscas foram realizadas nas bibliotecas digitais da ACM e IEEE Xplore. Foram revisados 591 artigos no período selecionado de 2009 a 2015. Neste artigo são destacados os resultados obtidos, bem como as metodologias utilizadas nas avaliações. Os dados demonstram que nas estratégias no ensino colaborativo entre professores ainda não são considerados os aspectos de colaboração de ensino, restringindo-se aos fatores de colaboração voltados às necessidades dos alunos ou na construção de material didático, os quais são geridos individualmente e não de forma colaborativa. A pesquisa também visa verificar tendências, estratégias e limitações por meio dos resultados identificados.*

1. Introdução

O desenvolvimento das atividades relacionadas com a aprendizagem colaborativa é frequentemente defendido no meio acadêmico. Uma das razões para essa conduta é o seu potencial de promover uma aprendizagem mais ativa por meio do estímulo ao pensamento crítico [Torres 2015]. O desdobramento de capacidades de interação, negociação de informações e resolução de problemas, também são fatores relevantes.

Métodos de aprendizagem colaborativa podem tornar os alunos mais responsáveis por sua formação, conduzindo a uma aquisição do conhecimento com maior autonomia

[Torres 2015]. Com isso, o processo de ensino-aprendizagem não está mais centrado na figura do professor e o aluno exerce um novo papel, onde o professor cria contextos e ambientes adequados para que o aluno possa desenvolver suas habilidades sociais e cognitivas de modo criativo, na interação em sala.

Neste sentido destaca-se a importância do compartilhamento de experiências entre professores. Esse compartilhamento pode favorecer o desenvolvimento da destreza na análise crítica, na resolução de problemas e na tomada de decisões [Nono and MIZUKAMI 2001].

O trabalho colaborativo multidisciplinar pode enriquecer a prática pedagógica, pois proporciona uma visão mais abrangente sobre as dificuldades enfrentadas e envolve um maior número de pessoas responsáveis pelo processo de escolarização dos alunos [Rabelo 2012]. [Pugach and Johnson 1989] salientam a importância de se realizar parcerias colaborativas na educação. Segundo [Damiani 2008], este cenário de colaboração não tem acontecido e os professores continuam trabalhando individualmente, indicando que esta tendência parece não ter mudado.

Alguns autores [Engeström 1994, Fullan et al. 2000, Thurler and Wolff 2001] descrevem a profissão docente como solitária. Já [Fullan et al. 2000] sugerem que o isolamento docente tem raízes em fatores como a arquitetura das escolas, a estrutura dos seus horários, a sobrecarga de trabalho e a própria história da profissão docente. Segundo [Pugach and Johnson 1989], o maior problema no trabalho colaborativo entre os professores são os mesmos abandonarem sua posição de perito, e compartilhar seu planejamento, medos, necessidades e potenciais com outros professores. Para [Gately and Gately 2001], o ideal seria que os professores pudessem compartilhar seu trabalho educativo, sem hierarquias, onde as responsabilidades seriam de todos os atuantes com aqueles educandos, visando o compartilhamento de planejamentos, as intervenções, monitorias, avaliações e replanejamento.

Neste trabalho foi realizado uma Revisão Sistemática da Literatura (RSL) afim de identificar o estado da arte acerca do ensino colaborativo. O estudo considerou trabalhos em inglês publicados nas bibliotecas digitais da ACM e IEEE Xplore, no período selecionado de 2009 a 2015. A revisão busca verificar iniciativas onde os professores colaboram entre si na troca ou na melhoria de experiências de ensino-aprendizagem.

O restante do trabalho está organizado da seguinte forma: A seção 2 descreve os trabalhos correlatos. A seção 3 descreve a revisão sistemática da literatura realizada. Já as seções 4 e 5 descrevem, respectivamente, a análise dos resultados obtidos por meio da revisão e as discussões sobre esses resultados. Finalmente, a seção 6 descreve as conclusões obtidas até o momento, bem como aponta os próximos passos desta pesquisa.

2. Trabalhos correlatos

Esta seção apresenta alguns trabalhos diretamente relacionados com a pesquisa descrita no presente artigo.

[Sa'don et al. 2014] observam a propagação de plataformas *Massive Open Online Courseware* (MOOC). De acordo com os autores, apesar do crescimento da literatura relacionada com MOOC, há poucos trabalhos dedicados a revisar e avaliar sistematicamente esse conteúdo produzido. Via mineração de dados, os autores encontraram mais

de 400 artigos publicados na área. Avaliaram e sintetizaram esse material, e identificaram a possibilidade de mapeamento de tendências de pesquisa, estratégias e limitações para a utilização destas plataformas.

Segundo [Sa'don et al. 2016], as iniciativas inovadoras que incentivam a disseminação do conhecimento precisam de uma estrutura de negócio sólida para a sustentabilidade dos cursos como MOOC. A proposta impacta as partes interessadas, as instituições renomadas e os usuários finais ao acesso de educação de qualidade.

[Lin and Lin 2010], ao notarem as reformas educacionais e curriculares em Taiwan, bem como a diversidade de condições de absorção de tais reformas por parte dos docentes, experimentaram uma metodologia colaborativa (*E-Teacher Community Competition* - ETCC) por meio da formação de equipes entre professores. A plataforma foi desenvolvida pela *Educational of Hualien County* (EBHC). Os autores estimularam a produção de conteúdos colaborativos entre os membros. Ao longo de 13 semanas, os autores perceberam que o ETCC se tornou uma ferramenta de incentivo e também um novo modelo para a formação de professores.

Algumas estratégias foram adotadas para o sucesso do programa, como: o site ETCC, projetado e desenvolvido pelo EBHC, serviu como uma plataforma para acelerar o aprendizado colaborativo entre a comunidade e-professor; todos os conteúdos dos e-professor foram mantidos, dessa forma, esses conteúdos podiam ser consultados ainda que após a competição, servindo como modelo para outros professores de como construir um ensino colaborativo; e a premiação para equipe vencedora.

[Lin et al. 2011], analisaram o processo após 3 anos do programa ETCC. Segundo os autores, a vantagem é que cada professor do time e-professor, pode promover suas habilidades profissionais, se tornar um aprendiz ao longo da vida e transformar toda a comunidade e-professor. Ainda de acordo com os autores, para que exista um ensino colaborativo é preciso se ter uma base robusta, investidores públicos ou privados que garantam a sustentabilidade do programa; e apoio ao corpo docente e discente.

3. Revisão Sistemática da Literatura

Segundo [Mafra and Travassos 2005], uma revisão sistemática da literatura deve seguir uma metodologia rigorosa, confiável e passível de auditoria. Além disso, a revisão deve ter bem definidos os protocolos de busca afim de possibilitar que a pesquisa seja repetida por outros pesquisadores. Nos trabalhos de [Keele 2007] e [Brereton et al. 2007], os autores apresentam uma sequência de passos bem definidos de como realizar tal revisão, também utilizados para desenvolvimento deste trabalho.

Na concepção de [Petticrew and Roberts 2008], a fim de se delimitar o escopo da pesquisa e auxiliar no desenvolvimento das suas questões, e da *string* de busca, optou-se por utilizar a estrutura PICOC que trata em responder mais de uma pergunta na mesma metodologia.

De acordo com os autores [Keele 2007] e [Petticrew and Roberts 2008], a estratégia PICOC trata dos critérios: População (*Population*): as pessoas ou grupo que se quer estudar; Intervenção (*Intervention*): o que se está interessado em rever; Comparação (*Comparison*): a intervenção é comparada com o que; Saída (*Outcomes*): em muitas intervenções sociais existe um conjunto grande de resultados e com isso a avaliação en-

volve alguns critérios; e Contexto (*Context*): quando a metodologia tem mais de uma pergunta na mesma avaliação. Nesta pesquisa não será utilizado o critério saída da estratégia PICOC.

Para [Keele 2007], a estratégia de busca deve permitir que a pesquisa seja repetida por outros pesquisadores. Para que isto seja possível, deve-se divulgar a *string* de busca utilizada, assim como levar em consideração as diferentes bases de dados analisadas. Segundo [Kitchenham et al. 2006], o elemento essencial ao se fazer uma revisão sistemática é fazer a *string* de busca, pois ela constitui a cadeia de pesquisa automática, determina os dados extraídos e restringe a busca. Outro fator importante é que a *string* de busca faz parte do protocolo e a partir do momento que o protocolo é aceito e validado, a *string* de busca não deve ser alterada.

3.1. Questão da Pesquisa

A questão de pesquisa se refere a colaboração do ensino entre professores e foca os meios colaborativos que os professores utilizam. Porém, além da questão da pesquisa principal, uma questão secundária visa investigar se tal colaboração entre os professores conta com o auxílio de software.

Questão principal (QP): Qual o papel para melhorar a colaboração entre professores no desenvolvimento de suas atividades?

Questão secundária (QS): Quais sistemas são adotados pelos professores para fomentar a colaboração com seus pares durante o desenvolvimento de suas atividades?

3.2. Processo da Pesquisa

A pesquisa foi realizada no período selecionado compreendido entre 2009-2015 para buscar o estado da arte na temática do ensino colaborativo entre professores. As bibliotecas digitais escolhidas foram *ACM* e *IEEEExplore*. Essa escolha é justificada por essas bibliotecas possuírem publicações de artigos e revistas; bem como possuem publicações relacionadas com a área de sistemas colaborativos e serem bases em ciências da computação.

Se buscou o estado da arte da temática seguindo a revisão sistemática e utilizando a *string* de busca: (“*collaborative learning*” OR “*collaborative teaching*”) AND (*teacher* OR *instructor* OR *educator* OR *professor*) and (“*didactic material*” OR “*educational activity*” OR “*pedagogical activity*” OR *courseware*).

Foram identificados critérios de inclusão e exclusão dos artigos relevantes ao tema de pesquisa aplicado a *string* de busca definida, afim de selecionar os trabalhos que tenham relevância para a pesquisa. Os resultados das buscas foram revisados por três pesquisadores envolvidos, e as discordâncias foram discutidas e resolvidas. Em casos de falta de consenso sobre o trabalho estudado o mesmo é incluído.

Os critérios de inclusão adotados neste trabalho foram: estudos que definem colaboração entre professores para o desenvolvimento de suas atividades, bem como estudos que demonstrem o uso de sistemas para colaboração entre professores nas atividades. Os critérios de exclusão adotados neste trabalho foram: estudos onde a cooperação ou colaboração não foram avaliadas, estudos repetidos ou duplicados foi considerado apenas o mais recente, estudos que apresentem texto, conteúdo e resultados incompletos, estudos pagos e estudos em *workshops*, resumos, cartazes, painéis, palestras e demonstrações.

A metodologia utilizada para selecionar os artigos foi realizada seguindo a metodologia de [Santos et al. 2012] tanto para os critérios de inclusão/exclusão como para os critérios de qualidade que constam na Tabela 1.

Foram aplicados os critérios de inclusão/exclusão em cada um dos passos para se manter o maior número de artigos relevantes no estado da arte da temática. [Santos et al. 2012] colocam quatro passos a serem seguidos da metodologia, são eles: Passo 1: Eliminação por Título; Passo 2: Eliminação por resumo; Passo 3: Eliminação por leitura diagonal; e Passo 4: Eliminação por leitura completa.

O escopo para aplicação desta RSL, diz respeito a métodos de colaboração entre professores, sejam eles no planejamento de suas aulas, na criação e compartilhamento de atividades, métodos de ensino ou experiências. Para que o estado da arte fosse delimitado aos métodos de avaliação para sistemas colaborativos optou-se em utilizar a metodologia da RSL de [Santos et al. 2012]. Nesta metodologia os autores utilizam 12 anos para se fazer as buscas do estado da arte da temática escolhida. Contudo, a estratégia desta pesquisa se baseia nos períodos de 2009-2015.

[Santos et al. 2012] mencionam que existe muito a ser estudado e validado em sistemas colaborativos; assim como muita coisa está surgindo. Outro ponto abordado pelos autores é a importância de se colocar critérios de qualidade para o método escolhido e seu respectivo peso. Os critérios de qualidade escolhidos para selecionar os artigos relacionados que demonstrem o estado da arte da temática desta pesquisa podem ser visualizados na Tabela 1. Esses critérios serão aplicados no passo 4, leitura completa dos artigos, e após tal seleção eles se encontram na etapa IV da Tabela 2 e de forma detalhada na Tabela 3. Esses critérios são adaptados da RSL de [Santos et al. 2012], assim como os pesos aplicados para seleção dos artigos.

Tabela 1. Critérios de Qualidade adaptados de [Santos et al. 2012]

Critérios de Qualidade	Peso
O estudo define claramente o objetivo da pesquisa (define questão de pesquisa)?	1
O artigo responde às questões de pesquisa definidas?	1
O artigo relata os passos do método que foram seguidos para a avaliação?	2
O artigo cita o contexto do sistema colaborativo em análise?	2
O artigo apresenta a motivação para a escolha do método?	1
Apresenta dificuldades para avaliar o sistema colaborativo?	2
O artigo apresenta o custo/benefício de utilizar o método estudado?	2
Total	11

A Tabela 2 mostra os artigos classificados em cada uma das etapas da RSL. Os artigos foram selecionados por meio da revisão sistemática utilizando a *string* de buscas nas bibliotecas citadas e de acordo com os critérios de eliminação expostos, por ordem de relevância.

Na Etapa 1 os artigos foram eliminados pelo Passo 1: eliminação pelo título, os artigos eliminados na Etapa 2 se deram decorrentes ao Passo 2: eliminação pelo resumo, os artigos eliminados na Etapa 3 ocorreram pelo Passo 3: leitura diagonal. Já para os artigos que foram selecionados na Etapa 4 foram analisados o Passo 4: leitura completa.

Tabela 2. Artigos Classificados nas Etapas da RSL

Base de Dados	Analisados	Etapa I	Etapa II	Etapa III	Etapa IV
<i>ACM</i>	411	46	14	10	2
<i>IEEE Xplorer</i>	180	126	80	30	3
Total	591	172	94	40	5

Após concluir todos os passos foram selecionados, cinco artigos relacionados com o ensino colaborativo entre os professores, conforme mostra a Tabela 2 na Etapa IV. A Tabela 3 contempla os cinco artigos selecionados com alguns detalhes como: sua referência (autores), o contexto que o mesmo está inserido, o meio utilizado no ensino colaborativo entre os professores.

Tabela 3. Referências dos Artigos Selecionados

ID	Artigos Selecionados	Autor(es)	Contexto	Meio utilizado
1	A Keyword Recommendation Experiment to Support Information Organization and Folksonomies in Edu-AREA	Cairo-Rodríguez	Compartilhamento de atividades de ensino pela WEB	Software Colaborativo (Edu-Area)
2	A Proposal Model to Monitor Interdisciplinary Research Projects in Latin American Universities	Heredia and Vinuesa	Interdisciplinarietà (IL)	Web Ontology
3	Opening Lesson Plans to Support Teaching Innovation and Open Educational Resources Adoption	Rodríguez	Desenvolvimento e compartilhamento de atividades de ensino	Software Colaborativo (Edu-Area)
4	A Collaborative Environment for the Design of Accessible Educational Objects	Bocacci, Ribaud and Mesiti	Ambiente colaborativo para desenvolvimento de material de ensino	VisualPedia
5	Open educational resources in computer science teaching	Dichev and Dicheva	Recursos educacionais abertos	OER

4. Análise dos Resultados

Após a conclusão das etapas de seleção a qual os artigos foram submetidos, foram obtidos cinco resultados relacionados com o ensino colaborativo.

As avaliações descritas neste artigo têm o intuito de melhorar a colaboração entre professores durante a realização de suas atividades, bem como identificar os sistemas utilizados para que possam melhorar o processo de ensino-aprendizagem.

A Tabela 3 mostra os artigos selecionados no ensino colaborativo, seus autores, o contexto que está inserido e o(s) meio(s) utilizado(s). Os artigos classificados tratam da QP: “qual o papel para melhorar a colaboração entre professores no desenvolvimento de suas atividades?” e da QS: “quais sistemas são adotados pelos professores para fomentar a colaboração com seus pares durante o desenvolvimento de suas atividades?”. As questões específicas foram todas analisadas e respondidas na revisão dos artigos na etapa IV.

Nos artigos 1 e 3, [Caeiro-Rodríguez et al. 2015b] e [Rodríguez 2014] descrevem a ferramenta Edu-Area. Nela, os professores podem criar ou modificar recursos, atividades e planos de aula para o seu contexto de ensino. Além disso, os professores também podem relatar experiências de atividades desenvolvidas com os alunos, podendo anexar fotos, vídeos, etc. Outro fator interessante é que o Edu-Area é uma ferramenta web, e o material disponibilizado é considerado um recurso educacional aberto (OER), permitindo sua utilização, replicação e adaptação quando necessário. Assim, segundo os autores, a ferramenta Edu-Area traz um novo desenvolvimento, apoiando a operação dos usuários de forma aberta e na organização social.

No artigo 2, [Heredia and Vinueza 2015] tratam da colaboração que pode existir dentro de uma universidade, mais em específico dentro de um curso onde acontece a interdisciplinaridade que pode gerar possíveis inovações na área pesquisada.

No artigo 4, [Boccacci et al. 2009] discorrem acerca de uma ferramenta web com o foco no desenvolvimento de material de ensino de forma colaborativa. Por meio da ferramenta, além de colaborar com o conteúdo do material, os professores contribuem no desenvolvimento de materiais que atendam pessoas com todo tipo de deficiência, como por exemplo, pessoas com deficiência visual necessitam de materiais em forma de áudio, ou pessoas com deficiência auditiva que necessitam de materiais de forma ilustrativa, as imagens utilizadas são tratadas por um algoritmo matemático no *VisualPedia*; o que pode ser desenvolvido por mais de um professor. A colaboração é unida a necessidade de inclusão das pessoas com diversas formas de deficiência, obtido com a integração da MediaWiki.

No artigo 5, [Dicheva and Dichev 2014] relatam sobre os OERs. Eles afirmam que a busca por materiais de ensino é frequente, porém muitas vezes os conteúdos disponibilizados não atendem ou não se adequam a realidade do professor, não atendendo as expectativas. Segundo os autores, não é suficiente ter apenas uma grande quantidade de conteúdo, também são necessários critérios de filtro e de busca para que esses conteúdos possam ser facilmente encontrados, porém ainda é baixo o conhecimento existente sobre OER. A pesquisa revela os diferentes pontos de vista dos usuários e das soluções fornecidas. Contudo, conforme os autores, não adianta apenas disponibilizar a informação. Também é preciso apoiar o ensino nas diferentes atividades, nos critérios de seleção, na facilidade de encontrar o desejado e na sua adaptação.

Neste artigo foi realizada uma revisão sistemática com um método consolidado, que especificam, passo a passo o que deve ser realizado. Contudo, o resultado obtido ainda não foi satisfatório. Dos 591 artigos resultantes da busca pelas palavras chaves apenas cinco destes foram selecionados como pertinentes ao estado da arte da temática. Assim, o resultado encontrado nos indica que a busca deve ser realizada em mais bibliotecas digitais,

não somente as bases de ciências da computação. Os artigos em português também devem ser incluídos, bem como os critérios de exclusão devem ser revistos, como por exemplo, a exclusão por título.

5. Discussões

Após a análise dos resultados, vista na seção anterior, as pesquisas relacionadas com os 5 artigos selecionados por meio da RSL foram estendidas com o intuito de complementar o estudo sobre a colaboração durante o processo de ensino-aprendizagem.

Nos artigos 1 e 3, [Caeiro-Rodríguez et al. 2015b] e [Rodríguez 2014] relatam sobre a ferramenta Edu-Area. Conforme [Bergman et al. 2008], quando uma organização categoriza a informação, melhora-se o acesso a informação, sua pesquisa e a sua recuperação por futuros usuários. Não apenas quem categoriza tem a informação, mas toda a comunidade de usuários. Segundo [Caeiro-Rodríguez et al. 2015a], não houve uma adoção da ferramenta Edu-Área, pois houveram problemas e barreiras no processo de sua adoção. Os principais problemas encontrados foram decorrentes da divergência do apoio efetivamente realizado do que foi proposto; e as diferentes abordagens ao conceito de organização e usuários.

No artigo 2, [Heredia and Vinueza 2015] analisam a colaboração que pode existir dentro de uma universidade, especificamente em disciplinas multidisciplinares. Os autores [Khalili et al. 2007] em pesquisas anteriores, mencionam que as equipes multidisciplinares precisam de tempo para que os alunos possam colocar seu projeto em uso e o testar em toda sua extensão. Outro ponto é o incentivo para o corpo docente, avaliações trimestrais para que os ajustes do curso possam ocorrer durante o mesmo e não apenas no final das atividades. Outra consideração é que para os alunos contemplarem as diferenças culturais, sociais e econômicas precisam aprender uma segunda língua, tais como espanhol ou inglês. Segundo [Ktoridou and Eteokleous 2013], ao incluir alunos de diferentes disciplinas, os alunos são mais inclinados a terem uma perspectiva integrada e as soluções encontradas são oriundas dessa multidisciplinaridade. Os alunos que vem de disciplinas diferentes acabam beneficiando a própria aprendizagem interdisciplinar (IL). Uma vantagem educacional dos programas com abordagem IL é excluir o fracionamento e as habilidades isoladas, os alunos têm uma motivação e uma adesão no aprendizado. Já os professores nesse ambiente podem enriquecer seu ensino com o conhecimento adquirido nos diversos cursos.

No artigo 4, [Bocacci et al. 2009] apresentam uma ferramenta web com o foco no desenvolvimento de material de ensino de forma colaborativa. O material foi desenvolvido para atender todo tipo de pessoas, incluso as pessoas com deficiência, sejam elas: cognitivas, visuais e auditivas. O visualPedia, conforme [Valtolina et al. 2012], é um sistema de e-learning multimídia e audiovisual colaborativo. Ele tem por objetivo a inclusão de alunos com deficiência, bem como apoiar os professores na construção dos materiais utilizados de forma personalizada. O sistema possui um controle de perfis configurável de acordo com as necessidades dos alunos. O processo de concepção e desenvolvimento da ferramenta VisualPedia é analisado em 3 pontos da usabilidade e acessibilidade: avaliação de usabilidade semiótica, análise da acessibilidade técnica e experiência do usuário. De acordo com os autores, nessa abordagem consegue-se produzir sistemas mais utilizáveis e acessíveis; e como trabalhos futuros precisam corrigir os problemas de acessibilidade

e usabilidade ainda existentes, assim como encontrar novas estratégias de colaboração para construção dos materiais. A ideia para construir esses objetos educacionais é envolver os alunos no preparo desses materiais, pois ninguém mais do que eles sabem quais os métodos que mais se adequam as suas necessidades, principalmente aqueles que possuem alguma deficiência, resultando no aumento da comunidade; e explorando a inserção dos usuários no uso das ferramentas sociais. Conforme os autores é preciso conquistar o usuário para que ele traga suas habilidades das redes sociais e trabalhe de forma ativa no melhoramento da ferramenta.

No artigo 5, [Dicheva and Dichev 2014] abordam alguns recursos educacionais abertos (OERs). Os autores, em pesquisas anteriores [Dichev and Dicheva 2012], relatam que a forma como os recursos são produzidos, a sua reutilização e distribuição é que são novos. Em decorrência disso, a quantidade de material existente é razoável, o problema é que boa parte desse material não é utilizada. A principal barreira é a sua *findability*, ou seja, localizar onde está o repositório e o conteúdo dentro do repositório. Para tal, foi mudado a forma como são realizadas as pesquisas de busca, por meio da utilização de um vocabulário específico. Segundo [Dicheva and Dichev 2015], ainda não se tem conhecimento das experiências de OER e por quem eles são utilizados, dessa forma não se sabe o impacto real dos OERs na aprendizagem das disciplinas. Dessa forma, se busca saber quem utiliza e como os OERs são utilizados quando apoiados por uma interface de busca que possibilite a exploração dos mesmos. Os usuários usam o conteúdo, quando sabem da existência dos seus repositórios e de uma ferramenta que permita explorar esse conteúdo. Interfaces que possibilitem uma maior interação resulta num maior engajamento dos usuários. Os autores afirmam que não adianta disponibilizar os OERs, se não houver: divulgação do repositório e de seus objetos, um mapeamento de quem vai utilizá-los e apoio aos mesmos; e por último, uma ferramenta de exploração com uma interface amigável e voltada para as necessidades dos usuários.

6. Conclusões

Este artigo apresenta uma revisão sistemática da literatura (RSL) de estratégias de ensino colaborativo entre professores. A revisão considerou o período selecionado correspondido entre 2009 a 2015. Inicialmente, 591 artigos foram resultantes da busca. Contudo, após seguir os passos subsequentes da RSL, que se deram em quatro etapas, obteve-se cinco artigos que contemplam o estado da arte da temática.

Os primeiros resultados encontrados demonstram algumas iniciativas relacionadas com o ensino colaborativo. Contudo, a principal constatação é que ainda existe muito a ser explorado, tanto em ferramentas que facilitem a colaboração entre os professores, quanto em atividades de que ressaltem para esses profissionais as vantagens da colaboração.

A RSL se mostrou interessante como um ponto de partida para a pesquisa, contudo, acreditamos que é importante complementá-la com busca de trabalhos relacionados com os encontrados na RSL. Assim, a identificação das citações relevantes encontradas nos artigos de encontrados na RSL, bem como a pesquisa por trabalhos que citam tais artigos permite refinar o resultado da pesquisa.

Como trabalhos futuros pretende-se complementar a revisão sistemática da literatura expandindo as buscas para outras bibliotecas digitais, assim como revisar e aprimorar a *string* de busca, complementando com os novos termos encontrados nos trabalhos rela-

cionados. Além disso, é importante que a busca seja realizada em mais bibliotecas digitais, não somente as bases de ciências da computação, bem como incluir artigos em português.

Também pretendemos investigar como está funcionando a colaboração entre o professor titular da disciplina e o intérprete ou segundo professor, para os casos dos alunos com deficiências auditivas. Assim, desejamos investigar a colaboração desses profissionais que auxiliam o professor da disciplina na preparação dos materiais de ensino-aprendizagem.

Referências

- Bergman, O., Beyth-Marom, R., and Nachmias, R. (2008). The user-subjective approach to personal information management systems design: Evidence and implementations. *J. Am. Soc. Inf. Sci. Technol.*, 59(2):235–246.
- Boccacci, P., Ribaudó, M., and Mesiti, M. (2009). A collaborative environment for the design of accessible educational objects. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 03, WI-IAT '09*, pages 207–210, Washington, DC, USA. IEEE Computer Society.
- Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., and Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *J. Syst. Softw.*, 80(4):571–583.
- Caeiro-Rodríguez, M., Llamas-Nistal, M., Fernández-Iglesias, M., Mikic-Fonte, F., and Lama-Penín, M. (2015a). Supporting real open educational resources in edu-area: Different views about open educational resources. In *Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE*, pages 1–8.
- Caeiro-Rodríguez, M., Santos-Gago, J. M., Lama, M., and Llamas-Nistal, M. (2015b). A keyword recommendation experiment to support information organization and folksonomies in edu-area. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 10(2):60–68.
- Damiani, M. (2008). Entendendo o trabalho colaborativo em educação e revelando seus benefícios. *Educar em Revista*, 0(31).
- Dichev, C. and Dicheva, D. (2012). Open educational resources in computer science teaching. In *Proceedings of the 43rd ACM technical symposium on Computer Science Education*, pages 619–624. ACM.
- Dicheva, D. and Dichev, C. (2014). Finding open educational resources in computing. In *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*, pages 22–24. IEEE.
- Dicheva, D. and Dichev, C. (2015). Cs/it oer: Availability, discoverability and use. In *Proceedings of the 16th Annual Conference on Information Technology Education, SIGITE '15*, pages 133–138, New York, NY, USA. ACM.
- Engeström, Y. (1994). Teachers as collaborative thinkers: Activity-theoretical study of an innovative teacher team. *Teachers' minds and actions: Research on teachers' thinking and practice*, pages 43–61.

- Fullan, M., Hargreaves, A., and Garcez, R. (2000). *A escola como organização aprendente: buscando uma educação de qualidade*. Biblioteca Artmed : fundamentos da educação. Artmed.
- Gately, S. E. and Gately, F. J. (2001). Understanding coteaching components. *Teaching exceptional children*, 33(4):40–47.
- Heredia, R. M. and Vinueza, P. C. (2015). A proposal model to monitor interdisciplinary research projects in latin american universities. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 10(3):102–108.
- Keele, S. (2007). Guidelines for performing systematic literature reviews in software engineering. In *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*.
- Khalili, N. R., Jacobius, T. M., Acevedo, J., and Nadal, E. O. (2007). Education without borders: Development of an interdisciplinary project based course between iit and itesm.
- Kitchenham, B., Mendes, E., and Travassos, G. H. (2006). A systematic review of cross-vs. within-company cost estimation studies. In *Proceedings of the 10th international conference on Evaluation and Assessment in Software Engineering*, pages 81–90. British Computer Society.
- Ktoridou, D. and Eteokleous, N. (2013). Interdisciplinary web-based learning practices in higher education. In *Global Engineering Education Conference (EDUCON), 2013 IEEE*, pages 536–539.
- Lin, C.-D. and Lin, M.-J. (2010). Promoting elementary school teachers’ professional development by e-teacher community competition. In *Educational and Information Technology (ICEIT), 2010 International Conference on*, volume 1, pages V1–135–V1–139.
- Lin, C. D., Lin, M. J., and Zhang, X. S. (2011). Can we transform the e-teacher community to be a learning organization to promote school teachers’ professional development. In *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, pages 245–249.
- Mafra, S. N. and Travassos, G. H. (2005). Técnicas de leitura de software: Uma revisão sistemática. *XIX Simpósio Brasileiro de Engenharia de Software, SBES*, 5.
- Nono, M. A. and MIZUKAMI, M. d. G. (2001). Aprendendo a ensinar: futuras professoras das séries iniciais do ensino fundamental e casos de ensino. *Reunião anual da ANPED*, 24.
- Petticrew, M. and Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons.
- Pugach, M. C. and Johnson, L. J. (1989). The challenge of implementing collaboration between general and special education. *Exceptional Children*, 56(3):232–235.
- Rabelo, L. C. C. (2012). *Ensino colaborativo como estratégia de formação continuada de professores para favorecer a inclusão escolar*. Universidade Federal de São Carlos.
- Rodríguez, M. C. (2014). Opening lesson plans to support teaching innovation and open educational resources adoption. In *Proceedings of The International Symposium on Open Collaboration, OpenSym ’14*, pages 47:1–47:4, New York, NY, USA. ACM.

- Sa'don, N. F., Alias, R. A., and Ohshima, N. (2014). Nascent research trends in moocs in higher educational institutions: A systematic literature review. In *Web and Open Access to Learning (ICWOAL), 2014 International Conference on*, pages 1–4.
- Sa'don, N. F., Alias, R. A., and Ohshima, N. (2016). In *Proceedings of the EUROPEAN STAKEHOLDER SUMMIT on experiences and best practices in and around MOOCs*, pages 173–184.
- Santos, N. S., Ferreira, L. S., and Prates, R. O. (2012). An overview of evaluation methods for collaborative systems. In *Collaborative Systems (SBSC), 2012 Brazilian Symposium on*, pages 127–135.
- Thurler, M. and Wolff, J. (2001). *Inovar no interior da escola*. Artmed.
- Torres, P. L. (2015). *Metodologias para a produção do conhecimento: da concepção à prática*. Curitiba : SENAR - PR.
- Valtolina, S., Barricelli, B. R., Mesiti, M., and Ribaldo, M. (2012). User-centered design of e-learning tools for users with special needs: The visualpedia case study. *IxD&A*, 13:47–55.

Avaliação por pares como ferramenta colaborativa na correção de redações: Um experimento com o ambiente educacional MeuTutor

Thyago Tenório¹, Ig Ibert Bittencourt²

¹Universidade Federal de Alagoas (UFAL) – Campus Arapiraca/Pólo Penedo
Av. Beira Rio – 57.200-000 – Penedo – AL – Brasil

²Instituto de Computação (IC) – Universidade Federal de Alagoas (UFAL)
Av. Lourival Melo Mota, s/n – 57.072-900 – Maceió – AL – Brasil

thyago.oliveira@penedo.ufal.br, ig.ibert@ic.ufal.br

Abstract. *While on-line learning environments provide scalable ways to present content, forums and evaluate student progress automatically, they are still limited in their ability to evaluate complex activities such as essays correction. In this kind of activity, the manually correction by teachers / tutors would quickly become infeasible due to the increased number of students. This paper presents a peer assessment mechanism and applies it in essays correction in MeuTutor educational environment. In the experiment conducted, it was concluded that there were no statistical variations between the grades obtained compared to the experts, the correction time was reduced, and the cost involved decreased 72.4 %.*

Resumo. *Embora os ambientes de aprendizagem online ofereçam maneiras escaláveis para apresentar um conteúdo, fóruns e avaliar o progresso dos alunos automaticamente, eles continuam limitados na capacidade de avaliar atividades complexas, como correção de redações. Nesse tipo de atividade, a correção de forma manual pelos professores/tutores se tornaria rapidamente inviável a medida que o número de atividades cresce. Este trabalho apresenta um mecanismo de avaliação por pares e aplica-o no contexto de redações no ambiente educacional MeuTutor. Com o experimento realizado, concluiu-se que não houve variações estatísticas entre as notas obtidas se comparado à especialistas, o tempo de correção foi reduzido, assim como o custo envolvido diminuiu 72.4%.*

1. Introdução

A aplicação de tecnologias da informação e comunicação (TIC) na educação tem sido cada vez mais destacada no processo de ensino-aprendizagem. Nos últimos anos, diversos países têm adaptado suas abordagens educacionais para promover e suportar a utilização de tecnologias, tanto em cursos presenciais quanto em cursos a distância. Contudo, enquanto novas tecnologias da Web permitem formas escaláveis para fornecer conteúdo de vídeo conferência, implementar fóruns sociais e acompanhar o progresso do aluno nestes ambientes, continuamos limitados em nossa capacidade de avaliar e dar feedback para trabalhos complexos dos estudantes e muitas vezes abertos (avaliações escritas), tais como provas matemáticas, projeto de problemas e redações [Piech et al. 2013].

Diante de tal dificuldade, estes ambientes educacionais online frequentemente oferecem soluções mais simples aos seus usuários, como, por exemplo, questões de múltipla escolha, uma vez que este tipo de problema pode ser processado de forma automática pelo computador. Com o crescimento do número de estudantes nestes ambientes e consequentemente, um maior número de atividades sendo feitas a cada momento, a correção delas pelos professores/tutores se tornaria rapidamente inviável. Para suprir este problema pode-se utilizar um conceito chamado de avaliação por pares. Avaliação por pares é um processo pelo qual os alunos ou seus pares atribuem notas com base em parâmetros de referência pré-definidos pelo professor [Sadler and Good 2006]. A prática é utilizada para salvar o tempo dos professores e melhorar a compreensão dos estudantes a respeito dos materiais do curso, bem como melhorar as suas habilidades meta-cognitivas [Malehorn 1994].

No contexto de ambientes de EAD, o sistema gerencia as correções feitas entre os próprios estudantes, isto é, os próprios alunos corrigem as atividades escritas de seus pares. Existem diversas abordagens de avaliação por pares publicados na literatura, como as abordagens formativas [Orsmond* et al. 2004], modelos probabilísticos [Piech et al. 2013] e até mesmo modelos usando redes bayesianas [Wang and Vassileva 2003]. Como consequência, o uso desta técnica resulta em um aumento da aprendizagem do aluno e, principalmente, a redução de sobrecarga do professor [Sadler and Good 2006]. No entanto, o uso de modelos de avaliação por pares aplicados em ambientes educacionais online com o propósito de correções de atividades escritas (redações) ainda é limitado. Os resultados de sua aplicação são fortemente ligados ao contexto envolvido e dependem de vários fatores externos, incluindo, por exemplo, o nível educacional dos avaliadores e as atividades que serão corrigidas.

Neste sentido, a solução proposta visa criar um mecanismo de avaliação por pares e aplicá-lo no contexto de provas discursivas nestes ambientes. Usando este conceito em um ambiente educacional EAD, o sistema ficará responsável pelo gerenciamento das correções que serão feitas pelos próprios estudantes. Usando avaliação por pares é possível desatrelar o custo das correções do número de usuários, e sem necessariamente aumentar o trabalho do professor. Ao se utilizar essa técnica, surgem vários problemas para o bom funcionamento da colaboração entre esses alunos, como por exemplo "como selecionar os pares mais adequados para a correção?" ou "como garantir um nível alto de confiança e reputação geral dos avaliadores?". Na solução aqui apresentada é incluído mecanismos para solucionar tais problemas dentro do modelo proposto.

Um experimento foi conduzido com o objetivo de avaliar o mecanismo de avaliação por pares criado e apresentado neste trabalho. Dessa forma, este foi implementado dentro de um ambiente educacional chamado MeuTutor. Nesse experimento, cerca de 30 alunos tiveram acesso ao sistema e tinham como objetivo realizar algumas redações que foram liberadas no ambiente para serem feitas e, em seguida, corrigir tais redações utilizando o mecanismo de avaliação por pares proposto, assim como o método tradicional (em que as correções foram feitas por professores especialistas). Como resultado, não houve variações estatísticas significativas entre as notas obtidas em ambos os modelos, o tempo de correção usando o modelo proposto se mostrou mais rápido que o tempo de correção por especialista e houveram evidências estatísticas suficientes que indicaram que o custo envolvido do modelo proposto se mostrou menor que o custo do modelo tra-

dicional. Através da criação de um modelo de regressão apresentado, mostrou-se que a redução no custo é de cerca de 72.4%.

Este documento está estruturado da seguinte forma: A Seção 2 apresenta alguns trabalhos relacionados. Em seguida, na Seção 3 é apresentado de forma resumida o ambiente educacional MeuTutor. Logo após, a Seção 4 apresenta o mecanismo de avaliação por pares criado. Finalmente, a Seção 5 apresenta o planejamento e execução do experimento que foi conduzido com o objetivo de avaliar o mecanismo criado e a Seção 6 apresenta as nossas conclusões, limitações e alguns trabalhos futuros.

2. Trabalhos relacionados

Os trabalhos discutidos aqui envolvem o uso de avaliação por pares em ambientes educacionais. Podemos citar alguns trabalhos como em [Dominguez et al. 2012] em que é aplicada a técnica para avaliação de trabalhos em um curso de engenharia. [Chang et al. 2012] aplica com o objetivo de avaliar portfólios em estudantes do ensino médio que cursam computação. Finalmente em [Kawai 2006] é utilizado a técnica para avaliação de mensagens de voz e cartas para alunos estudantes de língua estrangeira. Neste trabalho, a técnica é utilizada de maneira abstrata, no qual um modelo de utilização é proposto e que pode ser utilizado para correções de avaliações escritas em geral, independente de qual tipo de avaliação específica será realizada pelo professor.

Em [Sterbini and Temperini 2013] é apresentada uma ferramenta chamada OpenAnswer que tem como objetivo permitir o suporte a avaliações escritas usando técnicas de avaliações por pares, algo semelhante ao proposto aqui nesse trabalho. Contudo, essa ferramenta não é integrada ao ambiente, o que implicaria que os usuários tivessem que utilizar mais uma plataforma só para as atividades de avaliações escritas. Em nossa proposta, pretendemos que o uso de avaliação por pares seja incluído direto no ambiente, sem essa necessidade de outros sistemas. Dessa forma o trabalho aqui proposto não apresenta uma ferramenta que utiliza avaliação por pares e sim um modelo de avaliação por pares que poderá ser utilizado nos ambientes educacionais.

Em [Kahiigi Kigozi et al. 2012], os autores projetaram e modelaram um processo de revisão baseado em avaliação por pares para aprendizagem colaborativa. O modelo é baseado no uso do estudante, apoiando-o pedagogicamente sua aprendizagem. Em [Tosic and Nejkovic 2010], é proposto um novo método para avaliação por pares dos alunos com base no conceito de confiança. Outros trabalhos propõe modelos baseados em redes bayesianas como em [Sterbini and Temperini 2013] e [Wang and Vassileva 2003]. Todos esses modelos são implementados de forma própria e utilizados de acordo com suas necessidades, porém não tem como objetivo se integrar com os diversos ambientes educacionais e, portanto, se diferem desse trabalho nesse aspecto.

3. O ambiente educacional MeuTutor

O modelo foi implementado e integrado ao ambiente educacional MeuTutor. Podemos defini-lo como um sistema tutor inteligente, que tem como objetivo acompanhar a aprendizagem dos alunos de forma personalizada, garantindo uma qualidade no ensino e melhorando o desempenho dos seus usuários. O ambiente visa auxiliar alunos do ensino médio a se prepararem para o Exame Nacional do Ensino Médio (ENEM). Com isso, oferece cursos referentes às disciplinas do ensino médio.

O ambiente MeuTutor-ENEM foi construído tendo como ponto forte três bases: ambiente gamificado, aprendizagem personalizada e experiência social. O primeiro tem como objetivo motivar o aluno a continuar seus estudos no ambiente. O objetivo principal da Gamificação é aumentar o engajamento dos usuários por meio do uso de técnicas semelhantes àquelas presentes em jogos, fazendo com que os usuários se sintam no controle de suas ações e se motivem com as tarefas [Pavlus 2010]. Com isso, é possível observar na Figura 1 alguns elementos de Gamificação (jogos) presentes na plataforma, que foi projetada como se fosse um jogo para motivar a aprendizagem e o estudo do aluno. A plataforma disponibiliza um mecanismo de pontos e níveis personalizados, para o aluno evoluir em cada disciplina e no geral (na figura o aluno está no nível 5).



Figura 1. Elementos de gamificação do MeuTutor

Sob o segundo aspecto, a aprendizagem personalizada, no MeuTutor-ENEM as questões e vídeo-aulas se adequam aos alunos. Por meio de uma base de questões no estilo ENEM, o estudante pode praticar seu conhecimento resolvendo vários problemas de cada assunto nas disciplinas disponíveis no sistema. À medida que ele resolve corretamente um problema, seu progresso no assunto cresce. Se errar a questão, uma vídeo-aula será recomendada para ele suprir sua deficiência.

Por fim, no terceiro aspecto (experiência social), o MeuTutor mantém a preferência dos alunos tradicionais (discutir e estudar um determinado assunto ou conteúdo juntos, construindo conhecimento compartilhado), fornecendo recursos poderosos dentro do ambiente por meio de técnicas da aprendizagem colaborativa em sistemas computacionais. Entre os recursos colaborativos providos aos estudantes está a ideia de grupos de estudo. Dessa forma, um aluno pode convidar outros amigos para praticar um determinado assunto. A Figura 2 mostra a tela de grupo de estudos, em que os alunos respondem questões juntos depois de debater e entrar em um determinado consenso.

Figura 2. Estudos em grupo - Experiência social do MeuTutor

4. Mecanismo Avaliação por pares

Para entender melhor como o mecanismo se comporta dentro do sistema, a Figura 3 apresenta o fluxo de execução que deve ser incrementado no ambiente educacional.

O primeiro passo é a criação das atividades discursivas feitas pelo professor juntamente com a criação e definição do formulário de avaliação, contendo os critérios no qual as atividades serão avaliadas pelos alunos. As atividades e os formulários deverão, em seguida, serem enviados (cadastrados) ao sistema pelo professor em uma interface específica para isso. O sistema por sua vez irá disponibilizar as atividades aos alunos de acordo com o plano de avaliação pré-definido. A partir da disponibilização da atividade ao aluno para ser respondida, inicia-se o período de submissões - segundo passo. Esse período também está definido no plano de avaliação. Os estudantes responderão as atividades no próprio ambiente online.

Em seguida, o aluno submeterá sua resposta para o ambiente e este será responsável pelo gerenciamento da correção desta. No terceiro passo, em primeiro lugar, a gerência do sistema identifica os alunos que irão corrigir a atividade submetida e, a partir desse momento, o sistema enviará a atividade para estes alunos. Esse processo de escolha dos corretores deve ser feito por meio de uma lista de usuários que fizeram a mesma atividade e também possui atividades pendentes de correção, em ordem crescente



Figura 3. Fluxo de execução do modelo em detalhes

de atividades atribuídas, de acordo com o algoritmo a seguir.

É de salientar que o número mínimo de correções necessárias (`minCorrection` do Passo 1) para uma atividade é muito dependente do tipo de atividade que está sendo executada. Teoricamente, um maior número torna mais confiável os resultados da avaliação. Por outro lado, um maior número gera mais trabalho para os alunos, o que pode ser prejudicial para o resultado final do modelo. Portanto, há um trade-off na escolha deste número. O número de margem de segurança (`safetyMargin` a partir do Passo 2) serve como um armazenamento de usuários que serão recuperados para análise. Quanto maior o número de alunos, maior a precisão e mais lento o algoritmo. Por sua vez, o número máximo de atividades de correção pendentes a um único utilizador (`maxPendentCorrection` a partir do Passo 3) também deve ser escolhido cuidadosamente, uma vez que pode sobrecarregar os estudantes envolvidos.

O número mínimo de estudantes selecionado foi 2. A razão foi que, com duas correções é possível avaliar os resultados que sejam semelhantes. Se houver discordância, um terceiro estudante será associado. A nota final é a média das duas notas mais próximas dos alunos. Por outro lado, para que o modelo funciona corretamente o número máximo de atividades pendentes foi escolhido como 3, porque esperamos que cada usuário corrija, pelo menos, três redações. Em seguida, no quarto passo, os estudantes, de acordo com a distribuição na etapa anterior, por sua vez, irão avaliar as respostas para as atividades apresentadas para serem corrigidas, baseado no formulário de avaliação pré-definido e disponibilizado pelo professor. O prazo definido para a revisão (correção) das atividades no sistema é chamado de período de avaliação. É neste período que os alunos avaliarão seus pares e enviarão as notas para o sistema.

Algorithm 1 Recuperar usuários para correção de uma atividade

function RETRIEVEUSERS(minCorrection, safetyMargin, maxPendentCorrection)

/ minCorrection - Número mínimo de correção por estudantes; safetyMargin - Margem de segurança para buscar usuários; maxPendentCorrection - Número máximo de atividades pendentes de correção para um único usuário/

```

    List<User> users = retrieveUsersMadeActivity(idActivity, minCorrection + safetyMargin, "Ascending order");
    for i do 1 until size(users)
        qtdPendentCorrection = getAmountPendentCorrectionByUser(users.get(i));
        if qtdPendentCorrection > maxPendentCorrection then
            users.remove(user.get(i));
        end if
    end for
    if size(users) >= minCorrection then
        return users.sublist(0,minCorrection-1);           ▷ Retorna os primeiros
minCorrection elementos da lista
    else
        List<User> moreUsers = retrieveUsersHighProbabilityCorrect(minCorrection - size(users));
        users.addAll(moreUsers);
        return users;
    end if
end function

```

Para atribuir sua nota, o estudante receberá as respostas de um determinado aluno e o formulário de avaliação com os critérios definidos. Em seguida, irá analisar a atividade sob o ponto de vista dos critérios (assim como um professor faria) e em seguida basta preencher o formulário com suas respostas. Esse processo será repetido enquanto houver atividades existentes para a correção para aquele aluno. De acordo com o algoritmo apresentado, esse número tende a ser um média entre os alunos, para não sobrecarregar nenhum. Porém, é possível que haja pequenas variações.

No quinto passo, para se ter uma maior confiança na nota final, deverão ser avaliados a reputação geral de cada avaliador e o nível de competência de cada avaliador em cada critério. O cálculo da reputação geral de um avaliador deverá levar em consideração as seguintes atividades realizadas por um aluno (em que + significa um aspecto positivo e - significa um aspecto negativo): + Acesso frequente ao sistema (A); + Alto número de atividades discursivas realizadas (QP); + Alto número de correções de atividades feitas (QC); – Baixa frequência de acesso ao sistema; – Alto número de correções pendentes (PC); – Alto número de atividades feitas com baixo número de correções.

Para se calcular o nível de competência do avaliador em um determinado critério deve-se olhar seu histórico de atividades feitas que utilizaram aquele critério. Caso não possua dados, seu nível será considerado neutro. Caso possua dados no histórico, este de-

verá ser utilizado para determinar se o avaliador possui um aspecto positivo e/ou negativo a depender das notas dos critérios. A partir dos cálculos definidos, o sistema calculará a confiabilidade de cada avaliação, o nível de competência do avaliador nos critérios e poderá então calcular o resultado final usando a reputação geral (OR) de cada avaliador. Este processo de cálculo da reputação geral do aluno X acontece com as seguintes equações:

$$FA = \frac{A(X) - AVG(A)}{MAX(A)} \quad FQP = \frac{QP(X) - AVG(QP)}{MAX(QP)} \quad FQC = \frac{QC(X) - AVG(QC)}{MAX(QC)}$$

$$OR = \frac{FA*1 + FQP*2 + FQC*2}{5}$$

O intervalo para as funções FA, FQP e FQC é $-1 \leq F \leq 1$. Nessas funções, quanto mais alto os valores para a quantidade de acesso, quantidade de redações realizadas e quantidade de redações corrigidas, respectivamente, mais próximos os valores de sua função será 1. Da mesma forma, valores muito baixos para tais quantidades, implicará em valores próximos a -1. Se um estudante estiver na média, o resultado será 0 (neutro). Note que a função de reputação geral (OR) tem um comportamento similar, no intervalo $[-1,1]$, sendo definida como a média ponderada das funções anteriores.

Finalmente, no sexto passo, o sistema deverá gerar um relatório contendo o resultado final da avaliação que será apresentado ao aluno que submeteu a atividade. Este relatório deverá ser detalhado, apresentando as notas de cada critério, bem como a nota final da atividade. É possível a inclusão de comentários e possíveis alterações nas notas por parte do professor (opcional) antes de sua apresentação ao aluno. Este por sua vez irá identificar os pontos fracos e usar o sistema para aprender mais, através dos materiais do curso, realizando novas atividades e/ou corrigindo as atividades dos outros estudantes.

5. Experimento

Em ordem para avaliar a efetividade, nós projetamos e realizamos um experimento controlado no ambiente MeuTutor. Inicialmente, o MeuTutor não suportava redações. Por esta razão, os criadores do ambiente incluíram o mecanismo aqui apresentado, permitindo que os estudantes pudessem realizar redações dentro do ambiente. Essa inclusão do modelo nos levou a questionar a qualidade das correções com a seguinte pergunta: **“Como nós podemos avaliar a qualidade das correções feitas pelos alunos?”** e, principalmente, **“essas avaliações podem ser comparadas com uma avaliação de um especialista (professor)?”**. Além disso, também nos questionamos sobre **“Será que o modelo aplicado nesse contexto realmente soluciona os problemas citados?”** (Redução do custo e sobrecarga do professor, entre outros). Assim, surgiram as seguintes questões de pesquisa:

P1 - O uso do modelo de avaliação por pares proposto apresenta a mesma eficiência, isto é, apresenta resultados semelhantes se comparado ao método tradicional de correções? O que nos leva às seguintes hipóteses: **H1-0**: O uso do modelo de avaliação por pares proposto é equivalente ao método tradicional; **H1-1**: O uso do modelo de avaliação por pares proposto não é equivalente ao método tradicional

P2 - O uso do modelo de avaliação por pares proposto apresenta diferenças nas métricas de tempo em relação a não usar esse modelo? O que nos leva às seguintes hipóteses: **H2-0**: O uso do modelo de avaliação por pares proposto não traz diferenças de tempo em relação ao método tradicional; **H2-1**: O uso do modelo de avaliação por pares proposto traz diferenças significativas de tempo em relação ao método tradicional.

P3 - O uso do modelo de avaliação por pares proposto apresenta diferenças nas métricas de custo em relação a não usar esse modelo? O que nos leva às seguintes hipóteses: **H3-0**: O uso do modelo de avaliação por pares proposto não traz diferenças de custo em relação ao método tradicional; **H3-1**: O uso do modelo de avaliação por pares proposto traz diferenças significativas de custo em relação ao método tradicional.

Formalmente, as hipóteses descritas anteriormente podem ser definidas conforme a Tabela 1. As funções N, T e C, apresentadas na tabela, retornam respectivamente, o valor da nota final, o tempo gasto e o custo envolvido, com relação a utilização do modelo tradicional M1 ou a utilização do modelo proposto M2, quando aplicado no ambiente educacional MeuTutor E1 ou quando aplicado sem ambiente educacional E2. (N, T e C são as métricas e M (M1 = Modelo tradicional / M2 = Modelo proposto) e E (E1 = MeuTutor / E2 = Sem ambiente) são os fatores do nosso experimento.

Tabela 1. Definição formal das hipóteses de pesquisa.

Hipótese	Hipótese Nula	Hipótese Alternativa
H1	H1-0: $N(M1,E1) = N(M2,E1)$	H1-1: $N(M1,E1) \neq N(M2,E1)$
H2	H2-0: $T(M1,E1) = T(M2,E1)$	H2-1: $T(M1,E1) \neq T(M2,E1)$
H3	H3-0: $C(M1,E1) = C(M2,E1)$	H3-1: $C(M1,E1) \neq C(M2,E1)$

No nosso caso, como temos fatores com apenas 2 níveis cada, podemos utilizar um experimento fatorial 2^k sem repetição, retirando uma combinação que é impossível de ser realizada (que é a utilização do modelo proposto M2 em nenhum ambiente educacional E2). Cada execução do modelo tem um custo relativamente alto além de exigir bastante tempo para preparação do ambiente, alunos e demais esforços necessários para sua conclusão. Por este motivo, foi escolhido um projeto fatorial sem repetição. Nesse sentido temos apenas 3 ensaios possíveis, sendo executado cada um apenas uma vez, totalizando três tratamentos possíveis e consequentemente três execuções. Os tratamentos são: T1 (M1 e E1), T2 (M2 e E1) e T3 (M1 e E2).

Após a configuração do ambiente MeuTutor, o próximo passo foi a escolha dos alunos que iriam participar do experimento. Para isso, foram selecionados 30 usuários do MeuTutor que gostariam de participar desse experimento. Vale ressaltar que todos os usuários possuem o mesmo conhecimento na disciplina abordada neste experimento. Em seguida, foram convidados dois especialistas em correção de redações, indicados aqui por especialista 1 e especialista 2. Os usuários executaram seus devidos tratamentos e os especialistas corrigiram as redações em T1 e T3.

5.1. Análise de dados

A seguir, serão analisados os dados de cada uma das variáveis estudadas. A primeira variável analisada é a nota final (N). A Figura 4 apresenta o diagrama de caixa para a métrica da nota final N com relação aos tratamentos T1, T2 e T3.

Podemos ver na figura que a média (média (T1)=615 contra média(T2)=578) e mediana (mediana(T1) = 660 contra mediana(T2)=573), nos dois casos, estão bem próximas. Esta figura sugere que as notas obtidas com a aplicação desses tratamentos possuem variações estatísticas semelhantes, o que nos indica uma certa semelhança nos tratamentos. No entanto, ainda não foram geradas evidências estatísticas para afirmar isso. Tais afirmações só poderão ser feitas quando os testes estatísticos forem realizados. Como

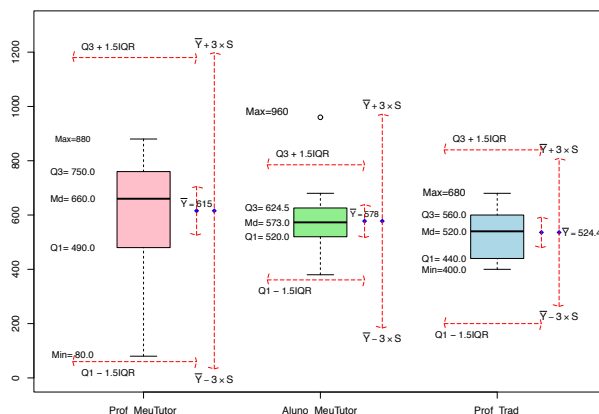


Figura 4. Diagrama de caixa com comparativo da métrica nota para os tratamentos T1, T2 e T3

uma de nossas questões de pesquisa é avaliar as notas obtidas pelo modelo proposto comparadas com as notas obtidas pelo modelo tradicional, então nós subtraímos T1 de T2. Podemos observar na figura que a diferença das notas do tratamento T1 e T2, na média (média(T1 - T2)= 37,8) e mediana (mediana(T1 - T2)= 40), estão bem próximas de 0 (zero), indicando que são bem equivalentes. No Geral, a diferença entre as notas T1 e T2 variou entre -79,75 (o resultado negativo indica que a nota do especialista no tratamento T1 foi menor que a nota do modelo no tratamento T2) a 110 (resultado positivo indica que a nota do modelo foi menor que a nota do especialista). Se considerarmos a escala de notas entre 0 e 1000, temos que as diferenças entre as notas variam entre 7.9% a 11%.

Contudo, ainda precisamos verificar a validade das hipóteses de pesquisa. Enfatizamos que para cada hipótese, deve-se indicar qual tratamento foi melhor. A primeira tarefa para se realizar a verificação de uma determinada hipótese de pesquisa é analisar a normalidade da distribuição dos dados que estão envolvidos em sua resposta. A normalidade dos dados é importante pois determina qual teste estatístico deve ser utilizado na análise. Existem alguns testes estatísticos para verificar a normalidade dos dados, porém o mais recomendado pelos estatísticos é o teste de Shapiro-Wilk [Shapiro and Francia 1972].

A primeira verificação de hipótese será feita com relação a métrica de nota. Os resultados da aplicação do teste de Shapiro-Wilk quando executado sobre os dados da métrica nota nos tratamentos estão apresentados na tabela 2. Neste caso, nós vemos que os dados vem de uma distribuição normal, uma vez que $W_{calculado} > W_{\alpha}$ e $P_{valor} > \alpha$ em todos os casos. Assim, nós aplicamos o T-Teste [Welch 1938], comparando os tratamentos. Os dados da execução do T-test estão apresentados na Tabela 3.

Podemos observar que os valores obtidos com a execução do T-test, são todos maiores que $\alpha = 0.05$. Portanto, com 95% de confiança, estatisticamente, não pode se afirmar que os valores das notas obtidos entre os tratamentos possuem diferenças significativas entre si, isto é, não há evidências estatísticas que mostrem a não equivalência das notas, não sendo possível gerar evidência estatística suficiente para refutar a hipótese

Tabela 2. Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Nota

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.92223	0.897	0.1416	0.05
T2	0.90344	0.897	0.06599	0.05
T3	0.96201	0.842	0.8086	0.05

Tabela 3. Resultado da aplicação do T-test com os dados da métrica Nota

Tratamento	t	df	p-value
T1 x T2	0.85383	17	0.4051
T1 x T3	1.4728	25.546	0.153
T2 x T3	0.98704	24.523	0.3333

nula H_1-0 , implicando que **O uso do modelo de avaliação por pares proposto é equivalente ao método tradicional**. A próxima verificação de hipótese será feita com relação a métrica Tempo (T). A Figura 5 apresenta um conjunto de diagramas de caixa para a métrica do tempo (T) com relação aos tratamentos T1, T2 e T3.

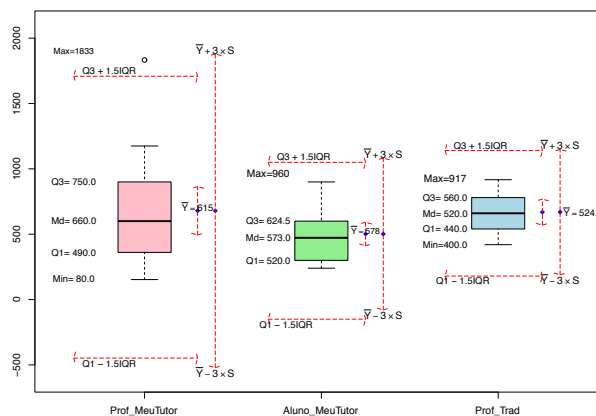


Figura 5. Diagrama de caixa com comparativo da métrica tempo para os tratamentos T1, T2 e T3

Ao se analisar os dados de média e mediana dos tratamentos em que há um professor especialista envolvido, podemos verificar que os dados são bem semelhantes. Se compararmos esses dados com os dados do tratamento T2, notamos que o tempo no tratamento T2 é significativamente menor que o tempo nos outros dois tratamentos. Porém, como os diagramas se interceptam, é necessário a execução dos testes de hipóteses para ter uma conclusão estatisticamente válida. Nesse caso, aplicou-se o teste de Shapiro-Wilk com os dados desses tratamentos para analisar a normalidade dos dados e o resultado pode ser visto na Tabela 4.

Ao se analisar os dados da tabela, temos um caso especial no tratamento T1, uma vez que $W_{calculado}(T1) = 0.89637 < W_{\alpha}(T1) = 0.897$ e $P_{valor}(T1) = 0.04969 < \alpha = 0.05$. Com isso, podemos refutar a hipótese nula do teste de Shapiro-Wilk e consequentemente, podemos afirmar com nível de significância de 5% que a amostra não provém de uma

Tabela 4. Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Tempo

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.89637	0.897	0.04969	0.05
T2	0.91724	0.897	0.1155	0.05
T3	0.97249	0.842	0.9129	0.05

população normal. Se analisarmos os dados do tratamento T2 e T3 vemos que esses provêm de uma distribuição normal. Nesse caso, como os dados do tratamento T1 possui uma distribuição não-normal, então quando os dados de comparação envolverem T1 deveremos usar um teste não-paramétrico, como o Teste de Wilcoxon. Nesse caso, apenas a combinação T2 x T3 envolverá o T-test, uma vez que ambas são distribuições normais. Os dados da execução desses testes estão apresentados na Tabela 5.

Tabela 5. Resultado da aplicação dos testes estatísticos com os dados da métrica Tempo

Tratamento	Teste Utilizado	v/w	t	df	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	122	-	-	0.1169
T1 x T3	Wilcoxon rank sum test with continuity correction	81.5	-	-	0.7007
T2 x T3	Welch Two Sample t-test	-	-2.4605	21.938	0.02222

Podemos observar, de acordo com a Tabela 5, que como $p\text{-value}(T1 \times T2) = 0.1169 > \alpha = 0.05$, então não é possível concluir que há diferenças estatísticas entre esses tratamentos. Contudo, ao analisarmos o p-value dos tratamentos T2 e T3, $p\text{-value}(T2 \times T3) = 0.02222 < \alpha = 0.05$, podemos concluir que há diferenças estatísticas significativas e podemos concluir, com 95% de confiança, que o tempo em T2 é menor que o tempo em T3. Como a nossa avaliação se baseia no comportamento dos dados do tratamento T2 (em que o modelo é aplicado), temos estatisticamente que o tempo envolvido nele é menor que o tempo do tratamento T3 (especialista sem ambiente) com 95% de confiança. Há evidências estatísticas suficientes de que o uso do modelo proposto traz diferenças de tempo em relação ao método tradicional e com isso, pode-se refutar a hipótese nula H2-0 (que não haveria diferença) e aceitar a hipótese alternativa H2-1, com 95% de confiança, implicando que **o uso do modelo de avaliação por pares proposto traz diferenças significativas de tempo em relação ao método tradicional.**

Finalmente, a última métrica a ser analisada é o custo da correção (C). Esse custo tem diferentes significados a depender do tratamento envolvido. Quando envolve professores, o custo refere-se ao preço cobrado (em média) pelos especialistas para corrigir cada redação. Isso inclui também dados que o próprio INEP disponibiliza como valor base para o pagamento de correção de redações. Já com relação ao custo do modelo, se aplica aos custos de manter o servidor com a aplicação funcionando mais o processo de configuração do ambiente pelo professor (pago por hora de trabalho). A figura 6 apresenta o diagrama de caixa para a métrica de custo envolvido (C) com relação aos tratamentos executados.

Podemos observar que as médias e medianas nos tratamentos T1 e T3 ($média(T1) = R\$6,03$ / $média(T3) = R\$4,83$) são parecidas, uma vez que ambos os tratamentos utilizam especialistas nas correções. A variação existente entre elas se dá pelo fato de que os especialistas cobraram preços diferentes. Há um espalhamento (variação) maior

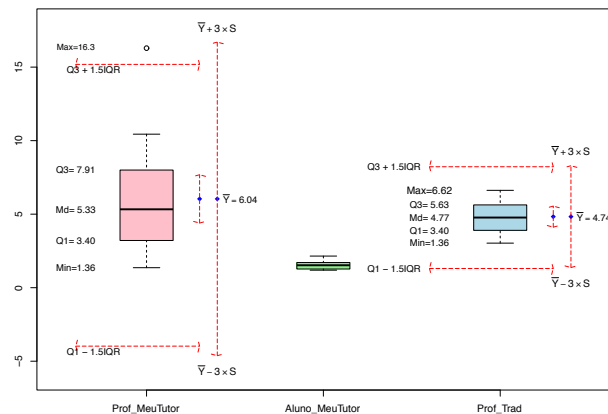


Figura 6. Diagrama de caixa com comparativo da métrica custo entre os tratamentos T1, T2 e T3

nos dados do tratamento T1, enquanto que esse comportamento não ocorre no tratamento T3, indicando que o custo por usuário no tratamento T3 é mais uniforme que o custo por usuário no tratamento T1. Finalizando a verificação de hipóteses, serão analisadas as hipóteses que envolvem a métrica de custo (C). Os resultados da aplicação do teste de Shapiro-Wilk com os dados dos tratamentos T1, T2 e T3 estão ilustrados na Tabela 6.

Tabela 6. Resultado da aplicação do teste de Shapiro-Wilk com os dados da métrica Custo

Tratamento	$W_{calculado}$	W_{α}	P_{valor}	α
T1	0.89635	0.897	0.04965	0.05
T2	0.25269	0.897	1.057e-08	0.05
T3	0.97276	0.842	0.9152	0.05

Analisando os dados da tabela, observamos agora dois casos de anormalidade dos dados. No caso do tratamento T1 e T2, uma vez que $W_{calculado} < W_{\alpha}$ e $P_{valor} < \alpha$. Por outro lado, T3 é normal. Os resultados da aplicação do Teste de Wilcoxon com os dados desses tratamentos estão apresentados na Tabela 7.

Tabela 7. Resultado da aplicação do teste de Wilcoxon com os dados da métrica Custo

Tratamento	Teste Utilizado	v/w	p-value
T1 x T2	Wilcoxon signed rank test with continuity correction	170	0.0002522
T1 x T3	Wilcoxon rank sum test with continuity correction	103	0.5486
T2 x T3	Wilcoxon rank sum test with continuity correction	0	1.111e-06

Com os dados da tabela, podemos observar que o $p\text{-value}(T1 \times T2) = 0.0002522 < \alpha = 0.05$, implicando que há diferenças estatísticas significativas entre o custo dos tratamentos T1 e T2. Com isso, podemos concluir que o custo do tratamento T2 é menor que o custo do tratamento T1 com uma confiança de 95%. Da mesma forma, se analisarmos o valor de $p\text{-value}(T2 \times T3) = 1.111e-06 < \alpha = 0.05$, concluímos que o custo também são diferentes, isto é, com 95% de confiança temos que o custo de T2 é menor que o custo de

T3. Por outro lado, o valor de $p\text{-value}(T1 \times T3) = 0.5486$ é maior do que $\alpha = 0.05$, o que implica que não há diferença significativa entre os tempos de T1 e T3.

Como nós queremos avaliar é o comportamento dos dados do tratamento T2, temos comprovado estatisticamente que o custo envolvido em T2 é menor que os custos envolvidos tanto no tratamento T1 quanto no tratamento T3 com uma confiança de 95%. Com isso, temos evidências estatísticas suficientes de que o custo são diferentes e portanto, conseguimos refutar a hipótese nula H3-0 (em que não haveria diferença) e aceitar a hipótese alternativa H3-1, com uma confiança de 95%, implicando que **o uso do modelo de avaliação por pares proposto traz diferenças significativas de custo em relação ao método tradicional**. Por meio de um modelo de regressão linear foi mostrado que o custo é 72.4% menor.

6. Conclusão e trabalhos futuros

O trabalho apresentou um mecanismo de avaliação por pares que tem como objetivo prover uma solução para a inclusão de avaliações escritas em ambientes educacionais online de forma eficiente. A necessidade da criação desse modelo surgiu do grande número de estudantes que estavam presentes nos ambientes e da grande dificuldade em prover provas discursivas neles, uma vez que isso gerava um alto custo e uma grande sobrecarga nos professores envolvidos no processo. Dessa forma, as atividades discursivas eram limitadas, dificultando a aprendizagem do aluno no processo.

O modelo proposto atingiu o objetivo, possibilitando a inclusão de avaliações discursivas nos ambientes online de maneira viável, como podemos observar em sua integração com o ambiente educacional MeuTutor e seu uso com provas de redação aplicadas para cerca de 30 alunos. Os experimentos apresentados também deram comprovações estatísticas suficientes de que os resultados obtidos com o modelo proposto são semelhantes aos resultados do modelo tradicional. Os resultados foram bastante satisfatórios, tendo em vista que foi comprovado, estatisticamente, que as notas dos modelos são equivalentes, permitindo que a substituição entre eles possa ser feita, sem comprometer os resultados finais. Da mesma forma, os resultados da métrica Tempo indicaram uma possibilidade de o tempo de correção ser menor usando o modelo proposto, contudo em alguns casos esse tempo seja semelhante. Por fim, os resultados da métrica Custo foram bem favoráveis a aplicação do modelo, tendo em vista que não é necessária uma correção por especialista no modelo proposto. O modelo de regressão criado mostrou que o custo é cerca de 72.4% menor.

Como uma contribuição secundária do modelo proposto, temos os algoritmos de seleção de usuários, bem como os cálculos de confiança e reputação dos usuários. Como trabalho futuro temos o planejamento de implantar o modelo em outros ambientes educacionais diferentes do MeuTutor, como por exemplo, o Moodle. Com a implementação/integração em outros ambientes, as dificuldades que por ventura poderão aparecer servirão como entrada para possíveis melhorias na implementação do modelo, deixando-o mais completo, flexível e compatível com a maioria dos ambientes educacionais.

Além disso, é preciso avaliar novas métricas de comparação do modelo proposto com o modelo tradicional que podem ser relevantes, como por exemplo, o nível de aprendizagem do aluno, melhorias das suas capacidades de julgamento, entre outros benefícios

que um modelo de avaliação por pares pode trazer. Com isso, pretende-se realizar novos experimentos mais completos a fim de avaliar tais métricas. Além disso, é preciso avaliar também questões subjetivas através da aplicação de questionários com os atores envolvidos no processo.

Referências

- Chang, C.-C., Tseng, K.-H., and Lou, S.-J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a web-based portfolio assessment environment for high school students. *Computers & Education*, 58(1):303–320.
- Dominguez, C., Cruz, G., Maia, A., Pedrosa, D., and Grams, G. (2012). Online peer assessment: An exploratory case study in a higher education civil engineering course. In *Interactive Collaborative Learning (ICL), 2012 15th International Conference on*, pages 1–8. IEEE.
- Kahiigi Kigozi, E., Vesisenaho, M., Hansson, H., Danielson, M., and Tusubira, F. (2012). Modelling a peer assignment review process for collaborative e-learning. *Journal of Interactive Online Learning*, 11(2):67–79.
- Kawai, G. (2006). Collaborative peer-based language learning in unsupervised asynchronous online environments. In *Creating, Connecting and Collaborating through Computing, 2006. C5'06. The Fourth International Conference on*, pages 35–41. IEEE.
- Malehorn, H. (1994). Ten measures better than grading. *The Clearing House*, 67(6):323–324.
- Orsmond*, P., Merry, S., and Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, 41(3):273–290.
- Pavlus, J. (2010). The game of life. *Scientific American*, 303(6):43–44.
- Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., and Koller, D. (2013). Tuned models of peer assessment in moocs. *arXiv preprint arXiv:1307.2579*.
- Sadler, P. M. and Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational assessment*, 11(1):1–31.
- Shapiro, S. S. and Francia, R. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337):215–216.
- Sterbini, A. and Temperini, M. (2013). Openanswer, a framework to support teacher's management of open answers through peer assessment. In *Frontiers in Education Conference, 2013 IEEE*, pages 164–170. IEEE.
- Tosic, M. and Nejkovic, V. (2010). *Trust-based peer assessment for virtual learning systems*. Springer.
- Wang, Y. and Vassileva, J. (2003). Trust and reputation model in peer-to-peer networks. In *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*, pages 150–157. IEEE.
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, pages 350–362.

Uma Abordagem Colaborativa para Aprendizagem de Programação Orientada a Objetos

Leonardo Davi Pereira Machado¹, Carla Diacui M. Berkenbrock¹,
Ivante Zuchi Siple², Celso M. Hirata³

¹Departamento de Ciência da Computação

²Departamento de Matemática
Universidade do Estado de Santa Catarina (UDESC)

³Divisão de Ciência da Computação
Instituto Tecnológico de Aeronáutica

leonardo.davi.machado@gmail.com, carla.berkenbrock@udesc.br

ivanete.siple@udesc.br, hirata@ita.br

Abstract. *This paper presents a collaborative approach called Collaborative Session Cycle. In this approach, students work using different strategies: individually followed by group work, or directly in the group in order to exchange experiences and learn Object Oriented Programming (OOP). In this study we developed the mobile application entitled CLinClass to support the approach and we conducted a case study to demonstrate the applicability of the approach. The results indicate the potential of the approach to support the learning of computer programming concepts.*

Resumo. *Este artigo apresenta uma abordagem colaborativa chamada de Ciclo de Sessão Colaborativa. Nesta abordagem os alunos trabalham usando diferentes estratégias: individualmente seguida por um trabalho em grupo, ou diretamente em grupo a fim de trocar experiências e aprender Programação Orientada a Objetos (POO). Neste trabalho foi desenvolvida a aplicação móvel intitulada CLinClass para apoiar a abordagem utilizada e foi conduzido um estudo de caso para demonstrar a aplicabilidade da abordagem. Os resultados alcançados indicam o potencial da abordagem para apoiar a aprendizagem dos conceitos de programação de computadores.*

1. Introdução

Nos cursos de computação, as disciplinas de Algoritmos, Linguagem de Programação e Programação Orientada a Objetos (POO) são reconhecidas como difíceis para os alunos e resultam num alto índice de reprovação e evasão [Barcelos et al. 2009, da Silva et al. 2009, Priesnitz Filho et al. 2012].

Pesquisadores e professores têm investigado e aplicado diversas técnicas, processos e estratégias, junto com recentes tecnologias tais como dispositivos móveis [Barcelos et al. 2009], lousa digital [Priesnitz Filho et al. 2012], jogos digitais [Rapkiewicz et al. 2006, da Silva et al. 2014] e ambientes virtuais [Salgado et al. 2013] a fim de facilitar a aprendizagem dos conceitos e técnicas de programação. Nós acreditamos que existe espaço para uma maior investigação e desenvolvimento de recursos

para apoiar a aprendizagem de programação. Conjecturamos que processos devem ser flexíveis para acomodar diferentes formas de estratégias para atingir o objetivo de aprendizado, bem como devem existir ferramentas que deem suporte adequado aos processos.

As organizações em geral exigem que os seus profissionais tenham cada vez mais uma postura colaborativa a fim de que eles contribuam para o desenvolvimento de competências do grupo. Contudo, aprender a programar de forma colaborativa não é uma tarefa fácil, pois de acordo com [Castro et al. 2008] os alunos não estão acostumados a trabalhar em grupo para esta finalidade e envolve diferentes processos, estratégias, técnicas e tecnologias.

Neste trabalho é definida uma abordagem colaborativa de processo que acomoda duas estratégias para a aprendizagem de programação orientada a objetos. Para mostrar a viabilidade do processo, foi desenvolvido um aplicativo para dispositivos móveis e Web, o CLinClass, que pode ser integrado à prática do professor. O CLinClass considera requisitos de formação de grupo, autonomia, estratégias e percepção entre outros requisitos que foram apresentados em trabalhos anteriores [Machado et al. 2015, Machado et al. 2016]. Apresenta-se também um estudo de caso realizado na disciplina Programação Orientada a Objetos do curso de Ciência da Computação na Universidade do Estado de Santa Catarina.

O artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos correlatos; a Seção 3 apresenta a abordagem utilizada para empregar a aprendizagem colaborativa em sala de aula; na Seção 4 é apresentado o aplicativo CLinClass; a Seção 5 relata o estudo de caso realizado; na Seção 6 é realizada uma discussão sobre os resultados alcançados; e por fim na Seção 7 são apresentadas as considerações finais.

2. Trabalhos correlatos

Pesquisadores têm empreendido esforços para aplicar a colaboração no contexto acadêmico com o objetivo de obter melhor desempenho na aprendizagem dos alunos. Na sequência, descrevemos esses trabalhos e alguns propiciaram características importantes para a abordagem proposta, bem como para o desenvolvimento do aplicativo.

[Siqueira and Alcântara 2003] inserem a colaboração no contexto da educação superior como meio de inovar a prática pedagógica. Os autores afirmam que “no papel de agente de transição cultural, cabe ao docente desenvolver as habilidades de trabalhar em grupo, para que o aluno possa contribuir para o grupo e para o seu próprio desenvolvimento”.

[Alcântara et al. 2004] aplicam a aprendizagem colaborativa em três cursos de graduação, utilizando uma abordagem pela qual os alunos realizavam atividades individuais e depois coletivas. Segundo os autores “utilizar os trabalhos individuais para desenvolver o trabalho em grupo acabou sendo uma das etapas mais produtivas do processo”.

[Castro et al. 2008] apresentam um esquema de progressão de aprendizagem de programação em grupo. O esquema foi avaliado em estudo de caso realizado na disciplina de introdução de programação na Universidade Federal do Amazonas. O esquema prevê uma transição gradual da prática individual para a programação em grupo. Os autores afirmam que tal transição é necessária, pois os estudantes não estão acostumados a

trabalhar em grupo. O estudo apontou uma boa aceitabilidade e rendimento dos alunos nas sessões de laboratório.

[Valdivia et al. 2009] aplicaram a aprendizagem colaborativa em duas disciplinas do curso de Ciência da Computação, Introdução a Programação e Fundamentos de Linguagens de Programação. Para apoiar a aprendizagem foi utilizado um aplicativo de PDA (*Personal Digital Assistant*), denominado CANA (*Collaborative Answer Negotiation Activity*), que disponibiliza aos alunos um conjunto de questões de múltipla escolha onde cada aluno deveria apresentar uma resposta para cada questão e depois em grupo, os alunos deveriam negociar uma resposta de consenso do grupo.

[Alvarez et al. 2011] apresentaram o CollPad, um aplicativo para PDA desenvolvido para apoiar a aprendizagem colaborativa. O aplicativo foi utilizado para fins de avaliação em duas disciplinas do curso de Ciência da Computação, Gestão do Conhecimento e Interação Humano Computador. O aplicativo permite dividir os alunos aleatoriamente em pequenos grupos, depois o professor envia uma questão aberta para os alunos. Cada aluno deve responder essa questão em seu dispositivo e depois o grupo deve escolher uma resposta para ser a resposta do grupo. Se desejarem, os alunos podem escrever uma nova resposta de forma coletiva, nesse caso o aplicativo seleciona um membro do grupo para fazer a escrita e os outros membros confirmam se aceitam essa nova resposta. Se não aceitarem a resposta é descartada e outro membro do grupo é selecionado para escrever uma nova resposta, repetindo o processo. Após ter a concordância de todos os membros, a resposta é encaminhada ao professor.

Os trabalhos citados apresentam diferentes abordagens de colaboração para a aprendizagem. Entretanto, algumas características comuns são evidenciadas na forma de organização das atividades realizadas: os alunos trabalham individualmente previamente e posteriormente trabalham em grupo com intuito de chegar a um consenso sobre a solução. Essa característica também foi adotada como uma opção na abordagem de processo que descrevemos na próxima seção.

3. Ciclo de Sessão Colaborativa

A aprendizagem colaborativa possibilita práticas educacionais diferenciadas daquela que, normalmente, o professor expõe um determinado assunto e na sequência submete os alunos a exercícios para realizar a fixação do conteúdo, depois torna a expor um novo conteúdo e novamente submete os alunos a novos exercícios, haja vista que as interações entre pares constituem elementos importantes na aprendizagem colaborativa.

No processo de ensino, a prática realizada pelo professor privilegia, em muitos casos, a repetição e memorização em detrimento da compreensão. Assim, os alunos seguem os exemplos resolvidos em sala de aula para resolver extensas listas de exercícios que seguem determinados padrões [de Oliveira 2014].

No contexto do ensino de POO, essa abordagem, em muitos casos, não favorece a troca de conhecimento e a discussão entre os alunos para encontrar a melhor solução para o problema proposto. Isso ocorre porque o desenvolvimento de um programa, principalmente usando o paradigma de orientação a objetos, exige além de lógica de programação, uma abstração diferente que inclui conceitos de objeto, classe, chamada de método, encapsulamento, composição e herança.

Por outro lado, percebe-se que os alunos quando estudam em grupo têm melhor rendimento na aprendizagem, pois essa estratégia de aprendizagem favorece a ocorrência e resolução de dúvidas, retenção do conhecimento e a satisfação do aluno [Koshino 2010, p. 18]. Dessa forma, propõe-se uma abordagem colaborativa de processo denominada de Ciclo de Sessão Colaborativa. Na abordagem proposta, inicialmente o professor define um problema aos alunos e o professor seleciona uma estratégia de trabalho. Na primeira estratégia, os alunos trabalham individualmente e depois trabalham em grupo. Na segunda estratégia, os alunos trabalham em grupo diretamente, conforme mostra a Figura 1. No trabalho em grupo, os alunos trabalham para chegar a uma solução de consenso.

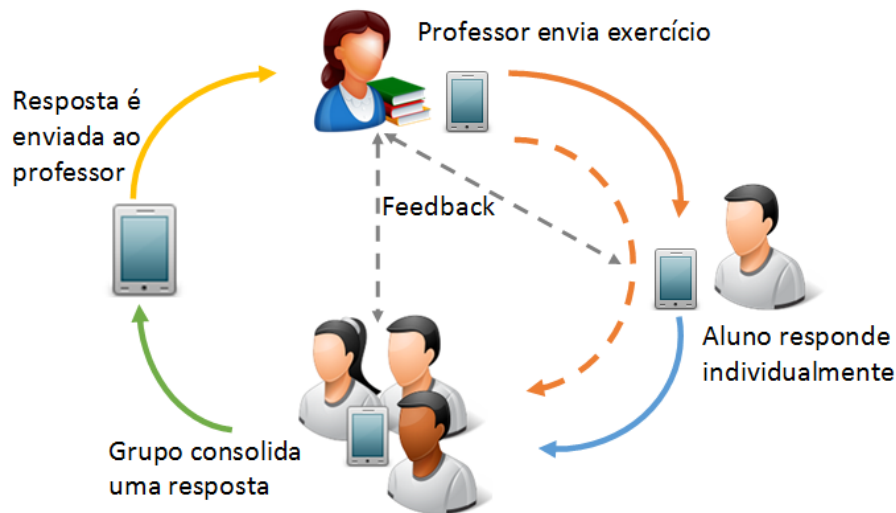


Figura 1. Ciclo de sessão colaborativa

A abordagem proposta fundamenta-se também no conceito de sala de aula invertida e na prática da aprendizagem colaborativa. O conceito de sala de aula invertida diz respeito a inversão que ocorre no método de ensino, onde o aluno estuda o conteúdo teórico antes da aula presencial e em sala de aula tira as dúvidas, discute os pontos indicados pelo professor e desenvolve os exercícios [Piva Jr. and Cortelazzo 2015, Valente 2014]. Do ponto de vista prático, pode-se citar os trabalhos de [Siqueira and Alcântara 2003, Alcântara et al. 2004, Valdivia et al. 2009, Alvarez et al. 2011], conforme apresentados na Seção 2, nas experiências relatadas nessas pesquisas os alunos trabalham individualmente e também em grupo.

Assim, o trabalho individual é executado pelo aluno quando ele estiver em casa, ou qualquer outro ambiente que não seja a sala de aula. Neste momento, o aluno pode consultar os materiais fornecidos pelo professor, livros ou outra fonte de informações que possibilite a execução das atividades propostas. Além disso, é um momento para o aluno refletir os conceitos teóricos e conseguir elaborar uma argumentação para expor aos seus colegas de grupo. Após terem realizado individualmente as atividades os alunos se reúnem em grupos, em sala de aula, para discutir o objeto de estudo a fim de chegar juntos a uma solução para o problema trabalhado. Nesse momento, o aluno tem a oportunidade de expor suas dúvidas, testar suas hipóteses, esboçar suas conjecturas e criar estratégias para resolver o problema. Essa discussão em grupo favorece a troca de conhecimento e experiência para a resolução de problemas. Assim o grupo constrói um consenso sobre

qual é a melhor solução frente ao problema abordado, e esta solução que é encaminhada para o professor como a resposta do grupo.

Nesse processo, o professor pode monitorar o andamento das atividades e, a qualquer momento, pode dar um *feedback* para um aluno ou grupo, fornecendo dicas, tirando dúvidas e motivando os alunos a continuarem a execução da tarefa.

Na aprendizagem colaborativa, o importante é a interação que ocorre durante o trabalho em grupo. Contudo, conforme as pesquisas mencionadas acima, o trabalho individual executado previamente ao trabalho em grupo também tem sua importância, pois pode dar ao aluno recursos para que a interação em grupo seja mais efetiva, permitindo que ele expresse seu pensamento e sua opinião de forma mais clara contribuindo para o aprendizado dos colegas e seu próprio aprendizado.

Em algumas situações, os desafios das atividades não requer os alunos trabalhem individualmente, por exemplo a atividade para se chegar a um consenso sobre conceitos de programação. Nesses casos, o professor deve ter a possibilidade de adotar a estratégia de colocar diretamente a atividade para trabalho dos alunos em grupo.

Pelos motivos apresentados é que o Ciclo de Sessão Colaborativa apresenta a possibilidade de trabalho individual seguido de trabalho em grupo ou trabalho em grupo diretamente.

O ciclo colaborativo apresentado pode ser utilizado mediado por diferentes recursos. Contudo, com a percepção de que os jovens têm facilidade no uso dos dispositivos móveis, o professor deve buscar formas de aproveitar essa tecnologia para melhorar o processo de aprendizagem. Dessa forma, foi desenvolvido o aplicativo chamado CLinClass que utiliza o Ciclo de Sessão Colaborativa a fim de apoiar a aprendizagem colaborativa com os alunos da disciplina Programação Orientada a Objetos. Na próxima seção, o aplicativo é descrito.

4. CLinClass

O aplicativo CLinClass (*Collaborative Learning in Classroom*) foi desenvolvido com o objetivo de apoiar o uso da aprendizagem colaborativa de programação de computadores de nível básico para o Ciclo de Sessão Colaborativa proposto. O aplicativo foi projetado para ser utilizado em *smartphones* e *tablets* que executem os sistemas operacionais Android, iOS ou Windows Phone. O aplicativo implementa um conjunto de requisitos funcionais e requisitos para apoiar a colaboração e a aprendizagem, tais requisitos são descritos nos trabalhos [Machado et al. 2015, Machado et al. 2016].

Como recurso pedagógico o aplicativo disponibiliza três tipos de atividades que são: questões de múltipla escolha; problema de Parsons; problemas abertos. Os problemas abertos são aqueles em que o aluno tem um campo de texto disponível para colocar livremente sua resposta, dependendo do enunciado do problema a resposta pode ser uma explicação ou o código de um algoritmo. O problema de Parsons é uma atividade que se assemelha a um quebra cabeça cujo o objetivo é reordenar os fragmentos de código a fim de chegar no algoritmo correto [Parsons and Haden 2006]. Na Figura 2 é apresentado um exemplo dos três tipos de atividades que o CLinClass disponibiliza.

Para favorecer a colaboração o aplicativo permite a comunicação entre os alunos e professor, por meio da troca de mensagens de texto, possibilitando que eles esclareçam

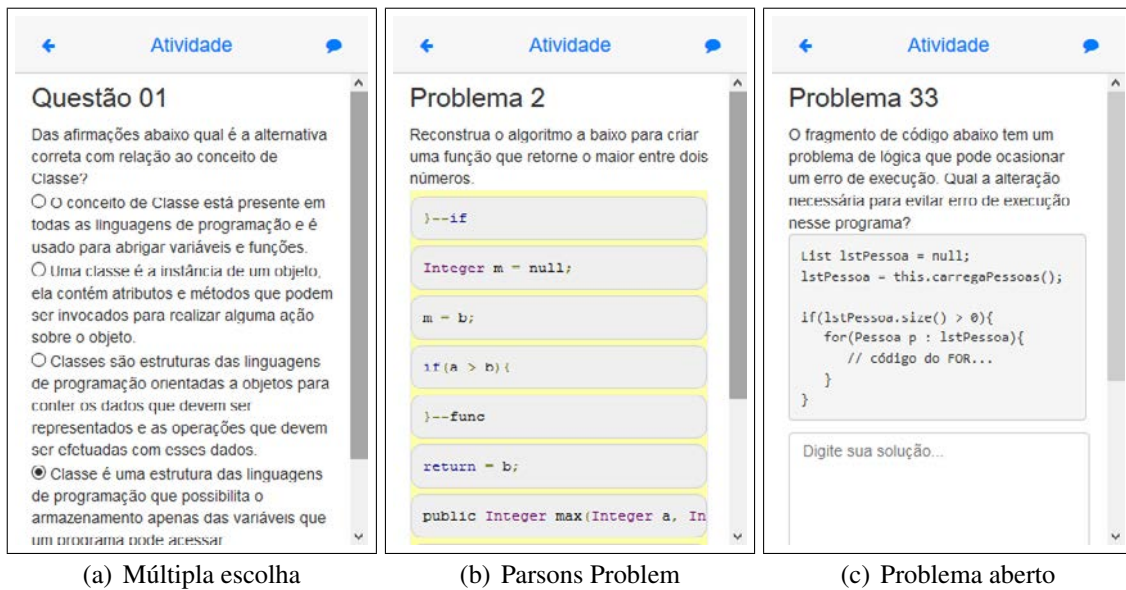


Figura 2. Tipos de atividades

dúvidas com os pares, ou solicitem orientações ao professor. O aplicativo possibilita também a percepção de trabalho dos outros membros do grupo, sendo possível visualizar quais atividades cada membro já realizou e qual a última vez que as mensagens foram lidas. Na Figura 3 são apresentadas as telas de mensagens e de informações do grupo.

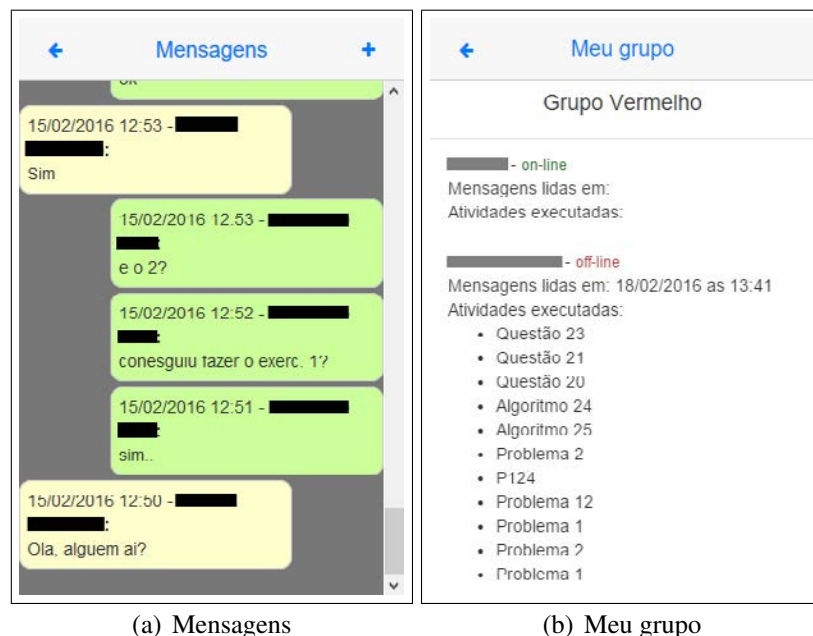


Figura 3. Telas CLinClass

5. Estudo de caso

Para avaliar a utilização da abordagem proposta, Ciclo de Sessão Colaborativa, e o aplicativo CLinClass, foi realizado um estudo de caso com 26 alunos da disciplina

de Programação Orientada a Objetos do terceiro semestre do curso de Ciência da Computação da Universidade do Estado de Santa Catarina. As duas possibilidades de estratégia foram investigadas: a estratégia em que os alunos trabalham diretamente em grupo e a estratégia em que os alunos trabalham individualmente e depois trabalham em grupo.

Inicialmente foi aplicado um questionário com o intuito de obter o perfil dos alunos quanto ao gênero, idade, dispositivos que possuíam e a configuração desses dispositivos. Todos os integrantes da turma são rapazes, com idade média de dezenove anos. Todos os alunos possuem *smartphones* e 24% possuem *tablets*.

Buscou-se também informações sobre a configuração, especialmente o sistema operacional, dos *smartphones* pois o aplicativo provia duas versões para utilização. Na Figura 4 percebe-se a predominância do sistema operacional Android nos dispositivos e os tamanhos de telas de 4" e 5".

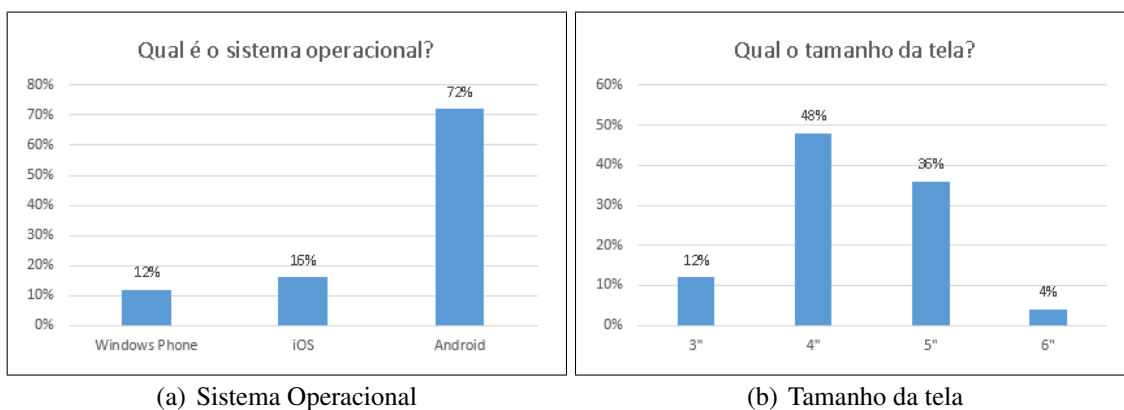


Figura 4. Configuração *smartphones*

Com essas informações pôde ser traçado a organização da prática realizada em sala de aula. Os alunos utilizaram duas versões do aplicativo. Os alunos que possuíam dispositivos com sistema operacional Android, fizeram a instalação do aplicativo no dispositivo, já os alunos que possuíam sistema operacional Windows Phone ou iOS utilizaram a versão Web acessado por meio do navegador web do dispositivo.

5.1. Caso 1

No primeiro encontro os alunos foram divididos em cinco grupos de até seis alunos, os próprios alunos puderam escolher o grupo em que queriam se inserir, pelo aplicativo. O professor da disciplina precisou intervir apenas em dois casos de alunos que estavam chegando na turma nesse mesmo dia e por esse motivo ainda estavam deslocados em relação a turma. Foi utilizada a estratégia dos alunos trabalharem diretamente em grupo para responderem dez questões de múltipla escolha versando sobre os conceitos teóricos de orientação a objetos. Na Figura 2(a) é apresentado um exemplo dessa atividade.

Após executarem todas as atividades, os alunos foram submetidos a um questionário para avaliar a satisfação do aluno com relação ao trabalho realizado e forma de se comunicar com colegas e professor. O questionário foi respondido por 22 alunos. As perguntas constantes no questionário são apresentadas na Tabela 1.

Tabela 1. Questionário Caso 1

Perguntas	Opções de resposta
1) Qual é o seu nível de satisfação em relação as atividades propostas?	Muito insatisfeito, Insatisfeito, Satisfeito e Muito satisfeito
2) Qual é o seu nível de satisfação em relação a forma de se comunicar com os colegas e professor?	
3) Qual é a sua opinião em relação a forma que os grupos foram criados?	Péssimo, Ruim, Bom e Muito bom
4) Qual é a sua opinião em relação a adequação do conteúdo ao tamanho da tela de seu dispositivo?	
5) Como você avalia o processo de construção da resposta do grupo para cada atividade e a geração do consenso?	
6) A ferramenta é flexível, permitindo trabalhar on-line e off-line, além de executar as atividades na ordem que me é mais conveniente?	Discordo plenamente, Discordo, Concordo e Concordo plenamente
7) O aplicativo fornece meios para perceber como está o andamento do trabalho dos colegas?	

Na Figura 5 percebe-se que 73% dos alunos se sentiram satisfeitos com as atividades propostas. Já quanto a forma de se comunicar com colegas e professor, 68% dos alunos se dizem satisfeitos.

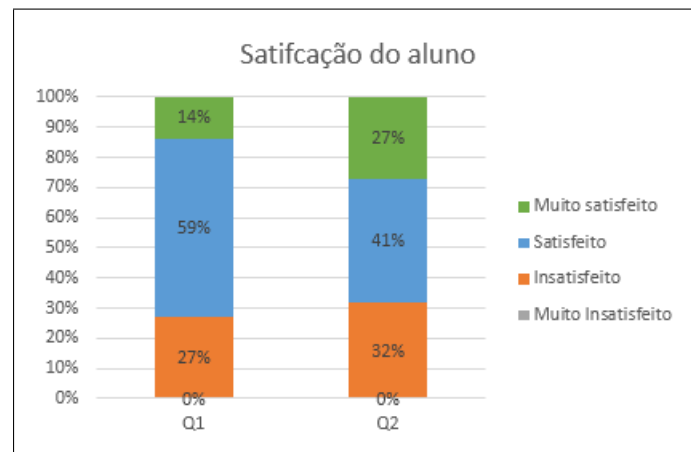


Figura 5. Satisfação do aluno com relação ao trabalho realizado (Q1) e forma de se comunicar com colegas e professor (Q2)

Como mencionado anteriormente os alunos puderam escolher livremente em qual grupo entrar, na Figura 6, percebe-se que 91% dos alunos acharam boa essa forma de criar os grupos. No que diz respeito a adequação do conteúdo ao tamanho da tela, 86% dos alunos acharam que o aplicativo apresenta uma boa adequação ao tamanho do dispositivo. O processo de construção da resposta do grupo foi avaliado como bom por 91% dos alunos que participaram da pesquisa.

As questões 6 e 7 têm por finalidade verificar a visão do aluno, no sentido de concordância ou não das afirmações apresentadas. Percebe-se na Figura 7, 82% dos alunos

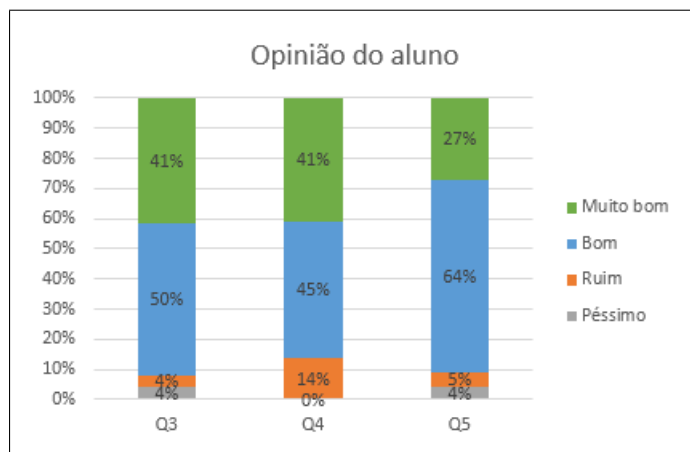


Figura 6. Opinião do aluno sobre a forma de criação dos grupos (Q3), adequação do conteúdo à tela (Q4) e processo de construção da resposta (Q5)

concordam que a ferramenta é flexível, permitindo que as atividades sejam executadas na sequência que seja mais adequada a cada aluno. Além disso, 81% dos alunos concordam que o aplicativo fornece meios para ter a percepção de como está o andamento do trabalho dos colegas.

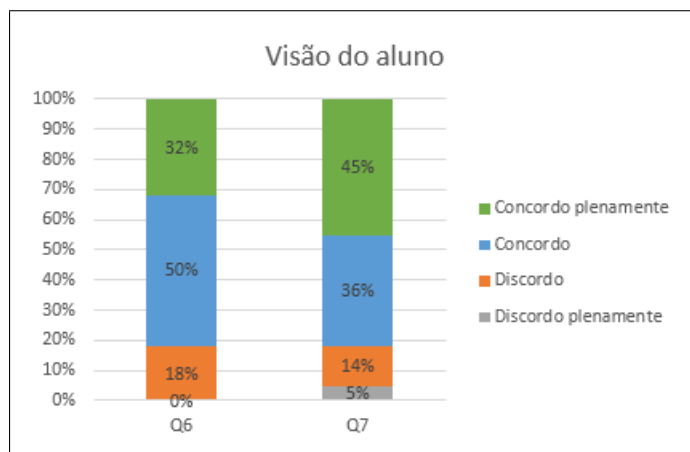


Figura 7. Visão do aluno quanto à flexibilidade da ferramenta (Q6) e percepção do andamento de trabalho (Q7)

5.2. Caso 2

No segundo caso estudado os alunos foram mantidos na mesma formação de grupos do Caso 1. Foram passados 5 problemas abertos para serem resolvidos, a Figura 2(c) apresenta um exemplo desse tipo de problema. Seguindo o conceito de sala de aula invertida, os alunos deveriam resolver os problemas em casa, elaborando individualmente uma solução para cada problema. Na aula seguinte os grupos voltaram a se reunir para gerar um consenso quanto a solução de cada problema.

Após os alunos trabalharem em grupo, o questionário foi aplicado novamente, porém com algumas adaptações, a fim de levantar a opinião dos alunos quanto ao trabalho realizado, a Tabela 2 apresenta as perguntas do questionário.

Tabela 2. Questionário Caso 2

Perguntas	Opções de resposta
1) Qual é o seu nível de satisfação em relação as atividades propostas?	Muito insatisfeito, Insatisfeito, Satisfeito e Muito satisfeito
2) Qual é o seu nível de satisfação em relação ao feedback fornecido pelo professor?	
3) Qual é o seu nível de satisfação em relação a forma de se comunicar com os colegas?	
4) Qual é o seu nível de satisfação em relação a forma de se comunicar com o professor?	
5) Qual é a sua opinião em relação a forma que os grupos foram criados?	Péssimo, Ruim, Bom e Muito bom
6) Qual é a sua opinião em relação a adequação do conteúdo ao tamanho da tela de seu dispositivo?	
7) Como você avalia o processo de construção da resposta do grupo para cada atividade?	
8) Como você avalia a geração do consenso para cada atividade?	
9) A ferramenta é flexível? Permite trabalhar on-line e off-line, além de executar as atividades na ordem que me é mais conveniente?	Discordo plenamente, Discordo, Concordo e Concordo plenamente
10) Trabalhar individualmente e depois em grupo contribuiu para a aprendizagem em relação ao assunto de cada atividade?	
11) O aplicativo fornece meios para perceber como está o andamento do trabalho dos colegas?	

O questionário foi respondido por 19 alunos. Quanto à satisfação, conforme mostra a Figura 8, percebe-se que os alunos ficaram mais insatisfeitos em relação as atividades propostas (53%) e ao *feedback* fornecido pelo professor (52%). No quesito comunicação os alunos apresentaram mais satisfação em relação a forma de se comunicar com colegas (94%) e com o professor (68%).

Quanto a opinião dos alunos, conforme a Figura 9, percebe-se que eles acharam boa a forma de geração do grupo (84%) e a adequação do conteúdo ao tamanho da tela do dispositivo (79%). Os alunos também acharam bom o processo de construção da resposta do grupo (84%) e a geração do consenso (84%).

Quanto a visão dos alunos, a Figura 10 mostra que 79% dos alunos concordam que a ferramenta é flexível. Com relação a sequencia de trabalho individual e trabalho em grupo, 79% dos alunos concordam que isso contribuiu para a aprendizagem. Com relação a percepção ao andamento do trabalho dos colegas, 85% dos alunos concordam que a ferramenta disponibiliza meios para se perceber quais atividades os colegas do grupo já executaram.

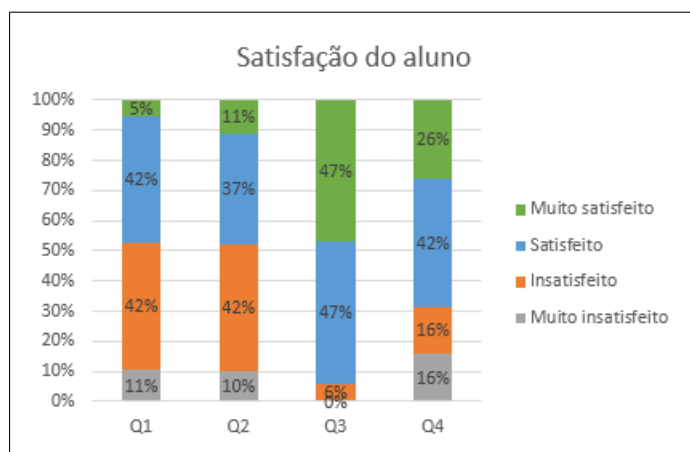


Figura 8. Satisfação dos alunos em relação as atividades propostas (Q1), feedback do professor (Q2), forma de comunicação com colegas (Q3) e forma de comunicação com professor (Q4)

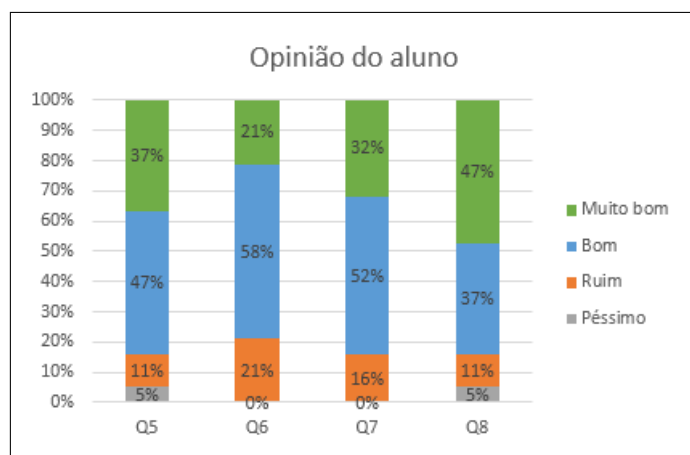


Figura 9. Opinião dos alunos com relação a forma de criação do grupo (Q5), adequação do conteúdo ao tamanho da tela (Q6), processo de construção da resposta (Q7) e forma de geração de consenso (Q8)

6. Discussão

Conforme apresentado na seção anterior, os dois casos estudados têm algumas diferenças. A primeira diz respeito a forma como os alunos trabalharam, no Caso 1 os alunos realizaram apenas o trabalho em grupo enquanto que no Caso 2 foi utilizado o conceito de sala de aula invertida, onde os alunos tiveram que trabalhar individualmente e depois em grupo. A segunda diferença é em relação ao tipo de atividade executada, no Caso 1 foram aplicadas questões de múltipla escolha e no Caso 2 foram trabalhados problemas abertos onde o nível de dificuldade de cada atividade era maior do que no Caso 1.

Analisando o resultado dos questionários percebe-se que a satisfação dos alunos em relação as atividades propostas apresentou diminuição de 26%, caindo de 73% no Caso 1 para 47% no Caso 2. Essa diminuição na satisfação dos alunos pode estar relacionada ao nível de complexidade que os problemas abertos apresentavam, pois eles foram elaborados para fazer com que os alunos pesquisassem novas informações e assim elaborassem uma solução para o problema. Os resultados do Caso 1 evidenciam que a al-

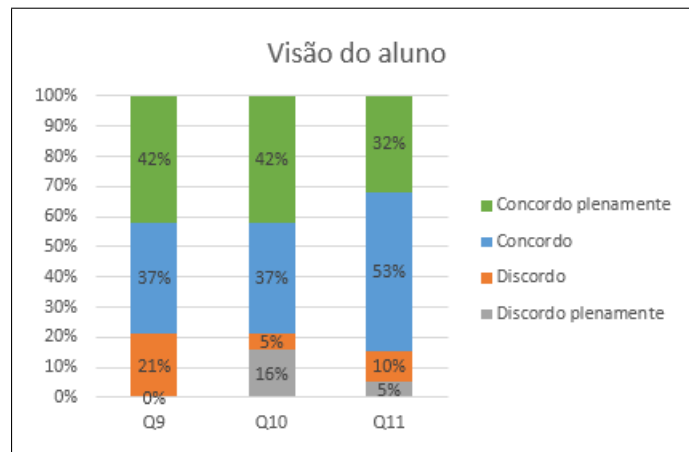


Figura 10. Visão dos alunos com relação à flexibilidade da ferramenta (online/offline) (Q9), sequência trabalho individual e trabalho em grupo (Q10) e percepção provida pela ferramenta das atividades realizadas pelos colegas (Q11)

ternativa de estratégia onde os alunos trabalham diretamente em grupo tem sua utilidade.

Antes de ser realizado o Caso 2 o professor da disciplina forneceu um *feedback* aos alunos em relação as atividades executadas anteriormente, passando quais foram os erros e os acertos de cada grupo. Porém, ao serem questionados sobre a satisfação em relação ao *feedback* fornecido pelo professor percebe-se um alto índice de insatisfação (52%). Esse fato precisa ser melhor estudado para compreender se a insatisfação dos alunos é em relação a forma que o *feedback* foi fornecido, ou seja, mensagem de texto enviado pelo professor por meio do aplicativo, ou se os alunos ficaram insatisfeitos com o desempenho que eles obtiveram nas atividades propostas.

O mecanismo para o fornecimento do *feedback* é a mensagem de texto, que é o mesmo utilizado para a comunicação entre aluno-aluno e aluno-professor. Quanto a comunicação, 94% dos alunos se dizem satisfeitos com a forma de se comunicar com os colegas e 68% se dizem satisfeitos com a forma de comunicação com o professor.

Quanto a flexibilidade da ferramenta, na visão dos alunos, 79% concordam que a ferramenta é flexível (online/offline) depois de realizado o Caso 2. Esse percentual foi de 82% no Caso 1. Quanto a percepção dos trabalhos realizados pelos colegas de grupo, no Caso 2, 85% dos alunos concordam que o aplicativo fornece meios para saber quais atividades os membros do grupo já realizaram. Com o trabalho individual realizado em casa é importante que os alunos consigam saber como está o andamento do trabalho dos colegas para poderem trocar informações sobre a resolução dos problemas. O objetivo não é que um aluno passe inteiramente a resposta para o outro, mas sim que consiga compartilhar os conhecimentos necessários para que o colega consiga por si mesmo elaborar uma solução para o problema trabalhado.

Conforme já mencionado, no Caso 2 os alunos primeiro resolveram individualmente os 5 problemas propostos e depois, em grupo, chegaram a um consenso da melhor solução para cada problema. Ao serem questionados se o trabalhado individual e depois em grupo contribuiu para o aprendizado do assunto abordado em cada atividade, 79% dos alunos se manifestaram concordando que houve uma contribuição no aprendizado.

7. Considerações finais

Este artigo apresentou uma abordagem colaborativa chamada de Ciclo de Sessão Colaborativa, na qual os alunos recebem atividades/problemas para serem resolvidos individualmente e depois é realizada uma discussão em grupo para criar um consenso sobre a melhor solução para o problema trabalhado. Nessa abordagem, duas alternativas de estratégia podem ser realizadas: os alunos podem trabalhar diretamente em grupo ou os alunos podem trabalhar individualmente e depois trabalhar em grupo. Foi apresentado também o aplicativo CLinClass, desenvolvido para apoiar a utilização do ciclo na disciplina Programação Orientada a Objetos, bem como o estudo de caso realizado para avaliar a utilização do ciclo e do aplicativo CLinClass.

Comparando o aplicativo CLinClass com os aplicativos CANA e CollPad, apresentados na seção 2, percebe-se que os três aplicativos fornecem a possibilidade de trabalhar com atividades de questões de múltipla escolha e com a interação de trabalho individual seguido de trabalho em grupo. Contudo, o CLinClass amplia as possibilidades do docente, permitindo os alunos trabalharem com problemas abertos e com os Parsons Problems, bem como permitindo o trabalho diretamente em grupo. Dessa forma o docente pode, conforme o conteúdo ministrado em aula, escolher quais tipos de atividades utilizar e qual a estratégia a ser aplicada para poder obter um melhor resultado na aprendizagem dos alunos.

Os resultados alcançados até o momento indicam que a abordagem utilizada tem o potencial de favorecer o aprendizado dos alunos. Contudo, o estudo está em andamento, sendo necessário realizar mais interações com os alunos para eles se acostumarem com a abordagem utilizada e com o trabalho colaborativo.

Referências

- Alcântara, P. R., Siqueira, L. M. M., and Valaski, S. (2004). Vivenciando a aprendizagem colaborativa em sala de aula: experiências no ensino superior. *Revista Diálogo Educacional*, 4(12):169 – 188.
- Alvarez, C., Alarcon, R., and Nussbaum, M. (2011). Implementing collaborative learning activities in the classroom supported by one-to-one mobile computing: A design-based process. *The Journal of System and Software*, 84:1961–1976.
- Barcelos, R., Tarouco, L., and Bercht, M. (2009). O uso de mobile learning no ensino de algoritmos. *RENOTE - Revista Novas Tecnologias na Educação*, 7(2):327–337.
- Castro, T., Castro, A., and Fuks, H. (2008). Aprendendo a programar em grupo. In *Sistemas Colaborativos, 2008 Simpósio Brasileiro de*, pages 45–54.
- da Silva, I., SILVA, I., and SANTOS, M. (2009). Análise de problemas e soluções aplicadas ao ensino de disciplinas introdutórias de programação. *IX Jornada de Ensino, Pesquisa e Extensão - JEPEX 2009*.
- da Silva, T. S. C., de A. R. Tedesco, P. C., and de Melo, J. C. B. (2014). A importância da motivação dos estudantes e o uso de técnicas de engajamento para apoiar a escolha de jogos no ensino de programação. In *XXV Simpósio Brasileiro de Informática na Educação - SBIE 2014*.

- de Oliveira, F. L. (2014). A produção de conhecimento matemático acerca de funções de duas variáveis em um coletivo de seres-humanos-com-mídias. Master's thesis, Universidade Federal de Ouro Preto.
- Koshino, P. (2010). A aprendizagem e as interações em um treinamento a distância. Master's thesis, Universidade de Brasília.
- Machado, L. D. P., Berkenbrock, C. D. M., and Siple, I. Z. (2016). Desenvolvimento de aplicativos para aprendizagem colaborativa apoiada por dispositivos móveis: uma análise dos requisitos. In *COMPUTER ON THE BEACH 2016*.
- Machado, L. D. P., Berkenbrock, C. D. M., Siple, I. Z., and Hirata, C. M. (2015). Utilizando dispositivos móveis para apoiar a aprendizagem colaborativa baseada em problemas. In *XII Simpósio Brasileiro de Sistemas Colaborativos - SBSC 2015*.
- Parsons, D. and Haden, P. (2006). Parson's programming puzzles: A fun and effective learning tool for first programming courses. In *Proceedings of the 8th Australasian Conference on Computing Education - Volume 52, ACE '06*, pages 157–163, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.
- Piva Jr., D. and Cortelazzo, A. L. (2015). Sala de aula invertida, ambientes de aprendizagem e educação online: a junção de três métodos para potencialização do ensino de algoritmos. In *IV Congresso Brasileiro de Informática na Educação - CBIE 2015*.
- Priesnitz Filho, W., Abegg, I., and Simonetto, E. (2012). Uma abordagem diferenciada no ensino de algoritmos através da utilização de uma lousa digital. *GEINTEC - Gestão, Inovação e Tecnologias*, 2(2).
- Rapkiewicz, C. E., Falkembach, G., Seixas, L., dos Santos Rosa, N., da Cunha, V. V., and Klemann, M. (2006). Estratégias pedagógicas no ensino de algoritmos e programação associadas ao uso de jogos educacionais. *RENOTE - Revista Novas Tecnologias na Educação*, 4(2).
- Salgado, N., Castro, T., and Castro, A. (2013). Aprendizagem colaborativa de programação com scratch e opensimulator. In *X Simpósio Brasileiro de Sistemas Colaborativos - SBSC 2013*.
- Siqueira, L. M. M. and Alcântara, P. R. (2003). Modificando a atuação docente utilizando a colaboração. *Revista Diálogo Educacional*, 4:1–13.
- Valdivia, R., Nussbaum, M., and Ochoa, S. (2009). Modeling a collaborative answer negotiation activity using ims-based learning design. *IEEE Transactions on Education*, 52:375–384.
- Valente, J. A. (2014). A comunicação e a educação baseada no uso das tecnologias digitais de informação e comunicação. *Revista UNIFESO – Humanas e Sociais*, 1(1):141 – 166.

Sistema multiagentes para recuperação e análise de informações como suporte à mediação pedagógica em ambientes colaborativos de aprendizagem

Sabrina Siqueira Panceri^{1,2}, Crediné Silva de Menezes^{1,3}

¹Programa de Pós-Graduação em Informática – Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514, Goiabeiras – Vitória – ES – Brasil – Caixa Postal 29075-910

²Campus Guarapari – Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo
Estrada da Tartaruga, s/n, Muquiçaba – Guarapari – ES – Brasil – Caixa Postal 29215-090

³Faculdade de Educação – Universidade Federal do Rio Grande do Sul
Av. Paulo Gama, s/n, Farroupilha – Porto Alegre – RS – Brasil – Caixa Postal 90046-900

sabrina.panceri@ifes.edu.br, credine@gmail.com

Abstract. *This paper presents a proposal of a Multi-Agent System, which aims to recover and analyze information through the mining of texts produced in virtual teaching and learning environments. The Educational Architecture Thesis Debate presented as favorable scenario the application of SMA. They are elicited pedagogical mediations within the context of a Thesis Debate. The development of functional prototype of SMA and its application to a specific case are described. Pedagogical mediations that are aided by the use of SMA are teased. Finally, the results achieved in the final working considerations are exposed.*

Resumo. *Este trabalho apresenta uma proposta de um Sistema Multiagente (SMA), que visa a recuperação e análise de informações, através da mineração dos textos produzidos em ambientes virtuais de ensino e aprendizagem. A Arquitetura Pedagógica Debate de Teses apresenta-se como cenário favorável a aplicação do SMA. São elicítadas as mediações pedagógicas dentro do contexto de um Debate de Teses. O desenvolvimento do protótipo funcional do SMA e sua aplicação a um caso específico são descritos. As mediações pedagógicas que são auxiliadas pelo uso do SMA são descritas com detalhes. Por fim, os resultados alcançados e as considerações finais do trabalho são expostos.*

1. Introdução

Diante dos avanços tecnológicos ocorridos na última década, a internet se apresenta como parte integrante do cotidiano da atual sociedade. Dessa forma, as diversas formas de comunicação adaptaram-se para sua integração e utilização como meio para propagação e compartilhamento de informações. Assim também as práticas educacionais incorporaram a internet como meio para promovê-las, difundindo cada vez mais a educação à distância com base nesse novo aporte computacional.

Com o objetivo de oferecer suporte à educação a distância, diversas soluções computacionais foram desenvolvidas, dentre elas destacam-se os Ambientes Virtuais de Ensino e Aprendizagem (AVEA), que são sistemas colaborativos concebidos para

favorecer a mediação de aprendizagens. Nos AVEA atuais, observa-se a presença de variadas ferramentas de autoria colaborativa, tais como blogs, *wikis* e fóruns.

As interações praticadas nas ferramentas de autoria são, em grande maioria, feitas através da produção de textos, seja na elaboração de uma resposta a uma pergunta, nas postagens realizadas nos fóruns, ou mesmo na produção de uma redação de forma conjunta com outros participantes. Este volume de textos produzidos subsidia um nicho de pesquisa ainda pouco explorado, a Recuperação de Informações em Textos produzidos dentro do contexto educacional.

Diante desse cenário, propõe-se o desenvolvimento de um Sistema Multiagente (SMA) que servirá como uma camada de análise e que poderá ser acessado por qualquer AVEA por meio de interfaces de comunicação predefinidas. O SMA é responsável por processar, por intermédio de seus agentes, os textos produzidos nas ferramentas de autoria, gerando seis tipos de resultados. Os resultados gerados pelo SMA servem para apoiar decisões e mediações pedagógicas.

Este artigo está organizado da seguinte maneira: Na Seção 2 apresenta-se brevemente a fundamentação teórica utilizada para a construção da proposta de solução que é detalhada na Seção 3. Na Seção 4 mostra-se a avaliação da proposta e sua aplicação a um caso real, bem como a análise dos resultados obtidos. Na Seção 5 estão os trabalhos correlatos ao contexto e proposta aqui apresentados e por fim, na Seção 6 estão as considerações finais deste trabalho.

2. Fundamentação Teórica

Esta seção apresenta os fundamentos teóricos utilizados para a construção de uma solução computacional inteligente, baseada na arquitetura de Sistemas Multiagente, para dar suporte tecnológico à aplicação de Arquiteturas Pedagógicas, especificamente, para aplicação da Arquitetura Pedagógica Debate de Teses. Bem como um breve resumo sobre a Recuperação de Informações em Textos que dá suporte a construção do núcleo de processamento dos textos que serão analisados pela solução proposta.

2.1. Arquiteturas Pedagógicas

Para subsidiar a formulação de propostas de uso das tecnologias digitais na educação, em [Carvalho et al. 2005] é apresentado o conceito de Arquiteturas Pedagógicas, que destinam-se a aplicação de abordagens pedagógicas baseadas na Epistemologia Genética de Jean Piaget [Ramos-Chiarottino 1998].

A proposta dos autores [Carvalho et al. 2005] tem por objetivo organizar as interações em ambientes virtuais de aprendizagem, adotando protocolos específicos, com a finalidade de auxiliar a construção de conhecimento dos indivíduos de forma colaborativa.

3.1.1 A Arquitetura Pedagógica Debate de Teses

Uma especialização das Arquiteturas Pedagógicas é apresentada em [Nevado et al. 2011] com a designação de Arquitetura Pedagógica Debate de Teses (APDT). A APDT propõe a construção de conhecimento de forma colaborativa por meio de interações orientadas. Uma visão geral sobre o fluxo de interações realizadas na APDT é apresentado pela Figura 1.

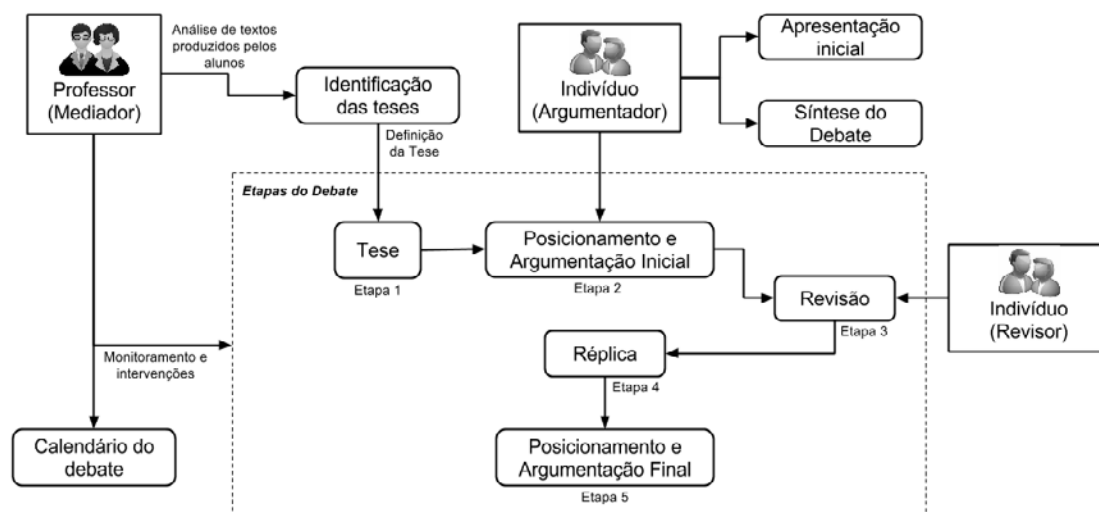


Figura 1: Fases de um Debate de Teses

Na APDT (Figura 1) o professor, no papel de mediador, pela definição do calendário (prazos) de cada etapa do debate, pelo monitoramento das interações realizadas pelos alunos, e pela Definição da Tese (Etapa 1) através da análise de textos produzidos anteriormente sobre um determinado tema. Uma tese é uma afirmação sobre qualquer assunto, sobre a qual o aluno pode concordar ou discordar. Finalizada as definições iniciais, o professor monitora as interações e realiza, se necessário, pequenas intervenções a fim de auxiliar os alunos a desenvolver da melhor maneira os textos de cada etapa do debate.

Os alunos são responsáveis pelas produções das demais etapas da APDT, iniciando sua participação no debate através da indicação de seu Posicionamento e Argumentação Inicial sobre a tese (Etapa 2). O posicionamento deve ser favorável ou contrário à tese em discussão, e o argumento deve ser desenvolvido de forma lógica e com base nos conhecimentos prévios e que justifique o posicionamento indicado.

A Revisão (Etapa 3) é realizada por outros dois alunos – revisão por pares. Estes alunos devem analisar a consistência dos argumentos, tendo por base verificar se há evidências que sustentem a argumentação em revisão. O aluno na etapa de revisão assume o papel de revisor.

Na etapa de réplica (Etapa 4), o aluno volta ao papel de argumentador, e deve responder aos levantamentos feitos por seus revisores. As Etapas 3 e 4 preparam o aluno para a reescrita de seu posicionamento e argumentação, que é concretizado na Etapa 5 – Posicionamento e Argumentação Final.

Nesta última etapa, o aluno indica seu posicionamento sobre a tese e reescreve sua argumentação, levando em conta as interações realizadas. Ao final da aplicação da APDT, espera-se que o aluno tenha consolidado seus conhecimentos prévios sobre o assunto debatido.

As interações nesta Arquitetura Pedagógica ocorrem mediante a autoria de textos, geralmente, textos curtos. Para que sejam aplicadas mediações pedagógicas pertinentes a cada etapa, o professor (mediador) deve ler e compreender todos os textos produzidos. Considerando uma turma com dez (10) alunos, e seguindo a Figura 1, para

cada tese proposta o professor terá dez (10) posicionamentos e argumentações iniciais, vinte (20) revisões, vinte (20) réplicas, e dez (10) posicionamentos e argumentações finais, num total de sessenta (60) textos para analisar, compreender e identificar as características intrínsecas de cada etapa.

Uma abordagem possível para dar suporte computacional a Arquitetura Pedagógica Debate de Teses é a configuração de um fórum com regras específicas. Ao desenvolver a APDT utilizando a ferramenta fórum como suporte, cada tese pode ser definida na forma de um tópico do fórum (etapa 1) e as respostas realizadas por cada aluno ao tópico equivalem a etapa 2 de posicionamento e argumentação inicial. Nesta resposta, o aluno deve manifestar sua concordância ou discordância sobre a afirmação proposta, e em seguida fazer sua argumentação de forma lógica. Para cada resposta enviada, são designados dois revisores, que apresentam problematizações ao autor da resposta, caracterizando assim a etapa 3 – revisão. Os argumentadores podem responder a cada uma das revisões (etapa 4) e finalmente devem produzir uma nova resposta à tese (etapa 5 – posicionamento e argumentação final). Uma visualização dessa abordagem está na Figura 2.

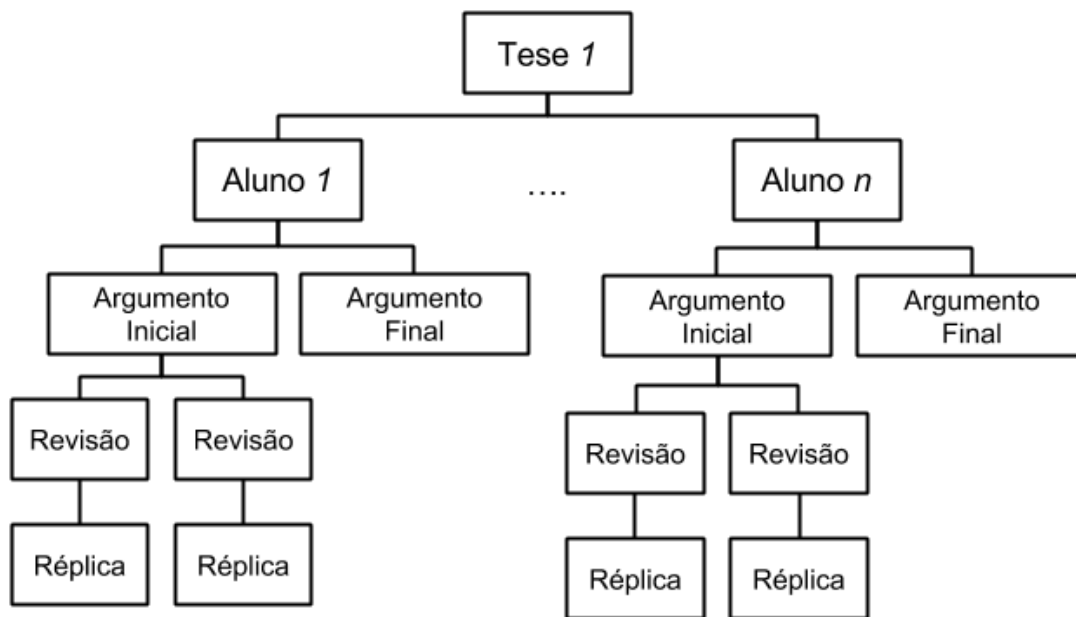


Figura 2: Árvore de interações para produção de um Debate de Teses utilizando a ferramenta Fórum

Contudo, a prática da APDT com o uso da ferramenta fórum fica prejudicada uma vez que a organização dos textos, o acesso as produções e demais interações ficam encadeadas ao tópico proposto, dificultando o monitoramento e possíveis intervenções do professor. Diante disso, um ambiente próprio foi desenvolvido para dar suporte a aplicação da APDT. Neste ambiente¹ o fluxo de interações segue a Figura 1. O objetivo da presente pesquisa é apresentar uma colaboração ao ambiente em funcionamento, voltada para apoiar o professor nas tarefas subsequentes a realização do debate, fornecendo suporte computacional às tarefas de análise dos textos para subsidiar a

¹ O ambiente pode ser acesso em <http://www.pead.faced.ufrgs.br/sites/cms/debate/>. Acesso em 25 de março de 2016.

aplicação das mediações pedagógicas.

2.2. Recuperação de Informações em Textos

A Recuperação de Informações (RI) tem como foco a busca de informações em documentos ou textos, e a classificação destas buscas a fim de retornar os documentos mais importantes em uma coleção, dada uma quantidade limitada de informações a serem resgatadas. De acordo com [Manning et al. 2008], “RI busca encontrar material de natureza não estruturada que satisfaçam uma necessidade de informação, a partir de grandes coleções”.

Outro importante aspecto das pesquisas em RI é a utilização do Modelo Espaço Vetorial para representação computacional das coleções que serão analisadas [Salton et al. 1975]. Este modelo representa os documentos de uma coleção como vetores em um espaço multidimensional. Para calcular a frequência ou ocorrência dos termos que compõem os documentos da coleção, aplica-se a medida TF-IDF - *Term Frequency x Inverse Document Frequency*. Nesta medida, a importância de um termo é proporcional à sua frequência de ocorrência em cada documento da coleção, e inversamente proporcional ao número de documentos em que o termo aparece. Esse cálculo é feito de acordo com a Fórmula 1:

$$TF-IDF(w, d) = TermFreq(w, d) \cdot \log(N / DocFreq(w)) \quad (1)$$

Onde $TF-IDF(w, d)$ é a frequência de cada palavra no documento, N é o número de todos os documentos da coleção, e $DocFreq(w)$ é o número de documentos que contém o termo w .

Dessarte, o vetor que representa cada documento contém todos os termos utilizados em todos os documentos da coleção. Sendo assim, quanto maior a coleção, maior a dimensionalidade do espaço vetorial que a representa. Para isso, a aplicação de técnicas de pré-processamento na coleção que será analisada pode melhorar a qualidade de representação vetorial da mesma, uma vez que há a redução nos termos que compõem a coleção. No contexto deste trabalho, utilizou-se as seguintes técnicas:

- Remoção de termos frequentes (*stopwords*): Essa técnica visa a remoção de termos comuns, geralmente pertencentes as classes gramaticais dos artigos ou preposições e de palavras pequenas (com menos de 3 caracteres). Essas palavras, geralmente, não possuem relevância semântica e não interferem na recuperação de informações nos textos.
- Capitalização (*case-folding*): Essa técnica troca os caracteres escritos em maiúsculo para minúsculo, para que palavras como por exemplo “Interessante” e “interessante” sejam consideradas as mesmas.
- Remoção de caracteres acentuados: A troca de caracteres acentuados (ã, é, õ) pelo seu correspondente sem acentuação (a, e, o) auxilia no processo de análise dos termos e na padronização dos termos que serão analisados.
- Redução ao radical (*stemming*): O objetivo do *stemming* é reduzir palavras que se encontram em formas derivadas para o seu radical de formação (*stem*), através da remoção dos afixos (prefixos ou sufixos).

O modelo de representação do corpus é o “*bag-of-words*”, no qual todas as

palavras que compõem o documento são consideradas suas características. Este modelo não considera a ordem em que as palavras aparecem no texto.

Segundo [Manning et al. 2008] a Clusterização é a forma mais comum de aprendizado não supervisionado, ou seja, não há interferência humana no processo. Utiliza-se algoritmo de clusterização *K-Means* para particionar o corpus em um conjunto pré-definido de grupos de semelhança (*clusters*). Os agrupamentos são feitos com base na semelhança encontrada entre os textos. Por padrão o *K-Means* utilizada a Distância Euclidiana para calcular a semelhança entre os textos, porém [Feldman e Sanger 2006] afirma que a Similaridade de Cossenos (*Cosine Similarity*) é a medida de similaridade mais utilizada para realizar o agrupamento de textos, dessa forma, utiliza-se neste trabalho uma adaptação do *K-Means*, na qual a Similaridade de Cossenos é empregada para calcular a similaridade dos textos, e assim criar os grupos de semelhança.

Para criar os dicionários de contexto e relações semânticas, utiliza-se a Análise Semântica Latente (LSA, do inglês *Latent Semantic Analysis*) [Dumais 2004]. A LSA é uma técnica estatística que visa a construção de matrizes de relações entre os documentos e seus termos, considerando que termos que estão num mesmo documento podem conter determinada relação semântica. Apesar de considerar a co-ocorrência dos termos num documento, LSA utiliza o modelo de representação “*bag-of-words*”, e não leva em consideração a sequência em que os termos aparecem no documento. Para que a aplicação do LSA seja bem-sucedida, é necessário que seja feito o treinamento da base que será usada para a análise de similaridade dos textos.

2.3. Sistemas Multiagente

De acordo com [Reis 2003], Sistemas Multiagente (SMA) são “sistemas computacionais em que dois ou mais agentes interagem ou trabalham em conjunto de forma a desempenhar determinadas tarefas ou satisfazer um conjunto de objetivos”.

Segundo [Wooldridge 2009] um agente é “um sistema de computador que está situado em algum ambiente, e que é capaz de ação autônoma neste ambiente, a fim de cumprir os objetivos para os quais fora projetado”.

Agentes inteligentes podem ser classificados em dois tipos: Agentes Inteligentes Reativos e Agentes Inteligentes Cognitivos. Os Agentes Inteligentes Reativos são aqueles que reagem mediante a condição percebida em um cenário real ou fictício. Os Agentes Inteligentes Cognitivos são aqueles que consideram uma série de fatores para realizar uma reação complexa e bem elaborada. Além disso, um agente inteligente deve possuir atributos que o possibilitem interagir com o meio, são eles os atributos sensores e atributos atuadores. Atributos sensores permitem que os agentes inteligentes capturem informações do meio sobre sua influência. Atributos atuadores possibilitam a aplicação das reações dos agentes inteligentes.

Os agentes inteligentes que compõem um SMA possuem uma área de atuação ou influência, que pode ser compartilhada com outros agentes inteligentes, e que fazem parte do SMA como um todo, como mostra a Figura 3

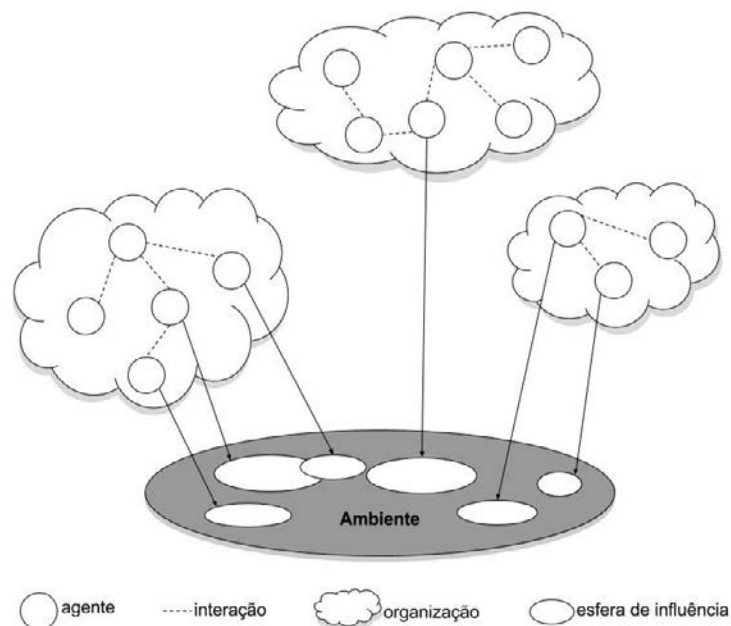


Figura 3: Estrutura típica de um Sistema Multiagente [Jennings 2000]

Segundo [Coppin 2004] em muitas situações, um SMA pode ser composto apenas por agentes reativos, uma vez os agentes não existem de forma isolada e na prática, os agentes são compostos por um conjunto de regras, logo suas ações são limitadas. Ainda assim, SMA são considerados sistemas complexos, tanto para sua modelagem, quanto para seu desenvolvimento.

3. Contexto do Problema

No contexto dos cursos de graduação ou pós-graduação realizados na modalidade a distância, as interações realizadas entre professores e alunos são feitas através da troca de mensagens, sendo o fórum um dos principais instrumentos apoiadores dessa comunicação. A forma mais comum de uso da ferramenta Fórum é o suporte ao esclarecimento de dúvidas (pergunta-resposta). Contudo, outras atividades de interação podem ser promovidas através dessa ferramenta como por exemplo a discussão sobre um assunto determinado, o desenvolvimento de histórias colaborativas, o esclarecimento de dúvidas diversas sobre o conteúdo das disciplinas. Em todas essas possibilidades há uma vasta produção de pequenos textos.

Os textos produzidos nas ferramentas de autoria dos AVEA são analisados pelos professores e/ou tutores, com o objetivo de oferecer o suporte necessário para apoiar a construção de conhecimento dos alunos. A análise destes textos demandam a dedicação de tempo e atenção de seus analisadores, que segundo [Mason e Grove-Stephensen 2002] essa atividade consome cerca de 30% do tempo de trabalho dos analisadores. Essa quantidade de tempo é expressiva e minimizá-la incentiva a criação de ferramentas computacionais que realizem a análise e recuperação de informações para auxiliar o professor neste processo de análise.

Buscando contribuir para a superação das dificuldades que um professor enfrenta

neste esforço de mediação, propõe-se o desenvolvimento de um Sistema Multiagente (SMA) que visa extrair informações relevantes através da análise dos textos produzidos. A arquitetura geral da proposta é apresentada na Figura 4. Nela o SMA é composto por seis agentes cada um deles produzindo resultados específicos.



Figura 4: Arquitetura Geral da Proposta

O SMA é responsável apenas pela análise dos textos, a extração dos textos contidos nas ferramentas de autoria deve ser feita de acordo com as configurações do AVEA. A interface de comunicação com o SMA é através do envio de arquivos de texto, no qual cada linha do arquivo será considerada uma produção distinta a ser analisada.

De acordo com o resultado desejado pelo usuário, os agentes são acionados e executam processos comuns à área de Recuperação de Informações em Textos. A seguir descreve-se as responsabilidades de cada agente previsto na Figura 4.

- Agente de Pré-processamento: responsável por coordenar os processos de pré-processamento textual, que são: remoção de *stopwords*; remoção de acentuação; aplicação de *stemming*; remoção de *links*;
- Agente de Sintaxe: Responsável por verificar a formação das orações que compõem os textos, classificar as palavras de acordo com sua ordem nas orações através da utilização do Projeto Floresta Sintática² [Afonso et al. 2001].
- Agente de Produção Textual: Responsável por realizar a análise de similaridade entre os textos, através do algoritmo de Similaridade de Cossenos. Utiliza do Agente de Contexto, a Análise de Semântica Latente (LSA) [Dumais 2004] para criar dicionários de contexto com base nos próprios textos que serão analisados, e a partir desses dicionários calcular as relações semânticas entre os termos e destacar os termos mais importantes da coleção.
- Agente de Clusterização: Responsável por aplicar algoritmos de clusterização, como o *K-Means*. Os grupos são definidos com base na análise de similaridade entre os textos.
- Agente de Análise textual: Responsável por identificar se os textos produzidos contém elementos novos, a fim de contribuir com a construção de conhecimentos dos outros membros do AVA.
- Agente de Contexto: Responsável por verificar se os textos produzidos são de

2 <http://www.linguateca.pt/Floresta/>

determinado tema. O tema deve ser informado através do envio de materiais didáticos para criação dos dicionários de contexto. Essa criação é realizada através da aplicação da Análise Semântica Latente (LSA).

Os agentes serão acionados conforme a solução desejada. Segue a descrição das soluções.

- Grupos de similaridade: O usuário deve fornecer a quantidade de grupos em que ele deseja dividir sua base de análise. Produto da aplicação dos Agentes de Pré-processamento e Clusterização, os resultados apresentados são grupos de textos semelhantes.
- Destaque de ideias: O Agente de Sintaxe em parceria com os agentes de Pré-Processamento, Contexto e Produção Textual, analisa os textos a fim de encontrar frases que ressaltam informações novas sobre o tema e as mostra para o usuário.
- Análise de plágio: O Agente de Produção Textual separa os textos que possuem um nível de semelhança muito alto e os apresenta ao professor, para que este o analise e confira se há uma relação de cópia/plágio entre os textos.
- Análise de contribuições: Verificar se as informações dos textos possuem relevância e apresentam novos elementos a fim de colaborar com a construção de conhecimento dos membros do AVEA. Utiliza o Agente de Sintaxe.
- Identificação de polaridade: Destaca as afirmações positivas ou negativas identificadas nos textos analisados. Essa análise utiliza os Agentes de Contexto e Produção Textual.
- Identificação de dúvidas: Verifica expressões de dúvidas, mesmo que não tenham sido escritas como uma. Utiliza os agentes de Sintaxe, Produção Textual e Contexto.

4. Avaliando o SMA proposto

Para apoiar a utilização e aplicação da Arquitetura Pedagógica Debate de Teses, foi desenvolvido um ambiente virtual de aprendizagem próprio apresentado em [Nevado et al. 2011]. Com a aplicação deste protótipo ao longo dos anos, diversas mediações pedagógicas foram identificadas, contudo aplicá-las sem o devido suporte computacional mostrou-se inviável. As mediações elicitadas são [Panceri e Menezes 2014a, Panceri e Menezes 2014b]:

1. Agrupar participantes por posicionamento e argumentação inicial;
2. Fazer comentários sobre a argumentação inicial para orientar o participante;
3. Verificar se as revisões feitas pelo participante são diferentes e relevantes;
4. Verificar se não houve comparações entre argumentações revisadas;
5. Conferir se as réplicas respondem aos questionamentos feitos nas revisões;
6. Verificar se existem novos argumentos que reforcem o posicionamento;
7. Identificar os “pontos importantes” para orientar a fase de reflexão;
8. Elaborar síntese das argumentações finais para realizar o *feedback*;

9. Agrupar argumentações finais semelhantes;
10. Verificar qual o grau de influência das revisões e réplicas na argumentação final;
11. Identificar quais teses podem ser trabalhadas a partir das argumentações finais;
12. Analisar o padrão de respostas de cada participante;
13. Verificar a evolução das argumentações;
14. Avisar sobre o prazo de envio das etapas;
15. Verificar se houve plágio nas revisões, réplicas e argumentações finais;
16. Verificar se houve cordialidade nas redações.

Aplicando o SMA para auxiliar nas mediações pedagógicas identificadas, considera-se como *corpus* de análise os textos produzidos em cada fase do Debate de Teses. Dessa forma, para cada *corpus* retirado de uma das etapas do debate, teremos resultados e análises diferenciadas, vinculados às mediações pedagógicas identificadas para cada etapa do debate.

Neste artigo, o foco é a apresentação da ferramenta desenvolvida para dar suporte as mediações pedagógicas “Agrupar participantes por posicionamento e argumentação inicial” e “Agrupar argumentações finais semelhantes”. Agrupar os alunos com base na análise do posicionamento e argumentação inicial serve para facilitar a distribuição de revisores e para identificar quais alunos iniciaram o DT com o nível de conhecimentos prévios semelhantes. Dessa forma, o professor mediador poderá sugerir materiais didáticos para auxiliar nas próximas fases do DT, e indicar revisores que estejam em grupos diferentes, para estimular que as revisões sejam feitas sob pontos de vistas diferentes. No caso do agrupamento com base no posicionamento e argumentação final, o professor terá os grupos de alunos que chegaram a conclusões semelhantes sobre o assunto discutido.

4.1. Instanciando o SMA para um Debate de Teses

Para comprovar a eficácia do SMA proposto, uma instância dele foi implementada e a ferramenta **Grupos de Similaridade** foi aplicada sobre os textos produzidos num debate de teses realizado no ano de 2013. O debate em questão foi realizado como parte do curso de extensão promovido na XIV Maratona de Empreendedorismo da UFRGS [Michels 2014]. As teses propostas foram:

- (1) “O mundo mudou, o perfil do consumidor mudou, e, conseqüentemente, a forma de se planejar negócios também mudou”
- (2) “O plano de negócios aumenta a probabilidade de sucesso dos negócios”
- (3) “A visão macro do negócio é desenvolvida com o auxílio do modelo de negócios e a visão micro com o auxílio do plano de negócios. Logo, o empreendedor precisa ter visão do todo e, se necessário, a visão de cada elemento do negócio”

A amostra analisada pela ferramenta é composta pelos 21 alunos que concluíram todas as etapas do debate. Os experimentos foram realizados para divisão do total de alunos em 3, 4 e 5 grupos, e os resultados dos agrupamentos são apresentados nas Tabelas 1, 2 e 3.

A Tabela 1 apresenta os resultados obtidos sobre o debate desenvolvido a partir da análise da Tese (1).

Tabela 1. Resultados dos experimentos realizados para agrupar as argumentações iniciais produzidas sobre a Tese (1)

Divisão em 3 grupos				
Grupos	G1	G2	G3	
Qtde alunos	6	9	6	
Similaridade média	75%	51%	65%	

Divisão em 4 grupos				
Grupos	G1	G2	G3	G4
Qtde alunos	3	9	4	5
Similaridade média	61%	49%	68%	78%

Divisão em 5 grupos					
Grupos	G1	G2	G3	G4	G5
Qtde alunos	3	3	4	8	3
Similaridade Média	54%	85%	80%	52%	61%

Analisando os resultados apresentados pela Tabela 1, observa-se que:

- ◆ A similaridade média entre os membros dos grupos formados com a formação de 5 grupos é maior que 50%.
- ◆ Para a formação de 4 grupos, houve uma variação maior entre a quantidade de membros entre os grupos.
- ◆ O G2 formado com a divisão do total em 3 grupos possuem uma quantidade maior de alunos e uma baixa similaridade média entre seus membros.

A Tabela 2 apresenta os resultados obtidos sobre o debate desenvolvido a partir da análise da Tese (2). Analisando os resultados apresentados pela Tabela 2, observa-se que:

- ◆ Nas três divisões realizadas, um grupo destaca-se por ter uma similaridade média entre seus membros de 92%
- ◆ Na divisão em 3 grupos, o grupo G1 possui os mesmos alunos do grupo G3 formado com divisão em 4 grupos. Ambos os grupos possuem uma similaridade média entre seus membros de 92%.
- ◆ A distribuição mais balanceada, em relação ao número de alunos por grupos, deu-se na divisão em 5 grupos de alunos.

Tabela 2. Resultados dos experimentos realizados para agrupar as argumentações iniciais produzidas sobre a Tese (2)

Divisão em 3 grupos					
Grupos	G1	G2	G3		
Qtde alunos	4	10	7		
Similaridade média	92%	82%	46%		

Divisão em 4 grupos					
Grupos	G1	G2	G3	G4	
Qtde alunos	9	5	4	3	
Similaridade média	81%	59%	92%	28%	

Divisão em 5 grupos					
Grupos	G1	G2	G3	G4	G5
Qtde alunos	3	3	6	5	4
Similaridade Média	44%	92%	85%	75%	41%

Tabela 3. Resultados dos experimentos realizados para agrupar as argumentações iniciais produzidas sobre a Tese (3)

Divisão em 3 grupos					
Grupos	G1	G2	G3		
Qtde alunos	4	12	5		
Similaridade média	82%	66%	47%		

Divisão em 4 grupos					
Grupos	G1	G2	G3	G4	
Qtde alunos	1	12	4	4	
Similaridade média	0%	66%	81%	58%	

Divisão em 5 grupos					
Grupos	G1	G2	G3	G4	G5
Qtde alunos	4	4	8	4	1
Similaridade Média	65%	81%	64%	58%	0%

A Tabela 3 apresenta os resultados obtidos sobre o debate desenvolvido a partir da análise da Tese (3). Analisando os resultados apresentados pela Tabela 3, observa-se que:

- Para a divisão em 4 ou 5 grupos, os agrupamentos apresentam uma argumentação destoante das outras, que fica isolada em grupo de 1 membro.
- Na divisão em 4 e 3 grupos, os grupos G2 contém os mesmos 12 alunos em ambas distribuições.

Sobre a aplicação da ferramenta Grupos de Similaridade, destaca-se:

(a) Os grupos formados com a aplicação do Agente de Clusterização, parte central da ferramenta Grupos de Similaridade, são suficientes para auxiliar o professor a fornecer *feedback* para os grupos, reduzindo assim o seu esforço na produção do *feedback* individual para todos os alunos;

(b) A utilização dos agrupamentos para colocar em prática os princípios Piagetianos de “desequilíbrios”, indicando revisores que não estão num mesmo grupo a fim de estimular o acesso a pontos de vistas diferentes sobre uma mesma tese, dessa forma o revisor irá contrapor os argumentos do outro aluno. Por consequente, na réplica, o aluno revisado terá acesso a contrapontos sobre sua argumentação, o que causará certo desequilíbrio, dessa forma o aluno revisado terá que buscar por novos conhecimentos para conseguir criar sua réplica, estabelecendo novamente um estado de equilíbrio sobre o assunto.

5. Trabalhos relacionados

Os seguintes trabalhos podem ser associados a esta proposta por utilizarem os textos produzidos em Ambientes Virtuais de Ensino e Aprendizagem (AVEA) para recuperação de informações necessárias para responder outras questões.

Em [Iandoli et al. 2014] é apresentada a ferramenta *Debate Dashboard*. Essa ferramenta é integrada ao sistema *Cohere*. No artigo são apresentados argumentos e evidências empíricas que mostram que as ferramentas para ambientes colaborativos com suporte computacional para visualização de argumentações podem ser desenvolvidas para apresentar mapas conceituais sobre as interações realizadas pelo participante com o objetivo de fornecer a cada um a possibilidade de acompanhar o seu desenvolvimento dentro do debate, quais foram suas contribuições, quais relações foram realizadas com os argumentos feitos por ele. Importante ressaltar, que o *Debate Dashboard* auxilia na construção de deliberações coletivas, ou seja, fornece suporte para retirada de afirmações relevantes para o tema em debate que sejam suficientes para criar um entendimento coletivo do que foi discutido.

O trabalho de [Azevedo et al. 2011] apresenta a aplicação do software *MineraFórum*, que realiza uma análise qualitativa das mensagens enviadas por alunos em fóruns de discussão. O *MineraFórum* utiliza técnicas de mineração de textos, e tem como foco principal analisar a relevância das mensagens enviadas pelo aluno em relação ao tema em discussão.

A relação entre os trabalhos de [Iandoli et al. 2014] e [Azevedo et al. 2011] e o proposto neste artigo está na aplicação de técnicas reconhecidas para extração de informações em textos produzidos em AVEA. [Azevedo et al. 2011] analisa textos produzidos em Português e utiliza técnicas de pré-processamento de texto para reduzir o *corpus* de análise. Já [Iandoli et al. 2014] utiliza técnicas de Recuperação de Informações para tratar os textos e criar mapas conceituais que apresentam as

contribuições dos alunos no debate. Apesar de nossa proposta ser aplicada a um ambiente específico para coleta dos dados e validação, a utilização da arquitetura Multiagentes nos possibilitou modelar a camada de inteligência de forma flexível para que a mesma possa ser acoplada em qualquer AVEA, desde que os protocolos de comunicação sejam respeitados e apenas textos sejam enviados para processamento.

6. Considerações Finais

Este trabalho apresentou a modelagem conceitual de um Sistema MultiAgente (SMA) capaz de auxiliar o processo de mediação pedagógica em Ambientes Virtuais de Ensino e Aprendizagem, por meio da análise e recuperação de informações sobre os textos produzidos durante as interações.

A fim de validar a proposta, um protótipo funcional de parte do SMA, a ferramenta Grupos de Similaridade, composta pelo Agente de Clusterização e pelo Agente de Pré-processamento, foram implementados. O Agente de Pré-processamento é composto pelos processos responsáveis por preparar a corpus de análise que serão analisados pelo Agente de Clusterização a fim de criar os grupos solicitados. O Agente de Clusterização é composto pelo algoritmo de clusterização *K-Means* combinado com a técnica de similaridade de textos – Similaridade de Cossenos.

O protótipo desenvolvido foi aplicado ao Ambiente Virtual de Aprendizagem Debate de Teses, com o objetivo de coletar dados de interações reais realizadas em 2013. Os textos analisados foram produzidos por 21 alunos, e esses escreveram seus posicionamentos e argumentados para fundamentá-los sobre 3 teses distintas. Dessa forma, foram analisados um total de 63 textos. Os resultados encontrados com a aplicação do protótipo desenvolvido mostram-se satisfatórios e estão em consonância com as Mediações Pedagógicas identificadas e apresentadas na seção 4, pois duas das 16 mediações elicitadas são plenamente apoiadas pelo protótipo, comprovando a importância da construção de ferramentas de suporte para prática da mediação pedagógica aliada a educação à distância.

Como proposta de trabalhos futuros, pretende-se realizar a implementação completa do SMA e utilizá-lo como uma camada de inteligência integrada a dois ou mais AVEA para coleta e análise dos dados e posterior validação.

Referências

- Afonso, S., Bick, E., Haber, R. e Santos, D. (2001). Floresta sintá(c)tica: um 'trebank' para o português. *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística (APL)*, páginas 533-545.
- Azevedo, B. F. T., Behar, P. A., e Reategui, E. B. (2011). Análise das mensagens de fóruns de discussão através de um software para mineração de textos. In *Anais do 22º Simpósio Brasileiro de Informática na Educação*, volume 1, páginas 20–29.
- Carvalho, M. J. S., Nevado, R. N., e Menezes, C. S. d. (2005). Arquiteturas pedagógicas para educação à distância: Concepções e suporte telemático. In *Anais do 16º Simpósio Brasileiro de Informática na Educação*, volume 1, páginas 351–360.
- Coppin, B. *Artificial Intelligence Illuminated*. Jones & Bartlett Learning, 2004.

- Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1):188–230.
- Feldman, R. e Sanger, J. (2006). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York, NY, USA.
- Iandoli, L., Quinto, I., De Liddo, A., e Buckingham Shum, S. (2014). Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. *Int. J. Hum.-Comput. Stud.*, 72(3):298–319.
- Jennings, N. R. *On agent-based software engineering*. Artificial Intelligence, v. 117, n. 2, p. 277-296, 2000.
- Manning, C. D., Raghavan, P., e Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- Mason, O. e Grove-Stephensen, I. (2002). Automated free text marking with paperless school. Loughborough University.
- Michels, A. B. (2014). Do fazer ao compreender no contexto da educação a distância: Uso de arquiteturas pedagógicas no processo de empreender. Dissertação de Mestrado, Programa de Pós-Graduação em Educação – UFRGS.
- Nevado, R. N., Menezes, C. S. d., e Vieira Jr, R. R. M. (2011). Debate de teses – uma arquitetura pedagógica. In *Anais do 22º Simpósio Brasileiro de Informática na Educação*, páginas 820–829.
- Panceri, S. S. e Menezes, C. S. d. (2014a). Alpes: Um sistema multiagente para análise de produções textuais no contexto de um debate de teses. In *Anais do 25º Simpósio Brasileiro de Informática na Educação*, volume 1.
- Panceri, S. S. e Menezes, C. S. d. (2014b). O suporte computacional como auxílio a mediação pedagógica em um debate de teses. In *TISE'14*, volume 1.
- Ramozzi-Chiarottino, Z. (1988). *Psicologia e Epistemologia Genética de Jean Piaget*. São Paulo: EPU.
- Reis, L. P. Coordenação em Sistemas Multi-Agente: Aplicações na Gestão Universitária e Futebol Robótico. 2003. Tese de Doutorado, Faculdade de Engenharia da Universidade do Porto. Portugal.
- Salton, G., Wong, A., e Yang, C.-S. (1975). *A vector space model for automatic indexing*. Communications of the ACM.
- Wooldridge, M. *An introduction to multiagent systems*. John Wiley & Sons, 2009.

O Uso de Storytelling para Apoiar o Compartilhamento de Conhecimento

Roberta Salim¹, Angélica F.S. Dias¹, Juliana B. S. França¹, Bruno B. Lage¹, Marcos R.S. Borges¹

¹Pós-graduação em Informática – Universidade Federal do Rio Janeiro (UFRJ)
Caixa Postal 68.530 – 21.941-590 – Rio de Janeiro – RJ – Brasil

robertasalim@hotmail.com, brunobarboza@gmail.com, {angelica, juliana.franca, mborges}@ppgi.ufrj.br

Abstract. *Public schools in Rio de Janeiro have recently adopted an approach, which is essentially communicative, concerning teaching English. The use of group storytelling supported by technology is an effective way of collecting data related to moments shared by a group. When such stories are formalized, the tacit knowledge can also be observed, making reality known. By using a tool called Sofia, specially made for collecting information through stories, students are supposed to work collaboratively so that a certain result can be achieved.*

Resumo. *Escolas públicas do município do Rio de Janeiro adotou recentemente uma abordagem essencialmente comunicativa para o ensino de língua inglesa. A contação de histórias apoiada pela tecnologia é uma forma bastante eficiente de se colher dados relacionados a algo que foi vivenciado por um grupo de pessoas. Quando essas histórias são formalizadas, o conhecimento tácito também pode ser observado, fazendo com que seja possível uma maior aproximação da realidade dos eventos. Através da Sofia, uma ferramenta especialmente criada para a recuperação de informações através de histórias, busca-se chegar a um resultado a partir da colaboração de um grupo de alunos.*

1. INTRODUÇÃO

Há aproximadamente cinco anos, o município do Rio de Janeiro, Brasil, oferece aulas de língua inglesa para crianças do primeiro ao quinto ano do ensino fundamental. Anteriormente, esse segmento não possuía acesso ao estudo de uma língua estrangeira. Contudo, devido aos eventos esportivos da Copa do Mundo de 2014 e dos Jogos Olímpicos de 2016, ambos no Rio de Janeiro (Brasil), a necessidade de preparar os alunos para tais eventos tornou-se evidente e o foco seria a comunicação entre eles e os estrangeiros. Na perspectiva da obtenção de bons resultados, foram confeccionados livros para atender a esse público e adotou-se uma abordagem essencialmente comunicativa. Nessa abordagem, o professor assume o papel de facilitador do processo de aprendizagem, a fim de que seus alunos usem a língua alvo como meio de comunicação [Larsen-Freeman 2000].

Neste novo momento, as aulas expositivas seriam totalmente substituídas por uma abordagem diferenciada que exigiria dos docentes não somente um extenso conhecimento da modalidade escrita da língua, mas também da oral. Para tanto, o quadro de professores

foi ampliado e estes foram submetidos a provas escritas e orais a fim de que a fluência do docente pudesse ser considerada.

Uma vez que a comunicação é o principal objetivo da abordagem adotada, o uso de ferramentas tecnológicas torna-se bastante pertinente para dar suporte a mesma.

Os avanços tecnológicos não somente criam novos hábitos, como também trazem novos conceitos e levam a novas formas de pensamento. Dentre tantas mudanças, o uso de computadores e dispositivos móveis em redes diversas destacam-se por ter promovido uma das mais radicais transformações sociais e comportamentais de todos os tempos.

A internet causou uma verdadeira revolução na vida dos cidadãos, conhecida como: Revolução da Internet, Revolução Digital, dentre outras [Fuks 2012]. Embora rápidas, essas mudanças ocorreram de forma contínua e foram marcadas por determinados períodos conforme os avanços tecnológicos. A Geração de *Baby Boomers*, compreendendo os que nasceram entre o fim da guerra mundial e o início dos anos de 1960, foi impactada pela televisão e valorização da juventude. A Geração X, abrangendo os que nasceram entre o início da década de 1960 e o fim da década de 1970, presenciou a popularização do computador pessoal, videogames e o início da Revolução da Internet. Já a Geração Y, ou Geração Digital, e a Geração da Internet, são as primeiras a crescer em um ambiente digital. Essas gerações, diferentes das anteriores, não precisaram se adaptar aos novos hábitos trazidos pelos computadores em rede visto que desde cedo já possuem contato com a internet.

O público alvo destacado neste artigo, jovens das últimas gerações mencionadas, apresentam um perfil bastante diferente das gerações anteriores, uma vez que têm acesso à informação com facilidade e rapidez; diferentemente do que ocorria há um tempo atrás quando os alunos precisavam ir até bibliotecas para acessar determinado conteúdo. Basta uma rápida pesquisa na Internet para encontrarem diferentes informações e conhecerem pontos de vista distintos sobre um mesmo assunto. Com isso, os jovens passaram a adotar uma postura investigativa diante das informações recebidas e a forma de aprender foi inteiramente modificada. Os jovens deste século trabalham o tempo todo conectados, compartilhando informação, atuando em conjunto e aprendendo com a colaboração de outras pessoas. Por isso, o método tradicional até então utilizado por muitas escolas precisou ser revisto e novas abordagens têm sido adotadas.

O ensino de línguas estrangeiras é fundamental para essa nova geração. Embora muitos alunos, sobretudo os da rede pública brasileira, não tenham condições financeiras de visitar outros países, eles certamente se comunicarão com pessoas de outras nações através da Internet. Para que essa comunicação fosse plenamente estabelecida, seria necessário que os usuários conhecessem o idioma uns dos outros. Entretanto, a fim de facilitar esse processo, a língua inglesa, por distintos fatores, acabou por ser a principal forma de comunicação entre habitantes de diferentes nações; ela passou a ser considerada língua franca.

Aproximadamente cinco anos se passaram desde que a nova abordagem foi adotada na disciplina de inglês nas escolas públicas da rede municipal do Rio de Janeiro. Contudo, os alunos precisaram adaptar-se a essa nova forma de aprender, uma vez que estavam habituados a um ensino mais tradicional. As aulas passaram a ser mais lúdicas, a interação e cooperação tornaram-se hábitos diários, diferentes mídias foram constantemente adotadas e a tecnologia passou a ser parte do conteúdo. Segundo Kumar (1996) ambientes educacionais centrados em promover interação entre os

indivíduos, tendem a buscar estratégias de colaboração para estimular o grupo a buscar melhores resultados no processo de aprendizagem. Destaca-se neste cenário a evolução do conhecimento baseado no trabalho coletivo, e não somente no individual. Voltado para os aspectos sócio técnicos, vinculados não somente a tecnologia, como também às relações humanas [Nonaka et al 2009].

Os alunos analisados, que cursavam o quinto ano, fizeram parte das primeiras turmas a terem acesso a uma língua estrangeira desde o 1º do ensino fundamental e foram os primeiros a experimentar a abordagem comunicativa na rede pública de ensino do município. Dito isso, viu-se a necessidade de obter-se um retorno por parte desses alunos a fim de que, através de relatos construídos por eles, conheçam-se os impactos causados pela abordagem no ensino da língua. Logo, a questão principal deste estudo é verificar a influência da abordagem comunicativa na aprendizagem da língua inglesa em escolas da rede municipal do Rio de Janeiro.

A fim de que os dados necessários fossem capturados de modo eficaz, utilizou-se a técnica de contação de histórias em grupo (*group storytelling*) que se originou de um dos meios mais utilizados para transmissão de conhecimentos, práticas e percepções pela humanidade: histórias. A história pode ser definida como uma narrativa – uma passagem de conhecimento, ou uma cadeia de eventos dito ou escrito em prosa ou verso [Valle 2003]. O *group storytelling* reforça a construção de conhecimentos e o compartilhamento de ideias através de narrativas com a formalização do conhecimento tácito. No caso dos alunos, buscam-se as narrativas de suas experiências e vivências. O uso da tecnologia serve como facilitador do processo de levantamento dos dados, mas também torna o momento mais significativo, prazeroso e memorável para as crianças.

2. A INFLUÊNCIA TAYLORISTA/FORDISTA NA EDUCAÇÃO

Para melhor compreendermos a situação da maioria das escolas no Brasil, faz-se necessário um breve relato histórico de tendências que influenciaram ou vêm influenciando as escolas.

De acordo com o estilo de produção em massa de Henry Ford, o trabalhador assumia tarefas simples e repetitivas que não exigiam muita qualificação de seus operários, pois não era necessário conhecer todo o processo no qual estavam envolvidos. O modelo de produção de carros bem como a forma como eram geridos influenciaram não somente a economia, mas a educação.

Segundo Ferreira (2013), as tendências pedagógicas conservadoras em todas as suas modalidades foram originadas pelos princípios tayloristas/fordistas. A divisão entre pensamento e ação, a organização e fragmentação dos conteúdos, a memorização como evidência de aprendizagem, a repetição, dentre outros, são apenas alguns exemplos da herança fordiana. As crianças não eram incitadas a pensar, contextualizar ou criticar. Não havia espaço para a criatividade. O conhecimento era apenas depositado nos alunos, que eram vistos como seres passivos, como máquinas que não possuem atividades cognitivas.

Embora o Fordismo date do início do século XX, algumas escolas, apesar de terem incorporado a tecnologia às suas práticas pedagógicas, ainda adotam um padrão instrucionista. De acordo com esse padrão, o computador é visto como o tradicional professor, que tem a função de fornecer informações e transmiti-las aos alunos. Apesar da tecnologia, o aluno ainda possui um papel inerte diante da máquina. Segundo Valente (2015), “a abordagem que usa o computador como meio para transmitir a informação ao

aluno mantém a prática pedagógica vigente. Na verdade, o computador está sendo usado para informatizar os processos de ensino que já existem. Isso tem facilitado a implantação do computador na escola, pois não quebra a dinâmica por ela adotada.”

Embora muitos esforços tenham sido feitos para que haja uma total mudança na educação, sabe-se que esse caminho é gradativo e muitas vezes lento. Algumas escolas no município do Rio de Janeiro têm procurado adotar novas posturas a fim de que seus alunos sejam amplamente atendidos, como o caso da abordagem comunicativa aplicada ao ensino de idiomas. Ainda assim, existe um longo caminho a ser percorrido.

2.1 A ABORDAGEM COMUNICATIVA

Embora a abordagem comunicativa seja inovadora no ensino de línguas, a mesma data da década de 1970: época em que os educadores começaram a questionar se estavam atingindo seus objetivos da forma correta. Alguns observaram que embora os alunos pudessem produzir frases gramaticalmente corretas em determinada lição, eles não eram capazes de se comunicar genuinamente fora de sala de aula. Outros perceberam que ser capaz de se comunicar exige mais que dominar estruturas linguísticas, isso requer que os alunos sejam capazes de executar funções, tais como: prometer, convidar, rejeitar convites, entre outros, de acordo com um determinado contexto [Larsen-Freeman 2000].

Seguem alguns dos princípios fundamentais da abordagem observados por Larsen-Freeman (2000) são: (a) a língua deve ser usada, sempre que possível, de forma autêntica e em um contexto real; (b) ser capaz de entender as intenções do falante ou do escritor é parte da competência comunicativa; (c) a língua alvo é um meio de comunicação e não objeto de estudo; (d) uma mesma função pode ter diferentes formas linguísticas; o processo de comunicação deve ser enfatizado em vez da forma; (e) alunos devem ser coesos e coerentes; jogos são importantes; o significado deve ser negociado; (f) trabalhar em pequenos grupos maximiza a prática comunicativa; (g) os alunos devem ter a oportunidade de expressar opiniões e ideias; (h) erros são tolerados e vistos como parte do desenvolvimento; o principal papel do professor é estabelecer situações a fim de que haja comunicação; (i) a interação motiva relações cooperativas entre os alunos e cria a oportunidade de negociação de significado; (j) o contexto social é essencial; aprender a usar estruturas da língua adequadamente é parte importante do processo comunicativo; (k) o professor atua como facilitador e como conselheiro durante as atividades; (l) na comunicação, o falante deve ter escolha não somente de o que dizer, mas de como dizer; a gramática e o vocabulário devem partir de funções, contexto situacional e dos papéis dos interlocutores; e (m) os alunos devem ter oportunidades de ouvir a língua como é usada na comunicação autêntica.

Se considerarmos as ideias de Ford e Taylor, e a abordagem comunicativa, fica evidente que há um imenso contraste entre o que era utilizado nas escolas municipais do Rio de Janeiro e o que tem sido realizado atualmente na disciplina de inglês. Outra importante questão refere-se à lacuna entre as metodologias abordadas por outras disciplinas e a abordagem sugerida especialmente para o ensino de língua inglesa; o que não indica que a forma de ensinar inglês seja superior a abordagem utilizada em outras matérias e vice-versa.

A comunicação deve ser autêntica em vez de inventada ou falseada. Dito isso, a tecnologia é fundamental, principalmente quando se leva em conta a internet, que é uma rica fonte de conhecimentos e um grande meio de comunicação entre pessoas de

diferentes nacionalidades. Portanto, sempre que possível, deve-se fazer uso da mesma a fim de que os seus benefícios possam ser explorados e que os alunos sejam amplamente atendidos.

2.2 A COLABORAÇÃO E O GROUP STORYTELLING

Estudos apresentam os “Sistemas Colaborativos”, “Groupware”, ou Computer Supported Cooperative Work (CSCW) como algumas nomenclaturas que designam o uso de tecnologias para o trabalho em grupo. Embora esses termos existam mesmo antes da web, eles estão relacionados a sistemas computacionais que apoiam a colaboração [Fuks 2012]. A última, CSCW, refere-se também a uma área de estudo que tem por base um modelo cujas dimensões apoiam um trabalho em grupo, o modelo 3C.

Ele é dividido em 3 (três) dimensões: comunicação, coordenação e cooperação. A comunicação refere-se à ação uma vez que, enquanto se comunicam, as pessoas negociam e tomam decisões. Já a coordenação, refere-se aos conflitos e organização das atividades. Muitas vezes, durante a cooperação, há a necessidade de renegociação, o que demanda comunicação e coordenação para a reorganização de tarefas. Nesse cenário, a questão da percepção é bastante relevante já que o indivíduo obtém feedback de suas ações e *feedthrough* das ações dos colegas [Fuks 2012]. Segundo Oliveira (2014), a diversidade de conhecimento, ou seja, diferentes níveis de conhecimento compartilhados em sala de aula, pode melhorar o processo de aprendizagem.

A fim de que o conhecimento seja compartilhado eficazmente é necessário que haja um foco interpretativo comum e que seu contexto seja estabelecido e compartilhado entre todos os membros. Logo, se um grupo de pessoas participou de alguma situação comum, esse contexto real pode auxiliar a entender melhor determinada situação. Esse conhecimento é chamado de tácito já que, em grande parte, ele está na mente das pessoas e não é registrado em documentos formais. Sendo assim, faz-se necessário capturá-lo e organizá-lo a fim de que possa ser (re)utilizado. Contudo, extrair e explicitar esse conhecimento não é uma tarefa simples. Para tal, pode-se utilizar a criação de histórias em grupo como uma ferramenta de registro poderosa nesse contexto compartilhado [Santoro et al 2006, 2013].

O *Group Storytelling* [Laporti et al 2009 e Lemos et al 2012] permite que as pessoas trabalhem em conjunto colaborativamente e pode ser apoiado pelas mais diversas tecnologias. As histórias podem ser uma abordagem poderosa para representar e fazer com que ideias multidimensionais sejam conhecidas; podem comunicar informação e emoção, tanto explícitas — Exemplo: referenciando documentos ou trechos destes — quanto tácitas; e também permitem determinar contextos. Logo, quando os sistemas colaborativos trabalham com histórias, podem ajudar as pessoas a relatarem melhor os eventos e a apoiarem a extração de informações.

Quando uma história é contada, é fundamental que ela seja contextualizada, pois uma ação ou evento ganha significado de acordo com seu contexto. Ações individuais não são suficientes para que haja total entendimento de determinada história. O contexto e conhecimento contidos em histórias, principalmente nas que contêm diversas visões, são extremamente complexos. Para compreendê-los um pouco mais, é necessário perceber alguns conceitos, tais como os definidos em [Santoro et al 2013] e expostos a seguir:

- Em uma atividade, o contexto é a soma de todo o conhecimento que os atores possuem dentro de certo assunto/conjuntura/tarefa.
- O conhecimento contextual é aquele que está relacionado ao presente, e é fundamental para o total entendimento de certa história. Tal conhecimento é despertado por certas situações e eventos e é determinado pelo foco do indivíduo.
- De acordo com certo contexto, o ator considera mais especificamente um subconjunto do conhecimento contextual. Esse conhecimento é então agrupado, organizado e estruturado em um contexto processual, que contém “partes de conhecimentos”.
- Embora o conhecimento contextual exista na teoria, ele é considerado implícito e latente, e não é utilizado até que um objetivo ou intenção surja. Quando um evento acontece, ele ganha o foco da atenção do ator, e é então processado.
- Já que o contexto se modifica de acordo com a realização de determinada tarefa, ele pode ser considerado dinâmico; ele também é relativo pois varia de acordo com o foco de atenção. O foco determina o que o contexto deve conter, e o contexto limita o foco.

Definindo melhor o contexto pela ótica da interação dos diferentes conhecimentos, tem-se que contexto é um espaço de conhecimento compartilhado, explorado pelos atores envolvidos na interação. Contudo, é necessário que haja a distinção entre conhecimento contextual de cada ator e contexto compartilhado. O objetivo de uma atividade, por exemplo, pertence ao contexto compartilhado; cada ator precisa negociar, utilizando seu conhecimento contextual, para que o contexto processual seja construído colaborativamente. Já o contexto compartilhado age como um filtro, definindo que conhecimentos devem ser considerados (explicitados) e que conhecimentos não devem ser compartilhados (conhecimento implícito). O contexto processual contém todas as informações discutidas e aceitas (ou negociadas) por todos os envolvidos na interação. Quando um trabalho em grupo é realizado, o contexto compartilhado é fruto da negociação do conhecimento contextual dos atores envolvidos na atividade. Isso não significa que pontos de vista devam ser uniformizados. Entretanto, trabalhar colaborativamente exige constante comunicação, coordenação e cooperação, como visto no Modelo 3C.

O conhecimento coletivo pode ser um instrumento em potencial para a solução de problemas. Quando esse conhecimento é registrado, ele pode ser recuperado para que seu resultado seja melhor analisado e reaproveitado. Entretanto, quando esse conhecimento não é formalizado, não há como conhecermos suas implicações. A contação de histórias pode ser usada para que informações importantes sejam recuperadas. Nesse cenário, a história é um evento real vivido por determinado grupo de pessoas onde seus participantes/personagens devem expressar suas memórias sobre certos acontecimentos. Os eventos representam “partes de conhecimentos”, que, quando conectados, resultam em uma história. Sendo assim, os ouvintes da mesma passam a ter acesso não somente aos eventos, como percebem todo o contexto que a envolve. Segundo Brooks (2003), contexto significa ver de um determinado ponto de vista. Este estudo busca conhecer o ponto de vista dos alunos em relação à abordagem essencialmente comunicativa proposta para a disciplina de inglês, através do *storytelling* [Salmon 2007], com o apoio da tecnologia.

3. APLICAÇÃO DE TÉCNICAS PARA A RECUPERAÇÃO COLETIVA DE CONHECIMENTO

O conhecimento pode ser distinto de duas formas: conhecimento tácito e explícito. Podemos entender que o tácito é o conhecimento que cada indivíduo traz consigo, obtido através da experiência. Já o conhecimento explícito envolve a captura e representação dos fatos [Nonaka 1991]. Estudos mostram que a barreira no processo de gestão do conhecimento pode estar centrada na externalização, ou seja, na transformação do conhecimento tácito para explícito. No entanto, faz-se necessário pensar em formas, ferramentas ou métodos de compartilhamento de conhecimento que possam apoiar este processo.

Uma história, ou relato, pode apresentar diferentes pontos de vista de acordo com posicionamentos e crenças de certos indivíduos. O relator pode descrever a versão da história que guardou em sua mente após ter participado do evento ou acompanhado o seu desenvolvimento; ele também pode externar seu conhecimento tácito ou acrescentar algo a uma versão que já é conhecida. Embora pouco provável, já que o ser humano é normalmente tendencioso e movido por emoções, o relator também pode descrever aquilo que é real [Carminatti et al 2006]. Embora o *storytelling* seja uma ferramenta eficaz, nem sempre a versão relatada retrata a versão conhecida ou real. O conceito real é muito abstrato já que varia de indivíduo para indivíduo; o que é realidade para certa pessoa, não necessariamente é para outra. Com isso, existem técnicas que visam aproximar essas diferentes visões a fim de que se chegue o mais próximo possível do que de fato ocorreu.

Fuks e Pimentel (2012) propõem três maneiras de se alcançar a realidade: a técnica da entrevista, da dinâmica de grupo e da dinâmica de grupo apoiada por sistemas. As três técnicas sugeridas devem contar com um editor, cujo papel é ler os relatos e organizar uma versão segundo sua lógica pessoal. Com isso, surge uma nova discussão em relação à visão do editor. A primeira técnica é a da entrevista, onde os participantes do evento devem responder perguntas e, então, o editor combina as informações e produz o relato final. A segunda técnica busca capturar o conhecimento através de dinâmicas de grupo, onde os participantes se reúnem em um certo local físico e externam suas ideias através da interação. Essa dinâmica requer um facilitador e um relator; o facilitador deverá conduzir a discussão e o relator deverá tomar nota do conhecimento compartilhado e torna-lo acessível. Já na terceira técnica, que também conta com a dinâmica de grupos, os participantes interagem para enriquecer a narrativa apoiados pela tecnologia; o que pode tornar a interação síncrona ou assíncrona. Nesse caso, não há a necessidade de um relator.

Segundo Perret (2004), as narrativas com começo, meio e fim, são uma forma adequada de dizer o que aconteceu e, ao mesmo tempo, pode exteriorizar o conhecimento tácito do grupo, permitindo aos indivíduos realizarem suas contribuições, principalmente em ambientes educacionais.

Considerando o perfil dos discentes da atualidade, o presente trabalho fez uso da terceira técnica. Os alunos da rede municipal do Rio de Janeiro usaram a ferramenta Sofia [Lage e Quintanilha 2010], onde a história foi escrita através de fragmentos associáveis compartilhados a todos, com a possibilidade de serem comentados e editados de acordo com dimensões do domínio estabelecidas pelo facilitador.

4. FERRAMENTA SOFIA

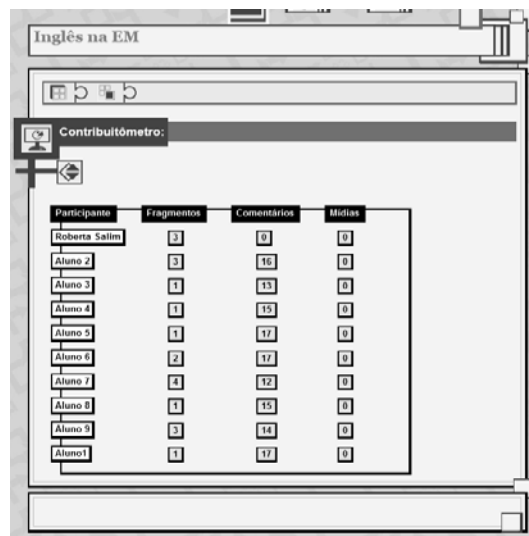
O arcabouço Sofia [Lage e Quintanilha 2010] foi especialmente elaborado para apoiar a técnica de *group storytelling*. Através da mesma, histórias podem ser contadas por diferentes atores de forma colaborativa. Sendo necessário que o participante seja o criador de uma história ou o mesmo seja convidado para uma. Dentro da história há definições sobre papéis. Além disso, a história pode ser definida como pública ou privada; ser gerenciada por um papel específico, normalmente o facilitador, ou ser auto gerenciável, em que todos os participantes dos mais diferentes papéis podem manipular toda a história; permitir a utilização de mídias; entre outras possibilidades. A Figura 1 mostra o cadastro de uma nova história.

Os usuários podem assumir diferentes papéis na história: contador, facilitador ou leitor. O contador é responsável pela construção da história, interação e cooperação com outros contadores — ele pode acrescentar fragmentos à mesma, bem como comentar os relatos de outros atores. O facilitador é responsável pela coordenação da história durante o seu processo; ele poderá participar diretamente ou indiretamente da história. Já o leitor, é responsável pela extração do conhecimento buscado ou pretendido da narrativa.

Figura 1: Criação de uma nova história na ferramenta.

Através do Sofia, várias histórias podem ser criadas (Figura 1). Além disto, uma mesma história pode conter diferentes narrativas (fragmentos), que podem conter inúmeros comentários. O facilitador tem acesso a todas as informações concernentes ao processo (Figura 2), tais como: histórias disponíveis; contribuímetro contendo número de participantes, número de fragmentos, número de comentários e mídias; e o participômetro, contendo o número total de contribuições da história e número total de contribuições acessadas.

A ferramenta pode ser facilmente usada e seu uso possui muitos benefícios. Contudo, seu aspecto colaborativo é essencial. O fato de várias pessoas poderem colaborar através de fragmentos e comentários torna a experiência extremamente rica, e faz com que resultados satisfatórios possam ser alcançados.



Participante	Fragmentos	Comentários	Médias
Roberta Salim	3	0	0
Aluno 2	3	16	0
Aluno 3	1	13	0
Aluno 4	1	15	0
Aluno 5	1	17	0
Aluno 6	2	17	0
Aluno 7	4	12	0
Aluno 8	1	15	0
Aluno 9	3	14	0
Aluno1	1	17	0

Figura 2: Participantes e quantitativo de suas contribuições.

Os indivíduos envolvidos na narrativa não são identificados através de nomes ou e-mails, mas números são atribuídos aos seus usuários. Dessa forma, o sigilo da autoria dos fragmentos e comentários é mantido.

5. MÉTODO DE PESQUISA ADOTADO

O levantamento da influência da abordagem comunicativa na aprendizagem da língua inglesa em escolas da rede municipal do Rio de Janeiro, Brasil foi conduzido através de uma pesquisa descritiva. Segundo Gil (2008), este método de pesquisa permite descrever aspectos de determinados eventos através de técnicas de coleta de dados como a observação sistemática. Para a condução desta pesquisa foi executado um experimento que é descrito na próxima seção.

5.1 Experimento

Como o projeto de língua inglesa foi iniciado em 2011, os alunos que cursavam o quinto ano do ensino fundamental, foram os primeiros a utilizarem a nova metodologia e o novo material proposto especialmente para a disciplina de inglês. Por essa razão, um grupo de alunos do quinto ano foi selecionado para a pesquisa. A princípio, buscou-se envolver toda uma turma. Entretanto, por problemas operacionais com transporte, doze alunos foram convidados para o experimento, que foi realizado em um dos laboratórios da UFRJ; dos doze, apenas onze puderam comparecer.

Alguns dias antes do experimento, o projeto foi explicado aos alunos bem como seus objetivos. Eles ficaram cientes de que deveriam falar sobre as experiências vividas nas aulas de inglês, e que poderiam narrar tanto aspectos positivos quanto negativos de acordo com seus diferentes pontos de vista. Os alunos também foram alertados da importância de serem totalmente sinceros durante o projeto, expressando seus contentamentos e descontentamentos. Eles também foram informados de que deveriam não somente criar fragmentos como também interagir com seus colegas através de comentários. Também foi explicitado aos alunos como a ferramenta Sofia os auxiliaria no processo.

Embora, de acordo com o planejamento inicial, o uso das máquinas devesse ser individual, duas duplas precisaram ser formadas por questões técnicas. Com isso, sete

alunos trabalharam individualmente em suas máquinas e duas duplas cada uma em uma máquina.

Os participantes do projeto foram devidamente cadastrados na ferramenta e passaram a ser identificados por números. Dessa forma, puderam acessar a ferramenta e selecionaram seus papéis na história — contador. A partir de então, eles puderam começar a criar seus fragmentos optando por estes serem classificados como um relato, fato ou descritor. Este último é voltado para quando alguém está detalhando algo em relação a outro fragmento existente. Isto ocorre por meio de associações.

O professor de inglês (que assumiu o papel de facilitador e editor), em certo momento, também interagiu com as crianças no ambiente. Embora três pessoas estivessem envolvidas no experimento, além dos alunos, em nenhum momento houve interferência em relação ao conteúdo das narrativas. O facilitador, que em dado momento também postou fragmentos a fim de incitar as crianças a discorrerem sobre certos tópicos, também não exerceu muita influência sobre o conteúdo dos mesmos.

Os alunos participaram eficazmente da atividade, contribuindo com diversos fragmentos e fazendo diversos comentários. O participômetro mostrou um número final equivalente a cento e cinquenta e seis participações, sendo vinte o número de fragmentos criados e cento e trinta e seis o número de comentários postados. Levando-se em conta a média de idade dos alunos (entre 11 e 12 anos) a participação foi considerada muito boa.



Figura 3: Participantes do estudo através do ambiente virtual Sofia.

Apesar de os alunos terem sido identificados por números a fim de diminuir a questão da influência, durante o experimento, as crianças notaram que os números haviam sido dados de acordo com suas respectivas posições. Com isso, algumas delas puderam identificar com quem estavam interagindo no ambiente virtual. Embora não esperado, as crianças sentiram-se confortáveis em fazer os comentários quando identificavam os colegas e também se divertiram com a questão da identificação. A Figura 3 mostra o ambiente onde o experimento ocorreu e os participantes do estudo.

Os alunos concluíram o experimento dentro do tempo esperado e se mostraram receptivas antes e após o projeto. Esboçaram várias vezes seus sentimentos em relação ao experimento através de palavras e gestos. A duração total do evento foi de quatro horas.

5.2 Dados Coletados

Após a realização do experimento, todos os fragmentos e comentários feitos pelos alunos, através da ferramenta Sofia, foram coletados para a edição. As informações coletadas foram consideradas pelo editor e dimensões abaixo foram criadas. Percebeu-se que alguns pontos comuns foram considerados por diferentes alunos. Cento e cinquenta e seis contribuições foram relatadas, grande parte delas tratava dos seguintes assuntos: material utilizado, relação professor-aluno, uso do caderno/quadro em sala de aula, “*sense of achievement e storytelling*” associado à tecnologia.

O material utilizado pelos alunos é fornecido pela Prefeitura do Estado do Rio de Janeiro, Brasil. O uso desse material, somado à aplicação da abordagem, faz com que as aulas sejam dinâmicas e eficazes. Ainda assim, embora o uso do quadro branco não seja frequente, alguns alunos mencionaram seu uso como ponto negativo; o que reforça o descontentamento dos alunos com um estilo de aula mais tradicional. O desagrado com o uso do quadro e do caderno pode ser percebido em três dos vinte fragmentos e não é mencionado em nenhum dos comentários. A seguir, são expostos alguns fragmentos:

- FRAG7: *“Na aula de inglês eu aprendi muita coisa legal com o Zip e a Stella [personagens] no livro de inglês. Eu gostava muito, já no caderno era muito ruim. Quando a professora X ensinava os nomes de lugar, comida e nome de pessoa era legal”.*
- FRAG12: *“Olá eu estou contando a minha história do Inglês. Eu acho que o inglês não tem falha a não ser pelo simples fato de às vezes ter que copiar do quadro. Eu gosto da apostila porque tem muitas coisas legais como o Zip e Stella [personagens] e seus companheiros que são muito legais. Se não tivesse que copiar seria muito mais legal e não teria defeito, na minha opinião. O inglês é a melhor matéria da escola”.*
- FRAG20: *“Gostava muito quando fazíamos trabalho em grupo, mas não gostava quando a tia escrevia o trabalho no quadro kkk [sic]”.*

Além do descontentamento com o uso do caderno, o FRAG20 apresenta um dos mais importantes princípios da abordagem comunicativa, que é a interação. É chamativo o modo como a aluna estabelece um contraste entre o trabalho no quadro e a atividade em grupo, ou seja, entre uma abordagem tradicional e a abordagem comunicativa. Esse é apenas um exemplo de feedback dentre muitos outros obtidos através do uso do storytelling apoiado pela tecnologia.

5.3 Discussão dos Resultados e Limitações

Os resultados coletados permitiram que alguns princípios da abordagem comunicativa fossem observados e, através da formalização do conhecimento tácito, ficou explícita a satisfação dos alunos em relação à abordagem comunicativa, e a insatisfação com um modelo mais tradicional de ensino. Eles mencionaram o fato de copiar matéria do quadro como algo negativo, embora não seja uma prática comum nas aulas de inglês. Mencionaram músicas, mídias, aspectos lúdicos, material, tecnologia e professor como pontos positivos. Alguns também expressaram seu desenvolvimento em relação à língua inglesa.

O número de participação dos alunos, considerando fragmentos e comentários feitos, foi excelente. A colaboração, coordenação e comunicação puderam ser percebidas durante o processo (modelo 3C). O uso da tecnologia foi imprescindível e foi um agente

motivador do experimento. Os alunos estavam extremamente felizes com o projeto e com toda a tecnologia que os cercava. Ficaram pasmados com os computadores do laboratório bem com a ferramenta.

O experimento realizado foi bem-sucedido uma vez que respostas foram fornecidas quanto à influência da abordagem comunicativa na aprendizagem de língua inglesa em escolas da rede municipal do Rio de Janeiro. Através das respostas com o ponto de vista dos estudantes, foi possível identificar que a abordagem comunicativa vem trazendo benefícios ao ensino da língua inglesa, contudo ainda são precisos mais experimentos com as demais turmas em formação para sua comprovação. Com os resultados alcançados, pode-se apontar que o *storytelling* e, conseqüentemente o *group storytelling*, é eficaz no levantamento de dados com crianças, uma vez que é possível incluí-las em um ambiente desafiador, o que as coloca numa posição participativa através de seus relatos.

Apesar dos resultados obtidos, a pesquisa foi realizada com um pequeno grupo de estudantes. Ainda que o número tenha sido reduzido frente aos 20 alunos que totalizam a turma, os participantes foram escolhidos de forma a representarem o perfil da turma segundo seus conhecimentos. A fim de que houvesse uma maior acurácia em relação aos resultados, seria interessante que os próximos experimentos contemplassem todas as turmas, e que os mesmos fossem moradores de diferentes áreas do município.

Outro aspecto a ser aprimorado está relacionado à interferência da professora no processo; seu papel no experimento poderia ser revisto a fim de que os alunos se sentissem mais confortáveis ao narrarem suas histórias. Da mesma forma, para a obtenção de um resultado mais neutro, a edição poderia ser feita por uma pessoa não diretamente ligada ao processo/alunos. Outra opção seria que a edição fosse realizada por mais de uma pessoa e que resultados fossem comparados. Dessa forma, as realidades traçadas poderiam ser confirmadas.

Embora mudanças e melhoramentos possam ser realizados buscando uma maior acurácia nos resultados, esta pesquisa pode indicar novos caminhos para o desenvolvimento educação verificando a perspectiva dos alunos quando nas mais diversas disciplinas e mais uma vez combinar colaboração, tecnologia e uso do *storytelling*. Ademais, a eficácia do experimento poderá servir como esboço de modelo para que seja utilizada em diferentes áreas visando diferentes objetivos.

6. CONCLUSÃO

A fim de competir na economia global, os jovens precisam possuir conhecimentos e habilidades concernentes ao perfil exigido pelo Século XXI. No passado, deter conhecimento significava ser detentor de poder. Hoje em dia, o foco está não apenas no que se conhece, mas também no que as pessoas podem fazer com o conhecimento/informação que pode ser acessado e nas habilidades interpessoais. As empresas buscam profissionais que saibam resolver problemas, que sejam comunicadores eficazes e conhecedores da tecnologia da informação. Há então uma urgência em que os alunos sejam preparados para atuar nesse novo tempo; e é papel da escola prepará-los até mesmo para profissões que ainda surgirão no futuro. Dito tudo isso, uma abordagem tradicional jamais se aplicaria a esse cenário; é necessário que haja interação, comunicação e colaboração.

Foi observado com os resultados alcançados neste estudo, que a abordagem

comunicativa tem sido um meio eficaz em viabilizar o ensino da língua inglesa em escolas da rede pública do município do Rio de Janeiro, Brasil. Essa constatação foi possível através da aplicação do *storytelling* com o suporte tecnológico para a coleta desses dados. Ficou claro também que a contação de histórias aliada a tecnologia constrói um ambiente propício para o levantamento de impressões e críticas do público infantil. Como próximos passos desse trabalho é prevista a aplicação do *storytelling* através da ferramenta Sofia em outros cenários a fim de constatar sua aplicabilidade no que tange o envolvimento de crianças na melhoria de estratégias em domínios que sejam participantes.

Agradecimentos

Agradecemos a Escola Municipal do Rio de Janeiro por apoiar o projeto e permitir a aplicação da pesquisa. Também gostaríamos de agradecer a autorização de pais que permitiram que seus filhos participassem do experimento.

Referencias

- Ferreira, Naura S. Carapeto. (2013) *Gestão Democrática da Educação, Atuais Tendências, Novos Desafios*. 8 ed. São Paulo: Cortez.
- Fuks, Hugo; Pimentel; Mariano (organizadores). (2012) *Sistemas Colaborativos*. Rio de Janeiro: Elsevier.
- Gil, A.C. (2008) *Como elaborar projetos de pesquisa*. 4. ed. São Paulo: Atlas.
- Kumar, V. S. (1996) *Computer-Supported Collaborative Learning: Issues for research*, In 8th Annual Graduate Symposium on Computer Science, University of Saskatchewan, Canada.
- Lage, B. B.; Quintanilha, C. B. C. (2010) *Sofia 2.0: A Maturação de um Arcabouço para Soluções de Group Storytelling através de uma Instanciação Avant-garde*. Orientador: Marcos Roberto da Silva Borges. Rio de Janeiro: UFRJ/IM/DCC. Projeto Final de Curso (Bacharelado em Ciência da Computação).
- Laporti, V, Borges, M., Barganholo, V. P. (2009) "Athena: A collaborative approach to requirements elicitation." *Computers in Industry* 60.6.367-380.
- Larsen-Freeman. Diane. (2000) *Techniques and Principles in Language Teaching*. Oxford: Oxford University Press.
- Lemos, J.; Alves, C.; Duboc, L.; Rodrigues, G. N. (2012) "A systematic mapping study on creativity in requirements engineering". In *Proc. of the 27th Annual ACM Symposium on Applied Computing* (pp. 1083-1088). ACM. Mar.
- Nonaka, I. (1991) "The knowledge-creating Company". *Harvard Business Review*, pp. 96-104.
- Nonaka, I.; Takeuchi, H. (1997) *Criação do Conhecimento na Empresa*. Rio de Janeiro, RJ: Elsevier, 358p.
- Oliveira, E. W.; Borges, M.R.S. (2014.) *A Influência da Diversidade de Conhecimento no Processo de Aprendizagem Colaborativa*. XI SBSC – Simpósio Brasileiro de Sistemas Colaborativo, Curitiba, PR.
- Perret, R. ; Borges, M. R.S; Santoro, F.M. (2014) *Applying Group Storytelling in Knowledge Management*. CRIWG 2004. Springer-Verlag Berlin Heidelberg .

- Richards, Jack C; Rodgers, Theodore S. (2001) *Approaches and Techniques in Language Teaching*. Cambridge: Cambridge University Press.
- Salmon, Christian.(2007) *Storytelling: la machine à fabriquer des histoires et à formater les esprits*. Paris: La Decouverte.
- Santoro, Flávia Maria; Brézillon, Patrick. (2006) *The Role of Shared Context in Group Storytelling*. *Computing and Informatics*, v. 25, 497-522.
- Santoro, Flávia Maria; BORGES, Marcos R. da Silva; SANTOS; Neide. (2013) *Um Framework para Estudos de Ambientes de Suporte à Aprendizagem Cooperativa*. Dissertação de Mestrado.
- Valente, José Armando. *Informática na Educação: instrucionismo X construcionismo*. <http://www.educacaopublica.rj.gov.br/biblioteca/tecnologia/0003.html> Acessado em 05/08/2015.
- Valle, C., Raybourn, E.M., Prinz, W., Borges, M.R.S. (2003) *Group Storytelling to Support Tacit Knowledge Externalization*. Proc. of the 10th International Conference on Human – Computer Interaction. Crete, Greece.

SECOMU

Apresentação

O SECOMU - Seminário de Computação na Universidade, evento que acontece desde 1971 (a partir de 1980 junto ao congresso da SBC), reúne a comunidade acadêmica de computação do Brasil para discutir temas político-científicos de interesse nacional. Ao longo dos anos, os painéis do SECOMU ajudaram a comunidade de computação do Brasil a construir suas posições políticas e acadêmicas em temas de grande relevância, contribuindo com soluções e estratégias para o desenvolvimento tecnológico brasileiro. O SECOMU é tradicionalmente estruturado em torno de painéis compostos por personalidades políticas, acadêmicas, representantes de agências de fomento, membros da indústria e da sociedade em geral.

No congresso da SBC de 2016, o SECOMU tem como foco "Computação e Interdisciplinaridade", explorando os vários aspectos dessa temática contemporânea, desde a utilização da computação como instrumento da interdisciplinaridade até o desenvolvimento da própria computação através da visão interdisciplinar. Para tanto, o SECOMU 2016 é composto por três painéis e uma sessão especial dedicada à assembleia geral da SBC.

O primeiro painel, denominado "Computação e Interdisciplinaridade: o papel das Associações Científicas", apresenta as visões de diferentes sociedades científicas acerca da temática indisciplinar envolvendo a computação. Nesse painel, teremos: o presidente eleito da IFIP (*International Federation for Information Processing*), Prof. Mike Hinchey; a presidente da Sociedade Brasileira de Informática em Saúde (SBIS), Profa. Beatriz de Faria Leão; o representante da Sociedade Brasileiro de Matemática (SBM), Prof. Álvaro Luiz de Bortoli; e o coordenador do Serviço de Informática do Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Dr. Marco Antonio Gutierrez.

O segundo painel, "Raciocínio Computacional como Instrumento para a Interdisciplinaridade", apresenta visões distintas, de convidados brasileiros e estrangeiros, sobre o significado de raciocínio computacional (*computational thinking*) e de como tal conceito vem sendo utilizado na construção de uma visão interdisciplinar da computação. Teremos, nesse painel, a presença dos professores: Ken Kahn, Oxford University, UK; David Cavallo, UFSB; Ricardo Anido, UNICAMP; e Leila Ribeiro, UFRGS.

O painel "Governança da Internet" completa o conjunto de painéis do SECOMU 2016. Sendo essencialmente uma temática interdisciplinar e multisetorial - com questões que vão desde aspectos políticos, passando por acadêmicos e até econômicos -, esse painel traz personalidades importantes no cenário nacional e mundial: Steve Crocker, ICANN *chair*; Demi Getschko, diretor-presidente do NIC.br; e Raúl Echeberría, Vice-Presidente de "Global Engagement" da Internet Society.

Todos os painéis são organizados através de apresentações de nossos convidados, seguidas de blocos dedicados a interações com a audiência - a partir de questões levantadas no final das apresentações.

Agradecemos aos painelistas convidados pela gentileza de terem aceitado nossos convites e desejamos bom proveito do SECOMU 2016 para toda nossa comunidade.

Raimundo José de Araújo Macêdo UFBA/UFSB
Coordenador geral

Painéis / Palestras

Painel 1: Computação e Interdisciplinaridade: o papel das Associações Científicas

Esse painel apresenta as visões de diferentes sociedades científicas acerca da temática indisciplinar envolvendo a computação.

Moderador: Raimundo Macêdo

Painelistas:

Prof. Mike Hinchey

President-Elect of IFIP (International Federation for Information Processing) and will serve as its President from 2016 to 2019. Director of Lero—the Irish Software Research Centre and Professor of Software Engineering at University of Limerick, Ireland. Prior to joining Lero, Professor Hinchey was Director of the NASA Software Engineering Laboratory. In 2009 he was awarded NASA's Kerley Award as Innovator of the Year. Hinchey holds a B.Sc. in Computer Systems from University of Limerick, an M.Sc. in Computation from University of Oxford and a PhD in Computer Science from University of Cambridge. The author/editor of more than 15 books and over 200 articles on various aspects of Software Engineering, at various times Hinchey previously held positions as Full Professor in Australia, UK, Sweden and USA. He is a Chartered Engineer, Chartered Engineering Professional, Chartered Mathematician and Chartered Information Technology Professional, as well as a Fellow of the IET, British Computer Society and Irish Computer Society.

Profa. Beatriz de Faria Leão

Presidente da Sociedade Brasileira de Informática em Saúde (SBIS). Médica (UFRGS 1977), Doutora em Medicina (UNIFESP 1988) com Pós-Doutorado em Informática em Saúde na Universidade Erasmus University em Rotterdam, Holanda (1989-1990). Atua na área de Informática em Saúde desde 1980. É fundadora da Sociedade Brasileira de Informática em Saúde em 1986. Beatriz alia experiência acadêmica (foi professora adjunta na UNIFESP e no Instituto de Informática da UFRGS), de governo (consultora no Ministério da Saúde- DATASUS e Programa Brasileiro de Telessaúde) e da área privada. Atualmente ela é consultora da Jhpiego Johns Hopkins University, atuando no projeto do Sistema de Informações para Recursos Humanos em Saúde de Moçambique, desde 2008 e consultora da UNB na área de padrões e interoperabilidade, desde setembro de 2012. Beatriz é coordenadora do Grupo 1 Arquitetura da Comissão Especial de Estudos em Informática em Saúde da ABNT e coordenadora do Grupo de Trabalho em Padrões da IMIA International Medical Informatics Association.

Prof. Álvaro Luiz de Bortoli

Representante da Sociedade Brasileiro de Matemática (SBM).É professor titular do Departamento de Matemática Pura e Aplicada da UFRGS. Tem doutorado em Aeroelasticidade (UFSC/Brazil - German Aerospace Center/Alemanha, 1995) e pós-doutorado em combustão (ITV-RWTH, Aachen, Alemanha, 2007). É pesquisador do CNPq e professor orientador dos programas de pós-graduação em Matemática Aplicada e em Engenharia Química da UFRGS. Ele é coordenador do Grupo Modelagem e Simulação em Dinâmica de Fluidos Reativos. Tem publicado artigos

em Matemática Aplicada, e publicou alguns livros em escoamentos reativos, o mais recente intitulado "Modeling and Simulation of Reactive Flows" (Elsevier, 2015). Tem atuado em projetos interdisciplinares de consultoria empresarial e representa o IM-UFRGS no conselho do Parque Científico e Tecnológico ZENIT-UFRGS

Dr. Marco Antonio Gutierrez

Engenheiro Eletrônico, com especialização em Ciência da Computação pela Universidade de São Paulo (USP) e Ph.D. em Sistemas Eletrônicos pela mesma Universidade. Foi pesquisador visitante do Image Science Institute, Utrecht University, Holanda e pesquisador associado do Department of Radiology, University of Southern California, EUA. Suas áreas de interesse incluem processamento de imagens médicas, processamento de sinais biológicos e sistemas de informação hospitalares. Atualmente coordena o Serviço de Informática do Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo.

Painel 2 - Raciocínio Computacional como Instrumento para a Interdisciplinaridade

Esse painel apresenta visões distintas, de convidados brasileiros e estrangeiros, sobre o significado de raciocínio computacional (computational thinking) e de como tal conceito vem sendo utilizado na construção de uma visão interdisciplinar da computação.

Moderador: Raimundo Macêdo

Painelistas:

Prof. Ken Kahn

Professor Ken began his academic career at the MIT Artificial Intelligence Lab. After more than a decade exploring logic and constraint programming at the University of Uppsala, Sweden and Xerox PARC he changed his focus to programming languages for children and technology enhanced learning. He founded Animated Programs to develop ToonTalk. ToonTalk is a programming language where children build programs in a game world by training virtual robots. He has been doing research on computers and education since 2003 at the University of Oxford and the University of London. He is the designer and developer of the Behaviour Composer, a web app to help non-programmers build agent-based computer models and simulations. Every winter he teaches a course on computational thinking and modelling at the National University of Singapore.

Prof. David Cavallo

Professor Titular Visitante at the Federal University of Southern Bahia (UFSB), a new public university serving the rural areas of the south of the state of Bahia. At UFSB he is collaborating with colleagues to design, develop and teach new courses on computational thinking which all students take as part of their general formation, as well as robotics for education, inter-disciplinary courses that use computation and computational environments, ways the university can help improve basic education in the region, and ways to use technology to support active, collaborative learning and to overcome the barriers of distance and the lack of experienced teachers for all students. Prior to joining UFSB Cavallo was the Vice-President for Education and Chief

Learning Architect at One Laptop per Child (olpc). Cavallo led the learning team, worked internally and with partners to design and develop the low-cost, low-power, ecological laptop for learning for children, and then worked with countries to develop materials and local teams to support the transformation of public education to enable equitable access to high-quality education. Cavallo was a Research Scientist and Director of the Future of Learning Group at the MIT Media Laboratory. His research focus is on how we can better facilitate human learning using computational technologies. Prior to MIT, Cavallo led the design and implementation of medical informatics as part of a reform of health care delivery and management at the Harvard University Health Services. He was a Software Principal Engineer in Digital's Artificial Intelligence Group and was also the founder of the Advanced Technology group for Digital's Latin American and Caribbean Region. Dr. Cavallo holds a Ph.D and Master of Science degree from the MIT Media Laboratory in Media Sciences focusing on learning and technology, and studied Computer Science at Rutgers University and University of California at Berkeley.

Prof. Leila Ribeiro

Possui graduação em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (1988), mestrado em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (1991) e doutorado em Informática - Technische Universität Berlin (1996). Atualmente é pesquisadora nível 2 do Conselho Nacional de Desenvolvimento Científico e Tecnológico e professor titular da Universidade Federal do Rio Grande do Sul. É membro do IFIP Working Group 1,3 (Foundations of System Specification). Tem experiência na área de Ciência da Computação, com ênfase em Métodos Formais, atuando principalmente nos seguintes temas: especificação e semântica formal, bioinformática, verificação formal e modelos de computação.

Prof. Ricardo Anido

Possui graduação em Engenharia Mecânica - Aeronáutica pelo Instituto Tecnológico de Aeronáutica (1978), mestrado em Ciência da Computação pela Universidade Estadual de Campinas (1983) e doutorado em Computing - Imperial College (1989). Atualmente é professor associado da Universidade Estadual de Campinas, atuando principalmente nos seguintes temas: algoritmos distribuídos, sistemas distribuídos, computação móvel, internet e desempenho

Painel 3 - Governança da Internet

Governança da Internet é um tema essencialmente interdisciplinar e multissetorial. Embora nos primórdios da Internet a sua governança estivesse concentrada principalmente nos aspectos técnicos de seu desenvolvimento e operação, hoje em dia ela também tem enorme entrelaçamento com questões políticas e econômicas, já que a Internet afeta todos os aspectos da vida econômica, social e cultural. Assim, as decisões técnicas necessárias a seu desenvolvimento têm forte relação com políticas públicas e legislações em múltiplos setores. Nos cenários nacional e internacional, a Governança da Internet é exercida por um conjunto distribuído de entidades e fóruns, reunindo governos, empresas de diferentes setores, entidades e indivíduos da sociedade civil, entidades técnicas e comunidades acadêmicas de diferentes áreas de conhecimento. O objetivo principal deste painel é introduzir as principais questões e

entidades envolvidas na Governança da Internet nos cenários nacional e internacional, enfatizando seu caráter interdisciplinar e multissetorial.

Moderador: Flávio Wagner

Painelistas:

Steve Crocker (videoconferência)

Chair do Board da ICANN – Internet Corporation for Assigned Names and Numbers. É considerado um dos pais da Internet, tendo feito parte do time que desenvolveu a ARPANET. Faz parte do “Internet Hall of Fame” da Internet Society. Biografia curta: ver https://en.wikipedia.org/wiki/Steve_Crocker

Demi Getschko

Diretor-Presidente do NIC.br – Núcleo de Informação e Coordenação do Ponto BR e conselheiro do CGI.br – Comitê Gestor da Internet no Brasil. É considerado um dos pais da Internet no Brasil, tendo sido responsável pela primeira conexão TCP/IP entre a FAPESP e o Fermilab. Também faz parte do “Internet Hall of Fame” da Internet Society. Biografia curta: ver https://pt.wikipedia.org/wiki/Demi_Getschko

Raúl Echeberría

Vice-Presidente de “Global Engagement” da Internet Society. É uma das personalidades mais importantes da América Latina na área de Governança da Internet, tendo desempenhado diferentes papéis em sua trajetória. Biografia curta: ver <https://www.internetsociety.org/who-we-are/staff/mr-raúl-echeberría>

SEMISH 2016 – XLIII Seminário sobre Desenvolvimento Integrado de Software e Hardware

Apresentação

O SEMISH, fórum científico mais tradicional do Congresso da Sociedade Brasileira de Computação (CSBC), está em sua quadragésima terceira edição e ocorre em Porto Alegre, nos dias 4 e 5 de julho, juntamente com o XXXVI CSBC. Este ano o Congresso da SBC discute “Computação e Interdisciplinaridade”. O SEMISH 2016 acompanha essa temática, tendo como base os documentos norteadores oriundos das discussões, na SBC, sobre os Grandes Desafios da Computação no Brasil.

Nesta edição a comunidade continuou a prestigiar o evento com uma quantidade expressiva de submissões: 70. Cada artigo foi revisado por três especialistas, resultando em um montante de 210 revisões. Do total de artigos completos submetidos, apenas 19 foram aceitos (todos os com média acima de 6,1) para serem apresentados durante o SEMISH, perfazendo uma taxa de aceitação de aproximadamente 27%.

Ao longo dos dois dias do evento serão apresentadas e discutidas soluções que avançam o estado-da-arte ou que apresentam direções de pesquisa efetivamente interdisciplinares para os problemas complexos que emergem na sociedade atual. Nos anais encontram-se os textos dos artigos completos selecionados, cobrindo tópicos como mobilidade urbana, desastres naturais e saúde. Acreditamos que esses tópicos refletem bem a diversidade e o vigor das iniciativas de pesquisa interdisciplinares em curso nas universidades, centros de pesquisa e empresas do País.

Nossos sinceros agradecimentos aos membros do Comitê de Programa e revisores pelo árduo trabalho realizado em um curto espaço de tempo. Um agradecimento especial ao coordenador geral do XXXVI CSBC – professor Avelino Zorzo – pela confiança depositada em nosso trabalho e pelo apoio prestado ao longo de todo o processo. Por fim, agradecemos a todos os autores de artigos e participantes por acreditarem e participarem do Seminário. Esperamos que apreciem a programação técnica e se inspirem com as boas idéias que, por certo, permearão os dois dias de trabalho!

Luciano Paschoal Gaspar (UFRGS)

César Augusto FonticIELha De Rose (PUCRS)

Comitê de Organização

Coordenação Geral

Luciano Paschoal Gaspar (UFRGS)

Coordenação Local

César Augusto FonticIELha De Rose (PUCRS)

Comitê de Programa

Altigran Soares da Silva (UFAM)

Carlos Ferraz (UFPE)

Claudia Motta (UFRJ)

Clodoveu Davis (UFMG)

Cristiano Costa (UNISINOS)

Cristiano Maciel (UFMT)

Daniela Brauner (UFRGS)

Edson Cáceres (UFMS)

Eliana Almeida (UFAL)

Fernando Osório (ICMC-USP)

Fernando Trinta (UFC)

Helio Lopes (PUC-Rio)

Jonice Oliveira (UFRJ)

Jussara Almeida (UFMG)

Leandro Buss Becker (UFSC)

Leandro Costalonga (UFES)

Leonardo Murta (UFF)

Lisandro Zambenedetti Granville (UFRGS)

Márcia Ito (IBM)

Raimundo José de Araújo Macêdo (UFBA)

Raul Ceretta Nunes (UFSM)

Renata Galante (UFRGS)

Ricardo Anido (UNICAMP)

Roberto Bigonha (UFMG)

Sérgio Luis Cechin (UFRGS)

Sergio Campos (UFMG)

Sergio Soares (UFPE)

Soraia Musse (PUCRS)

Vanessa Braganholo (UFF)

Weverton Cordeiro (UFRGS)

Revisores Associados

Abel Silva-Filho (UFPE)
 Adam Pinto (USP)
 Adriana Braun (PUCRS)
 Alexandre Carissimi (UFRGS)
 Aline Paes (UFF)
 Antonio Carlos Beck Filho (UFRGS)
 Bruno Feijó (PUC-Rio)
 Bruno Lopes (UFF)
 Cássio Freitas Pereira de Almeida (ENCE)
 Claudio de Farias (UFRJ)
 Daniela Trevisan (UFF)
 Davi Viana (UFMA)
 Eder John Scheid (UFRGS)
 Edison Pignaton de Freitas (UFRGS)
 Fabio Rangel (UFRJ)
 Fabricio Faria (UFRJ)
 Gabriel Nazar (UFRGS)
 Gerson Geraldo H. Cavalheiro (UFPE)
 Glederson Santos (IFSul)
 Gustavo Moreira (PUC-Rio)
 Henrique Foresti (UFPE)
 Igor Moraes (UFF)
 Jeferson Coelho (PUC-Rio)
 Joffre Filho (UFRJ)
 José Palazzo Moreira de Oliveira (UFRGS)
 José Viterbo (UFF)
 Josiel Figueiredo (UFMT)
 Juliana Freitag Borin (UNICAMP)
 Juliana Jansen Ferreira (IBM Research Brazil)
 Juliana Saraiva (UFPB)
 Julio Cezar Silveira Jacques Junior (PUCRS)
 Leandro Couto (USP)
 Leandro Dihl (PUCRS)
 Leandro Fernandes (UFF)
 Leandro Galvão (UFAM)
 Leonardo Bays (UFRGS)
 Leonardo Viana (UFAL)
 Lucas Bondan (UFRGS)
 Luciano Silva (UFPR)
 Marcelo Caggiani Luizelli (UFRGS)
 Marcelo Marotta (UFRGS)
 Marcelo Walter (UFRGS)
 Marcus Ritt (UFRGS)
 Marlon Alcantara (PUCRS)
 Mateus Rutzig (UFMS)

Mônica Cunha (IFAL)
Milton Heinen (UNIPAMPA)
Muriel Franco (UFRGS)
Natássya Silva (USP)
Osmar Marchi dos Santos (UFSM)
Paula Ceccon Ribeiro (PUC-Rio)
Paulo Engel (UFRGS)
Pedro Santos Neto (UFPI)
Rafael Berri (USP)
Rafael Escalfoni (UFRJ)
Rafael Moraes (UFRJ)
Rafael Rocha (PUC-Rio)
Raoni Kulesza (UFPB)
Raphael Guerra (UFF)
Raul Feitosa (PUC-Rio)
Rejane Frozza (UNISC)
Ricardo Pfitscher (UFRGS)
Ricardo Luis dos Santos (UFRGS)
Roben Lunardi (IFRS)
Roberto Pereira (UFPR)
Rodolfo Antunes (UFRGS)
Rodrigo Schramm (UFRGS)
Soraia Silva Prietch (UFMT)
Tsang Ing Ren (UFPE)
Vinícius Schaurich (UFRGS)
Vinicius Garcia (UFPE)

Palestras

Supreendendo a Intuição e Revelando Alternativas: A Importância da Observação e Experimentação Sistemáticas na Descoberta Científica e no Avanço Tecnológico

Manuel Menezes de Oliveira Neto (INF-UFRGS)

A computação desempenha um papel estratégico nas mais diversas atividades humanas, e este contexto interdisciplinar nos oferece inúmeras oportunidades para avanços científicos e tecnológicos. Em várias situações, a experimentação *in Silico*, realizada por meio de simulações computacionais, apresenta vantagens (como redução de custos e tempo) sobre métodos convencionais utilizados em áreas como biologia, medicina e engenharias. Nestes casos, a possibilidade de formular e testar hipóteses em condições que não seriam possíveis por meio de experimentação convencional (*e.g.*, por questões éticas, destruição de amostras, escala de tempo envolvida nos experimentos, etc.) tem um valor inestimável. Por sua vez, a disponibilidade de dispositivos móveis com considerável poder de processamento e a capacidade de prototipação oferecida pelas impressoras 3D oferecem novas oportunidades para avanços tecnológicos nas mais diversas áreas. Para se beneficiar de tais oportunidades, observação e experimentação sistemáticas são fundamentais. Esta palestra abordará a questão de interdisciplinaridade e a importância da observação e da experimentação sob a perspectiva do cientista da computação, ilustrando-a com diversos exemplos práticos. Por fim, serão mencionadas algumas tecnologias recentes, para as quais a audiência será convidada a descobrir novas oportunidades de exploração e de integração com outras áreas.

Cidades Inteligentes: Interdisciplinaridade, Software Livre, Dados Abertos e Startups

Fabio Kon (IME-USP)

Com o crescimento da população urbana, problemas de infraestrutura e de acesso limitado a recursos em diversas cidades ao redor mundo afetam negativamente a vida de bilhões de pessoas. Tornar as cidades mais inteligentes pode ajudar a melhorar os serviços urbanos aumentando a qualidade de vida de seus cidadãos. As Tecnologias da Informação e Comunicação (TIC) são meios fundamentais para esse objetivo desde que integradas de forma harmoniosa com outras disciplinas tais como arquitetura, urbanismo, engenharia, sociologia, psicologia, economia e política. Técnicas avançadas de desenvolvimento de sistemas de software podem ser usadas para facilitar enormemente a criação e integração de aplicações robustas para cidades inteligentes. Entretanto, ainda existem desafios técnicos, científicos, sociais e políticos significativos que necessitam ser enfrentados antes que essas plataformas possam ser amplamente utilizadas. Nesta palestra discutiremos os principais requisitos e desafios técnicos para a criação de infraestruturas de software robustas para cidades inteligentes. Abordaremos também o papel absolutamente essencial dos dados abertos e do software livre para a criação de ecossistemas saudáveis para o desenvolvimento de cidades inteligentes. Finalmente, discutiremos oportunidades para a criação de *startups* por bacharéis, mestres e doutores formados em Computação.

Trabalhos Aceitos

Enactive Systems & Computing: Mapping the Terrain for Human-Computer Interaction Research

Alessandro Arpetti (UNICAMP)

M. Cecília C. Baranauskas (UNICAMP)

Smart Control of Expenses Using Mobile Devices

Giovani R. F. Junior (UFOP)

Vicente J. P. Amorim (UFOP)

Thiago L. Gomes (UFOP)

Igor M. Pereira (UFOP)

Explorando Dados Abertos Governamentais Sobre a Mobilidade Urbana na Cidade do Rio de Janeiro

Sérgio Manuel Serra da Cruz (UFRRJ)

Luan Soares Andrade (UFRRJ e UFRJ)

Jonice Oliveira Sampaio (UFRJ)

Exploratory Analysis of Public Transportation Data in Curitiba

Nádia P. Kozievitch (UTFPR)

Tatiana M. C. Gadda (UTFPR)

Keiko V. O. Fonseca (UTFPR)

Marcelo O. Rosa (UTFPR)

Luiz C. Gomes-Jr (UTFPR)

Monika Akbar (University of Texas at El Paso)

Análise da Utilização de Interface Multimodal para Aplicação de TV Digital e Segunda Tela

Renan Patrick de C. Marçal (UFPA)

Cláudia Caroline L. de Oliveira (UFPA)

Marcos César da Rocha Seruffo (UFPA)

Utilização de Redes Bayesianas para Avaliação da Qualidade de Experiência do Usuário: Um Estudo de Caso com Segunda Tela

Joahannes Bruno Dias da Costa (UFPA)

Marcos Cesar da Rocha Seruffo (UFPA)

Marianne Kogut Eliasquevici (UFPA)

Daniel da Silva Souza (UFPA)

Um Framework para Coprojeto de Hardware e Software de Sistemas ADAS Baseados em Visão

Leandro A. Martinez (ICMC-USP)

Tiago Lobo (ICMC-USP)

Eduardo Marques (ICMC-USP)

Qualidade de Experiência do Usuário em uma Rede WiFi Municipal usando Streaming de Vídeo

Juliano Ratusznei (UFABC)

Waleska Zanotto (UFABC)

Carlos Kamienski (UFABC)

Proposta de Solução para Aplicação de Internet das Coisas nas Redes Elétricas Inteligentes Brasileiras Considerando as Limitações das Normas Vigentes

Efrem E. O. Lousada (IFNMG)

Raquel A. F. Mini (PUC Minas)

Um Veículo Aéreo Não Tripulado para Encontrar Focos de Poluição

Marcelo Maia da Silva (UFPA)

Heitor Freitas (ICMC-USP)

Bruno Faíçal (ICMC-USP)

Gerson Serejo (ITV)

Jó Ueyama (ICMC-USP)

Gustavo Pessin (UFPA e ITV)

Análise de Padrões Espaciais e Temporais da Mobilidade de Táxis em San Francisco e Roma

Cristiano Martins Monteiro (CEFET-MG)

Fábio Rocha da Silva (CEFET-MG)

Cristina Duarte Murta (CEFET-MG)

SEDNA – Sistema Emergencial para Desastres Naturais

Yuri Marinho Olivatti (FEI)

André Kurt Amann Pereira (FEI)

Rodrigo Filev Maia (FEI)

Ferramentas de Computação Visual para Apoio ao Treinamento de Expressões Faciais por Autistas: Uma Revisão de Literatura

Thiago Pradi (UFPR)

Luciano Silva (UFPR)

Olga R. P. Bellon (UFPR)

Gustavo M. S. Dória (UFPR)

Uma Contribuição à Reabilitação Cardíaca Explorando a Identificação de Situações na IoT

Douglas Scheunemann (UCPEL)

Adenauer Yamin (UCPEL)

João Lopes (UFRGS)

Cláudio Geyer (UFRGS)

Ciência de Contexto na IoT: Uma Contribuição à Pesquisa Agropecuária

Patrícia Davet (UFPEL)

Huberto Kaiser Filho (UFPEL)

Leonardo João (UFPEL)

Lucas Xavier (UFPEL)

Rodrigo Souza (UFRGS)

João Lopes (UFRGS)

Ana Marilza Pernas (UFPEL)

Nelsi Warken (Embrapa Clima Temperado)

Adenauer Yamin (UFPEL)

Avaliação de Cluster Raspberry Pi para Execução de Aplicações de Análise de Imagens Microscópicas Médicas

Rafael M. Ramos (UnB)

Célia Ralha (UnB)

George Teodoro (UnB)

Seleção Adaptativa de Operadores Aplicada ao Problema do Despacho Econômico de Energia Elétrica

Richard Aderbal Gonçalves (UNICENTRO)

Carolina Paula de Almeida (UNICENTRO)

Sandra Mara G. S. Venske (UNICENTRO)

Josiel N. Kuk (UNICENTRO)

Lucas M. Pavelski (UNICENTRO)

Arquitetura para Fog Computing em Sistemas de Middleware para Internet das Coisas

Matheus Crespi Schenfeld (PUCRS)

Leonardo Amaral (PUCRS)

Everton de Matos (PUCRS)

Fabiano Hessel (PUCRS)

Escalabilidade e Paralelismo em Ambiente de Nuvem: Uma Solução para Processar Trâmites Judiciais

Emmanoel M. Sousa Junior (UFRN)

Frederico Lopes (UFRN)

Idalmis Milian (UFRN)

Enactive Systems & Computing

Mapping the terrain for Human-Computer Interaction research

Alessandro Arpetti, M. Cecília C. Baranauskas

Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251, Cidade Universitária, Campinas/SP - Brasil, CEP 13083-852

{alessandro.arpetti,c.baranauskas}@ic.unicamp.br

***Abstract.** While the concept of enaction is not new, recent trends on the use of contemporary devices to interact with computational systems have attracted the attention in the field of interaction design. A quick look in the literature reveals different meanings for an enactive system, depending on the knowledge domain. This paper explores the subject by conducting a Systematic Literature Review to construct a big picture on the subject. Results reveal the paths being traced and issues still unexplored that offer opportunities for investigation in the human computer interaction field.*

1. Introduction

Over the past 25 years the concept of “enaction” had a spread in the scientific world, focusing mainly in the field of cognitive science and artificial intelligence (Vernon et al. 2007). In the common language, the term “enaction” derives from the verb “enact” that, in addition to the juridical meaning of establishing by law, means perform a play or put something into practice. Varela, Thompson and Rosch refer to this last meaning of the term to define their concept of cognition, starting the tradition of the enactivism (Varela et al. 1991). For these authors, “cognition is not the representation of a pre-given world by a pre-given mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs” (Ibid. p. 9). This new vision allows the overcoming of the Cartesian dualism of mind and body, reaffirming the importance of the body and the action in relation to both perception and knowledge (Flender 2008). More recently, the concept of enaction was approached to the field of information systems by offering new ways of interaction and involvement of users and systems. The development of new technologies that allow a more physical and immersive modality of interaction brought to the conception of an “enactive system”, that “is constituted by dynamically coupled human and technological processes, that is, a dynamic mind-technology embodiment” (Kaipainen et al. 2011). Nevertheless, although the concept of enaction seems to find fertile ground in the Human-Computer Interaction field and in the application of new ways of interaction, a generic search of the string “enactive system” in the main digital libraries returns very few occurrences (Table 1). This result suggests that the literature seems to show a gap between the concept of “enaction”, which has a relatively long and well-established history in the field of cognitive science, and its application in the field of information systems.

Table 1: Results of search for the string “enactive system”.

Source	Results
ACM Digital Library	2
IEEE Xplore	0
Science Direct	1
Scopus	9
Web of Science	3
	= 15 (total)
	- 4 (duplicated)
Unique results	11

The goal of this paper is to present results of a systematic review about enactive systems in the computing field. This paper is organized as follows: Section 2 describes the research method of the study; Section 3 gives an overview of the results of the study; Section 4 presents a discussion on the reported results. Finally, Section 5 summarizes this work, presents our conclusions and points out future works.

2. The Method

This study has been undertaken as a systematic literature review (SLR) following the guidelines for Evidence-based software engineering (EBSE) proposed by Kitchenham (2004). A SLR is a technique used to search for evidence in scientific literature that is conducted in a formal manner, applying well-defined steps, according to a previously elaborated protocol.

This systematic review method is composed by various steps organized in three main phases: phase 1: Planning (specify research questions, develop review protocol, validate review protocol); phase 2: Conducting (identify relevant research, select studies, assess study quality, extract required data, synthesize data), and phase 3: Reporting (write review report and validate report) (Brereton et al. 2007). During phases 1 and 2 the software "StArt" (Laboratory of Research on Software Engineering (LaPES) - Computing Department of the Federal University of São Carlos (DC/UFSCar) 2010) has been used to provide a support for the compilation and execution of the review. The goal of this review is to summarize the research in the topic of enactive systems and computing, and provide a consistent literature background in order to appropriately position new research activities.

2.1. Review Questions

The research questions addressed by this study are: RQ1 - How many studies concerning "enaction" and computing have been published since 2005? RQ2 - What research topics are being addressed? RQ3 - Who is leading these researches? RQ4 - Are these studies theoretical or practical? Do they involve some kind of software application? RQ5 - How the concept of "enaction" is interpreted?

Concerning RQ1, it has been identified the number of publications per year and the journal or conference where the study was published. For RQ2 it has been looked for the main topic area of the study. With respect to RQ3, it has been considered individual researchers, the organization to which researchers were affiliated and the organization country. In relation to RQ4, it has been classified the nature of the study and, eventually, the software applications used or proposed within the research. Finally, for

RQ5 the definition and use of the “enaction” concept has been observed, both in theoretical and practical applications.

2.2. Sources and Search Selection Criteria

The search process started with automatic search in digital libraries. Based on their importance in the computing literature, on their wide range of publication topics and on their popularity, five digital libraries were chosen: ACM Digital Library (<http://dl.acm.org>), IEEE Xplore (<http://ieeexplore.ieee.org>), Science Direct (<http://www.sciencedirect.com>), Scopus (<http://www.scopus.com>), Web of Science (<https://www.webofknowledge.com>). Since the search for the string “enactive system” returns, as mentioned above, a few number of results, a big concern for the definition of the search string was to add a greater number of results and to provide a broader basis for answering the research questions. The result was to select all the publications subsequent to 2004 that contained in the title, in the abstract or in the keywords at least one of the terms: “enaction”, “enactive” or “enactivism”.

2.3. Sources and Search Selection Criteria

For the definition of inclusion and exclusion criteria it has been taken into account the requirements due to the research questions and the linguistic possibilities of the researchers who carried out the selection. To be included in the review, the articles must be: peer-reviewed; published between Jan 1, 2005 and September 30, 2015; written in English, French, Italian, Portuguese or Spanish; related to the computing area; computing and Enaction must not have a marginal role in the study. Excluded articles are those that, in addition to those that did not meet any of the criteria mentioned above, were: short articles; duplicated studies.

2.4. Data Extraction

After the selection phase, it was the turn of to the data extraction process that was performed by reading the abstract and screening the full-text of each one of the selected studies. The data extracted from each study were: 1) Paper Information (Title, Year, Authors, Affiliation, Country, Source type [Journal, Conference]); 2) Research Area; 3) Topic; 4) Research type (Theoretical, Empirical); 5) Use or propose any tool? (Yes, No); 6) If use a technological tool: what type of tool? (Web application, Simulation software, Data visualization, NUI, Ontological tool, Robotics); 7) Interpretation of the “Enaction” concept (as body presence, as system interaction, as interface modality, as common vocabulary, as Varela’s definition).

3. Results

This section presents the results of the review, based on studies selection and data extraction processes. Firstly, the automatic search was conducted at each digital library. Then, an iterative process was applied to exclude duplicate and not relevant papers.

The result of the automatic search reached 1268 studies, mainly found in Scopus (52.6%) and Web of Science (38.3%) (Figure 1). With the selection process, 594 duplicate results were detected and 602 studies were rejected, based on inclusion/exclusion criteria and full-text screening. As final result of the selection process, 72 studies were selected and admitted to the data extraction process.

Concerning the type of publication, 27 studies were published in conference proceedings (37%), while 45 papers in journals (63%). As regards of the temporal distribution of the studies, it is possible to notice a growing linear trend, characterized by an irregular distribution with a greater or lesser number of publications depending on the year (Figure 2).

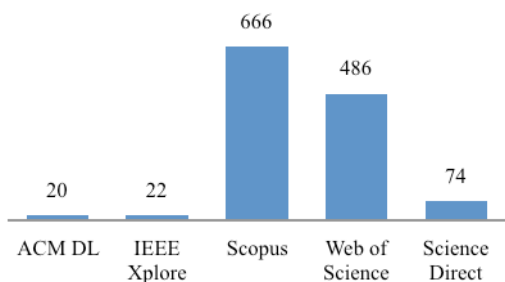


Figure 1: Automatic search results per source.

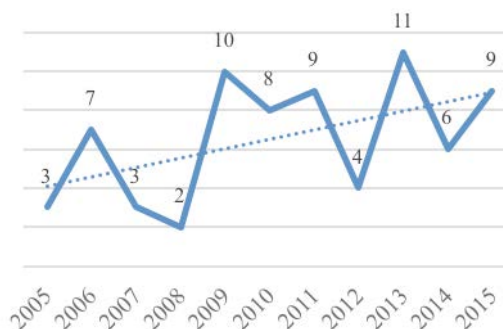


Figure 2: Distribution of publications per year.

The papers presented in conferences are distributed among 21 conferences (Table 2). Only 4 conferences show the presence of more than one paper, with the “Artificial Intelligence and Simulation of Behavior” (AISB) and the “ACM SIGCHI Creativity and Cognition” (SIGCHI CC) conferences with 3 papers, the “American Society Of Mechanical Engineers” (ASME) and the “International Society for Optics and Photonics” (IS&T/SPIE) conferences with 2 papers.

Table 2: List of Conferences.

AISB	3	CASYS	1	EDU LEARN	1
SIGCHI CC	3	CogInfoCom	1	ICDL	1
ASME	2	CORIA	1	ICHIM	1
IS&T/SPIE	2	CSCL	1	IROS	1
ALIFE	1	CW	1	T4E	1
BICA	1	CYBCONF	1	WCECS	1
CAiSE-DC	1	DEST	1	WCETR	1

These conferences are related to 10 research fields (Table 3) and it is notable that 48.1% of the conferences concern AI, Robotics and Cybernetics, while 40.7% are divided between Education, Art and Information Systems.

Table 3: Conferences research fields.

Artificial Intelligence	7	Cybernetics	2
Education	4	Cognitive science	1
Robotics	4	Computing	1
Art - Technology	3	Optical sciences	1
Information System	3	Virtual reality	1

The number of journals in which the papers are published is 33, with 8 journals presenting more than one publication. Among them there are “Kybernetes” with 4 papers, “Lecture Notes in Artificial Intelligence” and “Leonardo” with 3 papers, the “Journal of E-Learning and Knowledge Society”, “Virtual Reality”, “AI & Society”, “Constructivist Foundations” and “Digital Creativity” with 2 papers (Table 4).

Table 4: List of Journals.

Kybernetes	4
Lecture Notes in Artificial Intelligence	3
Leonardo	3
Journal of E-Learning and Knowledge Society	2
Virtual Reality	2
AI & Society	2
Constructivist Foundations	2
Digital Creativity	2
Adaptive Behavior	1
Artificial Life & Technology	1
Automated Systems Based on Human Skill and Knowledge	1
Cognitive Computation	1
Computer-Aided Design and Applications	1
Contemporary Music Review	1
Educação em Revista	1
Educational Research	1
Educational Technology Research and Development	1
Emotion Modeling	1
Frontiers in Artificial Intelligence and Applications	1
IEEE Transactions on Haptics	1
IEEE Transactions on Evolutionary Computation	1
International Journal of Interactive Design and Manufacturing	1
International Journal of Machine Consciousness	1
Journal of Experimental & Theoretical Artificial Intelligence	1
Journal of Interactive Learning Research	1
Journal of New Music Research	1
Journal of Research on Technology in Education (International Society for Technology in Education)	1
L'Année psychologique	1
Minds and Machines	1
Personal Ubiquitous Computing	1
Philosophy & Technology	1
PsychNology Journal	1
Systems Research and Behavioral Science	1

The list of research fields related to these journals amounts to 11 and, as for the conferences, almost half of the papers are published in the AI, Robotics and Cybernetics

field (42.2%), while the absence of Journals related to the Information Systems field is replaced by a bigger number of publications in Psychology and Cognitive Science fields (15.6%) (Table 5).

Table 5: Journals research fields.

Artificial Intelligence	14	Design	2
Education	7	Music	2
Art - Technology	5	Virtual reality	2
Cybernetics	4	Computing	1
Psychology	4	Robotics	1
Cognitive science	3		

The review holds contributions from 22 countries located in all the continents (Figure 3), but with a concentration in the Americas (23%) and Europe (68%) (Figure 4). It is interesting to note that only five countries (UK, Italy, France, Sweden and USA) gathered a total of 53 publications, accounting for 73.6% of the total.

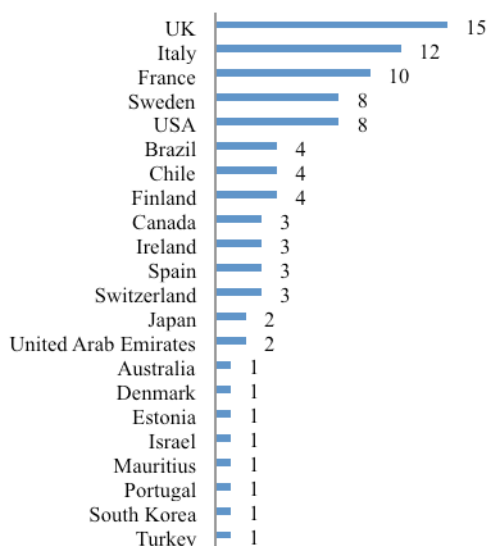


Figure 3: Distribution of publications per Country.

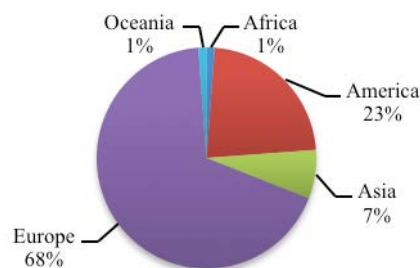


Figure 4: Distribution of publications per Continent.

The list of authors appears quite varied, with a total of 176 and an average of 2.4 authors per paper. 28 authors are present in more than one study, with 21 authors with 2 papers, 5 with 3 papers and 2 authors with 4 papers (Figure 5). Even for organizations to which the authors are affiliated, it is possible to notice a wide variety, with 82 different entities. Between them, 17 are repeated more than once. The most active organizations are the Informatics Research Centre of the University of Skovde, in Sweden, the Polytechnic of Milano, in Italy, the Department of Informatics of the University of Sussex, in UK and the Industrial Engineering Department of the University of Santiago de Chile, in Chile (Figure 6).

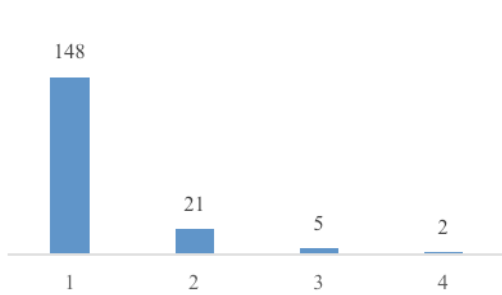


Figure 5: Number of papers per author.

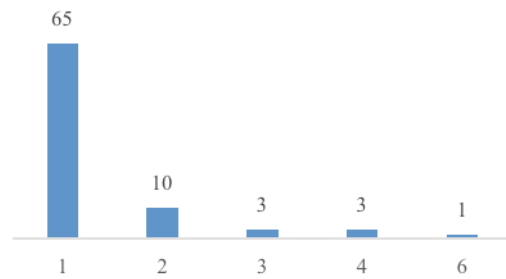


Figure 6: Number of papers per organization.

In relation to the research areas, 17 fields have been identified (Figure 7), with a primary focus of studies in Artificial Intelligence (AI) and Robotics (27.8%), Educational Technology (20.8%) and Human-Computer Interaction (HCI - 18.1%).

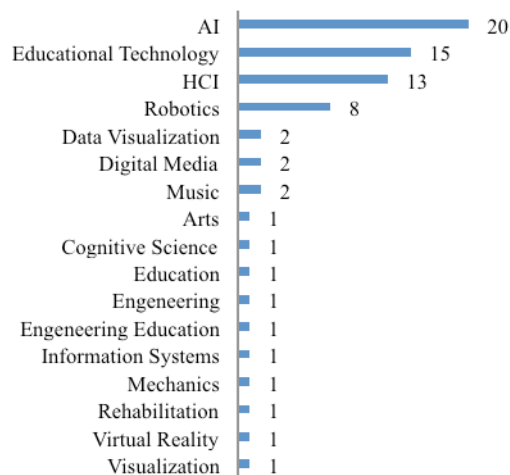


Figure 7: Research areas of publications.

The topics of the papers are more dispersed and it was possible to identify 53 different subjects, with a slight concentration for the modeling of emotions, for virtual reality, for autonomous and conscious cognitive systems, and for different types of interfaces (Table 6).

Table 6: Topics of selected studies.

Topic	N. of studies	Topic	N. of studies
Computational modeling of emotion	6	Digital narratives	1
Virtual reality	4	Educational methods	1
Enactive interfaces	3	Enactive systems	1
Machine consciousness	3	Engineering learning	1
Multimodal interfaces	3	Game-based learning	1
Autonomous cognitive systems	2	Gesture analysis	1

often coupled with virtual simulation technologies (25%). The application in robotics is fairly common (17.3%), while less frequent is the presence of Data visualization (7.7%) or Ontological tools (7.7%) (Figure 9). Finally, in relation to the use and the interpretation of the “enaction” concept, it is possible to notice that the majority of the selected studies (61.9%) follow the original definition of Varela et al. The remaining studies (38.1%) adopt an interpretation more linked to an embodied modality of interaction in which the presence of the body is related to new typology of interfaces and information systems (Figure 10).

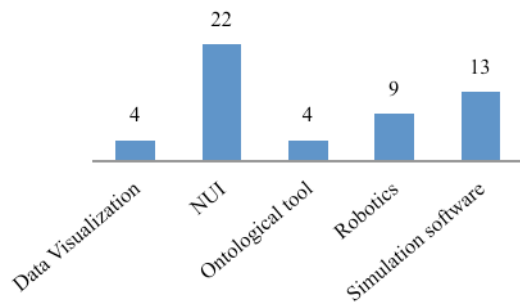


Figure 9: Kind of technological tools.

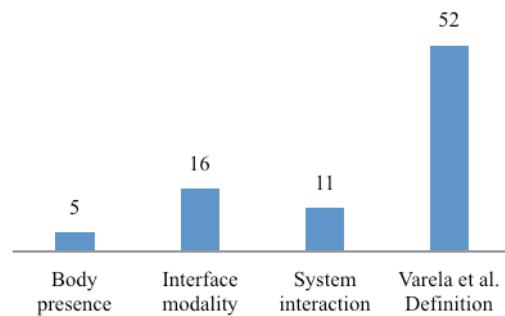


Figure 10: Interpretation of the “enaction” concept.

For example, an interpretation of the concept of enaction as body presence is that present in Di Tore et al. (2013) for whom the knowledge is codified in the form of motor responses and acquired in the action. An interpretation as an interface modality is offered instead by Froese et al. (2012), whom interprets the interfaces as augmented sense-making that serve as a transparent medium for augmenting our natural skills of interaction with the world. An interpretation of the enaction as system interaction is offered by Fedeli & Rossi (2011) for whom a system transforms itself thanks to the interaction between different subjects and to the mind-body-artifact-world continuum. Finally, an example of interpretation as the original Varela definition is proposed by Pugliese & Lehtonen (2011) that synthesize the concept as a continuous two-directional loop of bodily interaction. This process is a feedback loop: the actions performed by the enactor affect the medium that in turn affects the following actions of the enactor. Focusing the attention on the 13 papers related to the Human-Computer Interaction area, we can notice a bigger concentration in the latest 5 years, with a balanced distribution between journals and conferences (Table 7).

Table 7: Studies related to HCI.

Paper	Journal	Conference
(Bordegoni & Cugini 2013)	x	
(Bordegoni & Cugini 2009)		x
(Davis, Comerford, et al. 2015)		x
(Davis, Hsiao, et al. 2015)		x
(Flint & Turner 2015)	x	
(Froese et al. 2012)	x	
(Kaipainen et al. 2011)	x	
(Luyat & Regia-Corte 2009)	x	
(Pacheco & Souza-Concilio 2013)		x

(Pugliese & Lehtonen 2011)	x	
(Saidali et al. 2009)		x
(Todaro et al. 2007)		x
(van der Linden et al. 2011)		x

Topics of these papers are related to three main subjects: Interfaces (enactive interfaces, haptic interfaces, multimodal interfaces); Design (system design and evaluation, tool design and evaluation, digital environment design, enactive systems); Perception (appropriation and perception, motion capture, affordances).

Within these studies, the concept of enaction is interpreted as system interaction (4 studies), interface modality (6 studies) and as the definition of Varela et al. (8 studies). 4 of these studies are theoretical, while the other 9 are mainly practical and empirical. All the practical studies use or propose a technological tool; among them 3 simulation software, 1 data visualization and 7 NUI. Among the theoretical studies, one of them uses a software game.

4. Discussion

Although the initial search for “enactive system” returned few results, a more thorough search for “enaction”, “enactive” or “enactivism”, returned an expressive number of papers. It is worth noticing that most papers come from journals, where more mature research is usually presented. In the last ten years, a growing presence of the subject was detected. Most contributions come from Europe, mainly associated to the fields of Artificial Intelligence, Educational Technology and HCI. The Significant presence in AI seems related to the development of a new paradigm in robotics. The main topics addressed the subject associating it to computational modeling of emotion and virtual reality. There is a good balance among the papers between theoretical and empirical contributions, and also among the contributions that propose or use tools and those that do not. For those studies that embed tools, those usually refer to NUI (natural user interfaces). Experimentations with NUI and virtual reality are located especially in the educational technology field.

Within the field of Human Computer Interaction, the studies address issues mainly related to the design and evaluation of new types of interfaces (enactive, multimodal, haptics), and to an in-depth analysis of perception in relation to enaction and design solutions. This set of facts gives us a snapshot of the use of enactive systems in the computing field research in the last ten years.

5. Conclusions

The enaction may be understood as the quality of interactive systems that allow organization of knowledge obtained through a circular action. Examples are interfaces that couple a human with a machine to do things usually done unaided, such as shaping a three-dimensional object using multiple modality interactions with a data base or using interactive video to allow a student to visually engage with mathematics concepts (Bordegoni 2010; Tall et al 2008). Although several understandings for the term exist, depending on the knowledge domain area, the SLR revealed a wider adherence to the traditional concept of “enaction”, as proposed by Varela et al (1991). Within these coordinates, the concept of enaction links biology and phenomenology, emphasizing the interaction of the embodied organism with its environment. It is not a question of the

separation of the representation and the action in an objective world, but it is the same activity of the embodied organism that produces a world of meaning (Nehaniv et al. 2013). The presence of the majority of studies in Europe and America seems to suggest the complexity and multidisciplinary nature of the topic that, based on a large philosophical background, is positioned in a breaking line with the tradition. These factors, as well as the cultural conditions and freedom in experimenting with new solutions, can strongly influence the understanding and design of enactive systems. Moreover, the SLR exposed a lack of studies regarding the investigation of the "enaction" concept in relation to the web or social systems. This indicates an opportunity for research, especially in the field of HCI that may lead to innovative ways of understanding the social phenomena mediated by technology. Further work involves deepening the subject to include specificities of different knowledge domains to the concept and use of enactive systems.

Acknowledgements

This work was conducted during a PNPd/CAPES scholarship at the University of Campinas.

References

- Bordegoni, M., 2010. Design tools based upon enactive interfaces. In Shuichi Fukuda, ed. *Emotional Engineering: Service Development*. Springer. pp. 78.
- Bordegoni, M. & Cugini, U., 2009. Multimodal Perception-Action Interaction for the Exploitation of Enactive Knowledge. In *Volume 2: 29th Computers and Information in Engineering Conference, Parts A and B*. ASME, pp. 983–990.
- Bordegoni, M. & Cugini, U., 2013. Tools for Industrial Design: From Barriers to Enablers of Creativity. *Computer-Aided Design and Applications*.
- Brereton, P. et al., 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80(4), pp.571–583.
- Davis, N., Comerford, M., et al., 2015. An Enactive Characterization of Pretend Play. *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition - C&C '15*, pp.275–284.
- Davis, N., Hsiao, C., et al., 2015. Drawing Apprentice : An Enactive Co-Creative Agent for Artistic Collaboration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. pp. 185–186.
- Domingues, D. et al., 2014. Embodiments, visualizations, and immersion with enactive affective systems. In *Proc. SPIE 9012, The Engineering Reality of Virtual Reality 2014*.
- Fedeli, L. & Rossi, P.G., 2011. A study of teacher/student relations in a formal face-to-face university context supported by online tools within an enactive approach. *Procedia - Social and Behavioral Sciences*, 28, pp.673–678.
- Flender, C., 2008. Ecological Modelling of Information Systems. In *Proceedings of CAiSE-DC*. pp. 40–52.

- Flint, T. & Turner, P., 2015. Enactive appropriation. *Ai & Society*, pp.1–9.
- Froese, T. et al., 2012. The enactive torch: A new tool for the science of perception. *IEEE Transactions on Haptics*, 5(4), pp.365–375.
- Kaipainen, M. et al., 2011. Enactive Systems and Enactive Media: Embodied Human-Machine Coupling beyond Interfaces. *Leonardo*, 44(5), pp.433–438.
- Kitchenham, B., 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(TR/SE-0401), p.28.
- Laboratory of Research on Software Engineering (LaPES) -Computing Department of the Federal University of São Carlos (DC/UFSCar), 2010. StArt (State of Art through Systematic Review). Available at: http://lapes.dc.ufscar.br/tools/start_tool.
- van der Linden, J. et al., 2011. Haptic reassurance in the pitch black for an immersive theatre experience. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. pp. 4686–4691.
- Luyat, M. & Regia-Corte, T., 2009. The affordances : From James Jerome Gibson to the recent formal models of the concept. *Annee Psychologique*, 109(2), pp.297–332.
- Nehaniv, C.L. et al., 2013. Interaction and experience in enactive intelligence and humanoid robotics. *Artificial Life (ALIFE), 2013 IEEE Symposium on*, pp.148–155.
- Pacheco, B. a. & Souza-Concilio, I. a., 2013. Multimodal interfaces: An enactive approach. In *4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2013 - Proceedings*. pp. 51–58.
- Pugliese, R. & Lehtonen, K., 2011. A framework for motion based bodily enaction with virtual characters. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6895 LNAI, pp.162–168.
- Saidali, Y. et al., 2009. Aid to the interpretation of legal documents: A user-centered approach. In *CORIA 2009 - Conférence en Recherche d'Information et Applications*. pp. 457–468.
- Todaro, E. et al., 2007. The Hand Dance : A Didactic Performance Platform. In *International Cultural Heritage Informatics Meeting (ICHIM07): Proceedings*,.
- Di Tore, P.A., Discepolo, T. & Di Tore, S., 2013. Natural user interfaces as a powerful tool for courseware design in physical education. *Journal of E-Learning and Knowledge Society*, 9(2), pp.109–118.
- Tall, D, Smith D., Piez C. 2008. Enactive control. In Mary Kathleen Heid, Glendon W. Blume, eds. *Research on Technology and the Teaching and Learning of Mathematics*. Information Age Publishing Inc. pp. 213.
- Varela, F.J., Thompson, E. & Rosch, E., 1991. The Embodied Mind: Cognitive Science and Human Experience. *An International Journal of Complexity and*, 1992, p.328.
- Vernon, D. et al., 2007. A Survey of Artificial Cognitive Systems : Implications for the Autonomous Development of Mental Capabilities in Computational Agents. , 11(2), pp.151–180.

Smart Control of Expenses Using Mobile Devices

Giovani R. F. Junior¹, Vicente J. P. Amorim¹, Thiago L. Gomes², Igor M. Pereira¹

¹iMobilis Lab – Department of Computer and Systems(DECSI)

ICEA - Federal University of Ouro Preto (UFOP) - João Monlevade - MG - Brazil

gr.ferregetti@gmail.com, vjpmorim@icea.ufop.br, igormuzetti@gmail.com

²Department of Computer and Systems(DECSI)

ICEA - Federal University of Ouro Preto (UFOP) - João Monlevade - MG - Brazil

thiagoluange@gmail.com

Abstract. *It is estimated that only 3.1% of the Brazilian population controls their expenses through digital applications. While 8.9% does not use a digital platform due to a lack of knowledge, 8.1% do not have time. Considering the current usage levels, applications providing a more automated control of expenditures would simplify this task for an average user, making mobile applications a more attractive option. Using digital image processing techniques, optical character recognition (OCR), and post-processing a novel software was developed in Android for receipts automatic recognition allowing mobile users to monitor expenses using photos. Information recognized by the application replicated a real receipt with an acceptable level of similarity.*

1. Introduction

According to the study conducted by the Brazilian Credit Protection Service (SPC Brasil) [Bruno and Miret 2016], only two in five Brazilians properly take care of their expenses. Furthermore, only 3.1% of Brazilians control their spending using digital applications. Compared to the last annual study conducted by SPC Brasil, [Bruno and Miret 2016] and [Gandra 2015], the percentage of people that use digital applications to control spendings reduced from 4% to 3.1%. Also, the percentage of individuals who used paper for controlling spending increased from 23% to 29.8%.

In a separate study conducted through Connect, the Brazilian Institute of Public Opinion and Statistics (IBOPE) digital platform [Administradores 2013], 69% of Brazilians believe that mobile technologies, such as tablets and smartphones, positively contribute to improving population's quality of life. Additionally, 66% of respondents believed that mobile devices are essential to becoming informed consumers. Although that, consumers that are controlling their expenses are abandoning the use of these electronic devices. One hypothesis blames currently available applications, which fail to provide a satisfactory performance or function to meet consumer needs: difficult use, or require excessive amounts of time and effort to use[Zhel 2016][Google 2016c]. Besides that, from the user's point of view, inserting data from each receipt might become very annoying and time consuming.

To improve consumer use of technology to monitor their expenses, this paper presents the development of a novel way to control financial data through a mobile application. Using digital image processing techniques and optical character recognition

(OCR), the approach proposed here analyzes a photograph taken by a mobile device to retrieve receipt information. Thus, providing a simple system for consumers to digitally keep track of all expenses.

Some work has been done to improve Digital Image Processing, Optical Character Recognition and Post-Processing techniques such as the proposed by Hang[Heng 2013], Hasnat et. al.[Hasnat et al. 2009] and Bassil et. al.[Bassil and Alwani 2012]. Next section makes a comparison between these works.

The remaining of this paper is divided into the following sections: Section 2 present the works that are close to the one described here. Section 3 depict main motivations that led to this work. Section 4.1 explains the two sets of receipts created and their purpose. Section 4.2 describes how the image processing techniques were selected. Section 5 show how a test application was built to validate our idea. By its turn, Section 6 present the results while Section 7 conclude this work and describe possible future works.

2. Related Works

Application developed by Hang (2013) [Heng 2013] used user images to determine how a bill will be split amongst people at a table in a restaurant or bar. After taking the picture, the user must then manually set regions containing items and prices in the image. The application recognizes the data and lists each contribution value based on items a person has consumed. This study mentions the use of an adaptive algorithm for binarization, despite no specific technique is specified. After binarization, resulting image is recognized using the Tesseract library[Library 2016] without any training. After that, the text is processed, and items contained in the pictures are identified. Unlike the work presented here, this proposal trusts on a user intervention to select specific parts of a picture where the most important part of information is.

Hasnat et al. [Hasnat et al. 2009] designed an OCR process software for the Bengali language in combination with the Tesseract library, which was called BanglaOCR. This paper focused mainly on Tesseract training and post-processing techniques. Initially, the authors created fourteen different training sessions to determine what combination of training produced optimum results. Therefore, it developed a post-processing on two levels: correction recognition problems based on specific language's rules, for instance, the characters order; and the use of eighty thousand dictionary words, whose function is to check words spelling in the text. Final result of recognition was 93 % for high-quality images. This paper obtains the document picture using a scanner with a specific angle, different from what we proposed here, which is a receipt picture obtained by a mobile camera, with the angle determined by the user through the guidelines.

Bassil and Alwani [Bassil and Alwani 2012] proposed a post-processing algorithm to correct syntax and semantic mistakes based on their context. Presented algorithm uses an online word correction service maintained by Google. It divides resulting text of optical character recognition in five words blocks, which was judged to be enough to produce a context. These blocks are sent to Google's correction service. When a response did not contain the phrase "you mean ...", the five words block was considered as a correct block; and when the answer did contain that phrase, the old block is replaced by the suggested one. According to the authors, proposed algorithm improved error correction rate by 690% for a document in English and 403% for a text in Arabic. Although this tech-

nique can correct syntactic and semantic errors, it depends on the Google's service and an internet connection. Here, we propose a post-processing mechanism using a framework which is built inside Android and does not need a permanent internet connection.

Other projects have very similar proposals. As an example, there is a Brazilian project called NFSCAN [Application 2015], which uses digital image processing and optical character recognition to identify data from receipts to ease credits donation to Non-Governmental Organizations. However, the lack of a specific documentation about its implementation and results prevents a further discussion.

3. Motivations and Main Objectives

As presented before, expenditures control via mobile applications still not popular. Reasons for this unpopularity are on mobile software complexity, lack of usability or even due to the lengthy time required to manage provided information. Low user adoption of digital methods to control expenditures may also be caused by the lack of specific functionalities. To provide an easy way of taking control of expenses and prevent users from having to enter all the receipt's relevant information would reverse that digital abandonment. Therefore, a tool that can interpret receipt's data and perform necessary calculations automatically is essential. Indeed, the user can simply use a smartphone to photograph the receipt and thus keep track of expenses.

To implement an application based on photographic data recognition to control expenses, digital image processing methods and optical character recognition must obtain relevant information from the receipt. However, depending on image quality and character font used to write the receipt, these two techniques may not be enough. As a way to circumvent this fact, previous works [Hasnat et al. 2009] [Bassil and Alwani 2012] proposed a step called post-processing, known as a way to enhance information obtained by OCR through the use of online or offline spelling checkers. Since internet access is not available all the time for the users, any application that performs digital image processing, OCR and post-processing shall let it into account.

Considering the points raised above, the solution presented by this work covers the creation of an application to automatically identify data from a generic receipt allowing the users to control his expenses in a more fashionable way.

4. Data Sampling and Test Set

This section describes the selection process of receipts and techniques considered by this work.

4.1. Selection of Receipts

To apply those methods described in the next section and get final results from the method proposed here, we created two sets that include all types of receipts: new, old, crumpled, with different fonts and layouts.

The first set has 50 receipts which are used to compare image processing methods and create 53 training regiments. The second set has 30 receipts which are used to apply the post-processing and obtain the final recognition rate.

4.2. Selection of Techniques

This section describes main methods used during development of this project. Digital image processing methods, such as binarization and segmentation, were used to improve OCR results. Training methods are commonly used when is necessary to apply OCR over a non-standard fonts and idioms. By its turn, post-processing is used to correct spelling mistakes in the results coming from OCR tool.

4.3. Digital Image Processing

Several tests were performed involving filters and digital image processing techniques, for example: binarization, median filter, Gaussian filter, erosion and dilation, and transformed perspective. These techniques were taken into account as a way to improve the recognition rate of information retrieved from the receipts. However, some processes degraded the picture quality and its characters, even though it results in a greater recognition rate, that occurs owing to the noise reduction.

4.3.1. Binarization

Binarization algorithms compare each pixel value to a threshold value, that can be fixed, calculated using the whole image, or even calculated using different regions of a picture. So, if the pixel value is greater than a threshold value, the pixel is set to foreground. Otherwise, the pixel value is set to background[Gonzalez and Woods 2010].

Binarization method with a fixed threshold value proved to be the least efficient technique because it can not process lighting differences, such as shadows. Indeed, in the work presented here, binarization function provided by Tesseract library produced worst results comparatively to the top 4 algorithms, as shown by Figure 5.

In contrast, best results were obtained by both local and global adaptive techniques, which adjust threshold value to the image being binarized. Still, two best results were produced by Sauvola's and Wolf's algorithm, which represent the local adaptive binarization technique. Unlike global adaptive binarization method, the threshold of each region of an image adapts such region in a way that it will have a specific threshold value. Therefore, binarization algorithm chosen during application development was the one proposed by Sauvola due to its better results.

4.3.2. Segmentation

Segmentation is the process of dividing one image into different regions so that they can be treated separately later[Gonzalez and Woods 2010]. Tesseract library[Google 2016b] provides a function that allows image segmentation into lines, words and symbols, for images that contain text. However, our results showed that recognition was higher without segmenting the receipt, as shown by Figure 6. Consequently, a segmentation method was not considered by the application.

4.4. OCR Tool Training

Tesseract library can recognize common fonts from different languages. It is necessary to tell the library how to recognize new fonts or new languages, through a process called

training [White 2012].

Training occurs by representing characters in an image with a text symbol contained in that picture. For instance, an alphabet letter is represented by a particular image segment. Furthermore, it can be taught how this new “language” works by combining its symbols, forming a dictionary [White 2012]. This work have used the Tesseract library to recognize receipts text.

4.4.1. Training Files

Initially, 53 training regiments were created and tested to determine the success of a training file. From this set of training regiments, 50 were created using each test file, 1 was created using the whole set and the last 2 were obtained from Google’s page. Specifically at the initial phase, 50 receipts were considered for the first training set. These training files were obtained in two steps: Qt-box-editor[Qt 2016] tool was used to create box files, which was in charge of teaching Tesseract[Library 2016] what means each character symbol in the image, creating a binary file with a “tr” extension; moreover, the Tesseract library was used to generate the training file. Figure 7 presents 13 most significant training files.

Analyzing Figure 8, some of the training files produced better results recognizing numbers than characters. However, post-processing phase explained in section 4.5, correct words that are incorrectly spelled and cannot make any correction to the segments that contain numbers. Then, to perform a comparative analysis of the post processing impact, two training files were chosen to be tested with the second set that contains 30 receipts: the one with the highest percentage of numbers recognition, called “por” training file; and the one with second highest average recognition including all types of symbols, called “t16” training file.

4.5. Post-processing

Post-processing main objective is to detect and correct spelling mistakes created by optical character recognition tool.

As explained in detail by [Bassil and Alwani 2012], there are two types of errors: semantic and syntactic. Therefore, there are also three approaches of correction: **Manual error correction**: a group of people is hired to correct the errors manually. Although it is the easiest way, it requires continuous human intervention; **Dictionary-based correction**: corrects the mistakes based on a dictionary; and **Context-based correction**: correct the mistakes based on grammatical error and semantic context. In this paper, we implemented a dictionary-based correction method using Android’s Spelling Checker Framework[Google 2016a].

4.6. Levenshtein-Distance Algorithm

This algorithm calculates the minimum number of operations needed to transform one string of characters to another. Allowed operations are: insertion, substitution, and removal[Levenshtein 2016]. Therefore, from the number of required operations, is possible to calculate a similarity level between two characters sequence. Here, the Levenshtein

algorithm was used to calculate the similarity level between text automatically recognized and the original text – previously loaded to text files by a human developer.

5. Application Development

Bellow there is a description of methods used to implement the application main parts.

5.1. Implementation

As discussed in section 4.2, methods implemented by the application were: **Binarization Method**: Sauvola method provided best results; **Segmentation Method**: Omitted because recognition percentage worsened; **Training file**: “por” training file provided best number of recognitions among the training files explored with satisfactory character recognition; and **Post-processing**: Android’s Spelling Checker Framework was chosen because it provided an offline way to access the data.

5.2. Camera

This work proposes a test application called “NotaFiscal”, that has an initial screen similar to a common camera application. When the application starts, it is possible to see camera image, as shown in Figure 1. A green line, as presented by Figure 1, aims to assist the user in aligning the camera and the receipt. In Figure 2 there are two lines which also serve as a guide. Additionally, operation mode described above is complemented with an edge detection system to allow a better visualization and alignment.

With a screen tap, a receipt picture is taken, which will result in image saving and character recognition. To enable edge detection function, the user must execute a screen long tap for a few seconds.

5.3. Storage

To allow a later usage application saves data obtained from optical character recognizer right after it is processed by post-processing step. A new file is created in “recAppNotaFiscal” folder as shown by Figure 3.

Images, in a binarized form, are stored in picture’s folder, as shown in Figure 4. Storing the pictures in a binarized form is justified by its size, which is much lower when compared to the colored version.

Indeed, training file used in this project, “por” is stored inside “tessdata” folder as shown in Figure 3.

6. Evaluation and Results

This section discusses results of optical character recognition (OCR) and post-processing techniques using the two best training files, “por” and “t16”.

It is important to emphasize that OCR results only involve methods with best results in our tests. Those methods are Sauvola’s Algorithm for Binarization, without image segmentation, and the best training files. Furthermore, results of proposed post-processing method are based on Android’s internal dictionary.

As shown in Figure 9, the average result of OCR using “por” training file was 61.28%, with the best individual result being 78.0%. Post-processing enhanced similarity to 63.34% on average without changing best similarity in individual results.



Figure 1. Application initial screen presenting the receipt and guideline.



Figure 2. Application screen showing guide-lines and edge detection.

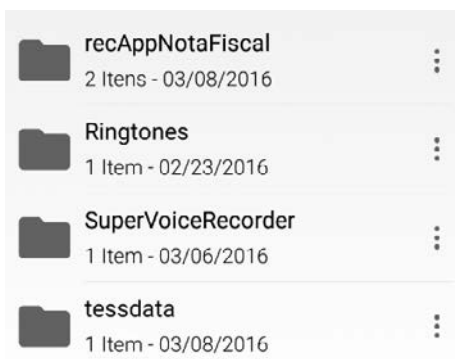


Figure 3. Application folders to store training files and post-processing results.



Figure 4. Album containing receipt photos taken by the user.

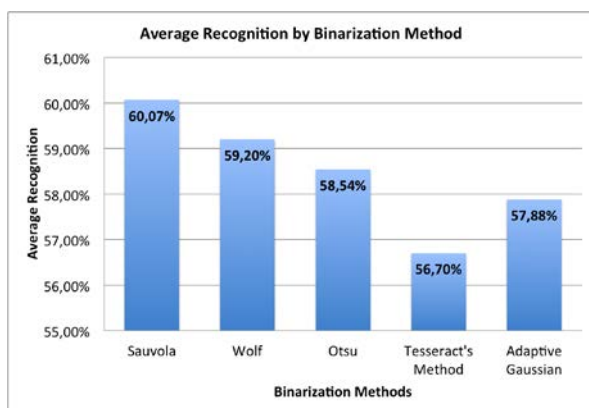


Figure 5. Tests results evaluating the best binarization method. For each set, 50 receipts were taken into account.

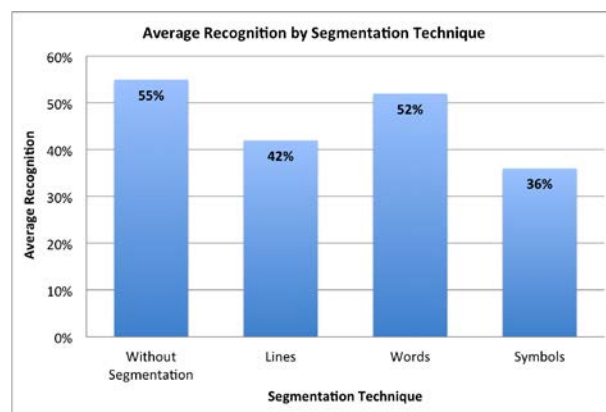


Figure 6. Tests results evaluating the best segmentation method. For each set, 50 receipts were taken into account.

Average result for “t16” training file was 63.18% (Figure 10). Post-processing enhanced similarity to 67.07%. In addition to that, best OCR result had its similarity decreased from 78.0% to 66.86%, so the best result after post-processing was 75.0%.

It was observed that OCR results can be affected by receipt’s characteristics. For instance, the first factor is receipt font non-standardization. Other influences include printing quality and how old receipt is. Thus, the receipt should be photographed as soon as possible to provide the best OCR results. Moreover, receipt physical state, e.g. scratched or cropped, also had a large impact on its recognition rate.

Hence, OCR results had the greatest impact on post-processing technique, specifically the exchange of letters, numbers and special characters. Techniques discussed here were not applied to correct sections of text that have numbers and special characters. This fact is mainly because it can have particular information about purchases. Indeed, the focus of this work was on syntactic and semantic errors correction.

This work also implemented an algorithm to split OCR results into sections of text through the recognition of blank spaces. If these sections only contain letters, the algorithm looks and replaces old section by the word(s) that has the biggest similarity level. Table 1 shows information regarding each receipt after post-processing. This information is divided into a total number of sections, number of words, number of correct words, and number of words that were replaced by a similar word.

It is possible to conclude that post-processing enhanced recognition rate for both training files, considering the average result. Interestingly, post-processing applied to OCR result of “t16” training file led to an enhancement up to 12.69% for one receipt. However, it also resulted in a decrease of 16.01% of similarity for another receipt. Regarding the “por” training file, post-processing led to an improvement up to 7.57%, while the similarity level decreases up to 0.27%, which is almost 60 times smaller than in training file “t16”.

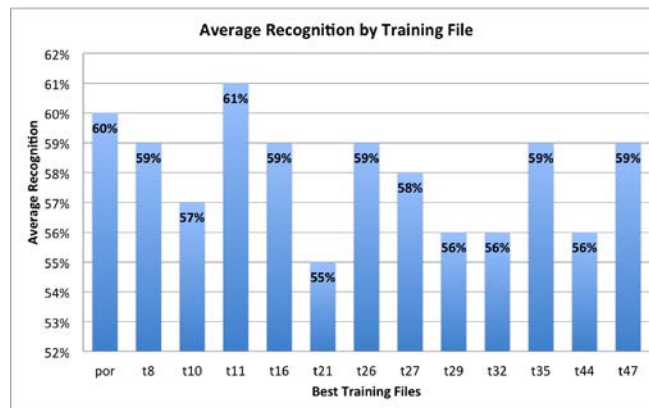


Figure 7. Tests results evaluating the best 13 training files from the set of 50 receipts.

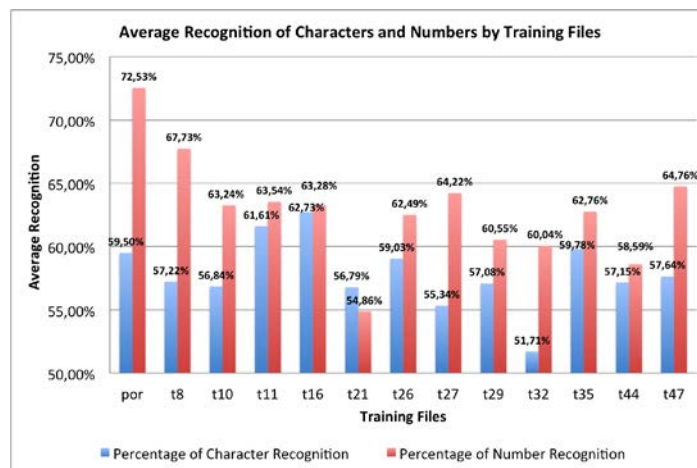


Figure 8. Tests results figuring out the percentage of character recognition and number recognition from the set of 50 receipts.

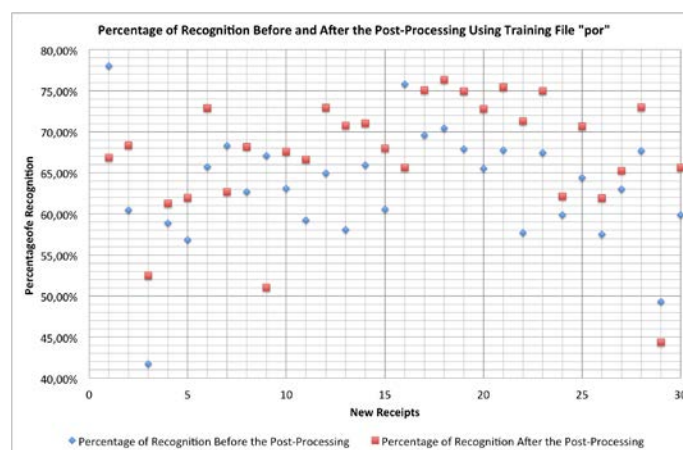


Figure 9. Results considering a new set of 30 receipts. Post-processing was applied on OCR results with “por” training file.

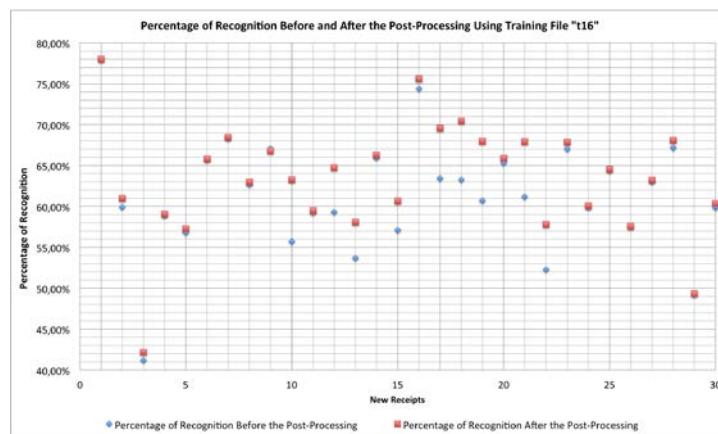


Figure 10. Results considering a new set of 30 receipts. Post-processing was applied on OCR results with “t16” training file.

7. Conclusions and Future Works

Recently, technology has evolved to provide a better quality of life for the population by simplifying everyday tasks. Activities that previously demanded hours to be completed can now be executed in few seconds with the aid of technology. This paper proposes a development of an application that can help consumers to take control of their expenses. Arduous task of creating spreadsheets and performing math can be exchanged by a simple task of taking a photo of the receipt in conjunction with the convenience of a smartphone.

When compared to similar works, application proposed here achieved good results. Our work specifically considers characters automatic recognition in the mobile environment without any user intervention. Furthermore, recognition rate obtained here encloses different formats of documents, which also use different sets of fonts.

This application development was made possible by the combination of best digital image processing, optical character recognition, and post-processing techniques. Results revealed a significant impact when considering a specific set of documents – receipts. In summary, for this specific case, best binarization method was proved to be the one proposed by Sauvola. Interestingly, image segmentation into lines, words, and characters resulted in a worse recognition rate. Also, the training file here called “por” provided by the Tesseract’s library web page resulted in the best percentage of characters recognition. Finally, the post-processing was demonstrated to be very effective on the similarity enhancement in a comparison between the recognized words and the real words held on receipt.

Each technique considered by application final version was chosen by decision making based on real data. Therefore, during the evaluation of each technique, it was possible to visualize an increased similarity between the original receipt and recognized data. Regarding techniques used in this study, it was observed that Sauvola’s algorithm results in better information recognition. In addition, the training file available on Google’s page, “por” training, has the best average character recognition rate, having a great importance in approach proposed by this work. It was also observed that no receipts segmentation technique resulted in better recognition rates. Finally, the post-processing anchored in Android’s platform native spelling checker produced a significant improvement in the

"por" Training					"t16" Training				
Receipt	Number of Segments	Number of Words	Exact Words	Replaced Words	Receipt	Number of Segments	Number of Words	Exact Words	Replaced Words
1	114	44	0	3	1	115	42	0	2
2	88	25	0	5	2	91	25	1	6
3	83	29	0	7	3	81	22	5	2
4	164	62	0	4	4	165	53	0	6
5	164	57	0	8	5	163	55	6	7
6	163	52	0	3	6	163	56	14	7
7	95	36	0	5	7	97	33	0	5
8	155	54	0	6	8	154	50	1	5
9	112	34	0	5	9	111	39	6	5
10	231	90	0	4	10	231	84	2	6
11	110	36	0	4	11	108	34	0	5
12	157	53	0	4	12	156	48	6	7
13	150	57	0	6	13	148	47	11	7
14	132	50	0	4	14	132	47	1	4
15	164	61	0	3	15	163	54	2	5
16	173	58	0	4	16	175	50	0	11
17	201	64	0	7	17	202	62	0	10
18	209	77	0	4	18	209	73	0	7
19	216	71	0	4	19	221	74	11	5
20	113	37	0	7	20	113	37	1	6
21	175	65	0	6	21	175	64	0	4
22	175	62	0	1	22	177	56	9	3
23	143	51	0	8	23	145	49	3	5
24	139	40	0	2	24	140	39	0	6
25	131	45	0	3	25	132	44	1	6
26	123	44	0	6	26	122	46	0	4
27	149	49	0	4	27	147	50	0	6
28	160	50	0	6	28	157	51	1	5
29	94	29	0	3	29	94	28	1	2
30	140	43	0	6	30	142	41	0	4
Mean	147.43	50.83	0	4.73	Mean	147.63	48.43	2.73	5.43

Table 1. Results of OCR using "por" and "t16" training files, number of segments that represent words and number of replaced words.

final recognition rate.

In addition to evaluation data, this research proposed an initial method of post-processing, however, this area has great potential for growth. Several other techniques can be implemented to enhance similarity of the recognized text further, considering the same mobile environment. Therefore, next possible step of the post-processing could use machine learning and/or another technique to correct recurring mistakes that arise from OCR. Future works could also analyze the impact of an automatic correction method based on machine learning. As proposed in [Hasnat et al. 2009], some correction methods can be implemented to correct spelling mistakes based on information that has a specific format and predefined rules, such as, date, hour, total amount and etc.

Finally, the application proposed here has a great potential to improve expenses management. In addition to providing an alternative and much less laborious method, this method provides organized data storage, which traditionally is made by several folders or spreadsheets. Therefore, it is expected that with some improvements made by future studies, this application may have a significant effect on the users' life quality boosting mobile digital applications to financial managements.

References

- Administradores (2013). Para brasileiros, smartphone e tablet melhoram a qualidade de vida. <http://www.administradores.com.br/noticias/tecnologia/pesquisa-para-brasileiros-smartphone-e-tablet-melhoram-a-qualidade-de-vida/75469/>. Last accessed 2016, March 09.
- Application, N. (2015). <http://www.nfscan.cc/?lang=pt>. Last accessed 2016, February 29.

- Bassil, Y. and Alwani, M. (2012). Ocr post-processing error correction algorithm using google's online spelling suggestion. *Journal of Emerging Trends in Computing and Information Sciences*.
- Bruno, V. and Miret, R. (2016). 46% dos brasileiros não controlam seu orçamento, revela pesquisa do spc brasil. <https://www.spcbrasil.org.br/imprensa/noticia/971-46dosbrasileirosnaocontrolamseuorcamentorevelapesquisadospcbrasil>. Last accessed 2016, March 09.
- Gandra, A. (2015). Spc: falta de disciplina é maior dificuldade para controle de gastos. <http://agenciabrasil.ebc.com.br/economia/noticia/2015-01/falta-de-disciplina-e-principal-dificuldade-dos-brasileiros-para-nao>. Last accessed 2016, February 02.
- Gonzalez, R. C. and Woods, R. E. (2010). *Processamento Digital de Imagens*. Pearson Education, 3 edition.
- Google (2016a). Android spelling checker framework. <http://developer.android.com/intl/pt-br/guide/topics/text/spell-checker-framework.html>. Last accessed 2016, February 29.
- Google (2016b). How to use the tools provided to train tesseract 3.0x for a new language accessed 08 mar. 2016. <https://github.com/tesseract-ocr/tesseract/wiki/TrainingTesseract>. Last accessed 2016, March 08.
- Google (2016c). Mobile app marketing insights: How consumers really find and use your apps. <https://think.storage.googleapis.com/docs/mobile-app-marketing-insights.pdf>. Last accessed 2016, April 01.
- Hasnat, M. A., Rahman, M., and Khan, C. M. (2009). An open source tesseract based optical character recognizer for bangla script. *10th International Conference on Document Analysis and Recognition*.
- Heng, A. (2013). L'addition: Splitting the check, made easy. https://stacks.stanford.edu/file/druid:yt916dh6570/Heng_LAddition_Restaurant_Check_Splitting.pdf. Last accessed 2016, March 09.
- Levenshtein (2016). The levenshtein algorithm. <http://www.levenshtein.net/>. Last accessed 2016, March 02.
- Library, T. O. (2016). Tesseract ocr library source code. <https://github.com/tesseract-ocr>. Last accessed 2016, march 01.
- Qt (2016). Qt box editor tool. <https://zdenop.github.io/qt-box-editor/>. Last accessed 2016, march 08.
- White, N. (2012). *Training Tesseract for Ancient Greek OCR*. Department of Classics and Ancient History 38 North Bailey Durham, DH1 3EU, UK, Email: nick.white@durham.ac.uk.
- Zhel, M. (2016). 10 reasons why people abandon your app. <https://mofluid.com/blog/10-reasons-why-people-abandon-your-app>. Last accessed 2016, April 01.

Explorando Dados Abertos Governamentais Sobre a Mobilidade Urbana na Cidade do Rio de Janeiro

Sérgio Manuel Serra da Cruz^{1,3}, Luan Soares Andrade^{1,2},
Jonice Oliveira Sampaio²,

¹Universidade Federal Rural do Rio de Janeiro - LabData/UFRRJ

²Universidade Federal do Rio de Janeiro - LabCores/UFRRJ

³Programa Educação Tutorial - PET-SI/UFRRJ

serra@pet-si.ufrrj.br, oluan@ufrj.br, jonice@dcc.ufrj.br,

Resumo *Um dos grandes desafios das cidades brasileiras é a busca de alternativas para melhorar a mobilidade urbana. O conceito de cidades inteligentes tem despontado como a combinação do ambiente digital com as comunidades reais com vistas a tornar mais eficiente a gestão pública dos espaços urbanos. Soluções baseadas em dispositivos móveis que utilizem dados abertos governamentais são consideradas importantes diferenciais. Este trabalho propõe uma arquitetura distribuída intitulada BusInRio e experimentos que utilizam dados abertos governamentais que permitem aos usuários de transporte público da cidade do Rio de Janeiro acompanharem rotas e a localização dinâmica dos ônibus de uma linha e o estado do trânsito.*

Abstract *One of the great challenges of the Brazilian cities is the search of best opportunities to improve the urban mobility. The concept of smart cities has emerged as a combination of the digital environment with real communities aiming to making the management public of urban space more efficient. This paper presents a mobile distributed architecture named BusInRio, its prototype and experiments. BusInRio uses open government data to aid user of public transport of the city of Rio de Janeiro monitor and evaluate routes and dynamically recognize the location of the bus lines and the traffic status.*

1. Introdução

A gestão pública vem se modernizando e recebendo diversas demandas. Um dos principais desafios para o avanço da democracia é o desenvolvimento contínuo das ferramentas que aprofundem a interação entre o governo e a sociedade. Neste sentido, ao lado das iniciativas de *e-Governo*, surgiu o conceito de Dados Abertos (Portal Dados Abertos, 2013) numa injunção para assegurar a disponibilidade de dados governamentais para a sociedade.

O conceito se refere aos dados que podem ser usados, modificados e compartilhados livremente por qualquer pessoa e por qualquer propósito, estando sujeito apenas, e no máximo, aos requisitos que preservem sua proveniência e compartilhamento pela mesma licença (Portal Dados Abertos, 2013, Open Data Definition, 2016). Diversos eventos recentes contribuíram para a disseminação do tema na sociedade brasileira.

Citamos a Lei de Acesso à Informação¹, a criação da Infraestrutura Nacional de Dados Abertos², além de diversas iniciativas internacionais tais como o Global Open Data Index (index.okfn.org) e Open Data Institute (theodi.org) que incentivam o uso das TIC (Tecnologia da Informação e Comunicação) e dados abertos como um esforço para tornar os governos mais transparentes, efetivos e confiáveis, com instituições que empoderam os cidadãos e escutam suas necessidades. Atualmente diversos países mantêm portais com diversos tipos de dados abertos governamentais, dentre eles se destacam: data.gov.uk, open.canada.ca, data.gov e data.gov.br.

O conceito de Cidades Inteligentes (Dohler et al. 2011, Schaffers et al, 2011) tem despontado como a combinação do ambiente digital com comunidades reais, tornando mais eficiente a gestão pública dos espaços urbanos, o conceito é bastante amplo e possui diferentes conotações que podem variar segundo as perspectivas do usuário, da tecnologia ou da economia. Apesar de ser um conceito relativamente recente, ele já se consolidou como tema fundamental na discussão global sobre o desenvolvimento sustentável, estima-se que movimentará um mercado global de soluções tecnológicas estimado em US\$ 400 bilhões até 2020.

Atualmente, cidades de países emergentes estão investindo bilhões de dólares em produtos e serviços baseados na economia do conhecimento para sustentar o crescimento econômico e as demandas materiais da nova classe média. No entanto, um dos principais problemas enfrentados pelos gestores e que afeta grande parte população é a precariedade da mobilidade urbana. Problemas tais como engarrafamentos, falta de investimentos em infraestrutura e qualidade precária dos serviços de transporte afetam diretamente a economia e tornam a vida do cidadão mais estressante. Com vistas a explorar as oportunidades de negócio advindas deste conceito, diversas empresas ligadas à economia do conhecimento (Uber, Cabify, Waze, entre outros) oferecem serviços centrados em plataformas digitais distribuídas baseados em dados proprietários que visam oferecer alternativas aos desafios da mobilidade urbana.

A cidade Rio de Janeiro, assim como outras grandes cidades do mundo, já apresenta iniciativas interessantes sobre os conceitos de Cidades Inteligentes e Dados Abertos, por exemplo, ele anualmente promove o concurso de criatividade RioApps³ que estimula o desenvolvimento de *apps* que contribuirão para tornar a cidade inteligente. Além disso, ele mantém o Centro de Operações (centrodeoperacoes.rio) que monitora a cidade e realiza o gerenciamento de crises (chuvas, deslizamentos, acidentes). Ela também mantém repositórios de dados abertos governamentais através do portal (data.rio) sobre os mais diversos assuntos tais como: educação, saúde, impostos administração pública, meteorologia e transportes em formatos abertos para que empreendedores e desenvolvedores criem suas aplicações de interesse público. Cada dataset possui sua estrutura sintática própria e são ofertados em vários formatos abertos, a saber: XML, JSON, CSV, entre outros.

Este trabalho tem como objetivo contribuir com a oferta de soluções tecnológicas relacionadas aos desafios da mobilidade urbana nas grandes cidades brasileiras.

¹http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

²<http://www.governoeletronico.gov.br/acoes-e-projetos/Dados-Abertos/inda-infraestrutura-nacional-de-dados-abertos>

³<http://portalrioapps.com.br/>

Apresentamos uma arquitetura distribuída denominada *BusInRio*, seu protótipo e experimentos baseados em nuvem de computadores, que utiliza dados abertos governamentais sobre as viagens dos ônibus que são transformados através de regras Extração, Transformação e Carga (ETL) em bases NoSQL interligadas que oferecem informações acerca das linhas de ônibus para que o usuário do serviço possa escolher o modal que irá utilizar em função não só da localização na cidade do Rio de Janeiro como também sobre o estado do trânsito por onde a linha circulará.

O trabalho está organizado da seguinte forma: a seção 2 apresenta os trabalhos relacionados. A seção 3 discute os fundamentos teóricos, já a seção 4 apresenta os materiais e métodos adotados. A seção 5 apresenta a arquitetura *BusInRio* e a seção 6 discute os principais resultados. Por fim, a última seção apresenta a conclusão, limitações e discute os trabalhos futuros.

2. Trabalhos Relacionados

Diversas cidades do mundo oferecem *apps* que utilizam dados abertos sobre mobilidade urbana. As cidades Nova Iorque, Chicago e Londres publicam uma variedade de dados através das suas autoridades de trânsito sobre as rotas dos transportes públicos para que desenvolvedores, pesquisadores, planejadores urbanos, jornalistas de dados e *startups* os utilizem.

Por exemplo, na cidade de Londres o *app CityMapper* também está integrado com o metrô, em Nova Iorque existe o *app OneBusAway* que apresenta dados de transporte público integrando dados abertos sobre os modais trem, metrô e ônibus. Algumas cidades do Brasil, dentre elas o Rio de Janeiro, começam a oferecer alguns *apps* comerciais que utilizam dados abertos governamentais (RioBus, Cadê o Ônibus?, Buus⁴). O *app* RioBus é um sistema colaborativo de monitoramento de ônibus em tempo real que provê algumas funcionalidades semelhantes às primeiras versões do Buus. Os *apps*, utilizam as APIs de dados abertos de mobilidade urbana disponibilizada pela Prefeitura da cidade do Rio de Janeiro. Devemos ressaltar que esses *apps* não provêm informações técnicas, publicações ou mesmo maiores detalhes sobre sua arquitetura, algoritmos, estrutura de dados nem mesmo sobre o funcionamento pois se tratam de programas protegidas por interesses comerciais.

O sistema UbiBus (Tito et al, 2012) tem como foco a cidade do Recife (PE), ele está baseado em um *middleware* que faz uso de informações contextuais de *tweets* e de computação ubíqua para se adaptar às situações do trânsito e recomendar rotas alternativas dos ônibus. O sistema Busão (Leite et al, 2013) tem como foco a cidade de Campina Grande (PB) e oferece aos usuários os itinerários dos ônibus e as melhores rotas dos ônibus. Ambos não utilizam dados abertos, mas dados do GPS coletados pelos passageiros.

Barbosa et al., (2014) e Bessa et al., (2015) desenvolveram soluções baseadas em dados abertos sobre os transportes do Rio de Janeiro, os autores oferecem recursos de visualização das trajetórias de ônibus do Rio de Janeiro. Este último trabalho tem como principal característica a detecção de *outliers* através de uso de redes neurais. Diferentemente, do *BusInRio*, essas soluções não discutem os aspectos arquiteturais das soluções apresentadas nem relatam possíveis integrações com aplicativos móveis nem

⁴ Desenvolvedor Tormenta Labs; Luis Picanço e Buus Serviços de Tecnologia (respectivamente).

representam os dados dinamicamente. Por fim, esses trabalhos não consideram a proveniência de dados abertos.

3. Fundamentação Teórica

A qualidade e a proveniência dos dados abertos (Mendonça et al., 2014) foram os principais fatores considerados para a estratégia de desenvolvimento do *BusInRio*. Nesta seção descreveremos os principais fundamentos teóricos utilizados neste trabalho.

3.1 Dados Abertos Governamentais

Os dados abertos governamentais dissem respeito a publicação e a disseminação das informações do setor público na Internet, sendo compartilhadas em formato bruto e aberto, compreensíveis logicamente, possibilitando o reuso por aplicações desenvolvidas pela sociedade.

A atividade de abrir dados ainda é realidade para uma pequena parcela dos governos nacionais. Se consideramos os governos municipais essa parcela é ainda menor. Para grande parte dos órgãos públicos, a falta de pessoal capacitado ainda é o principal motivo que contribui com essa realidade. Abrir dados não é uma tarefa trivial. Eles devem ser completos, primários, atuais, não proprietários, compreensíveis por máquina e não proprietários (OpenGovData, 2016). Os principais efeitos dos dados abertos nas políticas públicas são a inclusão social, a transparência governamental e a efetividade das políticas públicas.

3.2 Proveniência em Dados Abertos

A proveniência de dados representa a ancestralidade de um objeto e pode ser descrita em diferentes termos, dependendo do domínio abordado (Gil et al. 2010). Em geral, ela fornece a documentação essencial para registrar a origem, a qualidade e a autoria dos dados, assim como a capacidade de reproduzir e validar resultados de processos comerciais ou científicos. Os diferentes tipos de proveniência e seu nível de detalhamento (Cruz et al. 2009) são facetas que podem ser exploradas no contexto da qualidade dos dados abertos. Quanto menor a granulosidade, mais refinadas serão as possibilidades de investigação entre os dados publicados e seus descritores de proveniência. Nesse trabalho, adotamos a abordagem proposta por Mendonça et al. (2013) consideramos apenas a proveniência retrospectiva de baixa granulosidade relativa ao processo de ETL dos dados abertos.

3.3 Cidades Inteligentes

Apesar do conceito de Cidade Inteligente ser recente (Hernández-Munõz et al., 2011), ele já se consolidou como temática fundamental nas discussões sobre economia do conhecimento e mobilidade urbana. Ele diz respeito aos sistemas onde pessoas interagem e usam energia, materiais, serviços e financiamento para catalisar o desenvolvimento econômico e a melhoria da qualidade de vida. Esses fluxos de interação são considerados inteligentes por fazer uso estratégico das TICs juntamente com o planejamento e gestão urbana para buscar resposta às necessidades da sociedade (Dohler et al. 2011).

4. Materiais e Métodos

Esta seção apresenta não só a metodologia proposta para o desenvolvimento do *BusInRio*, como também descreve os datasets utilizados nesse trabalho.

4.1. Datasets e suas estruturas

Nesse trabalho utilizamos vários datasets abertos disponibilizados pela Prefeitura (percursos de cada linha de ônibus e dados em tempo real dos GPS de cada ônibus de cada linha). Resumidamente, os volumes de dados diários produzidos são medidos em dezenas de megabytes/dia. Os dados são de natureza textual e são caracterizados por milhares de registros produzidos pelos GPS de mais de 8.000 ônibus e arquivos das 490 linhas de ônibus que atendem o município.

Os dados contendo a localização das paradas (ponto de ônibus) de cada linha de ônibus são oferecidos pela FETRANPOR-RJ em formato CSV (<http://data.rio/dataset/pontos-dos-percursos-de-onibus>) e possuem o seguinte estrutura: Descrição de início e fim daquela linha; agência que disponibiliza os dados; número de sequência em que o ônibus passa em uma determinado ponto; coordenadas (latitude e longitude) do ponto de ônibus.

Os dados dos GPS instalados nos ônibus são de alta frequência, os dados de GPS com a localização atual de cada ônibus são produzidos a cada minuto e disponibilizados em tempo real em formato JSON (<http://data.rio/dataset/gps-de-onibus>). Este conjunto de dados cobre parte da frota das concessionárias da Cidade do Rio de Janeiro e algumas linhas intermunicipais. O *dataset* possui a seguinte estrutura: Data e hora da geração do dado; identificação alfanumérica encontrada na lateral do ônibus; identificador da linha do ônibus; coordenadas (latitude e longitude) do ônibus no trajeto; velocidade do ônibus no momento da coleta e medida em graus em relação ao norte, que representa a direção do veículo.

4.2. Metodologia proposta

Os dados de GPS das trajetórias dos ônibus apresentam alguns problemas tais como: ausência de informação da direção, erros de representação das coordenadas, erros de coleta (apesar da frequência ser a cada minuto, algumas medidas são superiores a 10~12 minutos). Neste trabalho, adotamos a abordagem ETL composta por três fases para garantir a qualidade e a integridade dos dados abertos, a saber:

1. *Coleta de dados brutos*: nesta fase os dados brutos são coletados a partir de diversas fontes do portal rio.data, eles são temporariamente armazenados em disco (*data staging area*) antes que qualquer tipo de manipulação;

2. *Limpeza e consolidação de dados*: nesta fase os dados brutos são corrigidos, transformados e consolidados em dados curados de acordo com os requisitos da aplicação e dos princípios do registro de proveniência de dados abertos ligados conforme apresentamos em trabalho anterior (Reiche, Cruz e Campos, 2014). Dentre as tarefas de limpeza realizamos exclusões de diversos tipos de registros que apresentam falhas: duplicados, falhos ou sem coordenadas; coordenadas erradas, trajetórias errôneas com distâncias superiores a 100 metros a partir da rota esperada; trajetórias com intervalos de GPS inferiores a 1 minuto ou superiores a 5 minutos. Como consolidação de dados calculamos a direção rota do ônibus e ligamos os dois datasets.

3. *Cargade dados curados*: nesta fase os dados consolidados são (re)agrupados em documentos ou família de colunas e carregados para os repositórios NoSQL apropriados.

4. *Anotação de proveniência retrospectiva*: nesta fase os metadados de proveniência do tipo retrospectiva de baixa granulosidade (Cruz et al, 2009) são produzidos pelos processos computacionais de limpeza e consolidação de dados abertos, os metadados também são armazenados nos repositórios de dados curados. As anotações de proveniência descrevem como o dado foi coletado e processado antes da publicação e em última análise garantem a transparência dos dados e das suas transformações.

5. Solução Proposta

Este trabalho propõe uma arquitetura, intitulada *BusInRio*, para a publicação de dados abertos curados e anotados com proveniência, baseia-se na descoberta de dados abertos governamentais da área de transporte público. O *BusInRio* é um serviço que não depende de plataformas tecnológica ou *demiddlewares* comumente utilizados em aplicações do tipo Cidades Inteligentes (Hernández-Munõz et al., 2011).

A arquitetura *BusInRio* é do tipo distribuída e escalável. A Figura 1 ilustra sua representação conceitual. Sua porção servidora executa em ambiente do tipo nuvem de computadores do tipo pública e modelo orientado a serviços SaaS (Chee, Franklin Jr, 2013) e a porção cliente executa em *smartphones* com diversos tipos de sistema operacional.

A arquitetura adota o padrão arquitetônico cliente-servidor. Os códigos da aplicação servidora foram desenvolvidos em linguagem PHP, já o cliente é totalmente desenvolvido tecnologias HTML5-JavaScript e os repositórios de dados curados e de descritores de proveniência são do tipo não relacional.

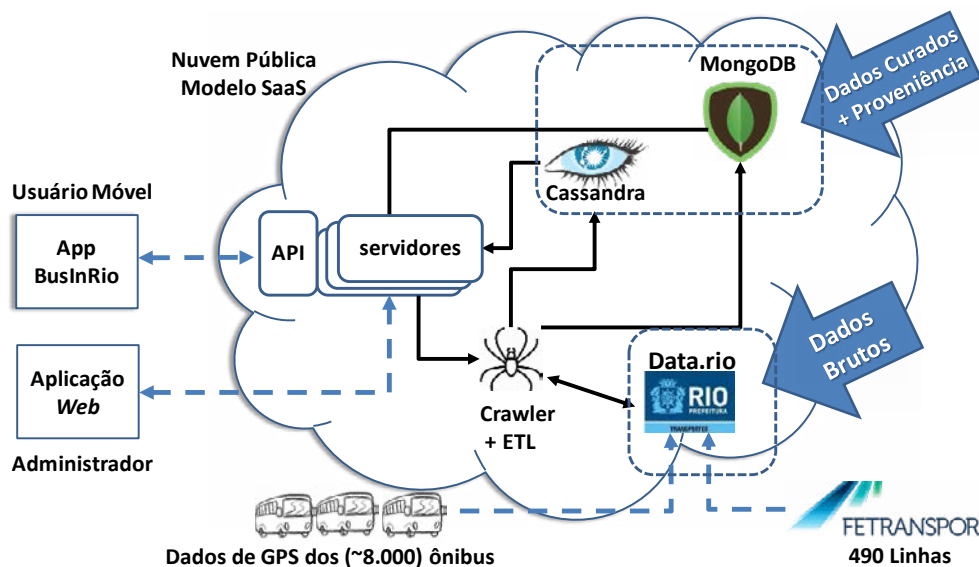


Figura 1. Representação conceitual da arquitetura *BusInRio*.

O servidor de aplicação tem como principais funcionalidades conectar-se ao *data.rio*, utilizar a abordagem ETL para coletar, tratar os dados abertos de transporte público e armazená-los juntamente com sua proveniência retrospectiva em suas

respectivas bases NoSQL (Cassandra e MongoDB) e disponibilizar uma API para que a aplicação móvel consulte e exiba os dados curados e mapas aos usuários.

O cliente é totalmente baseado do framework Ionic⁵, sendo capaz de oferecer uma aplicação híbrida capaz de operar em diversos tipos de *smartphones* ou navegadores web. O cliente possui as seguintes funcionalidades: coletar os dados através da API e exibir para o usuário lista de linhas de ônibus e as representações gráficas através de mapas das APIs do Google Maps que permitem a leitura e visualização de dados sobre o ônibus e sobre trânsito do entorno.

BusInRio é uma arquitetura que explora o princípio da elasticidade das nuvens, à medida que aumentam as requisições dos clientes ou percentagem de uso de CPU dos servidores e de memória variam acima (ou abaixo) de nível pré-definidos, novas instâncias são dinamicamente iniciadas (ou desabilitadas) conforme o caso.

O protótipo do *BusInRio* utiliza o serviço de nuvem pública baseado no modelo SaaS oferecido pelo Koding⁶ que disponibiliza instâncias em servidores da AmazonEC2 para que os clientes possam testar aplicações em nuvem. As instâncias utilizadas nesse trabalho foram do tipo t2.medium (2 CPU e 4GB de RAM virtuais), essa configuração de instâncias fornece um ambiente de uso geral de custo reduzido onde se pode hospedar servidores web, ambientes de desenvolvimento e os repositórios de dados e de proveniência.

5.1. *BusInRio* Crawler

O crawler é um módulo central da *BusInRio*, ele executa no servidor e consiste em acessar recursivamente os servidores do data.rio, recuperar e tratar os dados através das regras da abordagem ETL e retroalimentar as bases de dados NoSQL em tempo real com os dados curados sobre os GPS dos ônibus e suas linhas. Para isso, ele coleta e trata os dados em formato JSON e armazena-os sob a forma de documentos no MongoDB e os dados do percurso em CSV e armazena-os no Cassandra.

5.2. Bases NoSQL

Nessa arquitetura utilizamos dois bancos do tipo NoSQL. O Cassandra é um banco de dados não relacional distribuído altamente escalável que se apoia na arquitetura do DynamoDB, da Amazon Web Services e modelo de dados baseado no BigTable⁷. Este sistema apresenta grande velocidade pois guarda os dados curados na memória RAM. Além disso, é orientado a colunas e por esses motivos foi escolhido para representar os dados dos percursos dos ônibus. Foi criada uma *keyspace* chamada transporte e as tabelas "percurso" e "linha" para representar os dados. Em "linha" armazena-se os dados sobre a linha do ônibus (identificação e descrição) e no "percurso" os pontos pelo qual o ônibus passa.

O MongoDB também é uma aplicação não relacional de alta performance, código aberto, sem esquemas e orientado a documentos. Além de ser orientado a documentos,

⁵ <http://www.ionicframework.com>

⁶ <https://koding.com/>

⁷ <https://cloud.google.com/bigtable/>

ele é formado por um conjunto de documentos JSON. Muitas aplicações podem, dessa forma, modelar informações de modo muito mais natural, pois os dados podem ser aninhados em hierarquias complexas e continuar a ser indexáveis e mais fáceis de buscar. Graças ao formato JSON dos dados (visto que a maioria dos dados seriam utilizados em JSON pelo *app*) este servidor foi o escolhido para armazenar e manipular os dados de GPS dos ônibus. Foi criado um banco "transporte" no MongoDB e uma coleção "onibus" e, pode-se facilmente inserir um documento que representa as informações coletadas pelo GPS do ônibus.

5.3. API BusInRio

A API BusInRio é composta por funções acessíveis somente por programação, e que permitem utilizar características da arquitetura, ela permite que a aplicação cliente que executa no *smartphone* interprete e converta os dados dos transportes para o formato JSON que é mais facilmente interpretado pelas aplicações móveis e navegadores web.

Através da API BusInRio o aplicativo envia requisições HTML parametrizadas para o servidor de aplicação, por exemplo, número da linha escolhida pelo usuário e data; então o servidor responde com dados sobre os ônibus (em formato JSON) que será interpretado pelo aplicativo e disposto na interface para visualização do usuário.

5.4. Prova de Conceito

Nesta sub-seção, serão apresentados detalhes sobre a avaliação da solução proposta, incluindo tecnologias utilizadas para o desenvolvimento do protótipo do *app*.

O *app BusInRio* está baseado no framework Ionic. Este é uma SDK que permite o desenvolvimento de aplicativos híbridos capazes de executar em *smartphones* com diversos tipos de sistema operacional e navegadores web apenas utilizando os recursos do HTML5, Javascript e CSS. O framework é baseado no AngularJS e seu núcleo compõe funcionalidades do Apache Cordova. O aplicativo móvel *BusInRio* foi desenvolvido com auxílio deste framework utilizando os dados NoSQL disponibilizados pela API.

O *app BusInRio* permite ao usuário escolher, a partir de um *smartphone*, a linha de ônibus que desejar e consultar a localização dos ônibus daquela linha em tempo real. A Figura 2 ilustra algumas telas da aplicação. Por exemplo, quando o *app* passa como parâmetro o "número da linha" de ônibus, o servidor atende a requisição produzindo um documento JSON contendo as últimas informações sobre cada carro daquela linha, tais como "*ordem, latitude, longitude, data hora e velocidade*". Essas informações são então interpretadas pelo *app* que produz uma visualização amigável para o usuário (figura 2B).

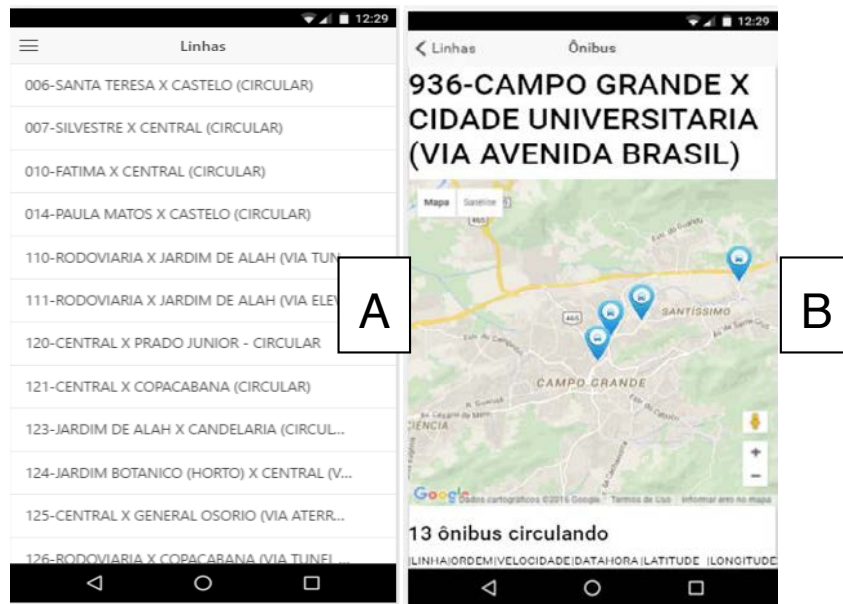


Figura 2. O app *BusInRio* smartphone - consulta às linhas disponíveis (A) e visualizaçãoda localização de cada ônibus da linha e total de ônibus da linha em operação(B).

6. Experimentos e Discussão

Através dos dados abertos disponibilizados pelo data.rio sobre os percursos dos ônibus e GPS em tempo real e com o auxílio de bancos de dados NoSQL MongoDB e Cassandra foi possível estabelecer a ligação entre esses diferentes datasets e fornecer informações úteis ao usuário e de maneira simples. Nossos experimentos utilizaram dados relativos ao mês de julho de 2015.

Através do app *BusInRio* pode-se exibir a localização exata de cada ônibus de uma linha no aplicativo. Ao analisar os dados de localização e velocidade de cada ônibus foi possível gerar um relatório do tipo gráfico exibindo um mapa de calor que representa o trânsito nas principais ruas da malha viária da cidade do Rio de Janeiro através do Google Maps.

Neste caso o usuário, através do navegador do seu *smartphone*, pode identificar claramente os pontos onde o tráfego é mais lento como mostra a Figura 3 (esquerda). As regiões onde o trânsito está mais intenso está representado na cor vermelha e onde flui em melhores condições há um gradiente (do vermelho ao verde). Caso a velocidade média de um conjunto de ônibus que percorram um determinado trajeto comum sejam inferiores a 10Km/h o trânsito na região é considerado intenso (vermelho), e superior 20km/h (verde) é considerado aceitável. Utilizou como valor de referência a velocidade média dos ônibus da cidade de Curitiba-PR (~17Km/h) por ser considerada uma cidade com transporte urbano modelo para o Brasil.

O *appBusInRio* utiliza os dados abertos governamentais curados, ligados e armazenados no MongoDB para estimar a velocidade média dos ônibus localizados nas principais regiões da cidade e para identificar as condições do trânsito, quanto menor a velocidade média do ônibus, maiores as chances de estar retido em um congestionamento.

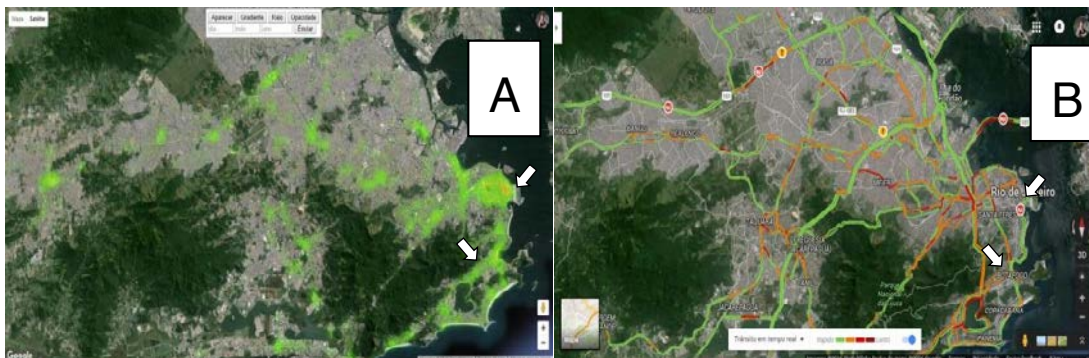


Figura 3. Mapa de calor com as condições do trânsito calculadas pelo *BusInRio* (A) em comparação com as condições do trânsito avaliadas pelo Google Maps (B). As setas indicam pontos geográficos equivalentes em mapas distintos (Centro do RJ e Zona Sul).

Verificou-se empiricamente se os cálculos do *BusInRio* (sobre as condições do trânsito e linhas do município do Rio de Janeiro) se aproximavam com os de outros produtos, efetuamos dezenas de consultas em dias/horários e comparamos de consultas contra o Google Maps sobre as condições do trânsito calculadas. Verificamos que houve grande correlação. Por exemplo, na Figura 3(A e B) é possível verificar trânsito lento no centro da Cidade do Rio no dia 06/04/2016 por volta das 20:35h, neste dia havia uma manifestação convocada pelos Servidores Públicos do Estado no centro do RJ. Através dos mesmos mapas de densidade de trânsito se verificou que ele estava pouco congestionado nas suas principais vias de escoamento à cidade (av. Brasil) e em diversas partes da Zona Sul da cidade.

7. Conclusões, Limitações e Trabalhos Futuros

Cidades inteligentes são resultantes de conjuntos de ações que objetivam tornar a gestão dos espaços públicos e recursos mais eficientes por meio das TICs. Diversas soluções têm sido propostas para melhorar o problema da mobilidade urbana. Contudo, essas soluções carecem de estudos mais aprofundados.

Este trabalho apresenta a arquitetura *BusInRio*, totalmente baseada em softwares e dados abertos governamentais e em tecnologias NoSQL. As regras da abordagem ETL mostraram-se adequadas, porém limitadas para o tratamento de dados abertos em ambiente de nuvem. Com relação aos repositórios de dados, verificou-se que o MongoDB se mostrou muito superior no que diz respeito a velocidade de resposta das consultas do *app*, especialmente considerando-se que tinha um volume muito maior de registros das viagens dos ônibus. Por questões de limitações de escopo, neste trabalho não apresentamos maiores detalhes sobre a geração de metadados de proveniência dos processos de transformação e ligação dos dados abertos, nem sobre desempenho e volumes de dados manipulados pela arquitetura.

Como trabalhos futuros é sugerido que se desenvolvam novas regras para detecção de *outliers* e dados ausentes. Além disso se buscará que aplicação integre dados de metrô, barcas e trens para oferecer ao usuário uma estimativa aproximada dos melhores percursos e das rotas comparando com outros modais de transportes públicos na cidade do Rio de Janeiro.

Referências Bibliográficas

- Barbosa. L.; Kormaksson, M.; Vieira, M. R.; Tavares, R. L.; Zadrozny. B.; (2014), “Vistradas: Visual Analytics for Urban Trajectory Data”. 15th Brazilian Symposium on Geoinformatics.
- Bessa, A.; Silva F. M.; Nogueira R. F.; Bertini. E, Freire, J. (2015), “RioBusData: Visual Data Analysis of Outlier Buses in Rio de Janeiro”. Symposium on Visualization in Data Science.
- Chee, B. J. S; Franklin Jr. Curtis (2013) “Computação em Nuvem – Tecnologias e estratégias” Ed. M Books.
- Cruz, S.M.S. et al. (2009) “Towards a Taxonomy of Provenance in Scientific Workflow Management Systems”, In: Congress on Services, IEEE, p. 259–266.
- Dohler, M.; Vilajosana, I.; Vilajosana, X.; Llosa, J. (2011) “Smart Cities: An action plan,” in Proc. Barcelona Smart Cities Congress, Barcelona, Spain, pp. 1–6.
- Gil, Y. et al. (2010) "Provenance XG Final Report", W3C, www.w3.org/2005/Incubator/prov/XGR-prov
- Hernández-Muñoz, J. M.; Vercher, J. B.; Muñoz, L.; Galache, J. A.; Presser, M.; Hernández Gómez, L. A.; Pettersson J., “Smart Cities at the forefront of the future Internet,” The Future Internet, LNCS., v. 6656, pp. 447–462
- Leite, D. F. B; Rocha, J. H.; Baptista, C. S.; (2013) “Busão: um Sistema de Informações Móvel para Auxílio à Mobilidade Urbana Através do Uso de Transporte Coletivo”. In: SBSI, 2013, João Pessoa.
- Mendonça, R. R.; Cruz, S. M. S, Campos, M. L. M (2014) “Gerência de Proveniência Multigranular em Linked Data com a Abordagem ETL4LinkedProv”. In SBBDD, 2014 Curitiba.
- Open Government Data (2016). <https://opengovdata.org/>. Acesso em março de 2016.
- Open Definition (2016) <http://opendefinition.org/>. Acesso em março de 2016.
- Portal Brasileiro De Dados Abertos. (2013) “Maturidade em dados abertos, entenda as 5 Estrelas”, <http://dados.gov.br/noticia/maturidade-em-dados-abertos-entenda-as-5-estrelas/>. Acesso em: março de 2016.
- Schaffers, H. et al. (2011) “Smart cities and the future internet: Towards cooperation frameworks for open innovation”. The Future Internet, LNCS., v. 6656, pp. 431–446.
- Tito, A. O., et al (2012) “UbiBus: Um Sistema de Informações Inteligentes para Transporte Público”. In: Workshop Tecnologias da Informação e Comunicação nos Grandes Eventos Esportivos (WTICEE), Aracaju-SE.

Exploratory Analysis of Public Transportation Data in Curitiba

Nádia P. Kozievitch¹, Tatiana M. C. Gadda¹, Keiko V. O. Fonseca¹,
Marcelo O. Rosa¹, Luiz C. Gomes-Jr¹, Monika Akbar²

¹ Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, PR – Brazil

²University of Texas at El Paso, El Paso, Texas, USA

{nadiap,tatianagadda,keiko,mrosa}@utfpr.edu.br,
gomesjr@dainf.ct.utfpr.edu.br, makbar@utep.edu

Abstract. *Smart transportation systems have been providing more data over time (such as bus routes, users, smartphones, etc.). Such data provides a number of opportunities to identify various facets of user behavior and traffic trends. In this paper we address some of the urban mobility challenges (already discussed by the Brazilian Computer Society), from a number of different perspectives, including (i) pattern discovery, (ii) statistical analysis, (iii) data integration, and (iv) open and connected data. In particular, we present an exploratory data analysis with GIS for public transportation toward a case study in Curitiba, Brazil.*

1. Introduction

The concept of smart cities is usually linked to efficiency in the use of natural resources [Souza et al. 2015]. Public transportation is a critical factor for the functioning of a city. It provides mobility to the masses and helps to mitigate traffic and pollution. With the advent of smart technologies, urban transportation systems are able to capture a lot of useful data. Such data can be used to shed light on a number of factors, including user trends and traffic patterns. These items are essential for urban planning, optimizing the transportation system (e.g., fuel, time), reducing environmental impacts of mass transport (e.g., noise pollution, air pollution).

Public transportation is one of the most critical areas of smart cities. In Brazil, the vehicle fleet in major cities grew more than the road structure ¹. Mobility challenges have already gained attention of computer science discipline in Brazil ². In particular, these challenges can be grouped in the following areas: (i) discovery of patterns, (ii) data statistics, (iii) data integration, (iv) location and tracking, (v) open and connected data, (vi) contextual information, (vii) security and privacy, (viii) energy and management, (ix) use of cloud resources, (x) trajectories with semantic information, among others.

In this paper we address the mobility challenge (by the public transportation view), from the top 5 perspectives previously mentioned, for a case study in Curitiba. Curitiba,

¹ <http://www1.folha.uol.com.br/cotidiano/2014/08/1503030-frota-de-veiculos-cresce-mais-rapido-que-a-estrutura-viaria-no-pais.shtml> – Last visited on 05/03/2015.

² <http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/802-grandesdesafiosdacomputaonobrasil> – Last visited on 05/03/2015.

a city located in the south of Brazil, with 1.8 million people in a total area of 430,9 km², according to the Brazilian Institute of Geography and Statistics (IBGE)³. This area encompasses 75 districts, and is surrounded by other 29 cities, known as the metropolitan region of Curitiba (with the Portuguese acronym *RMC*).

In particular, Curitiba was already considered one of the world's smartest cities⁴ and also belongs to a group of cities, the *C40 cities*⁵, which set ambitious targets to improve urban life quality and protect their environment. Curitiba has developed and implemented mass transport corridors, densification of land-use along these corridors, and mobility solutions using Bus Rapid Transit (BRT) systems in the 1970s, where one main feature of the success of the system is its complex network of feeder lines [Duarte et al. 2016].

Routes run in exclusive lanes using high capacity bi-articulated buses, which transport up to 250 passengers per bus. Most of the buses in this system are currently powered by diesel engines. However, due to environmental concerns and sustainable principles, they should be gradually replaced by buses that use other sources of energy, like BEV buses [Sebastiani et al. 2016].

The city has also been participating in the open data initiative, through several government stakeholders, such as Instituto de Planejamento de Curitiba (IPPUC)⁶ and the Municipality of Curitiba⁷. In the academy, although it has been studied under distinct perspectives [Zannin et al. 2002, Calixto et al. 2003, Zannin et al. 2003, Sebastiani et al. 2016], we believe that new clues about mobility can be derived from the analysis of new domains.

Briefly, the efforts of integrating different datasets towards exploring mobility can be addressed from different directions, including (i) from the GIS perspective, the challenge is how to integrate different sources, with data having single semantic along with efficiency (and to enable any future integration with other systems); (ii) from the integration perspective, the challenge is not to have a bottleneck; (iii) from the data perspective, the challenge is to detect limitations and standardize a unified view; and (v) from the pedagogical perspective, the challenge is to explore all this structure and have a interdisciplinary learning environment.

The objective here is not to understand the specific details (such as hardware [Suresh et al. 2009], networks [Park and Kim 2011] or integrate environment-related parameters, but to understand the sources and obstacles which create impact from a general overview, along with a general understanding of the behavior of the system. The main goal is to explore the mobility challenge with an exploratory data analysis with GIS for public transportation toward a case study in Curitiba, Brazil.

The rest of this document is organized as follows: Section 2 presents the related work. Data is presented within Section 3. Behavioral trends are presented in Section 4.

³ <http://www.ibge.gov.br> – Last visited on 14/05/2015.

⁴ <http://www.forbes.com/2009/12/03/infrastructure-economy-urban-opinions-columnists-smart-cities-09-joel-kotkin.html> – Last visited on 15/04/2016.

⁵ <http://www.c40.org> – Last visited on 30/05/2015.

⁶ <http://ippuc.org.br/Lastvisitedon15/05/2015> – Last visited on 05/03/2015.

⁷ <http://www.curitiba.pr.gov.br/DADOSABERTOS/> – Last visited on 15/05/2015.

Conclusions and future work are presented in Section 5.

2. Related Work

2.1. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a philosophical approach to data analysis [NIST/SEMATECH 2012]. The posed question is: what data can tell us about certain relationships, properties or structures. There are no imposed techniques to apply to the data set, but graphical visualization plays an important role in this approach [Hartwig and Dearing 1979]. The non-inferential approach in data analysis encourages the openness perspective required to integrate new domains. The availability of processing power and data storage provide new tools for handling massive amounts of data processing, allowing flexible search of evidences in the available data through designed experiments [Martinez et al. 2010].

2.2. GIS and GIS-T

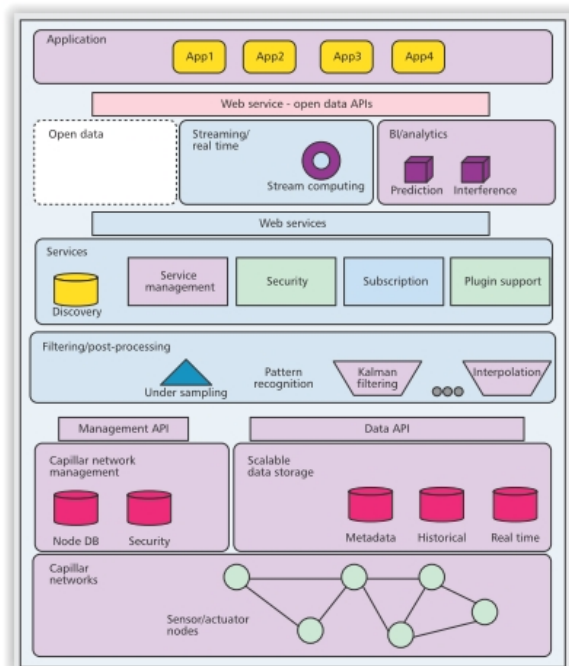


Figure 1. Generic platform showing the main components to support heterogeneous smart city applications [Vilajosana et al. 2013].

Geographic Information System (GIS) stores data from heterogeneous sources in several formats in the form of geodatabases representing spatial features, storing latitude and longitudinal positions. Geodatabases are increasing nowadays generating huge volume of data from satellite images, day traffic, among other sources. Many of these application systems are based on global positioning systems (GPS) [Stenneth et al. 2011], using specific data (such as the real time bus locations, spatial rail and spatial bus stop information) and specific techniques (such as spatial data mining [Mennis and Guo 2009]). An enterprise geodatabase can help solve common transportation challenges [Butler 2008]: the many origins, destinations, paths, and conveyances that may be present, along with their locations.

In particular, geographic information systems for transportation (GIS-T) are interconnected hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing, and communicating particular types of information about the earth [Butler 2008].

Within transportation systems, the GIS database for transportation is organized on the basis of the road network map. Each link can have as attributes topographic (e.g. nodes UTM coordinates, total length), toponomastic (street names), physical (traffic directions, number of lanes), transport (road typology by means of the speed-flow curves) and transit (description of public transport lines and corresponding frequencies) information [Arampatzis et al. 2004]. Here, generally origin destination matrices express the traffic demand.

Several models can be used in order to model the traffic: a deterministic model to simulate road traffic, the representation of the multi-modal transport problem (through simultaneous elastic demand-traffic assignment concepts model), the behavior of public transport users (modeled by using the concept of transit hyper-paths on a generalized network), among others.

Examples of applications include SIGGESC [de Oliveira et al. 2012], which associates the spatial and alphanumeric data of the bus lines with the occupation of the territory, mobility, census information, etc., available to the public through a web portal. General, data for these applications can explore open data, through different categories, such as public transit tracking (such as Chicago Transit Authority Bus Tracker ⁸), stakeholders information (such as Chicago Transit Authority ⁹), mobile phones [Stenneth et al. 2011], among others. Within the GIS-T challenges, we can mention standards, the representation within databases, unambiguous communication of locations, economic models, response to new technologies, and applications to other fields [Goodchild 2000].

2.3. Smart Cities

Smart Cities was already defined as the integration of technology into a strategic approach to sustainability, citizen well-being, and economic development ¹⁰. Smart city platforms (such as Figure 1) are designed to provide and agglutinate heterogeneous services to support a variety of application domains. In particular, the data integration (GIS, historical, real time) includes sensor nodes, actuators, software components, and other devices that gather data from the city infrastructures and citizens [Vilajosana et al. 2013]. The delivery of the information can take advantage of open data models, without charging any fee to developers and contributors. In Brazil, several communities and projects have been created toward open data ¹¹, smart cities ¹², and smarter use of resources ¹³.

From IT perspective, there are still challenges regarding the architecture of this infrastructure [Chourabi et al. 2012], including (i) lack of integration across government

⁸ <http://www.ctabustracker.com/> – Last visited on 04/04/2016.

⁹ <http://www.transitchicago.com/about/facts.aspx> – Last visited on 04/04/16.

¹⁰ <http://www.navigantresearch.com/media/smart-cities-5-security-areas-cios-should-watch> – Last visited on 07/04/2015.

¹¹ <http://br.okfn.org/category/dados-abertos/> – Last visited on 15/04/2016.

¹² smartcitiesforumbrasil.com.br/ – Last visited on 15/04/2016.

¹³ <http://redesinteligentesbrasil.org.br/o-projeto.html> – Last visited on 15/04/2015.

systems; (ii) restrictions regarding the integration capabilities of existing internal systems; (iii) lack of knowledge regarding interoperability; and (iv) availability and compatibility of software, systems, and applications.

The objective of this paper is toward this direction, exploring the mobility challenge with an exploratory data analysis with GIS for public transportation toward a case study in Curitiba, Brazil.

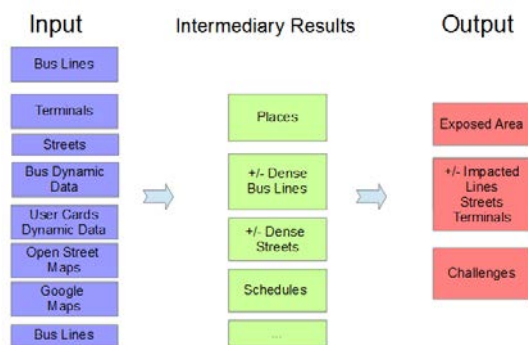


Figure 2. Overview of the dataset.

3. Data

The datasets used in this paper come from IPPUC, the Municipality of Curitiba, along with data from Open Street Map¹⁴. Figure 2 shows the input sources, divided between stationary and dynamic data. General statistics are presented in Figure 1¹⁵. Details are listed below.

Description	Quantity
Official Number of Bus Vehicles	1500
Official Number of Daily Passengers	1.620.000
Official Number of Bus Routes	250
Number of Bus Terminals	23
Number of Tube Bus Stations	342
Average Bus Routes KMs by day	480.000
Average Age of the Fleet	7
Number of Bus Vehicles Detected with data sample	1368
Number of Bus Stops Detected with data sample	9940
Average number of Bus User Cards using GIS Online Tracking	349.729
Average number of Bus Routes using GIS Online Tracking	269
Average number of Bus Vehicles using GIS Online Tracking	1522
Number of Roads within Curitiba	9.135

Table 1. General Statistics of the Data and Bus Network.

Stationary Data. Stationary (or static) data includes georeferenced data from bus lines, bus stops, terminals, and streets that do not change often.

Bus Routes. The city has 482 bus routes distributed within 11 categories (Metropolitano, Linha Direta, Expresso Ligeirão, Interbairros, Expresso, Troncal, Convencional, Turismo, Circular Centro, Interhospitais, Alimentador), as shown by green lines in Figure 3. The categories Alimentador and Convencional have the majority of lines, with 265 and 65 units respectively.

¹⁴ <http://www.openstreetmap.org> – Last visited on 14/05/2015.

¹⁵ <https://www.urbs.curitiba.pr.gov.br/> –Last visited on 04/04/2016.

Bus Stops. The city has 9940 bus stops detected within the stationary data. Curitiba has an additional bus stop category named Tube Bus Stations (in portuguese, *Estação Tubo*). The tube stations (officially 342) are bus stops which look like tubes, for specific bus routes, such as Expresso and Linha Direta. The districts named CIC and Centro have the majority of them, with a total of 1628 and 667 units each one.

Bus Terminals. The city has 23 terminals (buses) and one terminal which also use trains, as shown by red points in Figure 3. The oldest one is named Guadalupe, from January 1st, 1956. The districts CIC and Boqueirão concentrate their majority, with 3 and 2 units each one.

Streets. The city has 9.135 streets, divided among 17 categories (Anel Central, Central, Coletora 1, Coletora 2, Coletora 3, Externa, Linhão, Normal, Outras Vias, Pedestre, Prioritária 1, Prioritária 2, Rodovia Estadual Duplicada, Rodovia Estadual Simples, Rodovia Federal Duplicada, Rodovia Federal Simples, and Setorial). The districts CIC and Sítio Cercado have their majority, with 1217 e 524, each one.

Figure 5 presents the density of how bus lines are present in the city. Gray lines present streets with an average of 1-17 lines, the green lines represent an average of 17-37 lines, the pink lines represent an average of 37-55 lines, and the red lines represent streets with 55-73 lines. The zoom at the same figure presents the density of bus lines within the downtown district, having each edge the respective number of lines. For the same region with zoom within Figure 5, we calculated the *Eigenvector Centrality* [da F. Costa et al. 2007] of the the bus lines, as shown in Figure 6. The Eigenvector Centrality (also the basis for the PageRank algorithm) captures the influence of nodes in a network, taking into account that nodes with influential neighbors should have higher centrality values. In a bus line network, high centrality values indicate points of (potentially) intense traffic. Identifying the points with high centrality score in the Figure reaffirms our observation that traffic within the Guadalupe Bus Terminal (the green triangle at the Figure) concentrates intense traffic.

Dynamic Data Dynamic data includes data from bus lines, and user cards, transmitted at an average frequency of 5 minutes. In particular, the data selected was between October 19th, 2015 and October 26th, 2015.

User Cards. This set of data represents the commuting information comprising trip data per user card: the vehicle id, line code, the user card number, and the date of the trip. Within the date range selected, it was detected 349.729 different user cards, 269 different bus lines were present, divided among 1522 vehicles. Part of the vehicles and lines represented here have geolocated data, presented within the following set.

Vehicles and Respective Bus lines. This set of data represents the daily itinerary, having as data the vehicle id, line code, the datetime, and the latitude and longitude. Within the data range selected, it was detected 1623 distinct vehicles and 703 different bus lines.

The complete dataset were inserted in a PostGIS ¹⁶ database. Different sources were created as different tables in the database. Later, specific tablespaces and indexes were created in order to optimize the access. Since semantic errors were present (different sources presented different street names for the same location, for exam-

¹⁶ <http://www.postgis.net> – Last visited on 15/05/2014.

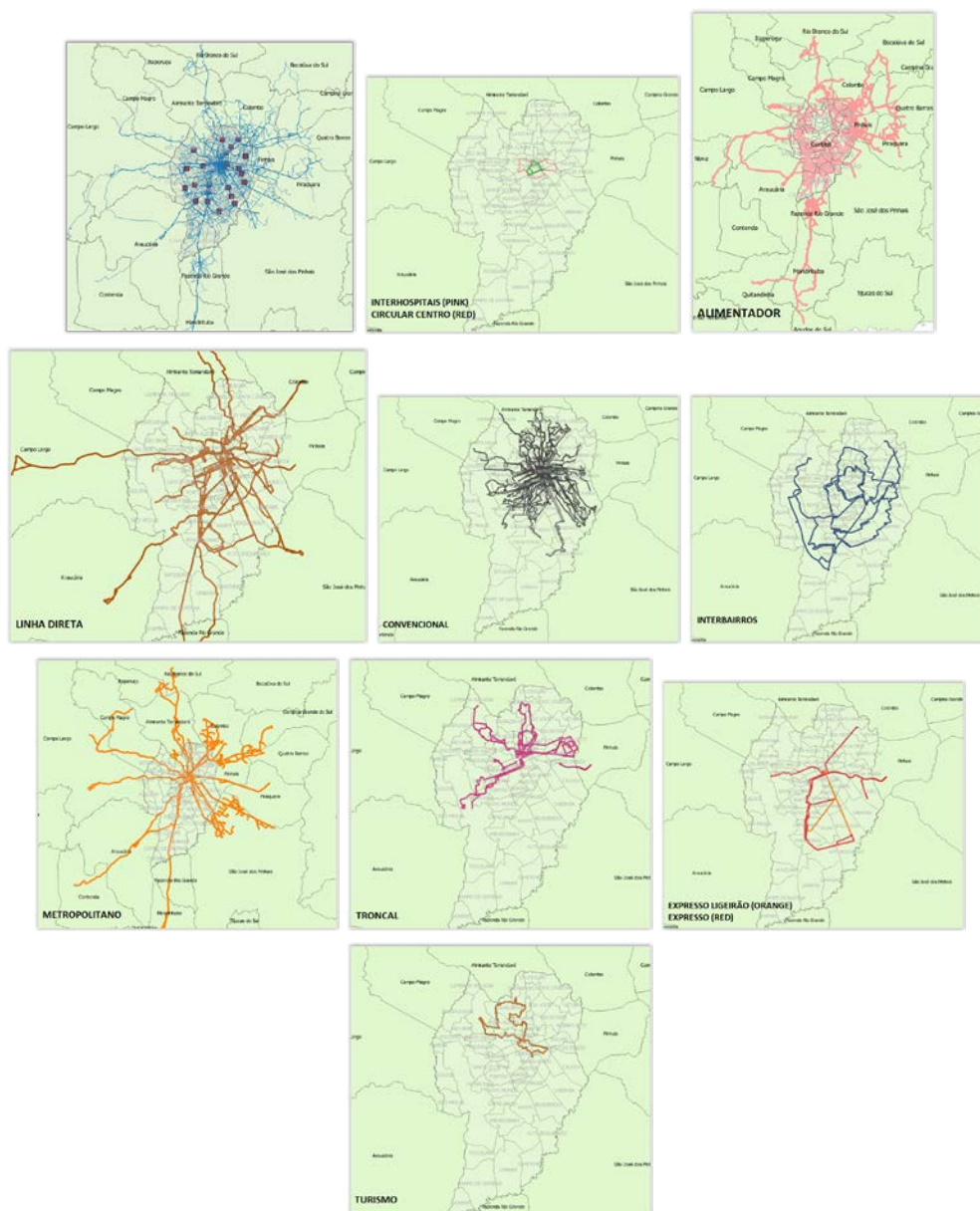


Figure 3. Representation of bus lines (blue) and bus terminals (red) on upper left, followed by the eleven bus categories.

ple [Barczynszyn 2015]), geolocation and specified time range were used to correlate the data.

4. Behavioral trends

The analysis of dynamic data from user cards indicated some trends: **(i)**: Over the week of data analyzed, there is an average of 350 bus lines, an average of 1500 vehicles, and an average of 250.000 user cards (Figure 7) per day. Sundays presented the lowest number of user cards. **(ii)**: The analysis of user cards by hour indicated that the peak hours for traffic are 7 AM, 5 PM, and 6 PM (Figure 8). Friday (October 23rd, 2015) presented the highest amount of different user cards (264.224), and Sunday (October 25th, 2015) presented the lowest amount of user cards (53.510). **(iii)**: The top 5 bus lines which had the majority of the user cards were during the time range were "Oper s/Linha", "Op. Contingencia",

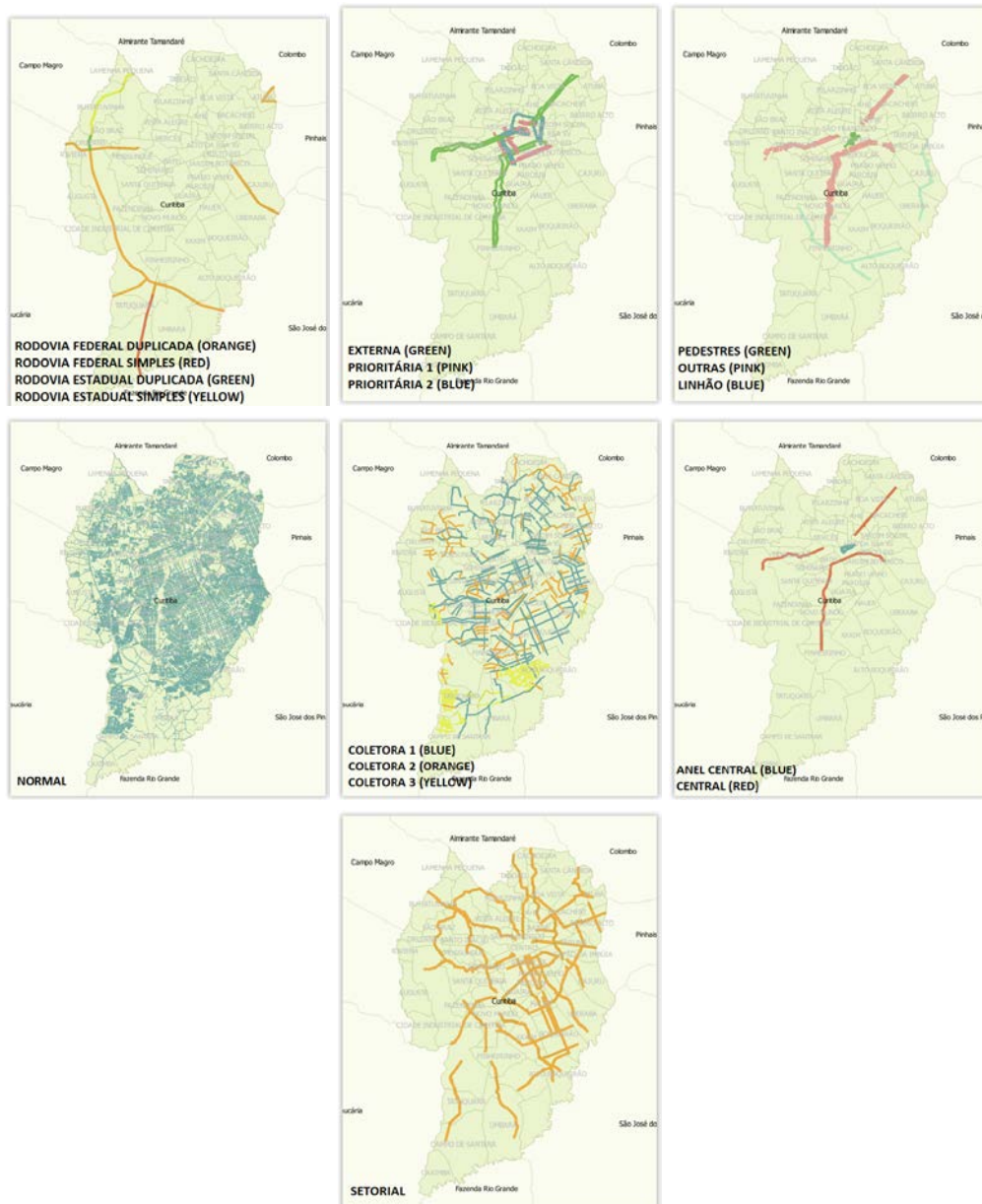


Figure 4. Representation of streets, by each category.

”Sistema Araucária”, ”Interbairros IV” and ”Interb. II Anti H”. The first one indicated bus operating without a specific line. The second one indicated extra buses to the already existing lines. The third one connects the metropolitan region to Curitiba. The last two lines operate within the Interbairros category. These lines connect several districts.

The analysis of the dynamic data from bus lines also presented particular characteristics:

- The analysis of Interbairros IV bus line data indicated that the vehicles were also used to other lines at the same category (as indicated in Figure 9: instead of just using the blue line indicated in the figure, vehicles were also using other lines).
- Routes which are not included within the respective line, were also present (Region A in Figure 9). The analysis of which streets are used by vehicles to connect different lines and how they are used during the peak hours could be used for the

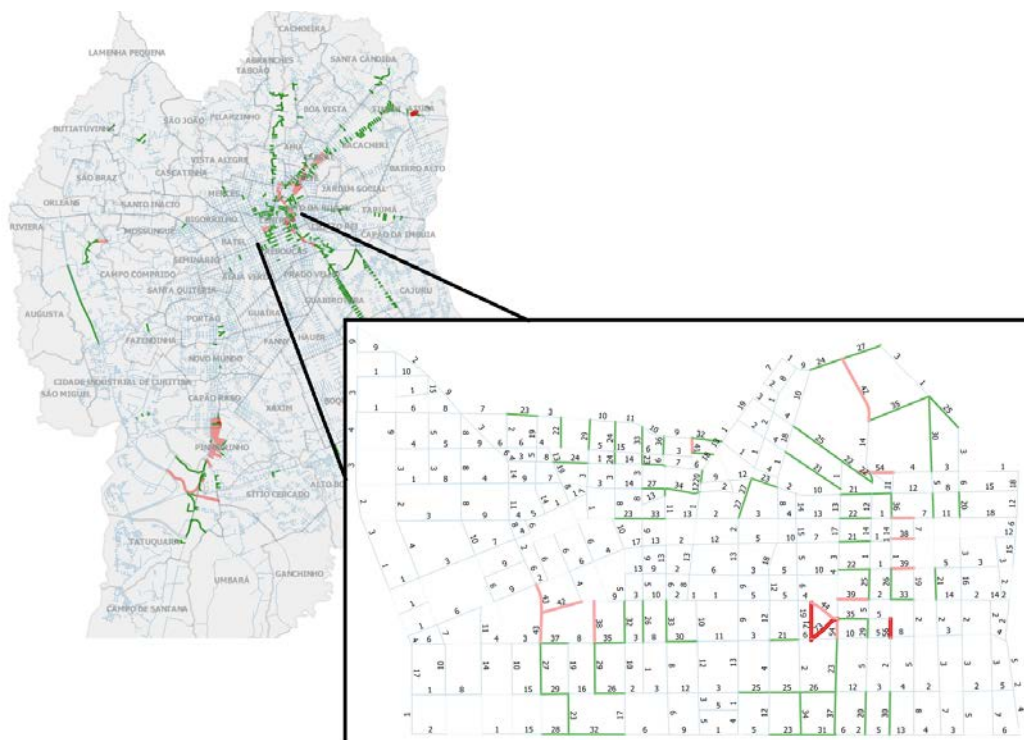


Figure 5. Density of Bus Lines within streets in Curitiba, with zoom in the central district. Pink and red streets present higher bus densities.

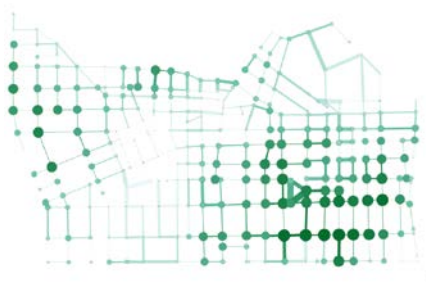


Figure 6. Eigenvector Centrality of the bus lines within the Central District. Larger and darker circles indicate higher centrality scores.

planning of the routes to optimize various parameters of the transportation system of the city.

- In parallel, routes which are regularly accessed, but are not present by main route were also detected (Region B in Figure 9). The behavior was present within both the peak and off-peak hours.
- The busiest route (Interbairros IV - blue line at Figure 9 - left) surrounds the central district, connecting several different neighborhoods and five bus terminals (Santa Felicidade, Campo Comprido, Fazendinha, CIC e Pinheirinho). The data indicates peaks for six A.M. (average of 902 users), seven A.M. (average of 938 users) and six P.M. (average of 604 users) within the bus line.

In summary, some of the detected patterns and associated challenges are as follows: (i) **Data Integration**: the different coordinate systems, formats, and information along the several government institutions keep the data integration as a challenge, even considering only Brazilian cities (such as São Paulo - through the Municipality of São

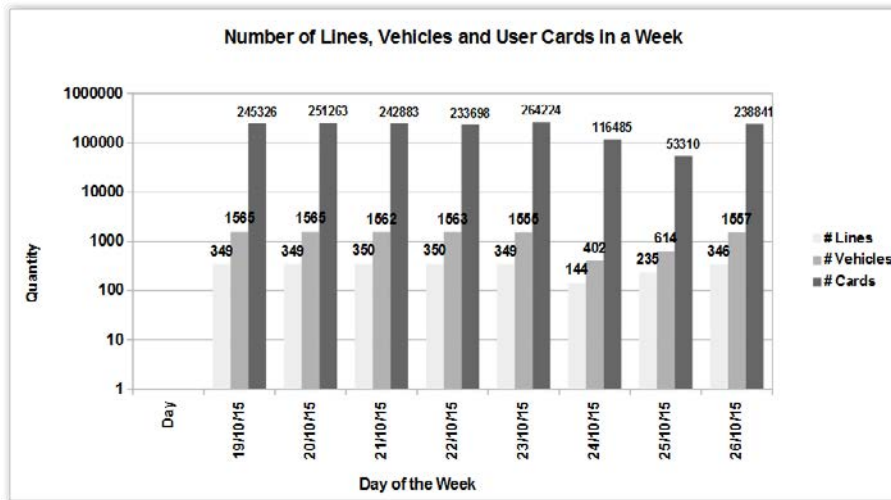


Figure 7. Number of Lines, Vehicles, and User Cards in a Week.

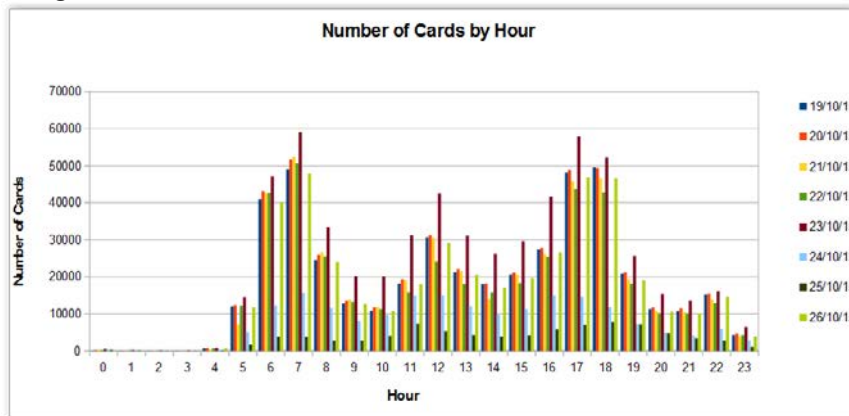


Figure 8. Number of user cards by hour.



Figure 9. Representation of Interbairros IV line, during all day (left), seven A.M. (center) and five P.M. (left).

Paulo¹⁷), Recife (through the Municipality of Recife¹⁸) and Rio de Janeiro (through Sistema Municipal de Informações Urbanas - RJ¹⁹)). In general, formats such as CSV, Excel,

¹⁷<http://dados.prefeitura.sp.gov.br/> – Last visited on 02/06/2015.

¹⁸<http://dados.recife.pe.gov.br/dataset> – Last visited on 15/05/2015.

¹⁹<http://portalgeo.pcrj.opendata.arcgis.com/> – Last visited on 15/05/2015.

Json and Shapefile are preferred (some use APIs and KML). Not all of them provide meta-data or the data visualization. Some services still have issues (such as the download of spatial files from the Municipality of Rio de Janeiro). In particular, the data within the use case had text, excel and shapefile formats; (ii) **Data Statistics**: initial data statistics indicated impacts by peak hours, days of the week, contingency and metropolitan vehicles, among others. Visualization indicated vehicles use among several bus lines, with intermediary routes to connect them; (iii) **Open and Connected Data**: the use case presented used Open Data resulted from decree 1135/2012, from law 12.527²⁰. Besides this policy, the municipality is also using the data for competitions named Hackatons²¹, as well as integration of partners for future projects (regional and international)^{22 23}.

5. Conclusion

While the number of vehicles operating in a city increases rapidly, the change in infrastructure is a rather slow process. In order to accommodate the users and their transportation needs, a city must carefully analyze data sources to determine the user needs and possible changes in transportation to support those needs. This paper presented an initial investigation in order to identify scenarios and implications from the public transportation data from the city of Curitiba, integrating several sources, including stationary and dynamic open data. Detected characteristics can be used for further analysis in order to optimize the transportation system. This work will be extended to provide recommendation and optimization for alternative and existing bus routes that minimizes various parameters including travel time and number of transfers. Lastly, we will investigate various approaches for integrating other data sources.

Acknowledgment

We would like to thank the Municipality of Curitiba, IPPUC, CAPES, CNPq. This work is partially funded by the National Science Foundation (NSF) grant HRD-1242122 and EU-BR EUBra-BigSea project (MCTI/RNP 3rd Coordinated Call).

References

- Arampatzis, G., Kiranoudis, C., Scaloubacas, P., and Assimacopoulos, D. (2004). A gis-based decision support system for planning urban transportation policies. *European Journal of Operational Research*, 152(2):465 – 475. New Technologies in Transportation Systems.
- Barczynszyn, G. L. (2015). Integração de dados geográficos para planejamento urbano da cidade de Curitiba. Universidade Tecnológica Federal do Paraná.
- Butler, J. A. (2008). *Designing Geodatabases for Transportation*. Esri Press.
- Calixto, A., Diniz, F. B., and Zannin, P. (2003). The statistical modeling of road traffic noise in an urban setting. *Cities*, 20:1–74.

²⁰ <http://multimidia.curitiba.pr.gov.br/2014/00147194.pdf> – Last visited on 07/04/2015.

²¹ <http://hackathon.curitiba.pr.gov.br/> – Last visited on 15/05/2015.

²² <http://www.curitiba.pr.gov.br/noticias/curitiba-e-holanda-vaotrabalhar-juntas-em-projetos-de-ciclomobilitate-para-a-cidade/37601> – Last visited on 14/04/2016.

²³ <http://multimidia.curitiba.pr.gov.br/2015/00166636.pdf> – Last visited on 14/04/2016.

- Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., Pardo, T. A., and Scholl, H. J. (2012). Understanding smart cities: An integrative framework. In *System Science (HICSS), 2012 45th Hawaii International Conference on*, pages 2289–2297.
- da F. Costa, L., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.
- de Oliveira, T. H. M., Painho, M., and Henriques, R. (2012). A spatial decision support system for the portuguese public transportation sector. In *IWGS '12*, pages 84–90, New York, NY, USA. ACM.
- Duarte, F., Gadda, T., Luna, C. A. M., and Souza, F. T. (2016). What to expect from the future leaders of bogotá and curitiba in terms of public transport: Opinions and practices among university students. *Transportation Research Part F: Traffic Psychology and Behaviour*, 38:7 – 21.
- Goodchild, M. F. (2000). Gis and transportation: Status and challenges. *Geoinformatica*, 4(2):127–139.
- Hartwig, F. and Dearing, B. (1979). *Exploratory Data Analysis*. 07. SAGE Publications.
- Martinez, W., Martinez, A., and Solka, J. (2010). *Exploratory Data Analysis with MATLAB, Second Edition*. Chapman & Hall/CRC Computer Science & Data Analysis. Taylor & Francis.
- Mennis, J. and Guo, D. (2009). Spatial data mining and geographic knowledge discoveryan introduction. *Computers, Environment and Urban Systems*, 33(6):403 – 408. Spatial Data Mining- Methods and Applications.
- NIST/SEMATECH (2012). *E-Handbook of Statistical Methods*, available at <http://www.itl.nist.gov/div898/handbook/>.
- Park, H.-S. and Kim, J.-D. (2011). Modeling and Analysis of DTN in Metropolitan Bus Network. In *ICUIMC '11*, pages 20:1–20:10, New York, NY, USA. ACM.
- Sebastiani, M. T., Luders, R., and Fonseca, K. V. O. (2016). Evaluating electric bus operation for a real-world brt public transportation using simulation optimization. *IEEE Transactions on Intelligent Transportation Systems*, PP(99):1–10.
- Souza, R., Oliveira, I. P., Junior, F., Sales, L., and Ferraz, F. (2015). Beyond efficiency: How to use geolocation applications to improve citizens well-being. In *The Fourth International Conference on Smart Systems, Devices and Technologies*, pages 37 – 40.
- Stenneth, L., Wolfson, O., Yu, P. S., and Xu, B. (2011). Transportation mode detection using mobile phones and gis information. In *GIS '11*, pages 54–63, New York, NY, USA. ACM.
- Suresh, D. C., Agrawal, B., Yang, J., and Najjar, W. (2009). Energy-efficient encoding techniques for off-chip data buses. *ACM Trans. Embed. Comput. Syst.*, 8(2):9:1–9:23.
- Vilajosana, I., Llosa, J., Martinez, B., Domingo-Prieto, M., Angles, A., and Vilajosana, X. (2013). Bootstrapping smart cities through a self-sustainable model based on big data flows. *IEEE Communications Magazine*, 51(6):128–134.
- Zannin, P. H. T., Calixto, A., Diniz, F. B., and C., J. A. (2003). A survey of urban noise annoyance in a large brazilian city: the importance of a subjective analysis in conjunction with an objective *Environmental Impact Assessment Review*, 23:245–255.
- Zannin, P. H. T., Diniz, F. B., and Barbosa, W. A. (2002). Environmental noise pollution in the city of Curitiba, Brazil . *Applied Acoustics*, 63(4):351 – 358.

Análise da Utilização de Interface Multimodal para Aplicação de TV Digital e Segunda Tela

Renan Patrick de C. Marçal, Cláudia Caroline L. de Oliveira, Marcos César da Rocha Seruffo

Laboratório de Desenvolvimento de Sistemas (LADES) – Faculdade de Computação –
Universidade Federal do Pará (UFPA) – Campus Castanhal
Caixa Postal 15.064 – 91.501-970 – Castanhal – Pará – Brasil

renancarvlho@gmail.com, carollemarc@hotmail.com, seruffo@ufpa.br

Abstract. *This paper aims to conduct a comparative use of natural user interfaces, through the use of an interactive application that streamlines the interrelationship between the user and the digital TV, using for this second screen and the kinect device. The evaluation of user experience quality is performed through quantitative and qualitative metrics. The purpose of this proposal is to find the means of more direct and natural interaction goal to the user's needs, so it does not interfere with the traditional way of watching TV.*

Resumo. *Este artigo tem como objetivo realizar um comparativo da utilização de interfaces naturais do usuário, através do uso de uma aplicação interativa que dinamiza a inter-relação entre o usuário e a TV digital, utilizando para isso segunda tela e o dispositivo kinect. A avaliação da qualidade de experiência do usuário é realizada através de métricas quantitativas e qualitativas. O objetivo desta proposta é encontrar o meio de interação natural mais direto e objetivo às necessidades do usuário, de forma que não interfira na maneira tradicional de assistir TV.*

1. Introdução

A interação homem-computador (IHC) tem sido um dos grandes focos dos desenvolvedores nos últimos tempos. Estudos relacionados à qualidade de experiência do usuário, seu nível de satisfação e facilidade de interação com o sistema, tem ganhado destaque em congressos, pesquisas e afins. Segundo [Karray 2008] o conceito de IHC formou-se juntamente com o surgimento dos computadores e mais recentemente das máquinas em geral, devido ao fato de que algumas máquinas precisam ser utilizadas de forma adequada pelo homem.

Com o crescimento de estudos das interfaces na área de IHC e o grau de influência que estas exercem sobre a interação com o usuário, desenvolveu-se este trabalho visando avaliar e medir o nível de interação do usuário através de Interfaces Naturais (INs). Optou-se por abordar a área de TV Digital Interativa (TVDi), devido sua consolidação e crescente ascensão no cenário atual da rede de televisão brasileira.

Segundo um estudo do IBOPE, realizado em 2014, encomendado pela Secretária de Comunicação Social (SECOM), 97% dos entrevistados veem TV enquanto que 47% têm o hábito de acessar a internet, porém, os valores diferem um pouco quando se trata de preferência, segundo o mesmo estudo, a TV é o meio preferido para 76,4% da população, enquanto que 13,1% preferem a internet. [IBOPE, 2014]

Em conjunto com a TVDi foi feito uso de segunda tela, conceito este que atualmente está ganhando maior visibilidade e interesse na área acadêmica, aumentando

o número de pesquisas e aplicações em desenvolvimento. A segunda tela ou second screen é utilizada como complemento da interação por INs, acarretando em uma interação híbrida por parte do usuário, através da disponibilização de conteúdo adicional.

Através da avaliação da utilização da interface multimodal, este trabalho visa uma comparação das interfaces naturais por gestos e por voz, por meio de métricas quantitativas e qualitativas. Observando a qualidade de interação e o grau de satisfação dos usuários.

Este artigo está dividido em sete seções. Na segunda seção é feita a revisão literária descrevendo de forma sucinta as tecnologias e conceitos que serviram de base para este trabalho. Na terceira seção são citados os trabalhos que serviram de fundamento para o desenvolvimento deste estudo. Na quarta seção é mostrada a arquitetura desenvolvida, onde é explicado o funcionamento da aplicação desenvolvida, algumas etapas do desenvolvimento. Na quinta seção é descrito o estudo de caso, o ambiente de testes utilizado e a forma como ocorreram os testes. Na sexta seção são citados os resultados obtidos pelos testes e questionários realizados com usuários. Por último a sétima seção é composta pela conclusão, relatando as considerações finais e as propostas de trabalhos futuros.

2. Revisão literária

Nesta seção serão abordados (de forma sucinta) as principais tecnologias e conceitos que foram utilizados para o desenvolvimento do aplicativo proposto neste artigo, pois, durante a aplicação do questionário na fase de testes, notou-se falta de conhecimento dos usuários sobre estes conceitos.

2.1 Interface Natural do Usuário

O papel básico da interface do usuário é permitir que se estabeleça a comunicação entre a máquina e o usuário. Quanto mais simples de manusear e menor o grau de dificuldade para aprender a manipular determinada interface, maior será a capacidade de interação humano-computador. “De fato, a solução perfeita seria aquela na qual o usuário nem percebesse a utilização de uma interface” [Tebbutt 1995].

A evolução das interfaces, associada à busca pela melhor qualidade e facilidade de interação do usuário com sistemas computacionais, caminha cada vez mais para a utilização de interfaces naturais (*Natural User Interface* - NUI), como a opção mais intuitiva e direta dentre as alternativas existentes. NUI segue um novo paradigma com o objetivo de respeitar a percepção humana. Segundo [Valli 2007], interagir com sistemas com Interfaces Naturais é extremamente intuitivo e agrada a todos envolvidos.

2.2 TV Digital Interativa

Duas outras tecnologias, abordadas no presente trabalho, ascendentes no campo dos desenvolvedores e pesquisadores da ciência da computação, são a TVDi (TVDi) e a segunda tela, tecnologias que também visam o conforto dos usuários. A característica básica das iTVs é a possibilidade de controle do usuário sobre a experiência com a TV, podendo definir o curso de interação com o conteúdo disponibilizado. [EY 2013]

Segundo [Fernandes 2004, p. 22], “Um sistema de televisão digital interativa deve adotar e integrar um conjunto de diferentes tecnologias de hardware e software para implementar suas funcionalidades. Os conteúdos de interatividade apresentados

durante a programação da TV podem ser exibidos na primeira tela (geralmente a TV), no menu desta ou através de uma Segunda Tela.

2.3 Segunda tela

O uso da internet aliado a conteúdos televisionados tem modificado as formas tradicionais de o usuário assistir TV, enquanto recebe informações da primeira tela o telespectador pode compartilhar e trocar informações com outros usuários através da segunda tela. É definido como segunda tela a ação de utilizar dispositivos móveis como meio de interação para compartilhamento e troca de informações com outros usuários enquanto assiste a programas de TV

A utilização de segunda tela é realizada através do acesso ao conteúdo adicional da programação televisionada na primeira tela. Para a entidade que dirige e que rege todas as ações e estratégias das empresas referentes a segunda tela, essa, é toda e qualquer “experiência de engajamento da audiência, que inclui a TV Social como um elemento integrante”. [SET, 2013].

Utilizou-se a aplicação (app) Dengue Fever (DFapp) para acesso do conteúdo adicional. O DFapp é um app que tem o foco na doença Dengue, uma doença que mata milhares de pessoas por ano no Brasil. Através da utilização deste app são disponibilizadas informações adicionais sobre a doença em questão.

Neste trabalho foi feito uso de autenticação através do uso do QR Code, por sua velocidade de processamento e sua interação com dispositivos móveis (segunda tela). O termo QR é acrônimo de Quick Response, que em português significa resposta rápida. É um código de barras 2D que pode ser detectado e traduzido por celulares que possuem uma câmera e um aplicativo específico para leitura deste código [Gazolla neto 2012]

3. Trabalhos Correlatos

As tecnologias abordadas nesta pesquisa estão sendo estudadas há algum tempo e têm ganhado destaque no âmbito das Tecnologias de informação e comunicação. Vários trabalhos se destacam nessa área, embora a maioria não una essas tecnologias em um trabalho apenas.

[Cruickshank 2007] desenvolveu um trabalho de pesquisa em que este analisou a qualidade da interação dos usuários com a TVDi e as respectivas barreiras encontradas na utilização das correntes interfaces de interação utilizadas, através disso, propôs uma solução para a problemática adicionando o uso de segunda tela, durante a utilização das TVDi's, diminuindo o número de informações na primeira tela e melhorando a interação dos usuários com a TV e a navegação pelos serviços disponíveis.

[Turk 2013] realizou uma análise sobre a interação multimodal, passando através da evolução, das possibilidades e dos desafios da área, focando em um dos pontos críticos no que se refere a interação multimodal; a integração dos diferentes modos. O autor pondera sobre as possibilidades de interação precoce ou tardia, analisando evidências dos diferentes tipos.

[Kong 2011] quantifica a preferência dos usuários por cada modalidade de interação (interface multimodal), e busca adaptar o modelo de interação baseado em requisitos de QoS analisados. Através da realização de testes, compara as modalidades de interação selecionadas automaticamente e as selecionadas aleatoriamente, obtendo resultados satisfatórios para a modalidade de interação selecionada automaticamente, baseada nos ajustes feitos através da análise das necessidades pessoais dos usuários.

[Bobeth 2014] realizou um trabalho semelhante ao proposto neste artigo, analisando a qualidade de interação em plataformas multimídias de TVDi de diferentes aplicações populares na área. O autor analisa três diferentes tipos de entrada de interação com a TV Digital (*tablet*, gesto mão-livre, controle remoto), analisando o efeito da performance em diferentes grupos etários, e a influência da idade na qualidade de experiência dos usuários. Através do estudo constatou grande influência da modalidade de dispositivo de entrada sobre a performance e qualidade de experiência, enquanto que a idade pareceu não influenciar na mesma.

A diferença deste trabalho para [Bobeth 2014], é que no presente trabalho são avaliadas duas formas de NUI (voz e gesto) levando em consideração a qualidade da experiência do usuário, quanto que Bobeth avalia diferentes tipos de interfaces levando em consideração a influência da idade dos usuários na qualidade de experiência dos diferentes métodos interativos.

Utilizou-se TV Digital e a tecnologia de segunda tela em conjunto com o uso de interfaces focando o estudo assim como [Bobeth 2014], na análise dos dispositivos de entrada, usando porém, comandos por gestos e por voz. É realizada ainda uma análise do perfil dos usuários através da mineração dos dados coletados por meio de testes, procurando delimitar fatores importantes que influenciam a performance e a qualidade de experiência dos usuários.

4. Arquitetura Desenvolvida

A arquitetura da aplicação proposta foi desenvolvida com o intuito de dar ao usuário da TVDi total liberdade de interação, sem afetar a usabilidade das funções padrões da TV ou prejudicar os comportamentos tradicionais dos usuários que assistem televisão.

A utilização de recursos de segunda tela através da aplicação desenvolvida pode ser feita a qualquer momento durante uma programação, sem a necessidade de redirecionamento para uma tela de autenticação, o que neste último caso, ocasiona perda do conteúdo televisionado, ou a necessidade de redimensionamento da tela para que o *QR Code* possa ser disponibilizado ao lado da programação corrente. O *QR Code* ficará oculto durante toda a programação, aparecendo ao comando (de voz ou gestual) e sendo ocultado novamente após o comando referente a essa ação.

Na Figura 1 é possível ter uma visão esquemática do funcionamento da aplicação, onde se observa a interação dos usuários com a televisão e simultaneamente com o dispositivo de sensor de movimentos Kinect, escolhendo um dos modos de interação disponíveis, gesto ou voz e depois autenticando com um segundo dispositivo, ilustrado, neste caso, como um celular. Pode-se observar que os componentes envolvidos no funcionamento da aplicação encontram-se conectados à uma rede.

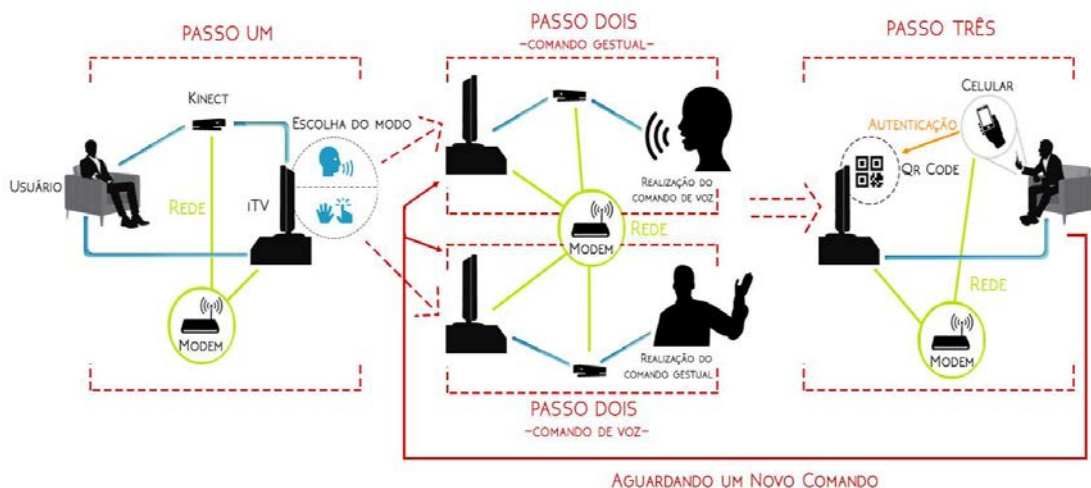


Figura 1- Visão Geral do Sistema

A aplicação foi desenvolvida através da IDE *Microsoft Visual Studio*¹, com o auxílio do SDK do sensor de movimentos *Kinect*, periférico escolhido para a intermediação da interação entre usuário e aplicação, devido a sua relativa facilidade de manuseio e baixo custo de aquisição. O desenvolvimento da aplicação foi feito de forma a criar um módulo de fácil adaptação e acoplamento em distintas aplicações, tendo como foco inicial a inter-relação com aplicativos desenvolvidos em HTML5, para TV Digital.

O *kinect* apesar de ter sido criado inicialmente para a área de jogos passou a ser utilizado em diversas outras áreas como tecnologia de desenvolvimento de interfaces naturais, devido ao baixo custo e a grande capacidade computacional. [Silva 2013].

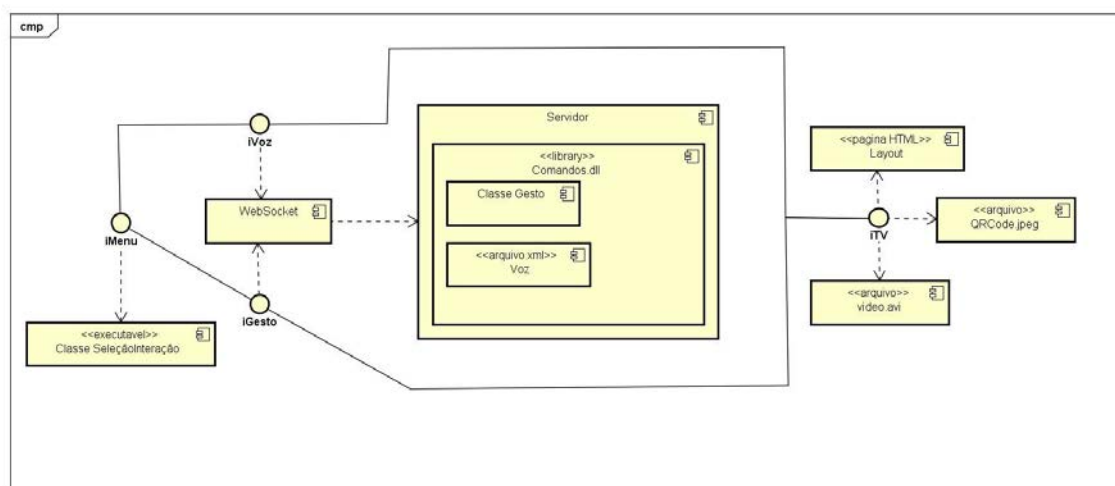


Figura 2 - Diagrama de Componentes do Sistema

A arquitetura da aplicação é demonstrada na Figura 2 através de um diagrama de componentes. É inicializada pela interface do menu de seleção do modo de interação desejado pelo usuário, que dependendo desta escolha é direcionado para a respectiva interface de cada interação natural, (voz ou gesto). A comunicação entre a aplicação da TV e o servidor é realizada através do *websocket*, no servidor está armazenada uma

¹ <https://www.visualstudio.com>

biblioteca que contém os comandos gestuais e as palavras reservadas para realização da interação por voz. Outra interface do sistema é a da TVDi, que é composta por um arquivo HTML que contém um determinado vídeo (programação).

Na escolha da interação por gestos o servidor ficará aguardando novos comandos que serão enviados a partir do reconhecimento de um movimento especificado através do sensor de movimentos, enquanto a programação pode ser exibida normalmente, o gesto será identificado, enviando uma mensagem para o servidor C#, através de *websocket*. O *websocket* retornará uma mensagem para a TV que então exibirá o *QR Code*.

Caso o usuário escolha a opção por comandos de voz o servidor ficará aguardando novos comandos que serão enviados a partir do reconhecimento da pronuncia de uma palavra especificada, enviando uma mensagem para o servidor C#, através de *websocket*. Novamente, o *websocket* retornará uma mensagem para a TV que então exibirá o *QR Code*, possibilitando ao usuário acessar os conteúdos adicionais através de seu dispositivo móvel.

Abaixo, na Figura 3 é apresentado o fluxo de execução em camadas de acesso, onde pode-se observar na camada superior, mais externa, as interfaces inicial, de gestos e de voz.

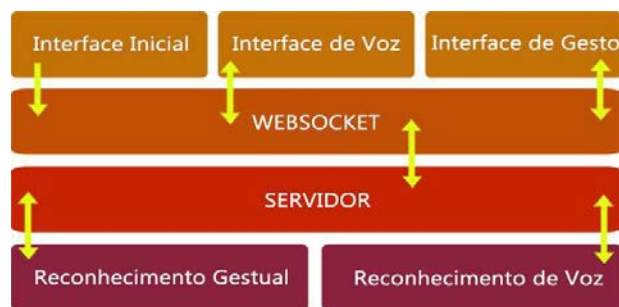


Figura 3. Fluxo de Execução em Camadas

Como demonstrado na Figura 3, a interface inicial é o ponto de partida. O fluxo segue então para a segunda camada, do websocket, onde poderá ser direcionado para uma das outras interfaces, podendo retornar a camada do websocket posteriormente e então passar para as camadas seguintes do servidor e finalmente a camada responsável pela identificação do comando realizado, retornando posteriormente por todas as camadas e exibindo a devida resposta a requisição na interface selecionada.

A proposta inicial era utilizar ambas as formas de interação na tela da interface inicial, porém, não foi possível realizar o preterido, devido à dificuldade de implementação dos dois modos de interação simultaneamente em uma interface, encontrada durante o desenvolvimento do aplicativo.

Parte da lógica das interfaces de gesto e voz é apresentada na Tabela 1, a seguir. Para a interação por voz, foram selecionadas três palavras base, sendo estas, “*Play*” para iniciar o vídeo da aplicação, “*Show*” para exibir o *QR Code* e “*Hide*” para ocultar.

Para interação natural por gestos o SDK do *kinect* disponibiliza as **Joints**, que são as articulações que o *kinect* consegue rastrear. Através da função de rastreamento do corpo é possível implementar diversos tipos de movimentos e poses utilizando **Joints**. Na Tabela 1 é demonstrado o uso das **Joints** (*handRight* e *head*). Na interface por gesto a pose da mão direita acima da cabeça foi implementada para ativar e desativar o *QR*

Code, podendo-se observar, de forma sucinta, ambas as formas de implementação das interfaces naturais utilizando o kinect.

Tabela 1. Código das Interface de Voz e Gestos

VOZ	GESTO
<pre>private void SpeechRecognized(object sender, SpeechRecognizerEventArgs e) { //Iniciar video const string Play = "PLAY"; //Remover QRcode const string Hide = "HIDE"; //Habilitar QRcode const string Show = "SHOW"; } </pre>	<pre>public class PoseQr : Pose { public PoseQr() { this.Nome = "PoseQr"; } protected override bool ValidamentoPos(Skeleton esqueletoUsuario) { Joint maoDireita = esqueletoUsuario.Joints[JointType.HandRight]; Joint cabeca = esqueletoUsuario.Joints[JointType.Head]; Bool maoDireitaAlturaCorreta = maoDireita.Position.Y > cabeca.Position.Y; return maoDireitaAlturaCorreta; } } </pre>

Ambas as implementações foram destacadas para dar ênfase no processo de criação de interfaces multimodal. Para interação por voz, utilizou-se o método que disponibiliza serviços de comandos de voz **SpeechRecognizer**, para criação e configuração da gramática utilizada no arquivo XML que contém as palavras bases para utilização da aplicação.

Na codificação da interface por comandos gestuais, a classe “PoseQr” que habilita e desabilita o *QR Code*, herda parâmetros de validação e correção da classe “Pose”, em caso da pose não estar completamente de acordo com o que foi pré-estabelecido para execução correta. O método “ValidamentoPos” rastreia as **joints** que estão sendo utilizadas, a variável “maoDireitaAlturaCorreta” verifica se a pose está sendo executada de forma correta, neste caso comparando a posição de ambas **joints** no eixo “Y” (vertical).

5. Estudo de Caso

Foram realizados testes com 26 usuários, selecionados aleatoriamente, entre alunos e professores dos diferentes cursos de graduação da Universidade Federal do Pará, campus Castanhal. A interatividade foi testada com mais de um usuário simultaneamente, desta forma, o aplicativo funciona a partir da interação natural apenas de um dos presentes usuários, então todos que estiverem assistindo a televisão podem interagir através da autenticação por *QR Code* com o conteúdo da segunda tela.

Para a realização dos testes procurou-se desenvolver uma atmosfera que simulasse o ambiente de utilização natural da televisão, tendo como base uma sala de estar, onde, buscando minimizar possíveis interferências e impactos negativos do ambiente de testes na interação do usuário com a televisão e o aplicativo, bem como obter resultados mais próximos possíveis do real, que poderiam ser observados em condições usuais.

Foram utilizados um sensor Kinect, para a detecção dos movimentos dos usuários, uma iTV onde foi disponibilizado o conteúdo de primeira tela, e um aparelho celular para a disponibilização do conteúdo de segunda tela, que consistia em conteúdo adicionais ao conteúdo exibido na primeira tela.

Os testes deram-se segundo uma ordem preestabelecida, com o intuito de guiar os pesquisadores envolvidos e manter o controle sobre a realização dos testes e a coleta de dados de forma sistemática. Em uma etapa prévia, os usuários eram convidados a preencher um questionário de teor socioeconômico, com campos de informações básicas. Em seguida os usuários eram direcionados, individualmente, à sala de testes, com dois desenvolvedores e uma câmera, onde recebiam as instruções básicas necessárias para a realização dos testes.

Cada usuário realizou o teste na interface por comando de voz e na interface por comandos gestuais. O usuário tinha total liberdade para escolher a interface de interação inicial. Toda a aplicação era igual para as diferentes modalidades. É possível observar a utilização das distintas interfaces, de gesto e de voz respectivamente, nas Figuras 4.1 e 4.2, abaixo.



**Figura 45.1 Realização dos Testes-
Comando por Gesto**



**Figura 4.2 Realização dos Testes-
Comando por Voz**

Inicialmente era exibida uma tela com a mensagem indicando ao usuário que opta-se por um dos modos de interação indicados em dois botões. A escolha ocorre através de comando gestual. Na tela da interface escolhida eram disponibilizadas as opções relacionadas a cada interface de interação, botões para comandos gestuais e gestos para interface gestual e as palavras reservadas para interface por voz.

Na interface de comando gestual eram exibidos dois botões, um para iniciar ou pausar o vídeo disponibilizado e outro para exibir ou ocultar o *QR Code*, sendo que o usuário poderia, também, exibir e ocultar o *QR Code* através de uma pose previamente estabelecida na aplicação. Na interface de comando por voz eram exibidas, ao lado do vídeo, as palavras preestabelecidas, responsáveis por iniciar o vídeo, exibir e ocultar o *QR Code*, respectivamente, *Play*, *Show* e *Hide*.

Após exibir o *QR Code* os usuários podiam realizar o pareamento com o aparelho de segunda tela disponibilizado ao início do teste e utilizar livremente o conteúdo extra, disponibilizado na segunda tela. Após o fim do teste, os usuários preencheram formulários relacionados a qualidade de interação, utilização e nível de conhecimento prévio acerca das tecnologias utilizadas, através dos quais foram obtidos os dados necessários para a realização dos estudos, análises e delineamento dos perfis dos usuários.

Além dos questionários, também foram realizadas observações por parte dos pesquisadores, relacionadas ao comportamento desempenhado pelos usuários no decorrer dos testes.

6. Resultados

Através da realização dos testes buscou-se definir padrões através da análise dos perfis de usuários, identificar o nível de satisfação dos usuários com relação à aplicação desenvolvida e detectar possíveis falhas e pontos a serem melhorados na mesma, desta forma disponibilizando este arcabouço de informações, novos desenvolvimentos que contemplam a mesma seara deste trabalho podem utilizar esta referência.

Os dados recolhidos através dos questionários foram tabulados para posterior análise. O perfil dos usuários da amostra dos 26 usuários convidados pode ser traçado, através da observação dos dados coletados, da seguinte forma na Figura 5: gênero dos usuários, a faixa etária, a renda mensal, o conhecimento prévio sobre NUI e *QR Code*.



Figura 5. Perfil do Usuário

Como mostrado na Figura 5, 73% dos usuários já haviam utilizado o *QR Code* anteriormente, a grande maioria para utilização do software whatsapp no computador, dessa forma apenas 27% não conhecia a tecnologia, desses mais da metade possuía renda de 1 a 3 salários mínimos. A tecnologia do *QR Code*, como mencionado anteriormente, foi utilizada para a autenticação e posterior utilização da segunda tela. A relação entre a renda dos usuários e o conhecimento prévio do *QR Code* é apresentada na Figura 6.

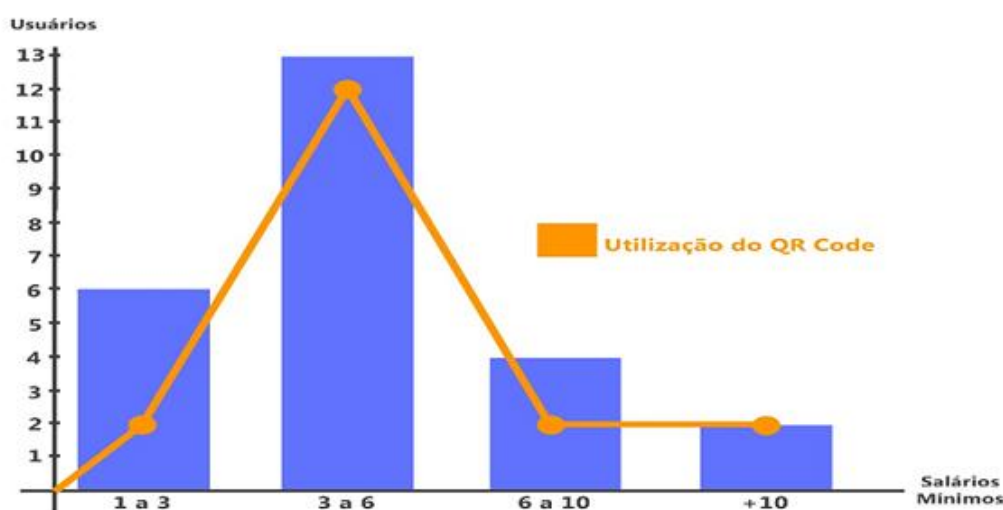


Figura 6. Relação entre nível de renda e conhecimento prévio sobre o QRCode

Dentre os participantes do teste, apenas 8%, nunca haviam utilizado nenhum tipo de interface natural, fato que não afetou o desempenho dos usuários durante a

interação com o sistema, embora a faixa de tempo utilizada na interação por gesto tenha sido relativamente alta, ela não fugiu ao padrão dos usuários com experiência prévia, da mesma forma com a interação por voz, obtendo, no entanto um tempo inferior à interação por gesto. Na Figura 7 é possível observar o tempo de execução para cada modo de interação.

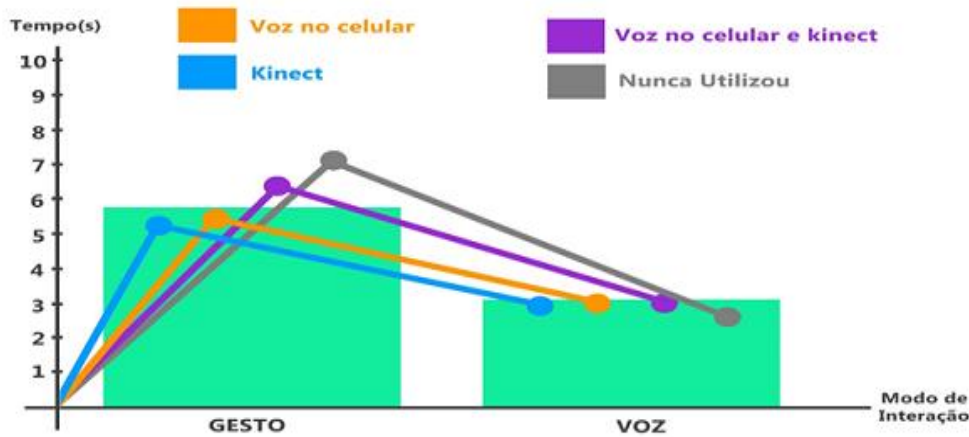


Figura 7. Tempo médio de execução para cada modo de interação e de cada segmento baseado na utilização prévia de NUI.

Os usuários levaram entre 5 e 6 segundos, para executar os comandos gestuais, enquanto que para efetuar os mesmos comandos por voz levaram cerca de 2 a 3 segundos, menos da metade do tempo que o primeiro como pode ser observado na Figura 8. Tal configuração era esperada, já que falar é muito mais intuitivo que fazer determinados gestos e poses predefinidas, porém, como mencionado anteriormente a escolha da pose pode, também, ajudar no desempenho do usuário, com base nisso a pose padrão foi substituída após a presente análise.



Figura 8. Área do tempo para primeira execução

Apesar de muito mais intuitiva a interface por comando de voz, teve como empecilho, durante a sua utilização, o fato das 3 palavras selecionadas serem em inglês, desta forma, o número de tentativas realizadas para que cada uma dos comandos predefinidos fossem executados com sucesso foi divergente entre eles (Figura 9), onde o comando utilizado com um maior grau de facilidade, ou seja, executado com um menor número de tentativas, foi o comando *Show Me*, seguido do comando *Play*, sendo que o comando de maior dificuldade de pronúncia e dessa forma de execução foi o comando *Hide*.

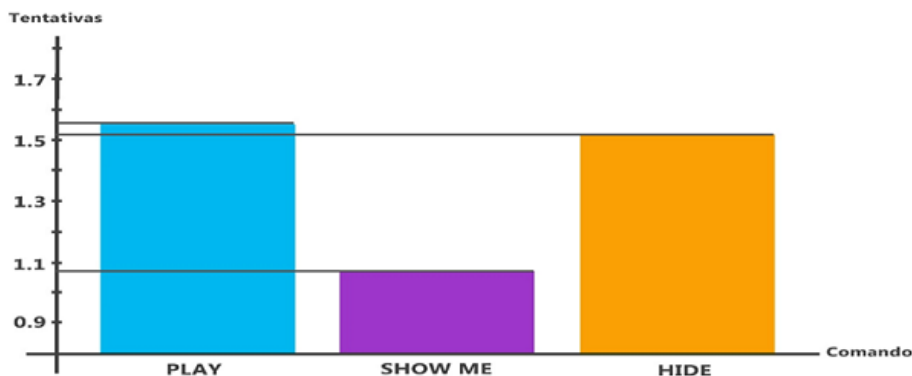


Figura 9. Média da variação do número de tentativas por comando

Um aspecto interessante relacionado à preferência dos usuários, com relação ao modo de interação, foi o fato de que todos os usuários que afirmaram preferir a interação por gesto, ainda não haviam utilizado essa tecnologia, os usuários que haviam utilizado interação gestual anteriormente além disso, foram unânimes ao declarar preferência pelo modo de interação por voz. (Figura 10)

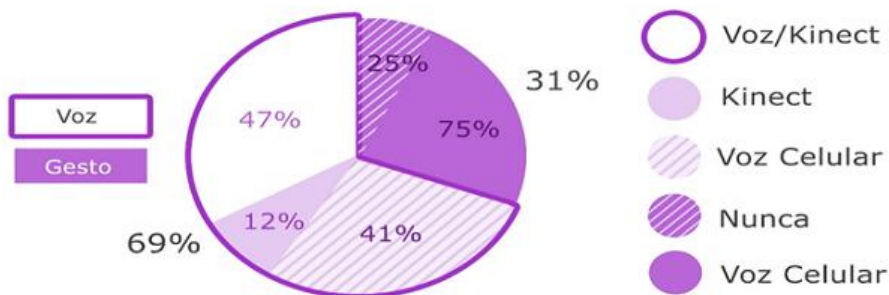


Figura 10. Preferência por interação relacionada ao uso anterior de NUI

De forma geral, os usuários não tiveram dificuldades em utilizar o aplicativo, mesmo os participantes que nunca haviam tido contato com Interfaces Naturais. O esforço e o tempo demandado para a execução do comando de voz foi inferior ao demandado para a execução dos comandos gestuais. O comando gestual mostrou-se bem aceito em um primeiro momento, porém, através da análise dos dados pode-se observar que quem já utilizou esse tipo de comando mais de uma vez tende a preferir o comando de voz por ser mais cômodo e eficaz.

7. Conclusão

O presente artigo expôs uma proposta utilizando aplicativo para TV Digital e Segunda Tela, para teste com interfaces como meios de interação, buscando dar ao usuário maior conforto, comodidade, e maior desempenho na utilização da TV Digital e da Segunda Tela.

A aplicação desenvolvida apesar de simples proporciona uma experiência nova para os telespectadores, e o aumento na qualidade de experiência do usuário, dando a este a possibilidade de utilizar sua TV segundo sua vontade, podendo operar ou não os recursos disponíveis.

Este trabalho permitiu a realização de uma análise distinta de interatividade do usuário, através de interfaces naturais para interação com a TV Digital e segunda tela,

possibilitando a utilização de aplicações de interações híbridas, configurando um novo nível de interação. E os resultados obtidos nos testes, podem ser utilizados para comparação com outras pesquisas na área de NUI e IHC. Ressalta-se que a aplicação de segunda tela utilizada neste trabalho foi desenvolvida pelo próprio grupo de estudo do Laboratório de Desenvolvimento de Sistemas (LADES) da UFPA.

No intuito de aprimorar o estudo desenvolvido, pretende-se estabelecer uma comparação relacionada ao desempenho de diferentes dispositivos de NUI, como *leap motion*, levando em consideração a qualidade de experiência dos usuários associada a utilização dos diferentes dispositivos e modos de entrada.

Referências

- Bobeth, J. Schrammel, S. Deutsch, M. Klein, M. Drobics, C. Hochleitner, and M. Tscheligi, Tablet, gestures, remote control?: influence of age on performance and user experience with iTV applications. In: Proc. of the 2014 ACM international conference on Interactive experiences for TV and online video, 139-146.
- Cruickshank, L., Tsekleves, E., Whitham, R., Hill, A. 2007. Making interactive TV easier to use: interface design for a second screen approach. *The Design Journal*, 10 (3).
- EY. Future of Television. EYGM Limited. 2013. Disponível em: [http://www.ey.com/Publication/vwLUAssets/EY_-_6_trends_that_will_change_the_TV_industry/\\$FILE/EY-6-trends-that-will-change-the-TV-industry.pdf](http://www.ey.com/Publication/vwLUAssets/EY_-_6_trends_that_will_change_the_TV_industry/$FILE/EY-6-trends-that-will-change-the-TV-industry.pdf)
- Fernandes, Jorge; Lemos, Guido; Silveira, Gledson. Introdução à televisão digital interativa: arquitetura, protocolos, padrões e práticas. In: Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação, 2004, Salvador. Anais do JAI-SBC. 56p.
- Gazolla neto, Alexandre, et al. 2012. Rastreabilidade aplicada à produção de sementes de soja. *Informativo ABRATES*. v.22, n.2
- IBOPE, DISPONÍVEL EM: < <http://agenciafosforo.com.br/2014/03/midias-vidabrasileiros-pesquisa-ibope-2014/> >, Acessado em: 16 DE FEVEREIRO, 2015
- Karray, F., M. Alemzadeh, J.A. Saleh, and M.N. Arab. Human-Computer Interaction: Overview on State of the Art. In *International Journal on Smart Sensing and Intelligent Systems*, Vol. 1, No. 1, March 2008.
- Kong, J., W. Y. Zhang , N. Yu , X. J. Xia. Design of human-centric adaptive multimodal interfaces, *International Journal of Human-Computer Studies*, v.69 n.12, p.854-869, December, 2011.
- SET. 2013 DISPONÍVEL EM: http://www.set.org.br/artigos/ed133/ed133_pag84.asp, acessado em 07 de abril, 2015
- Silva, Renato K. Interfaces Naturais e o Reconhecimento das Línguas de Sinais. 2013. 130 p. Dissertação (Mestrado)-Pontifícia Universidade Católica de São Paulo, 2013
- Tebbutt, David. In touch with tomorrow. In *PC PRO*, Feb 1995, p. 206-11. Dennis Publication, United Kingdom.
- Turk, M. Multimodal interaction: A review. *Pattern Recognition Lett.* (2013), <http://dx.doi.org/10.1016/j.patrec.2013.07.003>.
- Valli, Alessandro. Notes on Natural Interaction. 2007.

Utilização de Redes Bayesianas para Avaliação da Qualidade de Experiência do Usuário: Um Estudo de Caso Com Segunda Tela

Joahannes Bruno Dias da Costa¹, Marcos Cesar da Rocha Seruffo¹,
Marianne Kogut Eliasquevici², Daniel da Silva Souza¹

¹Faculdade de Computação (FACOMP) – Universidade Federal do Pará (UFPA)
68746-360 – Castanhal – PA – Brasil

²Instituto de Ciências Exatas e Naturais (ICEN) – Universidade Federal do Pará (UFPA)
66075-110 – Belém – PA – Brasil

{joahannes, seruffo, marianne, danielssouza}@ufpa.br

Abstract. *This article presents a study on evaluation of the quality of the user experience from the use of a second screen application that interacts with television programming. For this, it was employed a model of computational intelligence – Bayesian network – to correlate the main factors that influence the user experience, these being: human factors, and system context. As a result, a discussion that aims to promote the improvement of user experience through criteria to assess applications, either during the design phase or during its operation.*

Resumo. *Este artigo apresenta um estudo sobre avaliação da qualidade da experiência do usuário a partir do uso de um aplicativo de segunda tela que interage com a programação televisiva. Para isto, foi empregado um modelo de inteligência computacional – Rede Bayesiana – para correlacionar os principais fatores que influenciam na experiência do usuário, sendo estes: fatores humanos, sistema e de contexto. Como resultado, é apresentada uma discussão que visa promover a melhoria da experiência do usuário por meio de critérios que permitam avaliar os aplicativos, quer seja durante a sua fase de design ou durante sua fase de operação.*

1. Introdução

O crescimento do uso de múltiplas telas no momento da assistência à televisão (TV), no Brasil e no mundo, tem aberto perspectivas instigantes para os desenvolvedores interessados em entrar no mercado de segunda tela. A segunda tela nada mais é do que um recurso adicional para a primeira tela, tais como um notebook, tablet ou smartphone com um sistema operacional capaz de executar aplicativos instalados [Soskic 2014].

Para que além de os telespectadores simplesmente sentarem e assistirem o conteúdo televisivo, agora possuem a oportunidade de participar, comentar, escolher e compartilhar o que estão assistindo. Um fator que colabora para estas novas possibilidades é o barateamento do preço dos dispositivos móveis, aliado ao aumento do poder de processamento que os mesmos estão ganhando.

São várias as emissoras que estimulam a interação entre o telespectador e o programa televisivo por meio das redes sociais, fazendo com que esse mercado venha crescendo exponencialmente. Uma pesquisa indica que 57% dos brasileiros usam o Facebook ao mesmo tempo em que assistem TV. Dentre eles, 39% escolhem um dispositivo móvel como sua segunda tela. Nos horários entre 20h00min e 00h00min, oito em cada dez brasileiros usam a rede social enquanto estão vendo TV, representando 81% dos telespectadores [IpsosMediact 2014].

Durante a Copa do Mundo de 2014, por exemplo, enquanto os jogos eram transmitidos, os telespectadores teciam comentários nas redes sociais. No Twitter foram registrados 672 milhões de *tweets* sobre o Mundial, transformando o torneio no evento mais tuitado de todos os tempos; já no Facebook foram registradas três bilhões de interações entre usuários desde o início do evento, sendo a primeira vez que essa marca é atingida [Copa 2014].

Preocupadas em não perder mercado, emissoras brasileiras (Rede Globo, Rede Record, TV Cultura e o Sistema Brasileiro de Televisão – SBT) começam a se mobilizar dentro deste segmento para a TV aberta e lançam aplicativos para que os telespectadores acompanhassem e interagissem com a programação que estava sendo exibida. Dentre algumas dessas experiências de segunda tela no Brasil, se encontram: o aplicativo desenvolvido para o reality show Menino de Ouro, exibido pelo SBT e o reality show Superstar, exibido pela Rede Globo.

Toda essa transformação tem impacto na qualidade de experiência do usuário, o que implica na necessidade de se investigar de forma mais profunda a forma com que o usuário se comporta e interage com aplicações, ou mesmo o porquê do mesmo não ter interesse em interagir, para que se possa prover serviços que satisfaçam os diversos perfis de usuários existentes.

Diante deste contexto, a contribuição deste artigo é a proposição de um estudo sobre a qualidade da experiência do usuário no momento da interação¹ com um aplicativo de segunda tela, empregando a técnica de inteligência computacional Rede Bayesiana (RB) para correlação dos dados aferidos. Por meio da extração de conhecimento da base de dados obtida durante os testes realizados com o aplicativo, torna-se possível a montagem de um arcabouço que identifica fatores que avaliam o uso de aplicativos de segunda tela, quer seja durante a sua fase de design ou operação.

O restante do trabalho está organizado da seguinte forma: a Seção 2 apresenta a fundamentação teórica aplicada com conceitos de qualidade de experiência e sobre Redes Bayesianas; a Seção 3 mostra alguns trabalhos correlacionados que tratam de temas presentes nesse trabalho; a Seção 4 explicita o estudo de caso, com a arquitetura montada para avaliação, o aplicativo desenvolvido, o cenário de testes e os dados extraídos para análise; a Seção 5 apresenta os resultados e discussões da aplicação da metodologia proposta; e por fim, a Seção 6, apresenta conclusões sobre a pesquisa apresentada e apontamentos para trabalhos futuros.

¹Entende-se como interação qualquer ação que o usuário realiza sobre o dispositivo de segunda tela, que nesse caso, é no smartphone ou tablet que está com o aplicativo instalado.

2. Fundamentação Teórica

2.1. Qualidade de Experiência

Um conceito que tem sido muito empregado para avaliar serviços computacionais é a Qualidade da Experiência (QoE) que, para a *International Telecommunication Union* [ITU-T 2007], é a aceitabilidade geral de um aplicativo ou serviço, conforme percebida subjetivamente pelo usuário final, incluindo os efeitos de sistema (qualidade da rede, serviços de infraestrutura, entre outros).

No geral, aceitabilidade pode ser influenciada pelo contexto e expectativas do usuário [ETSI 2009]. Conforme [Geerts et al. 2010], as expectativas dos usuários, a satisfação e a QoE (percebida) estão sendo reconhecidas como determinantes para o sucesso de qualquer tecnologia, sendo ainda mais importante do que o desempenho isolado tecnológico e de excelência, em muitos casos definido como Qualidade de Serviço (QoS).

A QoE tem como intenção mapear os fatores que exercem influência na experiência de uso da tecnologia. A influência pode vir de qualquer característica de um usuário, sistema, serviço, aplicativo ou contexto. Assim, a literatura classifica QoE como uma interseção de fatores de influência, ou seja: Fatores Humanos (FH), Fatores de Contexto (FC) e Fatores de Sistema (FS).

2.2. Redes Bayesianas

A utilização de modelos de inteligência computacional na resolução de problemas é muito comum e eficiente, pois é sabido o alto nível de confiabilidade que os mesmos possuem. Neste estudo, foi utilizada a técnica de Rede Bayesiana (RB), que pode ser definida como uma rede probabilística constituída por uma estrutura gráfica e uma distribuição de probabilidades a ela associada [Sarabando 2011]. Sua representação utiliza grafos acíclicos direcionados com suas respectivas probabilidades condicionais.

Os eventos que influenciam uma variável influenciam diretamente a quantidade de informação na outra variável devido à correlação existente entre elas [Ferrari 2012]. Para que se possa obter essa correlação entre variáveis é preciso relacionar as suas probabilidades, e tal conexão é obtida a partir da aplicação do teorema de Bayes, base das RBs, já que ele investiga a ocorrência de novas probabilidades com base em outras existentes.

Como a quantização do grau de influência que uma variável tem sobre as outras é a base para medir a QoE, se buscou uma forma de medição dessa influência e foi então que se chegou ao Índice de Carga, ou Entropia Condicional, das variáveis. Sendo assim, quando as variáveis x e y estão correlacionadas, isto é, então, a entropia de cada uma das variáveis é condicionada a outra e isso mostra o quanto uma variável influencia na outra.

3. Trabalhos Correlacionados

Diante do contexto de grande utilização de diversos dispositivos juntamente com a TV, estudos vem surgindo para investigar de que forma tais serviços podem ser melhor ofertados para os usuários.

Dentre os autores que apoiam a ideia de considerar, não somente dados ligados ao sistema em si, como também fatores ligados ao ser humano (emoções, experiência prévia, etc.) ao se avaliar a experiência de uso de determinado sistema pode-se citar

[Piccolo et al. 2010], que apresenta uma metodologia para avaliação da QoE de serviços de TV interativa que considera tanto fatores subjetivos relacionados ao comportamento humano, quanto fatores objetivos referentes ao desempenho do serviço oferecido. E também [Yong 2013] que discute a elaboração de um quadro conceitual para a experiência do usuário com o objetivo de medir a resposta emocional, comportamental e reflexiva de quem utiliza determinado sistema.

O grau de satisfação do usuário em relação a um determinado serviço é crucial para o estreitamento das operadoras de serviços e os usuários. O trabalho de [Zhou 2013] descreve um estudo que comprova a importância de se avaliar serviços de TV móvel com base no usuário. A pesquisa identifica que pontos como a facilidade de utilização, a velocidade de acesso e a qualidade do conteúdo têm efeitos significativos sobre a experiência de utilização desses serviços. Esse tipo de estudo pode ser um forte aliado para análise mercadológica, em que se busca traçar o perfil do usuário e entregar um serviço direcionado.

Referente ao emprego de Redes Bayesianas, como técnica de inteligência computacional de aplicação eficaz em diversos domínios, que não necessariamente o de QoE. A pesquisa de [Xenos 2004], por meio da aplicação de RB, avalia o comportamento de estudantes de um curso de computação ofertado na modalidade à distância, para verificar o quanto os fatores presentes nesse cenário influenciam no rendimento acadêmico dos estudantes e prevê possíveis egressos com base em previsões que esse tipo modelo oferece. E também [Da et al. 2011] que emprega RB com o objetivo de criar um modelo de escolha de recursos de aprendizagem, de acordo com as características pessoais dos alunos, utilizando *feedback* dos resultados dos exames dos cursos de formação como amostras de dados.

O trabalho de [Zhu et al. 2015] reforça a ótica de avaliação da QoE levando em consideração não apenas as propriedades técnicas multimídias do serviço oferecido, mas também o contexto social e fatores ligados ao ser humano para que essa avaliação seja mais completa e que estime suficientemente a satisfação do usuário final. Isso nos motivou na construção deste estudo que busca a criação de uma metodologia de avaliação da QoE que considera fatores ligados ao sistema, fatores humanos e fatores ligados ao contexto em que esse usuário está inserido. Este trabalho serviu de base para o desenvolvimento deste artigo e ressaltou a importância de estudos desta natureza.

Conforme exposto, são vários os estudos sobre QoE e inteligência computacional que podem ser encontrados na literatura. Dentro deste enfoque, este artigo traz inovações por conseguir relacionar, em um mesmo trabalho, características não encontradas, ou encontradas de forma dispersa na literatura pesquisada para apresentar a construção de um modelo genérico para avaliação de QoE que leva em consideração tanto parâmetros de sistema, obtidos de forma dinâmica da aplicação à medida em que o usuário interage com a aplicação, como também parâmetros extraídos do preenchimento de questionários, que englobam fatores humanos e de contexto.

4. Estudo de Caso

4.1. Arquitetura Montada

Foi montada uma arquitetura, mostrada na Figura 1, para o estudo de caso que descreve a abordagem metodológica empregada, envolvendo desde a etapa dos testes com usuários

até o uso de RB para modelar a QoE de um aplicativo de segunda tela, tendo como base os fatores constantes da QoE.

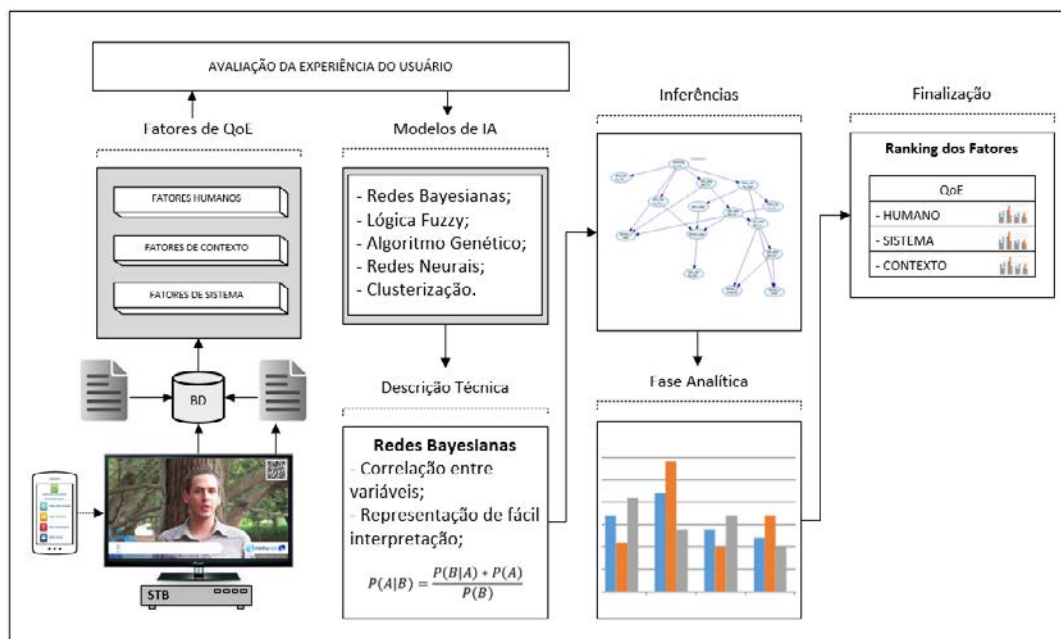


Figura 1. Arquitetura montada para testes e extração de conhecimento

Um programa televisivo foi exibido e o usuário podia interagir através de uma segunda tela (smartphone, por exemplo). Todo processo de interação dos usuários com o aplicativo DFapp foi monitorado e, ao final, foram aplicados instrumentos (questionários) avaliativos, para que se pudesse obter uma base de dados contendo o perfil do usuário e seus dados de interação, representado, assim, a experiência do usuário. A partir desta etapa, foram aplicados diferentes modelos de inteligência computacional, visando estabelecer qual mais se adequava ao objetivo proposto neste estudo e optou-se pela Rede Bayesiana como técnica de inteligência computacional a ser utilizada para a predição de resultados.

Tendo por base a modelagem das redes, inferências foram realizadas para obtenção do grau de correlação entre as variáveis (fatores) presentes, permitindo a visualização das tabelas de probabilidades de cada correlação observada. Por fim, foi feita uma análise para identificar o quanto determinado fator exerce de influência na experiência de uso do aplicativo de segunda tela, elencando os fatores envolvidos no sistema proposto.

4.2. Aplicativo Utilizado

O aplicativo Dengue Fever App (DFapp) foi desenvolvido por pesquisadores do Laboratório de Desenvolvimento de Sistemas (LADES), da Faculdade de Computação (FACOMP), do Campus de Castanhal da UFPA. Voltado para área de saúde (*t-health*), o DFapp permite levar informações sobre a Dengue, uma doença tropical com índices crescentes de ocorrência em algumas regiões do Brasil.

Na Figura 2 estão ilustradas: a tela inicial do aplicativo após a sincronização; (a tela com o vídeo que está sendo exibido no smartphone (segunda tela), de forma sin-



Figura 2. Telas do aplicativo de segunda tela – DFApp

cronizada ao que está sendo transmitido na televisão (primeira tela); e uma tela com conteúdo extra que fornece aos usuários mais informações sobre a Dengue. As outras opções do aplicativo: um jogo interativo em que o usuário tem que identificar e eliminar os focos da doença, tendo o tempo para esta ação cronometrado e exibido ao final da eliminação de todos os focos; e a possibilidade de o usuário acessar o Facebook e compartilhar informações sobre o aplicativo e sobre a doença [Seruffo et al. 2015], respectivamente.

A temática do aplicativo desenvolvido é da área da saúde que possui uma função social de levar informação a respeito da doença da Dengue à população. Caso o aplicativo desenvolvido fosse centrado em outra temática, os resultados apresentados poderiam ter sido diferentes. Isso significa dizer que analisar o conteúdo de uma aplicação é tão fundamental quanto os outros fatores estudados.

4.3. Definição do Cenário e Protocolo de Testes

O cenário montado para os testes foi composto por um dispositivo móvel e um aparelho de TV, ambos acessando uma rede sem fio por meio de um Ponto de Acesso conectado à Internet. Um servidor para representar um provedor de TV, responsável pela transmissão do vídeo online também fez parte do cenário. Os testes em laboratório foram realizados em ambiente controlado para facilitar a captura de informações. Durante um período de uma semana, se contou com a presença de 73 usuários, quantidade esta baseada em literaturas como [Norman 2004] e [Faulkner 2003]. É importante ressaltar que os testes foram concebidos a partir da perspectiva tanto qualitativa quanto quantitativa.

As idades variaram entre a faixa etária de 13 a 70 anos, sendo de diversas classes sociais e níveis de escolaridade, gerando assim uma base de dados consolidada e diversificada.

Visando sistematizar a bateria de testes, foi criado um protocolo que pudesse ser seguido pelos pesquisadores envolvidos. O processo de captura das métricas subjetivas, concernentes a cada usuário, tal como proposto em [ITU-T 2007], [Geerts et al. 2010] e [Menkovski et al. 2011]. Primeiro o usuário era submetido a um questionário de dados pessoais², o qual deveria ser preenchido de forma orientada, visando traçar um perfil socioeconômico. Este questionário foi armazenado com as informações preliminares de cada participante.

Após, o usuário era encaminhado à outra sala, chamada de sala de interação, na qual recebia uma explicação básica sobre o teste e quais os objetivos do projeto de pes-

²Questionário 01 – Disponível em <http://www.facompcastanhall.ufpa.br/lades/pesquisa/quest-dfapp.pdf>

quisa. A ideia foi ambientar os usuários antes de iniciar a utilização do aplicativo, contudo, não foi ensinado o uso da aplicação, deixando o usuário livre para experiência.

A partir daí a programação televisiva iniciava, sendo que a tela inicial de apresentação do aplicativo na TV exibia as seguintes mensagens: “A programação a seguir possui interatividade com segunda tela” e “Sincronize seu smartphone pelo QR Code que estará disponível”. Na tela subsequente, o usuário era convidado a sincronizar o smartphone com a TV pela leitura do QR Code através da seguinte mensagem: “Para continuar, sincronize com a TV utilizando o QR Code do smartphone”. Assim que o aplicativo era iniciado, o usuário poderia utilizar, de forma livre, as tecnologias disponíveis para os testes, supervisionado por uma pessoa da equipe do Laboratório. Estas mensagens foram inseridas na segunda versão do aplicativo, por ser constatado que mesmo com as explicações, poucos usuários entendiam a necessidade de sincronismo.

Ao final do teste de interação, o usuário era encaminhado para outra sala, para realizar uma entrevista em profundidade, norteada por perguntas contidas em um formulário previamente preparado³. O objetivo foi obter informações sobre: o experimento exercido pelo usuário; a facilidade da interação; a compreensão dos elementos das telas da aplicação nos diferentes dispositivos, TV e dispositivo móvel; o conhecimento prévio sobre aplicativos interativos; tempo de uso diário da TV e do serviço de Internet; e quais dispositivos mais usados junto com a TV, dentre outras questões, gerando uma base de dados sólida para o estudo⁴.

4.4. Métricas Utilizadas

Como métricas objetivas fez-se uso de 14 dados obtidos diretamente da utilização da aplicação (Tabela 1), tais como o tempo de interação do usuário no jogo, a quantidade de erros cometidos durante o uso da aplicação e durante o jogo mais especificamente, dentre outros, uma vez que a aplicação registra todos os dados das atividades realizadas pelo usuário. A definição dessas métricas foi baseada nos trabalhos correlatos apresentados na seção anterior.

Tabela 1. Variáveis do sistema

<i>erro_icone_jogo</i>	<i>qtd_selecao_video</i>
<i>erro_jogo</i>	<i>tempo_conteudo</i>
<i>erro_icone_social</i>	<i>tempo_jogo</i>
<i>erro_icone_video</i>	<i>tempo_social</i>
<i>qtd_selecao_conteudo</i>	<i>tempo_video</i>
<i>qtd_selecao_jogo</i>	<i>qtd_vezes_jogadas</i>
<i>qtd_selecao_social</i>	<i>erro_icone_conteudo</i>

Ainda sobre a seção de testes, foi solicitado a cada usuário o preenchimento de outro questionário⁵ com 12 perguntas (variáveis) referentes à interface do aplicativo e sua experiência de uso. Cada pergunta continha uma escala com cinco alternativas possíveis (1 - Péssimo/Muito Difícil, 2 - Ruim/Difícil, 3 - Razoável/Mediano, 4 - Muito Bom/Muito Fácil e 5 - Excelente/Extremamente Fácil). A soma destas 12 variáveis com as 13 pro-

³Questionário 02 – disponível em: <http://www.facompcastanhall.ufpa.br/lades/pesquisa/quest-dfapp.pdf>

⁴Base de dados – disponível em: <http://www.facompcastanhall.ufpa.br/lades/pesquisa/dados-dfapp.xlsx>

⁵Questionário 03 – disponível em: <http://www.facompcastanhall.ufpa.br/lades/pesquisa/quest-dfapp.pdf>

venientes do primeiro questionário, totalizaram as 25 métricas subjetivas (Tabela 2) que fizeram parte da modelagem das redes.

Tabela 2. Variáveis extraídas por questionários

<i>gênero</i>	<i>uso_paralelo_tv</i>
<i>conteudo_assistido</i>	<i>experiencia_uso_app</i>
<i>idade</i>	<i>interação_com_tv</i>
<i>escolaridade</i>	<i>interface_app</i>
<i>renda</i>	<i>memorizacao_interface</i>
<i>moram_com_vc</i>	<i>organizacao_infos_tv</i>
<i>tempo_uso_tv</i>	<i>organizacao_infos_celular</i>
<i>tipo_tv</i>	<i>comecar_app</i>
<i>costume_assistir_tv</i>	<i>sincronizacao_app_tv</i>
<i>uso_anterior_app</i>	<i>execucao_tarefas</i>
<i>tecnologia_acesso</i>	<i>tempo_adaptacao_app</i>
<i>uso_diario_tv</i>	<i>funcionalidade_app</i>
<i>uso_conjunto_tv</i>	

A partir da definição das métricas, visando aplicar o modelo de RB para identificar qual fator é mais significativo na experiência do usuário ao interagir com o aplicativo, as 39 métricas foram classificadas, de acordo com a literatura [ITU-T 2007] e [Callet et al. 2012] em fator humano, fator de contexto e fator de sistema (Tabela 3).

Tabela 3. Variáveis usadas para fatores humanos, contexto e sistema

Fator	Variáveis
Fator Humano	Genero, faixa etária, experiência prévia, coletividade, conteúdo assistido, uso anterior do app, experiência de uso do app, interação com a tv, interface do app, memorização da interface do app e tempo de adaptação do app.
Fator de Contexto	Tempo do vídeo, relações interpessoais, erros cometidos, uso paralelo, conteúdo assistido, renda, moram com você, tempo de uso da TV.
Fator de Sistema	Tipo da tv, tecnologia de acesso, quantidade de seleção do vídeo, quantidade de seleção jogo, quantidade de seleção do conteúdo, quantidade de seleção do social, erro no ícone do vídeo, erro no ícone do jogo, erro no ícone do conteúdo, erro no ícone do social, tempo no vídeo, tempo no jogo, tempo no conteúdo e tempo no social.

4.5. Resultados e Discussão

Com as inferências realizadas por meio das RBs montadas e aplicando o processo de entropia condicional nas variáveis, pôde-se perceber, conforme Tabela 4, que o fator de Contexto exerce maior influência na experiência do usuário, seguido pelo Fator Humano e Fator de Sistema.

Tabela 4. Influência dos fatores no DFapp

Fator	Influência
Contexto	40%
Humano	32%
Sistema	28%
TOTAL	100%

De forma mais específica, a variação entre as probabilidades de ocorrências das alternativas do conjunto de variáveis no fator de contexto foi maior e mais significativa

do que nos outros fatores. Por exemplo, em fator de contexto foram feitas 3 análises: erros cometidos, uso paralelo a tv e o costume de se assistir TV em conjunto, e dentre as três análises a que se mostrou mais impactante no contexto foi o uso paralelo da tv, pois de acordo com sua entropia condicional, influencia em 90% no contexto, ou seja, essa é a prática que mais influência na experiência do usuário, considerando apenas o contexto que o mesmo está inserido.

Ainda sobre o fator de contexto, usuários que não possuem acesso à Internet em suas residências, tendem a investir mais atenção no conteúdo transmitido na primeira tela (TV). Isto pode demonstrar quanto o processo de exclusão digital afeta a QoE, já que pessoas que não possuem acesso à Internet, ficam à margem do processo de apropriação de novas ferramentas que começam a se consolidar no mercado. Contudo, a partir do barateamento dos equipamentos eletrônicos, tais como smartphones, e a ampliação das tecnologias de acesso 3G e 4G, este cenário tende a sofrer alterações, gerando mais espaço para aplicações desta natureza.

Após análise das relações entre as variáveis, verificou-se que há uma tendência da assistência ao conteúdo televisivo de forma não coletiva. Entretanto, os usuários que assistem TV sozinhos passam menos tempo nesse tipo de atividade. Segundo [Zhu et al. 2015], “a co-visualização de vídeos com os amigos aumenta o nível de satisfação do usuário e aumenta a durabilidade da experiência”. Desta forma, os resultados alcançados neste trabalho, assim como os da literatura, apontam para uma maior satisfação por parte dos usuários quando assistem TV coletivamente.

Sobre os fatores de sistema, chama a atenção o fato de que os usuários que não possuem acesso à Internet em casa, passaram mais tempo na tela de conteúdo extra e cometem mais erros ao usar o aplicativo, este fato pode ser influenciado pois as pessoas ao interagirem com a aplicação, sentiram necessidade de usar a segunda tela para que obtivessem mais informações sobre o assunto, além disso, o elevado número de erros é ligado a falta de costume no uso de aplicações deste tipo, já que não há acesso à Internet. Por outro lado, pessoas com uma Internet de boa velocidade tendem a usar com frequência aplicações durante a assistência da TV, desta forma, essa experiência anterior refletiu nos testes, que apresentaram menos incidência de erros para esta categoria.

O sincronismo do vídeo da primeira com a segunda tela foi o recurso preferido dos usuários, que investiram mais tempo observando este funcionamento, isto é facilmente justificável pois trata-se de uma função ainda não disponibilizadas/disponíveis pelas emissoras de televisão do Brasil, portanto, gerando curiosidade do seu funcionamento pelos usuários, que ficaram surpresos sobre esta possibilidade de mobilidade.

De forma geral, pode-se extrair informações interessantes, como: qual gênero utiliza mais aplicativos e TV diariamente; o quanto as redes sociais influenciam e toma atenção dos indivíduos; a dificuldade que estudantes que não possuem Internet em suas casas têm ao utilizar aplicativos de segunda tela; o quanto utilizar diversos dispositivos ao assistir TV melhora o poder de percepção dos indivíduos, etc. Com as informações aqui apresentadas, o processo de análise do comportamento do usuário se torna mais fácil, já que o prisma de correlação entre as variáveis se mostrou mais claro e quantificável.

Outra observação importante que pôde ser realizada é o fato das mulheres que utilizaram o aplicativo erraram menos que os homens. Ou seja, tiveram mais facilidade

no manuseio do aplicativo. Essa facilidade essa encontrada pode ser embasada pelo que já foi exposto por [Charlton 2010], onde são apresentados dados que concluem a facilidade das mulheres para realização de determinadas atividades.

Dessa forma, ressalta-se que o fator que impacta mais intensamente na experiência do usuário é o de contexto. Sendo assim, o mundo ao redor do telespectador corrobora para a experiência de utilização que o mesmo tem. A carga de conhecimento que o usuário carrega, as suas relações interpessoais e o seu estado emocional em relação a determinado acontecimento são fundamentais neste processo de avaliação, conforme mostrado nos trabalhos correlatos. Considerar esses aspectos é de suma importância para os prestadores de serviços que tem como objetivo estarem sempre atrelados ao seu maior colaborador de *feedbacks*, que é o usuário final.

Salienta-se que estudos como este são fundamentais para a padronização de aplicativos que em sua fase de construção tendem atingir um público em específico. Vale ressaltar que o modelo aqui proposto não se restringe apenas a aplicativos de segunda tela, mas sim para qualquer tipo de aplicativo, já que esses podem gerar as mesmas informações de utilização que o DFapp gerou, contudo, para o estudo de caso apresentado, foram restritos os testes para segunda tela. Desta forma, o modelo é genérico e possui flexibilidade.

5. Conclusão

A evolução intrínseca da humanidade vem trazendo novas formas de se realizar tarefas, buscando sempre a comodidade de quem as realizam. A tecnologia vem para fortalecer essa evolução e proporciona velocidade importante nesse processo e atividades que antes eram simples e rotineiras hoje vem ganhando recursos que elevam a curiosidade e satisfação de uso.

Os modelos de avaliação da qualidade de experiência do usuário vêm tomando níveis altos de precisão e confiança. Com a utilização de técnicas computacionais que auxiliam nessa predição comportamental, tornam-se muito eficientes e precisos nas suas análises. Estudos recentes têm mostrado que a abordagem de considerar apenas parâmetros multimídia não pode suficientemente estimar a satisfação do usuário e que QoE depende de múltiplos fatores, além das propriedades técnicas de mídia. A contribuição deste trabalho está no ponto em que mescla fatores que antes não coexistiam, como, o quanto a percepção do usuário em relação à interface da aplicação é afetada dada sua condição financeira.

Esse trabalho contribui com a proposição de uma metodologia para avaliação da experiência do usuário para aplicativos utilizando como estudo de caso um aplicativo de segunda tela, mas não se restringe a avaliar apenas aplicativos desse tipo, traz um ganho para os desenvolvedores de aplicativos no que se refere a análise mercadológica de seus produtos e serviços, onde a preocupação que antes se direcionava apenas para qual tecnologia utilizar, hoje se mostra muito mais latente analisar o público que fará uso do produto final. Saber seus gostos, localização, bagagem de conhecimento e outros aspectos ligados ao contexto que esse usuário está inserido, como este estudo mostra, precisa de uma parcela maior de atenção na hora de se desenvolver um aplicativo.

O contexto que o usuário está inserido, bem como seu comportamento, traz informações importantíssimas para a definição do seu perfil como utilizador de certo sis-

tema. Obter essa informação de forma clara e concisa traz benefícios para quem pretende lançar um aplicativo ou até mesmo melhorar um já existente. Pensando nisso, deve-se buscar formas para avaliar essa experiência que o usuário tem ao utilizar aplicativos e este trabalho propõe um novo método de obtenção dessas informações utilizando RB.

Se conclui que para uma análise mais fidedigna ao contexto geral de uso de um aplicativo, é crucial considerar fatores tanto ligados ao sistema (tecnologia de acesso, qualidade da transmissão de vídeo, etc.) e também fatores ligados ao ser humano que faz uso desse sistema (idade, genero, experiência prévia, etc).

Como trabalhos futuros pretende-se utilizar outros métodos de inteligência computacional para comparar os resultados obtidos, visando consolidar ainda mais a ideia proposta e mostrar um novo rumo que as fórmulas de avaliação podem tomar, considerando sempre os diversos fatores que influenciam na observação de um aplicativo ou serviço.

Agradecimentos

Os autores agradecem à Universidade Federal do Pará (UFPA). E também à Faculdade de Computação (FACOMP) e ao Laboratório de Desenvolvimento de Sistemas (LADES), ambos do campus de Castanhal, pelo apoio no desenvolvimento deste trabalho.

Referências

- Callet, P. L., Möller, S., Perkis, A., et al. (2012). Qualinet white paper on definitions of quality of experience. *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*.
- Charlton, G. (2010). Why are women so intelligent? the effect of maternal iq on childhood mortality may be a relevant evolutionary factor. *Medical hypotheses*, 74(3):401–402.
- Copa (2014). Copa do mundo lidera estatísticas de 2014 do twitter e do facebook. <http://www.copa2014.gov.br/pt-br/noticia/copa-do-mundo-lidera-estatisticas-de-2014-do-twitter-e-do-facebook>. Acesso em: 10 de abril 2015.
- Da, M., Wei, W., Hu, H., and Guan, J. (2011). The application of bayesian classification theories in distance education system. *International Journal of Modern Education and Computer Science*, 3(4):9.
- ETSI (2009). Human factors (hf); quality of experience (qoe) requirements for real-time communication services. Technical Report TR 102 633.
- Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35(3):379–383.
- Ferrari, F. (2012). O estudo da entropia de transferência em sistemas dinâmicos discretos. Master's thesis, Mestrado em Ciências, Universidade Estadual de Ponta Grossa, Paraná.
- Geerts, D., Katrien, M., Istvan, K., An, J., Jan, B., Wout, J., Luc, M., and Lieven, M. (2010). Linking an integrated framework with appropriate methods for measuring qoe. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 158–163. IEEE.

- IpsosMediact (2014). Facebook reach and usage. Estudo comissionado pelo Facebook.
- ITU-T (2007). Definition of quality of experience. Technical Report SG12 109rev2, Geneva, Switzerland.
- Menkovski, V., Exarchakos, G., Liotta, A., and Antonio, S. (2011). Managing quality of experience on a commercial mobile tv platform. *International Journal on Advances in Telecommunications Volume 4, Number 1 & 2, 2011*.
- Norman, D. (2004). *Why we love (or hate) everyday things*. Emotional Design.
- Piccolo, L., Bonadia, G., and Tambascia, C. (2010). Aplicação de técnicas e métodos de ihc para avaliar a qoe de serviços de tv interativa. In *Proceedings of the IX Symposium on Human Factors in Computing Systems*, pages 225–230. Brazilian Computer Society.
- Sarabando, A. (2011). Um estudo do comportamento de redes bayesianas no prognóstico da sobrevivência no cancro da próstata. Master's thesis, Faculdade de Ciências e Medicina, Universidade de Porto, Portugal.
- Seruffo, M., Miranda, F., Oliveira, E., and Malcher, A. (2015). Dengue fever: Prevention through information. *TVX 2015*.
- Soskic, N. (2014). Second screen user experience: A new digital television frontier. *Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1057–1060.
- Xenos, M. (2004). Prediction and assessment of student behaviour in open and distance education in computers using bayesian networks. *Computers & Education*, 43(4):345–359.
- Yong, L. (2013). User experience evaluation methods for mobile devices. In *Innovative Computing Technology (INTECH), 2013 Third International Conference on*, pages 281–286. IEEE.
- Zhou, T. (2013). The effect of flow experience on user adoption of mobile tv. *Behaviour & Information Technology*, 32(3):263–272.
- Zhu, Y., Heynderickx, I., and Redi, J. (2015). Understanding the role of social context and user factors in video quality of experience. *Computers in Human Behavior*, 49:412–426.

Um framework para coprojeto de hardware e software de sistemas ADAS baseados em visão

Leandro A. Martinez¹, Tiago Lobo¹, Eduardo Marques¹

¹ Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo
Av. Trabalhador São-Carlense, 400 – CEP 13566-590 São Carlos, SP, Brasil

lmartinez@usp.br, tiagomelobo@gmail.com, emarques@usp.br

***Abstract.** This paper presents a framework for hardware and software co-design to building systems designed to driver assistant using computer vision. This work is part of a doctoral research project nearing completion. To validate the model, a modular pedestrian detection is implemented by comparing the results obtained with other design.*

***Resumo.** Este artigo apresenta um framework para co-projeto de hardware e software para a construção de sistemas destinados ao apoio de motoristas utilizando visão computacional. Este trabalho faz parte de um projeto de pesquisa de doutorado em fase de conclusão. Para sua validação, um sistema modular de detecção de pedestres é implementado comparando-se os resultados obtidos com outro projeto.*

1. Introdução

Evitar acidentes de trânsito ainda é um dos grandes problemas que o mundo enfrenta. Com as novas tecnologias de sistemas embarcados surgem diversas propostas que precisam ser exaustivamente validadas e testadas. Contudo, simular e testar novas propostas nos diversos possíveis ambientes unindo algoritmos de software com hardware reprogramável torna o problema complexo e multidisciplinar dentro da computação. Assim, com objetivo de contribuir nesta área de pesquisa, este artigo apresenta um framework para co-projeto de hardware e software para a construção de sistemas destinados ao apoio de motoristas que utilizam visão computacional.

Os processos e componentes utilizados para construir sistemas avançados para auxílio a motoristas (ADAS - Advanced Driver Assistance Systems) devem ser seguros e confiáveis. Validação e testes exaustivos acarretam em um enorme investimento tornando-se um dos aspectos mais desafiadores do desenvolvimento de ADAS, especialmente quando isso acontece com sistemas de visão. Para testar todos os cenários e produzir precisão 100%, com zero falsos positivos, e em todas as condições possíveis, necessita-se de uma grande quantidade de horas de cliques de vídeo para que possam ser reunidos e executados em banco de dados para teste de regressão (Kisačanin e Gelautz, 2014) (Bengler et al. 2014).

Nesse contexto, o framework apresentado, visa facilitar o desenvolvimento de aplicações permitindo melhor explorar o espaço de projeto, e assim contribuir para um ganho de desempenho no desenvolvimento de sistemas embarcados principalmente quando comparados à construção totalmente em hardware. Um dos recursos do sistema é a possibilidade da simulação da aplicação antes da síntese em uma plataforma reconfigurável.

Nota-se que a taxa de reuso em circuitos lógicos na indústria deverá crescer linearmente. Em 2011 essa taxa era de 54% e deverá chegar a 98% em 2026. Com estes pressupostos, mantendo constante esforço de projeto em SOC faz-se necessário uma melhoria da produtividade em projetos em dez vezes longo dos próximos dez anos. Para resolver o desafio da produtividade, várias abordagens devem ser combinadas. Em primeiro lugar, os níveis de abstração de projetos devem ser elevados. Em segundo lugar, deve ser aumentado o grau de automação, particularmente na verificação e implementação de projetos (ITRS, 2011)

2. Frameworks de Visão Computacional

Existem diversos frameworks para processamento de imagem usados em sistemas de visão automotiva e estão em constante evolução, no entanto, muitos destes projetos encontram-se com alguma desvantagem: alguns foram descontinuados, outros são inadequados para processamento de visão em tempo real com desenvolvimento rápido, muitos carecem de uma interface amigável com o usuário, e por fim os que não possuem design modular. A seguir, uma relação de alguns importantes sistemas:

O Baselabs fornece a infraestrutura para o desenvolvimento de ADAS de prototipagem rápida para desenvolver ADAS permitindo a captura, registo e reprodução de dados do sensor e da fusão de dados de sensores complexos. Disponibiliza uma interface gráfica do usuário com vários componentes (Baselabs, 2016).

Intempora RTMaps é um framework para prototipagem algoritmos com um ambiente modular para testar e avaliar funções com base em diferentes conjuntos de sensores e em diferentes configurações. Também fornece uma biblioteca de componentes preparados para desenvolver ADAS. É extensível permitindo a personalização de módulos pelo usuário (Intempora, 2016).

Comemso ADTF é também um framework para sistemas de processamento de imagem usado na indústria automobilística. Permite a criação de filtros de imagem e conexões de interfaces de hardware para sistemas embarcados. Também fornece ferramentas para teste e simulação (Comemso, 2016).

MontiVision fornece um conjunto de ferramentas que permite o desenvolvimento de aplicações de processamento de imagem de forma personalizada. O kit de desenvolvimento inclui vários módulos de processamento de imagem. Os filtros de imagens podem ser definidos e personalizados pelo usuário através de uma interface visual. Também inclui um conjunto exemplos para demonstrar o uso (MontiVison, 2016).

ImprovCV oferece um sistema modular de código aberto para processamento de fluxo de dados de visão. O software permite interatividade rápida com o usuário para o desenvolvimento de aplicações de visão ADAS (Boeing e Braunl, 2008).

NI LabVIEW é o software base da plataforma de projeto da National Instruments, para o desenvolvimento de sistemas de medição ou controle. Integrando diversas ferramentas e com um ambiente de desenvolvimento voltado à resolução de problemas, produtividade acelerada e inovação contínua (LabVIEW, 2016).

MATLAB HDL Coder e HDL Verifier fornece um ambiente de projeto integrado que acelera o desenvolvimento em FPGA e ASIC, integrando ferramentas de design e IP (Intellectual Property) para Xilinx e Altera. Este ambiente fornece um sistema de programação extensível, bem como uma arquitetura dataflow (MATLAB, 2016).

Infelizmente, a maioria desses sistemas não são de código aberto e poucos fazem a integração entre hardware e software. Dessa forma, a proposta deste trabalho é apresentar um ambiente de desenvolvimento que permita consolidar as principais vantagens desses sistemas. Destacando-se:

- Código fonte aberto para permitir a expansão do projeto;
- Simulação funcional da integração de IPs software com IPs hardware no mesmo ambiente;
- Opção para compilar o projeto criado pelo usuário para FPGA.

3. Características do sistema

Utilizando-se como plataforma base um software de código aberto (Blender, 2016) modificado com diversas implementações, o Framework Vision-ADAS apresenta ao usuário uma interface de fluxo de dados de fácil utilização (Figura 1). Usando nós de composição é possível fazer várias combinações de filtros em fluxos de vídeo ou imagens.

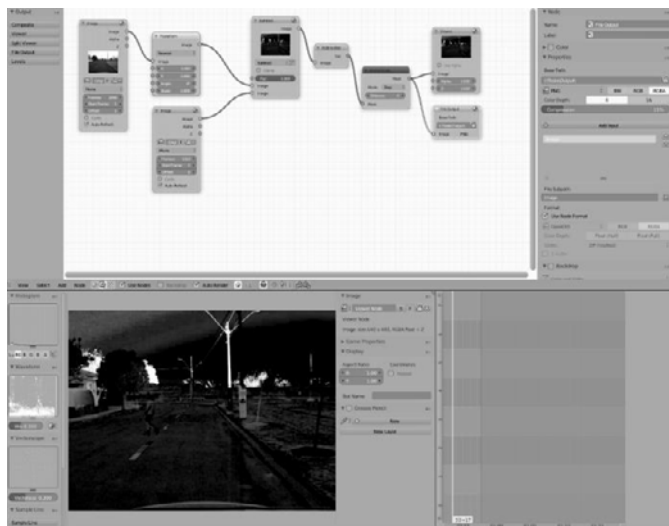


Figura 1. Tela principal da ferramenta de desenvolvimento

3.1. Ambiente de Projetos

A plataforma oferece muitos recursos para processamento de imagens. No conjunto, disponibiliza-se uma biblioteca própria de blocos que podem ser compilados pela

ferramenta para utilização em FPGA. Permite-se também utilizar filtros ainda não disponíveis para versões embarcadas de hardware e software, possibilitando simulações e entrada de dados para os outros blocos já disponíveis para embarcados.

Através de blocos IP (propriedade intelectual) o usuário pode arrastar e soltar filtros de processamento de imagem para a interface do usuário e exibir o seu efeito diretamente sobre o vídeo selecionado. Cada filtro tem parâmetros que podem ser ajustados em tempo real para fornecer feedback imediato. Permitindo testes interativos completos de sistemas de visão e simplificando ajuste de parâmetros.

3.2. Simulação

O nó de entrada de vídeo pode ser conectado permitindo comparações diretas de vários algoritmos de visão computacional. Uma vez que o aplicativo é um software de código aberto pode ser estendido ou adaptado para interoperar com outros sistemas.

Há três tipos principais de blocos:

- IP apenas para a simulação: Todos os nós podem ser utilizados para as simulações, contudo, alguns não estão disponíveis para serem embarcados diretamente em FPGA. Normalmente são codificados para ensaios e posteriormente convertidos. Como exemplo, podem ser disponibilizados bibliotecas de filtros provenientes do OpenCV (OpenCV, 2016), que podem não estar disponíveis para a compilação final. A grande vantagem deste tipo de bloco é a possibilidade de utilizar outras aplicações baseadas em C como protótipos para validar os resultados esperados.
- IP pronto para ser incorporado no processador (SoC): Os nós podem ter códigos pré-programados para serem usados em um processador embarcado. Durante a compilação, sub-rotinas de execução são inclusas com o código para ser usado pelo processador embarcado.
- IP pronto para ser sintetizado em hardware: São os blocos preparados para geração de hardware. Neste caso, o sistema utiliza códigos pré-compilados de um IP escritos em HDL (Hardware Description Language) para simular parte da aplicação. Na fase de compilação do projeto, um código de script é criado para geração do hardware, contendo os blocos e os controles de execução.

Neste ambiente, a criação de um novo IP pode ser validada por uma equipe de software, em seguida, convertido em uma versão de hardware ou até mesmo um software embarcado. Dessa forma, um bloco de software ou um bloco de hardware pode usar suas entradas e saídas para contribuir na construção e validação de seu complemento. Este método favorece uma melhor cooperação entre as equipes de desenvolvimento.

3.3. Arquitetura Geral do Framework

O objetivo final deste framework é o envio do projeto criado pelo usuário em uma arquitetura de hardware. Neste caso, após a validação em simulação funcional, existem algumas etapas a serem seguidas. A Figura 2 mostra um fluxo completo usando a estrutura para uma validação completa em FPGA.

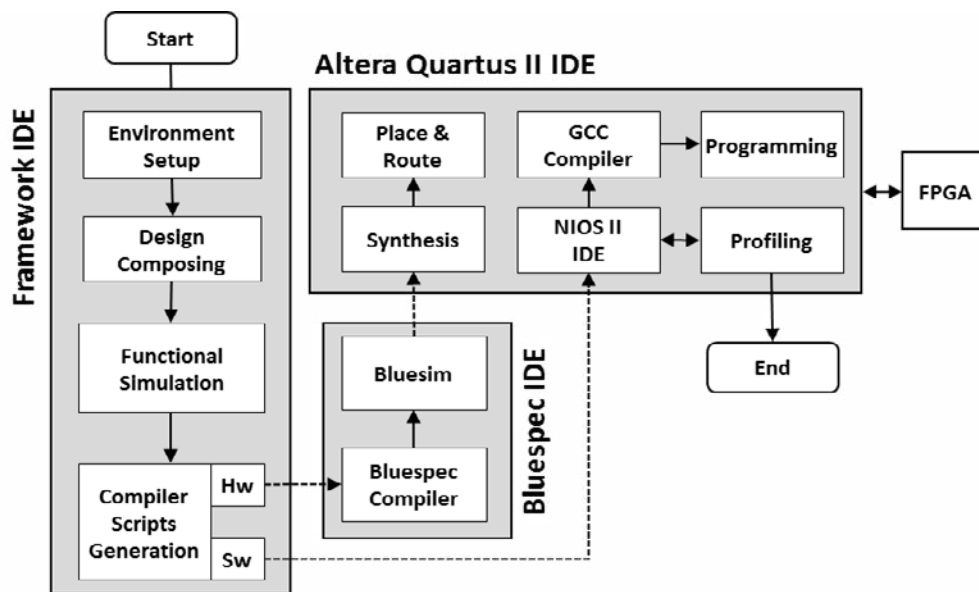


Figura 2. Fluxo completo usando o framework

Configuração do Ambiente: Ao iniciar a ferramenta, o primeiro passo é selecionar a placa FPGA a qual se destina o projeto. Em seguida, o usuário pode iniciar um projeto vazio ou selecionar um modelo base desenvolvido para permitir alterações no projeto. Os modelos são projetos validados que contêm exemplos de uso da ferramenta para vários ADAS.

Composição do Projeto: O sistema fornece uma biblioteca de IPs onde o usuário pode arrastar e soltar para o espaço de trabalho e ligá-los a outros blocos. O IP pode ser hardware ou software, dependendo da escolha do usuário e disponibilidade na biblioteca. Cada IP tem um conjunto de parâmetros que podem ser modificados durante a manutenção do projeto.

Simulação Funcional: Durante a elaboração do projeto, pode-se fazer uma simulação funcional do design. O sistema usa rotinas previamente compiladas de IPs para fazer a visualização de imagens processadas. O usuário também pode escolher vídeos com imagens sequenciais para comparação dos resultados.

Compilação para FPGA: No processo de compilação, o sistema verifica se todos os blocos utilizados na concepção permitem o uso embarcado. Neste caso, para IPs de hardware, o sistema gera scripts de código-fonte em linguagem Bluespec (Bluespec, 2016). Para os IPs de software, o compilador cria scripts código fonte para NIOS II. Esses scripts têm um formato padrão para a comunicação com blocos internos de controle de execução.

Enviar a Hardware: Para uma geração completa de hardware, o usuário deve compilar o projeto gerado pela ferramenta no ambiente de desenvolvimento Bluespec. Caso necessário, outras simulações de hardware utilizando-se a ferramenta BlueSim podem ser executadas. O Bluespec gera arquivos em formato Verilog em um projeto padrão. Este projeto poderá ser sintetizado pelo usuário na ferramenta Quartus II (Altera, 2016) onde poderá ser enviada para placa FPGA. Para concluir a implementação, deve-se compilar o projeto gerado dentro da ferramenta NIOS II (NIOS II, 2016) e enviar o software criado para o processador embarcado no FPGA.

3.4. Construção de IP

Para estender a ferramenta com o novos IP que possam ser embarcados, é necessário o desenvolvimento de estruturas de código em um formato padrão. Para garantir que os resultados obtidos na simulação sejam iguais aos resultados práticos, são exigidos alguns cuidados.

Para a validação de IP de software (Figura 3), deve-se usar um padrão pré-estabelecido de construção. O arquivo principal "IP.c" contém o algoritmo proposto e deve ser compatível com as bibliotecas utilizadas pelo NIOS II. Recomenda-se executar o script de teste "TestBench.c", isto permite-lhe ler as entradas de dados de um arquivo de origem "iStream.bin" para ser usado no teste de algoritmo e salvar a saída para um arquivo de verificação "oStream.out". Finalmente, com o sistema em execução no ambiente embarcado, um profiler pode ser feito para extrair várias características do algoritmo proposto.

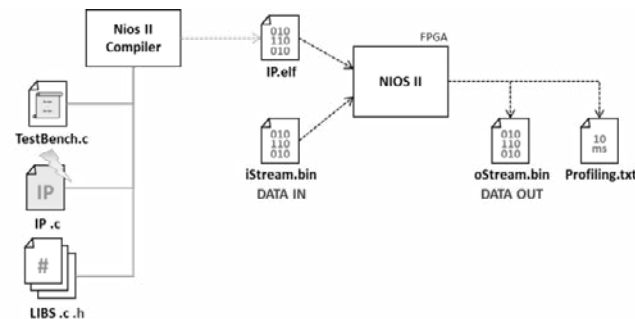


Figura 3. Validação de IP de Software

Para a validação de IP em hardware (Figura 4), o arquivo principal "IP.bsv" contém o algoritmo proposto em linguagem Bluespec. Caso necessário, a linguagem permite a reutilização de estruturas em Verilog usando um encapsulamento para essa finalidade. Uma vez compilado, gera-se um arquivo Verilog "IP.v" que pode ser adicionada a um template de validação no Quartus II. Através de técnicas de profiler, é possível extrair várias características do hardware gerado.

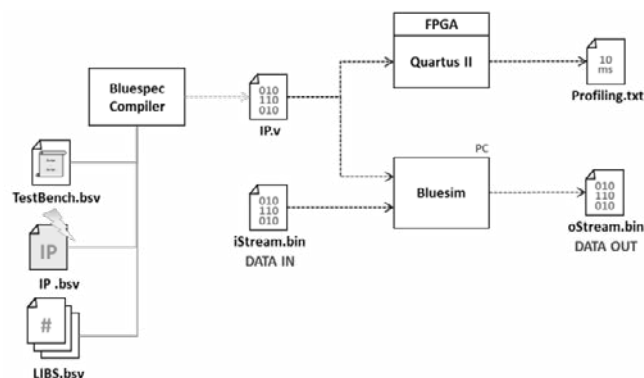


Figura 4. Validação de IP de Hardware

O BlueSim permite a validação do hardware gerado através de um programa escrito na própria linguagem "TestBench.bsv". Assim como os blocos validados em software, as entradas de dados podem ser lidas a partir de uma fonte de arquivo "iStream.bin" e a saída gerada em um arquivo de verificação "oStream.out". Em ambos

os casos, pode-se comparar os arquivos "fStream.bin" com a saída da implementação em software do PC.

4. Detecção de Pedestres

A proposta para a validação da ferramenta é a adaptação de um sistema de detecção de pedestres construído por Martinez (2011) (Figura 5) e baseado no trabalho de Jalalian (2008) e ao final, comparam-se os resultados finais com o trabalho original. Em seguida, modificar o sistema de estabilização de imagem para uma proposta de diferente autoria. No primeiro trabalho, o estabilizador é criado através de um vetor de valores pixels acumulados e o comparados com o mesmo vetor do quadro anterior. A segunda proposta consiste de uma implementação de fluxo óptico em hardware.

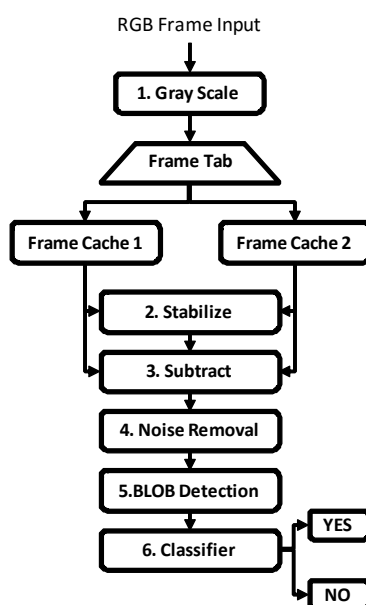


Figura 5. Modelo proposto para detecção de pedestres

4.1. Estabilização de Imagem com Vetor

Para comparar dois frames, as imagens devem estar alinhadas, dessa forma o IP de estabilização é essencial para o bom funcionamento do sistema. A detecção de objetos em movimento em um ambiente real é uma tarefa complexa pois, a imagem de fundo está mudando constantemente. Para isso, este bloco dimensiona o tamanho do vetor com a correspondente altura da imagem e armazena a soma dos valores de intensidade de pixel para cada linha (Figura 6). Para estabilizar, o vetor atual é comparado com o anterior trazendo como resultado o deslocamento do eixo Y. Utilizando-se apenas vetores, esta aplicação torna-se simples para ser construída em um sistema de hardware. Existem diversas abordagens para a estabilização de imagem, no entanto, muitas exigem grande quantidade de elementos lógicos para a sua construção.

Esta abordagem demonstrou ser eficiente, contudo foi somente implementada para identificar os deslocamentos verticais, de modo que o sistema não está preparado para detecções com o veículo em curva, permite apenas com o veículo movendo-se para frente.



Figura 6. Estabilização de Imagem com Vetor

4.2. Fluxo Óptico

Uma outra solução proposta para resolver o problema do alinhamento entre os quadros é a criação de vectores de movimento da imagem. Assim, para minimizar o efeito de trepidação da câmara a imagem é dividida igualmente em quatro regiões e calcula-se os quatro vectores que representam o vector geral de movimento. A principal desvantagem desta abordagem é o custo computacional quando comparado à estabilização por vetor. A fim de avaliar o sistema de estabilização de imagem, foi utilizado como uma base, o trabalho de Lobo (2013), que foi baseado na arquitectura de MIT (2011), o qual é construído um sistema de fluxo óptico (Figura 7).

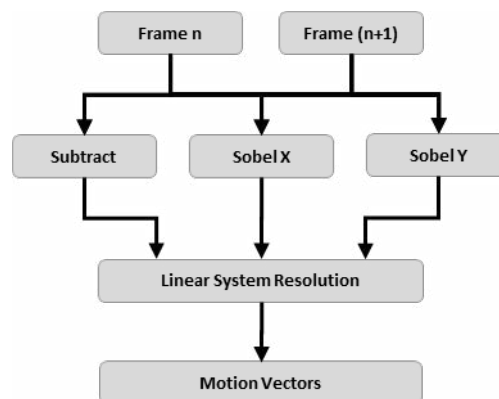


Figura 7. Fluxo Óptico Proposto

Usando convolução com filtro de Sobel, derivadas espaciais são calculadas. A derivada temporal é o resultado da diferença entre as dois frames. Com estes dados, um sistema de equação é resolvido utilizando múltiplos produtos por pixel e o resultado é soma da vizinhança de pixels (3 x 3), gerando o campo de fluxo óptico.

4.3. Remoção de Ruídos

Após a diferenciação de quadros de imagem subsequentes (Figura 8 esquerda) é possível detectar objectos que se moviam no processo de transição, contudo, o movimento irregular da câmara pode gerar ruído quando existe diferenciação. Para resolver este problema, um passo de pré-processamento é necessário para minimizar regiões de ruído. Assim, vários algoritmos de redução de ruído podem ser utilizados,

por exemplo, o filtro de média não linear, o que suaviza a imagem sem diminuir a resolução.

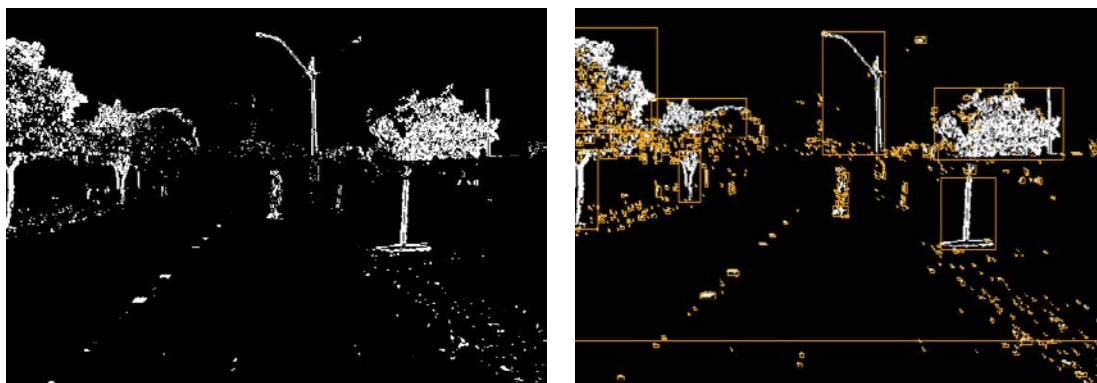


Figura 8. Resultado da diferença de dois frames (esquerda) e Componentes Conectados e Classificados (direita)

4.4. Identificação de componentes conectados e Classificador

O algoritmo de identificação de componentes conectados é usado para a detecção de objectos da cena. Este algoritmo é utilizado em visão computacional para detectar regiões conectadas em imagens digitais binárias.

Para classificação, adotou-se o uso de limitadores parametrizados dos objetos extraídos de componentes conectados. No projeto original, com o uso de uma placa de FPGA com recursos mais limitados, objetivou-se minimizar a utilização de recursos para atender ao requisito de execução em tempo real segundo a especificação de Gavrilá (2006).

4.5. Resultados da Detecção

Usando a mesma fonte de dados de imagem e com os algoritmos implementados segundo as especificações do projeto original, o resultado das detecções foram idênticas aos resultados do trabalho de base. Durante a síntese, foi necessário fazer alguns ajustamentos de temporização. Como esperado, o fluxo óptico para a estabilização de imagem tem boa vantagem, a Tabela 1 mostra a comparação entre estabilizadores de imagem:

Tabela 1. Detecção com diferentes estabilizadores

Estabilizador	Acertos	Falsos Positivos	Falsos Negativos
N/A	40%	20%	40%
Vetor	83%	9%	6%
Fluxo Optico	91%	2%	7%

A Figura 9 mostra um esboço do hardware gerado pelo framework.

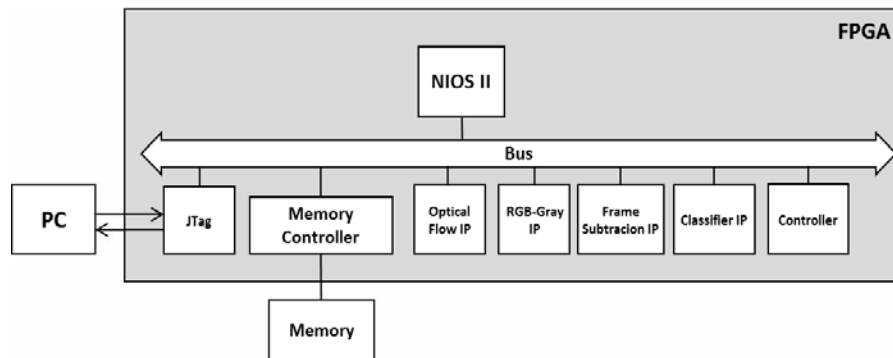


Figura 9. Esboço do hardware gerado

4.6. Plataforma de desenvolvimento do Hardware

A plataforma utilizada para a validação hardware gerado é a placa FPGA DE2i-150 da Terasic (Terasic, 2016). As principais especificações são: Dispositivo Cyclone IV EP4CGX150DF31; 149.760 Elementos Logicos; 720 M9K Blocos de Memória; 6.480 Kbits memória embarcada; 8 PLLs; 128MB (32Mx32bit) SDRAM; 4MB (1Mx32) SSRAM; 64MB (4Mx16) Flash modo 16-bit; Três osciladores de clock de 50MHz.

5. Conclusão

Este trabalho apresenta um framework para co-projeto de hardware e software para ADAS com visão computacional, sendo uma aplicação de código aberto baseada em componentes. As principais vantagens do aplicativo foram demonstradas através de um caso de uso para a detecção de pedestres. Apesar de ainda não ser uma ferramenta completa, o framework apresenta características promissoras através de validações de IPs e da execução de testes com diversos algoritmos. A partir dos dados apresentados na Tabela 1, nota-se que no esquema de detecção proposto, a estabilização de imagem usando o fluxo óptico é um fator crítico para melhores resultados. Observa-se também, que o vector para o algoritmo de estabilização, apesar da sua simplicidade, traz bons resultados em relação a resultados sem estabilização de imagem. Neste trabalho, utilizando-se de componentes baseados em outros projetos e com as mesmas amostras de fontes de dados foi possível obter os mesmos resultados originais, mostrando assim, que é possível migrar projetos externos para a ferramenta sem criar interferência nos resultados e permitindo a exploração do espaço de projetos.

Referências

- Altera (2016). Introduction to the Quartus II Software. web site. [Online] Acessado em abril 2016. http://www.altera.com/literature/manual/intro_to_quartus2.pdf
- Baselabs (2016). Baselabs Flexible ADAS Development web site. [Online] Acessado em abril 2016. <http://baselabs.de/flexible-adas-development-kopie.html>
- Bengler, K., Dietmayer, K., Färber, B., Maurer, M., Stiller, C., Winner, H. (2014). “Three Decades of Driver Assistance Systems: Review and Future Perspectives,” IEEE Intell. Transp. Syst. Mag., vol. 6, no. 4, pp. 6 – 22

- Blender (2016). Website. [Online] Acessado em abril 2016. <https://www.blender.org/>
- Bluespec (2016). Bluespec SystemVerilog User Guide, [Online]: <http://wiki.bluespec.com/Home/BSV-Documentation>
- Boeing, A.; Braunl, T. (2008) "ImprovCV: Open component based automotive vision," Intelligent Vehicles Symposium, 2008 IEEE , vol., no., p.297,302,4-6.
- Comemso (2016). Comemso web site. [Online] Acessado em abril 2016. <http://www.comemso.com/index.php/software>
- Gavrila D., Munder S. (2006). "Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle" International Journal of Computer Vision Springer Science.
- Intempora (2016). Intempora web site. [Online] Acessado em abril 2016. <http://www.intempora.com>
- ITRS (2011), International Technology Roadmap for semiconductors. [Online] Acessado em abril 2016. <http://www.itrs.net/Links/2011ITRS/2011Chapters/2011SysDrivers.pdf>
- Jalalian, A., Fathy, M. (2008). Pedestrian Detection from a Moving Camera with an Advanced Camera-Motion Estimator. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System.
- Kisačanin, B., Gelautz, M. (2014). Advances in Embedded Computer Vision, Springer.
- LabVIEW (2016). LabVIEW web site. [Online] Acessado em abril, 2016. <http://www.ni.com/labview>
- Lobo, T. M. (2009). Co-projeto hardware/software para cálculo de fluxo ótico. [Online] Acessado em abril 2016. <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-28082013-094816/en.php>
- Martinez, L. A. (2011). Projeto de um sistema embarcado de predição de colisão a pedestres baseado em computação reconfigurável. [Online] Acessado em abril 2016. <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-27022012-110356/pt-br.php>
- MIT (2011). MIT 6.375 web site. [Online] Acessado em abril 2016, http://csg.csail.mit.edu/6.375/6_375_2011_www/projects.html.
- NIOS II (2016). Altera NIOS II Hardware Tutorial web site. [Online] Acessado em abril 2016. http://www.altera.com/literature/tt/tt_nios2_hardware_tutorial.pdf
- MATLAB (2016). MATLAB HDL Coder web site. [Online] Acessado em abril 2016. <http://www.mathworks.com/fpga-design/hardware-software-codesign.html>
- MontiVision (2016). MontiVision Development Kit (SDK) web site. [Online] Acessado em abril 2016. <http://www.montivision.com>
- OpenCV (2016). web site. [Online] Acessado em abril 2016. <http://www.opencv.org>
- Stratix V (2013). Stratix V Development Board Reference Manual. [Online] Acessado em abril. http://www.altera.com/literature/manual/rm_svgx_fpga_dev_board.pdf
- Terasic (2016). Terasic DE2i-150 FGA Development Kit. web site. [Online] Acessado em abril 2016. <http://www.terasic.com.tw/cgi-bin/page/archive.pl?Language=English&CategoryNo=11&No=529&PartNo=1>

Qualidade de Experiência do Usuário em uma Rede WiFi Municipal usando *Streaming* de Vídeo

Juliano Ratusznei, Waleska Zanotto, Carlos Kamienski

Universidade Federal do ABC (UFABC)

{juliano.r, cak}@ufabc.edu.br, waleskacomk@gmail.com

Resumo. *Redes WiFi Municipais tem o objetivo de oferecer acesso gratuito à Internet através de rede sem fio, para incentivar a participação do cidadão e promover a inclusão digital. No entanto, essas infraestruturas em geral não possuem desempenho satisfatório, além de não estarem disponíveis resultados de avaliação de desempenho. A abordagem desse artigo é a realização de experimentos envolvendo a utilização de streaming de vídeo em praças digitais distribuídas pelas diversas regiões da cidade de São Paulo. Foi realizada uma análise de desempenho nas praças utilizando-se de métricas que avaliam tanto a Qualidade de Experiência do usuário, como a Qualidade de Serviço da rede. Os resultados mostram que para diferentes praças, dias e níveis de qualidades de vídeo, na maioria das vezes o resultado é satisfatório para os usuários.*

Abstract. *Municipal WiFi Networks aim at providing free Internet via wireless network infrastructures, for promoting citizen participation and bridging the digital divide. However, such infrastructures frequently suffer from poor performance, which is worsened by a general lack of results obtained in performance analysis studies. The approach of this paper is the use of video streaming in the so-called digital squares with free WiFi access, distributed throughout five regions of São Paulo City. A performance analysis study was conducted in squares using metrics such as Quality of Experience of users and Quality of Service of the network. Results show that for different squares, days and video qualities, in most cases the experience is satisfactory for users.*

1 Introdução

Devido a atual popularidade dos dispositivos móveis e sua crescente utilização em locais públicos, a disponibilização de Internet tem se tornado um serviço cada vez mais importante ao redor do mundo. No início dos anos 2000 acreditava-se que as cidades teriam redes WiFi Municipais com ampla cobertura de sinal para promover o acesso a Internet à população e a inclusão digital e incentivar a criação de serviços rumo às cidades inteligentes [Bar & Park 2006]. Principalmente nos Estados Unidos, onde a oferta de rede 3G chegou mais tarde, praticamente uma centena de cidades iniciaram projetos para oferecer WiFi gratuitamente à população, inclusive dentro de prédios residenciais. No final da década a maioria dos projetos já havia sido descontinuado ou então cancelado, devido a problemas regulatórios e técnicos, altos custos e eventualmente baixa demanda [Fraser 2010]. Atualmente existe um grande debate sobre as vantagens e desvantagens da redes WiFi Municipais [Potts 2014].

No entanto, várias iniciativas sobreviveram e inclusive muitas foram criadas a partir de 2010, justamente com foco em fornecer conectividade WiFi em espaços

públicos, tanto abertos como praças e parques, como fechados como bibliotecas e órgãos públicos. Nesse contexto se insere o programa WiFi Livre SP¹, iniciado em 2014 pela prefeitura de São Paulo para tornar a Internet mais acessível à população. O programa WiFi Livre SP implantou 120 praças digitais, onde o acesso é irrestrito e gratuito a qualquer cidadão. O programa foi projetado para oferecer 512 Kbps por usuário, para *download* e *upload* além de certas garantias de atraso mínimo. O serviço é prestado por duas empresas selecionadas através de um processo de licitação. As áreas de cobertura englobam as 5 regiões da cidade de São Paulo: Centro, Zona Norte, Zona Sul, Zona Leste, Zona Oeste.

O programa WiFi Livre SP vem sendo sistematicamente monitorado para compreender o uso das praças, a sua disponibilidade e desempenho [Ratusznei 2015]. A mensuração do efeito da política pública de oferta de conexão à Internet gratuita através de uma rede WiFi Municipal envolve uma equipe com atuação interdisciplinar para realizar avaliação qualitativa e quantitativa da qualidade das praças digitais sob a percepção dos usuários. Foram realizadas avaliações qualitativas com pesquisa de campo sobre a opinião dos usuários, cuja apresentação, no entanto, está fora do escopo desse artigo. Com relação à avaliação quantitativa, foram realizados estudos de avaliação de desempenho para averiguar os níveis de Qualidade de Serviço (QoS) usando métricas de rede, assim como para averiguar a Qualidade de Experiência (QoE) percebida pelo usuário, através de experimentos com *streaming* de vídeo. A metodologia utilizada tem por objetivo revelar informações que poderão ser úteis para aprimorar a abordagem da implementação técnica que baliza o projeto e instalação e operação das praças digitais, visando estudar os efeitos da Internet, em suas múltiplas dimensões, na vida dos cidadãos e das comunidades atendidas pela política de oferta de conexão gratuita de conexão WiFi.

Dada a atual importância das aplicações de vídeo, o presente artigo apresenta a avaliação de desempenho de aplicações de *streaming* de vídeo praças digitais de São Paulo, além de compará-las com dados de desempenho de rede. Os resultados confirmam que as métricas de QoS e QoE estão intimamente relacionadas e comprovam disparidade entre as praças com relação à qualidade de visualização de vídeos. Algumas praças permitem a visualização com qualidade adequada de vídeos com taxa de codificação de até 1200 Kbps, enquanto que outras geram desconfortos aos usuários (número e tempo de interrupções do vídeo) que impedem uma experiência satisfatória com bom nível de qualidade.

Na sequência do artigo, a sessão 2 apresenta os conceitos básicos e os trabalhos relacionados. A Sessão 3 apresenta a metodologia que foi adotada para avaliar o desempenho das aplicações de *streaming* de vídeo. A sessão 4 e 5 apresentam, respectivamente, os resultados obtidos no trabalho e uma discussão sobre os resultados. Finalmente, a sessão 6 apresenta a conclusão do trabalho.

2 Trabalhos Relacionados

Várias cidades de grande porte e importância no Brasil e no mundo disponibilizam acesso a redes WiFi públicas em larga escala, como São Paulo. Em termos da quantidade de pontos de acessos, destacam-se Roma, Barcelona, Zaragoza, Caracas,

¹ <http://wifilivre.sp.gov.br>

Adelaide, Xangai e Buenos Aires. No então, na maioria delas o acesso não é gratuito ou então o tempo de uso gratuito é limitado. Independente do tipo de acesso fornecido, não existem em geral resultados amplos de estudos que avaliam a qualidade da rede, de acordo com a percepção do usuário, como esse artigo se propõe a mostrar.

Existem várias referências na literatura sobre avaliação de WiFi e sobre avaliação de qualidade de vídeo, mas não foi encontrado nenhum resultado específico que contemple o uso de *streaming* de vídeo para compreender o funcionamento de um programa de WiFi público e gratuito. O trabalho desenvolvido por Sommers (2012) tem como objetivo entender o desempenho das redes móveis comparando o desempenho de redes celular e de WiFi utilizando de dados do Speedtest.net. O estudo considera o desempenho espaço-temporal (*upload/download*) da taxa de transferência e RTT. Mais de 3 milhões de testes foram realizados por usuários iOS e Android em 15 diferentes áreas metropolitanas. Um artigo correlato em termos de medida de larga escala em lugares de conexão sem fio aberta ao público foi o trabalho de Ghosh [2011] que obteve métricas de locais diversos como restaurantes, bares, empresas e cafeterias em duas cidades dos EUA com o intuito de propor um modelo de utilização das redes públicas.

Um estudo de grande porte envolvendo dezenas de milhares de redes sem fio WiFi em dois períodos de uma semana em janeiro de 2014 e janeiro de 2015 concluiu que a demanda pelo uso de redes sem fio mantém um crescimento rápido, principalmente pelo aumento da quantidade de dispositivos conectados e uso de aplicações de vídeo como YouTube e Netflix [Biswas et. al 2015]. Embora as redes estejam localizadas em ambientes particulares (cafeterias, bares, etc.) e portanto distintos das redes WiFi Municipais, o estudo contribui para ressaltar a importância da oferta de serviços públicos que atendam adequadamente as necessidades dos cidadãos, principalmente pela predominância do uso de aplicações de vídeo *streaming* que consomem considerável quantidade de recursos de rede.

A predominância do acesso a vídeo por usuários redes sem fio em dispositivos móveis tem despertado a atenção para a necessidade de compreender a Qualidade de Experiência (QoE) com foco na percepção do usuário, em vez de usar o mapeamento de métricas de Qualidade de Serviço (QoS) de rede para QoE [Baik et. al 2015]. A medição de qualidade de vídeo tradicionalmente tem utilizado métricas que comparam o vídeo recebido com o vídeo transmitido, ou seja, o resultado da transmissão com a referência armazenada no servidor, como PSNR (*Peak Signal to Noise Ratio*) [Serral-Gracia 2010], SSIM (*Structural Similarity*) [Wang 2002] e VQM (*Video Quality Metric*) [Pinson 2004]. Essas métricas auxiliam o processo de compreensão da experiência do usuário, mas não capturam o possível sentimento do usuário com relação à qualidade do vídeo assistido e por isso estão sendo criadas novas métricas que avaliam QoE [Balachandran et. al 2013].

A percepção do usuário é afetada por diferentes fatores que podem ser mensurados de forma objetiva e subjetiva. A relação da qualidade com o buffer tende a apresentar impacto significativo na percepção do usuário, como o tempo de início do vídeo (*playback start time*), a quantidade de vezes que a imagem é congelada (*number of interruptions*), a duração dos congelamentos de imagem (*duration of interruptions*) e o local no vídeo onde ocorrem os congelamentos [Juluri 2016] [Baik et. al 2015].

Nesse trabalho foi feita a opção de avaliar a qualidade de vídeo sob a percepção do usuário, assumindo como premissa que a interrupção do vídeo causada pelos

mecanismos de buferização é o fator que mais afeta negativamente a qualidade da experiência. Além desses fatores, a diminuição da taxa da codificação para vídeos adaptativos também afeta o usuário, mas nesse trabalho foram usados vídeos com codificação fixa, para possibilitar uma comparação justa entre praças, dias e horários.

3 Metodologia

Esse artigo se concentra em resultados de QoE com foco na percepção da qualidade de vídeo pelo usuário, uma vez que resultados de QoS com foco em desempenho de rede foram apresentados anteriormente em [Ratusznei 2015]. Apenas alguns resultados de QoS atualizados no período dos experimentos com vídeo são apresentados na seção 4.

3.1 Ferramentas

Existem ferramentas que mensuram métricas como PSNR e SSIM, como EvalVid [Klaue 2003], que no entanto não foram utilizadas pela escolha da técnica de percepção da qualidade baseadas nas interrupções do vídeo. Como nenhuma ferramenta foi encontrada que possibilitasse a medição das interrupções de vídeo, houve a necessidade de buscar soluções alternativas, que permitissem a coleta das métricas desejadas. A ferramenta encontrada que mais se aproxima da necessidade foi o VLC², na versão 2.2.1, um reprodutor multimídia multiplataforma de código aberto que transmite vídeo via *streaming* em várias codificações e fornece bibliotecas para customizar aplicações cliente. O servidor de vídeo foi configurado para iniciar a transmissão sob comando do cliente e foi instalado na UFABC.

O cliente de vídeo é responsável por se conectar ao servidor e solicitar a transmissão do vídeo, além de iniciar simultaneamente a medição de atraso (RTT) e perda de pacotes na rede usando mensagens ICMP através do programa *ping*. A biblioteca do VLC permite identificar quando o vídeo sofre interrupção (congelamento) e quando a transmissão normal é iniciada ou restabelecida. Além disso, permite que o cliente obtenha dados sobre a taxa de transmissão que está sendo efetivamente recebida (vazão), para compará-la posteriormente com a taxa de codificação do vídeo.

3.2 Procedimentos para Medição de QoS e QoE

As medições de QoS foram realizadas com dados coletados das empresas prestadoras do serviço e das medições realizadas pelo SIMET Box [Ratusznei 2015]. Esse artigo se concentra na metodologia e resultados de experimentos de medição de QoE de vídeo em cinco praças localizadas em cinco regiões de São Paulo. Foram realizados experimentos com duas taxas de codificação de vídeo nas cinco praças em todos os dias da semana, durante novembro e dezembro de 2015. Cada experimento foi replicado 30 vezes e os intervalos de confiança ao nível de 99% foram computados. Cada trecho de vídeo possui 1 minuto de duração, de modo que com dois tipos de vídeo (de qualidades diferentes) com 30 replicações somente o tempo dos experimentos consome uma hora. Além disso, em média foram necessários 40 minutos para troca de vídeo, intervalos entre replicações e eventuais falhas de conexão.

A Tabela 1 traz as lista dos parâmetros fixos utilizados nos experimentos. Para garantir a justiça na avaliação, o usuário (sempre o mesmo, para garantir justiça nas

² <http://www.videolan.org/vlc/>

avaliações) posicionou-se à mesma distância (10 metros) do Access Point (AP) em todos os experimentos. O valor de 10 metros foi escolhido após observação empírica do comportamento típico dos usuários nas praças. Na maioria das vezes, os usuários não tem consciência da localização do AP e utilizam a conexão WiFi de qualquer ponto da praça. Não houve variação da distância devido à complexidade logística, condições climáticas e demanda de tempo físico para a realização de novos experimentos. O vídeo utilizado tem duração de 60 segundos e foi editado a partir de um vídeo com grande número de visualizações no YouTube³. A ideia é permitir replicações com duração de um minuto e ao mesmo tempo utilizar um vídeo que misture cenas escuras e claras e com pouco e muito movimento.

A **Tabela 2** traz os fatores e níveis, ou seja, os parâmetros variados em experimentos diferentes e os valores que eles assumiram. Foram avaliadas cinco praças digitais, de acordo com a seção 3.3. Dois níveis de qualidade de vídeo foram avaliados, codificação com taxas de 1200 Kbps e 840 Kbps, com o objetivo de complementar os dados de uso e desempenho das praças com vídeos acima de 512 Kbps, a taxa média projetada por usuário. Houve coletas em todos os dias da semana, variando de 7h às 18h. Cada praça foi avaliada em todos os dias da semana em horários pré-estabelecidos. Novamente devido à complexidade logística da realização dos experimentos não foram realizadas medições em todas as horas do dia em todas as praças em todos os dias da semana.

Tabela 1: Parâmetros

Parâmetro	Valor
Distância do AP	≈ 10 metros
Servidor de destino	UFABC
Duração	60 segundos
Captura	1 segundo

Tabela 2: Fatores e Níveis

Fator	Nível
Praça	Centro, Norte, Sul, Leste, Oeste
Qualidade	1200 Kbps, 840 Kbps
Hora	7:00h às 18:00h
Dia	Todos os dias da semana

3.3 Seleção de Praças, Datas e Horários de Medição

Foram selecionadas cinco praças entre as 120 do programa WiFi Livre SP, uma de cada região. A Tabela 3 apresenta as praças e o número máximo de usuários para o qual elas foram projetadas, e a Figura 1 apresenta a sua localização no mapa.

- Utilização da praça: baseada no número de usuários conectados por hora do dia nos últimos três meses antes da realização dos experimentos;
- Segurança da praça: como as medições são realizadas in loco, foi considerada a taxa de criminalidade;
- Acessibilidade e deslocamento: as cinco praças foram visitadas todos os dias de medição e portanto foi considerado a facilidade de deslocamento entre elas.

Tabela 3: Praças Selecionadas

Região	Centro	Norte	Sul	Leste	Oeste
Praça	Biblioteca Dom José Gaspar	Praça Nossa Senhora dos Prazeres	Largo Dona Ana Rosa	Praça Silvio Romero	Praça Conde Francisco Matarazzo
Usuários	150	50	75	100	50

³ <https://youtu.be/05I6WZg8v8c>



Figura 1: Distribuição das praças na cidade de São Paulo

A Figura 2 apresenta o número médio de usuários conectados nas cinco praças selecionados durante um período de três meses. Esse parâmetro pautou tanto a escolha dessas praças quanto a concepção do cronograma e horários de medição em cada praça. Dada a diferença de perfil de usuário e de utilização entre dias de semana e finais de semana, foram realizadas medições em todas as praças em todos os dias. Devido a limitações logísticas, não foi possível realizar medições em várias horas do dia em vários dias da semana em uma mesma praça. O cronograma foi integralmente cumprido em aproximadamente 16 dias.

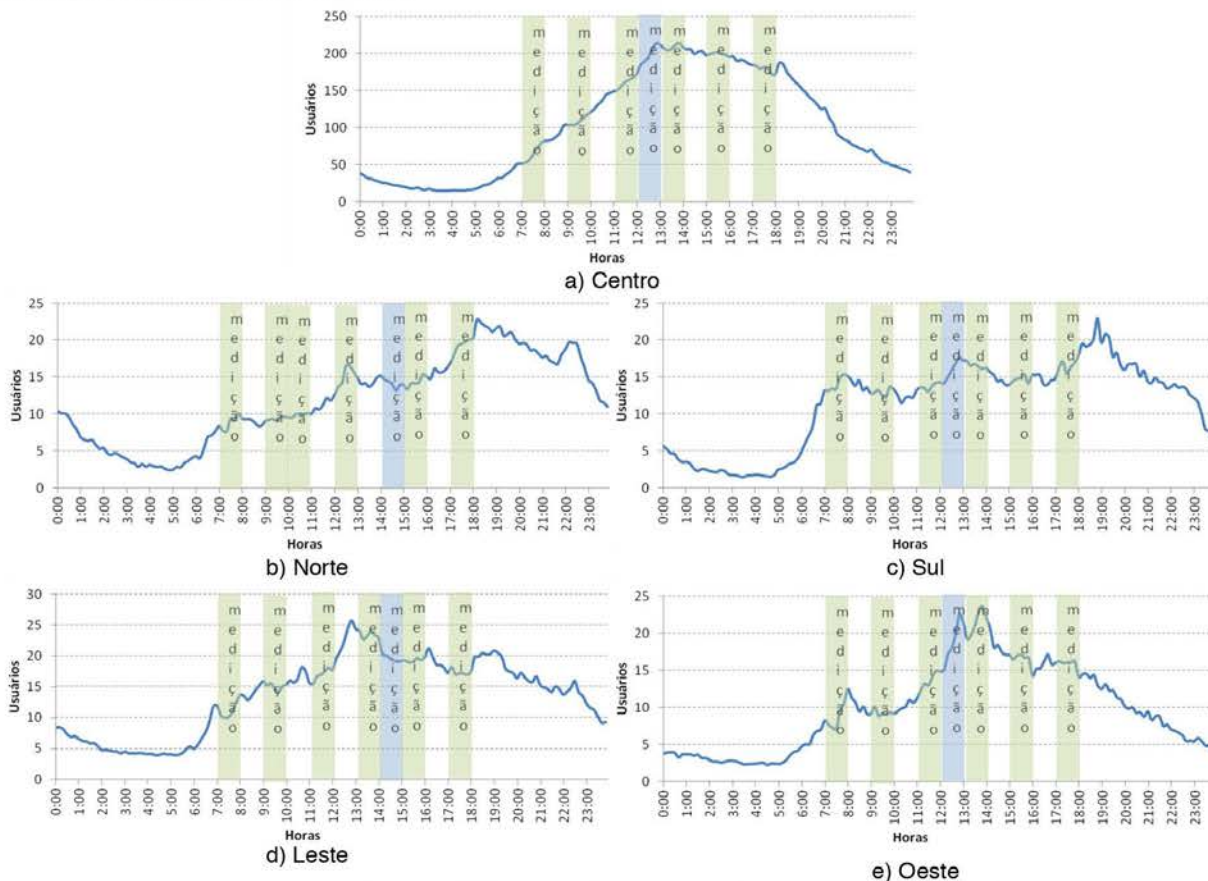


Figura 2: Número de usuários por horas do dia nas praças selecionadas

3.4 Métricas

As métricas utilizadas para avaliar as praças são divididas em duas categorias: QoS do serviço WiFi na praça QoE do usuário com relação ao vídeo.

- QoS: obtidos através da execução do comando ping durante as medições e através da coleta de dados provenientes das empresas prestadores do serviço.
 - Nível de utilização da praça: número de usuários que utilizam a praça dividido pelo número máximo de usuários previstos (Tabela 3). O número máximo determina a capacidade da rede WiFi da praça e do enlace de acesso, considerar uma capacidade média de 512 Kbps por usuário.
 - Taxa de entrada de dados por usuário: a taxa total de dados de entrada na praça dividida pelo número de usuários conectados naquele momento.
 - RTT: tempo de ida e volta de pacotes até o IP de destino.
 - Taxa de perda de pacotes: percentual de pacotes perdidos no caminho de ida ou de volta.
- QoE: visualização do vídeo com as médias dos dados estatísticos de:
 - Tempo de espera para carregamento de *buffer*: tempo despendido para encher o buffer e possibilitar o início da exibição do vídeo.
 - Número de congelamentos: número de vezes que o vídeo é interrompido devido a condições inadequadas de rede, para recompor o buffer de recepção.
 - Tempo de congelamento: tempo gasto para o vídeo se recuperar de uma interrupção (congelamento da imagem) em uma transmissão.
 - Taxa de recepção (vazão): taxa de recepção real, que na teoria deveria ser igual à taxa de codificação do vídeo.

4 Resultados

Os resultados são classificados de acordo com as métricas em dois grandes grupos: QoS e QoE. Devido à limitação de espaço são mostrados resultados de apenas duas praças, Praça Conde Francisco Matarazzo na Zona Oeste e Praça Silvio Romero na Zona Leste, com melhor e pior desempenho respectivamente.

4.1 Qualidade do Serviço (QoS)

Na Figura 3 é possível observar a taxa de utilização das praças de acordo com dados fornecidos pelas empresas no momento em que ocorreram os experimentos com vídeo. Na Figura 3(a) é possível observar que durante as medições realizadas no domingo a Praça Silvio Romero contava com apenas 10% do número máximo de usuários, devido ao horário da medição (7:00 horas da manhã), diferente da terça-feira, onde foi observado 40% do número máximo de usuários porque a medição foi realizada no início da tarde (13:00 horas). Além disso, na sexta-feira não há dados para essa métrica por motivo de falha na captura dos dados. A Figura 3(b) mostra que a praça Francisco Matarazzo apresenta utilização regular também considerando os dias e horários de medição. Para as demais praças ocorre comportamento similar ao da Figura 3.

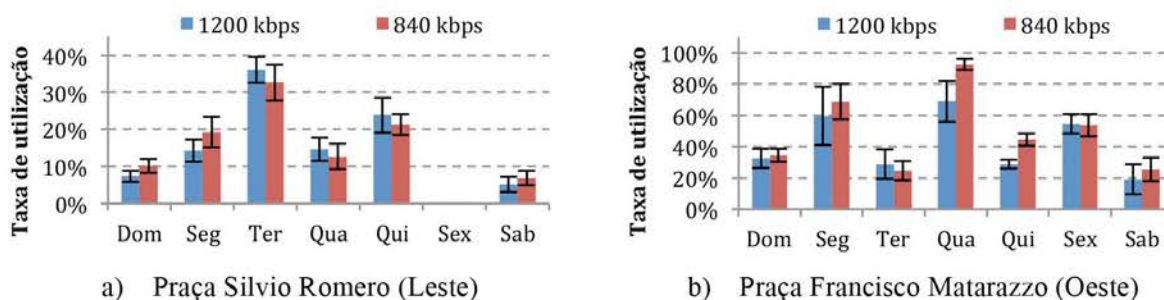


Figura 3: Nível de utilização das praças

Na Figura 4 é possível observar uma grande diferença entre as duas praças com relação à taxa de transferência de dados (entrada, ou seja *download*), onde há muito mais dados trafegados na Praça Francisco Matarazzo do que na Praça Silvio Romero. Esse fenômeno ocorre também para as demais praças analisadas, pois cada uma recebe usuários com comportamento e características de uso da rede próprios. A respeito da Figura 4(a) não há dados para a sexta em que foi realizado o experimento de vídeo.

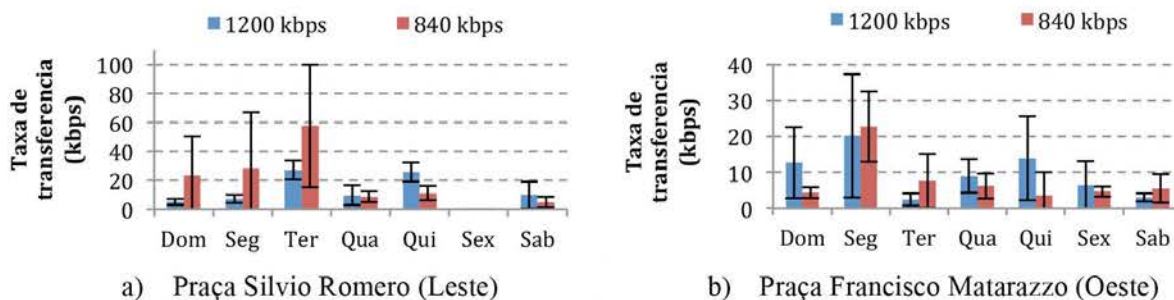


Figura 4: Taxa de transferência por usuário (download)

Para o atraso em duplo sentido (RTT) é possível verificar na Figura 5(a) que o atraso médio entre os dias da semana sofre alterações muito bruscas, além de grandes variações durante o dia ressaltadas pelo tamanho do intervalo de confiança. Comparando o domingo com a segunda-feira é visível um aumento de quatro vezes. Entretanto, na quarta-feira há uma redução brusca devido ao horário em que foi realizado o experimento. Alguns valores de RTT são claramente inadequados para uma qualidade adequada de utilização da maioria das aplicações na Internet. Esse comportamento ocorre nas praças do Centro e Zona Sul. Já na Praça Francisco Matarazzo na Zona Oeste, Figura 5(b) os valores e a variação são consideravelmente menores, situação similar à Zona Norte.

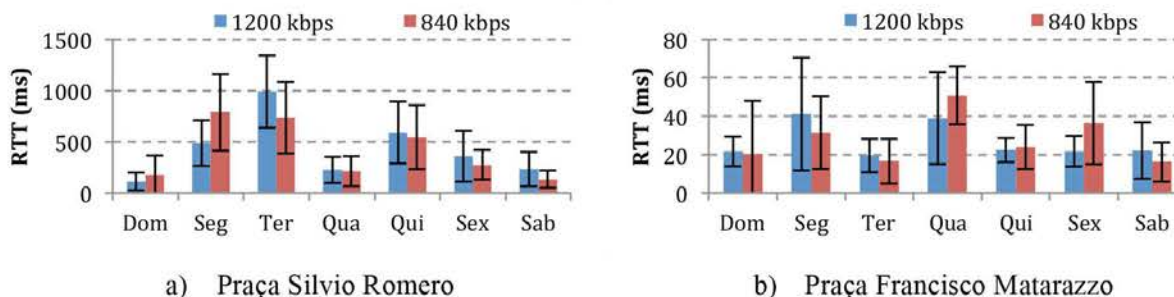


Figura 5: RTT (atraso em duplo sentido)

Para a perda de pacotes é possível verificar que no melhor caso, Figura 6(b) existe uma perda de no máximo 6% que ocorre na sexta-feira para o vídeo de maior taxa de

transmissão. Esse mesmo comportamento foi observado nas praças da Zona Norte e Zona Sul. A Figura 6(a) apresenta um comportamento aceitável para o domingo e a segunda-feira com perda de pacotes na faixa de 1% a 2%. Entretanto, nos outros dias ocorre uma elevação significativa na perda de pacotes, acima de 20% na terça-feira e no sábado. Comportamento semelhante a esse ocorre na praça da região central.

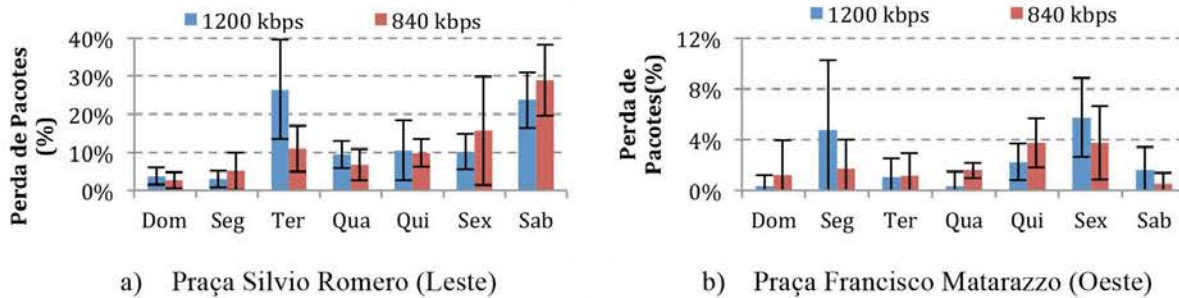


Figura 6: Perda de Pacotes.

4.2 Qualidade de Experiência

Essa seção explora os resultados de QoE, através do tempo de iniciar o vídeo, número e duração dos congelamentos de imagem e taxa de recepção. A Figura 7(a) mostra o tempo necessário para carregar o *buffer* na Praça Silvio Romero. É possível visualizar que para a maioria dos dias da semana, exceto na segunda-feira, ocorre um atraso significativo para que a exibição do vídeo seja iniciada. Esses dados são corroborados com a baixa taxa por usuário, alta atraso e alta taxa de perda de pacotes evidenciados na seção 4.1. Comportamento semelhante é observável em alguns dias da semana para as praças da zona norte, sul e central. Ao contrário do experimento, que foi executado até o final, para os usuários a percepção negativa de qualidade nessas condições é alta, gerando a desistência de assistir vídeos e impressão negativa da qualidade da praça.

Situação oposta ocorreu na praça Francisco Matarazzo onde o tempo observado para carregar o *buffer* e iniciar o vídeo permaneceu em torno de dois segundos, Figura 7(b). Mais uma vez, esses dados são corroborados com métricas de QoS mais favoráveis, como uma maior taxa por usuário, menor atraso e menor taxa de perda de pacotes evidenciados na seção 4.1.

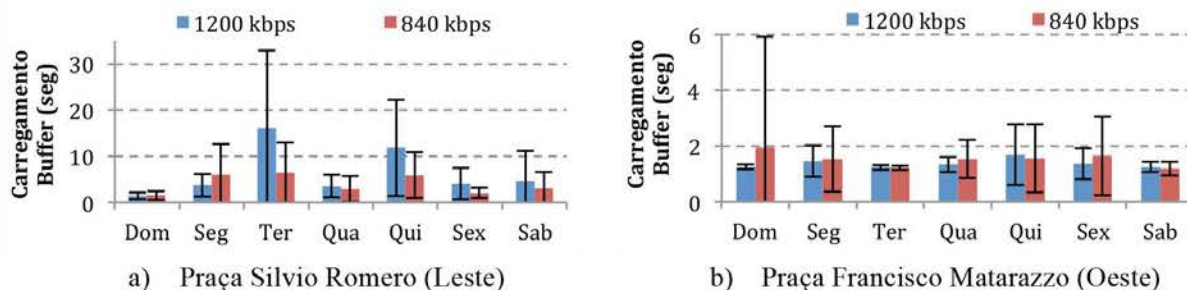


Figura 7: Carregamento de *buffer*

Para número de congelamentos do vídeo a Figura 8(a) representa o pior caso observado. Na maioria dos dias da semana é possível visualizar que para o vídeo de qualidade mais alta (1200 Kbps) ocorreram em média quatro congelamentos. Para as cinco praças analisadas a que teve o melhor comportamento foi a praça Francisco Matarazzo, Figura 8(b), onde em média ocorreu um congelamento. Novamente, a percepção de qualidade do usuário decai com as frequente interrupções do vídeo.

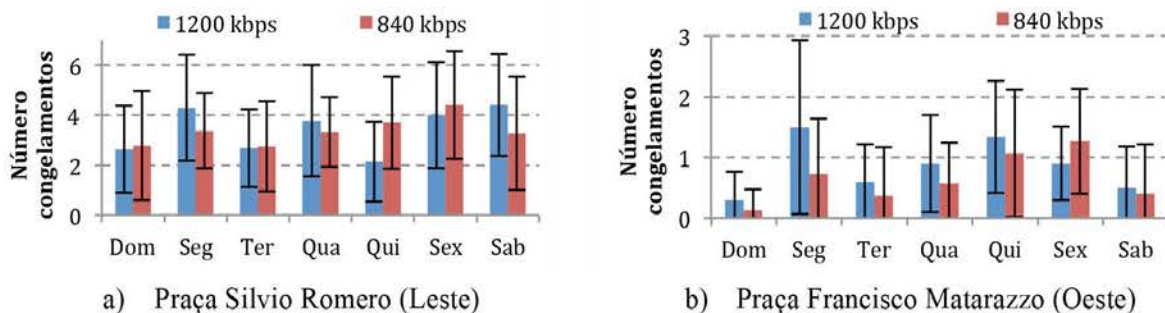


Figura 8: Número de congelamentos na exibição dos vídeos

Para a duração dos congelamentos, a Figura 9(a) mostra que houve um pico de 30 segundos de congelamento para o vídeo de qualidade alta na terça-feira na praça Silvio Romero. É possível visualizar para esse mesmo dia o vídeo de qualidade inferior possui tempo médio de congelamento de aproximadamente 20 segundos. A Figura 9(b) apresenta o melhor caso, com 2 segundos na maioria dos dias. Para as praças do centro e da zona sul ocorre fenômeno parecido com o da Figura 9(a). Já para a praça da zona norte é semelhante a Figura 9b.

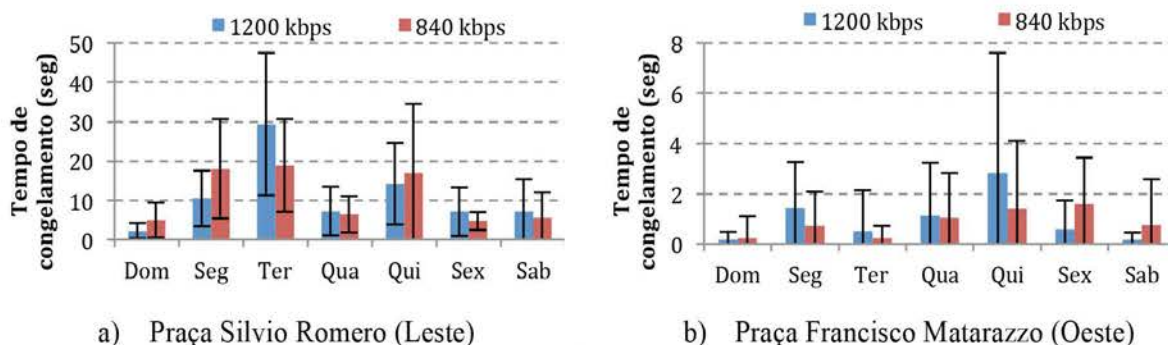


Figura 9: Tempo médio de congelamento por transmissão de vídeo

As informações sobre a taxa de recepção complementam e corroboram com as anteriores. A Figura 10(a) mostra um exemplo em que houve falha na exibição do vídeo, pois há variações significativas na taxa de transmissão. Por exemplo, na terça-feira a taxa deveria estar em 1200 Kbps mas o observado foi algo em torno de 200 Kbps, indicando falha grave na execução do vídeo. O caso ótimo é mostrado Figura 10(b) onde para todos os dias as métricas se mantêm estáveis em 1200 Kbps para o vídeo de melhor qualidade e 840 Kbps para o vídeo de baixa qualidade.

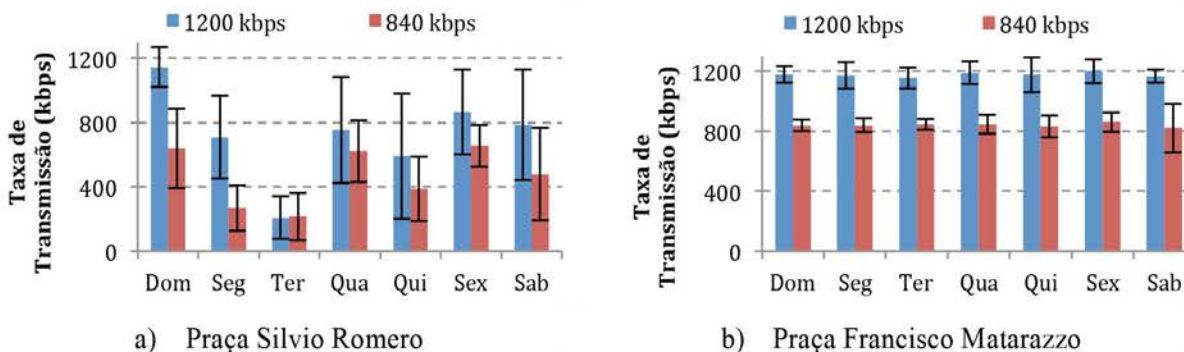


Figura 10: Taxa de transmissão dos vídeos

5 Discussão

Comparando os resultados de QoS e QoE, é possível observar que existe grande compatibilidade entre eles. Ou seja, quando as condições de rede estão inadequadas, não será possível obter qualidade aceitável na visualização de um vídeo. É importante lembrar que de acordo com as especificações do edital de contratação das empresas prestadoras do serviço da cidade de São Paulo, a taxa média de download para cada usuário deve ser de 512 Kbps. Enquanto isso, as duas taxas (e qualidades) de vídeo utilizadas são superiores a esse patamar. Isso significa que para um usuário individual, em algumas praças a qualidade está superior ao que foi contratado, enquanto em outras está significativamente inferior, como pode ser observado na Figura 10(a).

Foi demonstrado também que os resultados das métricas de QoE e QoS se relacionam, que corrobora com a visão intuitiva para essa situação. Por exemplo, o nível de utilização da praça está relacionado com as métricas de taxa de entrada por usuário, atraso e perda de pacotes, que por sua vez influenciam a estabilidade da taxa de recepção do vídeo e os congelamentos de imagem que tanto afetam a percepção de qualidade dos usuários. Além disso, por se tratar de um experimento de medição ativa (gera tráfego) o próprio experimento pode causar a degradação do serviço para os demais usuários da praça.

Para todas as praças foi observado no mínimo um dia em que os experimentos mostram resultados próximos ao caso ótimo, mesmo para a praça Silvio Romero, o que pode ser comprovado com a visualização as figuras (a). No melhor caso da Praça Silvio Romero, a taxa de transmissão chega aproximadamente nos níveis de 1200 Kbps para o vídeo de melhor qualidade e 840 Kbps para o vídeo de baixa qualidade, o que é corroborado pelo duração de congelamento de 5 segundos, ocorrência de congelamento próxima a 2 vezes e carregamento de buffer de 2 segundos, além de perda de pacotes de 4%, RTT abaixo de 180 ms, baixo nível de utilização, e baixo tráfego de entrada.

Nas cinco praças digitais do programa WiFi Livre SP avaliadas foi identificado um limiar empírico de cerca de 50 ms de RTT e 3% de perda de pacotes para garantir uma qualidade de vídeo aceitável para o usuário. Existe uma forte correlação entre esses limiares e a visualização de vídeos nas praças com cerca de dois segundos de carregamento de *buffer* e apenas um congelamento de aproximadamente no máximo de dois segundos. Isso possibilita que o usuário assista o vídeo sem muitas interrupções e com qualidade de taxa de transmissão muito próxima ao original transmitido.

6 Conclusão

Redes WiFi Municipais tem o objetivo de oferecer ampla conectividade gratuita aos cidadãos, mas já falharam no passado devido a inúmeros problemas técnicos e de regulação. A abordagem desse artigo é a utilização de *streaming* de vídeo em praças digitais do programa WiFi Livre SP distribuídas pelas cinco regiões da cidade de São Paulo. Foi realizada uma análise de desempenho nas praças utilizando-se de métricas que avaliam tanto a Qualidade de Experiência (QoE) do usuário, como a Qualidade de Serviço (QoS) da rede.

Para todas as praças foi possível verificar que em algum momento no mínimo uma transmissão teve algum tipo de perda, e também de no mínimo uma transmissão sem nenhuma falha. Em média os resultados mostram que para diferentes praças, dias e

níveis de qualidades de vídeo, na maioria das vezes o resultado é satisfatório para os usuários. Dada a magnitude do projeto, as avaliações até aqui demonstram que a qualidade é variável para praças diferentes.

Como trabalhos futuros, pretende-se efetuar experimentos com vídeo em um número maior de praças, dias e horários, além de utilizar novas métricas de qualidade, assim como identificar um índice de qualidade de vídeo para ambientes de Internet pública via WiFi.

Referências

- Bar, F., Park, N. (2006), “Municipal Wi-Fi Networks: The Goals, Practices, and Policy Implications of the U.S. Case”, *Communications & Strategies*, no. 61, p. 107.
- Baik, E., Pande, A., Stover, C., Mohapatra, P., (2015), “Video Acuity Assessment in Mobile Devices”, *IEEE INFOCOM 2015*.
- Balachandran, A., (2013), “Developing a Predictive Model of Quality of Experience for Internet Video Categories and Subject Descriptors”, *SIGCOMM*, p. 339–350, 2013.
- Biswas, S., Bakked, J., Wong, E., (2015), “Large-scale Measurements of Wireless Network Behavior”, *SIGCOMM 2015*, p. 153–165, 2015.
- Fraser, E., (2010), "A Postmortem Look at Citywide WiFi", *Journal of Internet Law*, 14(2), August 2010.
- Ghosh, A.; et. al. (2011), “Modeling and Characterization of Large-Scale Wi-Fi Traffic in Public Hot-Spots”, *IEEE INFOCOM 2011*.
- Juluri, P., Tamarapalli, V., Medhi, D., (2016), "Measurement of Quality of Experience of Video-on-Demand Services: A Survey", *IEEE Communications Surveys & Tutorials*, 18(1), First Quarter 2016.
- Klaue, J., Rathke, B., Wolisz, A., (2003), “EvalVid - A Framework for Video Transmission and Quality Evaluation”, *13th Intl Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pp. 255-272, 2003.
- Pinson, M. Wolf, S., (2004), "A New Standardized Method for Objectively Measuring Video Quality", *IEEE Transactions on Broadcasting*, 50(3), pp. 312-322, September.
- Potts, J., "Economics of Public WiFi", *Australian Journal of Telecommunications and the Digital Economy*, 12(1), Março de 2014.
- Ratusznei, J., Silva, W., Pinheiro, N., Melo, R., Kamienski, C. (2015). Uma Rede WiFi Aberta de Larga Escala como Infraestrutura para Cidades Inteligentes, *SEMISH 2015*.
- Serral-Gracia, R., Cerqueira, E., Curado, M., Yannuzzi, M., Monteiro, E., Masip-Bruin, X., (2010), “An Overview of Quality of Experience Measurement Challenges for Video Applications in IP Networks”, *Intl Conference on Wired/Wireless Internet Communications (WWIC 2010)*, pp. 252–263.
- Sommers, J.; Barford, P. (2012), “Cell vs. WiFi: on the performance of metro area mobile connections”, *2012 ACM conference on Internet Measurement Conference (IMC 2012)*, p. 301-314.
- Wang, Z., Lu, L., Bovik, A., (2002), “Video quality assessment using structural distortion measurement”, *International Conference on Image Processing*, 2002.

Proposta de Solução para Aplicação de Internet das Coisas nas Redes Elétricas Inteligentes Brasileiras Considerando as Limitações das Normas Vigentes*

Efrem E. O. Lousada¹, Raquel A. F. Mini²

¹Instituto de Informática – Instituto Federal do Norte de Minas Gerais (IFNMG)
Pirapora – MG – Brasil

²Departamento de Ciência da Computação
Pontifícia Universidade Católica de Minas Gerais – Belo Horizonte – Brasil

efrem.lousada@ifnmg.edu.br, raquelmini@pucminas.br

Abstract. *Advances in the areas of embedded systems, computing and networking have created thousands of heterogeneous devices. Electronic electricity meters are examples of these devices that interconnected can be part of a larger network called Smart Grids. Applying the concept of Internet of Things in these networks, meters should provide the collected information to all system users through servers in the cloud. This work specifies and implements a Brazilian smart meters network, considering the limitations imposed by Brazilian standards, and proposes a technique for calculating the duty cycle of the network dynamically in order to maximize the network lifetime. Experimental results show that the proposed solution, using dynamic duty cycle, increases significantly the network lifetime.*

Resumo. *Avanços nas áreas de sistemas embarcados, computação e redes têm criado milhares de dispositivos heterogêneos. Os medidores eletrônicos de energia elétrica são exemplos desses dispositivos que interconectados podem fazer parte de uma rede maior chamada de redes elétricas inteligentes. A partir da aplicação de internet das coisas nessas redes, os medidores passam a disponibilizar as informações coletadas a todos os usuários do sistema através de servidores na nuvem. Este trabalho especifica e implementa uma rede de medidores inteligentes brasileiros, considerando as limitações impostas pelas normas brasileira, além de propor uma técnica para cálculo do ciclo de trabalho da rede de forma dinâmica com o objetivo de maximizar o tempo de vida. Resultados experimentais mostram que a proposta de utilização de ciclo de trabalho dinâmico aumenta significativamente o tempo de vida da rede.*

1. Introdução

O crescente desenvolvimento das aplicações embarcadas tem proporcionado inteligência computacional a diversos tipos de dispositivos. Os medidores de energia elétrica, por exemplo, são microprocessados e deixaram de registrar somente o consumo de energia do consumidor. A possibilidade de comunicação remota com os medidores de energia

*Este trabalho foi parcialmente financiado pela empresa Nansen S/A Instrumentos de Precisão, CNPq e Fapemig.

elétrica e a criação de redes entre eles foram os primeiros passos para a criação das redes elétricas inteligentes ou *smartgrids*. Utilizando o conceito de geração e compartilhamento de informações relevantes para a sociedade, a partir de aparelhos computacionais interligados, chamado de Internet das Coisas (IoT - *Internet of Things*), a aplicabilidade das redes elétricas inteligentes pode ser expandida. As informações e serviços gerados pelos medidores eletrônicos de energia elétrica podem ser compartilhados com outros dispositivos ou pessoas interessadas nas informações [Bello and Zeadally 2014].

O grande desafio no desenvolvimento de uma rede inteligente de medidores eletrônicos de energia elétrica é atender à classe de medidores de BT (Baixa Tensão) e as normas que a define. Nessa classe, os recursos de hardware são limitados o que diminui as possibilidades de soluções. As fontes desses medidores conseguem prover no máximo 30 mA e qualquer solução implantada no medidor não poderá consumir mais do que isso. A rede deve ser capaz de funcionar mesmo sem a alimentação direta da rede elétrica, pois os nós devem enviar alarmes de falta de energia para as concessionárias atuarem de forma mais assertiva na solução do problema, atendendo às aplicações de IoT. Por isso, a alimentação por baterias é necessária quando faltar energia à rede de medidores.

Devido a utilização de baterias, é importante a aplicação de técnicas para maximizar o tempo de vida da rede. Entende-se como tempo de vida o tempo até a morte do primeiro nó. Uma das formas mais eficientes para maximizar o tempo de vida é desligar parte dos nós que não estão sendo utilizadas. Atualmente, os fabricantes de soluções para medição de energia elétrica devem manter os parâmetros e registradores de seus aparelhos por no mínimo 120 horas, quando estão sem energia e alimentados por baterias, conforme norma ABNT NBR (Norma Brasileira) 14519:2011 *Medidores eletrônicos de energia elétrica (estáticos)* [ABNT 2011]. Nesse caso, é necessário que o ciclo de trabalho seja o menor possível para aumentar o tempo de vida da rede.

Qualquer solução proposta para comunicação com medidores de energia deve atender a norma ABNT NBR (Norma Brasileira) 14522:2008 *Intercâmbio de informações para sistemas de medições de energia elétrica*. Esta norma é extremamente rígida quanto ao protocolo de comunicação utilizada pelos medidores e soluções para coleta dos dados de medição. Comandos de leitura de faturamento não podem sofrer qualquer alteração durante a transmissão e são checados durante a homologação do sistema no Inmetro. A agregação de dados, por exemplo, não é prevista e não pode ser utilizada por nenhuma solução de comunicação com os medidores [ABNT 2008].

O objetivo deste trabalho é projetar e desenvolver uma rede sem fio de medidores de energia que colete de forma automática as informações de medições de consumo e alarmes, disponibilizando em um servidor para acesso externo, atendendo as regulamentações brasileiras atuais. Além disso, a solução proposta realiza de forma dinâmica o cálculo do ciclo de trabalho de acordo com a formação e alterações na topologia, atendendo ao tempo mínimo de 120 horas de vida da rede, quando alimentada por baterias. Validações e testes realizados comprovaram o atendimento aos requisitos do setor.

O restante deste trabalho está organizado da seguinte forma: na seção 2, são apresentados os principais trabalhos relacionados à IoT aplicada as redes elétricas inteligentes. A seção 3 apresenta a descrição da solução proposta, especificando hardware, software e a técnica para cálculo dinâmico do ciclo de trabalho. A seção 4 apresenta as validações

e resultados experimentais da pesquisa e, por último, na seção 5, são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Sustentada pela massiva instalação de sensores, a popularização das aplicações de IoT proverá informações cada vez mais precisas e reais. Na medição de energia elétrica, os primeiros trabalhos relacionando IoT e as *smartgrids* focaram em entender as características das duas tecnologias e as possibilidades da interação entre elas. Em [Yun and Yuxin 2010], os autores analisam as tecnologias e apontam a necessidade de uma comunicação nos dois sentidos pelas *smartgrids* para que os serviços atendam aos requisitos da IoT.

Em [Chen et al. 2011], os serviços e os benefícios que as *smartgrids* podem prover foram especificadas considerando IoT. Segundo os autores, a automação da coleta de dados de faturamento permitirá aos usuários um maior controle dos gastos com energia elétrica. Aplicações de automação residencial poderão ser incorporadas às redes elétricas inteligentes. As linhas de transmissão podem ser monitoradas em tempo real e os veículos elétricos poderão ser monitorados com maior assertividade.

A partir da definição dos serviços que podem ser oferecidos pelas *smartgrids* baseados em IoT, simulações e aplicações práticas começaram a ser especificadas e implementadas. Na pesquisa desenvolvida em [Ou et al. 2012], os autores apresentam os resultados do desenvolvimento prático de IoT nas linhas de transmissão de alta tensão chinesas. Instalou-se sensores de velocidade dos ventos, temperatura dos cabos de transmissão, umidade, chuvas e movimento dos cabos. Todo o sistema é alimentado por baterias solar e a transmissão das informações coletadas é realizada através de redes 3G.

No trabalho desenvolvido em [Arif et al. 2013], os autores projetam e implementam um protótipo de medidor inteligente que utiliza GSM (*Global System for Mobile Communications*) e ZigBee para criar uma aplicação de *smartgrids* na Arábia Saudita. O medidor aproveita o amplo uso da rede GSM, com a sua capacidade de envio de SMS (*Short Message Service*), e da tecnologia ZigBee para enviar e coletar os dados de consumo de energia. No mesmo ano, a empresa de energia elétrica sueca Vattenfall apresentou um projeto para monitoramento em tempo real de interrupções de energia e eventos de qualidade de energia, descrito em [Garpetun 2013], com o objetivo de aprimorar o atendimento a seus consumidores. Nesse trabalho, utiliza-se PLC (*Power Line Communication*) para comunicação com os medidores. Leituras entre 1 e 10 minutos são realizadas e, caso um medidor não responda, o sistema considera que o mesmo está sem energia elétrica.

As tecnologias de comunicação para medidores eletrônicos de energia elétrica aplicadas em IoT são apresentadas em [Arnachalam et al. 2014]. Os autores comparam as tecnologias considerando frequência utilizada, taxa de transmissão, área de cobertura, aplicações suportadas pelas tecnologias e suas limitações. A pesquisa sobre tecnologias de comunicação para medidores eletrônicos aplicadas em IoT é importante como referência para escolha em implementações práticas. O trabalho não define uma tecnologia mais adequada para aplicação de IoT em *smartgrids*. Em [Spano et al. 2014], são propostos e desenvolvidos protótipos para medidores inteligentes de energia elétrica, roteadores, servidores distribuídos para IoT e uma interface de usuário acessível por *smartphones*. O medidor inteligente proposto pelos autores não faz parte da *smartgrid*, ele é um módulo a

parte que pode ser conectado às tomadas das residências. Para comunicação entre o medidor e o servidor IoT os autores utilizam um *gateway* ZigBee. A aplicação de *smartphone* acessa esses servidores IoT para coletar as informações geradas pelos medidores inteligentes.

Os trabalhos da literatura não apresentam solução para aumentar a vida útil da rede quando utiliza-se baterias para alimentação dos módulos de comunicação. Além disso, as aplicações práticas são voltadas para mercados específicos e, por isso, não atendem as normas que regem as redes elétricas inteligentes brasileiras.

3. Arquitetura da Rede Elétrica Inteligente

3.1. Funcionamento

Nesta seção, a arquitetura da rede inteligente para medição de energia elétrica é apresentada. A figura 1 apresenta a arquitetura da rede que se divide em três módulos: módulo concentrador de dados (1), módulo de comunicação sem fio (2) e módulo de medição de energia elétrica (3).

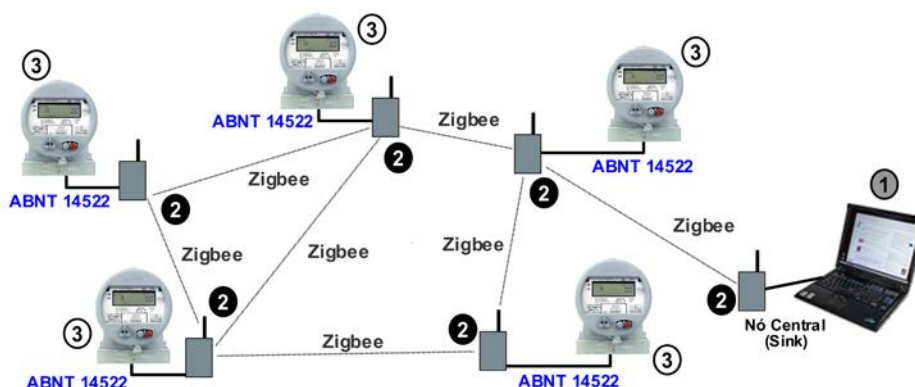
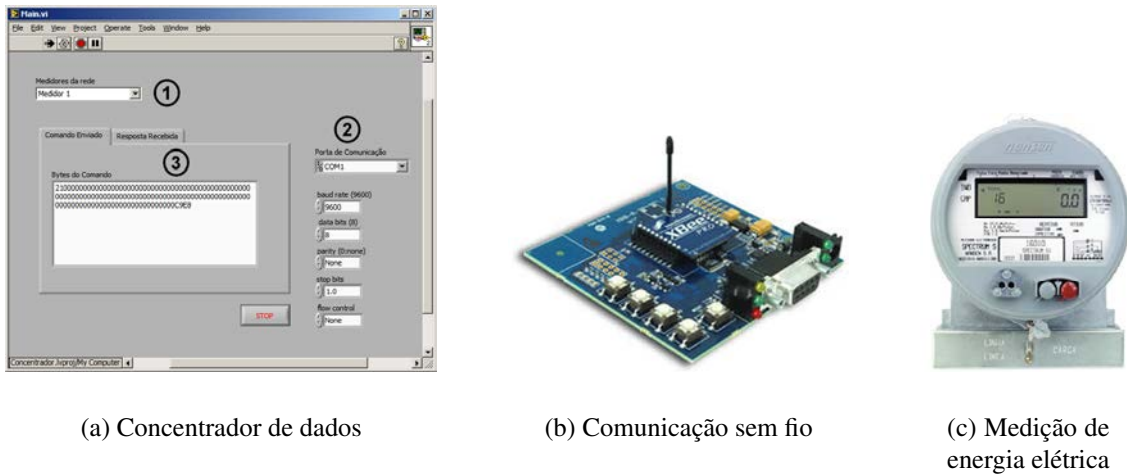


Figura 1. Rede sem fio de medidores eletrônicos de energia elétrica

O módulo concentrador de dados disponibiliza uma interface gráfica de usuário desenvolvida em LabView para controle das leituras de dados de consumo de energia elétrica dos clientes conectados à rede além de disponibilizar os dados em servidores na nuvem. A interface desenvolvida é apresentada na figura 2-(a). O número 1 indica o campo com os medidores componentes da rede. O número 2 representa os campos de configuração da porta serial. O número 3 apresenta os campos de comando e resposta.

O módulo de comunicação sem fio é responsável pelo tráfego de dados na rede, executando roteamento, controle de fluxo e criptografia dos dados. Dispositivos *transceiver* XBee, apresentados na figura 2-(b), foram utilizados para implementação. A árvore de coleção é formada e mantida pelo protocolo ZigBee. A prevenção de colisão é realizada através de CSMA/CA. O módulo de medição de energia elétrica é responsável pela medição dos dados de consumo dos clientes e disponibilização desses na rede. Para este módulo, utilizou-se o medidor eletrônico de energia elétrica Spectrum S, fabricado pela empresa Nansen S/A. A figura 2-(c) ilustra esse medidor. O medidor Spectrum S utiliza o DSP (*Digital Signal Processor*) TMS320C203 da Texas Instruments. As implementações desenvolvidas no sistema embarcado do medidor foram realizadas utilizando o compilador Code Composer 4.12.



(a) Concentrador de dados

(b) Comunicação sem fio

(c) Medição de energia elétrica

Figura 2. Módulos componentes da rede elétrica inteligente

Os módulos medição de energia elétrica e concentrador de dados comunicam-se através do módulo de comunicação sem fio. O módulo de medição de energia envia os dados de consumo registrados e a parametrização dos mesmos para a rede. O módulo de comunicação sem fio roteia esses dados, passando a informação por outros nós da rede, quando necessário, até o concentrador que disponibilizará os dados. Os dados coletados pela rede são disponibilizados na internet pelo sistema MECE (*Metering Center*) da Siemens S/A e os consumidores podem acompanhar seu consumo diariamente. As comunicações entre os módulos de medição de energia elétrica e o módulo de comunicação sem fio são realizadas utilizando a interface de comunicação serial RS232 dos medidores eletrônicos, conforme os padrões definidos na norma ABNT 14522 [ABNT 2008]. A comunicação entre os módulos de comunicação sem fio é feita utilizando ZigBee. A figura 3 apresenta as velocidades de transmissão entre os módulos.

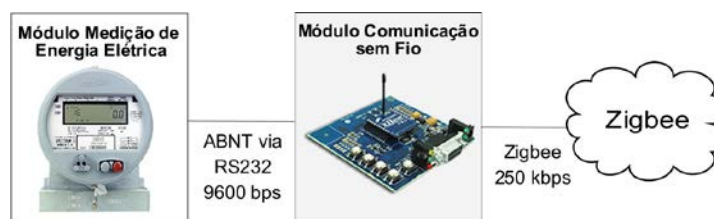


Figura 3. Velocidade de transmissão entre os módulos

Como os registradores dos medidores eletrônicos são atualizados a cada minuto e considerando a regulamentação da ANEEL que exige a atualização do mostrador de consumo do cliente sempre que o registrador de consumo do medidor for alterado, uma rede inteligente de medidores necessita de leituras dos dados de consumo e parâmetros do medidor a cada minuto. Outro fator importante que exige a leitura a cada minuto é o levantamento da curva de consumo do cliente, o que propiciará a aplicação da fatura horo-sazonal¹ e um melhor dimensionamento do sistema elétrico.

¹Aplicação de tarifas diferenciadas de consumo de energia elétrica e demanda de potência de acordo com a hora de utilização do dia e períodos dos anos.

Quanto ao regime de operação, deve-se levar em consideração o envio de 2 blocos de 258 bytes (comandos 21 e 23 da ABNT 14522 de tamanho fixo), informando os dados de medição e configuração dos medidores, a cada minuto por nó da rede. O comando 21 informa toda a parametrização do medidor. O comando 23 contém informações sobre o consumo instantâneo, demanda máxima e contador de pulsos de energia. Através da análise conjunta desses comandos é possível verificar se o medidor está enviando o consumo zerado devido a falta de energia, falha do medidor ou alteração indevida de terceiros. Os sistemas atuais de medição possuem algoritmos implementados que relacionam falta de energia, comando 21 e comando 23 para tomar decisões e disparar ações. Para isso, é requisito das concessionárias que o medidor continue enviando o consumo durante a falta de energia. Além disso, utiliza-se o envio a cada minuto para montar a curva de consumo, aplicar tarifação horosazonal e atender norma.

3.2. Ciclo de Trabalho Dinâmico

Esta seção apresenta o ciclo de trabalho dinâmico, solução proposta para maximizar o tempo de vida da rede inteligente para medição de energia elétrica, considerando o processo de formação da rede e as alterações na topologia que podem acontecer durante seu funcionamento. Inicialmente, todos os nós possuem ciclo de trabalho de 100%. O concentrador é parametrizado na concessionária com os nós componentes de sua rede, pois os clientes devem ser identificados para serem faturados. Assim checa-se, durante a formação, se todos os nós estão ativos. Caso um nó não entre na rede, a concessionária deverá verificar o problema e corrigi-lo, pois nenhum cliente pode ficar sem seu faturamento. Assim que todos os nós entram na rede, o primeiro ciclo de trabalho é calculado e enviado para a rede. Enquanto a rede está energizada os módulos de comunicação permanecem alimentados pela sua própria fonte e o ciclo de trabalho continua 100%. A rede passa a trabalhar dentro do ciclo de trabalho calculado quando ocorre uma falta de energia e os módulos passam a ser alimentados por baterias. No início de cada minuto, os nós acordam, transmitem as informações de consumo de energia elétrica e voltam ao estado inativo. Para isso, é necessário que todos os nós estejam sincronizados, acordando no tempo exato para enviar os dados. Essa sincronização é realizada pelo relógio interno do módulo medição. Para ativar ou desativar o módulo de comunicação sem fio, o módulo de medição envia um comando via RS232 para o mesmo. Foi necessário o desenvolvimento de um sistema embarcado específico para o medidor realizar as atividades descritas.

Sempre que a rede volta a ser alimentada pela rede elétrica o primeiro serviço realizado pelo módulo concentrador é a checagem do sincronismo dos nós da rede. Para isso, ele envia comandos específicos de leitura de hora e data para os módulos de medição de energia. Se necessário, um novo ciclo de trabalho é calculado considerando as diferenças de sincronismo. Medidores Spectrum S possuem relógio com precisão de 30 PPM/ano. Se a diferença for maior que a precisão esse medidor provavelmente está com problemas e um alarme é gerado no concentrador. Após esse alarme a concessionária pode realizar maiores checagens para entender o motivo da diferença e verificar se os medidores estão com problemas técnicos ou estão sendo alterados de forma indevida por terceiros.

A metodologia para cálculo do ciclo de trabalho deve ser aplicável a todas as configurações possíveis para redes de medidores de energia elétrica. A comunicação é realizada entre os módulos de comunicação e o módulo concentrador de dados. Como exemplo, considere uma rede de medidores formada de acordo com a topologia ilustrada

na figura 4. Para cada nó da rede, um número de saltos (d) até o concentrador de dados (*sink*) está apresentado. Essa profundidade é identificada durante a formação da rede.

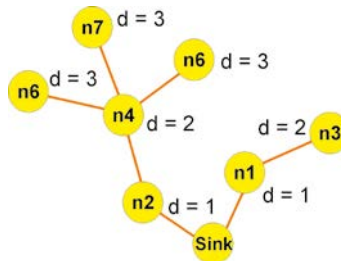


Figura 4. Exemplo de rotas entre nós

Quanto ao mecanismo de gestão dinâmica do ciclo de trabalho, sempre que um novo nó entra na rede um comando em difusão é enviado pelo mesmo, avisando a todos de sua entrada. Após receber o comando, o módulo concentrador calcula o novo ciclo de trabalho e envia o valor em difusão para todos da rede. O valor é assumido pelos nós e passa a ser o novo ciclo de trabalho da rede.

Para cálculo do valor do ciclo de trabalho, é necessário conhecer o tempo total necessário para que todos os nós da rede possam enviar seus dados. Neste trabalho, considera-se apenas o tempo de transmissão para o cálculo do tempo total de envio das informações. O tempo total necessário para enviar os dados é diretamente proporcional à quantidade de nós da rede (N), a profundidade de cada nó (d_i) e ao tempo de transmissão em um enlace (t). O tempo de transmissão em um enlace está relacionado ao tamanho do pacote e à velocidade de transmissão. O tamanho do pacote de dados é definido de acordo com a quantidade de dados que devem ser lidos de um nó da rede, sendo essa quantidade igual para todos os nós em uma rede para medição de energia elétrica.

Para que todas as informações cheguem até o concentrador, serão necessárias e transmissões, onde $e = \sum_{i=1}^N d_i$. Considerando que transmissões simultâneas não acontecem e também que todas as transmissões são bem sucedidas, o tempo total para transmissão seria: $T = t \times e$. Assim o ciclo de trabalho será suficiente para atender o número máximo de 40 nós da rede.

Entretanto, colisões e erros podem acontecer durante as transmissões. Considere p a probabilidade de sucesso de uma transmissão e Z uma variável aleatória com distribuição geométrica que representa o número de tentativas para envio com sucesso de um pacote em um determinado enlace, tem-se: $Z \sim Geo(p)$. O valor esperado da variável aleatória Z é: $E(Z) = \frac{1}{p}$. Sendo assim, no pior caso, o tempo total necessário para enviar todos os pacotes será definido por: $T = \frac{1}{p} \times t \times e$

Portanto, na solução proposta, no início de cada minuto, os nós ficam acordados durante T segundos para transmitirem as informações de consumo de energia elétrica e voltam ao estado inativo. O ciclo de trabalho deve ser atualizado sempre que um novo nó entrar na rede ou uma manutenção de rota for executada. Para realização da manutenção da rota, todos os nós continuam informando a sua profundidade durante o funcionamento, essa informação é registrada no concentrador e sempre checada. Caso ocorra uma alteração nesse valor, causado pela mudança de rotas, derivada de bloqueios ou mudança na qualidade do link, um novo ciclo de trabalho é calculado. Após realizar o

cálculo do novo valor de T , este deve ser atualizado e enviado para todos os nós da rede que passarão a utilizar o novo valor. O responsável pela execução do cálculo do ciclo de trabalho é o módulo concentrador de dados.

4. Validações e Resultados Experimentais

Esta seção apresenta as validações e resultados experimentais do sistema proposto neste trabalho. Os experimentos dividem-se em quatro grupos: medição do consumo de energia dos modos de funcionamento (seção 4.1), probabilidade de sucesso em cada transmissão (seção 4.2), tempo de vida da rede (seção 4.3) e o percentual de sucesso no recebimento das informações de consumo (seção 4.4). Todos os experimentos foram realizados na empresa Nansen S/A utilizando a plataforma XBee Série 2 - XB24BPKD juntamente com o amperímetro Minipa ET-1110 DMM para medição de consumo dos nós. Em todos os experimentos, utilizou-se a bateria Duracell Alcalina de 9 V padrão 6LR61 de 580 mAH como fonte de alimentação para os módulos de comunicação sem fio.

4.1. Medição do Consumo de Energia dos Modos de Funcionamento

O primeiro experimento foi realizado com o intuito de verificar o consumo de energia nos modos de funcionamento e o tempo de vida de um nó da rede em cada modo. Para este experimento, os nós 4 e 5, apresentados na figura 6-(b) foram utilizados. Um nó responsável pela transmissão e outro pela recepção. A distância entre os nós não foi alterada pois o ZigBee não possui controle de potência [Farahani 2008].

Para a coleta dos dados de consumo de energia, um circuito foi montado entre o módulo comunicação sem fio e sua alimentação. O circuito consiste em inserir um amperímetro em sequência com a bateria para medir o consumo da mesma. A cada cinco minutos, registrou-se o consumo do módulo de comunicação. A figura 5 ilustra o sistema montado e utilizado para medição das informações. Percebe-se na foto o módulo de medição de energia elétrica (1), medidor Spectrum S, o módulo comunicação sem fio ZigBee (2) e o amperímetro registrando o consumo (3).

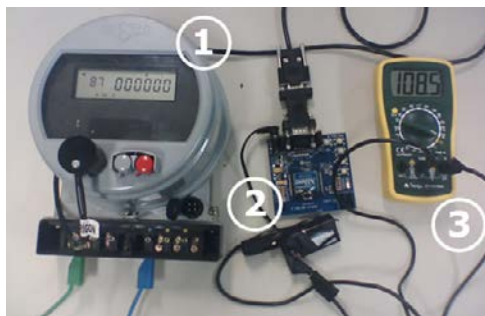


Figura 5. Registro de consumo do módulo de comunicação sem fio

Os experimentos foram realizados durante o período de 165 minutos, contabilizando 33 registros. Três modos de operação do módulo XBee foram avaliados: inativo, recepção e transmissão. No modo inativo, o rádio encontra-se desligado de forma que apenas a placa consome energia. O consumo do modo de funcionamento inativo é de aproximadamente 4,5 mA, pois é necessário manter os *leds* de funcionamento e a porta de comunicação RS232 para ativar ou destivar o módulo. Para execução do experimento

de medição do consumo durante a recepção de dados, o módulo comunicação sem fio permanece, durante todo o tempo, recebendo dados. Nesse modo, o consumo permanece constante na faixa de 87 mA. Durante a transmissão, o consumo aumenta consideravelmente se comparado com os modos anteriormente avaliados. O consumo para transmissão na placa de desenvolvimento permaneceu próximo a 110 mA.

A tabela 1 mostra o tempo de vida de um nó considerando seu modo de operação. O tempo de vida de um nó que permanece no modo inativo durante todo o período é 580 mAH (energia fornecida pela bateria) / 4,5 mA (consumo do modo de funcionamento), o que resulta em 128 horas e 53 minutos. Assim calcula-se o tempo de vida do nó para os demais modos de funcionamento.

Tabela 1. Tempo de vida de um nó

Modos de Operação	Tempo
Inativo	128 horas e 53 minutos
Recepção (Rx)	6 horas e 44 minutos
Transmissão (Tx)	5 horas e 16 minutos

A corrente necessária para transmissão do módulo de comunicação, 110 mA, é superior a fornecida pela fonte do medidor de energia elétrica que é 30 mA. Sendo assim, é impossível alimentar o módulo de comunicação com a fonte do medidor.

4.2. Probabilidade de Sucesso em cada Transmissão

O objetivo desse experimento é encontrar o valor da probabilidade de sucesso em cada transmissão (p) definido na seção 3.2. A figura 6-(a) ilustra a rede implementada, que utiliza a topologia *mesh*. Escolheu-se os nós 11 e 5 para esse experimento devido a maior quantidade de saltos para comunicação. O módulo de comunicação sem fio foi configurado para realizar apenas uma tentativa de envio do pacote pela rede. Caso ocorra algum problema durante o envio, não é realizada uma nova tentativa de transmissão. Além disso, o módulo permanece com o rádio ligado durante todo o experimento. Esse experimento foi realizado durante o intervalo de 5 horas, tempo suficiente para obter o valor de (p). A cada 5 segundos, o nó 11 envia os dados de consumo e configuração. Ao final, totalizou-se o envio de 7.198 comandos pelo nó 11 dos quais 7.133 foram recebidos pelo nó 5. Sendo assim, o percentual de sucesso em cada transmissão foi $p = 0,991$. Esse valor será utilizado nas próximas seções para o cálculo do valor do ciclo de trabalho.

4.3. Tempo de Vida da Rede

O objetivo dessa seção é avaliar o tempo de vida da rede para pilha de protocolos ZigBee e pilha de protocolos ZigBee com o ciclo de trabalho dinâmico proposto neste trabalho. Em todos os experimentos, a cada 60 segundos, o módulo de medição de energia envia os dados de consumo e configuração através do módulo de comunicação sem fio.

Para execução dos experimentos uma mesma tabela de roteamento foi utilizada. A figura 6-(b) ilustra as rotas definidas. Escolheu-se o nó 7 para medir o tempo de vida da rede devido a sua maior quantidade de filhos na rede, o que acarreta em um número maior de envio e recepção de dados. Os dados foram coletadas a cada 30 minutos, através da montagem especificada na figura 5. O tempo de vida da rede, para tecnologia ZigBee

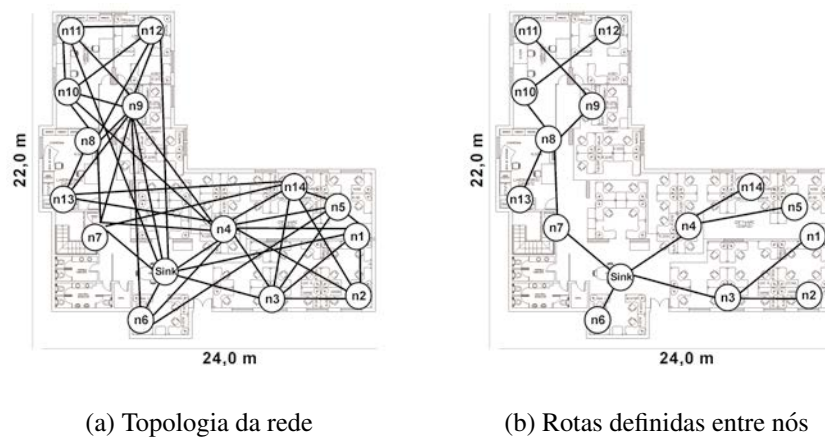


Figura 6. Topologia e Rotas

original, foi de 433 minutos ou 7 horas e 13 minutos. Um tempo de vida pequeno, mas dentro das expectativas devido às medições realizadas na seção 4.1.

Baseando-se na topologia apresentada na figura 6-(a) e na proposta para maximizar o tempo de vida da rede, calcula-se o valor do ciclo de trabalho. Considerando a velocidade de transmissão de 250 kbps do ZigBee, 516 bytes transmitidos, o valor do tempo (t) para cada transmissão realizada por enlace da rede será 0,016 segundos. Conforme apresentado na seção 3, o número total de transmissões realizadas na rede (e), considerando o pior caso, no qual não acontece transmissão simultânea entre os nós, é um somatório da profundidade de todos os nós da rede. Considerando a topologia apresentada na figura 6, o valor de e é igual a 31. De acordo com o experimento realizado na seção 4.2, a probabilidade de sucesso em cada transmissão é de $p = 0,991$. O tempo total necessário para enviar todos os pacotes será: $T = \frac{1}{p} \times t \times e = \frac{1}{0,991} \times 0,016 \times 31 = 0,50$ segundos.

Ao tempo T calculado é necessário acrescentar o tempo gasto para comunicação entre o módulo de medição de energia elétrica e o módulo de comunicação sem fio, pois conforme especificado na seção 3.1, a porta de comunicação RS232 dos medidores possui velocidade de 9600 bps. A comunicação entre os dois módulos acontece de forma simultânea em toda a rede, então basta acrescentar o tempo necessário para uma transmissão dos 516 bytes entre os módulos. De acordo com os dados apresentados, o valor de T será: $T = 0,50 + \left(\frac{516 \times 8}{9600}\right) = 0,93$ segundos.

Como cada nó deverá permanecer ligado durante 0,93 segundos em cada 60 segundos, conclui-se que o ciclo de trabalho para a rede analisada é de 1,55%. Estudos mostram que o valor de ciclo de trabalho deve ser próximos a 1% para que redes de sensores sejam funcionalmente viáveis [Guo et al. 2009].

O gráfico da figura 7 apresenta a comparação entre os tempos de vida das duas tecnologias analisadas, percebe-se que ao atingir 4,8 V o nó da rede para de comunicar, pois a plataforma XBee Série 2 - XB24BPK necessita de uma alimentação de no mínimo 4,8 V para o funcionamento de seus componentes. Observa-se também que o mecanismo para maximizar o tempo de vida da rede aumenta o tempo de vida em torno de 17 vezes. O tempo máximo de vida da rede será 126 horas e 30 minutos para o ZigBee com ciclo

de trabalho dinâmico, atendendo ao requisito da norma ABNT 14519 [ABNT 2011].

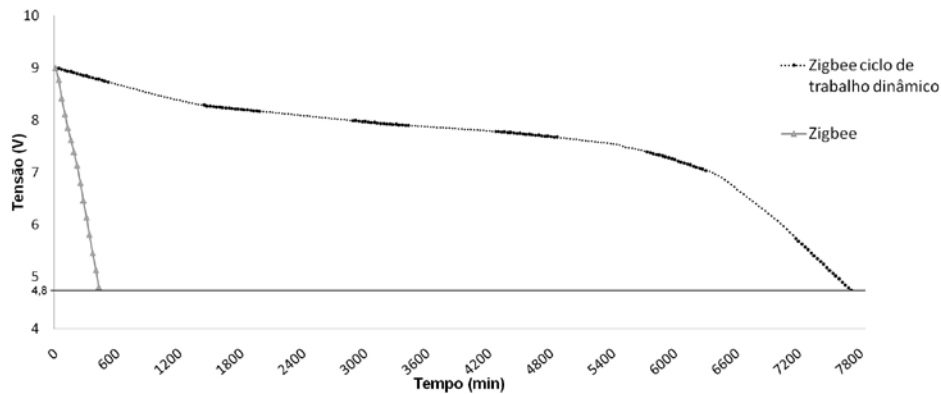


Figura 7. Tempo de vida das tecnologias analisadas

4.4. Percentual de Sucesso no Recebimento das Informações de Consumo

Para cálculo do percentual de sucesso no recebimento de informações de consumo utilizou-se a mesma topologia de rede dos experimentos de tempo de vida para o ZigBee com ciclo de trabalho dinâmico, ilustrada na figura 6-(b). Compara-se os dados recebidos pelo módulo concentrador de dados com os dados registrados nos módulos de medição de energia elétrica. Esse experimento foi realizado durante 8 horas, tempo maior que o utilizado no experimento da (seção 4.2), pois neste os envios são a cada minuto.

Os comandos trafegados na rede são compostos por 258 bytes. Os últimos dois bytes são utilizados para informar o CRC (*Cyclic Redundancy Check*) de 16 bits que é calculado sobre os demais 256 bytes. O módulo concentrador só considera um comando recebido como válido se o cálculo de seu CRC estiver correto.

Após as 8 horas de testes, o módulo de medição de energia elétrica interrompe o envio de comandos e o registro do envio dos mesmos. Em cada módulo de medição de energia elétrica da rede, registrou-se o envio de 958 comandos. Ao todo, 13.412 comandos foram enviados pela rede. O módulo concentrador de dados registrou o recebimento de 13.239 comandos corretos. Assim tem-se um percentual de sucesso de 98,7% no recebimento das informações.

5. Conclusões e Trabalhos Futuros

O trabalho apresentou uma aplicação prática para IoT nas redes elétricas inteligentes brasileiras utilizando as normas vigentes. A rede desenvolvida possibilitará um maior controle de gastos pelos clientes das concessionárias de energia elétrica devido ao acesso facilitado e em tempo real de consumo através de computadores na nuvem.

Importante ressaltar que, assim como as tecnologias, as normas brasileiras precisam avançar. Os protocolos definidos na norma ABNT 14522 são extremamente ultrapassados e rígidos. Propostas simples para diminuir o tráfego de dados na rede podem ser implementadas sem perda de confiabilidade. Agregação de dados e pacotes com tamanho variável, sem a obrigatoriedade de 258 bytes fixos, são exemplos.

Foi apresentada também uma metodologia para cálculo do ciclo de trabalho dinâmico cujo objetivo foi manter um ciclo de trabalho mínimo para leitura dos dados

de consumo, mesmo ocorrendo alterações na topologia da rede. A principal contribuição foi o sensível aumento do tempo de vida da rede. Com a utilização do ciclo de trabalho dinâmico, a rede passou a atender as normas vigentes. O ciclo de trabalho manteve-se dentro dos padrões definidos pela literatura, próximo à 1%. Além disso, a rede proposta mostrou-se eficiente quanto ao consumo máximo de um conjunto de medição definido nas normas vigentes para medição de energia elétrica. Considerando o percentual de sucesso nas transmissões realizadas, a ciclo de trabalho dinâmico mostrou-se eficiente atingindo um percentual de 98,7%, o que comprova a confiabilidade do sistema proposto.

Como trabalho futuro, pretende-se considerar, durante o cálculo do tempo necessário para enviar todos os pacotes, as transmissões simultâneas que poderão acontecer nos diversos enlaces da rede. Assim, o tempo necessário para enviar todos os pacotes pode ser diminuído. Outro trabalho futuro seria calcular o p dinamicamente. Esse valor seria utilizado no cálculo do ciclo de trabalho de forma que a solução proposta fosse adaptável às condições do meio de transmissão sem fio.

Referências

- ABNT (2008). *NBR 14522: Intercâmbio de informações para sistemas de medição de energia elétrica - Padronização*. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS ABNT, Rio de Janeiro.
- ABNT (2011). *NBR 14519: Medidores eletrônicos de energia elétrica (estáticos) - Especificação*. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS ABNT, Rio de Janeiro.
- Arif, A., Al-Hussain, M., Al-Mutairi, N., Al-Ammar, E., Khan, Y., and Malik, N. (2013). Experimental study and design of smart energy meter for the smart grid. In *Renewable and Sustainable Energy Conference (IRSEC), 2013 International*, pages 515 – 520, Ouarzazate. IEEE.
- Arnachalam, V., Jain, S., Kumar, N., Paventhan, A., Chinnaiyan, V., and Pradish, M. (2014). Survey on smart grid technologies- smart metering, iot and ems. In *Electrical, Electronics and Computer Science (SCEECS), 2014 IEEE Students' Conference on*, pages 1 – 6, Bhopal. IEEE.
- Bello, O. and Zeadally, S. (2014). Intelligent device-to-device communication in the internet of things. *Systems Journal, IEEE (Issue: 99)*, PP:1 – 11.
- Chen, X., Liu, J., Li, X., Sun, L., and Zhen, Y. (2011). Integration of iot with smart grid. In *Communication Technology and Application (ICCTA 2011), IET International Conference on*, pages 723 – 726, Beijing. IET - IEEE.
- Farahani, S. (2008). *Zigbee Wireless Networks and Transceivers*. Elsevier, Oxford, UK.
- Garpetun, L. (2013). Smart meters in operation center. In *Electricity Distribution (CIRED 2013), 22nd International Conference and Exhibition on*, pages 1 – 3, Stockholm. IET - IEEE.
- Guo, S., Gu, Y., Jiang, B., and He, T. (2009). Opportunistic flooding in low-duty-cycle wireless sensor networks with unreliable links. In *MobiCom '09: Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 133–144, New York, NY, USA. ACM.
- Ou, Q., Zhen, Y., Li, X., Zhang, Y., and Zeng, L. (2012). Application of internet of things in smart grid power transmission. In *Mobile, Ubiquitous, and Intelligent Computing (MUSIC), 2012 Third FTRA International Conference on*, pages 96 – 100, Vancouver, BC. IEEE.
- Spano, E., Niccolini, L., Pascoli, S., and Iannaccone, G. (2014). Last-meter smart grid embedded in an internet-of-things platform. *Smart Grid, IEEE Transactions on (Volume:PP, Issue: 99)*, PP:1.
- Yun, M. and Yuxin, B. (2010). Research on the architecture and key technology of internet of things (iot) applied on smart grid. In *Advances in Energy Engineering (ICAEE), 2010 International Conference on*, pages 69 – 72, Beijing. IEEE.

Um Veículo Aéreo Não Tripulado para Encontrar Focos de Poluição

Marcelo Maia da Silva¹, Heitor Freitas², Bruno Faiçal²,
Gerson Serejo³, Jó Ueyama², Gustavo Pessin^{1,3}

¹Instituto de Ciências Exatas e Naturais
Universidade Federal do Pará (UFPA) – Belém, PA, Brasil

²Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo (USP) – São Carlos, SP, Brasil

³Laboratório de Computação Aplicada
Instituto Tecnológico Vale – Belém, PA, Brasil

marcelomaia@ufpa.br, {heitorfv,bsfaical,joueyama}@icmc.usp.br
gerson.serejo@pq.itv.org, gustavo.pessin@itv.org

Resumo. *O presente trabalho tem como objetivo o desenvolvimento de um Veículo Aéreo Não Tripulado (VANT) para encontrar focos de poluição. Um protótipo de baixo custo foi construído para obtenção de concentração de poluentes (CO) e variáveis ambientais (umidade e temperatura) e duas metodologias de coleta em voo foram empregadas para coleta dos dados. Os testes de coleta foram realizados na área do Campus 2 da USP/ICMC. Como resultado, foi identificada a área com maior concentração de calor por meio de um algoritmo de interpolação. Por fim, notou-se que o protótipo de baixo custo foi capaz de monitorar poluentes atmosféricos e variáveis ambientais. Nota-se que o modelo serve também para outros tipos de medições como de agrotóxicos e radiações. Como contribuição deste trabalho, tem-se uma alternativa barata para monitoramento de poluentes direcionadas para contextos mais específicos.*

1. Introdução

Normalmente, estações de monitoramento de qualidade do ar tipicamente instaladas por agências de proteção ao meio ambiente são equipamentos caros, localizados em posições fixas e responsáveis por monitorar grandes áreas [Devarakonda et al. 2013]. Entretanto, veículos aéreos não tripulados (VANTs) podem ser empregados como estações móveis, a fim de permitir a percepção de variáveis de interesse em regiões mais específicas. Além disso, VANTs podem ser mais baratos e apresentam flexibilidade no planejamento do voo com uma maior possibilidade de aproximação a um alvo de interesse [Yan et al. 2010]. Este trabalho propõe uma forma barata e flexível de análise de qualidade de ar, onde dois modelos de coleta (que são complementares) são propostos para encontrar focos de poluição. Com o uso do kit móvel proposto neste trabalho, coletas e análises para o entendimento dos padrões das poluições serão facilitadas para diversas localidades e contextos. O kit apresentado é um dispositivo de computação móvel transportado num VANT, que é composto por sensores que coletam variáveis climáticas (como temperatura e umidade) e poluentes atmosféricos (CO).

Este trabalho tem a seguinte estrutura: a Seção 2 apresenta uma revisão da literatura na área de qualidade de ar e VANTs, visando relacionar trabalhos da literatura

com este; na Seção 3, é descrito o modelo proposto para o monitoramento, detalhando o protótipo de *hardware*, *software* e as metodologias de coleta. Na sequência, a Seção 5 apresenta os resultados alcançados e, finalmente, na Seção 6, são feitas as considerações finais e os trabalhos futuros.

2. Trabalhos Relacionados

Nesta Seção são apresentados trabalhos que usam VANT, com a finalidade de compará-los à esse trabalho. Por fim, é apresentada uma discussão sobre os trabalhos correlatos. Faiçal et al. [Faiçal et al. 2014] descrevem uma arquitetura orientada ao problema de auto ajuste de rotas do VANT ao pulverizar produtos químicos em uma plantação. Onde um algoritmo é proposto e avaliado (com o uso de algoritmo de agrupamento DBSCAN), considerando o retorno de informação oriunda de uma rede de sensores sem fio simulada, variações de vento em força e direção, observou-se que a arquitetura em questão pode notavelmente reduzir o desperdício de pesticidas e fertilizantes. São evidenciados na Figura 1 os testes feitos em simulação, onde foram usados algoritmos de agrupamento para determinar os locais adequados para aplicação de pesticida.

Roldán et al. [Roldán et al. 2015] criaram e validaram uma plataforma sensorial móvel para monitoramento de estufas. A plataforma consiste um conjunto de sensores acoplados a um quadrotor, os objetivos desse sistema incluem capturar medidas de concentração de temperatura, umidade, luminosidade e CO₂ e gerar mapas dessas variáveis. O posicionamento dos sensores foram determinados por meio de um estudo de aerodinâmica do quadrotor e a influência dos fluxos de ar e de seus motores. A contribuição primária do trabalho é a determinação do lugar ótimo para cada sensor, considerando o fluxo de ar. No trabalho em questão, foi mapeada a concentração de calor dentro de uma estufa como mostra a Figura 2.

Caltabiano et al. [Caltabiano et al. 2005] conduziram o projeto ROBOVOLC para exploração e análise de fenômeno vulcânico sem nenhuma interação humana. O veículo aéreo não tripulado teve de ser autônomo. Devido a duas dificuldades: (i) distância segura da pluma (nuvem) vulcânica e (ii) dificuldade de identificação visual do veículo.

2.1. Discussão sobre os trabalhos relacionados

Observando os trabalhos relacionados, verifica-se que análises de qualidade do ar com sensores e mapeamentos têm sido implementados, contudo os estudos apresentados possuem lacunas. Faiçal et al. [Faiçal et al. 2014] implementaram uma arquitetura de auto ajuste de rota de VANT para pulverização de produtos químicos, com geração de mapas de concentração de pesticidas, todavia, a validação da arquitetura foi em ambiente simulado. Roldán et al. [Roldán et al. 2015] implementaram uma plataforma sensorial para estufas, definiram através de simulação os locais ótimos dos sensores em um quadrotor, contudo, o experimento não foi realizado considerando as variações de altura. Caltabiano et al. [Caltabiano et al. 2005], implementaram um VANT para análise de pluma vulcânica, porém, não foram implementados mapas de concentração de poluentes. Por estas razões, este trabalho implementa um estudo em campo com dados reais coletados por um VANT, com a finalidade de encontrar poluição em uma determinada área ou altura.

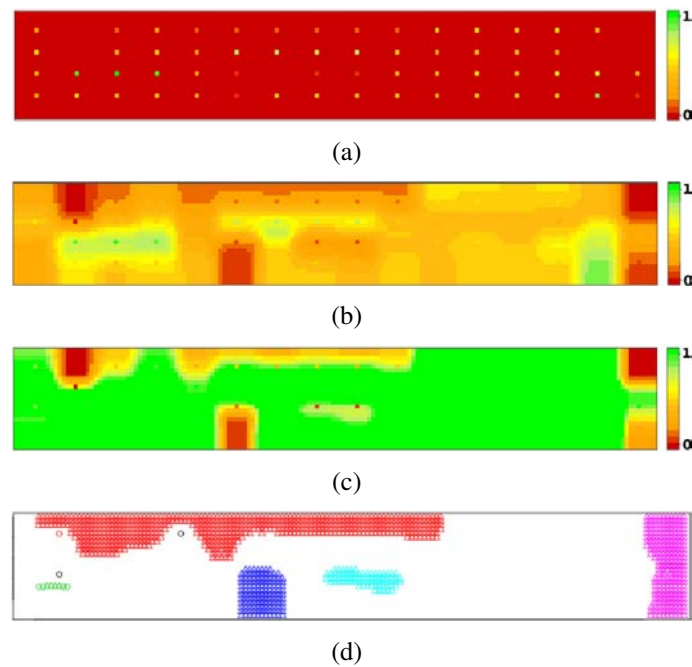


Figure 1. (a) Simulação de um campo de plantação representado numa matriz 2D apresentado por [Faiçal et al. 2014]. A variação de vermelho para verde significa menos pesticidas recebidos. A maior parte da área está vermelha por que os sensores são pontuais. A distribuição de pesticidas está desigual devido à variação de vento na simulação. (b) Mapa após aplicação de técnica de interpolação. (c) Mapa após a aplicação de um limiar, que determina onde há menos pesticida. (d) Identificação por agrupamento usando o algoritmo DBSCAN .

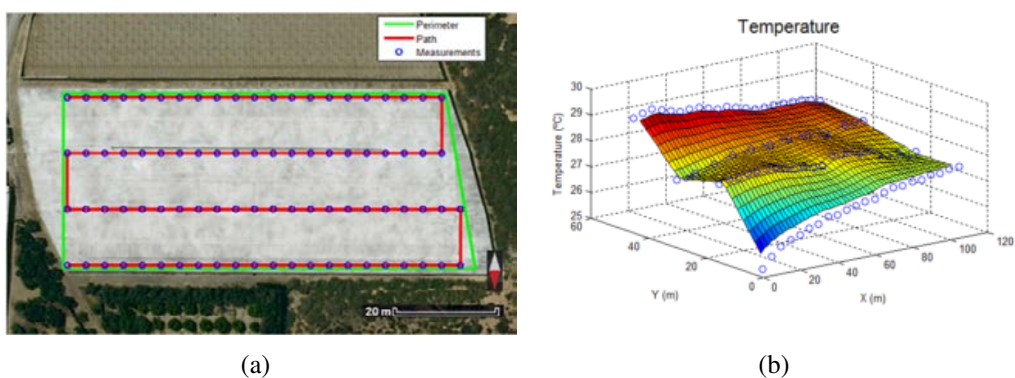


Figure 2. (a) O caminho de coleta do quadrotor dentro da estufa, os pontos no traçado são os locais de coleta. (b) Mapa de temperatura dentro da estufa [Roldán et al. 2015]

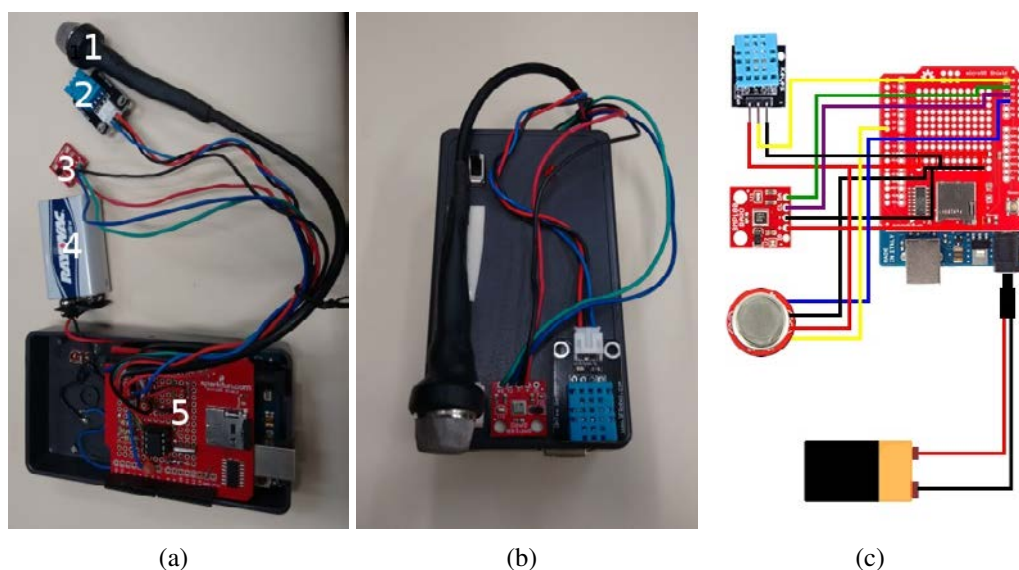


Figure 3. (a) Protótipo de coleta de CO: (1) detector de CO MQ-2, (2) sensor de temperatura e umidade DHT11, (3) barômetro BMP180 (4) bateria 9v e (5) Arduino Uno, DS1307 e cristal 32.768 usados para o relógio do Arduino e SD-card. (b) Protótipo fechado. (c) Esquema de circuito do protótipo.

3. Materiais e métodos

Foram implementados um protótipo de baixo custo e duas metodologias de coletas, ambos descritos a seguir.

3.1. Protótipo

O protótipo de coleta de poluentes e variáveis ambientais de baixo custo (Figura 3), consiste de um Arduino UNO, dois possíveis sensores de monóxido de carbono, sendo MQ-2 e MQ-7, um sensor de umidade e temperatura DHT11, um sensor de pressão e temperatura BMP180, um *shield* SD, um *timer* para ter acesso à hora e por fim o *software* de controle do protótipo. Esse protótipo móvel foi acoplado a um VANT (Figura 4), em colaboração com o ICMC-USP.

A Figura 4(c) mostra o VANT MikroKopter Okto¹ com vista de cima, o módulo UAV *Brain* que é o dispositivo computacional do VANT é exposto na Figura 5 com detalhes de sua implementação. O ZigBee, das Figuras 5(a)i e 5(a)ii, é o responsável pela comunicação sem fio, o Arduino, em conjunto com a Power Meter Shield (Figuras 5(b)ii e 5(b)iv respectivamente), atuam no monitoramento do consumo de energia elétrica do ZigBee, há um sensor DHT22 de temperatura e umidade para registro de dados climáticos 5(c)v. O Raspberry Pi, exposto na Figura 5(b)iii, é o núcleo do agente computacional UAV *Brain*, ao qual são delegadas as funções computacionais do VANT como registro de logs. O objetivo da integração do VANT com o protótipo móvel de coleta é implementar dois modelos de coleta: (i) zigue-zague e (ii) subida/descida, os quais serão explicados a seguir.

¹O VANT utilizado neste trabalho foi disponibilizado pelo pesquisador colaborador desta pesquisa Jó Ueyama, do ICMC-USP. O mesmo foi utilizado durante uma visita técnica a USP entre 15/06 até 04/07 de 2015.

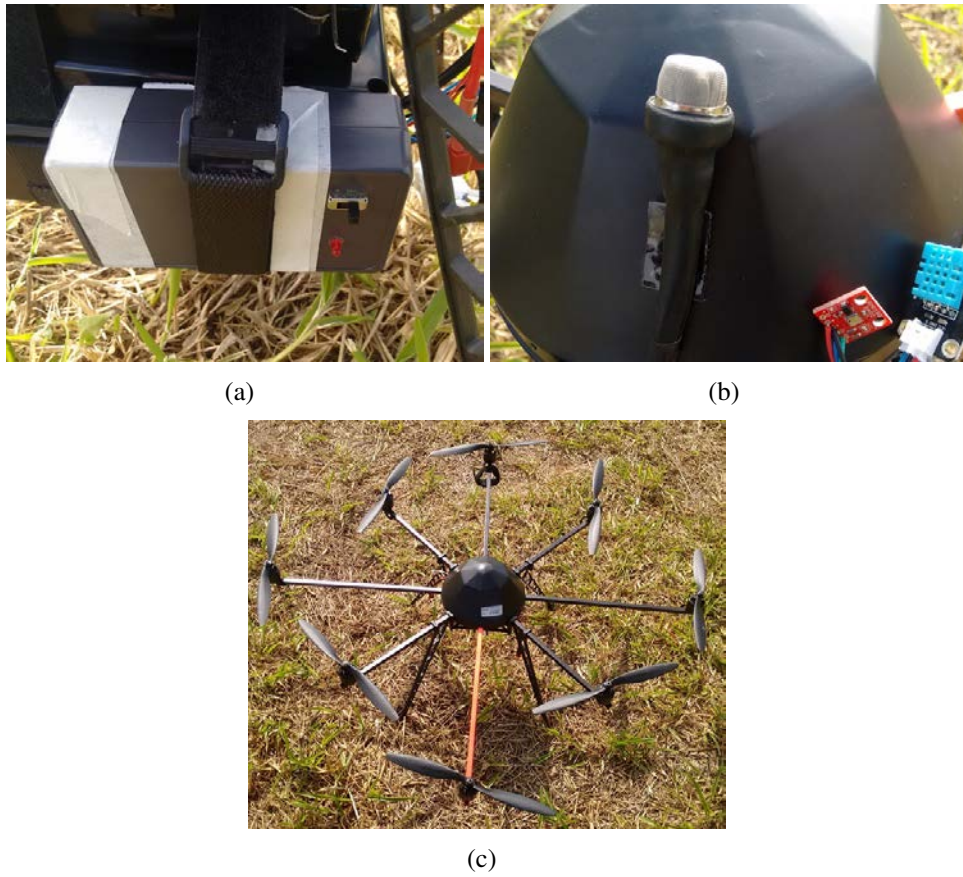


Figure 4. (a) O controlador Arduino em uma caixa de proteção. (b) Sensores MQ-2, BMP180 e DHT11 (esquerda para direita). (c) VANT MikroKopter Okto, com oito hélices classe III.

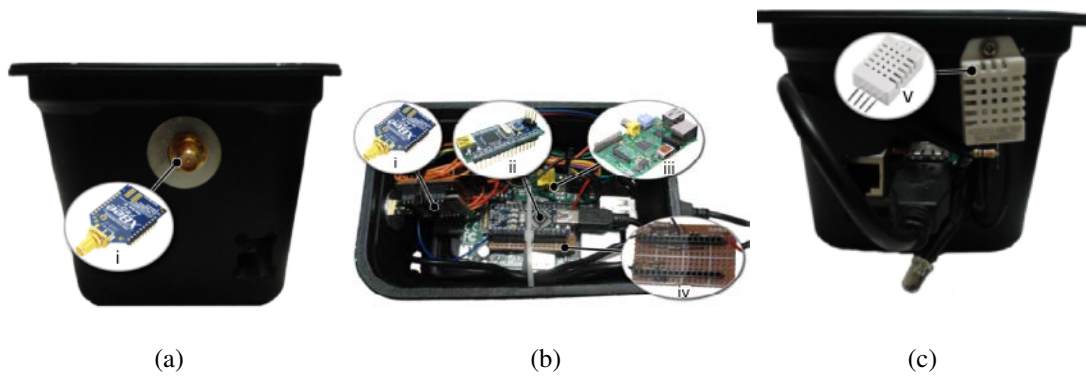


Figure 5. (a) Vista frontal. (b) Vista superior. (i) Rádio transceptor ZigBee. (ii) Arduino Nano. (iii) Raspberry Pi. (iv) Power Meter Shield [Ueyama et al. 2014].

4. Metodologia de coleta e análise empregada em plataforma móvel (VANT)

Foram elaborados dois planos de voo: (i) zigue-zague e (ii) variação de altura, o modelo de coleta em zigue-zague consiste de um voo em forma de zigue-zague em uma altura fixa a fim de coletar dados em uma determinada área. O modelo de subida/descida consiste de subir ou descer com paradas periódicas para coleta de dados, a fim de detectar as maiores concentrações de poluentes em uma determinada altura. Foi verificado por [Roldán et al. 2015] que o melhor lugar para a localização dos sensores que sofrem com a influência de vento é no topo do centro do VANT, desta forma, é indicado que sensores de gases, temperatura e umidade fiquem nesta posição. Os sensores de temperatura e umidade devem ficar afastados de fontes de calor, como o sensor de CO MQ-2 (mostrado no centro da Figura 4(b)) que aquece. Este método de coleta tem como objetivo realizar reconhecimento de concentração de poluentes ou variáveis climáticas. Após um voo inicial, o VANT pode indicar uma altura ou uma área específica onde deve-se realizar um outro voo de reconhecimento. O modelo propõe mapear e encontrar focos de poluição, contudo, observa-se que além de poluentes podem ser medidos quaisquer variáveis, como pesticidas e radiações, desde que sejam consideradas as interferências de fatores durante o voo, como vento, calor, umidade, dentre outros, nas leituras dos sensores.

5. Experimentos e resultados

Os dois modelos de coleta (zigue-zague e com variação de altura) foram implementados usando o dispositivo móvel acoplado ao VANT, devido a necessidade de realizar os voos afastados de um centro urbano (pois o VANT representa risco de morte). As variações de CO foram mínimas, desta forma, foram analisados os dados de temperatura, especificamente com o sensor BMP180, pois, houveram variações de temperatura ao longo do voo. A taxa de leitura do sensor BMP180 é 12 Hz, como a velocidade média do VANT nestas coletas foram de 3 m/s o sensor BMP180 registrou na média 4 leituras a cada metro, que possibilitou uma boa amostragem por espaço.

5.1. Coleta em zigue-zague

A coleta (i) zigue-zague destacada na Figura 6, foi realizada em uma altura fixa de 25 m para cobrir uma área de $1600 m^2$ percorrendo 1039 m em 5 min e 46 s. Foram desconsiderados os dados coletados durante a subida e descida, pois não faziam parte deste modelo de coleta. Com os dados obtidos, que são pontos de latitude, longitude com as medidas de CO, temperatura e umidade para cada ponto, foi aplicado um método de interpolação linear², objetivando encontrar concentrações de dados coletados, neste caso a temperatura. Como resultado desta interpolação dos dados, foi obtido o gráfico da Figura 6(a), o qual mostra as concentrações de temperatura (maiores temperaturas tendem ao vermelho e menores ao azul). Voos em sequência podem ser realizados em áreas específicas, com uma maior densidade de pontos em um menor espaço, com o objetivo de mapear e entender com maior grau de resolução poluentes/variáveis climáticas de uma determinada área.

²Algoritmo de interpolação linear, especificamente o método `LinearNDInterpolator()` do pacote `scipy` da linguagem Python. A interpolação é construída por triangular os dados de entrada com `Qhull` e, em cada triângulo realiza uma interpolação baricêntrica linear.

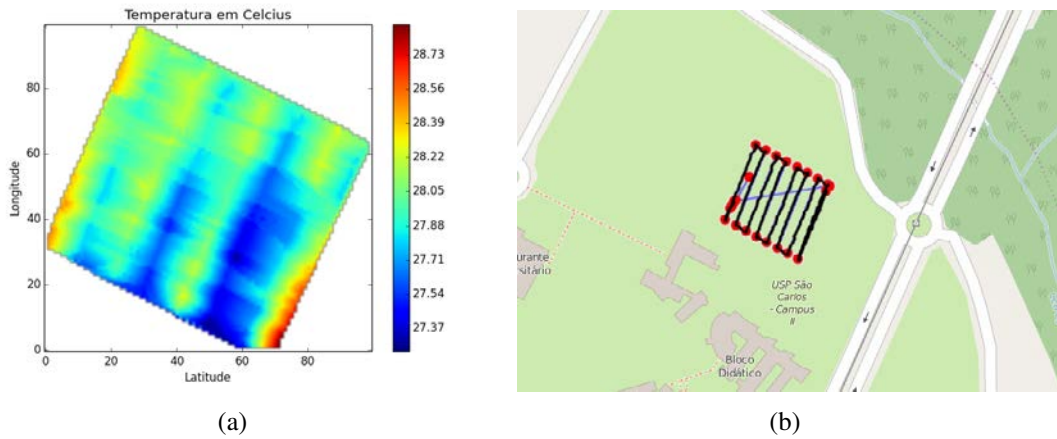


Figure 6. (a) Mapa de concentração de temperatura com o sensor BMP180. (b) USP São Carlos campus 2, área onde foi realizada a coleta e o conjunto de *way-points* usados pelo VANT em conjunto com seu *software* de controle para realizar a coleta.

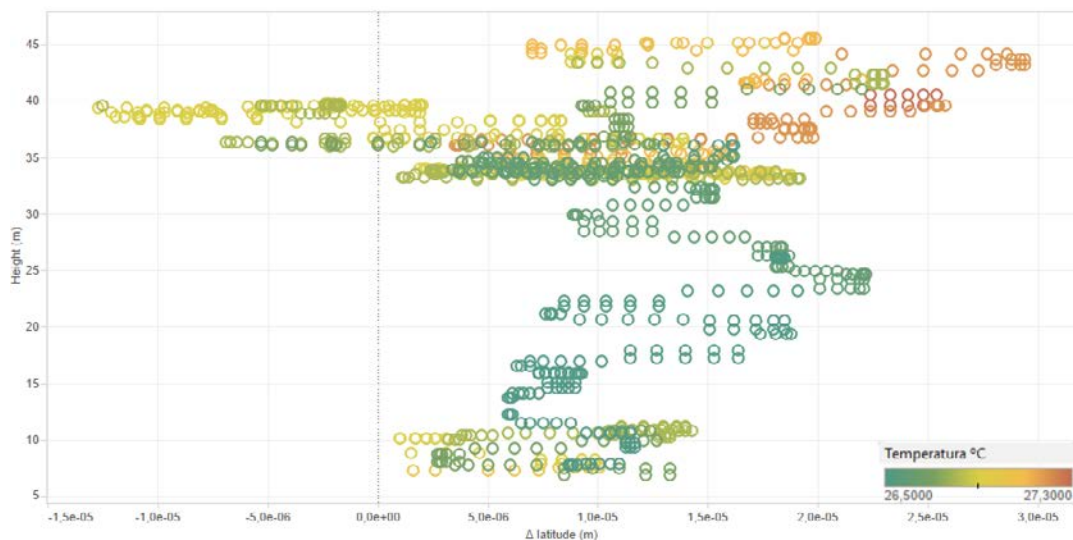


Figure 7. Dados coletados de temperatura considerando a variação de altura no ICMC campus 2. A cor dos pontos indicam a temperatura num dado ponto de coleta. O eixo x é a variação do VANT em relação ao seu ponto inicial na latitude, o eixo y diz respeito a altura. Podemos ver que a concentração de temperatura mostrou-se maior na altura de 40 m, também que a variação em torno do eixo x foi maior devido ao vento ser mais forte de acordo com a altura.

5.2. Coleta com variação de altura

A coleta (ii) com variação de altura evidenciada na Figura 7 foi realizada com uma variação de 5 m a 40 m de altura num total de 2417 leituras em 3.3 min, onde a cada 10 m houve uma parada de 30 s para coletar dados de temperatura. O eixo x é referente ao desvio do VANT na latitude em relação ao seu ponto inicial, o eixo y é referente à altura. Nesta coleta foi usado o sensor BMP180 para medir as variações de altura e temperatura ao longo do voo, percebe-se que a concentração de temperatura e a variação em torno do eixo x mostraram-se maiores na altura de 40 m (a variação em torno do eixo x é maior em função do vento). O VANT poderia ser então empregado para encontrar a altura que apresenta maior concentração de poluentes e a partir disso realizar o voo em zigue-zague. Com este modelo de coleta é possível ter uma alternativa aos balões meteorológicos, onde o VANT apresenta uma flexibilidade de voo que o balão não possui.

6. Considerações finais e trabalhos futuros

Neste estudo foi proposto um protótipo móvel de baixo custo para análise de qualidade de ar, objetivando entender as variações de poluentes em diferentes contextos. O protótipo foi acoplado a um VANT e testes com dois métodos de leitura foram realizados. Informações sobre monitoramento de qualidade de ar não são fáceis de serem obtidas; e mesmo onde há monitoramento, como se pode ver nos relatórios da CETESB [CETESB 2014], as análises são feitas em resolução de cidade/bairro. Generalizando um amplo espaço, são análises que não apresentam refinamento de localização. Comparando nosso trabalho com a abordagem tradicional, aplicada por agências ambientais oficiais, o presente trabalho mostrou a capacidade de monitorar os poluentes em áreas de interesse mais específicos de forma mais prática.

Como resultados, foram realizadas duas caracterizações de uma localização dentro do ICMC Campus 2, onde foram analisadas as concentrações de temperatura (não houve variação de CO neste ambiente de coleta), como dito anteriormente. O modelo proposto pode servir para encontrar concentração de poluentes, agrotóxicos, radiação, ilhas de calor, etc. Uma observação importante é que se deve considerar o posicionamento dos sensores e os fatores durante o voo que podem afetar em suas leituras.

6.1. Trabalhos Futuros

Diversas outras atividades ainda são previstas, entre elas, a integração do protótipo móvel de coleta com GPS (usou-se o GPS do VANT), análises em tempo real para encontrar focos de poluição em tempo de voo e a aquisição de outros sensores como de pesticidas, radiação, e de material particulado. Diversos são os pontos em abertos neste tipo de pesquisa. Em termos de inovação científica, existe um vasto campo a ser explorado na área, tanto em desenvolvimento de *hardware* como no desenvolvimento de *software*. Em termos de monitoramento ambiental, a pesquisa pode ter um desdobramento mais específico, como no mapeamento de concentração de pesticidas, de material particulado e até de radiação.

7. Agradecimentos

Os autores gostariam de agradecer aos colegas Bruno Ferreira (UFPA), Eduardo Carvalho (ITV), Leon Cardoso (ITV) e Geraldo Pereira (USP) por inspirações e auxílios no desen-

volvimento do trabalho. Ainda, os autores gostariam de agradecer ao apoio financeiro recebido através da Chamada 59/2013 MCTI/CT-Info/CNPq, processo 440880/2013-0.

References

- [Caltabiano et al. 2005] Caltabiano, D., Muscato, G., Elettrica, I., Doria, V. A., Federico, C., Giudice, G., Palermo, S., and Malfa, V. U. L. (2005). Architecture of a UAV for volcanic gas sampling. 1:739–744.
- [CETESB 2014] CETESB (2014). Relatório de qualidade do ar no estado de São Paulo.
- [Devarakonda et al. 2013] Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L., and Nath, B. (2013). Real-time Air Quality Monitoring Through Mobile Sensing in Metropolitan Areas.
- [Façal et al. 2014] Façal, B. S., Costa, F. G., Pessin, G., Ueyama, J., Freitas, H., Colombo, A., Fini, P. H., Villas, L., Osório, F. S., Vargas, P. a., and Braun, T. (2014). The use of unmanned aerial vehicles and wireless sensor networks for spraying pesticides. *Journal of Systems Architecture*, 60(4):393–404.
- [Roldán et al. 2015] Roldán, J., Joossen, G., Sanz, D., del Cerro, J., and Barrientos, A. (2015). Mini-UAV Based Sensory System for Measuring Environmental Variables in Greenhouses. *Sensors*, 15(2):3334–3350.
- [Ueyama et al. 2014] Ueyama, J., Freitas, H., Faical, B. S., Filho, G. P., Fini, P., Pessin, G., Gomes, P. H., and Villas, L. A. (2014). Exploiting the use of unmanned aerial vehicles to provide resilience in wireless sensor networks. *IEEE Communications Magazine*, 52(12):81–87.
- [Yan et al. 2010] Yan, L., Gou, Z., and Duan, Y. (2010). A uav remote sensing system: Design and tests. In *Geospatial Technology for Earth Observation*, pages 27–44. Springer.

Análise de Padrões Espaciais e Temporais da Mobilidade de Táxis em San Francisco e Roma

Cristiano Martins Monteiro¹, Fábio Rocha da Silva¹, Cristina Duarte Murta¹

¹Departamento de Computação – CEFET-MG

Resumo. A disponibilidade de registros de localização e tempo de serviços de táxi possibilita a análise de padrões de mobilidade em centros urbanos e melhor compreensão das rotinas da população local. Este trabalho utiliza um conjunto de dados com mais de trinta milhões de registros de localização e tempo de táxis de San Francisco e Roma para comparar padrões de mobilidade urbana nessas duas metrópoles. As rotas anômalas dos táxis em ambas as cidades foram encontradas e discutidas, assim como as rotas mais frequentes em San Francisco. Os resultados indicam a viabilidade do algoritmo utilizado para corrigir as distâncias das rotas com erros de GPS em que o táxi estava parado.

Abstract. The availability of location and time records from taxi services enables the analysis of mobility patterns in urban centers and better understanding of the routines of the local population. This work uses a data set with over thirty million location and time records of taxis of San Francisco and Rome to compare urban mobility patterns in both metropolises. The anomalous routes of taxis in both cities were found and discussed, as well as the most frequent routes in San Francisco. The results show the feasibility of the algorithm used for correct the routes's distances with GPS errors when the taxi was stopped.

1. Introdução

A proliferação do uso de dispositivos digitais móveis, em particular com habilitação de localização, produz rastros digitais que podem ser agrupados para estudos de padrões de mobilidade urbana. Pesquisas recentes utilizam dados urbanos abertos para entender e observar padrões nas rotinas diárias das pessoas, compondo modelos de mobilidade em grandes centros urbanos. Tais padrões são identificados por meio de análises espaciais e temporais dos dados. Esses dados são normalmente obtidos via dispositivos GPS, redes sociais, histórico de chamadas de celulares, circulação de veículos em geral ou trajetórias de transporte público, dentre outros.

As pegadas digitais dos táxis oferecem uma visão rica e detalhada da dinâmica da movimentação da população de uma cidade. Diferentemente dos ônibus e metrô, os trajetos de táxis não estão limitados a itinerários e horários fixos. Dessa forma, as rotas dos táxis representam bem as dinâmicas e costumes de determinada região [Castro et al. 2013]. Este trabalho utiliza dados de táxi das metrópoles de San Francisco, nos EUA, e Roma, na Itália, com o objetivo de explorar os padrões espaciais e temporais da mobilidade de táxis nessas cidades. As análises nos permitem visualizar as áreas de ambas as cidades com maior operação de táxis; encontrar e analisar rotas anômalas; calcular as distâncias percorridas pelas rotas de táxi; listar as rotas mais frequentes dos táxis com passageiro em San Francisco; em suma, inferir, pelos dados, o

funcionamento dos serviços de táxis nestas cidades ao longo do dia e da semana, inclusive padrões culturais.

Para inferir padrões de mobilidade, foram analisados mais de 30 milhões de registros espaço-temporais coletados durante um mês em cada cidade. Entendemos que se refere a uma análise de *Big Data* pois, além do volume de dados, trata-se de análise tridimensional de dados em espaço (localização georreferenciada) e tempo, os quais são complexos para se trabalhar. Uma das dificuldades de trabalhar com dados de tempo e de localização está na necessidade de se tratar os erros gerados por imprecisões nas coletas de dados e no cálculo das distâncias. Esse desafio é evidenciado ao calcular a distância percorrida por um táxi levando o passageiro ao seu destino, ou circulando até encontrar um novo passageiro. Nesse segundo caso, o taxista pode decidir estacionar o táxi por um determinado período, aguardando no ponto até que um passageiro surja. Uma vez que as localizações de latitude e longitude registradas pelo sistema de GPS apresentam erros [Valero et al. 2014], os quais geram localizações aparentemente instáveis, é importante que as distâncias percorridas pelos táxis sejam corrigidas. Para contornar esse problema, aplicamos um algoritmo para eliminar essas distorções antes de calcular as distâncias das rotas. Os resultados indicam a viabilidade dessa inferência no tratamento dos dados.

A descoberta dos padrões de mobilidade dos táxis e das rotinas diárias dos taxistas e cidadãos que usam o serviço de táxi pode ajudar as próprias agências de táxi ou mesmo serviços concorrentes, tais como o *Uber*. A informação pode beneficiar também comerciantes locais, empresas de turismo, setores de planejamento de transporte urbano, podendo produzir estimativas de fluxo de pessoas e veículos nas regiões.

A seção seguinte discute os trabalhos relacionados; a seção 3 apresenta as bases de dados utilizadas neste trabalho e a metodologia; a seção 4 aborda as análises espaciais, temporais e rotas frequentes de San Francisco e Roma; e a seção 5 finaliza o trabalho.

2. Trabalhos Relacionados

O rápido desenvolvimento das tecnologias da informação tem gerado uma grande massa de dados de trajetórias espaço-temporais, que representam a mobilidade de uma diversidade de objetos como pessoas, veículos e animais [Zheng et al. 2015]. Em anos recentes, o estudo da mobilidade humana foi impulsionado principalmente pelo crescimento expressivo do volume de dados georreferenciados resultante do aumento do uso de sistemas de navegação, como o GPS [Siła-Nowicka et al. 2015, Kellner and Egger 2016].

Vários trabalhos recentes utilizam traços de GPS obtidos por trajetórias de táxis a fim de descobrir padrões de dinâmicas comunitárias [Chen 2014]. Em [Castro et al. 2013] é apresentada uma vasta pesquisa sobre o uso de traços de GPS obtidos principalmente por táxis. Os autores organizaram os trabalhos da área em três categorias principais: dinâmicas sociais, dinâmicas de tráfego e dinâmicas operacionais. A categoria “dinâmicas sociais” lista trabalhos sobre o comportamento da população de uma cidade baseando em sua mobilidade. As “dinâmicas operacionais” apresentam estudos sobre a operação dos taxistas nas ruas. E as “dinâmicas de tráfego” se referem ao fluxo dos veículos nas redes das estradas.

Outro objetivo do estudo da circulação dos táxis nas estradas é avaliar o funcionamento e desempenho de VANETs (*Vehicular Ad hoc Networks*). No trabalho

de [Amici et al. 2014] é avaliado o desempenho de um protocolo epidêmico de roteamento de mensagens em VANETs comparando os resultados de simuladores de mobilidade com as bases de dados de San Francisco, de Roma, e de Shenzhen, na China. Em [Cunha et al. 2015] são analisados padrões de conectividade possíveis entre os carros que compõem as redes veiculares no decorrer do dia e são aplicadas técnicas estatísticas, de teoria dos grafos e de análise de redes nas bases de dados de San Francisco e Roma.

Padrões de comportamento dos taxistas e das rotas de táxis foram estudados por [Chen 2014]. O autor apresenta três principais contribuições: um algoritmo para identificar trajetórias anômalas dentre as demais com o mesmo local de origem e destino; um algoritmo que encontra regiões candidatas a novos pontos de ônibus a partir dos locais de início e fim das rotas de táxi noturnas; e um *framework* para gerar passeios personalizados, interativos e conscientes em relação ao tráfego de veículos de uma cidade.

A procura de passageiros pelos taxistas assim como a relação de táxis ocupados e desocupados e a renda desses taxistas também são temas de pesquisa recentes [Li et al. 2011, Zhang 2014]. O trabalho de [Murashkin and Geroliminis 2015] estima a quantidade de táxis procurando passageiros pelas ruas da cidade de Shenzhen, China. Os autores também avaliam o comportamento dos taxistas nas ruas diferenciando a velocidade e distância média percorrida com os táxis ocupados e procurando passageiros. Em [Liu et al. 2015] é estudada a decisão dos taxistas sobre como conseguir um passageiro. Os resultados demonstram, por exemplo, que taxistas experientes preferem estacionar o veículo e esperar os passageiros aparecerem que rodar pelas ruas procurando-os. No trabalho de [Ding et al. 2015] são analisados padrões de taxistas em Xangai, China, ao procurar passageiros ou parados, esperando clientes. Os autores também estimaram a renda obtida pelas corridas de táxi. Em [Zhu et al. 2012], os autores utilizam padrões históricos das rotas de táxi em uma base de dados de Beijing para inferir o *status* de ocupação dos táxis e momentos em que o veículo ficou estacionado.

O presente trabalho se difere dos trabalhos relacionados nos seguintes aspectos: inferimos os momentos em que os táxis estavam parados para reduzir erros no cálculo das distâncias; discutimos estatísticas das distâncias das rotas de San Francisco; avaliamos diferenças quantitativas entre rotas com o táxi ocupado e com o táxi desocupado; e contextualizamos os resultados com o fato dos dados de San Francisco serem sobre táxis e dos dados de Roma serem sobre taxistas.

3. Bases de Dados e Metodologia

Neste trabalho foram utilizadas duas bases de dados de movimentação de táxis, uma da cidade de San Francisco¹, EUA, abordada em [Piorkowski et al. 2009]; e outra de Roma², Itália, apresentada em [Bracciale et al. 2014]. Tais bases foram escolhidas por terem conteúdos semelhantes e estarem disponíveis para *download*. Não foram encontradas bases de dados similares de cidades do Brasil.

A base de San Francisco é composta por 536 arquivos, um para cada táxi, somando 11.219.955 linhas. Cada linha contém um registro com os seguintes dados: a identificação do táxi; a localização georreferenciada (latitude e longitude); o *status* de ocupação (1 para

¹<http://crawdad.org/epfl/mobility/20090224/>

²<http://crawdad.org/roma/taxi/20140717/>

táxi com passageiro e 0 para táxi sem passageiro); e o *timestamp* no formato *Unix Epoch* do momento de aquisição desses dados. A base de dados de Roma possui 21.817.851 registros em um único arquivo sobre 316 taxistas diferentes contendo, para cada registro, a identificação do taxista; a localização georreferenciada; a data e a hora de aquisição.

Utilizamos o termo “ocupado” para representar um táxi com passageiro; “desocupado” para representar um táxi sem passageiro; “registro” ao se referir a cada linha das bases de dados; “trecho” para definir o movimento do táxi a cada dois registros consecutivos da base de dados; e “rota” para representar uma sequência de registros com o mesmo *status* de ocupação. É importante ressaltar que na base de dados de San Francisco há a identificação dos táxis e não dos taxistas. Já em relação à Roma, há a identificação dos taxistas mas não há a identificação dos táxis nem um registro informando quando o táxi estava ocupado ou desocupado. Os trechos de San Francisco são registrados a cada 60 segundos em média e o intervalo médio entre os registros de Roma é de 7 segundos.

Ambas as bases de dados foram filtradas para eliminar registros que aparentemente apresentam erros de localização e para tratar possíveis erros das distâncias das rotas. As rotas de táxi com medidas ou localizações muito distintas das demais são denominadas como “rotas anômalas”. Foram retirados todos os registros posicionados fora de uma região com estradas (localizados no mar, por exemplo); e registros fora da região das cidades de Roma ou San Francisco e que não formam uma trajetória.

Assim, foram excluídos dez registros da base de dados de San Francisco localizados em um mesmo ponto da montanha canadense chamada “Stein Valley Nlaka’pamux Heritage Park”. Esses registros foram obtidos em momentos e veículos diferentes e se distanciam num raio de aproximadamente 1.400 km do próximo registro não anômalo. Também foi excluído um registro de uma rota anômala com o táxi ocupado dentro da cidade de San Francisco. Neste caso, o táxi teria se deslocado num raio de 600 km em 75 segundos, alcançando o estado de Nevada, e retornado outros 600 km até a cidade de San Francisco nos próximos 49 segundos. Outros 585 registros foram retirados somente da análise das distâncias das rotas devido a API do Google Maps não conseguir aproximar o ponto para uma estrada ou encontrar uma rota de carro possível para o trecho. Após esse tratamento, a base de dados de San Francisco perdeu 897 registros, o equivalente a menos de 0,008%.

Em relação à base de Roma, foram retirados 18 registros de uma rota anômala com o taxista parado num ponto localizado em Londres. Nesse caso, passou tempo o suficiente para terem deslocado o localizador GPS da Itália até a Inglaterra, porém, essa rota foi excluída por não haver uma trajetória ligando-a às demais rotas desse taxista. Foram retirados outros 3 registros de um taxista parado num raio de 400 km de distância do registro não anômalo mais próximo. Outros 2 registros foram excluídos por indicarem que um taxista movimentou num raio de 53 km em somente 15 segundos. Não foram calculadas distâncias das rotas utilizando a API do Google Maps porque a base de dados de Roma não possui o *status* de ocupação informando quando as rotas iniciaram ou terminaram. Portanto, não há registros retirados devido à API. A base de dados de Roma perdeu 23 registros, o equivalente a cerca de 0,0001%.

Utilizamos a *Google Maps Distance Matrix API* para estimar as distâncias trafeadas pelos táxis por considerar as ruas e o sentido das vias para carros, e porque essa

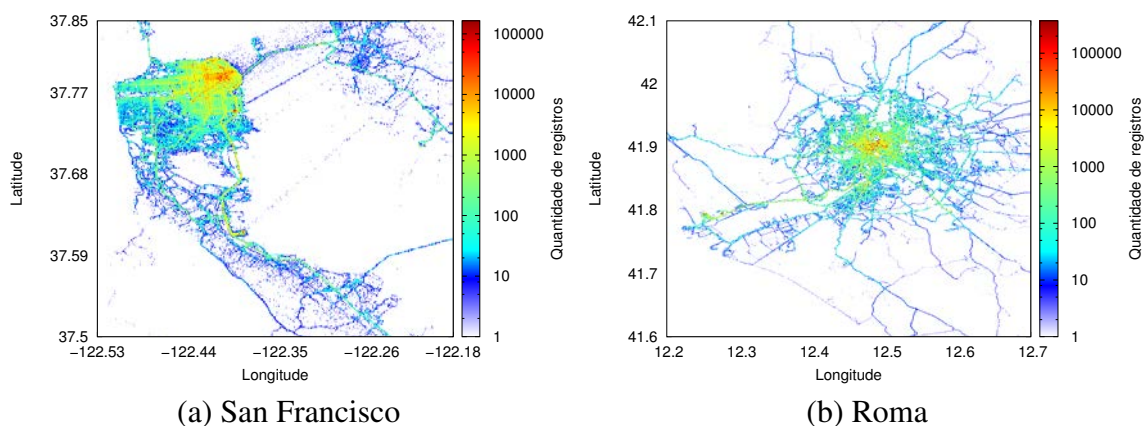


Figura 1. Mapas de calor das cidades considerando toda a base de dados

ferramenta usa os mesmos mapas do *My Maps*, utilizado para apresentar as rotas mais frequentes. Assumimos que o táxi percorreu a menor distância entre cada um dos trechos registrados na base de dados de San Francisco. No entanto, o limite de uso gratuito é de 2.500 requisições diárias, um valor bastante pequeno para a quantidade de dados em questão. As latitudes e longitudes da base de dados de San Francisco possuem 5 casas decimais, o que significa uma precisão de 1,132 metro [Rossi et al. 2015, Wikipedia 2015]. Acreditamos que tamanha precisão não é necessária para este trabalho. Para reduzir a quantidade de buscas na API do Google Maps, arredondamos as coordenadas geográficas de San Francisco de 5 para 4 casas decimais, reduzindo a precisão para 11,32 metros [Wikipedia 2015], o que entendemos ser razoável para um automóvel.

Avaliamos se há diferença significativa entre as distâncias convencionais e após a inferência de táxis parados, utilizando o teste estatístico de Kolmogorov-Smirnov para duas amostras. Esse teste foi escolhido por ser não paramétrico, sendo então mais robusto [Gibbons and Chakraborti 2003].

4. Resultados

Esta seção aborda os resultados das análises das bases de dados de San Francisco e Roma. A seção 4.1 trata da distribuição espacial dos registros em ambas as cidades e o cálculo das distâncias das rotas de San Francisco. A seção 4.2 discute a quantidade de táxis em atividade em cada base de dados no decorrer dos dias e da semana.

4.1. Análise Espacial

Frequentemente, os veículos não se deslocam aleatoriamente, e sim seguem padrões em suas trajetórias. Um dos objetivos do monitoramento e rastreamento de veículos é possibilitar a análise de mobilidade humana, para que seja possível detectar anomalias e prever comportamentos futuros. Uma vez obtidas, essas informações poderão ser utilizadas por gestores responsáveis em ações de planejamento do sistema de transporte com intuito de melhorar os deslocamentos dentro das cidades.

A Figura 1 apresenta a região abrangida pelas bases de dados de San Francisco e Roma utilizando mapas de calor em escala de intensidade logarítmica para ambas as cidades. O mapa de calor é uma representação gráfica de pontos em uma matriz colorida. No caso da Figura 1, é representada a quantidade de registros da base de dados para cada

posição geográfica (latitude e longitude). Quanto menor for a quantidade de registros em uma posição do mapa, mais próximo de branco e azul será a cor. No caso inverso, quanto maior for a quantidade de registros em uma posição, mais próximo de vermelho será a cor registrada no mapa.

Na Figura 1(a), nota-se maior intensidade de registros no nordeste da cidade de San Francisco, incluindo principalmente as rodovias 101 e 280 que conectam a região central ao “Aeroporto Internacional de San Francisco” (latitude 37.62 e longitude -122.38); a famosa ponte “San Francisco-Oakland Bay Bridge” (latitude 37.8 e longitude -122.38); e a rodovia 80 em direção a Oakland (latitude 37.8 e longitude -122.3). A maior quantidade de registros da base de dados está centralizada na rua “*Yellow Cab Access Rd.*” (latitude 37.751 e longitude -122.412), onde está localizada a “*San Francisco Yellow Cab Co-op*”, companhia de táxis amarelos e seu local de estacionamento.

Em relação ao mapa de calor de Roma apresentado na Figura 1(b), pode-se notar uma concentração maior no centro da cidade. Uma hipótese para essa maior concentração de registros de táxi está na mobilidade urbana da população local e de turistas para pontos famosos na região, tais como o “Coliseu”, o “Vaticano”, o “Pantheon” e as basílicas. A concentração de registros em Roma diminui à medida em que se distancia do centro. Porém, essa concentração se mantém em auto estradas importantes como a “Autostrada Roma - Aeroporto de Fiumicino” que conecta Roma ao “Aeroporto Internacional Leonardo da Vinci” localizado na coordenada latitude 41.8 e longitude 12.24.

A análise visual dos mapas de calor evidencia a maior demanda dos serviços de táxi para as regiões centrais e aeroportos internacionais de ambas as metrópoles. Esses padrões espaciais também retratam a importância das vias que ligam o centro das cidades aos aeroportos para os serviços de transporte público. Além disso, percebe-se que pode haver uma relação entre o número dos táxis e os pontos turísticos de ambas as cidades.

4.1.1. Inferência de Táxis Parados e Rotas Anômalas

A inferência de táxis parados é importante para o cálculo das distâncias percorridas pelas rotas porque erros de GPS podem alterar as coordenadas geográficas dos táxis mesmo que o veículo não se mova. Considerar as distâncias estimadas devido a variações de GPS quando o táxi estava parado pode resultar em valores anômalos.

Como exemplo de rota com valores anômalos está uma feita pelo táxi “new_ejesbay”. Essa rota ocorreu com o táxi sem passageiro, durou 15 horas, 28 minutos e 14 segundos e mediu 111,263 km. Todos os registros dessa rota são localizados na quadra onde está sediada a “*San Francisco Yellow Cab Co-op.*”. Conjecturamos que o comprimento da rota chegou a 111,263 km porque, mesmo com o táxi estacionado, as localizações do GPS variavam ao redor daquele mesmo local dando a impressão de que o táxi estava se movendo aleatoriamente pela região. Esse problema é ainda intensificado por possíveis imprecisões da API utilizada e pela distância ser calculada seguindo as vias e sentido permitidos. Nesse caso, uma variação de GPS que em linha reta mediria apenas 12 metros, teve a distância estimada em 958 metros pela API do Google Maps. Para contornar esse problema, aplicamos um processo de inferência para identificar se o táxi estava estacionado ou não, antes de somar as distâncias dos trechos de uma rota.

Os trabalhos [Zhu et al. 2012, Ding et al. 2015] propõem métodos de clusterização baseados em densidade para inferir se o táxi estava ou não estacionado. Uma vez que para este trabalho não é necessário identificar o local onde os táxis estavam estacionados, mas apenas se o táxi estava parado ou não, adaptamos a heurística proposta em [Zhu et al. 2012] ao cálculo das distâncias percorridas pelos táxis. Os autores definem um táxi como parado se após τ minutos (no mínimo) o veículo movimentou menos de δ metros. Os parâmetros utilizados no artigo citado e também neste trabalho foram $\tau = 3$ minutos e $\delta = 50$ metros. A medida de distância empregada nessa heurística foi a distância de Haversine [Robusto 1957] considerando o raio médio da Terra igual a 6.371 km, conforme também implementado em [Thang et al. 2012].

Uma vez que as bases de dados não possuem a informação de quando o táxi estava parado, não foi possível obter as quantidades de elementos relevantes, verdadeiros positivos e falsos positivos para calcular as medidas de *precision* e *recall* da inferência. Portanto, após estimar as distâncias corrigidas para as rotas, avaliamos se houve mudança significativa entre as distribuições de distâncias antes e após a inferência de táxis parados.

Considerando que é provável que um táxi estacionado esteja sem passageiro, e que é pouco provável que o táxi fique estacionado durante uma rota com passageiro, espera-se que haja diferença maior para as rotas com o táxi desocupado e que não haja diferença significativa para as rotas com o táxi ocupado. Avaliamos a hipótese nula de igualdade entre as distâncias das rotas antes e após a inferência de táxis parados. O teste foi realizado de três formas: entre todas as distâncias de rotas da base de dados; somente entre rotas com o táxi ocupado; e somente entre rotas com o táxi desocupado. Para o conjunto de rotas e para as rotas com o táxi desocupado o p-valor retornado com o teste de Kolmogorov-Smirnov no software R foi menor que $2,2 \times 10^{-16}$. Então, podemos dizer, com 5% de significância, que há evidências estatísticas suficientes para se rejeitar a hipótese nula de igualdade entre as distribuições. Porém, avaliando somente as distâncias com o táxi ocupado, o p-valor foi de 0,4296. Portanto, com 5% de significância, não temos evidências estatísticas para rejeitar a hipótese nula de igualdade entre as distribuições.

A técnica implementada tem efeito estatisticamente significativo nas distâncias dos trajetos e nas distâncias das rotas com o táxi desocupado, no sentido que as distribuições das distâncias antes e após a inferência de táxis parados são diferentes. Porém a técnica não causa impacto significativo nas distâncias das rotas com o táxi ocupado, uma vez que o teste utilizado não rejeitou a hipótese de igualdade entre as distâncias das rotas antes e após a inferência.

Em relação ao exemplo da rota anômala com o táxi desocupado, mencionada anteriormente, a distância caiu de 111,263 km para 1,696 km após a aplicação da técnica. Porém, não houve alterações relevantes nos quartis, médias, moda e mediana das distâncias convencionais e das distâncias após a inferência de táxis parados. Esses resultados indicam que a técnica implementada é válida para o tratamento das distâncias percorridas pelos táxis. No geral, não há impactos nas estatísticas das distâncias das rotas, com exceção das rotas anômalas obtidas com táxis estacionados.

A menor distância e também a moda das distâncias das rotas com o táxi ocupado e desocupado, antes e após a inferência foram iguais a zero. Rotas medindo zero metro mesmo antes da inferência de táxis parados podem ter ocorrido por falhas no dispositivo

que registra o *status* de ocupação das rotas, equívocos do taxista ao operar esse dispositivo, ou em situações quando o taxista termina uma rota com passageiro e logo após inicia outra rota com outro passageiro [Castro et al. 2013]. Nesse último caso, pode ser que não tenha passado tempo o suficiente para obter ao menos um registro com o táxi desocupado, e assim, separar essas duas rotas.

A situação em que não houve ao menos um registro para separar as duas rotas pode explicar a existência de rotas com distâncias enormes mesmo após a inferência de táxis parados. Embora a média das distâncias com o táxi ocupado seja 6,305 km antes da inferência e 6,283 km após a inferência, o percentil 99 para essas mesmas medidas foi de 50,723 km e 50,579 km, respectivamente.

Em relação às distâncias com o táxi desocupado, a média antes da inferência de táxis parados foi de 6,935 km e após a inferência foi de 6,684 km. Os percentis 99 foram de 77,646 km e 76,162 km para antes e após a inferência, respectivamente. A maior rota com o táxi desocupado teve 1.391.593 km antes da inferência e 1.327.492 km após a inferência de táxis parados. Esses resultados indicam que mesmo tratando a questão dos táxis parados, rotas com distâncias anômalas ainda são presentes nos dados de San Francisco. As hipóteses para a causa dessas medidas anômalas são a falta de uma marcação do fim das rotas de táxi e equívocos dos taxistas ou erros do dispositivo ao registrar o *status* das rotas.

4.1.2. Análise das Rotas Mais Frequentes de San Francisco

A análise das rotas mais frequentes dos táxis ocupados em uma cidade é importante para sugerir novos pontos de táxi, orientar melhor os taxistas na busca do próximo passageiro, propor melhorias no transporte público e para planejamento das vias urbanas. Como a base de dados de Roma não possui alguma informação indicando o início e final das rotas, foram analisadas as rotas mais frequentes somente da cidade de San Francisco.

A Figura 2 apresenta as dez rotas com maior número de ocorrências. As linhas em azul representam as rotas de menor percurso de carro entre os pontos de origem e fim segundo o Google Maps. Os locais em verde são o início e os locais em vermelho são os destinos das rotas. O local em amarelo representa a rota, dentre as mais frequentes, que iniciou e terminou na mesma região. Na Figura 2(a) são mostradas todas as dez rotas identificadas. Há basicamente dois grupos de rotas: um na região central da cidade, mostrado na Figura 2(b), e outro na parte inferior do gráfico localizado no “Aeroporto Internacional de San Francisco”. A rota em amarelo ocorreu na “*Yellow Cab Acces Rd.*”, rua que dá acesso ao “San Francisco Yellow Cab Co-op” mencionado anteriormente.

Alguns pontos da cidade alternaram-se como locais de início e fim das dez rotas mais frequentes. Na Figura 2(b), os pontos que aparecem mais de uma vez dentre as rotas mais frequentes são: a praça “Union Square” (ponto em verde no centro horizontal do mapa), o *shopping* “Pier 39” (na parte superior), e o hotel “San Francisco Marriot Marquis” (na parte inferior do mapa). Em relação ao hotel San Francisco Marriot Marquis, as duas rotas mais frequentes ligavam o hotel ao Aeroporto Internacional da cidade, tendo o início e fim dessas rotas situados em saídas diferentes do hotel e do aeroporto. Outros dois locais apareceram entre as dez rotas mais frequentes somente como termos de

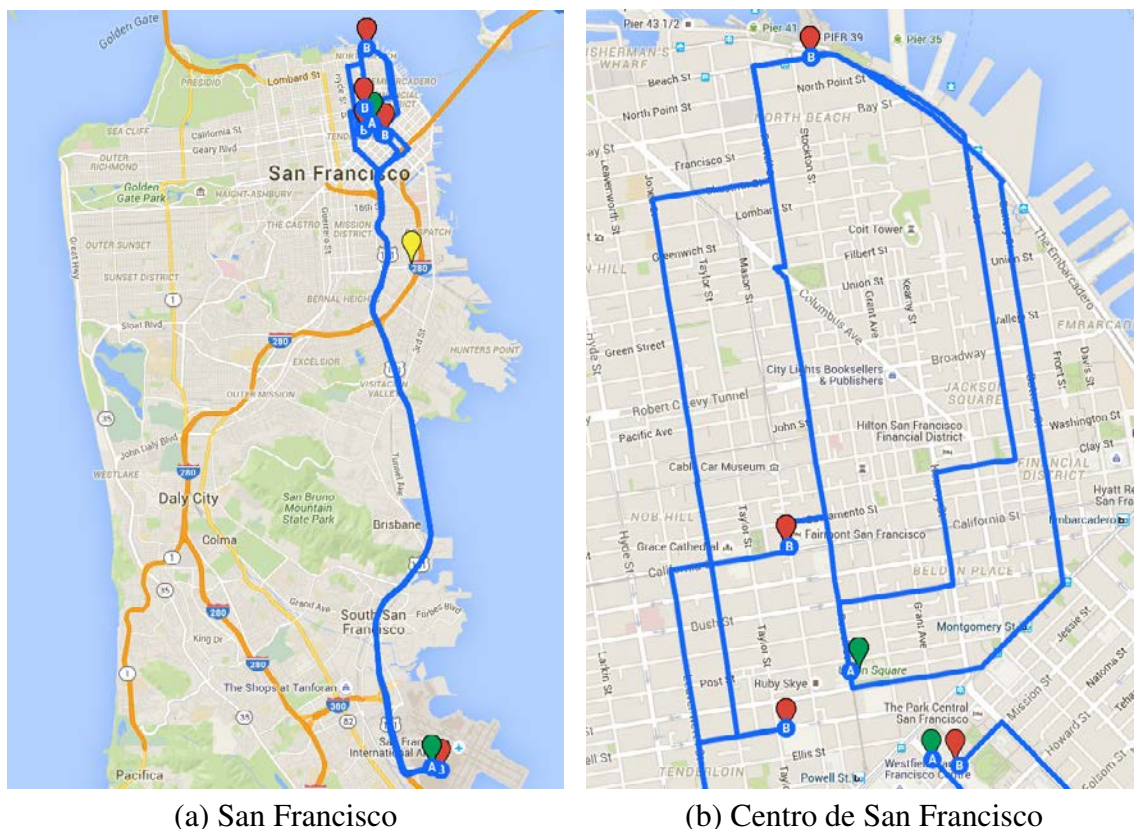


Figura 2. Rotas mais frequentes de San Francisco

rotas: “Pacific-Union Club” e o hotel “Hilton San Francisco Union Square”.

Todas as rotas identificadas ficam localizadas na região leste da cidade. Para todas as rotas que ligam o centro de San Francisco ao aeroporto, o Google Maps sugeriu utilizar a rodovia 101. Pela diferença de cores no mapa de calor da Figura 1(a), imagina-se que os táxis também escolhem a rodovia 101 para realizar esse trajeto.

O fato da maioria das rotas mais frequentes compartilharem os mesmos locais de início e fim indica que certos pontos de interesse de San Francisco como o Aeroporto Internacional, hotéis, o Union Square e o *shopping* “Pier 39” representam fontes importantes de clientes para os serviços de táxi. Além disso, as análises revelam uma relação de rotas entre o aeroporto e os hotéis da cidade. Acreditamos que essa relação se deve principalmente por turistas durante a chegada e saída de San Francisco.

4.2. Análise Temporal em Ambas as Bases de Dados

Compreender as dinâmicas dos táxi no decorrer do dia e da semana é importante para o gerenciamento das próprias agências táxi, para o planejamento de transportes públicos concorrentes, e para residentes e visitantes que utilizam ou dependem do serviço de táxi.

A madrugada é o momento de menor quantidade de registros diários em ambas as bases de dados de San Francisco e Roma. A menor presença de táxis na rua em San Francisco foi entre 03:00 e 04:00 horas (2,1% da quantidade diária). Em relação à Roma, a menor quantidade diária foi das 04:00 às 05:00 (1,05% do total diário). As menores taxas de táxis ocupados da base de dados de San Francisco ocorrem das 03:00 até antes

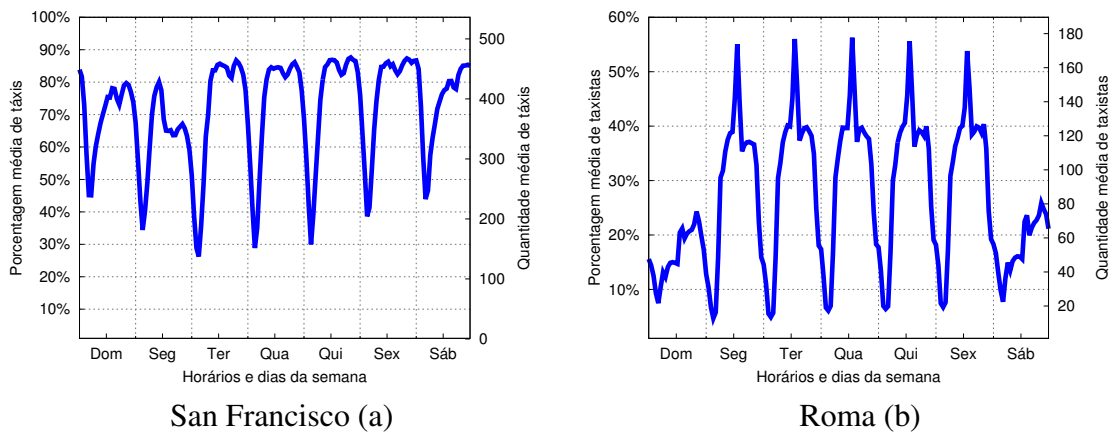


Figura 3. Quantidade de táxis e taxistas no decorrer da semana

das 07:00; nesse período, as taxas de ocupação por hora oscilam de 26,1% a 27,94%. Após as 04:00, o número de registros aumenta regularmente e se estabiliza por volta das 09:00 em ambas as bases de dados. Acreditamos que esse comportamento se deve à menor demanda de passageiros durante a madrugada. Portanto, além de haver menos rotas com os táxis ocupados, os taxistas parecem preferir estacionar o táxi e esperar até que um passageiro surja em vez de trafegar pelas ruas procurando o próximo passageiro.

Em San Francisco, a quantidade de registros oscila pouco das 09:00 até as 16:00. Uma hipótese para a demanda de táxis não ter aumentado durante o horário do almoço é de que a população local evita almoçar em locais distantes do serviço, não precisando utilizar táxis. A Figura 3(a) apresenta uma queda na quantidade de táxis em todos os dias da semana (principalmente na segunda-feira) próximo ao horário de almoço. Imaginamos que essa queda seja causada por troca de taxistas que dirigem o mesmo táxi, ou uma pausa dos próprios taxistas para almoçar. A quantidade de registros cresce regularmente das 16:00 até as 20:00, horário da maior porcentagem diária (5,02%). Das 17:00 até as 23:00 a quantidade de táxis ocupados ultrapassa a quantidade de táxis desocupados. Após as 18:00, a quantidade de registros cai até as 04:00 do dia seguinte. Essa demanda de táxis, principalmente ocupados, pode ser o reflexo das pessoas deixando os locais de trabalho após o horário comercial e dirigindo-se para suas casas ou outros destinos.

Em relação à Roma, o maior percentual de registros ocorre por volta das 13:00 (5,95%), tendo uma queda às 15:00 (4,65%) e voltando a subir até às 17:00 (5,93%), de onde cai até o menor valor no dia seguinte. A Figura 3(b) ilustra esse comportamento com a quantidade média de taxistas no decorrer da semana. Nota-se que, diferentemente de San Francisco, há um pico de demanda de taxistas no horário do almoço e os dados seguem um padrão muito semelhante de segunda a sexta, variando somente nos sábados e domingos, quando a porcentagem média de taxistas não chega a 30%. Acreditamos que essa diferença pode ser explicada devido à população de Roma ter mais costume de almoçar em locais distantes de onde trabalha, e pelos dados não serem referentes a táxis, e sim a taxistas que seguem jornadas de trabalho controladas durante o dia e semana.

A segunda-feira é o dia da semana com menor quantidade média de táxis em San Francisco, com a média de táxis variando entre 65% e 80% do total. A média dos táxis não alcança 90% dos 536 táxis em nenhum dia da semana. Para todos os dias da

semana, a quantidade média de táxis desocupados é maior que a quantidade média de táxis ocupados. As menores porcentagens diárias diminuem do começo de domingo até terça-feira. Após a terça-feira, essas menores porcentagens diárias começam a aumentar até o final da semana. Um comportamento diferente se percebe com as maiores porcentagens diárias, as quais oscilam pouco de terça à sábado tendo uma queda no domingo e segunda. Esse padrão pode ser o reflexo de eventos sociais da população de San Francisco que aumentam à medida em que o final de semana se aproxima.

Outro ponto diferente entre as duas cidades está nas porcentagens de uso do sistema de táxi. Enquanto em Roma as porcentagens médias não chegam à 60%, em San Francisco a porcentagem ultrapassa os 80% ao menos uma vez por dia. Em nenhum momento o percentual de taxistas operando em Roma supera a quantidade de táxis de San Francisco. Embora essas duas bases de dados tenham características semelhantes, os padrões temporais analisados das rotas de táxi indicam costumes e rotinas diferentes dentre as duas metrópoles.

5. Conclusão

Neste trabalho foram comparados os padrões espaciais e temporais da circulação de táxis de San Francisco e Roma; foi discutido o efeito da inferência de táxis parados ao calcular a distância das rotas; e foram apresentadas as rotas mais frequentes da base de dados de San Francisco. As análises espaciais mostram a maior atividade de táxis nos centros das cidades e que se distribuem pelas maiores avenidas de trânsito rápido das regiões. As análises temporais evidenciam diferenças no padrão de utilização dos serviços de táxis nas duas cidades, que podem refletir diferenças culturais ou mesmo legais desse serviço.

Em relação às rotas de táxi, foram apresentados resultados da eficiência de inferir táxis parados para evitar erros no cálculo de distâncias das rotas. Assim como o esperado, foram encontradas evidências estatísticas com 5% de significância de que as distâncias das rotas geradas após inferir e desconsiderar os trechos com o táxi parado são diferentes que as distâncias convencionais.

Como trabalhos futuros, sugere-se ajustar os parâmetros τ e δ da inferência de táxis parados para diferentes situações do tráfego de veículos nas estradas tais como congestionamentos. Também propõe-se que, embora haja o *status* de ocupação nos dados de San Francisco, seja inferido quando cada rota iniciou e terminou a fim de reduzir a quantidade de rotas anômalas da base de dados. Além disso, sugerimos que seja apresentada a variação dos locais de início e fim das rotas mais frequentes de San Francisco no decorrer dos horários dos dias e no decorrer dos dias da semana.

6. Agradecimentos

Os autores agradecem ao CEFET-MG e aos financiadores dos projetos INCT InWeB (MCT/CNPq 573871/2008- 6) e MASWeb (FAPEMIG/PRONEX APQ-01400-14).

Referências

- Amici, R., Bonola, M., Bracciale, L., Rabuffi, A., Loreti, P., and Bianchi, G. (2014). Performance Assessment of an Epidemic Protocol in VANET Using Real Traces. *Procedia Computer Science*, 40:92–99.
- Bracciale, L., Bonola, M., Loreti, P., Bianchi, G., Amici, R., and Rabuffi, A. (2014). CRAWDAD dataset roma/taxi (v. 2014-07-17). Downloaded from <http://crawdad.org/roma/taxi/20140717>.

- Castro, P. S., Zhang, D., Chen, C., Li, S., and Pan, G. (2013). From Taxi GPS Traces to Social and Community Dynamics: A Survey. *ACM Comput. Surv.*, 46(2):17:1–17:34.
- Chen, C. (2014). *Understanding social and community dynamics from taxi GPS data*. PhD thesis, Evry, Institut national des télécommunications.
- Cunha, F. D., Alvarenga, D. A., Viana, A. C., Mini, R. A., and Loureiro, A. A. (2015). Understanding Interactions in Vehicular Networks Through Taxi Mobility. In *Proceedings of the 12th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, pages 17–24. ACM.
- Ding, L., Fan, H., and Meng, L. (2015). Understanding taxi driving behaviors from movement data. In *AGILE 2015*, pages 219–234. Springer.
- Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric statistical inference*. Marcel Dekker, New York.
- Kellner, L. and Egger, R. (2016). Tracking Tourist Spatial-Temporal Behavior in Urban Places, A Methodological Overview and GPS Case Study. In *Information and Communication Technologies in Tourism 2016*, pages 481–494. Springer.
- Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., and Yang, Q. (2011). Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 63–68. IEEE.
- Liu, S., Wang, S., Liu, C., and Krishnan, R. (2015). Understanding taxi drivers' routing choices from spatial and social traces. *Frontiers of Computer Science*, 9(2):200–209.
- Murashkin, M. and Geroliminis, N. (2015). Analysis of taxi drivers' behavior with GPS data from a megacity. In *hEART*, Lyngby. European Association for Research in Transportation (hEART), Department of Transport at Technical University of Denmark (DTU Transport).
- Piorkowski, M., Sarafjanovic-Djukic, N., and Grossglauser, M. (2009). CRAWDAD dataset epfl/mobility (v. 2009-02-24). Downloaded from <http://crawdad.org/epfl/mobility/20090224>.
- Robusto, C. (1957). The cosine-haversine formula. *The American Mathematical Monthly*, 64(1):38–40.
- Rossi, L., Walker, J., and Musolesi, M. (2015). Spatio-temporal techniques for user identification by means of gps mobility data. *EPJ Data Science*, 4(1):1–16.
- Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., and Fotheringham, A. S. (2015). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, pages 1–26.
- Thang, N. V., Thang, P. M., and Tan, T. D. (2012). The performance improvement of a lowcost ins/gps integration system using the street return algorithm. *Vietnam Journal of Mechanics*, 34(4):271–280.
- Valero, B., Luis, J., Julián, A., Belén, A., Villén, G., and Natalia (2014). *GNSS. GPS: Fundamentos y aplicaciones en geomática*. Editorial de la Universidad Politécnica de Valencia, Valencia.
- Wikipedia (2015). Decimal Degrees. https://en.wikipedia.org/wiki/Decimal_degrees.
- Zhang, Y. (2014). How do taxis work in beijing? Master of urban and regional planning, University of California, Los Angeles, Los Angeles.
- Zheng, Z., Rasouli, S., and Timmermans, H. (2015). Two-regime Pattern in Human Mobility: Evidence from GPS Taxi Trajectory Data. *Geographical Analysis*.
- Zhu, Y., Zheng, Y., Zhang, L., Santani, D., Xie, X., and Yang, Q. (2012). Inferring taxi status using gps trajectories. *arXiv preprint arXiv:1205.4378*.

SEDNA – Sistema Emergencial para Desastres Naturais

Yuri Marinho Olivatti, André Kurt Amann Pereira, Rodrigo Filev Maia¹

¹ Centro Universitário FEI (FEI) – São Bernardo do Campo – SP – Brazil

yuri_olivatti@hotmail.com, {kpereira, rfilev}@fei.edu.br

Abstract. *With the population growth in large urban centers, the occupation of areas at risk of natural disasters is constantly growing and thus making the work for evacuation in emergency times more and more complex. Currently, one of the most widespread mechanisms of communication are cellular mobile devices. This paper proposes using mathematical models to identify the possibility of a risk occur and alert, through a smartphone application, the population in the risk area and designate a safe place for evacuation. The expected result of this work is that the sensors at the risk area and the model utilized indicate potential natural disaster occurrences and that this information may be sufficient to save the local population.*

Resumo. *Com o crescimento populacional em grandes polos urbanos a ocupação de áreas com risco de desastres naturais cresce constantemente, tornando o trabalho para evacuação, em momentos emergenciais, cada vez mais complexo. Atualmente, um dos dispositivos de comunicação mais difundidos são os celulares. Este trabalho propõe, utilizando modelos matemáticos, identificar a possibilidade de uma situação de risco ocorrer e acionar, por meio de um aplicativo móvel, a população que se encontra na área de risco e indicar uma área segura para evacuação. Espera-se, com o modelo utilizado, identificar potenciais ocorrências de desastres naturais e que estas informações possam ser suficientes para salvar a população local.*

1. Introdução

1.1. Revisão Bibliográfica

Muitos casos de mortes e perdas materiais devido a catástrofes naturais são registrados anualmente no Brasil. Segundo Collins (2004), mais de 29 milhões de brasileiros residem em áreas suscetíveis às inundações e 100 pessoas são vitimadas anualmente no Brasil. Esses números são decorrentes do crescimento populacional não planejado, resultando em moradias em áreas impróprias. Outras catástrofes que afetam a maioria das regiões brasileiras são as de deslizamentos de encosta. O processo de erosão e os deslizamentos, juntos à ação humana, estão constantemente causando a alteração da superfície terrestre. Esses eventos estão em contínuo aumento no Brasil, principalmente nas regiões urbanas, construções irregulares, e são responsáveis por grandes acidentes, causando perdas materiais e de vidas (Brito, 2013).

Estima-se que no mundo nos anos entre 1970 a 2012, os desastres naturais causaram cerca de 1,94 milhões de mortes e danos econômicos de 2,3 trilhões de dólares globalmente, onde na América do Sul foram mais de 50 mil vidas perdidas. (OMM, 2014).

Muitas regiões não possuem monitoramento e métodos de evacuação dessas catástrofes, ou são ineficazes. Mesmo em grandes cidades é possível encontrar sistemas bem consolidados para mapeamento de regiões de risco, porém poucas soluções de monitoramento e evacuação são encontradas. Atualmente, os alertas que são enviados para a população são realizados por alguns métodos comuns entre organizações, onde o principal utilizado no Brasil é o das sirenes. Elas emitem um som para que a população da região evacue a área; entretanto, foi identificado pela Defesa Civil do estado do Rio de Janeiro que apenas 10% da população atende as sirenes de evacuação (Gomide, 2013).

A Defesa Civil é um conjunto de órgãos e entidades, públicos e privados, de ordem municipal a nacional, que visam à redução de desastres, a diminuição da ocorrência dos mesmos e seus efeitos. O trabalho do Sistema Nacional de Defesa Civil (SNDC) é garantir a prevenção de desastres naturais ou provocados pelo homem, com programas e planos para proteger a população e minimizar danos no caso de ocorrências, alertando, dando assistência e reabilitando pessoas, e recuperando as áreas deterioradas. Um importante trabalho é o de monitoração: prevendo-se um desastre natural a partir das condições de uma determinada região, é possível evacuar a população de forma organizada e evitar danos físicos e mortes. Esse monitoramento normalmente acontece com especialistas observando as condições das áreas de risco e prevendo de forma subjetiva sua ocorrência, por exemplo, durante uma forte chuva. Porém, o monitoramento também pode ocorrer de forma determinística, através de modelos matemáticos e uma série de parâmetros obtidos da região (Nelson, 2001).

Diante dessa situação, onde o monitoramento das regiões de risco e a evacuação através de sirenes se mostraram muitas vezes ineficazes, este projeto tem como proposta a obtenção de dados através de sensores e sua utilização em modelos matemáticos/geomorfológicos a fim de perceber e alertar a população de possíveis desastres naturais, de forma intuitiva e com amplo alcance, através de um aplicativo móvel.

A obtenção dos dados para monitoração das áreas é possível através de sensores específicos. Instalados apropriadamente, poderão coletar dados referente ao solo, clima, nível de água, entre outros parâmetros, para monitoração em tempo real e aplicação dos modelos matemáticos. Os sensores podem comparar seus dados para detectar movimentações no solo e quais áreas sofreram maior variação, como funcionam alguns sistemas de predição de deslizamentos (Terzis, et al. 2006), ou enviar dados a uma central, que os usará como parâmetros de entrada para algum modelo matemático de previsão de deslizamentos, como o SHALSTAB (Ramos, et al. 2002). O índice de criticidade de ocorrência do desastre e os dados medidos da região possibilitarão ao operador do órgão responsável pelo monitoramento observar as alterações e identificar se será necessário emitir um alerta à população.

1.2. Objetivo

O objetivo geral deste projeto é monitorar áreas de risco através de sensores e indicar o potencial risco de deslizamentos de terra e inundação, utilizando modelos matemáticos de avaliação em tempo real do estado do solo e índices pluviométricos, permitindo a evacuação das pessoas em risco.

Pretende-se oferecer a análise em tempo real do risco em que as regiões monitoradas se encontram, fornecendo medições de umidade do solo e respectivos índices de estabilidade. Também se espera agilizar o processo de evacuação, através de um aplicativo para dispositivos móveis para a população, que indicará os locais seguros para fuga tendo como base a indicação dos centros de gerenciamento de emergências.

2. Modelos Geomorfológicos

Um modelo matemático geomorfológico é um modelo com aplicações na análise de terrenos ou relevos com diversos objetivos, entre eles a análise de possíveis deslizamentos em um determinado espaço geográfico. Modelos geomorfológicos determinísticos, em particular, apresentam algumas limitações com relação ao universo de conhecimento dos especialistas. Um modelo determinístico não necessariamente deve ser completo (não apresentando obrigatoriamente todos os parâmetros necessários para a solução ótima da saída), onde a precisão do modelo está relacionada com as variáveis que são utilizadas e a correta configuração da equação em si.

O SHALSTAB, utilizado nesse artigo, é um modelo matemático geomorfológico determinístico, bastante difundido por parte dos estudos geomorfológicos e com precisão razoável, que tem como objetivo apresentar locais em um espaço geográfico onde podem ocorrer os deslizamentos rasos, sendo possível prever um valor para a chuva crítica deflagradora da ruptura do solo (Ramos, et al. 2002).

$$\frac{Q}{T} = \frac{\text{sen}\theta}{(a/b)} \left[\frac{\rho_s}{\rho_w} \left(1 - \frac{\tan\theta}{\tan\phi} \right) \right]$$

Equação 1. Modelo Matemático SHALSTAB

O modelo SHALSTAB, retratado na equação 1, utiliza como parâmetros o índice de chuva crítica para ruptura do solo (Q), a transmissividade do solo (T), a densidade da água (ρ_w), a densidade do solo úmido (ρ_s), a área de contribuição por comprimento de contorno (a/b), o ângulo de inclinação da encosta (θ) e o ângulo de atrito interno (ϕ). O log (Q/T) representa o índice de estabilidade do solo, que quando maior que -2,2 indica uma região estável, e quando menor representa uma situação com diferentes graus de instabilidade (Ramos, et al. 2002).

3. Módulo de Sensores

O módulo de sensores é composto pelos sensores adequados para medição dos parâmetros requeridos pelo modelo matemático utilizado, bem como de microcontroladores para processamento dos dados e envio das medições para o servidor web.

Para realizar essa medição, foram utilizados alguns sensores próprios como o sensor ultrassônico, capaz de medir o nível da água, o acelerômetro, que percebe vibrações e movimentações do solo e o de umidade do solo (*soil moisture sensor*). Estes sensores estão ligados a um dispositivo integrado de microcontroladores, com um dispositivo Arduino e um Raspberry Pi, juntos em um único módulo, em uma arquitetura onde o Arduino faz a leitura e controle dos sensores e dispositivos conectados a ele, retornando quaisquer valores medidos ao Raspberry Pi, que com conexão à Internet pode enviá-lo a algum servidor ou outro dispositivo. Com essa

proposta, o Arduino nas portas digitais tem conectado o sensor ultrassônico, e nas portas analógicas o acelerômetro e o sensor de umidade, fazendo a constante medição, conversão de grandezas e envio dos dados ao Raspberry Pi, e este tem a função de consolidar as informações a aplicar no modelo matemático SHALSTAB, o qual realiza o estudo dos dados (medidos pelos sensores) e determina a probabilidade de ocorrer um desastre natural na região de onde as medições se originaram. O Raspberry Pi também administra a conexão com a Internet que é feita por um modem de conexão 3G/4G. Desta forma, todas as informações medidas pelos sensores são enviadas para um servidor remoto e armazenadas em uma base de dados.

É importante ressaltar que esta arquitetura permite paralelismo entre a leitura dos dados do sensor (Arduino) e a execução dos cálculos e envio ao servidor web (Raspberry Pi), evitando assim que dados medidos pelos sensores sejam perdidos enquanto o dispositivo está ocupado com operações de cálculo ou envio à Internet.

4. Implementação

A arquitetura do sistema se baseia em três pilares principais: o módulo de sensores, que realiza as medições da região monitorada e os cálculos do modelo matemático; o servidor, responsável por armazenar os dados em uma banco de dados e disponibilizar a interface de monitoramento para a análise de um especialista; e o aplicativo móvel, que exibe as condições da região ao usuário e o notifica quando estiver em perigo, exibindo rotas de fuga de acordo com a indicação do especialista, da central de monitoramento.

Para o módulo de sensores foram utilizadas as funcionalidades dos microcontroladores Arduino e Raspberry Pi, junto com um conjunto de sensores. O Arduino, responsável pela integração dos sensores, é ligado ao Raspberry Pi, responsável pelo modelo matemático e envio das informações ao servidor remoto (figura 1).

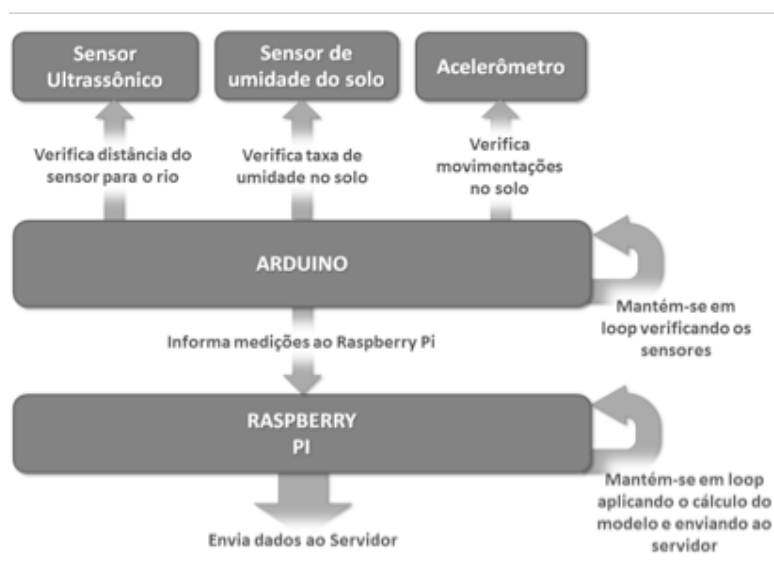


Figura 1. Funcionamento do Módulo de Sensores

O servidor é composto pela interface do operador do centro de monitoramento de emergências, dois webservices de comunicação, um para o aplicativo e outro para o módulo de sensores, e o banco de dados, como ilustrado na Figura 2.

Os *webservices* de comunicação são responsáveis por salvar as medições dos módulos de sensores enviadas, e fornecer aos aplicativos nos dispositivos móveis as informações a respeito da posição em que estão localizados, incluindo locais de fuga caso a região esteja em perigo; implementados em *PHP*, recebem requisições *HTTP* dos módulos de sensores, salvando as medições em banco *SQL Server*, e dos aplicativos, fornecendo ao mesmo a partir da base de dados as condições da região.

Com a interface de monitoramento, é possível que um especialista na central possa analisar a condição das regiões e alertar a população se percebido algum perigo, definindo locais seguros para evacuação. A interface foi desenvolvida em *ASPX* e *HTML5* e diversos códigos *javascript*, possibilitando a exibição de uma página dinâmica, em constante atualização com as medições e condições das regiões monitoradas.

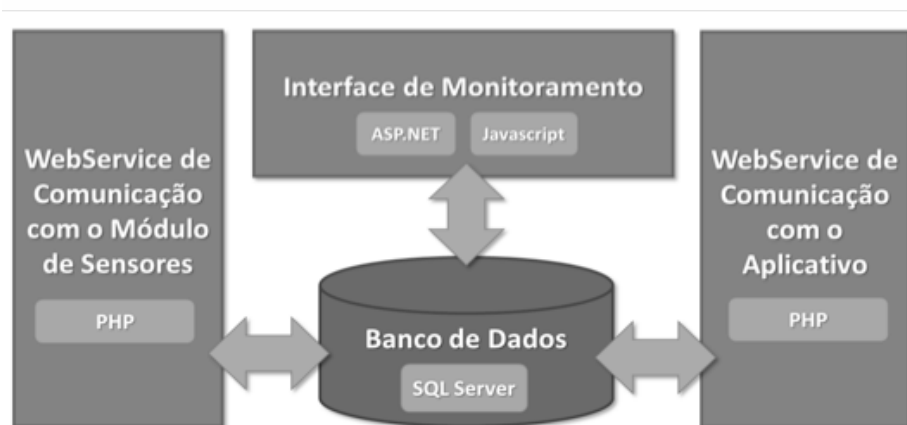


Figura 2. Arquitetura básica do servidor de monitoramento

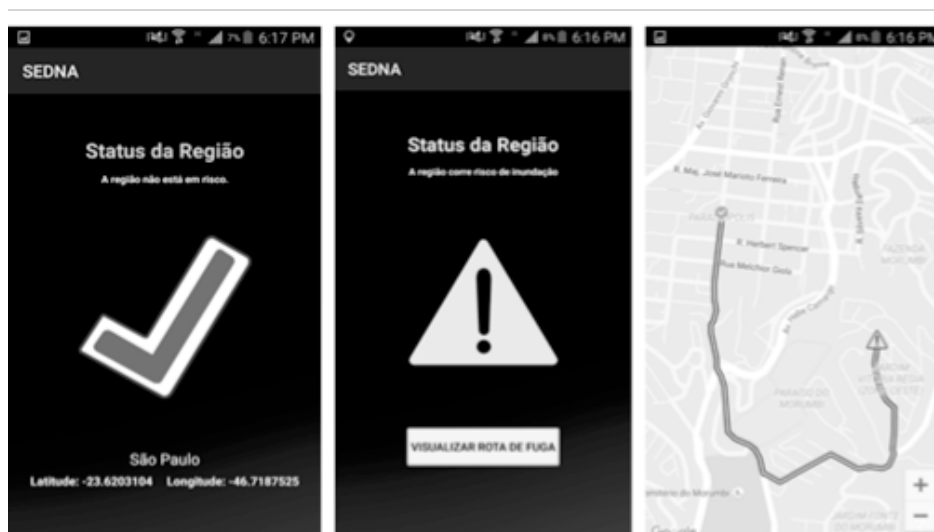


Figura 3. Telas do Aplicativo

O aplicativo móvel destinado a smartphones (figura 3) é o elemento do sistema que possibilitará a população em risco ser notificada. Desenvolvido para a plataforma Android, em Java, é utilizado o recurso de GPS dos smartphones para obter a posição do usuário e enviado constantemente requisições *HTTP* ao servidor solicitando as condições naquela localização; caso a região esteja em risco, também receberá indicações de locais seguros, onde através da API do Google Maps, integrada com o sistema Android, exibirá um mapa com a rota para evacuação. Também há um processo em background em execução, onde sempre que a região se encontrar em perigo o aplicativo abrirá e um som de alerta será disparado.

5. Resultados

Para a realização dos testes foi desenvolvido uma maquete representando uma situação real de região de risco, detalhada na seção 5.1. A solução do sistema emergencial teve resultados satisfatórios e seus dois modos, para deslizamento e inundação. Dois resultados relevantes devem ser destacados. O primeiro diz respeito quanto à reação do sistema para diferentes valores enviados pelos sensores. Sempre que o nível pluviométrico ou o índice de estabilidade apresentam valores considerados perigosos, o sistema reage adequadamente e a interface de monitoração mostra apropriadamente o risco na região em que o sensor está posicionado. O segundo diz respeito às medições dos sensores e resultado dos modelos calculados em relação ao risco real naquela região; os testes mostram que os resultados dos cálculos efetivamente indicam o risco antes do desastre ocorrer, sendo possível realizar a evacuação.

5.1. Maquete para testes

Para realização dos testes e verificação da metodologia aplicada, foi desenvolvido uma maquete para simulação dos desastres de inundação e deslizamento de terra (figura 4). A maquete representa uma área em torno de uma cidade paulista, onde há um rio em frente de um prédio que representa um risco de inundação, e mais a frente um morro com alta inclinação (45 graus), representando um perigo de deslizamento.

A maquete possui uma base onde estão todos os componentes eletrônicos (o módulos de sensores), e alguns orifícios por onde os fios dos sensores passam e são posicionados na montanha e próximo ao rio; nessa base também há um reservatório de água com bombas submersíveis, que através de mangueiras podem encher o rio, umedecer o solo do morro por baixo e irrigar o solo por cima em uma espécie de chuva, através de furos na mangueira. Assim, é possível simular a situação dos dois desastres em escala, e verificar o funcionamento do sistema desenvolvido.

Para os testes foram preparadas amostras de solo típicas de uma região da cidade estudada; o solo foi preparado a partir de uma mistura de solo siltoso (entre areia e argila, possuindo partículas pequenas e corpo rígido), adequado para a simulação de deslizamento, menos rígido que o solo argiloso, mas mais coeso que o puramente siltoso. Dados como coesão e peso específico do solo foram medidos para configuração correta do módulo de sensor, com os parâmetros necessários para realização dos cálculos do modelo. É importante ressaltar que a configuração exata do solo da região não foi mapeada para a maquete, limitando-se apenas à inclinação e ao tipo de solo (siltoso), representativos de um risco de deslizamento.



Figura 4. Imagem da maquete, com a mistura de solo aplicada e sensores posicionados

5.2. Testes de deslizamento

O primeiro teste foi realizado utilizando os parâmetros constantes da maquete, como a inclinação da encosta, porém simulando os dados medidos pelos dos sensores, como indicado na Figura 5.

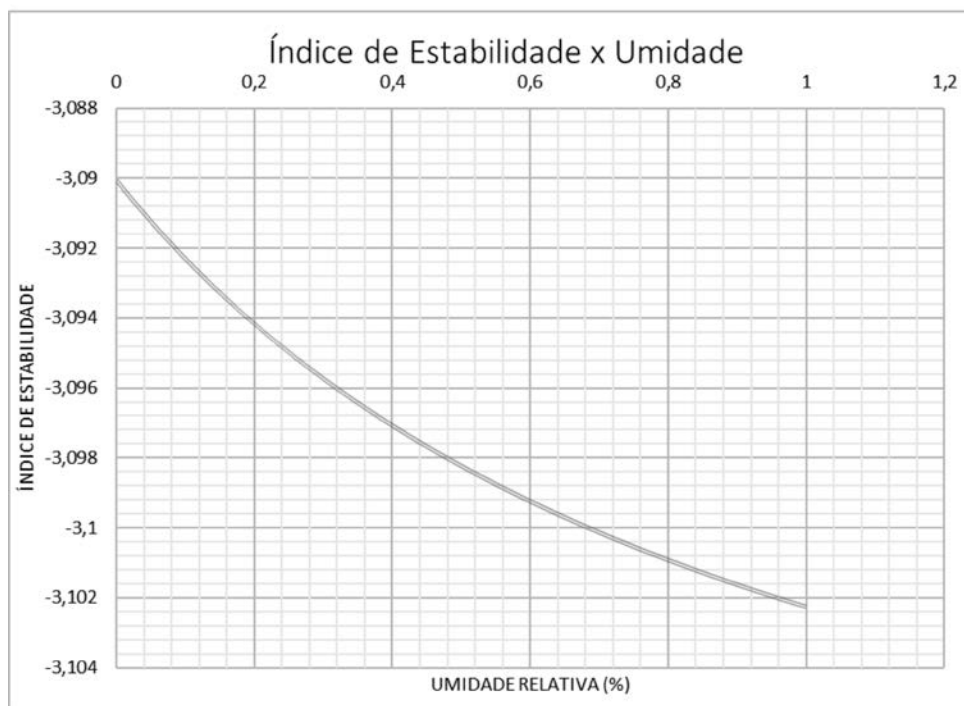


Figura 5. Gráfico do Índice de Estabilidade x Umidade

Utilizando o modelo matemático SHALSTAB e parâmetros pré-estabelecidos, verificou-se a reação do sistema. Conforme a umidade variou de 0 a 1 (0% a 100%), o resultado calculado pelo modelo matemático também mudou, indicando maior risco

conforme a umidade aumentava, e o sistema corretamente reagiu à variação, mudando a cor da região na interface e indicando que estava em perigo.

De acordo com o modelo SHALSTAB, os resultados do modelo de estabilidade indicam o seguinte grau de instabilidade para os intervalos calculados, com risco variando de 1 a 7:

- Índice de risco 1, região incondicionalmente estável
- **Maior que -2,2:** Índice de risco 2 – região medianamente estável
- **Entre -2,2 e -2,5:** Índice de risco 3 – região pouco estável
- **Entre -2,5 e -2,8:** Índice de risco 4 – região pouco instável
- **Entre -2,8 e -3,1:** Índice de risco 5 – região medianamente instável
- **Menor que -3,1:** Índice de risco 6 – região muito instável
- Índice de risco 7, região incondicionalmente instável

Assim, para os resultados obtidos no gráfico, com umidade de 0% a região tem índice de risco 5, aumenta gradualmente e quando se torna 0,7, o índice passa a ser 6. É importante ressaltar que estamos considerando uma região de que já é de risco, devido sua alta inclinação. Cada região monitorada pode possuir parâmetros diferentes, onde a umidade influencia mais ou menos e a variação do índice de risco é diferente.

Posteriormente os testes de deslizamento foram realizados na maquete, com o solo adequado para tais experimentos, em uma superfície com ângulo de inclinação de 45°. No decorrer dos testes foram utilizados dois sistemas para umedecer o solo e causar o deslizamento: o sistema de infiltração por dentro da maquete e o sistema de mangueiras para simular a chuva. Os testes mostraram que com os cálculos do modelo é possível perceber o alto risco de deslizamento algum tempo antes do desastre ocorrer, possibilitando a evacuação.

Em um dos testes, os sistemas de infiltração e chuvas foram ligados (instante x), com o solo apresentado a umidade inicial de 65%. Decorridos dois minutos e cinquenta segundos ($x+2:50\text{min}$), o nível de risco subiu para 6, o que já representa um alto risco de deslizamento conforme detalhado acima, e após mais um minuto e doze segundos ($x+4:02\text{min}$) o deslizamento ocorreu. Antes de ocorrer, quando o índice subiu para 6, foi enviado através da interface de monitoramento um alerta aos usuários na região, exibido corretamente no aplicativo. No teste verificou-se o funcionamento da comunicação fim-a-fim, desde a medição dos sensores até o alerta ao usuário, bem como a validade do modelo SHALSTAB para os valores de umidade medidos, mostrando que a região se encontrava em risco 1 minuto e 12 segundos antes do desastre ocorrer, permitindo a evacuação. A figura 6 mostra a variação das medições de umidade em função do tempo, do instante inicial de testes (0 seg) até o final (241 seg), e a figura 7 mostra como o índice de risco, como definido acima, variou nesse mesmo período.

Vale ressaltar que o tempo do instante em que o risco subiu para 6 até o momento em que o deslizamento iniciou foi curto, propositalmente, pois o cenário montado já era classificado como uma área de risco, devido sua alta inclinação e a mistura de solo preparada, processo acelerado pela entrada de água por diversos pontos em todo o sistema, com alta vazão, que pode representar uma situação real de chuva

muio intensa, de forma proposital a causar rapidamente a ocorrência do evento de deslizamento. O especialista pode analisar e julgar se a região está em risco ou não de acordo com os cálculos do modelo e a umidade que vai aumentando, e realizar o alerta aos usuários se necessário. Assim, não há um índice de risco em específico em que se pode afirmar com certeza que o desastre ocorrerá; essa é uma decisão passível de julgamento humano.

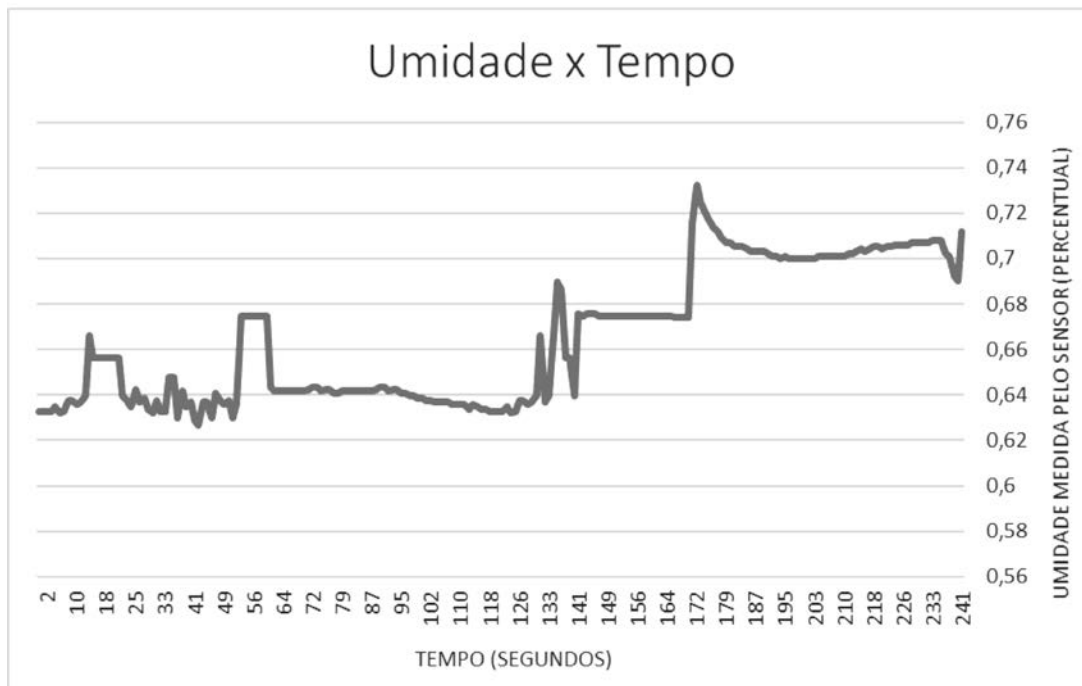


Figura 6. Gráfico da Umidade do Solo x Tempo

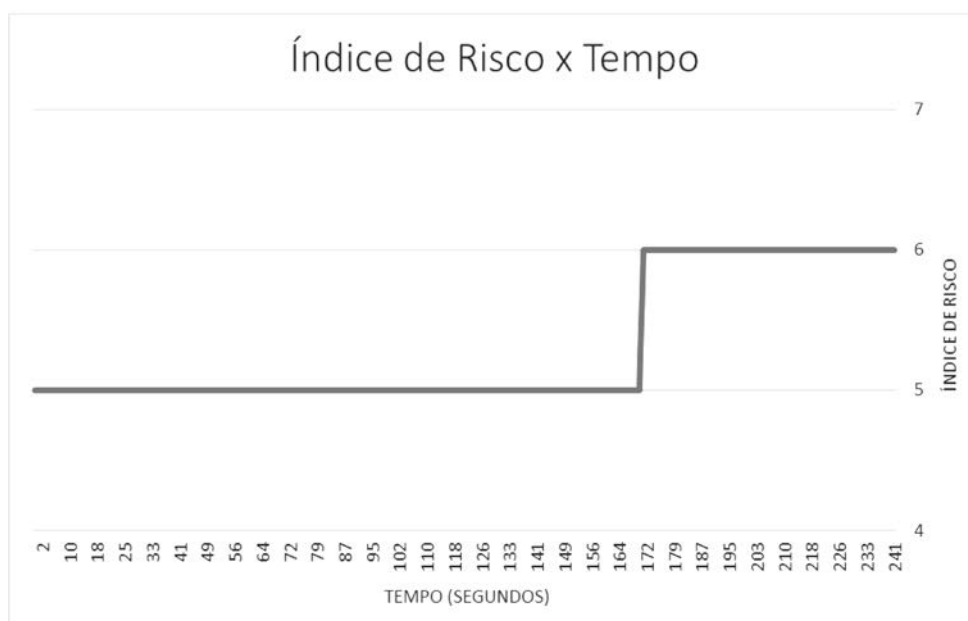


Figura 7. Gráfico do Nível de Risco x Tempo

5.3. Testes de Inundação

Os testes simulados mostraram que o sistema reage corretamente aos valores do índice pluviométrico medido. O índice é calculado com base na profundidade do rio, e o nível de água presente, sendo um valor percentual do quanto o mesmo está cheio. A partir de testes empíricos, os seguintes intervalos foram definidos:

- Índice de 0 a 0,5: Região segura
- Índice de 0,5 a 0,75: Região em alerta
- Índice superior a 0,75: Região em risco

Definindo o intervalo dessa forma, é possível manter a central de monitoramento atenta às regiões em alerta, passíveis de inundações, porém ainda não em risco, onde com nível superior a 75% a central de monitoramento será alertada, e de acordo com o seu julgamento o alerta poderá ser emitido. É importante perceber que tais intervalos devem ser estabelecidos pela autoridade de monitoramento, analisando e definindo o risco em que a região se encontrará para diferentes índices pluviométricos. Para a maquete apresentada o sistema respondeu adequadamente, indicando o risco conforme o nível pluviométrico subiu.

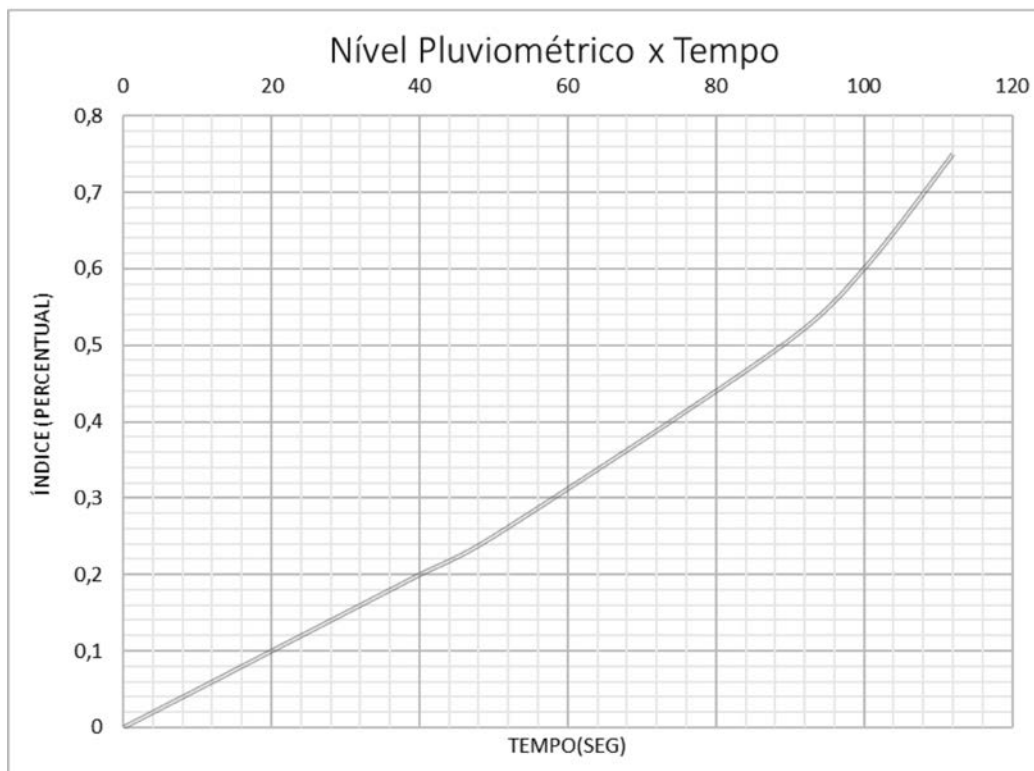


Figura 8. Gráfico do Nível Pluviométrico x Tempo

Testes realizados de inundação foram realizados na maquete, no local criado para tal experimento. A profundidade do rio é de 6,5cm e a altura do sensor ultrassônico é de 14,2cm. Assim, com o sensor medindo uma distância de 14,2cm, se conclui que o nível de água é 0, e o índice pluviométrico, portanto é de 0%. A partir desse nível do rio, a bomba foi ligada para a entrada de água; após o período de aproximadamente 1 minuto e 29 segundos, o sistema indicou a região como em alerta, com um índice

pluviométrico em 50%, e em 1 minuto e 52 segundos a região entrou em risco de inundação, representando 75% do rio cheio, com um crescente aumento, sendo assim necessário ao operador emitir um alerta aos usuários; esse teste está representado no gráfico da figura 8. O sistema reagiu corretamente ao aumento do nível, onde verificou-se a eficácia do sistema no monitoramento do desastre. O tempo adequado para evacuação deve ser definido pela autoridade responsável, onde os intervalos de risco para os níveis do rio devem ser ajustados de forma a indicar a situação perigosa com a antecedência desejada.

6. Considerações Finais

Com base nos resultados observados através da implementação do sistema e testes com a maquete, é possível concluir que o sistema consegue atender as necessidades e os objetivos previamente estipulados com relação a efetividade de evacuação de regiões de risco.

Nos testes realizados em laboratório o sistema consegue monitorar o estado atual do solo e indicar possíveis riscos com antecedência à ocorrência do evento, para a central de monitoramento, responsável por emitir o alerta e evacuar a região através do aplicativo para smartphones.

Em termos de implementação e adoção do sistema em larga escala, é importante perceber que em relação ao monitoramento de inundações são necessárias poucas alterações no modelo utilizado para garantir o funcionamento em qualquer tipo de margem de rio, lago ou córrego. Para desastres como deslizamento de terra, o sistema computacional funciona, porém com algumas limitações: a necessidade de um mapeamento prévio do solo e a adaptação do modelo matemático para cada local de implantação. Portanto, o modelo nem sempre deve ser aplicado indiscriminadamente para qualquer região.

O sistema foi testado e validado somente com um único módulo de sensores, através do desenvolvimento da maquete. Para o monitoramento de uma região real, é importante que sejam utilizados diversos módulos para obtenção das condições em diferentes pontos da encosta e do rio e a utilização de técnicas de redes de sensores para um processamento central no servidor e agregação dos dados, permitindo a análise da situação geral de toda a região, e não de somente alguns pontos.

Pode-se concluir que o sistema desenvolvido é aplicável em cenários reais de forma com que o sistema como um todo (sensores, interface de monitoramento e aplicativo), com os ajustes necessários, garanta uma otimização do tempo para evacuação das pessoas em uma região, evitando assim a perda de vidas.

7. Agradecimentos

Os autores agradecem ao Laboratório de Inovação em IoT FEI Telefônica Vivo pelo apoio financeiro à pesquisa e pela parceria Universidade - Empresa; convênio TA-004/14.

8. Referências

- Brito, G. (2013) “Modelo de Monitoramento de Deslizamento de Encostas Por Meio de Sensor Multiparamétrico”, Dissertação (Pós-graduação em Desenvolvimento em Processos Ambientais) – Universidade Católica de Pernambuco, Recife.
- Collins, T. (2004) “Disaster Risk for Floods: 1980 – 2000”, United Nations University. Tokyo: 2004. p.6. 56 f., <http://www.unu.edu/news/ehs/floods.doc> , 29 março 2016
- Gomide, R. (2013) “Só 10% dos moradores atendem a sirenes de evacuação em áreas de risco do Rio”. *Último Segundo, Rio de Janeiro*, 17/01/2013. <http://ultimosegundo.ig.com.br/brasil/rj/2013-01-17/so-10-dos-moradores-atendem-a-sirenes-de-evacuacao-em-areas-de-risco-do-rio.html>, 29 março 2016.
- Nelson, F. F., Guimarães, R. A., Montgomery, D. (2001), “Condicionantes Geomorfológicos dos Deslizamentos nas Encostas: Avaliação de Metodologias e Aplicação de Modelo de Previsão de Áreas Susceptíveis”. *Revista Brasileira de Geomorfologia, Rio de Janeiro*, v. 2, 2001, http://repositorio.unb.br/bitstream/10482/15860/1/ARTIGO_CondicionantesGeomorfologicosDeslizamentos.pdf, 29 março 2016.
- Nullis, C. (2014), “Atlas de Mortalidades e Perdas Econômicas por fatores de Clima e Desastres Naturais”. *Organização Meteorológica Mundial (OMM), Geneva*, n. 998, 11 Julho 2014. https://www.wmo.int/pages/mediacentre/press_releases/pr_998_en.html, 29 março 2016.
- Ramos, V. M., et al. (2002) “Aplicação do Modelo Shalstab, em ambiente Arcview, para o mapeamento de áreas susceptíveis a escorregamento raso na região do quadrilátero ferrífero (MG)”. *Espaço & Geografia*, v. 5, 2002. Disponível em: http://repositorio.unb.br/bitstream/10482/10656/1/ARTIGO_AplicacaoModeloShalstab.pdf, 29 março 2016
- Terzis, A. (2006) “Slip Surface Localization in Wireless Sensor Networks for Landslide Prediction”, <http://www.cs.jhu.edu/~terzis/landslide-camera.pdf>, 29 março 2016

Ferramentas de computação visual para apoio ao treinamento de expressões faciais por autistas: uma revisão de literatura

Thiago Pradi¹, Luciano Silva¹, Olga R. P. Bellon¹, Gustavo M. S. Dória²

¹Departamento de Informática – Grupo IMAGO de Pesquisa
Universidade Federal do Paraná – Curitiba, PR – Brasil.

²Departamento de Medicina Forense e Psiquiatria
Universidade Federal do Paraná – Curitiba, PR – Brasil

tpradi@inf.ufpr.br, {luciano, olga}@ufpr.br, gustavodoria@brturbo.com.br

Abstract. *Advances on research and application of visual computing on health sciences have become important allies in solving problems faced by individuals with Autism Spectrum Disorder (ASD). The difficulty in recognizing and producing emotions through the face often makes the social relationships of this population less adaptive and functional, which prejudice their autonomy. This paper reviews and discusses state-of-the-art works that use visual computing resources to provide solutions to challenges related to ASD. From this broad study, we raise questions and provide future directions for development of new tools to train recognition and production of facial expressions by children with autism, aiming both diagnosis and treatment.*

Resumo. *Os avanços nas pesquisas e aplicações de computação visual em ciências da saúde se tornaram importantes aliados na resolução de problemas enfrentados por indivíduos com Transtorno do Espectro Autista (TEA). Esta população apresenta limitações em reconhecer e expressar emoções através da face, o que torna suas relações sociais menos adaptativas e funcionais. Este artigo apresenta uma análise de trabalhos que utilizam recursos de computação visual para apoiar o incremento da capacidade de reconhecimento de emoções expressas pela face. A partir deste estudo, levanta-se questões e indica-se direções futuras para o desenvolvimento de novas ferramentas para treinar o reconhecimento e a produção de expressões faciais em crianças com autismo.*

1. Introdução

O autismo pertence ao grupo dos transtornos do neurodesenvolvimento, nos quais ocorre uma ruptura em processos fundamentais de socialização, comunicação e aprendizado do indivíduo [Klin 2006]. Novos estudos [Matson and Kozlowski 2011] apontam que o número de casos diagnosticados de autismo vem crescendo nos últimos anos. Em pesquisa recente do Centers for Disease Control and Prevention (CDC), estimou-se uma prevalência de uma em cada quarenta e cinco crianças, nos EUA, para o transtorno [Zablotsky et al. 2015]. Este crescimento está relacionado a diversos fatores, sobretudo às mudanças nos critérios de diagnóstico e conscientização da população em geral [Matson and Kozlowski 2011].

Entre as limitações apresentadas por indivíduos com TEA, pode-se citar a dificuldade na tarefa de reconhecimento e produção de expressões faciais

[Baron-Cohen et al. 1985]. Paul Ekman [Ekman and Friesen 1978] sugere que as habilidades para reconhecer e expressar emoções através da face são intrínsecas na raça humana. As emoções expressas através da face apresentam um papel importante na comunicação e interação entre seres humanos. Desde os primeiros meses de vida, recém-nascidos já interagem a partir de expressões faciais, utilizando-as como dicas essenciais para entendimento da comunicação [Picard et al. 2001].

Apesar do autismo não ter cura, abordagens terapêuticas podem ser eficazes na melhora do quadro, conferindo maior qualidade de vida e prevenindo o agravamento da doença. Por exemplo, recursos tecnológicos vêm sendo explorados com sucesso no caso de autistas que apresentam facilidade no uso de computadores, tablets e similares. Dentre essas iniciativas, um dos focos é incrementar a assertividade no reconhecimento e produção de expressões faciais.

Este trabalho apresenta uma revisão de trabalhos que utilizam recursos de computação visual para o treinamento em reconhecimento e produção de expressões faciais por autistas. As publicações inclusas nesta revisão foram avaliadas sob um conjunto de cinco critérios definidos pelos autores, dando ênfase nas contribuições dos recursos e métodos utilizados para o incremento desta habilidade. Este trabalho está organizado como se segue. A Seção 2 apresentando um comparativo das principais iniciativas que utilizam recursos computacionais para treinamento no reconhecimento e produção de expressões faciais. A Seção 3 discute os achados levantados com o presente estudo comparativo. A Seção 4 conclui o texto e sugere direções futuras.

2. Treinamento em reconhecimento e produção de expressões faciais por autistas

Sabe-se que autistas apresentam limitações que prejudicam a formulação da sua própria Teoria de Mente [Baron-Cohen et al. 1985]. A Teoria de Mente é a capacidade do indivíduo em atribuir estados mentais a si próprios e aos outros, diferenciar estes estados entre si, e interpretá-los. Conforme experimento [Gordon et al. 2014], a limitação que autistas apresentam para produzir expressões faciais não tem ligação com problemas motores ou musculares do rosto, mas sim com uma desconexão entre o estado mental de uma emoção e sua representação facial. Em [Tanaka et al. 2012] argumenta-se que autistas apresentam dificuldades em generalizar expressões faciais entre indivíduos, o que acaba agravando esse déficit.

Diversas pesquisas [Gordon et al. 2014, Tanaka et al. 2010, Deriso et al. 2012, Bölte et al. 2002, Silver and Oakes 2001] comprovam através de experimentos que é possível treinar as habilidades de reconhecimento e produção de expressões faciais por autistas, com a ajuda de ferramentas de computação visual capazes de reforçar a ligação entre o estado mental e a expressão facial. Para conhecer as iniciativas existentes, catalogá-las e identificá-las, foram efetuadas buscas nas bases PubMed, Web of Science e Google Scholar, utilizando os termos: autismo, computação, imagens, expressões faciais e reconhecimento de emoções. Nos resultados obtidos pelas buscas, foram selecionados somente trabalhos que descrevem ferramentas e softwares voltados à minimização do déficit apresentado por autistas no reconhecimento e produção de expressões faciais, com ano de publicação entre 2000 e 2015. No total, foram encontrados vinte e um trabalhos que atenderam aos critérios definidos.

Para obter uma visão geral das publicações encontradas, a Tabela 1 apresenta os artigos ordenados por data de publicação, avaliados a partir dos seguintes critérios: público alvo; forma de aprendizado de expressões faciais; tecnologia e interatividade; expressões faciais utilizadas e funcionalidades para avaliação da evolução do paciente. Em seguida, cada um dos critérios avaliados é detalhado (Seções 2.1 a 2.5).

2.1. Público Alvo

A escolha do público-alvo correto é um dos fatores chave para o sucesso de ferramentas computacionais para treinamento no reconhecimento e produção de expressões faciais. Como o público autista varia sob diversos aspectos - de acordo com a gravidade do quadro e idade dos indivíduos, por exemplo - acredita-se que funcionalidades e interfaces desenvolvidas de acordo com as preferências específicas de uma população possam gerar melhores resultados.

Ao avaliar o público alvo, o primeiro critério utilizado para comparação é a faixa etária. Como o autismo não tem cura e está presente durante toda a vida do indivíduo, é importante que ferramentas sejam definidas com contexto apropriados para cada faixa etária. A importância da definição da idade apropriada está relacionada ao fato de que crianças autistas possuem interesses e capacidades motoras diferentes de jovens e adultos autistas.

Outro critério considerado para comparação é a gravidade do transtorno. Neste quesito, baseado no DSM-5 [American Psychiatric Association 2013], o autismo pode ser dividido em três níveis: exige apoio (1), exige apoio substancial (2) e exige apoio muito substancial (3). A principal diferença entre estes níveis é o grau de comprometimento das funções básicas. Autistas do nível 1 apresentam nível intelectual dentro da normalidade e déficit intelectual leve ou inexistente, possibilitando sua alfabetização. Os autistas do nível 2 e 3 são indivíduos com os casos mais grave do transtorno, onde o déficit nas características afetadas pelo transtorno é maior, e geralmente está associado a um grau maior de deficiência intelectual, o que acaba comprometendo suas habilidades de comunicação e aprendizado.

Conforme apresentado no comparativo da Tabela 1, sete dos vinte e um trabalhos são focados em autistas do nível 1. Têm-se ainda os doze trabalhos que focam no público autista geral, porém apresentam interfaces onde a alfabetização do indivíduo é necessária para sua utilização, excluindo autistas do nível 2 e 3 do público de usuários. Existem ainda duas ferramentas que podem ser utilizadas por autistas do nível 2 e 3, porém estes trabalhos representam uma minoria, haja vista as dificuldades apresentadas no desenvolvimento de ferramentas computacionais para indivíduos que apresentam limitações profundas em suas habilidades cognitivas, de comunicação e de aprendizado.

Avaliando a faixa etária do público alvo, percebe-se que dezoito dos vinte e um trabalhos têm foco em crianças, enquanto três deles estudam adolescentes e jovens adultos autistas. Acredita-se que as ferramentas sejam focadas em crianças devido à crescente demanda por instrumentos de reabilitação para o público autista infantil, o que foi impulsionado pelo incremento diagnósticos precoces nos últimos anos [Matson and Kozlowski 2011]. Outro ponto importante é que a infância é a fase mais adequada para intervenções focadas no treinamento de expressões faciais e emoções, apresentando melhores resultados quando comparadas com intervenções efetuadas em

jovens e adultos [Aman 2005].

2.2. Forma de treinamento de expressões faciais

A forma de treinamento de expressões faciais compreende a metodologia utilizada para desenvolvimento de exercícios que melhorem a assertividade nas habilidades de autistas em reconhecer e produzir expressões faciais. Estas metodologias baseiam-se em achados da literatura médica, onde através de experimentos são demonstradas a efetividade dos métodos para treinamento de expressões faciais.

Categorizando-se as metodologias, pode-se estabelecer duas formas principais para treinar expressões faciais: reconhecimento de expressões faciais e produção de expressões faciais. No reconhecimento, são apresentados conteúdos (texto, imagens, sons ou vídeos) que estão associados a uma determinada emoção. A natureza dos conteúdos exibidos varia desde imagens da face de seres humanos até diálogos de situações sociais. Este conteúdo apresentado deve ser relacionado com uma expressão facial pelo autista. Esta metodologia é utilizada tanto em intervenções utilizando computadores quanto em intervenções através de fotografias, haja vista sua facilidade para ser executada.

O treinamento através da produção das expressões faciais é um tópico que vem sendo explorados em estudos recentes [Deriso et al. 2012, Gordon et al. 2014]. Esta metodologia só foi passível de inclusão em ferramentas computacionais devido aos avanços em métodos de visão computacional, que permitiram que computadores reconheçam expressões faciais em tempo real. A diferença nela é que ao invés do indivíduo reconhecer a expressão facial a partir de uma imagem, ele deve produzir, em sua própria face, a expressão requerida. O treinamento através de produção foi comprovado como sendo eficaz na melhoria das habilidades de reconhecimento de expressões faciais em [Deriso et al. 2012] e [Gordon et al. 2014]. Ainda, [Oberman et al. 2007] comprova que nossa habilidade de reconhecer expressões faciais está diretamente ligada com a habilidade de imitação das mesmas, o que reforça a importância da produção de expressões faciais.

Com base na Tabela 1, é possível visualizar que dezesseis das vinte e uma ferramentas encontradas treinam expressões faciais através do reconhecimento, enquanto somente cinco dos vinte e um trabalhos promovem o treinamento através da produção de expressões faciais. Acredita-se que esses resultados estejam relacionados com os esforços necessários para implementação do treinamento através da produção de expressões faciais, que só é possível ao implementar-se métodos de visão computacional para certificar-se que a expressão produzida está correta.

2.3. Tecnologia e interatividade

A tecnologia utilizada e sua forma de interação merece atenção ao desenvolver softwares para o público autista. Autistas apresentam atraso no desenvolvimento de habilidades motoras, com conseqüente dificuldade para a manipulação de objetos presentes na interação com computadores tradicionais, como mouses e teclados [Jasmin et al. 2009]. Esses problemas motores motivam o desenvolvimento de estratégias alternativas de interação, como telas sensíveis ao toque e métodos não motores utilizando câmeras.

Com base na Tabela 1, observa-se que quatorze dos vinte e um trabalhos foram desenvolvidos na plataforma Desktop, porém este padrão está sendo mudado ao longo

dos anos. Devido à popularização de novos dispositivos, como tablets e óculos de realidade virtual, novas ferramentas vêm se tornando disponíveis para auxiliar autistas no treinamento de expressões faciais [Xu and Tanaka 2014]. Visualiza-se uma mudança semelhante do ponto de vista da interação com o usuário, sendo que cada vez mais interfaces multimodais são utilizadas ao invés dos tradicionais mouses e teclados. Dezenas de ferramentas já utilizam interfaces multimodais para comunicação com o usuário, representando quarenta e sete por cento dos trabalhos. Essas interfaces estão

Artigo	N	I	FA	FI	P	EF	CS	RAP	CDP
[Silver and Oakes 2001]	G	C, O	R	MT	D	FTRS	X		
[Bölte et al. 2002]	1	O, U	R	MT	D	FTRNMSE	X		
[Tanaka et al. 2003]	G	C	R	MT	D	FTRNM			X
[Moore et al. 2005]	G	C, O	R	MT	D	FTRM	X		
[Beaumont and Sofronoff 2008]	1	C	R	MT	D	FTRNMS	X		
[Cockburn et al. 2008]	G	C	P	C	D	F			
[Grynszpan et al. 2008]	1	C	R	MT	D	HSAFSD	X		
[Madsen et al. 2008]	1	C	R	C, TS	M	Complex.			
[Finkelstein et al. 2009]	G	C, O	R	MT	D	FTRNMS	X		
[Abirached et al. 2011]	G	C, O	R, P	C, TS	M	FTRNMS	X		
[Hopkins et al. 2011]	1, 2	C	R	MT	D	FTRNMS			
[Deriso et al. 2012]	G	C	P	C	D	FTRNMS	X		
[Ryokai et al. 2012]	G	C	P	C, TS	M	-	X		
[Bekele et al. 2013]	G	O	R	MT	D	FTRNMSD			X
[Gay et al. 2013]	G	C	R	C, TS	M	FTRNMS	X	X	
[Kandalaf et al. 2013]	1	U	R	MT	D	-	X		
[Christinaki et al. 2014]	G	C	R	C	C	FTRMS			
[Harrold et al. 2014]	1	C	P	C	M	FTRNMS			
[Serret et al. 2014]	1, 2	C	R	MT	C	FTRNMSOE			
[Wang et al. 2014]	G	C	R	ORV	D	-			
[Chen et al. 2015]	1	C	R	C	D	FTRNMS	X		

Table 1. Comparação entre os trabalhos para treinamento de expressões faciais.

- Cabeçalho: N - Nível; I - Idade; FA - Forma de Aprendizado; FI - Forma de Interação; P - Plataforma; EF - Expressões Faciais; CS - Contexto Social; RAP - Relatório para Acompanhamento do Paciente; CDP - Coleta de Dados do Paciente.
- Nível: G - Geral; 1- Exige apoio; 2- Exige apoio substancial; 3- Exige apoio muito substancial.
- Idade: C - Crianças; O - Adolescentes; U - Adultos.
- Forma de Aprendizado: R - Reconhecimento; P - Produção.
- Forma de Interação: C - Câmera; ORV - Óculos para Realidade Virtual; TS - Touch Screen; MT - Mouse e Teclado.
- Plataforma: D - Desktop; ORV - Óculos para Realidade Virtual; C - Console; M - Dispositivos Móveis.
- Exp. Faciais: F - felicidade; T - tristeza; R - raiva, N - nojo, M - medo; S - surpresa; D - desprezo; O - dor; E - Neutro; Complex - Emoções complexas.

sendo utilizadas com auxílio da visão computacional, através do reconhecimento de gestos e expressões nos utilizadores com ajuda de câmeras de vídeo.

2.4. Expressões faciais

Ao treinar habilidades de reconhecimento e produção de expressões faciais em autistas, é importante definir quais expressões faciais serão utilizadas. Conforme mencionado previamente, autistas apresentam limitações na formação da teoria da mente, decorrentes de dificuldades no reconhecimento e produção de expressões faciais. Tais expressões englobam não apenas as seis básicas definidas por Ekman (felicidade, tristeza, surpresa, medo, raiva e nojo) [Ekman and Friesen 1978], mas também expressões complexas que podem representar dor [Minio-Paluello et al. 2009].

Com a mesma importância que o reconhecimento da expressão facial para interpretação de sua emoção associada, está o entendimento do contexto social em que ela está inserida. Por exemplo, uma pessoa em uma situação amedrontadora demonstrando raiva pode ser compreendida como expressando medo por observadores [Carroll and Russell 1996]. Autistas também apresentam limitações no entendimento de situações sociais em comparação com indivíduos não autistas [Santos et al. 2012], o que dificulta o entendimento de pistas importantes para reconhecimento da expressão facial e de seu significado. Quanto a estas pistas, seu treinamento utilizando ferramentas computacionais é efetivo em autistas, visto que podem prover um ambiente controlado para que ele pratique suas habilidades sem a pressão e imprevisibilidade do mundo real.

Dos vinte e um trabalhos encontrados, três não apresentaram dados sobre quais as expressões faciais utilizadas. Dezessete dos vinte e um trabalhos utilizaram as seis expressões faciais básicas descritas por Ekman [Ekman and Friesen 1978], ou uma variação das mesmas. Entre as variações encontradas, podem-se citar seis trabalhos que removeram uma ou mais expressões faciais. Outra variação das seis expressões faciais básicas é a inclusão de expressões como as associadas a dor [Serret et al. 2014] e desprezo [Bekele et al. 2013]. Como variação total das expressões básicas definidas por Ekman, pode-se citar o trabalho de [Madsen et al. 2008], que ensina seis expressões faciais complexas.

2.5. Funcionalidades para avaliação da evolução do paciente

Além das funcionalidades para o treinamento de expressões faciais, ferramentas computacionais podem prover funcionalidades importantes para que profissionais da área de saúde possam acompanhar a evolução de seus pacientes. Estas funcionalidades também possibilitam o registro longitudinal de dados que podem contribuir para um melhor entendimento do problema. As informações utilizadas para o acompanhamento podem ser de diversos tipos, como dados sobre o tempo para execução de determinado desafio, informações sobre erros e acertos do paciente ao executar determinada tarefa e vídeos do paciente utilizando a ferramenta. Para acessar estas informações de forma simplificada, relatórios são incluídos nas ferramentas, possibilitando seu acesso pelo médico ou terapeuta responsável. Estes relatórios devem estar devidamente protegidos por mecanismos de autenticação, garantindo a privacidade do paciente.

Para obter mais informações sobre a presença de relatórios para visualização de dados obtidos nas ferramentas, foram buscadas menções que descrevam esta

funcionalidade nos trabalhos encontrados. Conforme apresentado na Tabela 1, somente a pesquisa de [Gay et al. 2013] conta com um relatório para acompanhamento da evolução do paciente. Esse relatório apresenta informações sobre os acertos e erros do paciente ao reconhecer as expressões faciais.

No contexto da coleta de dados dos pacientes, os únicos trabalhos que apresentam alguma funcionalidade são os desenvolvidos por Bekele [Bekele et al. 2013] e Tanaka [Tanaka et al. 2003]. Em [Bekele et al. 2013], são utilizados equipamentos para coletar dados fisiológicos e de rastreamento ocular do paciente ao reconhecer a expressão facial apresentada. Já no trabalho de Tanaka [Tanaka et al. 2003], foi desenvolvido um mecanismo de coleta de dados do paciente em um estudo mais moderno da ferramenta [Tanaka et al. 2010], porém os dados coletados são somente sobre erros e acertos do paciente.

3. Discussão

Com base nas ferramentas encontradas e avaliadas com o objetivo de treinar habilidades de reconhecimento e produção de expressões faciais por autistas, é possível concluir que este campo está sendo aperfeiçoado ao longo dos anos, de diferentes formas. Observando o comparativo entre os trabalhos, percebe-se que essas melhorias foram efetuadas principalmente nos critérios de plataforma e interatividade, tornando as ferramentas mais inclusivas e facilitando o seu uso pelo público autista. Apesar de haver melhorias significativas em alguns critérios, percebe-se pouca atenção para a evolução de outros quesitos.

Do ponto de vista tecnológico, é possível visualizar que nenhuma das ferramentas foi desenvolvida para a plataforma web. Esta plataforma apresenta distribuição e acesso simplificados, não necessitando da instalação de nenhum software além do navegador de internet para sua utilização. Como navegadores estão disponíveis na maioria dos dispositivos, ferramentas desenvolvidas na plataforma web podem ser utilizadas em notebooks, celulares e tablets. Com o recente avanço do padrão HTML5, o acesso a câmeras e o armazenamento de dados no navegador de internet foram simplificados, permitindo que ferramentas computacionais explorem estes recursos.

Ao avaliar a metodologia para treinamento de expressões faciais, percebe-se que a produção de expressões faciais vem ganhando espaço, sendo explorada como alternativa aos métodos tradicionais de reconhecimento de expressões faciais. Tanto o reconhecimento [Tanaka et al. 2010] quanto a produção [Deriso et al. 2012] possuem trabalhos na literatura que comprovam sua efetividade na melhoria das habilidades de reconhecimento de expressões faciais. Acredita-se que a combinação de ambos seja o melhor caminho, provendo diferentes exercícios para o treinamento no reconhecimento e produção de expressões faciais.

Do ponto de vista da psicologia, a grande maioria das ferramentas continuam utilizando somente as seis expressões faciais básicas definidas por Ekman. No campo da psicologia, o conhecimento sobre expressões faciais vem sendo aperfeiçoando através de novas descobertas. Atualmente, existem trabalhos apontando que expressões faciais e seu entendimento não são culturalmente universais [Gendron et al. 2014], o que reforça a ideia de que ferramentas devem ser customizadas para seu público alvo. Fatores como contexto social em que a expressão facial é apresentada devem ser levados em conta no

desenvolvimento dos desafios das ferramentas.

No quesito de dados para avaliação da evolução do paciente, nota-se pouca atenção para o desenvolvimento dessas funcionalidades nas ferramentas encontradas. Somente o trabalho de [Gay et al. 2013] apresentou relatórios para visualização deste tipo de dados. Estes relatórios são extremamente importantes para profissionais da área da saúde, porque permitem que o profissional tenha acesso, de forma simples, aos dados coletados sobre a utilização da ferramenta pelo paciente. Deve-se ressaltar a importância da inclusão de relatórios que permitam que esses dados sejam avaliados de forma global, auxiliando no entendimento de questões relacionadas ao transtorno autista como um todo.

Aprofundando-se nos dados coletados para avaliação da evolução do paciente, outros dados não relacionados a acertos e erros nas tarefas podem ser processados por ferramentas computacionais. Como exemplo, pode-se citar a gravação de vídeos do usuário executando desafios e informações sobre a trajetória do rastreamento ocular ao reconhecer uma expressão facial. Quanto aos vídeos gravados, seu principal uso seria na futura avaliação do paciente e da população autista em geral. Como os vídeos de utilização das ferramentas podem ser muito longos, tornando a análise laboriosa, acredita-se no potencial de métodos de computação visual para sumarizar e extrair as partes mais relevantes dos vídeos.

Quanto ao rastreamento ocular, diversos trabalhos na literatura reforçam sua importância como critério para melhor entendimento do autismo [Falck-Ytter et al. 2013]. Focando na temática de expressões faciais, sabe-se que o rastreamento ocular é eficaz para auxiliar na compreensão do método utilizado por pacientes autistas para reconhecer expressões faciais [Bekele et al. 2014, Grynszpan et al. 2012]. Em ferramentas computacionais, seu uso pode ser ainda maior, visto que o percurso efetuado pelo autista durante a realização dos desafios propostos pode ser utilizado para otimizar a interface da ferramenta.

A implementação do rastreamento ocular é um desafio enfrentado pela área de visão computacional, sendo que [Hansen and Ji 2010] apresentam uma revisão de literatura dos principais métodos utilizados para resolução do problema. Existem hardwares específicos para efetuar rastreamento ocular disponíveis no mercado, como o Tobii X120 utilizado em [Bekele et al. 2013], que apresentam alta acurácia e confiabilidade nos dados obtidos. Apesar dos pontos positivos, este tipo de equipamento apresenta alto custo e não está integrado na maioria dos computadores. Webcams e câmeras de profundidade também podem ser utilizadas como dispositivos para rastreamento ocular, obtendo acurácia e confiabilidade similares aos dispositivos comerciais [San Agustin et al. 2010].

Baseado na melhoria dos aspectos apontados, visualiza-se um caminho futuro de novas ferramentas para treinamento de em reconhecimento e produção de expressões faciais por autistas. Ao incorporar as melhorias nos critérios citados, as novas ferramentas desenvolvidas possuem potencial na modernizando de intervenções terapêuticas para incremento das habilidades de pacientes autistas em reconhecer e produzir expressões faciais.

4. Conclusão

Este trabalho apresenta uma revisão dos estudos que utilizam recursos computacionais no desenvolvimento de ferramentas para o treinamento no reconhecimento e produção de expressões faciais por autistas, procurando oferecer uma visão crítica da área para novos pesquisadores. Foram encontrados vinte e um artigos que satisfizeram as condições buscadas, sendo eles avaliados com base em cinco critérios: público alvo; forma de aprendizado de expressões faciais; forma de interação com a ferramenta; expressões faciais utilizadas e funcionalidades para avaliação da evolução do paciente.

Com base no trabalho comparativo efetuado, este artigo apresenta critérios de melhoria em quatro dos cinco quesitos avaliados nas ferramentas. Entre as melhorias, são citadas as seguintes: inclusão de novos conceitos descobertos pela área de psicologia; inclusão da produção de expressões faciais como parte do treinamento; utilização da tecnologia web para permitir a portabilidade e o acesso simplificado às ferramentas; e a inclusão de funcionalidades médicas para obtenção de dados que auxiliem na avaliação da evolução do paciente, como a captura de vídeos do usuário e dados de rastreamento ocular ao avaliar determinada expressão facial.

Como um dos trabalhos futuros na área, aponta-se para o desenvolvimento de uma ferramenta computacional em um projeto multidisciplinar, envolvendo profissionais das áreas de ciência da computação e ciências da saúde. Esta ferramenta deve incorporar as melhorias nos critérios apontados, possibilitando uma intervenção moderna e eficiente. Através da ferramenta, o profissional responsável poderá acompanhar o progresso do paciente ao longo de sua utilização, possibilitando avaliar sua evolução e adequar a intervenção para as necessidades daquele paciente.

References

- Abirached, B., Zhang, Y., Aggarwal, J. K., Tamersoy, B., Fernandes, T., Miranda, J., and Orvalho, V. (2011). Improving communication skills of children with asds through interaction with virtual characters. In *Serious Games and Applications for Health (SeGAH), 2011 IEEE 1st International Conference on*, pages 1–4. IEEE.
- Aman, M. G. (2005). Treatment planning for patients with autism spectrum disorders. *Journal of Clinical Psychiatry*, 66:38.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Autor, Washington, DC, 5th ed. edition.
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind?". *Cognition*, 21(1):37–46.
- Beaumont, R. and Sofronoff, K. (2008). A multi-component social skills intervention for children with asperger syndrome: The junior detective training program. *Journal of Child Psychology and Psychiatry*, 49(7):743–753.
- Bekele, E., Crittendon, J., Zheng, Z., Swanson, A., Weitlauf, A., Warren, Z., and Sarkar, N. (2014). Assessing the utility of a virtual environment for enhancing facial affect recognition in adolescents with autism. *Journal of autism and developmental disorders*, 44(7):1641–1650.

- Bekele, E., Zheng, Z., Swanson, A., Crittendon, J., Warren, Z., and Sarkar, N. (2013). Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):711–720.
- Bölte, S., Feineis-Matthews, S., Leber, S., Dierks, T., Hubl, D., and Poustka, F. (2002). The development and evaluation of a computer-based program to test and to teach the recognition of facial affect. *International Journal of Circumpolar Health*, 61.
- Carroll, J. M. and Russell, J. A. (1996). Do facial expressions signal specific emotions? judging emotion from the face in context. *Journal of personality and social psychology*, 70(2):205.
- Chen, C.-H., Lee, I.-J., and Lin, L.-Y. (2015). Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Research in developmental disabilities*, 36:396–403.
- Christinaki, E., Vidakis, N., and Triantafyllidis, G. (2014). A novel educational game for teaching emotion identification skills to preschoolers with autism diagnosis. *Computer science and information systems*, 11(2):723–743.
- Cockburn, J., Bartlett, M., Tanaka, J., Movellan, J., Pierce, M., and Schultz, R. (2008). Smilemaze: A tutoring system in real-time facial expression perception and production in children with autism spectrum disorder. In *ECAG 2008 Workshop Facial and Bodily Expressions for Control and Adaptation of Games*, page 3.
- Deriso, D. M., Susskind, J., Tanaka, J., Winkielman, P., Herrington, J., Schultz, R., and Bartlett, M. (2012). Exploring the facial expression perception-production link using real-time automated facial expression recognition. In *Computer Vision—ECCV 2012. Workshops and Demonstrations*, pages 270–279. Springer.
- Ekman, P. and Friesen, W. V. (1978). *Manual for the facial action coding system*. Consulting Psychologists Press.
- Falck-Ytter, T., Bölte, S., and Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of neurodevelopmental disorders*, 5(1):28.
- Finkelstein, S. L., Nickel, A., Harrison, L., Suma, E., Barnes, T., et al. (2009). Emotion: A new game design to teach emotion recognition and programming logic to children using virtual humans. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 249–250. IEEE.
- Gay, V., Leijdekkers, P., Agcanas, J., Wong, F., and Wu, Q. (2013). Capturemyemotion: helping autistic children understand their emotions using facial expression recognition and mobile technologies. *BLED 2013 Proceedings*.
- Gendron, M., Roberson, D., van der Vyver, J. M., and Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, 14(2):251.
- Gordon, I., Pierce, M. D., Bartlett, M. S., and Tanaka, J. W. (2014). Training facial expression production in children on the autism spectrum. *Journal of autism and developmental disorders*, 44(10):2486–2498.

- Grynszpan, O., Martin, J.-C., and Nadel, J. (2008). Multimedia interfaces for users with high functioning autism: An empirical investigation. *International Journal of Human-Computer Studies*, 66(8):628–639.
- Grynszpan, O., Nadel, J., Martin, J.-C., Simonin, J., Bailleul, P., Wang, Y., Gepner, D., Le Barillier, F., and Constant, J. (2012). Self-monitoring of gaze in high functioning autism. *Journal of autism and developmental disorders*, 42(8):1642–1650.
- Hansen, D. W. and Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(3):478–500.
- Harrold, N., Tan, C. T., Rosser, D., and Leong, T. W. (2014). Copyme: a portable real-time feedback expression recognition game for children. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1195–1200. ACM.
- Hopkins, I. M., Gower, M. W., Perez, T. A., Smith, D. S., Amthor, F. R., Wimsatt, F. C., and Biasini, F. J. (2011). Avatar assistant: improving social skills in students with an asd through a computer-based intervention. *Journal of autism and developmental disorders*, 41(11):1543–1555.
- Jasmin, E., Couture, M., McKinley, P., Reid, G., Fombonne, E., and Gisel, E. (2009). Sensori-motor and daily living skills of preschool children with autism spectrum disorders. *Journal of autism and developmental disorders*, 39(2):231–241.
- Kandalaf, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., and Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders*, 43(1):34–44.
- Klin, A. (2006). Autismo e síndrome de asperger: uma visão geral. *Revista Brasileira de Psiquiatria*, 28(1):3–11.
- Madsen, M., El Kaliouby, R., Goodwin, M., and Picard, R. (2008). Technology for just-in-time in-situ learning of facial affect for persons diagnosed with an autism spectrum disorder. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 19–26. ACM.
- Matson, J. L. and Kozlowski, A. M. (2011). The increasing prevalence of autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5(1):418–425.
- Minio-Paluello, I., Baron-Cohen, S., Avenanti, A., Walsh, V., and Aglioti, S. M. (2009). Absence of embodied empathy during pain observation in asperger syndrome. *Biological psychiatry*, 65(1):55–62.
- Moore, D., Cheng, Y., McGrath, P., and Powell, N. J. (2005). Collaborative virtual environment technology for people with autism. *Focus on Autism and Other Developmental Disabilities*, 20(4):231–243.
- Oberman, L. M., Winkielman, P., and Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social neuroscience*, 2(3-4):167–178.
- Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1175–1191.

- Ryokai, K., Raffle, H., and Kowalski, R. (2012). Storyfaces: pretend-play with ebooks to support social-emotional storytelling. In *Proceedings of the 11th International Conference on Interaction Design and Children*, pages 125–133. ACM.
- San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W., and Hansen, J. P. (2010). Evaluation of a low-cost open-source gaze tracker. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, pages 77–80. ACM.
- Santos, A., Chaminade, T., Da Fonseca, D., Silva, C., Rosset, D., and Deruelle, C. (2012). Just another social scene: evidence for decreased attention to negative social scenes in high-functioning autism. *Journal of autism and developmental disorders*, 42(9):1790–1798.
- Serret, S., Hun, S., Iakimova, G., Lozada, J., Anastassova, M., Santos, A., Vesperini, S., and Askenazy, F. (2014). Facing the challenge of teaching emotions to individuals with low-and high-functioning autism using a new serious game: a pilot study. *Molecular autism*, 5(1):37.
- Silver, M. and Oakes, P. (2001). Evaluation of a new computer intervention to teach people with autism or asperger syndrome to recognize and predict emotions in others. *Autism*, 5(3):299–316.
- Tanaka, J. W., Lincoln, S., and Hegg, L. (2003). A framework for the study and treatment of face processing deficits in autism. *The development of face processing*, pages 101–119.
- Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M. D., and Schultz, R. T. (2010). Using computerized games to teach face recognition skills to children with autism spectrum disorder: The let’s face it! program. *Journal of Child Psychology and Psychiatry*, 51(8):944–952.
- Tanaka, J. W., Wolf, J. M., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S. S., South, M., McPartland, J. C., et al. (2012). The perception and identification of facial emotions in individuals with autism spectrum disorders using the let’s face it! emotion skills battery. *Journal of Child Psychology and Psychiatry*, 53(12):1259–1267.
- Wang, X., Desalvo, N., Gao, Z., Zhao, X., Lerman, D. C., Gnawali, O., and Shi, W. (2014). Eye contact conditioning in autistic children using virtual reality technology. In *Pervasive Computing Paradigms for Mental Health*, pages 79–89. Springer.
- Xu, B. and Tanaka, J. W. (2014). Teaching children with autism to recognize faces. In *Comprehensive Guide to Autism*, pages 1043–1059. Springer.
- Zablotsky, B., Black, L., Maenner, M., Schieve, L., and Blumberg, S. (2015). Estimated prevalence of autism and other developmental disabilities following questionnaire changes in the 2014 national health interview survey. *National health statistics reports*, (87):1–21.

Uma Contribuição à Reabilitação Cardíaca Explorando a Identificação de Situações na IoT

Douglas Scheunemann¹, Adenauer Yamin¹, João Lopes², Cláudio Geyer²

¹Universidade Católica de Pelotas (UCPEL)
Pelotas – RS – Brasil

²Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

scheunemann.d.a@gmail.com, adenauer.yamin@ucpel.edu.br

{jlblopes, geyer}@inf.ufrgs.br

Abstract. *The Internet of Things (IoT) has influenced the development of computational systems, enabling a more proactive interaction with users, expanding features as mobility and availability. In this scenario increases the demand for applications that can recognize the user's context and can provide situation-based services. The identification of situations is a research challenge for applications in the IoT, given the complexity of the relationships that must be established and processed. This paper presents an approach for the identification of situations, which is integrated with the EXEHDA middleware. The proposed approach provides the collaboration between the middleware and the applications. A case study in the health area was developed to evaluate the architecture. This case study was focused on cardiac rehabilitation.*

Resumo. *A Internet das Coisas (IoT) vem influenciando a maneira como os sistemas computacionais são desenvolvidos, possibilitando uma interação mais proativa com os usuários, expandindo características de mobilidade e disponibilidade. Nesse cenário, cresce a demanda por aplicações que possam reconhecer o contexto do usuário e fornecer serviços baseados em sua situação. A identificação de situações representa um desafio de pesquisa para aplicações em IoT, dada a complexidade das relações que precisam ser estabelecidas e processadas. Neste trabalho é apresentada uma abordagem de processamento para identificação de situações integrada ao middleware EXEHDA. A abordagem de processamento proposta propicia a colaboração entre o middleware e as aplicações na identificação de situações. Como forma de avaliação da arquitetura foi executado um estudo de caso na área de saúde com foco em reabilitação cardíaca.*

1. Introdução

A reabilitação cardíaca pode envolver diversas terapias, incluindo administração de medicamentos, aconselhamento nutricional e também a prescrição de atividades físicas. A reabilitação cardíaca através de atividades físicas é considerada uma terapia central. Estudos indicam que a reabilitação baseada em exercícios físicos foi associada a uma redução de 20 a 30% nas taxas de mortalidade, quando comparada com cuidados sem exercício [Rabelo et al. 2006].

Um aspecto que deve ser considerado na terapia através de exercícios físicos diz respeito a respostas desproporcionais na frequência cardíaca. Isto pode indicar uma situação de risco para o paciente. Desta forma, o reconhecimento da atividade física e sua correlação com a frequência cardíaca pode permitir uma recuperação mais segura [Negrão and Barreto 2010].

Nesse sentido, o emprego de sistemas computacionais para detecção autônoma de atividades e a correlação destas com parâmetros fisiológicos pode minimizar a necessidade de intervenção do próprio paciente ou do terapeuta na identificação de situações de risco.

Avanços em tecnologias de sensores, atuadores e protocolos de rede sem fio tem permitido o desenvolvimento de soluções de computação embarcada para diferentes finalidades. Dispositivos como utensílios domésticos, carros, equipamentos médicos e *smartphones* podem ser conectados e interagir a qualquer momento, em qualquer lugar com qualquer pessoa ou outro dispositivo através de uma rede unificada, denominada de Internet das Coisas (IoT) [Razzaque et al. 2015].

O crescimento do número de dispositivos conectados, propiciado pelos avanços em IoT, fez com que a quantidade de dados coletados e disponibilizados para as aplicações crescesse significativamente. Porém, para muitas aplicações os contextos de interesse podem ser obtidos através da fusão de dados de sensores, estabelecendo uma visão mais abstrata do contexto atual, permitindo a inferência da situação do usuário e do próprio sistema [Perera et al. 2013].

Nesse cenário, um dos principais desafios de pesquisa em IoT é a ciência de situação. Uma situação corresponde a uma visão compreensível e em alto nível de abstração do contexto de interesse da aplicação. Esta visão é resultante de dados de contexto coletados de sensores distribuídos pelo ambiente [Bellavista et al. 2012, Knappmeyer et al. 2013].

A principal contribuição deste trabalho é uma abordagem de processamento, integrada ao middleware EXEHDA (*Execution Environment for Highly Distributed Applications*) [Lopes et al. 2014a], chamada EXEHDA-CS (EXEHDA - Collaborative Situations), a qual provê ciência de situação para aplicações em IoT, permitindo a colaboração entre o middleware e as aplicações na composição e parametrização dos mecanismos de processamento de contexto utilizados na identificação de situações.

A arquitetura proposta foi avaliada através de um estudo na área de saúde com foco em reabilitação cardíaca. Este estudo de caso permitiu que funcionalidades relativas a parametrização de mecanismos de processamento e colaboração entre a aplicação e middleware fossem avaliadas.

O texto deste artigo está estruturado da seguinte forma: na seção 2 são abordadas características do middleware EXEHDA e da abordagem de processamento proposta. Na seção 3 é apresentado o estudo de caso que foi elaborado para avaliar a abordagem proposta. Na seção 4 são apresentados os trabalhos relacionados. Na seção 5 são apresentadas as considerações finais sobre o trabalho.

2. EXEHDA-CS: Concepção da Abordagem Proposta

O middleware EXEHDA possui dois tipos de servidores: (i) Servidor de Borda, responsável por interagir com ambiente através de sensores e atuadores; e (ii) Servidor de Con-

texto, responsável por prover funcionalidades para ciência de situação. Estes servidores são alocados em células do ambiente gerenciado pelo EXEHDA, onde cada célula possui um Servidor de Contexto e pode possuir vários Servidores de Borda. Na Figura 1 pode ser vista a disposição destes servidores no ambiente.

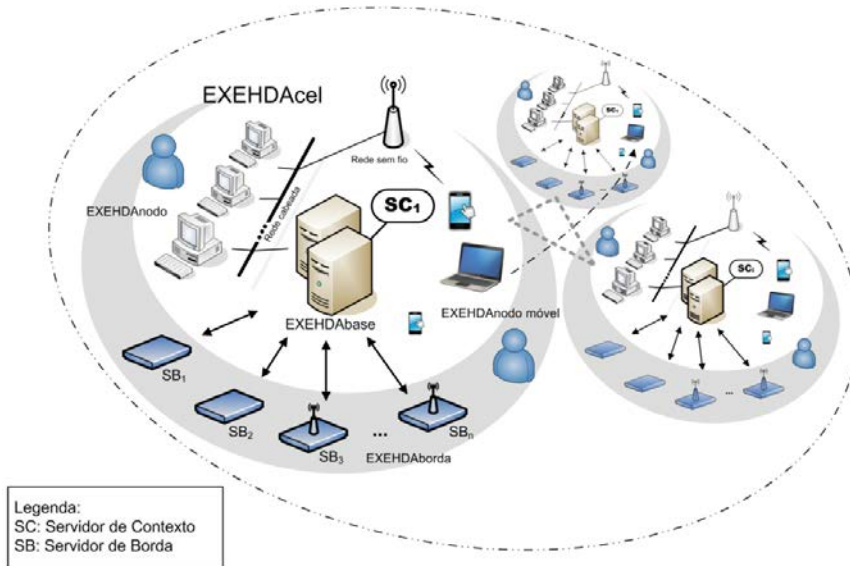


Figura 1. Ambiente gerenciado pelo EXEHDA.

O foco de deste trabalho é a criação de mecanismos para configuração e acesso ao Servidor de Contexto, permitindo que a identificação de situações possa ser feita de maneira colaborativa com as aplicações. Na Subseção 2.1 é apresentada uma visão geral sobre o Servidor de Contexto do EXEHDA e na Subseção 2.2 é apresentada a abordagem de processamento proposta e as APIs criadas para possibilitar a configuração e o uso do middleware pelas aplicações.

2.1. Servidor de Contexto

O Servidor de Contexto é formado por cinco módulos: Aquisição, Atuação, Notificação, Comunicação e Processamento [Lopes et al. 2014b], os quais podem ser visualizados na Figura 2 e são descritos a seguir.

Módulo de Aquisição: é responsável por prover suporte à captura das informações contextuais, coletadas pelos Servidores de Borda considerando sensores lógicos, implementados através de interfaces de software, e sensores físicos providos por interfaces de hardware.

Módulo de Atuação: controla a ativação, desativação e configuração dos atuadores, como consequência de uma notificação de outros módulos do Servidor de Contexto. Esse módulo recebe o identificador do atuador envolvido e os parâmetros operacionais a serem utilizados e interopera com os Servidores de Borda para disparo dos atuadores pertinentes.

Módulo de Processamento: tem como principal função realizar as tarefas pertinentes ao tratamento das informações contextuais, bem como dos eventos para identificar situações de interesse das aplicações. Para execução destas tarefas, os componentes de

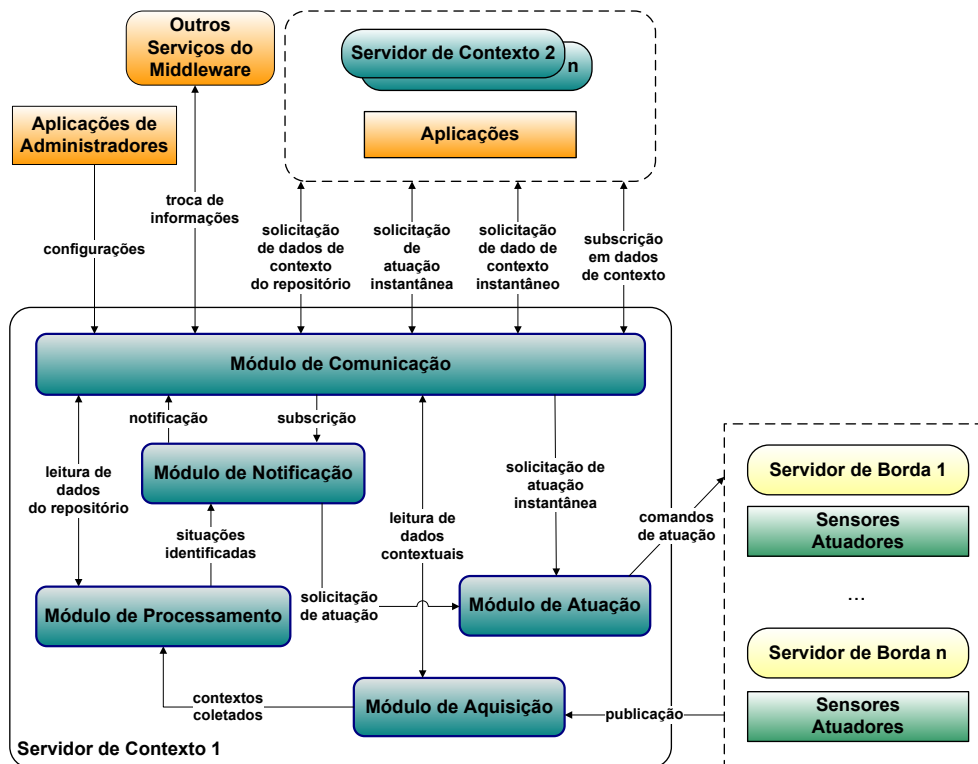


Figura 2. Arquitetura do servidor de contexto.

software que constituem esse módulo interagem com os modelos de representação do contexto.

Módulo de Notificação: é responsável por notificar o resultado do processamento das informações de contexto e da consequente identificação de situações realizada pelo Módulo de Processamento. Esse módulo opera recebendo subscrições dos serviços e/ou aplicações que desejem notificações a respeito dos estados contextuais e das situações identificadas, interoperando através do Módulo de Comunicação.

Módulo de Comunicação: é utilizado por Servidores de Contexto remotos e/ou aplicações para solicitação de situações constituídas por contextos de interesse correlacionados, dados contextuais e/ou o disparo de atuadores. Na arquitetura proposta neste trabalho será explorada uma API (*Application Programming Interface*) do tipo REST (*Representational State Transfer*) criada para suportar as etapas de configuração e acesso através do módulo de comunicação.

2.2. Abordagem Colaborativa para Identificação Situações

A abordagem de processamento colaborativa proposta neste trabalho contempla dois fluxos de dados contextuais: (i) o middleware executa a coleta de dados contextuais do meio de forma autônoma; (ii) as aplicações podem disponibilizar dados de contexto e/ou critérios que serão considerados quando do processamento combinado dos dados contextuais coletados.

A estrutura de processamento foi modelada para permitir acesso através de APIs do tipo REST [Fielding 2000]. Na Figura 3 pode ser vista a representação de um componente de processamento elementar da arquitetura <Componente-Processamento/id>.

Para compor o fluxo de processamento contextual são criadas instâncias dos componentes de processamento, alocadas em URIs (Uniform Resource Identifier) do tipo **<Instancia-Processamento/id>**.

A execução das instâncias de processamento é disparada por eventos de modificação nos dados de entrada. Cada instância deve possuir como entrada ao menos uma URI para sensor, contexto processado, ou contexto externo. A saída do processamento é armazenada em uma URI do tipo **<Contexto-Processado/id>**. Em cada instância de processamento estão envolvidos quatro tipos de URIs:

- **Sensores** <Sensor/id>: sensor disponibilizado por um Servidor de Borda.
- **Contextos Processados** <Contexto-Processado/id>: saída de uma instância de um componente de processamento, representa uma informação de contexto processado que poderá ser utilizada na identificação de situações ou composição de outros contextos processados.
- **Contextos Externos** <Contexto-Externo/id>: dados de contexto obtidos de aplicações externas, recebidos através do Módulo de Comunicação.
- **Componente de Processamento** <Componente-Processamento/id>: rotina de software armazenada no repositório do middleware.

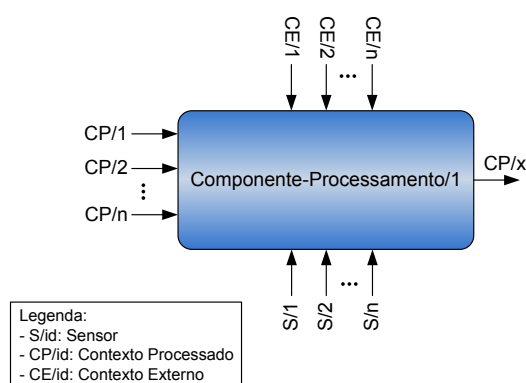


Figura 3. Componente de processamento elementar.

Como pode ser visto na Figura 3, os componentes de processamento de contexto podem receber como dados de entrada variáveis de contexto processadas por outros módulos, permitindo a criação de arranjos de processamento conforme mostra a Figura 4.

A descrição da API REST criada pode ser vista na Tabela 1. Os recursos quando acessados sem um ID, possuem somente o comando GET, para o qual é retornada a lista de recursos disponíveis. Quando é especificado um ID, estão disponíveis os comandos GET, POST, PUT e DELETE, os quais possuem o comportamento padrão do protocolo HTTP. Os sensores possuem somente o comando GET porque o seu gerenciamento é feito através de uma API do servidor de borda.

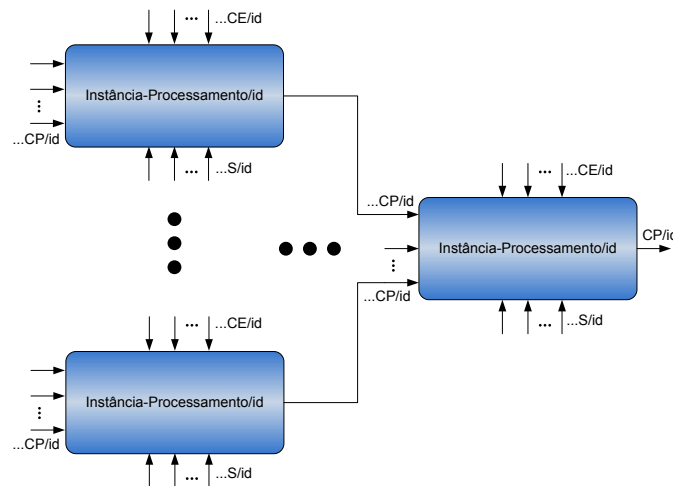


Figura 4. Modelo de processamento de contexto combinando múltiplas instâncias de componentes de processamento.

Tabela 1. Recursos da API REST da arquitetura proposta.

URI	Comandos	Descrição
Sensor	GET	Lê o conjunto de Sensores Disponíveis.
Sensor/id	GET	Lê os dados de um sensor.
Contexto-Processado	GET	Lê o conjunto de Contextos Processados cadastrados.
Contexto-Processado/id	POST, GET, PUT e DELETE	Operações sobre cadastros de Contextos Processados, incluindo leitura do valor atual.
Contexto-Externo	GET	Lê o conjunto de Contextos Externos cadastrados.
Contexto-Externo/id	POST, GET, PUT e DELETE	Operações sobre cadastros de Contextos Externos, incluindo leitura do valor atual.
Componente-Processamento	GET	Lê o conjunto de Componentes de Processamento cadastrados.
Componente-Processamento/id	POST, GET, PUT e DELETE	Operações sobre Componentes de Processamento.
Instancia-Processamento	GET	Lê o conjunto de instâncias de Componentes de Processamento criadas para a aplicação.
Instancia-Processamento/id	POST, GET, PUT e DELETE	Operações sobre uma dada instância de um componente de processamento.

3. EXEHDA-CS: Avaliação da Arquitetura

Para realizar a avaliação das funcionalidades da arquitetura foi prototipada uma aplicação para monitoramento de pacientes em reabilitação cardíaca após um acidente vascular. Estudos na área de cardiologia indicam que exercícios físicos podem reduzir o tempo de recuperação destes pacientes. Porém, respostas desproporcionais no batimento cardíaco podem indicar situações de risco para o paciente. Desta forma, o reconhecimento da atividade física e sua correlação com a frequência cardíaca pode permitir uma recuperação mais segura [Negrão and Barreto 2010]. O risco inerente à situação do paciente foi classificado no domínio linguístico: baixo, moderado e alto.

A aplicação proposta explora a capacidade de colaboração entre middleware e aplicações para a composição de situações. O middleware executa a identificação da atividade física do paciente e coleta a informação de frequência cardíaca do mesmo. A aplicação fornece ao middleware os parâmetros de normalidade do paciente, utilizados na etapa de conversão dos dados de contexto para o domínio fuzzy. Na Figura 5 pode ser visto um exemplo de funções de pertinência utilizadas para a conversão da frequência de batimento cardíaco para o domínio fuzzy.

A lógica fuzzy é adequada para a criação de regras para a classificação de situações, tendo em vista os seus mecanismos para tratamento de dados imprecisos e construção de algoritmos através de modelos interpretáveis, facilitando a criação de regras por especialistas da área de aplicação [Sobrevilla and Montseny 2003].

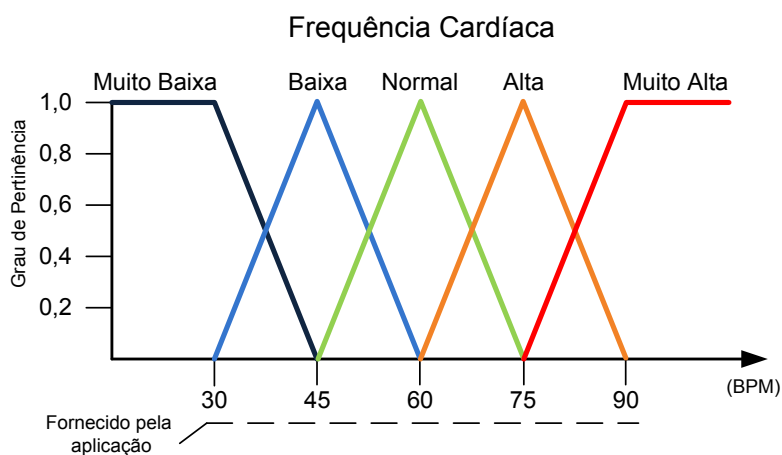


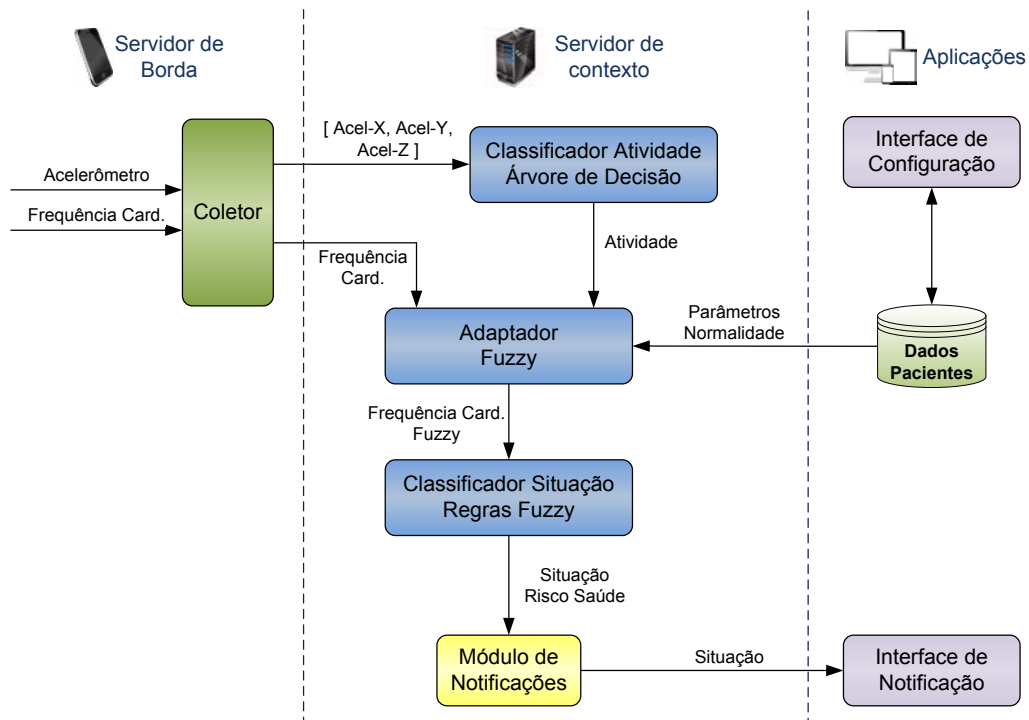
Figura 5. Representação da frequência cardíaca através de lógica fuzzy, com padrão de normalidade fornecido pela aplicação.

O processamento de contexto para identificação de situações, mostrado na Figura 6, baseou-se no uso da técnica de aprendizado Árvore de Decisão para identificação da atividade física realizada pelo paciente. As regras para inferência do risco de saúde foram criadas utilizando lógica fuzzy, considerando a atividade física e a frequência cardíaca do paciente. Os recursos alocados através da API REST são apresentados na Tabela 3.

O treinamento e o teste do componente para classificação de atividades foi feito utilizando a base de dados disponibilizada no trabalho de Kwapisz et al. [Kwapisz et al. 2011], no qual foram capturados sinais de acelerômetros de *smartphones* de 29 voluntários durante a execução das seguintes atividades físicas: caminhando,

Tabela 2. Matriz de confusão para classificação de atividades utilizando Árvore de Decisão.

	Repouso	Movimento lento	Movimento rápido
Repouso	97,8%	0,4%	0,1%
Movimento lento	2,2%	97,8%	3,2%
Movimento rápido	0,0%	1,8%	96,7%
Total correto	97,5%		

**Figura 6. Visão geral do estudo de caso proposto.**

correndo, subindo escada, descendo escada, sentado e de pé. Com a finalidade de adequar as atividades para a aplicação proposta as mesmas foram agrupadas como:

- **Repouso:** sentado e de pé;
- **Movimento lento:** caminhando, descendo escada e subindo escada;
- **Movimento rápido:** correndo.

Após o treinamento foi gerada uma Árvore de Decisão com 62 ramos e 123 elementos, sendo utilizadas 43 características extraídas do sinal do acelerômetro para composição das decisões. Dentre as características utilizadas, pode citar-se: frequência fundamental, módulo da aceleração resultante, aceleração média e desvio padrão em cada um dos três eixos do acelerômetro. O sinal do acelerômetro foi obtido com uma taxa de 20 amostras por segundo e o processamento das características foi executado em janelas temporais de 10 segundos. O resultado da classificação de atividades para a Árvore de Decisão é apresentado na Tabela 2. O percentual total de classificação correta obtido foi de 97,5%.

A frequência cardíaca apropriada para cada atividade física é obtida através de

Tabela 3. Recursos alocados para o estudo de caso.

URI	Descrição
Contexto-Processado/Atividade	Atividade inferida no classificador baseado em Árvore de Decisão.
Contexto-Processado/Freq-Cardiaca-Fuzzy	Frequência cardíaca em domínio fuzzy.
Contexto-Processado/Situacao	Saída que indica o risco de saúde atual do paciente.
Contexto-Externo/Parametros-Normalidade	Informações de normalidade do paciente fornecidas pela aplicação.
Instancia-Processamento/Adaptador-Fuzzy	Instância de componente de processamento para conversão da frequência cardíaca para domínio fuzzy.
Instancia-Processamento/Classificador-Atividade	Instância de componente de processamento para classificação de atividade utilizando Árvore de Decisão.
Instancia-Processamento/Classificador-Situacao	Classificador de situação utilizando regras em lógica fuzzy.

uma interface de **Contexto Externo** para a aplicação em que é feito o gerenciamento dos pacientes. O domínio linguístico para frequência cardíaca é dado por: muito baixa, baixa, normal, alta e muito alta. No bloco **Adaptador Fuzzy** os valores numéricos obtidos do sensor de batimento cardíaco são transformados para o domínio fuzzy aplicando funções de pertinência triangulares, parametrizadas com os padrões de normalidade do paciente.

Na instância de processamento em que é feita a classificação da situação de risco de saúde do paciente é aplicado um conjunto de regras utilizando lógica fuzzy. Estas regras podem ser vistas na Figura 7.

```

GET freq_cardiaca
IF freq_cardiaca IS muito_alta OR freq_cardiaca IS muito_baixa THEN
    risco IS alto
IF freq_cardiaca IS alta OR freq_cardiaca IS baixa THEN
    risco IS moderado
IF freq_cardiaca IS normal THEN
    risco IS baixo

```

Figura 7. Algoritmo para inferência da situação de risco de saúde do paciente.

4. Trabalhos Relacionados

O estudo de trabalhos relacionados foi feito considerando as especificações do EXEHDA-CS: arquitetura utilizada em diferentes domínios de aplicação; arquitetura distribuída; abordagem colaborativa de processamento de contexto e suporte para identificação de situações.

Yuan e Herbert [Yuan and Herbert 2014] propõem uma arquitetura para suporte ao monitoramento e reconhecimento de atividades e hábitos para aplicações na área de vivência assistida. Neste trabalho é previsto o uso de um algoritmo de aprendizado baseado em casos e as regras da aplicação são criadas utilizando lógica fuzzy. No EXEHDA-CS a abordagem de processamento foi estruturada de forma a suportar a inclusão de novos algoritmos através de Componentes de Processamento. Desta forma, diversas técnicas de aprendizado e especificação podem ser combinadas.

Haghighi et al. [Haghighi et al. 2014] apresentam uma arquitetura para uma aplicação específica na área de monitoramento de pacientes hipertensos. Nesta aplicação não é provido suporte ao processamento distribuído. Já o EXEHDA-CS foi especificado para ser independente do domínio de aplicação e sua arquitetura foi construída para poder operar de forma distribuída.

Cimino et al. [Cimino et al. 2012] propõem uma arquitetura de software para ciência de situação. Os autores propõem um modelo de processamento que considera o uso de regras e de modelos semânticos para representação e processamento do contexto. No EXEHDA-CS considerou-se que o middleware deve prover suporte para diversos tipos de representação de contexto e modelos de processamento. Desta forma, através da API proposta para o Servidor de Contexto, a técnica de processamento e o padrão de representação do contexto são definidos através dos Componentes de Processamento.

Os trabalhos avaliados enfatizam a transparência dos mecanismos para processamento e modelagem de contexto. Considera-se, no entanto, que esta abordagem “caixa-preta” dos trabalhos relacionados não é apropriada para o desenvolvimento de aplicações com o grau de flexibilidade e generalidade previstas para o EXEHDA. Desta forma, a proposta para minimizar o conflito entre transparência e a capacidade de identificação de situações é a colaboração entre a aplicação e middleware no processamento contextual. Com isso, busca-se estabelecer uma abordagem flexível para o processamento e modelagem do contexto.

5. Considerações Finais

A principal contribuição do EXEHDA-CS é uma arquitetura para prover ciência de situação, através de uma abordagem que permite a colaboração entre o middleware e as aplicações. Na API proposta foram considerados os aspectos de flexibilidade na criação de mecanismos para o processamento das informações de contexto visando à identificação de situações.

Através do estudo de caso foi possível demonstrar que a abordagem proposta permitiu a colaboração entre o middleware e a aplicação para identificação de situações, otimizando a construção dos modelos de raciocínio fuzzy (vide Figura 7).

Como trabalho futuro pretende-se expandir a API proposta neste trabalho para dar suporte aos mecanismos de adaptação previstos no EXEHDA.

Referências Bibliográficas

Bellavista, P., Corradi, A., Fanelli, M., and Foschini, L. (2012). A survey of context data distribution for mobile ubiquitous systems. *ACM Comput. Surv.*, 44(4):24:1–24:45.

- Cimino, M. G. C. A., Lazzerini, B., Marcelloni, F., and Ciaramella, A. (2012). An adaptive rule-based approach for managing situation-awareness. *Expert Syst. Appl.*, 39(12):10796–10811.
- Fielding, R. T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral dissertation, University of California.
- Haghighi, D. P., Averì, P., Indrawan-Santiago, M., and Minh, H. T. (2014). Situation-aware mobile health monitoring. *ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)*, pages 248–256.
- Knappmeyer, M., Kiani, S., Reetz, E., Baker, N., and Tonjes, R. (2013). Survey of context provisioning middleware. *Communications Surveys Tutorials, IEEE*, 15(3):1492–1519.
- Kwapisz, J. R., Weiss, G. M., and Moore, S. a. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12:74.
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014a). A middleware architecture for dynamic adaptation in ubiquitous computing. *Journal of Universal Computer Science*, 20(9):1327–1351.
- Lopes, J., Souza, R., Pernas, A., Yamin, A., and Geyer, C. (2014b). A distributed architecture for supporting context-aware applications in ubicomp. In *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*, pages 584–590.
- Negrão, C. E. and Barreto, A. C. P. (2010). *Cardiologia do Exercício: do Atleta ao Cardiopata*. Manole, Barueri, SP - Brazil, 3 edition.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2013). Context Aware Computing for The Internet of Things : A Survey. *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, X(X):1–41.
- Rabelo, D., Gil, C., and Araújo, S. D. (2006). Reabilitação cardíaca com ênfase no exercício: uma revisão sistemática. *Revista Brasileira de Medicina do Esporte*, 12(5):279–285.
- Razzaque, M. A., Milojevic-Jevric, M., Palade, A., and Clarke, S. (2015). Middleware for Internet of Things: a Survey. *Internet of Things Journal, IEEE*, PP(99):1.
- Sobrevilla, P. and Montseny, E. (2003). Fuzzy Sets in Computer Vision : an Overview. *Mathw. Soft Computing*, 10:71–83.
- Yuan, B. and Herbert, J. (2014). Context-aware hybrid reasoning framework for pervasive healthcare. *Personal and Ubiquitous Computing*, 18(4):865–881.

Ciência de Contexto na IoT: Uma Contribuição à Pesquisa Agropecuária

Patrícia Davet¹, Huberto Kaiser Filho¹, Leonardo João¹, Lucas Xavier¹, Rodrigo Souza², João Lopes², Ana Marilza Pernas¹, Nelsi Warken³, Adenauer Yamin¹

¹Universidade Federal de Pelotas (UFPEL) – Pelotas, RS – Brazil

²Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre, RS – Brazil

³Embrapa Clima Temperado – Pelotas, RS – Brazil

{ptdavit, hkaiser, ldrsjoao, lmdsxavier, marilza, adenauer}@inf.ufpel.edu.br,
{rssouza, jlblopes}@inf.ufrgs.br, nelsi.warken@cpact.embrapa.br

Abstract. *The Context-Awareness has been gaining attention in the scenario of the IoT due to increasing need for autonomous behavior both from the devices such as of the applications. This paper presents the PlenUS project and software architecture for Context-Awareness giving you supports it, called EXEHDA-IoT. This architecture was designed to promote the acquisition, the storage and the processing of the context information, so distributed, independent of the applications in an autonomic approach based on rules. To evaluate the functionalities of the EXEHDA-IoT, a case study at Embrapa Temperate Climate is presented, having been obtained promising results.*

Resumo. *A Ciência de Contexto vem ganhando destaque no cenário da IoT devido a crescente necessidade de comportamento autônomo por parte tanto dos dispositivos como das aplicações. Este artigo apresenta o projeto plenUS e a arquitetura de software para Ciência de Contexto que lhe dá suporte, denominada EXEHDA-IoT. Esta arquitetura foi concebida para prover a aquisição, o armazenamento e o processamento das informações de contexto, de forma distribuída, independente das aplicações, em uma perspectiva autônoma baseada em regras. Para avaliar as funcionalidades do EXEHDA-IoT, um estudo de caso junto a Embrapa Clima Temperado é apresentado, tendo sido obtidos resultados promissores.*

1. Introdução

O projeto plenUS (*plentiful Ubiquitous Systems*) tem como premissa o emprego do middleware EXEHDA com o objetivo de disponibilizar Sistemas Ubíquos, que explorem relações pró-ativas entre usuários, softwares e equipamentos, auxiliando nas pesquisas em desenvolvimento na Embrapa Clima Temperado. O mesmo vem sendo desenvolvido por um consórcio de pesquisa envolvendo o LUPS (*Laboratory of Ubiquitous and Parallel Systems*) da UFPEL e a Embrapa Clima Temperado.

De modo mais específico o projeto plenUS tem como foco atender as demandas referentes ao provimento de serviços computacionais cientes de contexto, permitindo um registro histórico dos estados contextuais em que se encontram os equipamentos dos seguintes laboratórios da Embrapa: (i) Laboratório de Qualidade do Leite (Lableite), (ii)

Laboratório de Reprodução Animal e (iii) Laboratório de Análise de Sementes Oficial (LASO), como também outras necessidades de sensoriamento em ambientes administrativos como o do Núcleo de Tecnologia da Informação (NTI), e de uma Sala de Testes do plenUS.

Com a implementação do sistema de qualidade nestes laboratórios aumentaram as exigências no controle dos equipamentos onde são realizadas as análises, visando precisão e confiabilidade nos resultados. Deste modo, tornam-se necessários registros das leituras de grandezas físicas como temperatura, umidade e luminosidade dos equipamentos envolvidos (germinadores, câmaras frias, salas de trabalho, etc.), bem como uma atuação automatizada em função das informações sensorizadas, sempre que for o caso.

A infraestrutura computacional proporcionada pela Internet das Coisas (*Internet of Things* - IoT) permite que estes equipamentos possam trocar informações entre si e/ou com usuários via Internet, permitindo o desenvolvimento de aplicações com maior valor agregado. Isto é possível devido à incorporação de dispositivos computacionais a estes equipamentos, unicamente identificados, e com capacidade de conexão à Internet, sensoriamento, armazenamento e processamento [Pires et al. 2015].

Por sua vez as aplicações IoT, enquanto ubíquas e assim dotadas de comportamento autônomo, necessitam ter ciência das informações contextuais que lhe interessam. Para o desenvolvimento de aplicações cientes de contexto no cenário da IoT, há uma série de funcionalidades que devem ser providas, envolvendo desde a aquisição de informações contextuais adquiridas por sensores, a partir de um grande número de fontes heterogêneas e distribuídas, até a representação dessas informações, seu processamento, armazenamento, e realização de inferências para seu uso em procedimentos de atuação [Perera et al. 2013].

Neste sentido, este trabalho tem como objetivo a concepção de uma arquitetura de software comprometida com a premissa de fornecer mecanismos para o desenvolvimento de aplicações cientes de contexto na IoT, denominada EXEHDA-IoT. Esta arquitetura é baseada no CoIoT (Context+IoT) e sua avaliação aconteceu junto ao projeto plenUS. O CoIoT especializa o middleware EXEHDA (*Execution Environment for Highly Distributed Applications*) no tratamento das questões de Ciência de Contexto na IoT [Souza et al. 2015].

O artigo está organizado da seguinte forma: a Seção 2 descreve a modelagem e principais funcionalidades da arquitetura de software proposta. A Seção 3 apresenta a avaliação da arquitetura pelo projeto plenUS. A Seção 4 discute os trabalhos relacionados, e, por fim, a Seção 5 apresenta as considerações finais.

2. EXEHDA-IoT: Arquitetura e Funcionalidades

O EXEHDA é um middleware baseado em serviços que visa criar e gerenciar um ambiente ubíquo, bem como promover a execução de aplicações sobre esse ambiente. No EXEHDA, as condições de contexto são pró-ativamente monitoradas e o suporte à execução deve permitir que tanto a aplicação como ele próprio utilizem essas informações na gerência da adaptação de seus aspectos funcionais e não-funcionais. O mecanismo de adaptação proposto para o EXEHDA emprega uma estratégia colaborativa entre aplicação e ambiente de execução, através da qual é facultado ao programador individualizar

políticas de adaptação para reger o comportamento de cada um dos componentes que constituem o software da aplicação [Lopes et al. 2014].

O fato do EXEHDA ter uma origem enquanto middleware para UbiComp, faz com que o tratamento da Ciência de Contexto seja inerente à sua arquitetura. Para que o EXEHDA continue provendo Ciência de Contexto tendo em vista os desafios do cenário da IoT, onde a aquisição das informações de contexto ocorre a partir de um grande número de dispositivos heterogêneos, distribuídos dinamicamente e com capacidade restrita, está em desenvolvimento o CoIoT [Souza et al. 2015].

Considerando isso, o EXEHDA-IoT foi modelado empregando os seguintes pressupostos do CoIoT:

- *automatização*: implica que um middleware para IoT deve suportar a coleta de informações, processamento e inferência contextual de forma autônoma;
- *inteligência*: dispositivos IoT devem estar capacitados a operar, com algum grau de adaptação as variações contextuais (variações nas infraestruturas de rede, aplicações e usuários, por exemplo);
- *dinamicidade*: um dispositivo pode se mover de um lugar a outro, necessitando que o middleware esteja apto para reconhecer esta mudança e conseqüentemente realizar os ajustes necessários;
- *zero configuração*: para simplificar os procedimentos de integração e/ou remoção de dispositivos, recursos *plug and play* devem estar disponíveis, reduzindo os esforços de gerência por parte dos usuários e facilitando o crescimento descentralizado que é inerente nas infraestruturas da IoT.

O ambiente ubíquo disponibilizado pelo EXEHDA-IoT é organizado em células de execução, nas quais os dispositivos computacionais são distribuídos (vide Figura 1). Cada célula é constituída dos seguintes componentes: (i) EXEHDAbase, o elemento central da célula, sendo responsável por todos serviços básicos e constituindo referência para os demais elementos; (ii) o EXEHDA nodo que corresponde aos dispositivos computacionais responsáveis pela execução das aplicações; (iii) o EXEHDA nodo móvel, um subcaso do anterior, que corresponde aos dispositivos tipicamente móveis que podem se deslocar entre as células do ambiente ubíquo, como *notebooks*, *tablets* ou *smartphones*, (iv) o EXEHDAborda, responsável por fazer a interoperação entre os serviços do middleware e os diversos tipos de *gateways*; e (v) o EXEHDAgateway, que consiste no elemento responsável por setorizar pontos de coleta e/ou atuação distribuídos, disponíveis no meio físico, realizando a interação destes com os outros componentes do middleware.

Para provimento de Ciência de Contexto na IoT, o EXEHDA-IoT se vale de dois tipos principais de servidores: Servidor de Borda e Servidor de Contexto (vide Figura 1). O Servidor de Borda é instanciado em um equipamento do tipo EXEHDAborda e se destina a gerenciar a interação com o meio físico através de *gateways*, bem como a execução de regras de contingência e armazenamento temporário das informações contextuais coletadas, em caso de falha de comunicação. O Servidor de Contexto, por sua vez, é alocado no EXEHDAbase e atua no armazenamento e no processamento das informações contextuais, integrando informações históricos e aquelas provenientes de diferentes Servidores de Borda distribuídos no ambiente ubíquo.

Na proposta do EXEHDA-IoT a premissa é que os sensores e/ou atuadores sejam incorporados ao Servidor de Borda através de *gateways*. Os *gateways* são utilizados

Considerando a dinamicidade e escalabilidade do cenário da IoT, o EXEHDA-IoT possibilita a identificação e descoberta automática dos dispositivos de sensoriamento e/ou atuação, a partir dos *gateways* integrados a arquitetura, facilitando a gerência dos dispositivos, quando da sua inclusão e/ou remoção. Para tal, foi utilizada a estratégia de comunicação *plug and play* do protocolo UPnP (*Universal Plug and Play*) [UPnP 2015].

A arquitetura de software do EXEHDA-IoT é baseada no CoIoT (vide Figura 2) e provê comunicação: (i) entre os *gateways* e o Servidor de Borda; (ii) entre os Servidores de Borda e o Servidor de Contexto; (iii) entre os Servidores de Contexto localizados em diferentes células do ambiente ubíquo; e (iv) com outros serviços do *middleware* ou aplicações. As linhas tracejadas na representação dos *gateways* indicam que algum dos tipos de *gateways* podem não estar presentes na arquitetura, a qual possibilita utilizar somente *gateways* físicos (Nativos ou Proprietários), somente Gateway Virtual, ou ambos.

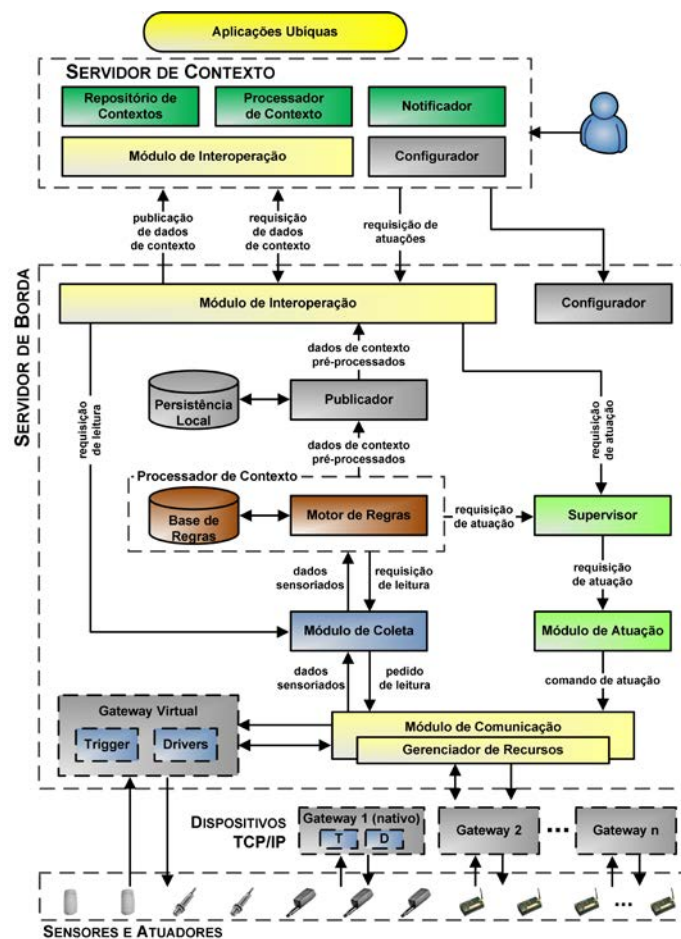


Figura 2. Arquitetura EXEHDA-IoT

2.1. Servidor de Contexto

O Servidor de Contexto no EXEHDA-IoT tem suas funcionalidades organizadas através dos seguintes módulos arquiteturais.

Repositório de Contextos: armazena dados de configuração da arquitetura, e os valores sensoriados publicados pelos Servidores de Borda existentes no ambiente ubíquo.

Processador de Contexto: realiza tarefas de manipulação e dedução das informações contextuais. Baseia-se em regras do tipo ECA (Evento-Condição-Ação), de forma a tratar eventos gerados a partir de mudanças nos estados dos contextos de interesse, que podem acarretar no disparo de ações pertinentes em função do estado contextual.

Notificador: tem a função de gerar notificações necessárias para a operação do EXEHDA-IoT. As notificações são consequências do processamento das informações contextuais coletadas. Este módulo utiliza uma estratégia de notificação baseada no modelo *publish/subscribe*, em que recebe subscrições de todos os serviços e/ou aplicações que requerem notificações a respeito das mudanças nos estados contextuais. Também é função do Notificador enviar comandos de atuação aos Servidores de Borda quando necessário, bem como a realização de notificação pela aplicação Web de novo gateway descoberto, o qual só será utilizado mediante autenticação pelo usuário interessado.

Módulo de Interoperação: considerando o perfil inerentemente distribuído das aplicações alvo do EXEHDA-IoT, o Módulo de Interoperação foi concebido para gerenciar a troca de informações com os Servidores de Borda, bem como os demais serviços do middleware em execução em outros equipamentos, como também as aplicações, oferecendo para isso uma interface padronizada e bem definida. A concepção deste módulo teve como referência o estilo arquitetural REST, sendo utilizado o protocolo HTTP e suas operações (GET, PUT, POST, DELETE) no suporte às trocas de informações.

Configurador: aglutina todos os procedimentos de configuração necessários para o funcionamento do EXEHDA-IoT. Através deste módulo é disponibilizada uma aplicação através da qual é possível gerenciar dispositivos *gateways* e regras de processamento de contexto, bem como realizar todo tipo de cadastro necessário ao funcionamento da aplicação IoT ciente de contexto.

2.2. Servidor de Borda

No EXEHDA-IoT o Servidor de Borda tem suas funcionalidades sistematizadas através dos seguintes módulos arquiteturais.

Módulo de Interoperação: é destinado a gerenciar a comunicação com o Servidor de Contexto, tendo também como referência o estilo arquitetural REST. Com isso, todas as funcionalidades associadas aos *gateways*, sensores e atuadores são mapeadas na forma de recursos e acessadas via URIs.

Configurador: no Servidor de Borda o módulo Configurador é responsável pelos ajustes necessários para o seu funcionamento. Entre as funcionalidades oferecidas tem-se: definição do Servidor de Contexto para a publicação das informações sensorizadas, bem como a inserção, atualização, ativação e desativação dos *gateways* e seus serviços por meio do protocolo de comunicação UPnP.

Publicador: é responsável por coordenar o principal fluxo de informações entre os Servidores de Borda e o Servidor de Contexto, promovendo a publicação de todas as informações contextuais coletadas e garantindo uma **Persistência Local** das mesmas nos períodos que a publicação ficar inviabilizada. As transações ocorrem sobre o protocolo HTTPS (*HyperText Transfer Protocol Secure*), o que garante que as informações serão transferidas criptografadas, evitando acessos indevidos aos mesmos.

Processador de Contexto: é responsável por tratar regras de contingência, com a

intenção de evitar que os dispositivos envolvidos atinjam estados críticos, o que faz com que este módulo adquira elevada importância quando a comunicação com o Servidor de Contexto for interrompida por problemas na infraestrutura de rede.

Supervisor: possui como objetivo aglutinar os comandos de atuação. Uma vez recebidos os parâmetros para controle da atuação, o módulo Supervisor avalia eventuais conflitos entre as requisições de atuação oriundas de diferentes fontes, para após enviar o comando de atuação ao Módulo de Atuação.

Módulo de Coleta: tem a função de direcionar os pedidos de coleta ao Módulo de Comunicação que envia aos respectivos *gateways*, sob demanda tanto do Motor de Regras, como do Servidor de Contexto ou das aplicações a qualquer momento. O módulo faz a recepção assíncrona das solicitações e repassa o ID do sensor a ser lido pelo gateway responsável. Este módulo também tem a função de direcionar aos demais elementos da arquitetura as informações sensorizadas enviadas pelos *gateways*, produzidas a partir de eventos do ambiente.

Módulo de Atuação: possui um funcionamento análogo ao Módulo de Coleta. Ele recebe do Supervisor comandos com o ID do atuador e os correspondentes padrões de operação (tempo de duração, potência de ativação, etc.), e os envia ao Módulo de Comunicação a fim de se serem interpretados e direcionados ao gateway responsável pelo efetivo tratamento.

Gateway Virtual: realiza o interfaceamento entre sensores e/ou atuadores (dispositivos sem conexão Internet) ligados diretamente ao Servidor de Borda. Possui dois módulos básicos: **Drivers** e **Trigger**. Os Drivers são responsáveis pelo encapsulamento individualizado de aspectos tecnológicos, tanto de sensores como atuadores, o que evita que as suas diferenças operacionais se projetem nos demais módulos da arquitetura. O Trigger é um módulo que tem a finalidade de gerenciar a leitura dos sensores através de eventos. Consiste em um tipo simplificado de regra de produção, que dá suporte a um conjunto limitado de funções para o tratamento de condições e ações.

Módulo de Comunicação: tem a função de prover aos demais módulos da arquitetura um acesso padronizado aos recursos disponibilizados pelos *gateways* no que diz respeito ao tratamento de sensores e/ou atuadores. Gerencia as comunicações com os Gateways Nativos e Gateways Virtuais através do protocolo UPnP, que se utiliza de princípios REST como forma de prover interoperabilidade entre dispositivos cliente e servidor. Por outro lado, o acesso aos serviços dos Gateways Proprietários é realizado através das APIs disponibilizados pelos seus fabricantes, que em muitos casos baseiam-se em protocolos típicos da IoT, como CoAP¹ e MQTT².

Gerenciador de Recursos: foi concebido com o objetivo de administrar procedimentos, como a entrada e saída de operação de dispositivos, bem como à reposição por avaria ou necessidade de adicionar outro ponto de monitoramento. É neste módulo que é instanciado o Ponto de Controle UPnP, que possui como tarefa descobrir os dispositivos UPnP (Gateways Nativo e Virtual) presentes na rede, armazenando as informações sobre os dispositivos e seus respectivos serviços descobertos (sensores e/ou atuadores) em um diretório, para eventual controle destes.

¹<http://coap.technology/>

²<http://mqtt.org/>

3. EXEHDA-IoT: Avaliação no Domínio Agropecuário

As funcionalidades propostas para o EXEHDA-IoT, foram avaliadas através de estudo de caso no domínio agropecuário relacionado ao projeto plenUS. Os ambientes priorizados pelo projeto, para recebimento de infraestrutura para ciência do contexto em que se encontram seus equipamentos, bem como os tipos de grandezas mensuradas e a disposição dos equipamentos utilizados podem ser visualizados na Figura 3.

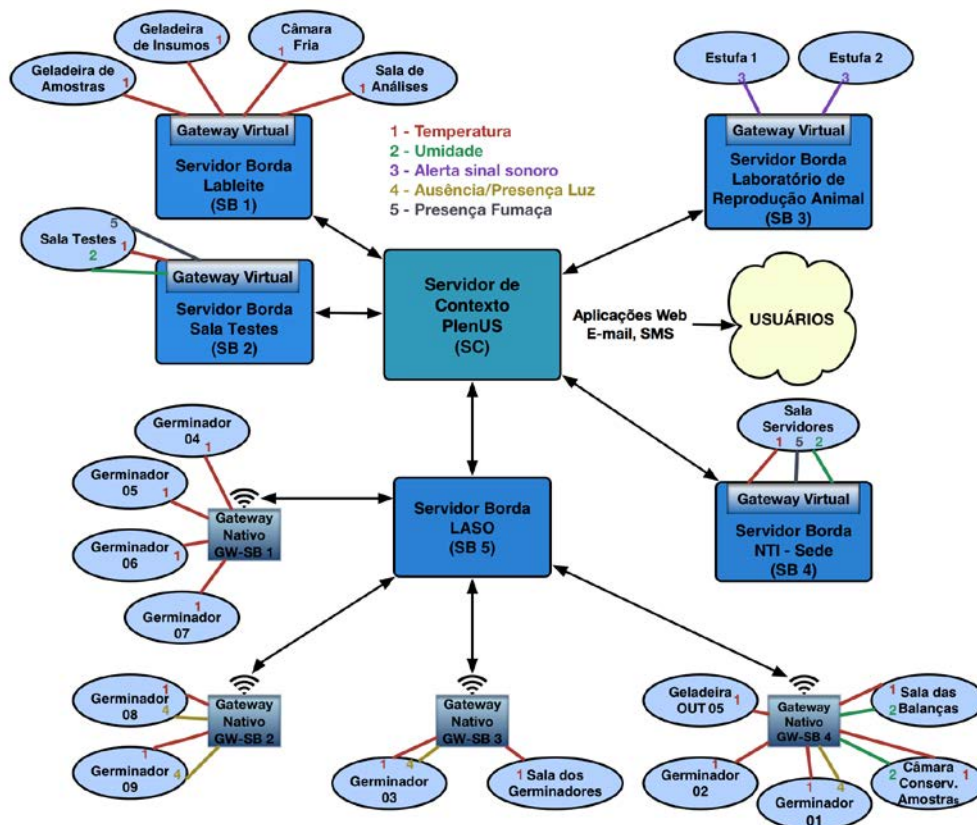


Figura 3. Estudo de Caso - Projeto plenUS

Considerando a estrutura física do estudo de caso, um Servidor de Contexto, cinco Servidores de Borda e quatro Gateways Nativos foram instalados para atender as demandas da Embrapa como um todo (vide Figura 3), constituindo uma célula de execução do ambiente ubíquo.

O Servidor de Contexto e o Servidor de Borda 4 (SB 4) foram posicionados no NTI, localizado na unidade Sede da Embrapa Clima Temperado, enquanto os outros Servidores de Borda foram posicionados na unidade Estação Experimental Terras Baixas (ETB), distante 21Km.

O hardware utilizado para o Servidor de Contexto foi um desktop com processador Intel Pentium E2160-1.8GHz de dois núcleos, com 3Gb de memória RAM e com o Sistema Operacional Ubuntu Server. Já os Servidores de Borda foram prototipados sobre o Sistema Operacional Raspbian, sendo usado como hardware a Raspberry PI Modelo B+³ (vide Figura 4a).

³<http://www.raspberrypi.org>

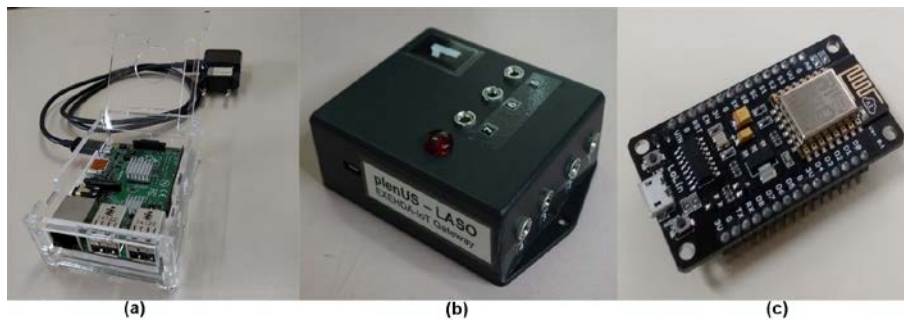


Figura 4. Hardwares: (a) Raspberry PI; (b) Gateway Nativo; (c) ESP8266-12

Enquanto aos *gateways* previstos pelo EXEHDA-IoT e utilizados para o estudo de caso, estes são de duas naturezas: **Gateways Virtuais**, internos ao Servidor de Borda (SB), e **Gateways Nativos** com conexão wi-fi com o SB. Os Gateways Nativos foram utilizados no caso onde as demandas de sensoriamento introduziram a sua necessidade, seja pela disposição física dos equipamentos e/ou salas ou pela quantidade significativa de sensores.

Para a escolha do hardware utilizado na concepção do Gateway Nativo (vide Figura 4b) foram considerados aspectos como: baixo consumo energético, possibilidade de operação em regime sem fio (wi-fi) e emprego de plataforma de software disseminada. Após revisão de literatura a escolha recaiu sobre o processador ESP8266-12 [Kurniawan 2015] (vide Figura 4c).

A operacionalização do cenário propiciado pelo estudo de caso é realizada através de regras distribuídas do tipo ECA (Evento-Condição-Ação) entre os Servidores de Borda e Contexto. Uma contextualização da forma de processamento das regras tipo ECA no cenário de avaliação é apresentado na Figura 5 por meio de uma notação BPMN [Group 2015].

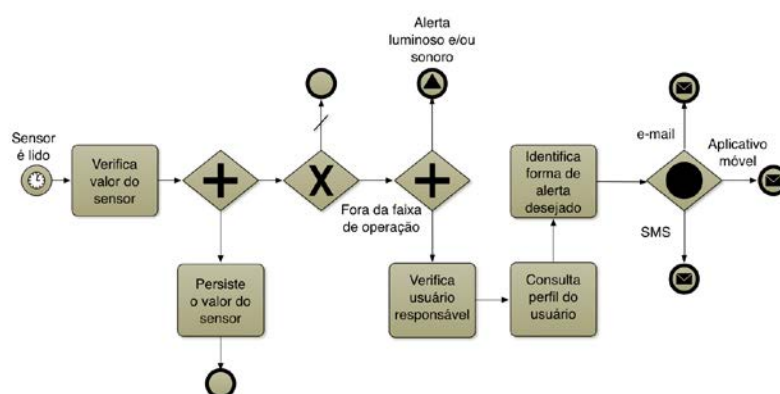


Figura 5. Processamento de Regra Tipo ECA no Cenário de Avaliação (Notação BPMN)

O cenário consiste em monitorar uma determinada grandeza e produzir registro histórico. Enquanto a grandeza é coletada e armazenada, a mesma tem seu valor avaliado tendo em vista as faixas de operação para cada dado de contexto coletado. Alertas são enviados aos usuários envolvidos, quando os valores estiverem fora da faixa de operação

especificada para o experimento.

A interação do usuário com o cenário é disponibilizada por meio de uma aplicação desenvolvida para navegadores web, que contempla tarefas de gerenciamento dos dispositivos de sensoriamento e/ou atuação, com suporte a descoberta automática, o que agregou funcionalidades de cadastro e/ou remoção automática dos dispositivos *gateways*, bem como de seus sensores quando de sua disponibilidade. Além disso, a aplicação também permite a visualização do registro histórico das informações contextuais, tanto de forma gráfica como textual.

O relatório gráfico (vide Figura 6) oferecido pelo sistema permite a visualização simultânea das informações de vários sensores. A seleção dos sensores é feita a partir de um menu com suporte a múltipla seleção. Também é disponibilizado um recurso de inspeção que permite a comparação dos valores em um determinado instante do tempo.



Figura 6. Projeto plenUS - Relatório gráfico

Para atender o fato da rotina de trabalho dos laboratoristas implicar em uma mobilidade nos diversos recintos do laboratório, a aplicação Web desenvolvida possui características responsivas, o que veio agregar maior funcionalidade ao seu uso, pois possibilitou que as configurações e testes dos dispositivos de sensoriamento e/ou atuação pudessem ser realizadas no local de instalação por meio de *smartphones*, devido ao seu *layout* se adaptar ao dispositivo computacional utilizado.

4. Trabalhos Relacionados

Uma análise dos trabalhos relacionados EcoDif [Pires et al. 2014], S³OIA [Vega-Barbas et al. 2012], LinkSmart [LinkSmart 2015] e Carriots [Carriots 2015] em relação à satisfação dos requisitos IoT considerados neste trabalho é apresentada na Tabela 1, onde o símbolo “X” denota que o requisito é completamente atendido, o símbolo “O” indica que o requisito é parcialmente atendido, e o símbolo “-” indica que o requisito não é atendido.

Tabela 1. Análise dos Trabalhos Relacionados

Plataformas de middleware	Interoperabilidade	Descoberta e Gerenciamento de Dispositivos	Interfaces de Alto Nível	Ciência de Contexto	Escalabilidade
EcoDiF	X	O	X	O	X
S³OIA	X	X	-	X	-
LinkSmart	X	O	-	X	-
Carriots	X	O	X	O	X
EXEHDA-IoT	X	X	X	X	X

Registre-se que o único requisito endereçado de uma maneira ou outra por todos os trabalhos relacionados, foi o de interoperabilidade. Isto é devido ao fato deste ser um requisito primordial e que deve ser imperativamente endereçado por qualquer plataforma de middleware para IoT.

Dentre os trabalhos analisados, o S³OIA destaca-se por ser o único a apresentar uma estratégia de descoberta de dispositivos automática, *plug and play* e de autoconfiguração para o gerenciamento de sensores e/ou atuadores. Os outros atendem ao critério de descoberta e gerenciamento de dispositivos de forma parcial, pressupondo que os dispositivos e seus recursos requeridos estejam disponíveis por meio de um cadastro prévio. No EXEHDA-IoT, por outro lado, se utiliza o UPnP como estratégia de autoconfiguração e descoberta, tendo o gateway como elemento básico do mecanismo de descoberta e a partir dele, os sensores e atuadores são facilmente descobertos e gerenciados. Além disso, a abordagem utilizada considera todo o potencial de protocolos de uso consolidado.

A ciência de contexto é contemplada por todos os trabalhos relacionados, porém o suporte oferecido, como no caso do EcoDiF e Carriots mostra-se limitado. Já o EXEHDA-IoT fornece uma organização distribuída para coleta e processamento de contexto através de regras, como também para atuação em ambientes com características heterogêneas e escaláveis. Além disso, por ter sido desenvolvido a partir de um middleware para Ubi-Comp, o EXEHDA, também provê suporte a outros serviços ubíquos disponibilizados pelo middleware, como adaptação, comunicação desacoplada, migração de código, etc.

Por fim, outra abordagem adotada pelo EXEHDA-IoT, que trouxe benefícios principalmente no que tange à otimização do tráfego de rede e ao consumo de energia, é a proposição do uso de *triggers* para gerenciar o fluxo de informações transmitidas. A abordagem de *triggers* foi concebida para permitir a personalização da coleta das informações através de eventos considerando as características de variabilidade física de cada grandeza monitorada, o que proporciona minimizar o fluxo das informações entre os *gateways* e o Servidor de Borda.

Além disso, a abordagem distribuída do processamento do contexto entre os Servidores de Borda e Contexto proposta pelo EXEHDA-IoT também contribui com a minimização do fluxo de informações contextuais em relação as abordagens centralizadas, empregadas pelos trabalhos relacionados.

5. Considerações Finais

O principal diferencial do EXEHDA-IoT em relação aos trabalhos relacionados diz respeito a possibilidade de gerenciar a aquisição, armazenamento e processamento das

informações de contexto, de forma distribuída, independente das aplicações, em uma perspectiva autônoma baseada em regras.

Os resultados obtidos pelo estudo de caso, no domínio agropecuário atenderam as demandas dos usuários, além de fornecer um *feedback* para avanços futuros, estimulando a continuidade das pesquisas na área.

Como trabalhos futuros, os seguintes aspectos foram priorizados: (i) explorar estudos de caso em que as regras de processamento contextual utilizem outros mecanismos de inferência de mais alto nível; e (ii) implantação de um framework para gerenciamento do motor de regras, tanto no Servidor de Borda como no Servidor de Contexto.

Referências

- Carriots (2015). Carriots iot application platform. Disponível em :<<https://www.carriots.com/>>. Acesso em julho de 2015.
- Group, O. M. (2015). Business Process Model and Notation. Disponível em :<<http://www.bpmn.org>>. Acesso em maio de 2015.
- Kurniawan, A. (2015). *Sparkfun ESP8266 Thing Development Workshop*. PE PRESS, Canada.
- LinkSmart (2015). Middleware for networked embedded systems. Disponível em :<<https://linksmart.eu/redmine/projects/linksmart-opensource/wiki>>. Acesso em julho de 2015.
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014). A middleware architecture for dynamic adaptation in ubiquitous computing. *Journal of Universal Computer Science*, 20(9):1327–1351.
- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2013). Context aware computing for the internet of things: A survey. *Communications Surveys & Tutorials, IEEE*, 16(1):414–454.
- Pires, P. F., Cavalcante, E., Barros, T., Delicato, F. C., Batista, T., and Costa, B. (2014). A platform for integrating physical devices in the internet of things. *Proceedings of the 12th IEEE International Conference on Embedded and Ubiquitous Computing*, pages 234–241.
- Pires, P. F., Delicato, F. C., Batista, T., Barros, T., Cavalcante, E., and Pitanga, M. (2015). Plataformas para a internet das coisas. *Livro Texto de Minicursos - SBRC 2015*.
- Souza, R. S., Lopes, J. L. B., Davet, P. T., Garcia, C. G., Gadotti, G. I., Yamin, A. C., and Geyer, C. F. R. (2015). Context awareness in ubicomp: an iot oriented distributed architecture. *Proceedings of 2015 IEEE International Conference on Electronics, Circuits, and Systems*.
- UPnP, F. (2015). Fórum upnp. Disponível em :<<http://upnp.org/>>. Acesso em julho de 2015.
- Vega-Barbas, M., Casado-Mansilla, D., Valero, M., López-de Ipina, D., Bravo, J., Flórez, F., et al. (2012). Smart spaces and smart objects interoperability architecture (s3oia). *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, pages 725–730.

Avaliação de Cluster Raspberry Pi para Execução de Aplicações de Análise de Imagens Microscópicas Médicas

Rafael M. Ramos¹, Célia Ralha¹, George Teodoro¹

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Campus Darcy Ribeiro, Edifício CIC/EST – Caixa Postal 4466
Asa Norte – CEP 70910-900 – Brasília – DF – Brasil

rafaelmr@gmail.com, ghedini@unb.br, teodoro@unb.br

Abstract. *The health care area, in particular, the clinical pathology with automated analysis of microscopic images is a field with growing demand for computing power. A barrier to the effective use of computing in this scenario is the unavailability of enough computational resources due to the high related costs. In this paper, we evaluate the use of a low-cost architecture named Raspberry Pi 2 to build computer clusters with 64 cores and 16GB RAM and their use in medical applications in the analysis of microscopic images. Our evaluation's goal the identification of the potential cost benefit of this platform compared to other processors, taking into account runtime, hardware cost and energy consumption. The experimental results showed that the use of a Raspberrys cluster is 10× and 2× faster than the execution of the application, respectively, on a Core2Duo and I7 machine. Even at full capacity, the cluster is more power efficient than other processors only in their idle mode. As such, the Raspberry Pi 2 has proved to be an excellent and promising platform for running our class of target applications.*

Resumo. *A área de saúde e, em especial, a patologia clinica com análise automatizada de imagens microscópicas é um campo com demanda crescente por poder computacional. Uma barreira para a utilização eficaz de computação nesse cenário é a indisponibilidade de recursos computacionais suficientes, devido aos altos custos relacionados. Dessa forma, nesse artigo, avaliamos a utilização de uma arquitetura de baixo custo e gasto energético formada por equipamentos Raspberry Pi 2 para construção de aglomerados de computadores com 64 cores e 16GB RAM para utilização em aplicações médicas em análise de imagens microscópicas. Nossa avaliação tem como objetivo identificar o potencial custo benefício dessa plataforma em comparação com outros processadores, levando em consideração tempo de execução, custo do hardware e gasto energético. Os resultados experimentais mostraram que a utilização de um cluster de Raspberrys é 10× e 2× mais rápido que a execução da aplicação, respectivamente, em máquinas Core2Duo e I7. Mesmo em sua capacidade máxima, o cluster foi energeticamente mais econômico do que os demais processadores. Dessa forma, o Raspberry Pi 2 mostrou-se uma excelente e promissora plataforma para execução de nossa classe de aplicações alvo.*

1. Introdução

A utilização de computação na resolução de problemas da área de saúde é um dos grandes desafios de pesquisa no país e no mundo. Notadamente, o desenvolvimento da área de

saúde vem cada vez mais contanto avanços tecnológicos advindos da aplicação das mais diversas técnicas computacionais. Além disso, com os recentes avanços em diversos tipos de sensores, como microscópicos eletrônicos, têm aumentado rapidamente a quantidade de dados disponíveis para análise. Como consequência, a demanda computacional na área é cada vez maior e isso leva a necessidade de analisar a adequação das diversas alternativas computacionais disponíveis para essa classe de aplicações.

Assim, nesse trabalho, avaliamos o uso de clusters (aglomerados) de computação formados por 16 equipamentos Raspberry Pi 2 [FOUNDATION. 2015] com 64 cores e 16GB RAM em aplicações reais de análise de imagens microscópicas [Kong et al. 2013, Cooper et al. 2010, Teodoro et al. 2014, Kurc et al. 2015]. Essa avaliação inclui a comparação desses processadores de baixo custo com CPUs multicore, levando em conta aspectos tais como tempo de execução, custo do equipamento e gasto energético. O Raspberry tem atraído bastante atenção da comunidade científica e industrial pelo fato de apresentarem um baixo custo e gasto energético [Cox et al. 2013]. Essas características ganham ainda mais importância no atual e complicado cenário econômico nacional.

Entretanto, prover recursos computacionais poderosos com baixo custo parece ser uma contradição e o caminho para tal feito é complexo. Apesar de o Raspberry ser um componente básico promissor, ainda não é claro se o poder computacional desse processador é adequado para as altas demandas das aplicações alvo desse estudo. Além disso, o acréscimo de poder computacional nesse ambiente é obtido através da adição de nós em clusters de computadores, mas o impacto dessa estratégia na escalabilidade das aplicações também precisa ser avaliado. Nossas avaliações são feitas utilizando uma aplicação real de análise de imagens microscópicas de câncer cerebral.

As nossas avaliações utilizam métricas tais como o gasto energético, custo e performance do cluster de Raspberrys proposto, comparado com uma plataforma Core2Duo e I7. Em relação a gasto energético cluster Raspberry foi muito melhor, em sua capacidade máxima, o consumo foi menor do que as máquinas Core2Duo e I7 em modo ocioso (executando apenas o sistema operacional). O custo do cluster foi um pouco mais caro que os demais, porém ele conseguiu ser 10x mais rápido que o Core2Duo e pouco mais que 2x mais rápido comparado com o I7. Dessa maneira, o cluster foi proporcionalmente melhor pela eficiência gerada. Também identificamos limitações na arquitetura Raspberry e conseqüentemente no cluster. Operações que possam usar demasiadamente disco ou que percorram intensivamente, por exemplo, visitando uma vizinhança muito grande para cada ponto analisado em uma imagem incorrem em desvantagens consideráveis comparada com as outras arquiteturas, devido à memória consideravelmente mais lenta do Raspberry.

Apesar de esse artigo ser focado em nossa análise de imagens microscópicas médicas, essa aplicação pertence a uma classe maior de aplicações científicas que analisam dados em baixa dimensionalidade (2D ou 3D) com objetivo de identificar e analisar objetos. Essa classe maior de aplicações inclui aquelas que processam dados de satélites, caracterizam reservatórios subterrâneos e analisam dados de telescópios em astronomia [Klie et al. 2006, Kurç et al. 2005, Parashar et al. 2005, Beynon et al. 2001, Shock et al. 1998, Chang et al. 1997]. Assim, acreditamos que os resultados obtidos das avaliações experimentais podem ser generalizados para outras classes de aplicações.

O restante desse artigo é organizado da seguinte forma. Na Seção 2 descrevemos

a aplicação motivadora em detalhes e as plataformas avaliadas. A implementação da aplicação é descrita na Seção 3 e os resultados experimentais são apresentados na Seção 4. Finalmente, na Seção 5 concluímos o artigo.

2. Aplicação Motivadora e Plataformas Avaliadas

Essa seção apresenta a aplicação real de processamento de imagens microscópicas de tecidos humanos utilizada nesse trabalho, assim como as plataformas nas quais avaliamos experimentalmente a aplicação motivadora.

2.1. Aplicação Motivadora

Alterações morfológicas nos tecidos nas escalas celulares e subcelulares fornecem informações valiosas sobre os tumores, complementando a informação genômica e clínica, podendo levar a uma melhor compreensão da biologia do tumor e resultado clínico [Cooper et al. 2010]. A aplicação motivadora tem sido utilizada no estudo de múltiplos tipos de tumores, que incluem o cerebral e pulmonar.

Os avanços nas tecnologias de captura e digitalização de imagens na última década foram muito significativos e simplificou o processo de capturas de imagens de tecido. Atualmente, é possível obter rapidamente uma imagem de amostra em alta resolução de tecido utilizando microscópio digital. Isso fortaleceu a área de estudos na busca de características morfológicas em escala celular e subcelular. Essa área relativamente nova em comparação as técnicas de ressonância e tomografia computadorizada é conhecida como micrologia digital.

Apesar da facilidade na obtenção das imagens em alta resolução, o processamento das mesmas não é trivial e é agravado pela grande quantidade de dados disponíveis. Uma única imagem pode ter da ordem de $100K \times 100K$ pixels. Esse cenário torna extremamente complexo o processamento dessas imagens, sendo possível identificar milhões de células na mesma. O processamento de uma única imagem pode levar diversas horas ou dias em um processador sequencial. Um estudo em larga escala pode envolver milhares de imagens de alta resolução, que podem ser processadas múltiplas vezes para mensurar o impacto de estudos de parâmetros. Logo, um estudo em larga escala levaria décadas em um processador sequencial, o que inviabilizaria o processo. Desse modo, esses estudos são viáveis somente com a utilização de processamento paralelo, que estão disponíveis hoje em diversas plataformas.

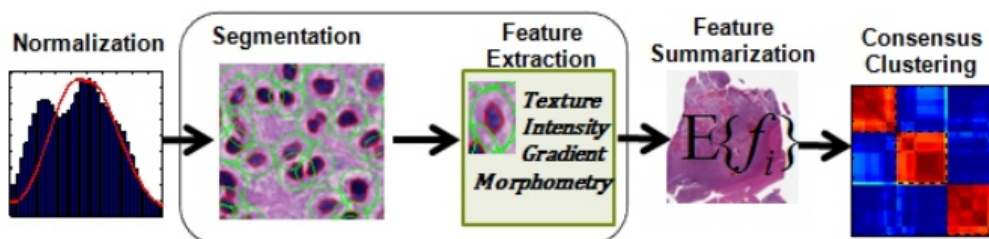


Figure 1. Fluxo de funcionamento da avaliação microscópica.

As fases da aplicação são apresentadas na Figura 1. As imagens são tipicamente dadas como entrada para um estágio de normalização, que remove possíveis ruídos advindo de diferenças na captura da imagem. Em seguida, as células e os núcleos

são detectados e delineados durante a fase de segmentação. Nesta etapa aplica-se uma cascata de operações que incluem limites de valor de pixel e reconstrução morfológica para identificar objetos candidatos, preenchimento de buracos dentro de objetos e definição das áreas onde os objetos não são de interesse. Além disso, objetos são separados utilizando operações como distance transform e watershed. A fase de extração de características calcula um conjunto de propriedades espaciais e textura por objeto, incluem pixels e estatísticas de gradiente, borda e as características morfométricas [Teodoro et al. 2014, Teodoro et al. 2013a, Teodoro et al. 2015]. As etapas de sumarização de características e clusterização são tipicamente mais baratas, pois trabalham em uma representação mais curta das imagens com um vetor de características por imagem.

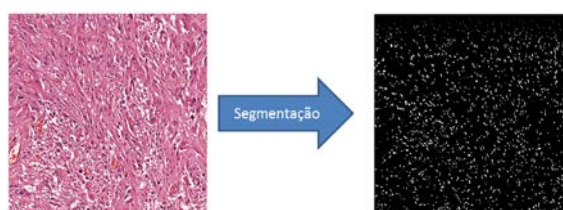


Figure 2. Aplicação da segmentação em uma amostra de tecido cerebral.

As avaliações feitas nesse trabalho utilizam o estágio de segmentação de células da aplicação, pois esta é a etapa mais cara de todo o processo. Na Figura 2 podemos observar um exemplo de entrada (amostra de tecido cerebral) e o resultado da segmentação para a mesma. A descrição do funcionamento dessa fase e as técnicas de programação aplicadas para paralelizar essa fase serão abordadas na sessão de implementação.

2.2. Plataformas Avaliadas

Tivemos grandes evoluções nos últimos anos na computação científica com o crescimento do número de arquiteturas alternativas para execução de aplicações com alta demanda computacional. Como exemplo, podemos citar os coprocessadores como GPUs e o Intel Xeon Phi, além de CPUs multicore com um número cada vez maior de núcleos de processamento. Embora esses processadores tenham conseguido aumentar significativamente a capacidade de processamento de máquinas modernas, eles são encontrados em suas versões mais modernas em valores financeiros elevados. Assim, apresentamos nessa seção, uma arquitetura com um cluster de Raspberry como uma alternativa mais barata em relação a esses processadores, assim como descrevemos duas CPU que são utilizadas para comparação.



Figure 3. Raspberry Pi 2 Modelo B [FOUNDATION. 2015].

A plataforma foi construída utilizando Raspberry Pi 2 B [FOUNDATION. 2015] apresentado na Figura 3, com a configuração apresentada na Tabela 1. Na proposta de um cluster devemos observar com mais atenção processadores, memória RAM e as interconexões de comunicação entre os nós do mesmo. O Raspberry é equipado com um processador Quad Core ARM7 Cortex 900MHz, 1 GB de Memória RAM e interface de rede de 100Mbps/s. O equipamento suporta a utilização de cartão micro SD para o uso de armazenamento de dados, sendo que em nossa solução usamos uma memória micro SD classe 10 Ultra de velocidade até 80MB/s [Corporation 2016b]. Alguns destaques desse equipamento são seu baixo gasto energético (3W) e ausência de sistema de resfriamento, como coolers e dissipadores de calor.

Table 1. Especificações do Raspberry Pi 2 [FOUNDATION. 2015] utilizado.

Ethernet	10/100 BaseT Ethernet socket
USB	4 x USB 2.0 Connector
CPU	Quad-core ARM Cortex-A7 900MHz
card slot	Micro SDIO
Chip	Broadcom BCM2836 SoC
Memory	1GB LPDDR2
Video out	HDMI v1.4
Power	Micro USB socket 5V, 2A

O cluster que construímos é composto por 16 Raspberry Pi2 B com cartão de memória para armazenamento secundário de 16GB classe 10 Ultra. Os processadores são interligados através de um Switch Ethernet de 100MB/s com 24 portas. O sistema operacional utilizado é o Raspian, uma variação do Linux Debian compilado para arquitetura Arm7 [Raspian 2016]. A comunicação entre processos e gerenciamento dos mesmos na nossa solução utiliza a plataforma MPI (Message Passing Interface) [mpi].

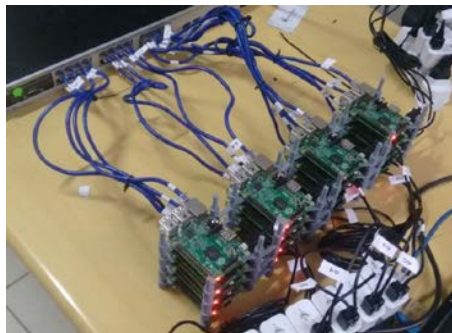


Figure 4. Cluster de baixo custo montando utilizado processadores Raspberry.

Como podemos observar na Figura 4, o cluster foi construído com quatro torres com quatro Raspberrys cada. Temos dois Raspberrys conectados em cada fonte de energia compatível com carregadores de celular 2A. As torres dos processadores foram impressas utilizando uma impressora 3D. Para comparar a eficiência de nosso cluster, decidimos comparar nossa capacidade de processamento com os processadores CPU Core2Duo e I7 apresentados na Tabela 2.

Outro item de nossa avaliação é o consumo de energia de cada um dos equipamentos, que são apresentados na Tabela 3. Como visto, o consumo de energia dos sistemas

Table 2. Configuração dos equipamentos comparados.

Equipamento	Proc.(MHz)	Núcleos	RAM	RAM(MHz)	Processador
Core2Duo	2300	2	4GB	667	Intel(R) Core(TM)2 Duo CPU E6550
I7	3000	8	8GB	1600	Intel(R) Core(TM) i7-4765T
Raspberry Pi 2 B	1000-700	4	1GB	500-400	Armv7

com CPU são maiores até que os do cluster de Raspberrys, mesmo quando comparamos os gastos das CPUs ociosas com os valores máximos dos Raspberrys. Alguns dos fatores que contribuem para o baixo consumo dos Raspberrys é a baixa frequência dos processadores e arquitetura mais simples, que inclusive fazem com que o mesmo não precise de dissipadores ou coolers para o resfriamento. Quando avaliamos o custo financeiro, nota-se que o valor unitário do Raspberry é significativamente menor que dos outros equipamentos, enquanto um cluster com 16 processadores Raspberry incluindo todos os equipamentos para sua montagem é um pouco mais caro que o valor em reais do I7.

Table 3. Consumo de energia e sistema de ventilação. O consumo dos processadores ociosos e trabalhando com carga máxima dos processadores foi extraídos de [Page 2009, Butler 2012]. Os preços dos equipamentos CPU foram extraídos de [DELL 2016], visitado em fevereiro de 2016.

Equipamento	Potência (W)		Preço(R\$)	Observações
	Ocioso	Max.		
Core2Duo	114	163	1.200,00	HD Sata 500GB, 4G RAM
I7	100	538	3.000,00	HD Sata 1TB, 8G RAM
Raspberry Pi2 B	-	3	250,00	1 fonte de 2A, 1 cartão micro SD classe 10 Ultra 16GB
Cluster RB	-	48	3.440,00	8 fontes de 2A, 16 cartão micro SD classe 10 Ultra 16GB, switch 24 portas 10/100 Dlink

3. Implementação

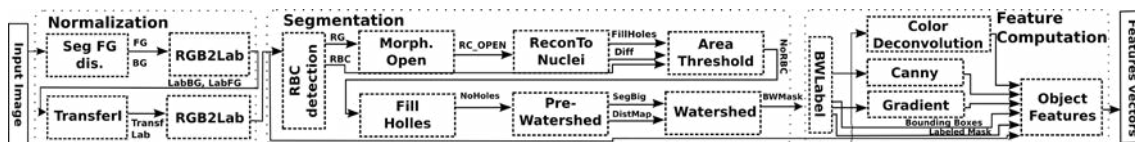
Nosso estudo se baseou na fase de segmentação da aplicação motivadora, cujas operações fundamentais são apresentadas na Tabela 4. Estas operações são categorizadas de acordo com o padrão de acesso a dados e a intensidade de computação. Sempre que possível, usamos implementações sequenciais das operações disponíveis com OpenCV [Bradski 2000] ou de outros grupos de pesquisa, tal como apresentado na coluna Fonte CPU da Tabela 4.

As operações na fase de segmentação realizam computações em elementos do domínio de dados de entrada, pixels em nosso caso. As operações realizadas podem ser categorizadas quanto ao acesso aos dados: regulares e irregulares. Os regulares acessam os dados em regiões contínuas e os irregulares acessam dados distribuídos de maneira irregular. Em alguns casos alguns dados que necessitam serem processados só são definidos em tempo de execução como resultado de alguma computação e por isso teremos um acesso irregular (pseudo-randômico) [Teodoro et al. 2013b, Teodoro et al. 2012]. Ainda em relação ao acesso de dados, temos a classificação em relação à computação de um dado

Table 4. Tabela de operações fundamentais da fase de segmentação

Operação	Descrição	Padrão de Acesso a Dados	Computação	CPU Source
Convert RGB to grayscale	Converter a imagem RGB em escala em cinza	Regular, multicanal local	Moderada	OpenCV
Morphological Open	Abrir e remover pequenos objetos, e preencher buracos no primeiro plano	Regular, vizinhança (disco 13x13)	Baixa	OpenCV
Morphological Reconstruction	Ajustar a imagem a uma máscara.	Irregular, vizinhança (4-/8-connected)	Baixa	Vincent [Vincent 1993]
Area Threshold	Remover objetos fora da área de interesse	Mista, vizinhança	Baixa	Desenvolvido
FillHoles	Preencher buracos em objetos usando pontos selecionados	Irregular, neighborhood (4-/8-connected)	Baixa	Vincent [Vincent 1993]
Distance Transform	Computar a distância do ponto mais próximo do fundo para cada pixel do primeiro plano.	Irregular, vizinhança (8-connected)	Moderada	Desenvolvido
Connected Components Labeling	Rotular com os mesmos pixels de valor em objetos de imagens de uma máscara	Irregular, global	Baixa	Desenvolvido

e pode ser local, operação depende apenas do seu valor, multicanal local, uma variação do local com a possibilidade do acesso aos dados em outras camadas do mesmo índice, vizinhança, operações sobre vizinhos espaciais ou temporais, global, operação dentro da área delimitada. As operações de segmentação variam de baixa a moderada no grau de computação. O fluxo de trabalho detalhado da aplicação é apresentado na Figura 5.

**Figure 5. Fluxo de trabalho detalhado da aplicação.**

A implementação da fase de segmentação foi feita utilizando C++ e sua paralelização utiliza MPI [mpi] para instanciar processos e gerenciar a comunicação entre eles. Nessa implementação, a execução é feita utilizando o modelo Mestre/Escravo, onde o mestre armazena a informação sobre as imagens que precisam ser processadas e assinala as mesmas para execução nos escravos sob demanda. Cada escravo então lê a imagem a ser processada e invoca o estágio de segmentação que é executado sequencialmente. Múltiplos núcleos de processamento em uma máquina ou em um ambiente distribuído são utilizados através criação de várias cópias do processo escravo.

4. Resultados

A avaliação experimental foi realizada utilizando a aplicação motivadora, implementada em C++ com MPI, utilizando 512 imagens de $1K \times 1K$ pixels como entrada. As imagens somam um total de 6GB de dados.

4.1. Avaliação do desempenho dos processadores para diferentes configurações

Uma preocupação inicial com a execução no Raspberry era a velocidade de leitura de arquivos, já que a primeira etapa da aplicação lê a imagem antes de repassar a mesma para computação na segmentação. Ao contrário dos outros equipamentos que usam HD SATA, o Raspberry utiliza uma memória micro SD que possui uma velocidade inferior. Logo, avaliamos variações do micro SD para analisar o impacto do mesmo para o desempenho da aplicação. As versões analisadas foram o micro SD Classe 10 Ultra que chega a velocidade de 80MB/s [Corporation 2016b] e micro SD Classe 10 Extreme que chega a 95MB/s [Corporation 2016a]. O Raspberry possui um processador Arm7 que tem velocidade variável de 700MHZ a 1000MHZ. Também avaliamos a velocidade mínima e máxima do processador para verificar seu impacto nos resultados.

Table 5. Proc. 512 imagens com apenas um único processo trabalhador (1 core apenas é utilizado para processamento).

Equipamento	Vel.(MHz)	Total(min.)	Img Read(seg.)
Core 2 Duo	2300	100,07	79,84
RB SD Classe 10 Extreme	1000	186,43	338,23
RB SD Classe 10 Ultra	1000	187,25	345,93
RB SD Classe 10 Extreme	700	227,95	339,78
RB SD Classe 10 Ultra	700	227,45	345,36
I7	3000	49,17	49,55

A Tabela 5 apresenta o tempo total de execução e tempo de leitura de arquivo de imagem na sua execução sequencial da aplicação em cada processador para as diferentes configurações. A diferença de velocidade pelo uso do micro SD Classe 10 Ultra ou Extreme não alterou significativamente os resultados. Como cada arquivo é lido uma única vez no início da execução e os resultados gravados ao fim da mesma, o ganho com o micro SD mais rápido é pequeno, pois as operações de E/S representam uma parte pequena do tempo de execução. Entretanto, a comparação dos tempos de leitura do RB com as outras plataformas mostra uma grande diferença, que pode ser impactante para outras aplicações mais intensivas em E/S. A comparação do tempo de execução total da aplicação mostra que cada core do I7 é cerca de $3,8\times$ mais rápido que um core do RB, enquanto essa diferença diminuiu para aproximadamente $1,8\times$ para o Core 2.

4.2. Desempenho por tipo operação

Nessa seção apresentamos os dados referentes ao custo de execução de cada uma das operações principais do estágio de segmentação nas plataformas alvo da nossa análise. Os valores apresentados na Tabela 6 referem-se ao custo médio de cada etapa na execução das 512 imagens utilizadas como entrada.

Table 6. Tempo médio de cada etapa da fase de segmentação em segundos

Equipamento	RBC	Morph.Open	ReconToNuclei	AreaThreshold	FillHolles	Pré-Watershed	Watershed
Core2Duo	247	418	2282	241	1238	1966	519
I7	92	11	1196	116	633	1047	301
Raspberry Pi2B	1533	1891	1735	515	1155	5178	978

Conforme pode ser visto, existe uma grande variabilidade em relação ao desempenho relativo entre os processadores para as diversas operações. Em especial, o Raspberry apresentou um desempenho significativamente inferior as outras plataformas na execução do Morph.Open, que é uma convolação que utiliza um disco de raio relativamente grande. Assim, para cada ponto da imagem, um grande número de vizinhos precisa ser acessado. Entretanto, como a memória do Raspberry é mais lenta e seu cache menor, logo, o desempenho do processador para essa operação é ruim. Inicialmente essa operação apresentava um tempo de execução de 18 segundos para cada imagem. Como essa operação se tornou um gargalo para nossa solução, realizamos uma modificação em sua implementação. Pesquisamos alternativas a essa função e simulamos algumas configurações. Descobrimos que poderíamos gerar o mesmo resultado diminuindo o tamanho do disco e alterando o tamanho da vizinhança que deveria ser acessada. O resultado então pode ser melhorado para aproximadamente 3 segundos para cada imagem.

Com isso diminuimos o acesso intensivo a memória minimizando o gargalo de memória do Raspberry.

O desempenho relativo do Raspberry em relação aos outros processadores é melhor para operações irregulares, pois nenhuma arquitetura é otimizada para esses tipos de acesso a dados e, inevitavelmente, nesse caso o cache vai ter menos eficiência e o custo de acesso a memória é o limitador.

4.3. Avaliação de escalabilidade e melhor desempenho em cada plataforma

Nessa seção apresentamos os resultados onde os equipamentos são avaliados instanciando um processo MPI trabalhador por núcleo de processamento, sendo assim possível processar múltiplas imagens em paralelo. A Figura 6 apresenta o speedup obtido na execução da aplicação no cluster Raspberry. Como visto, a aplicação atingiu um speedup de aproximadamente $47\times$ em relação a sua versão sequencial rodando no mesmo processador, o que leva a uma eficiência de aproximadamente 75%. Adicionalmente, a eficiência na execução com 4 núcleos em um único Raspberry é de 80%. Logo, as perdas de eficiência não são relacionadas a utilização de múltiplas máquinas, mas em relação ao uso dos núcleos de processamento em um Raspberry. Isso se deve ao fato de que essa aplicação é sabidamente intensiva em acesso a dados e algumas de suas operações fazem acesso irregulares à memória, o que leva a criação de um gargalo nos subsistemas de memória e limitação da escalabilidade de um processador.

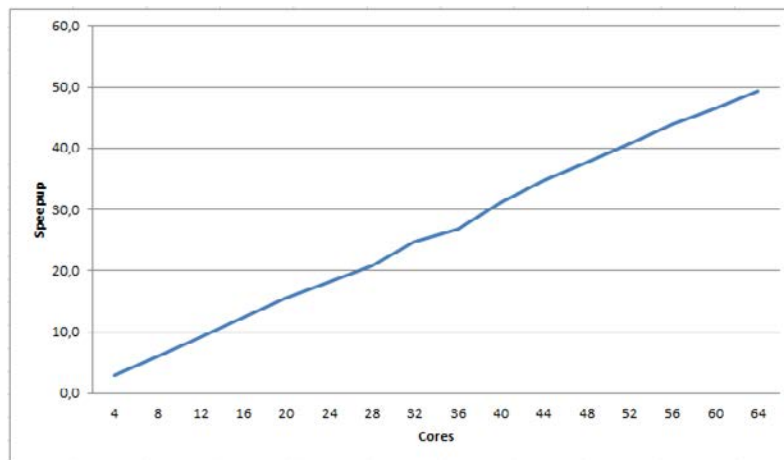


Figure 6. Speedup pela quantidade de trabalhadores (núcleos de processamento)

A Tabela 7 apresenta resultados comparativos das plataformas avaliadas. Como visto, o I7 atingiu a pior eficiência entre os processadores analisados. Esse resultado é consequência da maior quantidade de núcleos de processamentos desse processador, que são acoplados ao mesmo subsistema de memória que se torna um gargalo rapidamente devido ao fato das operações serem intensivas em dados. Esse efeito é menor no cluster de Raspberry, pois cada uma das memórias funciona independentemente e criam um ambiente mais adequado para aplicação. Adicionalmente, como visto, o cluster apresentou o melhor tempo de execução dentre as plataformas avaliadas. Ele é $10,62\times$ mais rápido que o Core2Duo e $2,32\times$ mais rápido que o I7.

Discussão: Em relação ao consumo de energia, apesar de não termos medido exatamente o que os equipamentos consomem, nos baseamos na documentação encontrada

Table 7. Resultados de tempo na capacidade máxima dos equipamentos.

Equipamento	Núcleos	Tempo(min)	speedup	Eficiência
Core 2 Duo	2	51,33	1,95	97%
I7	8	11,22	4,38	55%
Cluster Raspberry	64	4,83	47,16	78%

sobre as plataformas. Cada Raspberry Pi2 B na pior hipótese irá consumir cerca de 3W, ou seja, e baseado nisso o cluster com os 16 Raspberrys irão consumir na pior hipótese 48W. Buscando as informações de consumo do Core2Duo e do I7, apenas os processadores, desconsiderando o resto dos equipamentos necessários para montar a CPU, em modo ocioso esses processadores irão consumir 114W e 100W respectivamente. Logo o cluster irá consumir $2\times$ menos energia que qualquer outro equipamento no pior caso e o mesmo tempo de execução. Entretanto, essa diferença é ainda maior na prática porque as CPU não estarão ociosas e elas demoram mais tempo para executar a aplicação. Ainda no aspecto físico, o Raspberry não necessita de dispositivos de ventilação o que diminui custos para criação de racks ou salas cofres com sistema especial de resfriamento.

5. Conclusão e Trabalhos Futuros

Este trabalho avaliou alternativas computacionais para executar aplicações de processamento de imagens médicas microscópicas de tecidos, capazes de oferecer um bom custo benefício e os requisitos computacionais necessários para executar eficientemente essas aplicações. Para tanto, focamos no estudo de plataforma de baixo custo e comparamos custos, consumo energético e performance com outros processadores. Dentre essas plataformas, se destaca o Raspberry, que tem todos os atributos anteriores, mas com poder computacional limitado. Assim, para atingir a capacidade de computação desejada construímos um cluster com 16 máquinas Raspberrys, que foi comparada com processadores Core2Duo e I7.

O estudo realizado utiliza a fase de segmentação de nossa aplicação motivadora por ser a que demanda mais recurso. Verificamos que o cluster em seu consumo máximo de energia era de 48W enquanto o consumo de energia do Core2Duo e I7 respectivamente 114 e 110 em modo ocioso. Levantamos os custos financeiros de cada plataforma e constatamos que o cluster era um pouco mais custoso que o I7 e cerca de quase $2x$ mais caro que o Core2Duo. Entretanto, a avaliação de desempenho comparativa das plataformas mostrou que o cluster com Raspberry é $2x$ mais rápido que o I7 e $10x$ mais rápido que o Core2Duo. Assim, o custo benefício cluster em função do desempenho alcançado é muito melhor que das outras plataformas.

Embora o desempenho geral da aplicação tenha sido melhor no cluster com Raspberry, o comportamento de desempenho relativo as plataformas em função das operações internas da aplicação pode variar conforme a forma de acesso a dados e intensidade de computação. Se tivermos um algoritmo que explora muito a memória RAM pode ser que o Raspberry seja inferior por ser equipado com memória RAM mais lenta. Outro potencial gargalo dessa máquina pode ser a leitura e escrita de arquivos em uma aplicação que realize muitas operações E/S.

Durante nossos testes, nos deparamos com alguns resultados ruins na execução do cluster Raspberry. A etapa de Morphological Open apresentava um resultado de tempo

muito ruim. Avaliando mais a fundo, descobrimos que essa operação precisa percorrer de maneira intensa a memória RAM. Como o Raspberry Pi2 B possui uma memória RAM bem mais lentas que os demais equipamentos, foi necessário realizar alterações. Verificamos que o tempo estava relacionado diretamente com o tamanho do disco de entrada para realização da função Morphological. Conseguimos ajustar esse disco e chegar a um resultado muito melhor de desempenho sem alterar o resultado da aplicação. Com essa situação, ficou claro que devemos pesquisar mais a fundo o funcionamento das etapas de segmentação para melhorarmos sua eficiência com foco na plataforma de nosso cluster. Outra linha de trabalho que devemos continuar a desenvolver são as demais fases de nossa aplicação, normalização e extração de características.

Agradecimentos. Gostaríamos de agradecer ao CNPq e CAPES pelo apoio ao trabalho e a empresa CleverSystems (<http://www.cleversystems.com.br>), que nos forneceu os equipamentos Rasperrys Pi2 B e a estrutura para realização dos testes.

References

The Message Passing Interface (MPI).

Beynon, M. D., Kurc, T., Catalyurek, U., Chang, C., Sussman, A., and Saltz, J. (2001). Distributed processing of very large datasets with DataCutter. *Parallel Comput.*, 27(11):1457–1478.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Butler, H. (2012). Intel core i7-3770k power consumption. <http://www.bit-tech.net/hardware/2012/04/23/intel-core-i7-3770k-review/8>.

Chang, C., Moon, B., Acharya, A., Shock, C., Sussman, A., and Saltz, J. H. (1997). Titan: A High-Performance Remote Sensing Database. In *Proceedings of the Thirteenth International Conference on Data Engineering, ICDE '97*, pages 375–384, Washington, DC, USA. IEEE Computer Society.

Cooper, L. A. D., Kong, J., Gutman, D. A., Wang, F., Cholleti, S. R., Pan, T. C., Widener, P. M., Sharma, A., Mikkelsen, T., Flanders, A. E., Rubin, D. L., Meir, E. G. V., Kurc, T. M., Moreno, C. S., Brat, D. J., and Saltz, J. H. (2010). An integrative approach for in silico glioma research. *IEEE Trans Biomed Eng.*, 57(10):2617–2621.

Corporation, S. (2016a). Cartao de memoria sandisk extreme microsdhc/microsdx uhs-i. <https://www.sandisk.com.br/home/memory-cards/microsd-cards/extreme-microsd>.

Corporation, S. (2016b). Cartao de memoria sandisk ultra microsdhc/microsdx uhs-i para camaras. <https://www.sandisk.com.br/home/memory-cards/microsd-cards/ultra-microsd-for-cameras>.

Cox, S. J., Cox, J. T., Boardman, R. P., Johnston, S. J., Scott, M., and O'Brien, N. S. (2013). Iridis-pi: a low-cost, compact demonstration cluster. *Cluster Computing*, 17(2):349–358.

DELL (2016). Ecommerce dell. <http://www.dell.com.br>.

FOUNDATION., R. P. (2015). Raspberry pi 2 model b. <https://www.raspberrypi.org/products/raspberry-pi-2-model-b/>.

Klie, H., Bangerth, W., Gai, X., Wheeler, M. F., Stoffa, P. L., Sen, M. K., Parashar, M., Catalyürek, Ü. V., Saltz, J. H., and Kurç, T. M. (2006). Models, methods and middleware for grid-enabled multiphysics oil reservoir management. *Eng. Comput. (Lond.)*, 22(3-4):349–370.

- Kong, J., Cooper, L. A. D., Wang, F., Gao, J., Teodoro, G., Mikkelsen, T., Schniederjan, M. J., Moreno, C. S., Saltz, J. H., and Brat, D. J. (2013). Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates. *PLoS ONE*.
- Kurc, T., Qi, X., Wang, D., Wang, F., Teodoro, G., Cooper, L., Nalisnik, M., Yang, L., Saltz, J., and Foran, D. J. (2015). Scalable analysis of big pathology image data cohorts using efficient methods and high-performance computing strategies. *BMC Bioinformatics*, 16(1):1.
- Kurç, T. M., Çatalyürek, Ü. V., Zhang, X., Saltz, J. H., Martino, R., Wheeler, M. F., Peszynska, M., Sussman, A., Hansen, C., Sen, M. K., Seifoullae, R., Stoffa, P. L., Torres-Verdín, C., and Parashar, M. (2005). A simulation and data analysis system for large-scale, data-driven oil reservoir simulation studies. *Concurrency - Practice and Experience*, 17(11):1441–1467.
- Page, M. (2009). Cpu power consumption tests. <http://www.pcstats.com/articleview.cfm?articleid=2395&page=2>.
- Parashar, M., Matossian, V., Bangerth, W., Klie, H., Rutt, B., Kurç, T. M., Çatalyürek, Ü. V., Saltz, J. H., and Wheeler, M. F. (2005). Towards Dynamic Data-Driven Optimization of Oil Well Placement. In *International Conference on Computational Science (2)*, pages 656–663.
- Raspian (2016). Raspian. <http://www.raspian.org/>.
- Shock, C. T., Chang, C., Moon, B., Acharya, A., Davis, L., Saltz, J., and Sussman, A. (1998). The design and evaluation of a high-performance earth science database. *Parallel Computing*, 24(1):65 – 89.
- Teodoro, G., Kurc, T., Andrade, G., Kong, J., Ferreira, R., and Saltz, J. (2015). Application performance analysis and efficient execution on systems with multi-core CPUs, GPUs and MICs: a case study with microscopy image analysis. *International Journal of High Performance Computing Applications*.
- Teodoro, G., Kurç, T. M., Kong, J., Cooper, L. A. D., and Saltz, J. H. (2014). Comparative Performance Analysis of Intel (R) Xeon Phi (TM), GPU, and CPU: A Case Study from Microscopy Image Analysis. In *2014 IEEE 28th International Parallel and Distributed Processing Symposium, May 19-23, 2014*, pages 1063–1072.
- Teodoro, G., Pan, T., Kurc, T. M., Cooper, L., Kong, J., and Saltz, J. H. (2012). A fast parallel implementation of queue-based morphological reconstruction using GPUs. *Emory University, Center for Comprehensive Informatics, CCI-TR-2012-2*.
- Teodoro, G., Pan, T., Kurc, T. M., Kong, J., Cooper, L. A., Podhorszki, N., Klasky, S., and Saltz, J. H. (2013a). High-throughput Analysis of Large Microscopy Image Datasets on CPU-GPU Cluster Platforms. In *IPDPS '13: Proceedings of the 2013 IEEE International Symposium on Parallel and Distributed Processing*.
- Teodoro, G., Pan, T., Kurc, T. M., Kong, J., Cooper, L. A., and Saltz, J. H. (2013b). Efficient Irregular Wavefront Propagation Algorithms on Hybrid CPU-GPU machines. *Parallel Computing*, 39(4-5):189–211.
- Vincent, L. (1993). Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Transactions on Image Processing*, 2:176–201.

Seleção Adaptativa de Operadores Aplicada ao Problema do Despacho Econômico de Energia Elétrica

Richard Aderbal Gonçalves¹, Carolina Paula de Almeida¹,
Sandra Mara G. S. Venske¹, Josiel N. Kuk¹, Lucas M. Pavelski¹

¹Departamento de Ciência da Computação – Universidade Estadual do Centro-Oeste
UNICENTRO – Guarapuava – PR – Brasil

{richard, carol, ssvenske, jnkuk, lmpavelski}@unicentro.br

Abstract. *The Economic Load Dispatch Problem is one of the most important problems in the generation and distribution of electrical energy. Differential Evolution is an efficient evolutionary algorithm for continuous optimization. Distinct Differential Evolution operators are adequate to solve problems with distinct characteristics, however the choice of which operator is more adequate is a complex task. In this work, two adaptive operator selection techniques (Adaptive Pursuit and Probability Matching) are investigated to select the most appropriate operator to solve the Economic Load Dispatch Problem during execution time. The proposed algorithms are evaluated in problems with 13 and 40 thermal generators with valve-point loading effect. The proposed methods achieved better results than recent metaheuristics, finding the best known minimum cost value for all considered systems.*

Resumo. *O Despacho Econômico de Energia Elétrica é um dos mais importantes problemas na área de geração e distribuição de energia elétrica. A Evolução Diferencial é um algoritmo evolutivo eficiente para otimização contínua. Diferentes operadores da Evolução Diferencial são adequados para a resolução de problemas com características diferentes, contudo a escolha do operador mais adequado é uma tarefa complexa. Neste trabalho são investigadas duas técnicas de seleção adaptativa de operadores (Adaptive Pursuit e Probability Matching) para escolher em tempo de execução qual o operador mais eficiente para a resolução do Despacho Econômico de Energia Elétrica. Os algoritmos propostos são validados em problemas de teste que consideram 13 e 40 geradores térmicos e levam em consideração efeitos de ponto de válvula. Os métodos propostos superam os resultados reportados na literatura obtidos por metaheurísticas modernas, sendo capazes de encontrar o melhor valor de custo mínimo conhecido para todos os sistemas de teste considerados.*

1. Introdução

O Despacho Econômico de Energia Elétrica (DEEL) é um dos mais importantes problemas a serem resolvidos durante o planejamento e operacionalização de um sistema de geração de energia elétrica [Park et al. 2006]. O objetivo do despacho econômico é minimizar o custo total de combustível de geradores de energia elétrica sujeito a restrições operacionais. Neste trabalho, são considerados os efeitos de ponto de válvula na geração de energia, tornando o problema de otimização não-linear e multimodal. O DEEL na

prática corresponde a um problema de otimização complexo no qual o uso de técnicas de otimização clássicas mostram-se ineficientes e, portanto, técnicas de inteligência computacional [Engelbrecht 2007], têm se tornado populares na sua solução.

Neste trabalho é explorada a técnica de Evolução Diferencial (ED)[Storn and Price 1997]. Diferentes operadores da ED se mostram mais eficientes de acordo com as características dos problemas a serem resolvidos. A determinação do operador mais adequado é uma tarefa árdua e, além disso, a adequação dos operadores pode depender do estado atual do processo de otimização, ou seja, diferentes operadores podem ser mais eficientes em diferentes etapas da otimização. A seleção adaptativa de operadores visa determinar em tempo de execução qual ou quais operadores devem ser utilizados em cada etapa do processo de otimização de maneira automática [Gong et al. 2011]. Duas das principais técnicas de seleção adaptativa de operadores são a *Adaptive Pursuit* (AP) e *Probability Matching* (PM) [Thierens 2005]. Portanto, esse trabalho visa investigar a adequação do uso da AP e da PM para a seleção de operadores da ED para a resolução do Problema do Despacho Econômico de Energia Elétrica.

O restante deste trabalho é organizado conforme a seguir. Na Seção 2 o problema do Despacho Econômico de Energia Elétrica é brevemente descrito. A Seção 3 apresenta os principais conceitos da Evolução Diferencial e de seleção adaptativa de operadores. Os métodos propostos são descritos na Seção 4 enquanto a Seção 5 apresenta os experimentos e os resultados obtidos com os algoritmos implementados, comparando-os com algoritmos da literatura. Finalmente, a Seção 6 contém as conclusões e trabalhos futuros.

2. Despacho Econômico de Energia Elétrica

Este trabalho foca na otimização da geração de energia elétrica, ou seja, na minimização dos custos de produção de energia. Tal otimização tem grande importância econômica na sociedade moderna.

Uma usina elétrica possui vários geradores, sendo cada gerador responsável por produzir parte da energia demandada (P_D). Cada gerador tem atributos diferentes e para que se tenha a melhor eficiência de todos os geradores, deve-se ajustar os parâmetros de geração de energia de cada gerador, de modo que o custo conjunto de todos seja minimizado. O problema de despacho de energia elétrica pode ser descrito de maneira formal do seguinte modo [Basu 2002]:

$$\begin{aligned} & \text{Minimizar } C = \sum_{j=1}^n C_j(P_j) \\ & \text{sujeito a } \sum_{j=1}^n P_j - P_D = 0 \quad \text{e} \\ & \quad P_j^{min} \leq P_j \leq P_j^{max} \end{aligned} \quad (1)$$

tal que, C é o custo total considerando todos os geradores, P_j é a energia produzida pelo j -ésimo gerador, $C_j(P_j)$ é a função de custo do j -ésimo gerador (em \$/hr), P_j é a potência de saída da j -ésima unidade, n é o número de unidades geradoras no sistema, P_D é a demanda total de potência, P_j^{min} e P_j^{max} são, respectivamente, a potência de saída mínima e máxima da j -ésima unidade. A restrição de igualdade da Equação 1 é chamada de restrição de balanceamento de potência enquanto a restrição de desigualdade é chamada de restrição operacional. De acordo com as restrições, a potência total gerada deve ser igual a demanda total do sistema e a potência gerada por cada unidade deve estar

entre os seus limites mínimos e máximos.

O cálculo de custo (C) é dado pela soma do custo de cada gerador. Neste trabalho, o custo dos geradores leva em consideração o efeito de ponto de válvula, que é a interferência no custo por conta da abertura de válvulas [Basu 2002]. Considerando o efeito de ponto de válvula, o custo de cada gerador é afetado por uma função senoidal, conforme a Equação 2:

$$C_j(P_j) = a_j * P_j^2 + b_j * P_j + c_j + |e_j * \sin(f_j * (P_j^{min} - P_j))| \quad (2)$$

tal que a_j , b_j , c_j , e_j e f_j são os coeficiente empíricos de custo de combustível da j -ésima unidade com efeito de ponto de válvula.

3. Evolução Diferencial: clássica e adaptativa

A Evolução Diferencial (ED) [Storn and Price 1997] é um algoritmo evolutivo para otimização numérica que, basicamente, adiciona a diferença ponderada entre dois vetores da população (chamados de vetores de diferença) a um terceiro vetor (chamado de vetor alvo). A ED trabalha com alguns parâmetros como tamanho da população (N), constante de cruzamento (CR) e peso do vetor de diferenças - constante de mutação (F).

O algoritmo da ED possui algumas variações, conhecidas como estratégias de mutação. A notação adotada para caracterizar as variações é ED/ $x/y/z$. Estas estratégias diferem especialmente na forma em que os vetores utilizados no cálculo das diferenças são escolhidos (x), no número de vetores de diferença utilizados (y) e na forma em que o ponto de cruzamento é determinado (z). Neste trabalho são utilizadas as estratégias ED/rand/1/bin, ED/rand/2/bin, ED/current-to-rand/1 e ED/current-to-rand/2. Basicamente, indivíduos aleatórios (*rand*) ou o indivíduo corrente e outros aleatórios (*current-to-rand*) são selecionados, há $y = 1$ ou $y = 2$ pares de soluções que são aleatoriamente escolhidas para calcular a mutação diferencial e o cruzamento utilizado é o binomial ($z=bin$) ou específico (ED/current-to-rand/1 e ED/current-to-rand/2) [Storn and Price 1997].

O desempenho do algoritmo de evolução diferencial é altamente dependente da escolha dos seus parâmetros. O método de “tentativa e erro” para ajustar os parâmetros e escolher a melhor estratégia de mutação a ser utilizada requer tempo e conhecimento tanto sobre a técnica quanto sobre o problema abordado. Na literatura, há alguns algoritmos com controle adaptativo de parâmetros para a ED, tais como: SaDE [Qin and Suganthan 2005, Qin et al. 2009], JADE [Zhang and Sanderson 2009], EPSDE [Mallipeddi et al. 2011], AdapSS [Gong et al. 2010, Gong et al. 2011], AdapMODE [Li et al. 2011], ENS-MOEA/D [Zhao et al. 2012], entre outros. Algumas destas abordagens utilizam o conceito de conjunto (ou *pool*) de estratégias. Neste trabalho, a parte adaptativa da ED adota os métodos de *Probability Matching* (PM) e *Adaptive Pursuit* (AP) [Gong et al. 2011, Thierens 2005, Goldberg 1990] para atualizar a probabilidade de aplicação de cada uma das quatro estratégias de mutação da ED adotadas.

Considerando $S > 1$ estratégias no conjunto de operadores $\varsigma = \{s_1, \dots, s_S\}$ e o vetor de probabilidade $\mathbf{p}(g) = (p_1(g), \dots, p_S(g))$, sendo $p_s(g)$ a probabilidade de uso da s -ésima estratégia na geração g , $\sum_{s=1}^S p_s(g) = 1$; $\forall g : p_{min} \leq p_s(g) \leq 1$, sendo que p_{min} é o valor de probabilidade mínimo de cada estratégia, utilizado para garantir

que não se perca nenhuma estratégia [Thierens 2005]. Há muitos métodos para atualizar adaptativamente a probabilidade $p_s(g)$ da s -ésima estratégia, geralmente baseados na sua qualidade estimada empiricamente $q_s(g)$. Na maior parte do tempo, $q_s(g)$ é atualizada por um valor de *recompensa* recebido. Denota-se $r_s(g)$ como a recompensa que a s -ésima estratégia recebe depois de ser adotada na geração g . A recompensa adotada nesse trabalho é dada pela Equação 3.

$$r_s(g) = fitness_{melhor} / fitness_{filho} * |fitness_{pai} - fitness_{filho}| \quad (3)$$

A qualidade estimada empiricamente $q_s(g)$ da s -ésima estratégia na geração g pode ser atualizada como a seguir [Thierens 2005]:

$$q_s(g+1) = q_s(g) + \alpha * [r_s(g) - q_s(g)] \quad (4)$$

tal que $\alpha \in (0, 1]$ é a taxa de adaptação. Quanto maior a taxa de adaptação, maior a contribuição das recompensas mais recentes.

Com base nesta medida estimada, PM e AP diferem na forma em que estes métodos utilizam esta informação ($q_s(g+1)$) a fim de atualizar a probabilidade do uso de cada estratégia. O método *Probability Matching* calcula $p_s(g+1)$ como a seguir [Gong et al. 2011, Thierens 2005, Goldberg 1990]:

$$p_s(g+1) = p_{min} + (1 - S * p_{min}) * \frac{q_s(g+1)}{\sum_{s=1}^S q_s(g+1)} \quad (5)$$

Claramente, $\sum_{s=1}^S p_s(g+1) = 1$. A partir da Eq. 5 pode ser notado que quando somente uma estratégia obtém valor de recompensa durante um longo período de tempo, sua probabilidade de seleção converge para $p_{max} = p_{min} + (1 - S * p_{min})$.

O método *Adaptive Pursuit* foi inicialmente projetado para seleção adaptativa de operadores no contexto de Algoritmos Genéticos [Thierens 2005]. AP calcula $p_s(g+1)$ com a seguir:

$$p_{s^*}(g+1) = p_{s^*}(g) + \beta * [p_{max} - p_{s^*}(g)] \quad (6)$$

e $\forall s \neq s^* : p_s(g+1) = p_s(g) + \beta * [p_{min} - p_s(g)]$, com $s^* = \operatorname{argmax}_s(q_s(g+1))$ e $p_{max} = p_{min} + 1 - S * p_{min}$.

Esta restrição garante que se $\sum_{s=1}^S p_s(g) = 1$, então $\sum_{s=1}^S p_s(g+1) = 1$ [Thierens 2005]. O método AP tem uma taxa de aprendizado $\beta \in (0, 1]$, que controla quão gulosa será a estratégia adaptativa.

4. Algoritmos Propostos - ED_{AP} e ED_{PM}

Neste trabalho são propostas duas variações da Evolução Diferencial utilizando seleção adaptativa de operadores: ED_{AP} e ED_{PM}. A diferença entre ambas propostas reside somente na técnica de seleção e, portanto, o Algoritmo 1 descreve os passos tanto do ED_{AP} quanto do ED_{PM}.

O primeiro passo do algoritmo consiste na inicialização aleatória das soluções, respeitando as restrições operacionais de cada gerador. Como os indivíduos da população

inicial podem não respeitar a restrição de demanda, o procedimento de reparação é chamado (ver Algoritmo 2). A seguir o custo de todas as soluções é calculado, conforme as Equações 1 e 2. Na sequência o algoritmo entra em seu laço principal, o qual é repetido até que o critério de parada seja satisfeito (neste trabalho corresponde a um número máximo de avaliações). No laço principal cada solução passa pelas etapas de seleção e aplicação dos operadores, avaliação e reparação das soluções geradas, cálculo das recompensas e das qualidades empíricas, bem como a atualização das probabilidades de aplicação de cada operador.

A seleção do operador da ED a ser aplicado é estocástica, dependendo apenas da probabilidade atual de aplicação de cada operador (inicialmente todos os operadores possuem a mesma probabilidade). De acordo com a operação selecionada são escolhidos M indivíduos para a aplicação do operador. O valor de M é 3 para os operadores ED/rand/1/bin e ED/current-to-rand/1 e 5 para os operadores ED/rand/2/bin e ED/current-to-rand/2. Finalmente, o operador da Evolução Diferencial é aplicado para a geração de uma nova solução.

A nova solução pode ser ineficaz e, portanto, deve ser reparada através do procedimento descrito no Algoritmo 2. Após a reparação a nova solução é avaliada e a seleção é aplicada. No caso da ED a seleção é um torneio entre a solução original e a nova solução gerada, onde a melhor é mantida na população.

Os cálculos da recompensa e da qualidade empírica do operador utilizado são feitos de acordo com as Equações 3 e 4, respectivamente. Finalmente, no último passo do laço principal, as probabilidades de aplicação de cada operador são atualizadas de acordo com a Equação 5 caso a técnica *Probability Matching* esteja sendo utilizada ou pela Equação 6 caso a *Adaptive Pursuit* esteja sendo empregada.

5. Resultados

Foram realizadas 50 execuções independentes de cada algoritmo nos três casos de testes considerados. Para todos os experimentos, o tamanho da população foi igual a dez vezes o número de geradores ($10n$) conforme sugerido em [Storn and Price 1995] e o critério de parada adotado foi 2.000.000 de avaliações. Os valores de F e CR foram 0,5 e 0,1, respectivamente¹. São utilizados três estudos de casos para avaliar a qualidade dos algoritmos propostos: dois com 13 geradores e um com 40 geradores.

5.1. Resultados para o Primeiro Estudo de Caso - Caso 13 Geradores e Demanda de 2520 MW

O primeiro estudo de caso considera 13 geradores com efeito de ponto de válvula e demanda de 2520 MW [Coelho and Mariani 2006]. A Tabela 1 apresenta os resultados obtidos pelos algoritmos propostos (ED_{AP} e ED_{PM}) e os da literatura.

Para este caso, os algoritmos propostos (ED_{AP} e ED_{PM}), obtiveram os melhores valores para os três indicadores de custo: mínimo (24164,05), médio (24164,05) e máximo (24164,05) com desvio padrão nulo (0,00), indicando que todas as execuções conseguiram atingir o valor mínimo conhecido. Destaca-se que dentre os algoritmos da literatura apenas o SDE consegue achar o valor mínimo, ou seja, considerando todos os

¹Os valores de F e CR foram empiricamente determinados.

Algoritmo 1: ED_{AP} e ED_{PM} **Entrada:** Parâmetros NP, CR e F.**Saída:** Melhor solução encontrada ao longo da evolução.**início**

Inicializar aleatoriamente a população inicial de tamanho NP;
 Aplicar o Procedimento de Reparação;
 Avaliar cada solução da população inicial;

repita**para cada solução faça**

Selecionar o Operador da ED utilizando AP ou PM;
 Selecionar M soluções da população;
 Aplicar o Operador da ED selecionado;
 Aplicar o Procedimento de Reparação;
 Avaliar o Custo do novo indivíduo;
 Aplicar a Seleção;
 Calcular a Recompensa e a Qualidade Empírica;
 Atualizar as Probabilidades dos Operadores da ED;

fim**até** que o critério de parada seja satisfeito;**fim****Algoritmo 2:** Procedimento de Reparação**Entrada:** Solução**Saída:** Solução Factível**início****para cada componente j da solução i faça****se** $sol_{i,j} < P_j^{min}$ **então**| $sol_{i,j} = P_j^{min}$ **senão se** $sol_{i,j} > P_j^{max}$ **então**| $sol_{i,j} = P_j^{max}$ **fim se****fim****repita**

Selecionar aleatoriamente um componente j

se $(\sum_{i=1}^n P_i < PD)$ **então**| Adicionar uma quantidade a $sol_{i,j}$ que não viole a restrição operacional e minimize a violação**senão**| Subtrair uma quantidade de $sol_{i,j}$ que não viole a restrição operacional e minimize a violação**fim se****até** $\sum_{i=1}^n P_i = PD$;**fim**

Tabela 1. Resultados para o Primeiro Estudo de Caso. Custos em US\$

Caso com 13 geradores e demanda de 2520 MW				
Método	Custo Mín	Custo Médio	Custo Máx	Desvio
ED_{AP}	24164,05	24164,05	24164,05	0,00
ED_{PM}	24164,05	24164,05	24164,05	0,00
SDE [Balamurugan and Subramanian 2007]	24164,05	24168,28	24200,05	-
DTSA [Khamsawang et al. 2004]	24169,95	-	-	-
DE [Noman and Iba 2008]	24169,91	24169,91	24169,91	-
SSA [Yu and Li 2015]	24169,91	24169,91	24169,91	-
DPO [Han et al. 2015]	24169,90	24169,90	24169,90	-
HCRO-DE [Roy et al. 2014]	24164,82	24164,98	24165,34	0,06
TLBO [Banerjee et al. 2015]	24197,00	-	-	-
ORCSA [Nguyen and Vo 2015]	24169,91	-	-	-

indicadores de custo os algoritmos propostos tiveram desempenho superior aos reportados na literatura. A Tabela 2 apresenta o melhor resultado obtido pelos algoritmos propostos. A solução é codificada por um vetor P_i , $i = 1, \dots, 13$ com custo mínimo de 24164,05.

Tabela 2. Custo Mínimo para o 1º Estudo de Caso utilizando ED_{AP} e ED_{PM} .

Potência	Geração	Potência	Geração	Potência	Geração
P_1	628,3185	P_6	159,7331	P_{11}	77,3999
P_2	299,1993	P_7	159,7331	P_{12}	92,3998
P_3	294,4846	P_8	159,7331	P_{13}	92,3999
P_4	159,7331	P_9	159,7329	$\sum_{j=1}^{13} P_j =$	2520,00
P_5	159,7331	P_{10}	77,3998	Cost =	24164,05

5.2. Resultados para o Segundo Estudo de Caso - 13 Geradores e Demanda de 1800 MW

O segundo estudo de caso considera o mesmo conjunto de 13 geradores do estudo de caso anterior, mas com demanda de 1800 MW [Coelho and Mariani 2006]. Este problema do despacho econômico de energia possui vários mínimos locais e seu mínimo global é difícil de ser determinado [Coelho et al. 2009].

A Tabela 3 apresenta a comparação dos resultados obtidos pelos algoritmos propostos neste trabalho (ED_{AP} e ED_{PM}) com abordagens recentes da literatura. Neste caso é possível observar que os algoritmos propostos obtiveram melhores valores para o custo mínimo (17960,36) do que as abordagens de comparação. A versão ED_{PM} obteve o melhor custo médio (17960,54) enquanto o custo médio da versão ED_{AP} foi melhor do que dos algoritmos reportados na literatura. No quesito custo máximo as abordagens FAPSO-VDE e DPO obtiveram o melhor valor (17963,83). Destaca-se que o custo médio alcançado pelas abordagens propostas é inferior ao custo mínimo reportado por todas as demais abordagens e, portanto, é possível afirmar que as abordagens propostas fazem parte do estado-da-arte para este estudo de caso. A Tabela 4 apresenta o melhor resultado obtido pelos algoritmos propostos codificado por um vetor P_i , $i = 1, \dots, 13$ com custo mínimo de 17960,36.

5.3. Resultados para o Terceiro Estudo de Caso - 40 Geradores e Demanda de 10500 MW

O Terceiro Estudo de Caso considera 40 geradores com efeito de ponto válvula e demanda de 10500 MW [Ghorbani and Babaei 2016]. O espaço de soluções deste estudo de caso contém múltiplos ótimos locais [Sayah and Hamouda 2013].

Tabela 3. Resultados para o Segundo Estudo de Caso. Custos em US\$

Caso com 13 geradores e demanda de 1800 MW				
Método	Custo Mín	Custo Médio	Custo Máx	Desvio
ED_{AP}	17960,36	17961,26	17969,34	2,72
ED_{PM}	17960,36	17960,54	17969,34	1,27
DEC-SQP [Coelho and Mariani 2006]	17963,94	17973,13	17984,81	1,97
EP-SQP [Victoire and Jeyakumar 2004]	17991,03	18106,93	-	-
HQPSO5 [Coelho and Mariani 2008]	17963,95	18273,86	18633,04	123,22
HDE [Wang et al. 2007]	17975,73	18134,80	-	-
ST-HDE [Wang et al. 2007]	17963,89	18046,38	-	-
DE [Noman and Iba 2008]	17963,83	17965,48	17975,36	-
CDEM [Coelho et al. 2009]	17961,94	17974,68	18061,41	20,30
GSA [Swain et al. 2012]	17963,84	18041,21	18910,31	-
FAPSO-VDE [Mohammadi-Ivatloo et al. 2012]	17963,82	17963,82	17963,83	-
FA [Yang et al. 2012]	17963,83	18029,16	18168,80	148,54
θ -PSO [Hosseinnezhad and Babaei 2013]	17963,82	17965,20	17980,20	4,38
SSA [Yu and Li 2015]	17963,76	-	-	-
DPO [Han et al. 2015]	17963,82	17963,82	17963,83	-
TLBO [Banerjee et al. 2015]	18115,00	-	-	-
ORCSA [Nguyen and Vo 2015]	17963,83	-	-	-

Tabela 4. Custo Mínimo para o 2º Estudo de Caso utilizando DE_{AP} e DE_{PM}.

Potência	Geração	Potência	Geração	Potência	Geração
P_1	628,3185	P_6	60	P_{11}	40
P_2	149,5996	P_7	109,8665	P_{12}	55
P_3	222,7493	P_8	109,8665	P_{13}	55
P_4	109,8666	P_9	109,8665	$\sum_{j=1}^{13} P_j =$	1800,00
P_5	109,8665	P_{10}	40	Custo =	17960,36

A Tabela 5 apresenta a comparação entre os resultados obtidos pelo ED_{AP} e ED_{PM} e resultados reportados na literatura recente. O algoritmo ED_{AP} obteve o melhor valor para o custo mínimo (121406,91) enquanto o ED_{PM} obteve os melhores valores para o custo médio (121411,92), custo máximo (121412,12) e desvio padrão (0,07). Destaca-se que o custo mínimo (121411,89) e o custo médio (121411,92) obtidos pelo ED_{PM} são inferiores ao menor custo mínimo reportado pelos algoritmos de comparação (121412,53, obtido pelo EMA e ORCSA). Portanto, é possível afirmar que o ED_{PM} é um algoritmo do estado-da-arte. A Tabela 6 demonstra o melhor resultado obtido com o ED_{AP} codificado por um vetor P_i , $i = 1, \dots, 40$ com custo igual a 121406,91.

6. Conclusões e Trabalhos Futuros

Neste artigo foram investigadas técnicas de seleção adaptativa de operadores da Evolução Diferencial para resolver o problema do Despacho Econômico de Energia Elétrica. As técnicas investigadas foram a *Adaptive Pursuit* e a *Probability Matching* e o conjunto de operadores foi composto pela ED/rand/1/bin, ED/rand/2/bin, ED/current-to-rand/1 e ED/current-to-rand/2.

De acordo com os resultados obtidos, os dois métodos implementados (ED_{AP} e ED_{PM}) tiveram bom desempenho nos três casos de teste considerados, obtendo melhores valores que os algoritmos reportados na literatura para o custo mínimo, custo médio e custo máximo em todos os estudos de caso (com exceção do segundo estudo de caso no custo máximo, onde FAPSO-VDE e DPO foram melhores). Destaca-se que o custo médio obtido pelo algoritmo ED_{PM} em todos os casos é inferior ao menor custo mínimo reportado na literatura e, portanto, é possível concluir que o ED_{PM} é um algoritmo do estado-

Tabela 5. Resultados para o Terceiro Estudo de Caso. Custos em US\$

Caso com 40 geradores e demanda de 10500 MW				
Método	Custo Mín	Custo Médio	Custo Máx	Desvio
ED_{AP}	121406,91	121414,43	121417,51	2,37
ED_{PM}	121411,89	121411,92	121412,12	0,07
CCDE [Ghasemi et al. 2015]	121412,54	121412,99	121414,69	0,09
EMA [Ghorbani and Babaei 2016]	121412,53	121417,13	121426,15	-
HCRO-DE [Roy et al. 2014]	121412,55	121413,11	121415,68	0,13
SQPSO [Hosseinnezhad et al. 2014]	121412,57	121455,70	121709,56	-
ORCSA [Nguyen and Vo 2015]	121412,53	121472,45	121596,18	0,49
FAPSO-VDE [Mohammadi-Ivatloo et al. 2012]	121412,56	121412,61	121412,78	-
SSA [Yu and Li 2015]	121412,55	-	-	-
QPSO [Meng et al. 2010]	121448,21	122225,07	-	-
NDS [Lin et al. 2011]	121647,39	121647,39	121647,39	-
FA [Yang et al. 2012]	121415,05	121416,57	121424,56	1,78
QGS [Moradi-Dalvand et al. 2012]	121412,55	121423,52	121438,68	-
θ -PSO [Hosseinnezhad and Babaei 2013]	121420,90	121509,84	121852,42	92,39
DEPSO [Sayah and Hamouda 2013]	121412,56	121419,31	121468,25	-
DEC-SQP [Coelho and Mariani 2006]	121741,97	122295,12	122839,29	386,18
NPSO-LRS [Selvakumar and Thanushkodi 2007]	121664,43	122209,31	122981,59	-
CEP-PSO [Sinha and Purkayastha 2004]	123670,00	124145,60	124900,00	-
PSO [Victoire and Jeyakumar 2004]	123930,45	124154,49	-	-
PSO-SQP [Victoire and Jeyakumar 2004]	122094,67	122245,25	-	-
HDE [Wang et al. 2007]	121813,26	122705,66	-	-
ST-HDE [Wang et al. 2007]	121698,51	122304,30	-	-
DE [Noman and Iba 2008]	121416,29	121422,72	121431,47	-
CDMD [Coelho et al. 2009]	121423,40	121526,73	121696,98	54,86
CSO [Selvakumar and Thanushkodi 2009]	121461,67	121936,19	122844,53	-
BBO [Bhattacharya and Chattopadhyay 2010]	121426,95	121508,03	121688,66	-
ICA-PSO [Vlachogiannis and Lee 2009]	121413,20	121428,14	121453,56	-

da-arte para o Problema do Despacho Econômico de Energia Elétrica. Na comparação entre os algoritmos propostos é possível dizer que o ED_{PM} foi melhor que o ED_{AP}. O bom desempenho dos algoritmos propostos se deve principalmente à capacidade dos algoritmos propostos em escolher o melhor operador para cada estágio da evolução.

Em trabalhos futuros pretende-se analisar a auto-adaptação dos parâmetros da Evolução Diferencial (F e CR) em conjunto com a seleção adaptativa dos operadores.

Referências

- Balamurugan, R. and Subramanian, S. (2007). Self-adaptive differential evolution based power economic dispatch of generators with valve-point effects and multiple fuel options. *International Journal of Electrical, Electronic Science and Engineering*.
- Banerjee, S., Maity, D., and Kumar, C. (2015). Teaching learning based optimization for economic load dispatch problem considering valve point loading effect. *International Journal of Electrical Power and Energy Systems*, 73:456–464.
- Basu, M. (2002). Fuel constrained economic emission load dispatch using hopfield neural networks. *Electric Power Systems Research*.
- Bhattacharya, A. and Chattopadhyay, P. K. (2010). Biogeography-based optimization for different economic load dispatch problems. *IEEE Transactions on Power Systems*.
- Coelho, L. and Mariani, V. (2008). Particle swarm approach based on quantum mechanics and harmonic oscillator potential well for economic load dispatch with valve-point effects. *Energy Conversion and Management*, 49(11):1373–1378.

Tabela 6. Custo Mínimo para o 3º Estudo de Caso utilizando DE_{AP}.

Potência	Geração	Potência	Geração
P_1	110,8087	P_{21}	523,2795
P_2	110,8242	P_{22}	523,2880
P_3	97,4105	P_{23}	523,2821
P_4	179,7335	P_{24}	523,2831
P_5	87,6980	P_{25}	523,2890
P_6	139,9999	P_{26}	523,2747
P_7	259,6546	P_{27}	10,0027
P_8	284,6115	P_{28}	10,0000
P_9	284,6543	P_{29}	10,0000
P_{10}	130,0021	P_{30}	87,8369
P_{11}	94,0000	P_{31}	189,9998
P_{12}	94,0000	P_{32}	189,9994
P_{13}	214,7616	P_{33}	189,9996
P_{14}	394,2789	P_{34}	164,7942
P_{15}	394,2789	P_{35}	199,9983
P_{16}	394,2880	P_{36}	194,2579
P_{17}	489,2835	P_{37}	110,0000
P_{18}	489,2768	P_{38}	110,0000
P_{19}	511,2856	P_{39}	110,0000
P_{20}	511,2810	P_{40}	511,2817
$\sum_{j=1}^{40} P_i =$	10500	Custo =	121406,62

- Coelho, L. S. and Mariani, V. C. (2006). Combining of chaotic differential evolution and quadratic programming for economic dispatch optimization with valve-point effect. *IEEE Transactions on Power Systems*, 21(2):989–996.
- Coelho, L. S., Souza, R. T. T., and Mariani, V. C. (2009). Improved differential evolution approach based on cultural algorithms and diversity measure applied to solve economic load dispatch problems. *Mathematics and Computers in Simulation*, 79:3136–3147.
- Engelbrecht, A. P. (2007). *Computational Intelligence: An Introduction*. Wiley Publishing, 2nd edition.
- Ghasemi, M., Taghizadeh, M., Ghavidel, S., and Abbasian, A. (2015). Colonial competitive differential evolution: An experimental study for optimal economic load dispatch. *Applied Soft Computing*, 40:342–363.
- Ghorbani, N. and Babaei, E. (2016). Exchange market algorithm for economic load dispatch. *International Journal of Electrical Power and Energy Systems*, 75:19–27.
- Goldberg, D. E. (1990). Probability matching, the magnitude of reinforcement, and classifier system bidding. *Mach. Learn.*, 5:407–425.
- Gong, W., Álvaro Fialho, Cai, Z., and Li, H. (2011). Adaptive strategy selection in differential evolution for numerical optimization: An empirical study. *Inform. Sciences*, 181(24):5364 – 5386.
- Gong, W., Fialho, A., and Cai, Z. (2010). Adaptive strategy selection in differential evolution. In *Conference on Genetic and Evolutionary Computation*, pages 409–416.
- Han, L., Romero, C. E., and Yao, Z. (2015). Economic dispatch optimization algorithm based on particle diffusion. *Energy Conversion and Management*, 105:1251–1260.
- Hosseinnezhad, V. and Babaei, E. (2013). Economic load dispatch using θ -pso. *Energy Conversion and Management*, 49:160 – 169.

- Hosseinnezhad, V., Rafiee, M., Ahmadian, M., and Ameli, M. T. (2014). Species-based Quantum Particle Swarm Optimization for economic load dispatch. *International Journal of Electrical Power & Energy Systems*, 63:311–322.
- Khamsawang, S., Pothiya, S., and Boonseng, S. (2004). Distributed tabu search algorithm for solving the economic dispatch problem. In *TENCON*, pages 484–487, Chiang Mai, Thailand. TENCON.
- Li, K., Fialho, A., and Kwong, S. (2011). Multi-objective differential evolution with adaptive control of parameters and operators. In *International Conference on Learning and Intelligent Optimization*, pages 473–487.
- Lin, W.-M., Gow, H.-J., and Tsai, M.-T. (2011). Combining of direct search and signal-to-noise ratio for economic dispatch optimization. *Energy Conversion and Management*, 52(1):487 – 493.
- Mallipeddi, R., Suganthan, P., Pan, Q., and Tasgetiren, M. (2011). Differential evolution algorithm with ensemble of parameters and mutation strategies. *Appl. Soft Comput.*, 11(2):1679–1696.
- Meng, K., Wang, H. G., Dong, Z., and Wong, K. P. (2010). Quantum-inspired particle swarm optimization for valve-point economic load dispatch. *IEEE Transactions on Power Systems*, 25(1):215–222.
- Mohammadi-Ivatloo, B., Rabiee, A., Soroudi, A., and Ehsan, M. (2012). Iteration {PSO} with time varying acceleration coefficients for solving non-convex economic dispatch problems. *International Journal of Electrical Power & Energy Systems*, 42(1):508 – 516.
- Moradi-Dalvand, M., Mohammadi-Ivatloo, B., Najafi, A., and Rabiee, A. (2012). Continuous quick group search optimizer for solving non-convex economic dispatch problems. *Electric Power Systems Research*, 93(0):93 – 105.
- Nguyen, T. T. and Vo, D. N. (2015). The application of one rank cuckoo search algorithm for solving economic load dispatch problems. *Applied Soft Computing Journal*, 37:763–773.
- Noman, N. and Iba, H. (2008). Differential evolution for economic load dispatch problems. *Electric Power Systems Research*, 78(3):1322–1331.
- Park, J., Jeong, Y., and Lee, W. (2006). An improved particle swarm optimization for economic dispatch problems with non-smooth cost functions. In *IEEE Power Engineering Society General Meeting*, Montreal, Que.
- Qin, A. K., Huang, V. L., and Suganthan, P. N. (2009). Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Trans. Evol. Comput.*, 13:398–417.
- Qin, A. K. and Suganthan, P. N. (2005). Self-adaptive differential evolution algorithm for numerical optimization. In *IEEE Congress on Evolutionary Computation*, pages 1785–1791.
- Roy, P. K., Bhui, S., and Paul, C. (2014). Solution of economic load dispatch using hybrid chemical reaction optimization approach. *Applied Soft Computing*, 24:109–125.

- Sayah, S. and Hamouda, A. (2013). A hybrid differential evolution algorithm based on particle swarm optimization for nonconvex economic dispatch problems. *Applied Soft Computing*, 13(4):1608 – 1619.
- Selvakumar, A. and Thanushkodi, K. (2009). Optimization using civilized swarm: solution to economic dispatch with multiple minima. *Electric Power Systems Research*, 79(1):8–16.
- Selvakumar, A. I. and Thanushkodi, K. (2007). New particle swarm optimization solution to nonconvex economic dispatch problems. *IEEE Transactions on Power Systems*, 22(1):42–51.
- Sinha, N. and Purkayastha, B. (2004). Pso embedded evolutionary programming technique for nonconvex economic load dispatch. In *IEEE PES - Power Systems Conference and Exposition*, pages 66–71, Brasil. IEEE PES.
- Storn, R. and Price, K. (1995). Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces.
- Storn, R. and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.*, 11(4):341–359.
- Swain, R., Sahu, N., and Hota, P. (2012). Gravitational search algorithm for optimal economic dispatch. *Procedia Technology*, 6(0):411 – 419.
- Thierens, D. (2005). An adaptive pursuit strategy for allocating operator probabilities. In *Conference on Genetic and Evolutionary Computation*, pages 1539–1546.
- Victoire, T. and Jeyakumar, A. (2004). Hybrid pso-sqp for economic dispatch with valve-point effect. *Electric Power Systems Research*, 71(1):51–59.
- Vlachogiannis, J. G. and Lee, K. Y. (2009). Economic load dispatch - a comparative study on heuristic optimization techniques with an improved coordinated aggregation-based pso. *IEEE Transactions on Power Systems*, 24(2):991–1001.
- Wang, S. K., Chiou, J. P., and Liu, C. W. (2007). Non-smooth/nonconvex economic dispatch by a novel hybrid differential evolution algorithm. *IET Generation, Transmission & Distribution*, 1(5):793–803.
- Yang, X.-S., Hosseini, S. S. S., and Gandomi, A. H. (2012). Firefly algorithm for solving non-convex economic dispatch problems with valve loading effect. *Applied Soft Computing*, 12(3):1180 – 1186.
- Yu, J. J. Q. and Li, V. O. K. (2015). A social spider algorithm for solving the non-convex economic load dispatch problem. *Neurocomputing*, 171:1–11.
- Zhang, J. and Sanderson, A. C. (2009). JADE: adaptive differential evolution with optional external archive. *IEEE Trans. Evol. Comput.*, 13:945–958.
- Zhao, S.-Z., Suganthan, P. N., and Zhang, Q. (2012). Decomposition-based multiobjective evolutionary algorithm with an ensemble of neighborhood sizes. *IEEE Trans. Evol. Comput.*, 16(3):442–446.

Arquitetura para Fog Computing em Sistemas de Middleware para Internet das Coisas

Matheus Crespi Schenfeld¹, Leonardo Amaral¹, Everton de Matos¹, Fabiano Hessel¹

¹Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

{matheus.schenfeld, leonardo.amaral, everton.matos.001}@acad.pucrs.br,

{fabiano.hessel}@pucrs.br

Resumo. *Internet das Coisas (IoT) é considerada uma evolução computacional que preconiza a existência de uma infinidade de objetos físicos embarcados com sensores e atuadores, conectados por redes sem fio e que se comunicam através da Internet. Diversas pesquisas vêm sendo realizadas tentando mitigar os desafios existentes na IoT. Assim, cloud computing ganha espaço nesses cenários. Cloud computing refere-se ao uso de memória, armazenamento e processamento de recursos compartilhados, interligados pela Internet. No entanto, cloud computing trouxe problemas para aplicações IoT sensíveis à latência de comunicação. Para tentar minimizar esse problema, foi introduzido o conceito de fog computing, cuja ideia principal é a de distribuir serviços nos dispositivos de computação localizados nas extremidades da rede. Um grande desafio de fog computing na IoT é a definição de uma arquitetura de sistema que possa ser usada em diferentes domínios de aplicação, como saúde, cidades inteligentes entre outros. Esse trabalho apresenta uma arquitetura de sistema para dispositivos IoT capaz de habilitar o processamento de dados nos próprios dispositivos ou o mais próximo deles, em diferentes domínios, melhorando a Qualidade dos Serviços (QoS) e autonomia na tomada de decisão, mesmo se os dispositivos estiverem temporariamente desconectados da rede (offline). Apresentamos também uma breve revisão de trabalhos relacionados.*

Abstract. *Internet of Things (IoT) is considered a computational evolution that advocates the existence of a infinity of physical objects embedded with sensors and actuators, connected by the wireless networks and communicating through the Internet. Several researches have been made trying to mitigate the existing IoT challenges. Thus, cloud computing gains space in these scenarios. Cloud computing refers to the use of memory, storage and processing as shared resources, interconnected by the Internet. However, cloud computing brought problems for communication latency sensitive IoT applications. Trying to solve this problem, fog computing was introduced, whose main idea is to distribute services in the computing devices located at the ends of the network. One big challenge of fog computing in IoT is a definition of a system architecture that could be used in different domains, such as healthcare, smart cities and others. This paper presents a system architecture for IoT devices able to allow data processing on the devices themselves or as close to them, in different domains, improving Quality of Services (QoS) and autonomous decision-making, even if the devices are temporarily disconnected from the network (offline). We also present a brief review of related work.*

1. Introdução

O termo Internet das Coisas (IoT) foi introduzido em 1998 [Atzori et al. 2010] e definido como o paradigma de computação que permite que pessoas e coisas (dispositivos de computação) possam ser conectados a qualquer hora, em qualquer lugar, com qualquer coisa e qualquer um, usando qualquer caminho de rede ou serviço [Perera et al. 2014]. Conectividade entre os dispositivos é uma funcionalidade crítica que é necessária para cumprir a visão da IoT. Nesse sentido, existem estatísticas atuais de mercado e previsões que mostram um rápido crescimento em implementações de dispositivos relacionados a ambientes IoT.

A visão da IoT está em todos os lugares e pode ser aplicada em diversas áreas. Portanto, ela é fundamental no cenário brasileiro para alavancar soluções relacionadas aos grandes desafios da computação do país. Olhando para o futuro, com a evolução da tecnologia, a quantidade de dispositivos IoT só vai crescer. Em 2020, estima-se que haverá de 50 a 100 milhões de dispositivos conectados à Internet [Sundmaeker et al. 2010]. Nesse sentido, a seleção e a configuração manual desses dispositivos, juntamente com a escolha de quais os dispositivos que são relevantes para os usuários e aplicativos, não são consideradas tarefas triviais. [Perera et al. 2014].

Diversas pesquisas e avanços têm sido feitos tanto no meio acadêmico quanto no da indústria para tentar mitigar os desafios existentes na IoT. O objetivo desse esforço é tentar criar um ambiente onde todos os dispositivos a nossa volta estejam conectados à Internet, além de facilitar a comunicação entre eles e prover interações com menor intervenção humana possível.

O conceito de *cloud computing* surge da necessidade de, primeiro, colocar sistemas computacionais em ambientes altamente distribuídos e escaláveis em termos de desempenho e utilização de recursos, e segundo, de tornar simples o ambiente de configuração e reúso dos recursos. Esse termo refere-se à utilização da memória e da capacidade de armazenamento e cálculo de computadores e servidores compartilhados e interligados por meio da Internet [Patidar et al. 2011]. Entretanto, *cloud computing* trouxe problemas para sistemas IoT e aplicações sensíveis a latência de comunicação, tais como, aplicações de segurança crítica, aplicações militares, sistemas médicos emergenciais, entre outros. Além disso, aplicações consumidoras de informações e dispositivos geradores de dados estão em ambientes altamente distribuídos, o que resulta na necessidade de sistemas que melhor gerenciem a alta demanda de dados, mas com uma menor latência na comunicação.

Para tentar solucionar esse problema surge então o conceito de *fog computing*, cuja ideia principal é distribuir serviços e funções para os dispositivos computacionais (como as coisas ou dispositivos da IoT) localizados nas extremidades ou bordas da rede [Bonomi et al. 2012]. *Fog computing* tem como uma de suas principais características, criar uma camada federada, ou seja, virtual, que tem como finalidade aproximar de forma eficiente as camadas do sistema, provedoras e consumidoras de informação, e com isso, tentar minimizar o grande volume de comunicação.

Processar informação o mais perto possível dos dispositivos IoT pode trazer vantagens às aplicações requisitantes, pois ajuda a reduzir o volume de comunicação e a carga de trabalho nos nodos da rede que processam e produzem informação (como é

o caso dos sistemas de *middleware* IoT), além de também reduzir o tráfego na rede, latência, utilização de energia, e por fim melhorar a Qualidade do Serviço (QoS) provido às aplicações e usuários da IoT.

Nesse artigo é proposto o desenvolvimento de uma arquitetura de sistema para dispositivos IoT que seja capaz de habilitar o processamento de dados nos próprios dispositivos ou o mais perto destes, provendo uma melhor Qualidade do Serviço (QoS) provido e mais autonomia na tomada de decisões, mesmo que os dispositivos estejam momentaneamente desconectados (*offline*), baseando-se em regras pré-estabelecidas e controladas pelo processamento de eventos complexos (CEP).

O restante desse artigo está organizado como a seguir. Na Seção 2 são apresentadas algumas definições importantes, tais como IoT, *cloud computing* e *fog computing*. A proposta de arquitetura para *fog computing* é apresentada na Seção 3. Os trabalhos relacionados e uma breve discussão sobre a arquitetura proposta são apresentados na Seção 4. Por fim são apresentados a conclusão e os trabalhos futuros na Seção 5.

2. Referencial Teórico

Nessa seção são brevemente apresentados alguns conceitos comumente citados no decorrer do trabalho. Em 2.1 é apresentada a definição de Internet das Coisas. Na subseção 2.2 é apresentado o conceito de *cloud computing*. Por fim, a subseção 2.3 apresenta a definição de *fog computing*.

2.1. Internet das Coisas

A criação da internet revolucionou o funcionamento tradicional das sociedades modernas, e com o uso de protocolos de endereçamento, tornou-se possível a comunicação entre dispositivos separados fisicamente, rompendo diversos paradigmas da distância e do tempo. A possibilidade de quaisquer dispositivos - como, Identificação por Radio Frequência (RFID), tags, sensores, atuadores, smartphones, entre outros - conectados à rede, em qualquer lugar e a qualquer hora, iniciou a era da ubiquidade, principal característica e pilar fundamental para o surgimento da Internet das Coisas ou em inglês, Internet of Things (IoT) [Tiburski et al. 2015].

O conceito da IoT é uma evolução de diversos elementos tecnológicos (sensores, *hardware*, semântica, armazenamento, processamento, comunicação, entre outros), que ao serem unidos em um mesmo ambiente, representam o futuro da computação e das comunicações [Tan 2010]. A IoT tem o objetivo final de criar um mundo melhor para todas as pessoas, onde objetos a nossa volta tenham conhecimento e autonomia, agindo de forma inteligente sem instruções explícitas [Perera et al. 2014].

Uma das funcionalidades essenciais para o funcionamento de um sistema IoT é a camada de *middleware* [Atzori et al. 2010]. Dentre as suas características, a principal é a de abstrair os protocolos técnicos, tais como, protocolo de comunicação, tipo de porta de comunicação, entre outras, das diferentes tecnologias encontradas nos dispositivos, tornando assim a tarefa do programador de aplicações algo mais fácil e prazeroso. O *middleware* tem ganhado muito mais importância nos últimos anos, visto que facilita o desenvolvimento de novos serviços e a integração de tecnologias já existentes com as novas em desenvolvimento [Amaral et al. 2015]. Isto tira a necessidade do programador

de ter o conhecimento específico dos dispositivos, que atualmente têm variados tipos de protocolos de comunicação e funcionalidades [Schenfeld et al. 2014].

Entretanto, a possibilidade de poder estar conectado a diversos dispositivos e independente de localização, assim como prover para o desenvolvedor de sistemas a chance de poder utilizar recursos que facilitam o ambiente de configuração, adaptação e monitoramento de serviços, motivou a utilização de tecnologia de *cloud computing* para melhorar ainda mais a IoT.

2.2. Cloud Computing

Cloud Computing ou Computação em Nuvem, é o resultado da evolução e adoção de tecnologias e paradigmas computacionais consolidados [Zhou et al. 2010]. A principal tecnologia que permite a computação em nuvem é a virtualização, que separa um dispositivo físico de computação em um ou mais dispositivos virtuais, fazendo com que cada dispositivo possa ser facilmente utilizado e gerenciado para executar tarefas. Com a virtualização em nível de sistema operacional, a criação de um sistema escalável de vários dispositivos de computação independentes, além de recursos computacionais ociosos, podem ser atribuídos e utilizados de forma mais eficiente. O objetivo da computação em nuvem é permitir que os usuários tirem benefício de todas essas tecnologias, sem a necessidade de conhecimento profundo ou experiência com cada uma delas [Hamdaqa and Tahvildari 2012].

Entretanto, mesmo com diversas vantagens, *cloud computing* tem apresentado algumas limitações [Patidar et al. 2011], fazendo com que aplicações sensíveis a latência de comunicação sejam afetadas, visto que aplicações consumidoras de informações e dispositivos geradores de dados estão em ambientes distribuídos. Problemas com largura de banda e conseqüentemente os custos decorrentes disso também fazem parte das limitações. Para solucionar estes e outros problemas surge o conceito de *fog computing* [Bonomi et al. 2012].

2.3. Fog Computing

Fog Computing surge como uma infraestrutura de computação na qual aplicações e serviços podem ser tratados tanto em servidores *cloud*, como na própria rede [Yi et al. 2015]. O conceito *fog* surge visando atender a três objetivos principais:

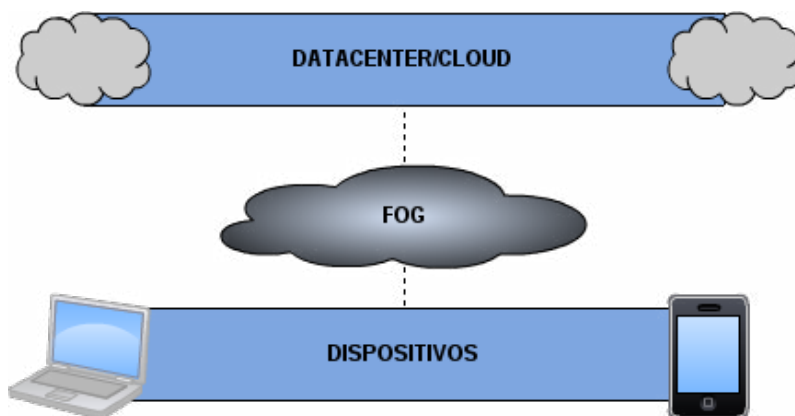


Figura 1. Camada virtual de fog computing.

- Melhorar a eficiência e reduzir a quantidade de dados que precisam ser transmitidos para que seja feito o processamento, análise e armazenamento.
- Aproximar o consumidor de informações com o provedor de dados.
- Prover segurança e conformidade para a transmissão de dados.

Em *fog computing*, grande parte do processamento ocorre nos próprios dispositivos que geram os dados, tendo em vista que nos últimos anos o seu poder de processamento e armazenamento teve significativa evolução. A arquitetura *fog* cria uma plataforma virtual que fornece serviços de processamento e de armazenamento entre a nuvem *cloud* e os dispositivos. Em outras palavras, cria uma camada federada, isto é, a união de diversas funcionalidades que formam uma camada virtual em ambiente descentralizado e mais próximo dos dispositivos de borda (ou *Edge Devices*), reduzindo a latência na rede e largura de banda, solucionando ao menos parcialmente os problemas encontrados em *cloud computing* [Lewis et al. 2014]. A Figura 1 ilustra a camada virtual criada em *fog computing* aproximando os dispositivos e a camada *cloud computing*.

3. Arquitetura do Sistema

Tendo como objetivo solucionar o problema exposto anteriormente, além de atender algumas das tendências futuras para sistemas IoT descritos por L. Tan et al. [Tan 2010] - dispositivos com inteligência; arquitetura dirigida a eventos; grande quantidade de dispositivos conectados ao sistema, bem como o volume de dados produzidos por eles - é proposto nesse artigo o planejamento e desenvolvimento de uma infraestrutura com suporte para *fog computing* em arquiteturas de *middleware* Orientados a Serviços (SOA) que provê mais autonomia para os dispositivos.

A arquitetura proposta, que pode ser vista na Figura 2, é constituída por cinco camadas: Camada de Aplicações, responsável por entregar os dados processados; Camada de *Middleware*, que executa as principais funções da arquitetura; Camada de *Fog Computing*, elemento principal da pesquisa desse trabalho; Camada de Comunicação, responsável pelas conexões entre diferentes plataformas, e por fim a Camada das Coisas, ou em outras palavras, camada dos dispositivos heterogêneos.

3.1. Camada de Fog Computing

Iniciamos a descrição da arquitetura pela camada de *fog computing*, visto que é o principal elemento do trabalho. A camada de *fog* tem como objetivo criar uma camada federada, ou seja, virtual, que tem como principal função aproximar de forma eficiente as camadas de aplicações e as camadas de dispositivos. Na arquitetura proposta, essa camada virtual será alocada em ambiente de *System-on-a-Chip* (SoC), um sistema computacional completo num único chip. Tendo como justificativa o grande poder de processamento, bem como a variedade de tipos de conexão, tais como, WiFi, *Bluetooth*, *Ethernet*, serial, entre outras, cada SoC atuará como um *Edge-Gateway*, que consiste em um concentrador de processamento de um ou mais dispositivos. Por exemplo, um *Edge-Gateway* de uma *SmartHome*, que pode estar conectado com o sistema de iluminação, eletroeletrônicos, carros, entre outros. Os dispositivos por sua vez, não mais se comunicam com o *middleware* central, mas sim com o *Edge-Gateway*, criando um ambiente virtual de *fog computing*.

3.2. Camada das Coisas

Essa camada consiste em sensores, sistemas embarcados, RFID tags e quaisquer outros objetos que possam estar conectados pela rede. Essas entidades são as que capturam os dados do ambiente.

3.3. Camada de Comunicação

A camada de comunicação conecta diferentes tipos de fontes de dados e serviços localizados em ambiente altamente distribuído. O primeiro estágio de tratamento dos dados coletados acontece nessa camada. Responsável pelas mensagens de roteamento, serviços *Publish/Subscribe*, e também por realizar a comunicação entre plataformas, se necessário.

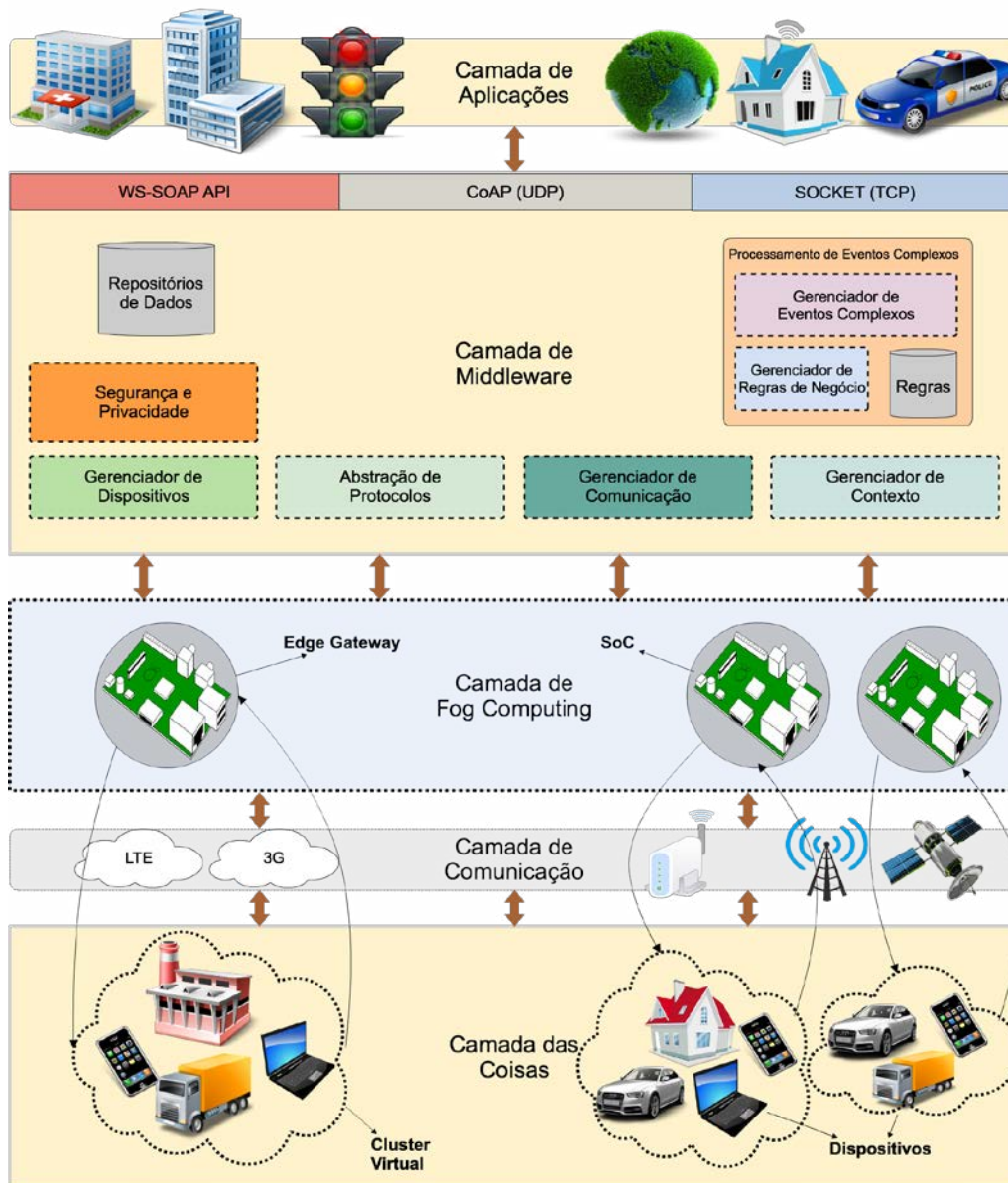


Figura 2. Arquitetura de fog computing

3.4. Camada de Middleware

Essa é a camada responsável por realizar as principais funções do sistema e opera em modo bidirecional. Ela age como uma interface entre a camada de *hardware* e a camada das aplicações e é responsável por funções essenciais, tais como, gerenciamento de dispositivos, gerenciamento de informações, filtragem dos dados, agregação de valores, análise de semântica, controle de acesso, além da abstração dos protocolos de acesso dos diversos dispositivos da camada de dispositivos. Essas funções são divididas em módulos, listados a seguir:

- **Abstração de protocolos:** Tendo em vista que os dispositivos IoT conectados à rede são heterogêneos, é esperado que cada um tenha características próprias, como protocolos de conexões, tipos de dados, ou linguagem de comunicação. É necessário que exista uma abstração desses protocolos para tornar o acesso aos dispositivos mais harmônico.
- **Gerenciador de dispositivos:** Módulo responsável por prover suporte à conexão de novos dispositivos através de informações de protocolos de comunicação e conexão de dispositivos já conhecidos.
- **Gerenciador de comunicação:** Módulo responsável por administrar as comunicações entre as camadas consumidoras e provedoras de serviços. Responsável também por definir o melhor tipo de protocolo de aplicação para ser usado. Exemplos de protocolos são: *Socket* (TCP), WS-SOAP (API) e o *Constrained Application Protocol* (CoAP).
- **Gerenciador de contexto:** O contexto é qualquer informação que possa ser extraída de uma entidade a fim de caracterizá-la. Essa entidade pode ser um usuário, um dispositivo ou um nodo *fog*. O contexto pode ser usado para prover serviços de informações relacionados a entidades. O módulo gerenciador de contexto é responsável por armazenar informações de entidades e também realizar procedimentos com essas informações (raciocínio de contexto), e assim obter informações contextualizadas.
- **Segurança e privacidade:** É essencial a presença de mecanismos de segurança em ambientes IoT. Muitas vezes as informações que trafegam em redes IoT são sigilosas, como por exemplo informações médicas ou informações pessoais. O *middleware* para IoT precisa ter funções de segurança que protejam os dados trafegados e também que não permita a alteração dos mesmos.
- **Processamento de Eventos Complexos:** O módulo para processamento de eventos complexos (CEP) opera através da avaliação da confluência de eventos, e consequentemente, da execução de uma ação. Os eventos podem ocorrer por um longo período de tempo. A correlação de eventos pode ser ocasional, temporal ou espacial. CEP requer o emprego de sofisticados intérpretes de eventos, detecção de padrões de acontecimento, e técnicas de correlação. CEP é comumente usado para detectar e responder a anomalias de negócios, ameaças e oportunidades. Geralmente, motores de ontologias e regras são usados para fornecer vocabulários comuns de um domínio e conhecimento comportamental, expressando restrições e reação a eventos [Laliwala and Chaudhary 2008].

4. Trabalhos Relacionados e Discussão

No âmbito da Internet das Coisas, diversas plataformas de *cloud computing* podem ser utilizadas para facilitar o desenvolvimento e a implantação de sistemas e aplicações para IoT. Exemplos conhecidos dessas plataformas são: ThingWorx, OpenIoT, Xively, Nimbits, Axeda, entre outras. A Tabela 1 resume algumas das características mais importantes destas e de outras plataformas de *cloud computing* para IoT (na tabela, "+" significa suporta e "-" significa não suporta).

As métricas de avaliação utilizadas na Tabela 1 incluem: suporte a *Edge-Gateway*, fazendo uma ponte entre a camada de *fog* e camada de *middleware*. Gerenciamento, que diz respeito ao suporte à descoberta de dispositivos, entrega das informações, configuração e ativação de aplicações e serviços, garantindo a autonomia. Suporte a qualidade e confiabilidade de aplicações e serviços (Veracidade). E por último, o suporte de protocolos de comunicação padronizados.

Xively, por exemplo, representa uma das primeiras plataformas para IoT disponíveis na web. Xively tem como objetivo conectar os dispositivos com as aplicações, garantindo a segurança em tempo real. Xively fornece uma Plataforma como Serviço (PaaS) para os desenvolvedores de aplicações IoT, além dos prestadores de serviços. Ela é capaz de integrar os dispositivos com a plataforma através do uso de bibliotecas prontas (como ARM mbed, *Electric Imp* e iOS/OSX) e também de facilitar a comunicação via HTTP(S), ou *Sockets/WebSocket* [Yang et al. 2013]. É possível também integrar Xively com outras plataformas que utilizam bibliotecas Java, JS, Python e Ruby.

Como outro exemplo, Nimbits é uma Plataforma como Serviço (PaaS) de código aberto que conecta dispositivos inteligentes com a nuvem [Doukas 2012]. Ele também realiza a análise de dados na nuvem, gerando alertas, além de se conectar com redes sociais e planilhas. Além disso, ele se conecta a sites e pode armazenar, compartilhar e recuperar dados dos sensores em diversos formatos, incluindo numérico, texto, GPS, JSON ou XML. A troca de dados ou mensagens XMPP é um serviço embutido no Nimbits. O núcleo do Nimbits é um servidor que fornece serviços Web REST para o registro e recuperação de dados brutos e processados.

Axeda é uma outra plataforma para integração de dados e desenvolvimento de aplicações, bem como para conexão de dispositivo e outros serviços comumente oferecidos em serviços de *cloud computing* [Bjelica et al. 2014]. Essa plataforma oferece suporte a *Edge-Gateway* desde que os dispositivos a serem conectados atendam a algumas especificações de Firewall. Axeda também dá suporte ao controle de eventos em tempo real. Além disso, ela também provê ao desenvolvedor uma interface de alto nível, tornando o ambiente de configuração e reutilização de recursos facilitado. A plataforma também dispõe de serviços e aplicações com funções de gerenciamento de dados, juntamente com aplicações web e APIs que tem como finalidade estender a plataforma usando Java SDK.

Como pode ser visto nos trabalhos relacionados da Tabela 1, mesmo que existam diversas soluções até mesmo comerciais, elas ainda não atendem por completo aos requisitos de uma arquitetura de *cloud e fog computing* explanados em Al-Fugaha et al. [Al-Fuqaha et al. 2015]. Exemplos desses requisitos (ou deficiências) são: nenhuma plataforma oferece suporte a CoAP, mesmo ele sendo uma padronização que é baseada em

Tabela 1. Plataformas Cloud IoT e Suas Características

Plataforma	Gateway	Gerenciamento	Veracidade	REST	CoAP
Arkessa	-	+	+	+	-
Axeda	+	+	+	+	-
LittleBits	-	-	-	+	-
NanoService	+	+	+	+	-
NinjaBlocks	+	-	-	+	-
Nimbits	-	-	-	+	-
RealTime.io	+	+	-	+	-
TempoDB	-	-	-	+	-
Thingworx	-	+	+	+	-
Xively	+	+	+	+	-

REST, porém com diversas melhorias. Em cinco plataformas, não há suporte para a conexão dos *edge-gateways*, o que impede que um conjunto de aplicações possa ser executada na plataforma, por exemplo, em alguns cenários de aplicações existe a necessidade de utilizar as capacidades de um *edge-gateways*, dentre elas, agregação, eliminação e filtragem dos dados, processamento em tempo real, análise de dados, entre outras. Além disso, também há pouco suporte no quesito de manter a qualidade e confiabilidade das informações.

A arquitetura de *cloud e fog computing* apresentada neste trabalho (Figura 2) identifica várias vantagens em comparação com trabalhos similares encontrados. Entre elas pode-se destacar a conexão com diversos tipos de equipamentos e dispositivos através do uso de diferentes redes de comunicação, como por exemplo, WiFi, *Bluetooth*, LTE, entre outras. A utilização do CoAP como uma das alternativas de protocolos de aplicações também pode ser citada, tendo em vista que representa um método simplificado para a transferência de dados entre servidores e consumidores através do HTTP. Diferente do REST, o CoAP está vinculado ao UDP (não ao TCP) por padrão, o que o torna mais viável para aplicações IoT. As capacidades da camada de *middleware* também tornam-se uma vantagem, pois, garantem a qualidade e confiabilidade do sistema. Entre as capacidades pode-se citar o suporte a descoberta, gerenciamento, configuração e adaptação, tanto de dispositivos como de contexto dos mesmos. O módulo de segurança também garante a transferência de dados de forma segura e privada. Além disso, a utilização de um SoC como *Edge-Gateway* torna-se um ponto positivo devido ao baixo custo e ao alto poder de processamento.

Além das vantagens apresentadas, a arquitetura também está preparada para atender algumas das tendências futuras para IoT, como por exemplo, estar preparada para o tamanho da rede, visto que algumas pesquisas mostram que até 2020 existirão entre 50 e 100 bilhões de dispositivos conectados a internet [Sundmaeker et al. 2010].

5. Conclusão e Trabalhos Futuros

Nesse trabalho foi proposta e detalhada uma arquitetura de *fog computing* para IoT que tem como objetivo aproximar de maneira eficiente as camadas consumidoras e provedoras de serviços.

É importante mencionar que *fog computing* não é um substituto de *cloud com-*

puting, mas sim, que essas duas tecnologias complementam uma a outra. A principal motivação do trabalho são as funções complementares que *fog* e *cloud* são capazes de proporcionar, as quais permitem ao usuário experimentar uma nova geração da computação, e também servem como requisito para que aplicações de tempo real e de baixa latência possam ser executadas nas bordas da rede. Além disso, essa combinação também permite suporte para análises completas de uma grande quantidade de dados na camada principal da rede.

Na continuidade dos trabalhos, a arquitetura proposta na Figura 2 terá sua implementação finalizada, bem como a realização de testes e validações em casos de uso reais. Os testes serão realizados no *Smart City Innovation Center* da PUCRS em parceria com a empresa Huawei. Tanto a PUCRS quanto a Huawei têm investido significativamente no desenvolvimento de novas tecnologias para impulsionar o crescimento da IoT voltado para cidades inteligentes.

Referências

- Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., and Ayyash, M. (2015). Internet of Things: A Survey on Enabling Technologies, Protocols and Applications. *IEEE Communications Surveys & Tutorials*, PP(99):1–1.
- Amaral, L. A., Tiburski, R. T., de Matos, E., and Hessel, F. (2015). Cooperative middleware platform as a service for internet of things applications. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15*, pages 488–493, New York, New York, USA. ACM Press.
- Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer Networks*, 54(15):2787 – 2805.
- Bjelica, M. Z., Golan, G., Radovanovic, S., Papp, I., and Velikic, G. (2014). Adaptive device cloud for internet of things applications. In *Consumer Electronics-China, 2014 IEEE International Conference on*, pages 1–3. IEEE.
- Bonomi, F., Milito, R., Zhu, J., and Addepalli, S. (2012). Fog Computing and Its Role in the Internet of Things. *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pages 13–16.
- Doukas, C. (2012). *Building Internet of Things with the ARDUINO*. CreateSpace Independent Publishing Platform.
- Hamdaqa, M. and Tahvildari, L. (2012). *Cloud Computing Uncovered: A Research Landscape*, volume 86.
- Laliwala, Z. and Chaudhary, S. (2008). Event-driven service-oriented architecture. In *Service Systems and Service Management, 2008 International Conference on*, pages 1–6. IEEE.
- Lewis, G., Echeverria, S., Simanta, S., Bradshaw, B., and Root, J. (2014). Tactical cloudlets: Moving cloud computing to the edge. In *Military Communications Conference (MILCOM), 2014 IEEE*, pages 1440–1446.
- Patidar, S., Rane, D., and Jain, P. (2011). A survey paper on cloud computing. *Proceedings - 2012 2nd International Conference on Advanced Computing and Communication Technologies, ACCT 2012*, pages 394–398.

- Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the internet of things: A survey. *Communications Surveys Tutorials, IEEE*, 16(1):414–454.
- Schenfeld, M. C., Vargas, F. D., Rebonatto, M. T., and Paixão, O. (2014). Middleware para equipamentos médicos em System on a Chip. *Brazilian Congress on Biomedical Engeneering*, pages 2608–2611.
- Sundmaeker, H., Guillemin, P., and Friess, P. (2010). *Vision and challenges for realising the Internet of Things*. Number March.
- Tan, L. (2010). Future internet: The Internet of Things. *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, pages V5–376–V5–380.
- Tiburski, R. T., Amaral, L. A., Matos, E. D., and Hessel, F. (2015). The importance of a standard security architecture for SOA-based iot middleware. *IEEE Communications Magazine*, 53(12):20–26.
- Yang, K.-p., Alkadi, G., Gautam, B., Sharma, A., Amatya, D., Charchut, S., and Jones, M. (2013). Park-a-lot: An automated parking management system. *Computer Science and Information Technology*, 1(4):276–279.
- Yi, S., Li, C., and Li, Q. (2015). A Survey of Fog Computing. *Proceedings of the 2015 Workshop on Mobile Big Data - Mobidata '15*, pages 37–42.
- Zhou, M., Zhang, R., Zeng, D., and Qian, W. (2010). Services in the cloud computing era: A survey. *2010 4th International Universal Communication Symposium, IUCS 2010 - Proceedings*, pages 40–46.

Escalabilidade e Paralelismo em Ambiente de Nuvem: Uma Solução para Processar Trâmites Judiciais

Emmanoel M. Sousa Junior, Frederico Lopes, Idalmis Milian

Instituto Metr pole Digital - Universidade Federal do Rio Grande do Norte
Natal - RN - Brasil

emmanoeljr@gmail.com, fred@imd.ufrn.br, idalmismilian@ect.ufrn.br

Abstract. *This paper proposes a scheduling mechanism for data processing in cloud computing environments. Such mechanism analyzes some specific variables in the business context of a company incubated at Metropole Digital Institute incubator, located at Federal University of Rio Grande do Norte. The main goal of this mechanism is to fulfill the sazonal demand using IaaS services and always considering two policies: (i) the maximum execution time allowed by the application may not be exceeded and (ii) the data have to be processed considering the lowest possible monetary cost. The proposed solution generates strategies to select the best set of virtual machines to process the current bunch of data. Such selection considers the amount of data, the estimated execution time for each specific strategy and the monetary cost of the virtual machines sets. In the context of this work, the strategy concept means the schedule of a set of virtual machines to process a specific amount of data, load balacing decisions and the paralelism of application's execution flow. The solution proposed in this work has resulted in great impact for that company since it allowed the vertiginous increase of the amount of clients served.*

Resumo. *Este artigo prop e um mecanismo de escalonamento para o processamento de dados em ambientes de computa o em nuvem. Tal mecanismo analisa algumas vari veis espec ficas no contexto de neg cios de uma empresa incubada no Instituto Metropole Digital incubadora, localizada na Universidade Federal do Rio Grande do Norte. O principal objectivo deste mecanismo   atender a demanda sazonal e altamente vari vel da empresa usando servi os de IaaS. O mecanismo deve sempre considerar duas pol ticas concomitantemente: (i) o tempo m ximo de execu o permitido pela aplica o n o pode ser ultrapassado e (ii) os dados devem de ser processados considerando o menor custo monet rio poss vel. A solu o proposta gera estrat gias para selecionar o melhor conjunto de m quinas virtuais para processar o grupo atual de dados. Tal sele o considera a quantidade de dados, o tempo de execu o estimado para cada estrat gia espec fica e o custo monet rio dos conjuntos de m quinas virtuais. No contexto deste trabalho, o conceito estrat gia significa a programac o de um conjunto de m quinas virtuais para processar uma quantidade espec fica de dados, decis es de balanceamento de carga e o paralelismo de fluxo de execu o da aplica o. A solu o proposta resultou em grande impacto para referida empresa uma vez que permitiu o aumento vertiginoso da quantidade de clientes atendidos.*

1. Introdu o

A computa o em nuvem   uma vis o conceitual que vem sendo utilizada nos  ltimos anos para descrever uma nova forma de se trabalhar a infraestrutura de TI, atrav s da rede mundial de computadores, ou em ambientes internos das organiza es, obedecendo crit rios de alta disponibilidade, escalabilidade e elasticidade dos recursos computacionais para apoiar um modelo de consumo dos recursos de acordo com as demandas funcionais dos sistemas. Este consumo n o envolve apenas recursos de

infraestrutura e armazenamento, mas também níveis de processamento de dados, serviços de *software* pré-configurados com pagamento de acordo com o consumo. Segundo Veras(2012), a visão dos principais serviços de um provedor de nuvem pode ser dividida nas seguintes camadas: SaaS (*Software as a Service*) aplicações de interesse de uma grande quantidade de usuários passam a ser hospedadas na nuvem como uma alternativa ao ambiente local; Paas (*Plataform as a Service*) capacidade oferecida pelo provedor ao desenvolvedor de aplicação que serão executadas e disponibilizadas na nuvem e Iaas (*Infrastructure as a Service*) capacidade que o provedor tem de oferecer uma estrutura de processamento e armazenamento de forma transparente.

O NIST (*National Institute of Standards and Technology*) argumenta que a computação em nuvem é um paradigma em evolução, um modelo que possibilita acesso, de modo conveniente e sob demanda, a um conjunto de recursos computacionais configuráveis que podem ser rapidamente adquiridos e liberados com mínimo esforço gerencial ou interação com o provedor de serviços (Mell and Grance 2009). Este contexto de TI surge da necessidade de trabalhos complexos, onde os usuários têm que realizar instalação, configuração e atualização de sistemas de *software* em uma arquitetura de solução mais inteligente, buscando uma melhor relação entre *hardware* e *software*.

O escalonamento de aplicações é outro aspecto importante no ambiente de nuvem. Selecionar, dentro de uma lógica controlável, os recursos necessários para a execução de determinado trabalho/tarefa, obedecendo a critérios de definição de ambiente como tempo e volume preestabelecidos. A escolha da melhor ou da mais apropriada estrutura de execução para o trabalho é determinada pelos algoritmos de escalonamento. Segundo TANEMBAUM (2008), dentro de uma visão de consumo de CPU, o escalonador é responsável por decidir qual trabalho será executado primeiro, caso haja vários trabalhos definidos para executar competindo pelo uso da CPU. Sendo também responsável por decidir qual é o processador mais adequado para executar cada tarefa, decidir qual é o intervalo de tempo que cada tarefa executará em cada processador e alocar os recursos necessários para cada tarefa; além de disparar a sua execução. Segundo Borges e Augusto (2011) é possível observar o comportamento do escalonador ao ver como a política utilizada afeta os recursos e os consumidores. Seguindo esta visão, este artigo mostra a necessidade de uma estratégia de escalonamento na nuvem, analisando o comportamento da aplicação no ambiente atual e como este atinge os clientes envolvidos. A nova estratégia deve estar baseada em um inteligenciamento da relação demanda x capacidade de processamento disponível no ambiente.

2. Descrição do problema

A empresa ATI *Service* LTDA, para a qual se desenvolveu este trabalho, tem sua estrutura de negócio relacionada a serviços que visam atender a provisão de informações jurídicas para advocacia em geral. Ela atua coletando processamento e gerindo informações relacionadas a processos judiciais de forma inteligente para disponibilizar aos interessados, através de uma plataforma de *software* em ambiente de nuvem. As informações enviadas variam em termos de volume de texto por registro, quantidade de registro por usuários responsável e quantidade de usuários responsáveis por clientes a cada dia. Estas informações são disponibilizadas até às 09 horas da manhã, devendo ser processadas e enviadas no formato do e-mail até às 10 horas do dia corrente, pois elas são fundamentais para tomada de decisão e trâmite de ações operacionais pelos clientes, relacionadas a prazos administrativos e judiciais. Segundo Maciel et al. (2005), o uso das novas tecnologias tem apoiado as ações do governo

eletrônico, em especial no eixo de serviços, onde o cidadão, advogados e gestores acessam processos judiciais via Internet.

A Tabela 1 apresenta os dados levantados e demonstram que a estrutura atual, operando com um volume de 120 de clientes, já não atende em alguns dias o requisito de processamento das informações para envio dos e-mails aos clientes, em relação ao contexto ideal que a empresa deseja atingir (abrangência nacional). No momento analisado, o sistema chega a um mínimo de dez vezes o volume atual de clientes, que é de no máximo 1 hora de processamento após a disponibilização, como pode ser visualizado do dia 20 ao dia 27 da amostragem.

Tabela 1. Amostragem de desempenho da rotina

Data	20/02/2015	23/02/2015	24/02/2015	25/02/2015	26/02/2015	27/02/2015	02/03/2015	03/03/2015
Registros	909	913	1765	1432	1406	1529	1531	1108
Hora início	09:00:00	09:00:00	09:00:00	09:00:00	09:00:00	09:00:00	09:00:00	09:00:00
Hora fim	11:41:02	10:38:25	10:53:50	10:45:30	10:05:50	10:05:27	09:52:22	09:44:38
Tempo	02:41:02	01:38:25	01:53:50	01:45:30	01:05:50	01:05:27	00:52:22	00:44:38

Esta amostragem do acompanhamento deixa explícita a necessidade de uma revisão da arquitetura do sistema. O objetivo de tal revisão é construir uma estrutura de operação do serviço que possa ser facilmente redimensionada em relação ao crescimento da demanda. Tal necessidade de reengenharia do sistema ficou ainda mais evidenciada quando, no dia 11/05/2015, a empresa vivenciou a situação em que, por um problema de captura das informações no sistema, o número de registros disponibilizados para envio aos clientes ultrapassou 9.000 para um único dia. Esse evento resultou em uma sobrecarga no processamento do servidor, provocando o atraso na entrega das mensagens aos clientes e a lentidão das aplicações que estão operantes no mesmo ambiente. Esta situação do dia 11/05/2015 demonstrou na prática que o sistema atual não estava estruturado para atender um nível de demanda tão alto. Um ponto positivo foi que, devido a estudos preliminares já realizados neste trabalho até a data do ocorrido, foi possível identificar o problema rapidamente e montar uma estratégia de ação imediata para minimizar o transtorno junto aos clientes.

3. Em busca de uma solução para o problema

Para achar um padrão mínimo de infraestrutura que possibilitasse o devido planejamento das atividades e um escalonamento controlado dos serviços, em relação a demanda na nuvem, foram utilizados os dados do universo apresentado na Tabela 1. Neste cenário, foi possível observar a quantidade de registros em relação ao tempo de operação para montar as mensagens a serem enviadas aos responsáveis de cada cliente. Para realizar as simulações necessárias, a empresa resolveu utilizar o ambiente de nuvem da *Amazon* (AMAZON WEB SERVICES), devido a mesma já possuir alguns serviços operando neste ambiente. A *Amazon* apresenta uma proposta de escalonamento de infraestrutura bem ágil e com custo de utilização por tempo ativo do serviço, o que facilita o controle do investimento. Neste momento, o principal foco foi buscar o entendimento dos fatores e pesos que podiam influenciar no resultado do serviço, segundo a visão de escalonamento no ambiente de nuvem. A configuração do ambiente selecionado é apresentada na Tabela 2.

Tabela 2. Dados de configuração do ambiente original do serviço

Modelo	Memória	Processamento	Sistema Operacional.	Serv. WEB	Linguagem
t2.micro	1 GiB	1 vCPUs, 2.5 GHz, Intel Xeon Family	Win 2003	IIS	PHP

Para atender as condições deste cenário e analisando a aplicação operando tanto, em termos das tecnologias utilizadas, como também em relação a versão destas tecnologias;

foi selecionado o sexto dia da amostragem da Tabela 1 que contem 1.529 registros como referência para a simulação. Observou-se que o sexto dia (27/02/2015) já estava no limite aproximado das necessidades de negócio da empresa, ou seja, com um tempo aproximado de uma hora, replicando em um desempenho similar ao quinto dia (26/02/2015) de operação do serviço com 1.406 registros. Para achar uma solução a este problema foram implementadas sequencialmente, quatro fases distintas e complementares apresentadas a continuação.

3.1 Fase 1 – Melhoria da Infraestrutura de *hardware*.

Com o ambiente de *software* equiparado e utilizando a mesma base de dados, foram executadas simulações do processamento dos dados para melhorar o perfil da máquina hospedeira e observar se o *software* teria um melhor desempenho. A partir do experimento realizado foram obtidos os seguintes resultados apresentados na Tabela 3.

Tabela 3. Desempenho do dia 27/04/2015 em ambiente de simulação

Registros	Mensagens	Modelo/Memória/ Processamento	Sist. Opera cional	Serv. WEB	Linguagem	Data do teste	Tempo
1.529	842	t2. <i>micro</i> / 1 GiB / 1 vCPUs, 2.5 GHz, Intel Xeon Family	Win 2003	IIS	PHP	14/03/2015	1:04:59
1.529	842	t2. <i>small</i> / 2 GiB / 1 vCPUs, 2.5 GHz, Intel Xeon Family	Win 2003	IIS	PHP	14/03/2015	0:56:13
1.529	842	t2. <i>medium</i> / 4 GiB / 2 vCPUs, 2.5 GHz, Intel Xeon Family	Win 2003	IIS	PHP	15/03/2015	1:01:52

O desempenho da rotina original, sem alterações em sua estrutura, no ambiente de simulação, replicou de forma equiparada ao sexto dia de execução da rotina, utilizando o modelo t2.*micro* do serviço de nuvem da Amazon, que opera com 1Gb de memória e um núcleo de processamento. Quando utilizado o modelo t2.*small* contendo 2Gb de memória e mantendo um núcleo de processamento, ocorreu uma variação de quase 10 minutos no processamento da rotina. Entretanto, ocorreu uma variação menor de desempenho, cerca de três minutos, quando realizado o upgrade para t2.*medium* contendo dois núcleos de processamento e 4Gb de memória. A estrutura de processamento sequencial na qual a rotina está construída, demonstrou que o processamento segue todos os passos para todos os clientes, mesmo não existindo registros para serem enviados no dia para um determinado cliente. Isto sobrecarrega o processador da máquina, considerando que apresenta dois núcleos de processamento.

Uma verificação importante no sistema operacional foi durante a execução realizada com a máquina t2.*medium* (com dois núcleos de processamento), onde a rotina utilizou somente um núcleo deixando o outro ocioso ou livre. Este caso mostrou que a simples melhora do ambiente de *hardware* pode não influenciar no desempenho da rotina se a aplicação não estiver preparada, pois a mesma não consegue tirar o melhor proveito do potencial disponível no ambiente. Este cenário indicou como a aplicação poderia explorar o paralelismo do *hardware* para assim obter um melhor desempenho.

3.2 Fase 2 – Particionamento de dados para execução da rotina de forma paralela por grupo de clientes.

Durante esta fase foram definidas regras que deviam ser utilizadas na aplicação, em relação a estrutura da rotina e da demanda apresentada para o dia na empresa. Foram realizadas algumas simulações para descobrir as variáveis que teriam influência no desempenho e para obter uma proposta de reengenharia da arquitetura do *software*. O objetivo neste caso era processar as informações em cerca de 1 hora. Visto que o *software* trabalha em forma sequencial, foi realizada uma simulação da rotina para

observar como seria o desempenho da mesma com uma abordagem paralela simples. Para isto, foram divididos os clientes em 5 grupos, executando a rotina original sem alterações estruturais e sendo chamada em instâncias concorrentes na mesma máquina.

Nesta primeira solução, foi utilizado o comando *exec* do PHP, linguagem em que já estava desenvolvida a rotina, num arquivo *.cmd* para chamar as instâncias do *software*. O escalonamento fica a cargo das regras de processamento interno do próprio sistema operacional da máquina. O critério utilizado para estimar a distribuição em cinco grupos de dados, foi que o tempo atual deveria ser de no máximo 20% do tempo total, já que as cargas de dados estariam sendo tratadas em paralelismo de atividades, através das instâncias do *software* dividido em cinco processos. Na primeira simulação (cenário 1) como mostram as Tabelas 4 e 5, queria se verificar se o tamanho dos textos tratados no processamento teria algum peso relevante. Neste cenário todos os clientes são agrupados por tamanho dos registros em bytes, para entender se o volume de texto pode estar influenciando no processamento.

Tabela 4 – Grupos de distribuição de cliente do cenário 1

Grupo	Cientes	Registros	Tamanho (bytes)
A	1	1.241	3.741.276
B	5	164	712.493
C	12	66	486.535
D	16	58	120.538
E	86	0	0
Total	120	1.529	5.060.842

Neste primeiro cenário identificou-se que o tamanho dos registros em bytes não tem influência sobre o processo, visto que o Grupo E contendo clientes sem registros foi o que consumiu mais tempo. O fato que mereceu destaque foi a simples utilização de 5 grupos de processamento paralelo no modelo de máquina *t2.micro*, que apresentou uma elevação de desempenho de 50% , caindo o tempo de execução em 30 minutos o que representa um resultado considerável para a empresa. A avaliação deste cenário (Figura 1) aponta que o número de clientes tem uma influência considerável sobre o desempenho da rotina. Desta forma, decidiu-se realizar outra simulação, cenário 2, onde o número de clientes fosse constante e tratados somente clientes com registros a serem processados. Neste segundo cenário, apresentado nas Tabelas 6 e 7, são considerados grupos com quantidade de registros e clientes iguais.

Tabela 5 – Resultados de simulação do cenário 1

Registros / Modelo	T2.MICRO		T2.SMALL		T2.MEDIUM	
	Grupos	0:30:20	Grupos	0:26:03	Grupos	0:29:51
1.241	A	0:10:26	A	0:10:15	A	0:08:56
164	B	0:15:15	B	0:14:49	B	0:15:33
66	C	0:07:21	C	0:06:59	C	0:08:52
58	D	0:09:55	D	0:09:58	D	0:11:41
0	E	0:30:20	E	0:26:03	E	0:29:51

Neste novo cenário de equalização das variáveis de número de clientes e número de registros, a simulação apresenta os mesmos modelos trabalhados na primeira fase.

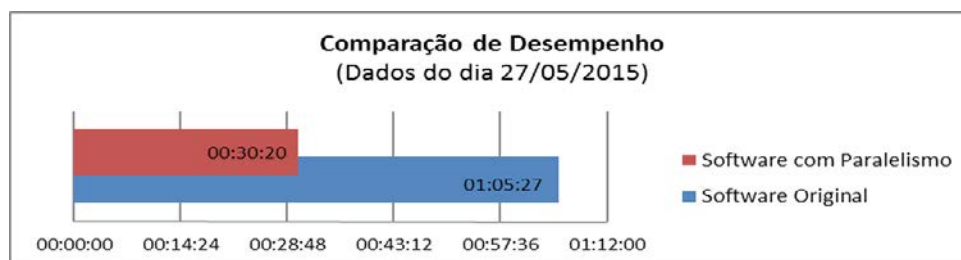


Figura 1 – Gráfico comparativo do primeiro cenário de simulação

Tabela 6 – Grupos de distribuição de cliente do cenário 2

Grupo	Cientes	Registros
A	5	1.598
B	5	1.598
C	5	1.598
D	5	1.598
Total	20	6.392

Tabela 7 – Resultados de Simulação do cenário 2

Registros / Modelo	t2.MICRO		t2.MEDIUM	
6.392	Grupos	2:36:09	Grupos	0:22:53
1.598	A	1:53:28	A	0:10:51
1.598	B	1:28:33	B	0:08:38
1.598	C	2:36:09	C	0:22:53
1.598	D	1:51:47	D	0:11:19

Demonstra-se aqui que a *t2.MICRO* sofreu o impacto da elevação do número de registros, extrapolando-se o tempo de 1 hora de execução em todos os grupos. Como a máquina possui somente um núcleo de processamento, tornou inviável a tentativa com o modelo *t2.SMALL*. Já com *t2.MEDIUM* o tempo de execução caiu de forma considerável, o que mostrou que a equalização do número de clientes e o número de registros podia ser utilizado como variável de parametrização para definir o modelo de infraestrutura de máquina empregada em um ambiente que comporte o paralelismo do processamento. Neste cenário foram identificados pontos que podiam ser trabalhados para evitar processamentos desnecessários. Este caso mostrou também que a estratégia de escalonamento, utilizando as variáveis de número de clientes e número de registros iguais, podia definir bem que tipo de infraestrutura de máquina devia ser selecionada de acordo com a demanda do dia, para assim executar o melhor resultado em termos de processamento.

3.3 Fase 3 – Reengenharia da estrutura do *software* buscando o melhor balanceamento das cargas de dados.

Esta fase apresenta uma proposta de reengenharia do *software* para se obter o melhor desempenho dentro do ambiente original. Foi utilizado balanceamento de carga durante o processamento, através das variáveis “número de clientes” e “número de registros”. A Figura 2 mostra o novo fluxo que foi proposto para o *software*.

Este novo fluxo apresentou três pontos fundamentais para a melhoria do desempenho do *software*: 1. Separação das empresas/clientes que possuem e que não possuem registros a serem enviados no dia. Desta forma, as empresas que não possuem registros para o dia não precisam percorrer o processamento de todos os seus usuários, sendo definido um modelo padrão de relatório a ser enviado para todos eles. 2. Separação das empresas/clientes que possuem registros para o dia em grupos de cliente com quantidade de registros equivalentes. Desta forma, é possível montar um paralelismo da rotina e diminuir assim o tempo de processamento geral como apresentado nas simulações realizadas anteriormente. 3. Dentro do processamento das mensagens de cada cliente, separar os usuários que podem ver todas as mensagens do cliente dos que só podem ver as que estão sobre sua responsabilidade. Desta forma, diminuí esforço de montagem de várias mensagens com grande número de registros, montando somente uma mensagem padrão e enviando para todos os usuários envolvidos em uma única mensagem. Para resumir as simulações realizadas no *software*, a Tabela 8 mostra a evolução até chegar na versão atual.

Tabela 8 – Resumo da evolução das versões do *software*

Versão	Descrição	Resultado
1	<i>Software Original</i> : sem alterações na estrutura de execução.	Em alguns momentos não atende o requisito de 1 hora no tempo de processamento das informações para envio das mensagens.
1.5	<i>Software com Paralelismo</i> : sem alterações na estrutura de execução, mas sendo executados em grupos subdivididos de clientes por número de registros, em formato paralelo, dentro da mesma máquina.	Extrema melhoria de desempenho, chegando a diminuir 50% no tempo do processamento do volume atual de informações.
2	<i>Software Otimizado e com Paralelismo</i> : alterações estruturais que realizam o balanceamento de carga em número de clientes e número de registros, além de evitar esforços desnecessários no tratamento das informações.	Extrema melhoria de desempenho, chegando a diminuir 90% no tempo do processamento do volume atual de informações.

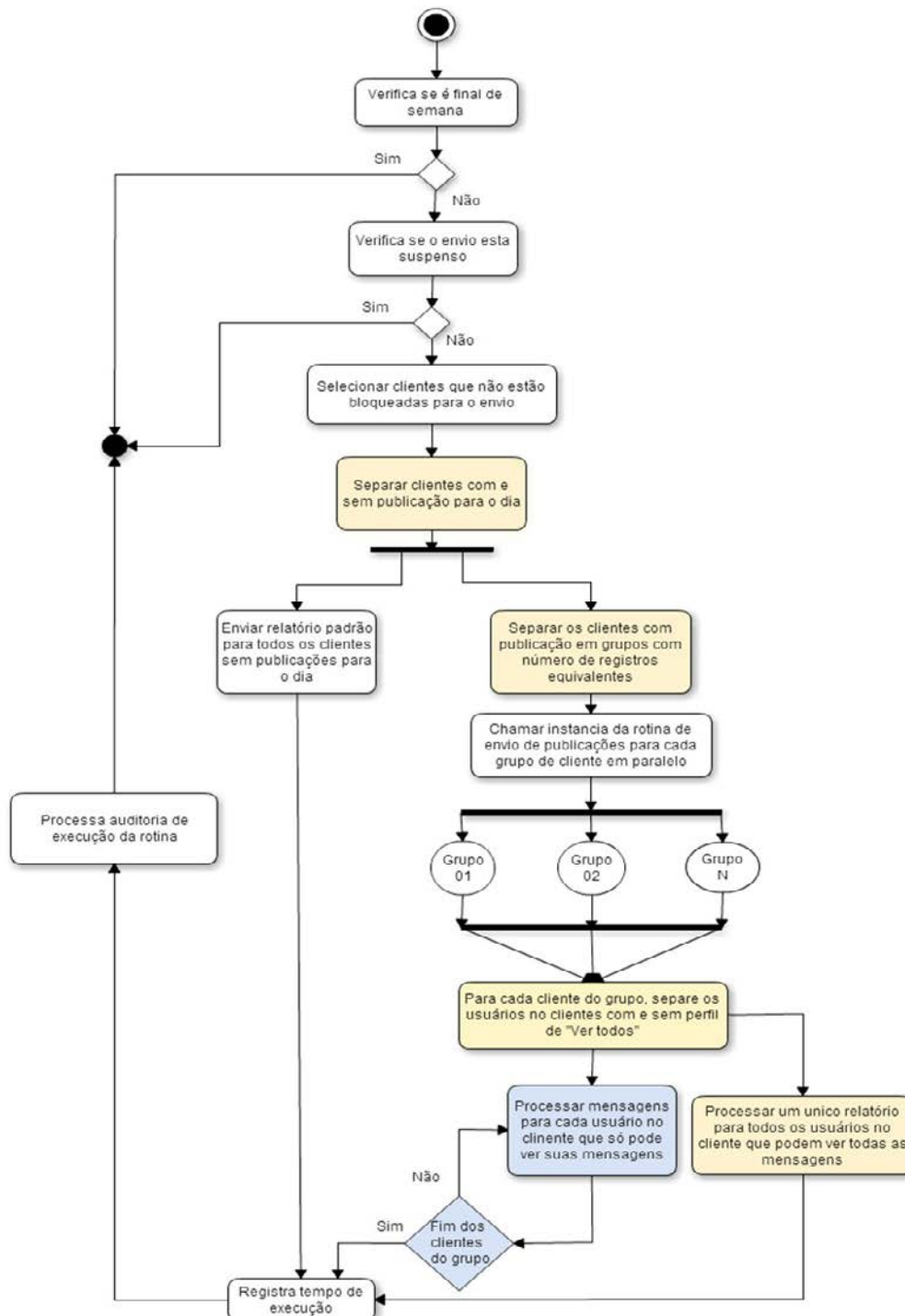


Figura 2 – Proposta de reconfiguração do *software*.

A comparação do desempenho da rotina pura com a rotina otimizada é apresentada na Figura 3. Nesta fase de testes, no ambiente onde o servidor não está dedicado totalmente, já apresenta os resultados com os dados da última semana do mês de maio. Foi concluído que, para o *software* obter resultados com maior exatidão em relação às variáveis que estão sendo trabalhadas, deve ser implantado em um servidor dedicado sem interferências de outras aplicações durante a sua execução.

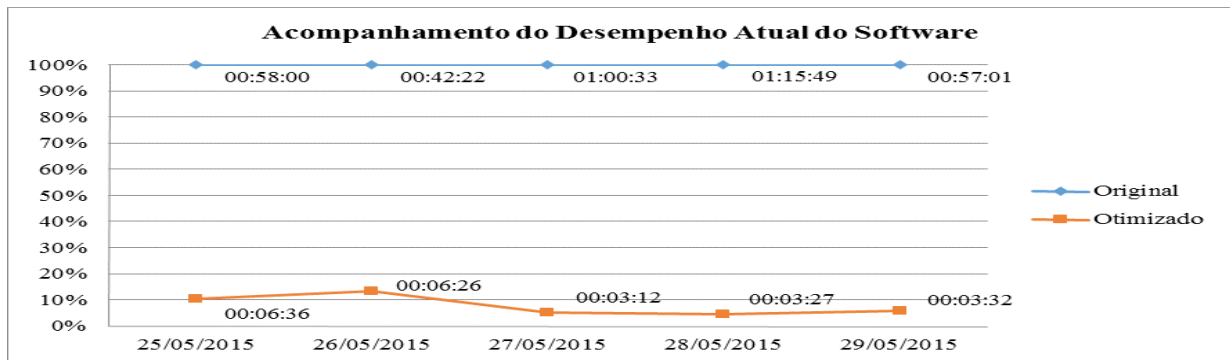


Figura 3 – Gráfico do acompanhamento de desempenho do *software* atual em relação ao *software* otimizado

Entretanto, já se obteve nesta fase um resultado extremamente importante para empresa, com um ganho real que pode atender com maior margem de segurança aos seus atuais clientes. O problema tratado até o momento tentou resolver a situação de otimização do *software* para uma realidade de demanda que pode ser atendida por uma máquina virtual com configuração pré-definida. Agora, e se o crescimento da demanda alcançar o limite de operação da máquina, mesmo com todas as otimizações realizadas no *software*? A resposta conclui a necessidade de uma estratégia de escalonamento *para o software* sobre a infraestrutura considerada, observando questões de memória e processamento.

3.4 Fase 4: Construção de uma ferramenta para o escalonamento da infraestrutura de nuvem de acordo com a demanda do dia.

A proposta nesta fase foi construir uma ferramenta, que em uma situação de demanda que supera o potencial da máquina padrão, realize uma seleção automática do modelo de máquina para a execução do *software* dentro dos modelos disponíveis da nuvem. Esta ferramenta deve ler as informações necessárias para que o *software* possa apontar qual o melhor modelo de máquina para processar os dados. Desta forma, o escalonamento do *software* deve atender os requisitos de desempenho estabelecidos pela empresa. O diagrama do algoritmo proposto para o escalonamento na nuvem é apresentado na Figura 4, com os passos que necessários para selecionar os recursos. A próxima seção descreve os testes que foram realizados e os resultados obtidos utilizando a estratégia de escalonamento proposta.

4. Ambiente de testes e resultados do escalonamento da aplicação na nuvem

A nuvem da *Amazon* possui uma lista com 39 modelos de máquinas virtuais disponíveis, com uma escala crescente de formatações em termos de memória, processamento e valor por tempo/hora de utilização. O custo da utilização por hora é proporcional ao potencial de processamento de cada modelo. Para as simulações de escalonamento realizadas, foi utilizada como referência os modelos trabalhados nas fases 1 e 2. Para identificar a potência de processamento de registros de cada tipo de máquina, foram realizados testes individuais com cada tipo de máquina, variando as cargas de registros (Tabelas 9 e 10). Com o objetivo de fazer uma simulação completa da arquitetura de

inteligência, para escalonamento e paralelização do processamento utilizando as máquinas no ambiente da Amazon proposto, construímos um cenário controlando o número de registros a serem processados e relatamos o comportamento de cada máquina virtual selecionada. Na construção deste cenário ampliamos o volume de registros para a equivalência de 40 dias, contemplando o período de 20/09/2015 à 30/10/2015.

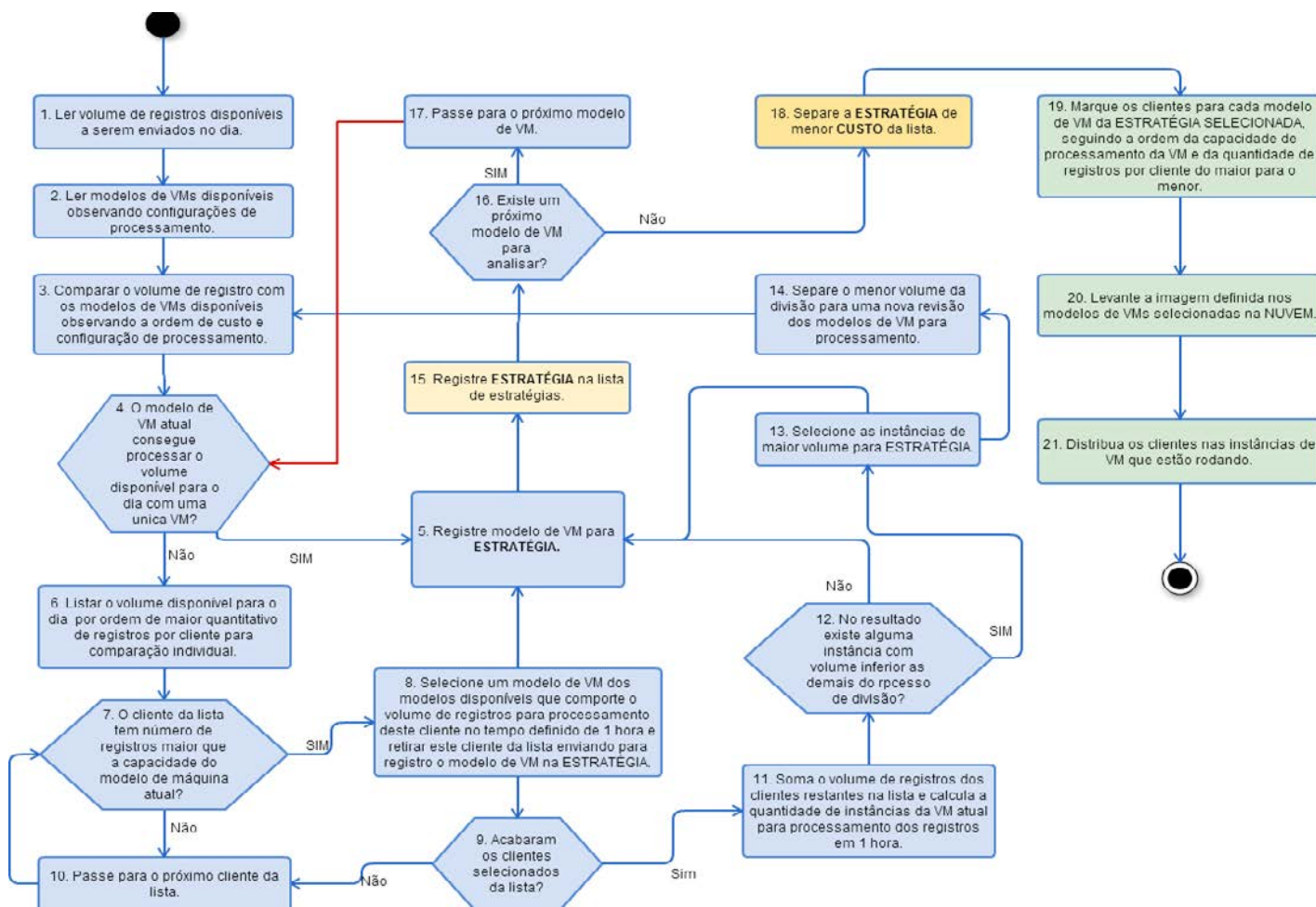


Figura 4 – Algoritmo de escalonamento.

Tabela 9 – Custos e capacidade de processamento por configuração de máquina virtual na Amazon.

MODELO	TIPO	CPU	MÉM.	CUSTO HORA (US)	REGISTROS/HORA
1	t2.micro	1	1	0.018	10.000
2	t2.small	1	2	0.036	15.000
3	t2.medium	2	4	0.072	20.000
4	t2.large	2	8	0.134	60.000

Tabela 10 – Opções de estratégia para o cenário na Amazon.

Opção	Estratégia	Custo (\$)
A	7 máquinas (t2.micro) com 10.000 registros e 1 máquina (t2.small) com 921 registros	0.162
B	5 máquinas (t2.small) com 15.000 registros e 1 máquina (t2.micro) com 8.921 registros	0.198
C	4 máquinas (t2.medium) com 20.000 registros e 1 máquina (t2.micro) com 3.921 registros	0.30
D	1 máquinas (t2.large) com 60.000 registros e 1 máquina (t2.large) com 23.921 registros	0.26
E	1 máquinas (m4.large) com 83.921 registros	0.25

Neste período foram identificados 136 clientes para o volume de registro a ser processado e o algoritmo de escalonamento gerou as seguintes opções para resolução da demanda. A Tabela 11 apresenta o resultado em que a opção de estratégia A tem o menor custo de processamento em relação a escala dos cenários seguintes. Também neste caso *t2.large*, na opção de estratégia D, não consegue processar toda a demanda em uma hora sozinha com custo maior do que o da estratégia A. A observação destes dois cenários iniciais apontam a política de custo da *Amazon*, em que os modelos de máquinas segue uma escala de dobrar o custo diante de cada mudança sequencial de configuração de modelo. Os cenários seguintes sempre serão mais caros que o cenário anterior, pois a capacidade de processamento da máquina não consegue acompanhar a mesma escala de progressão dentro dos cenários de ajustes realizados.

Tabela 11 – Resultado da estratégia selecionada para o cenário na *Amazon*.

Data	Cientes	Registros	Tempo Plan.	Custo Plan.	Custo Total
01/12/2015	136	83.921	60	((8 X \$ 0,018) + \$ 0,036)	(\$) 0,18
Máquina	Hora estratégia	Início processo	Fim processo	Tempo real	Custo real
t2.small	12:48:41	12:50:45	13:31:43	00:40:58	(\$) 0,024
t2.micro	12:48:42	12:50:39	13:21:38	00:30:59	(\$) 0,0093
t2.micro	12:48:47	12:50:45	13:34:28	00:43:43	(\$) 0,0129
t2.micro	12:48:50	12:50:41	13:27:13	00:36:32	(\$) 0,0108
t2.micro	12:48:43	12:50:39	13:36:13	00:45:34	(\$) 0,0135
t2.micro	12:49:22	12:50:46	13:55:17	01:04:31	(\$) 0,0192
t2.micro	12:48:45	12:50:37	13:02:17	00:11:40	(\$) 0,0036
t2.micro	12:48:48	12:50:39	14:05:17	01:14:05	(\$) 0,0222
t2.micro	12:48:49	12:50:49	14:05:06	01:14:17	(\$) 0,0222
Custo real total:				01:14:17	(\$) 0,1377

Objetivando ter resultados diferenciados para comparação, foi construída uma tabela fictícia (Tabela 12), de um suposto provedor de nuvem em que a escala de custos não siga a mesma regra de valoração da *Amazon*. Nesta nova tabela mantivemos as nomenclaturas e capacidade de processamento da *Amazon*, só que ajustamos a escala de agregação de custos onde o valor seguinte seja 20% menor que o dobro do valor anterior. O cenário da nuvem fictícia tem a mesma ampliação do volume de registros para a equivalência de 40 dias demanda, contemplando o período de 20/09/2015 à 30/10/2015. Neste período foram identificados 136 clientes para o volume de registro a ser processado e o algoritmo de escalonamento as mesmas opções do cenário anterior da *Amazon*, seguindo os novos critérios de custos para seleção da estratégia mais viável.

Tabela 12 – Custos e capacidade de processamento por configuração de máquina virtual em uma nuvem fictícia.

Modelo	Tipo	CPU	Mémoria	Custo Hora (US)	Registros/Hora
1	t2.micro	1	1	0.018	10.000
2	t2.small	1	2	0.028	15.000
3	t2.medium	2	4	0.044	20.000
4	t2.large	2	8	0.070	60.000

A execução do algoritmo de escalonamento para 83.921 registros neste novo contexto de nuvem fictícia, de custo por modelo de máquina, indica a opção de estratégia D como a mais viável, veja Tabela 13. Observe que apresenta custo menor que as opções anteriores e não encontrou nenhuma combinação posterior a disposição, que possa processar a demanda com menor custo de comparação. Apesar do algoritmo de

escalonamento ter gerado as estratégias de forma coerente em relação as questões de demanda e custo preestabelecidos nas regras, os resultados da nuvem com valores fictícios não puderam ser executados, devido a incompatibilidade da tecnologia hoje utilizada pela empresa. A mesma ainda trabalha com sistema operacional *Windows 2003 Server R2* para este serviço; o qual não suporta a configuração de *hardware* superior a 4 Gb de memória e dois núcleos de processamento. Como a partir da *t2.large* a *Amazon* já utiliza 8 Gb de memória, a imagem que é utilizada como padrão, em ambiente *Windows 2003 Server R2*, não consegue subir as instâncias de máquinas virtuais para o processamento da demanda.

Tabela 13 – Opções de estratégia para o cenário 2 utilizando a nuvem fictícia.

Opção	Estratégia	Custo (\$)
A	7 máquinas (<i>t2.micro</i>) com 10.000 registros e 1 máquina (<i>t2.small</i>) com 921 registros	0.154
B	5 máquinas (<i>t2.small</i>) com 15.000 registros e 1 máquina (<i>t2.micro</i>) com 8.921 registros	0.158
C	4 máquinas (<i>t2.medium</i>) com 20.000 registros e 1 máquina (<i>t2.micro</i>) com 3.921 registros	0.194
D	1 máquinas (<i>t2.large</i>) com 60.000 registros e 1 máquina (<i>t2.large</i>) com 23.921 registros	0.14

Esta situação reforça o contexto teórico tratado no início deste trabalho de que as empresas devem tratar suas aplicações de *software* observando o contexto da nuvem onde pretende operar seus serviços para que possam realmente tirar o melhor proveito dos potenciais de escalabilidade propostos. Segue-se então a recomendação que a empresa atualize a tecnologia de base de seu serviço para tirar melhor proveito dos provedores de nuvem disponíveis hoje no mercado.

5. Trabalhos relacionados

J. NAMJOSHI (2009) aborda a necessidade de empresas de viagens acessar a um grande número de redes de hotéis e serviços das companhias aéreas em todo o mundo, acessando suas tarifas mais recentes e informações de disponibilidade que irão ser atualizadas dinamicamente. As datas de viagem podem mudar a qualquer momento e os usuários podem precisar de ajustes rápidos em itinerários envolvendo passagens aéreas e reservas de hotel. Para atender a demanda, a ferramenta implementada no ambiente escalável da computação em nuvem distingue-se da visão típica da virtualização e destaca o poder que a provisão dos serviços na nuvem pode oferecer aos usuários. O sistema lida com aumento da demanda de clientes e fornecedores de viagens a medida que o negócio cresce, e busca a possibilidade de ajuste dinâmico do uso de infraestrutura de TI em relação a variação das necessidades do negócio. Os serviços de nuvem utilizados visam atender os requisitos de escalabilidade, disponibilidade e desempenho.

Já Lago (2012) apresenta um escalonador que leva em consideração a quantidade de energia necessária para executar determinada tarefa. Essas informações serão fornecidas ao escalonador EES (*Energy Efficient Scheduler*), que escalonará o job para o *cluster* que compõem a grade que obtiver o menor gasto de energia (informação fornecida pelo mecanismo que calcula o consumo), utilizando um dos algoritmos de escalonamento disponíveis. O projeto apresenta uma solução de escalonamento em ambiente de nuvem buscando a otimização do tempo com foco na economia de energia das máquinas virtuais na nuvem. Foi desenvolvido um algoritmo de regra de escalonamento de prioridade permitindo a redução do tempo de processamento de cargas com alta prioridade, sem comprometer significativamente o consumo de energia. O objetivo foi a organização das cargas de maior prioridade para seleção das máquinas virtuais mais aptas para seu processamento, sem prejudicar criticamente as cargas com baixa prioridade e praticamente, sem aumento do consumo de energia no escalonamento.

6. Conclusões

As ações exploratórias que foram desenvolvidas nas diferentes fases, geraram uma sequência de resultados evolutivos mostrando que a arquitetura de sistemas em nuvem requer uma visão multidisciplinar e contextualizada da relação entre negócios, *software* e *hardware*. As organizações que pretendem operar seus serviços baseados em *software* no ambiente de nuvem, devem observar se suas aplicações estão preparadas para alcançar os reais benefícios disponíveis neste ambiente. Em relação ao processamento paralelo se observou que o simples fato da paralelização do *software* em grupos de clientes e número de registros, já gerou um ganho real para empresa em termos de desempenho no tempo de execução, mas se a escala de clientes/registros tiver um crescimento considerável, em que a estrutura de máquina padrão não der suporte, será necessário melhorar a infraestrutura da máquina. Uma grande vantagem de se trabalhar com o ambiente de nuvem é a facilidade e agilidade de escalonar a infraestrutura disponível em relação as regras de negócio que se apresentarem de forma controlada. O algoritmo de escalonamento criado conseguiu definir a melhor estratégia para a demanda de registros e assim sugerir a infraestrutura virtual de forma automática, selecionando o modelo de máquina que atenda a demanda do momento. O desafio para empresa agora é seguir este modelo de arquitetura, atualizando as tecnologias do serviço atual e adequando os demais serviços de envio de informações automáticas a seus clientes, seguem os mesmos padrões da problemática tratada neste caso. Esta arquitetura de escalonamento pode evoluir para trabalhar o processamento paralelo de forma mais eficiente e também em nível de interoperabilidade entre nuvens, onde o algoritmo de escalonamento pode se utilizar da questão do valor ofertado pelos provedores do serviço de nuvem como parâmetro de seleção das máquinas virtuais a serem utilizadas.

7. Referências

- AMAZON WEB SERVICES. < <http://aws.amazon.com/>> Acesso em: 08 de novembro de 2015.
- BORGES, Carlos Augusto Lima. Escalonamento de tarefas em uma infraestrutura de computação em nuvem federada para aplicações em bioinformática. Brasília: UnB, 2011, 115p;
- J. NAMJOSHI; A. GUPTE. Service oriented architecture for cloud based travel reservation Software as a Service. In IEEE CLOUD, pages 147-150, 2009;
- LAGO, D. G., MADEIRA, E. R. M., and BITTENCOURT, L. F. (2012). Escalonamento com prioridade na alocação ciente de energia de máquinas virtuais em nuvens. In ' Anais do XXX SBRC. <http://www.lrc.ic.unicamp.br/~bit/arquivos/SBRC2012escalonamento_com_Prioridade_na_Alocacao_Ciente_de_Energia_de_Maquinas_Virtuais_em_Nuvens.pdf> Acesso em 12 de outubro de 2015.
- MACIEL, Cristiano; NOGUEIRA, José Luiz T.; GARCIA, Ana Cristina Bicharra. An x-ray of the brazilian e-gov web sites. In: **Human-Computer Interaction-INTERACT 2005**. Springer Berlin Heidelberg, 2005. p. 1138-1141.
- MELL, P. and Grance, T. (2009). Draft NIST Working De-finition of Cloud Computing. National Institute of Standards and Technology. <<http://csrc.nist.gov/groups/SNS/cloud-computing>> Acesso em: 19 de dezembro de 2015.
- SHAW, M., GARLAN, D. *Software Architecture: perspectives on an Emerging Discipline*. New Jersey: Prentice-Hall, 1996. TANEMBAUM, A.; WOODHULL, A. *Sistemas operacionais, projeto e implementação*. Tradução João Tortello. 3. ed. Porto Alegre: Bookman, 2008, 992p.
- TURION, Cezar. *Cloud Computing. Computação em nuvem. Transformando o Mundo da Tecnologia da Informação*. Rio de Janeiro: Brasport, 2009.
- VAQUERO, L. M., Rodero-Merino, L., Caceres, J., and Lindner, M. A break in the clouds: towards a cloud definition. SIGCOMM Comput. Commun. Rev. 2008.
- VERAS, Manoel. *CLOUD COMPUTING: a nova Arquitetura da TI*. Brasport - São Paulo, 2012.

WCAMA - Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais

Apresentação

O 7º Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA) teve como objetivo promover a integração efetiva da área de computação (metodologias e ferramentas) com temas relacionados ao meio ambiente (política e gestão).

Como a gestão dos recursos naturais é uma atividade complexa e dinâmica, ela requer a integração entre atores nos campos social, político e tecnológico para serem efetivamente desenvolvidas e implantadas.

Neste contexto, busca-se discutir do ponto de vista da computação o desenvolvimento de metodologias e ferramentas para gestão do meio ambiente e recursos naturais. Mais especificamente, trabalha-se no sentido de tratar eficientemente i) o gerenciamento e a comunicação entre grandes volumes de dados; ii) o desenvolvimento de técnicas para análise destes dados; e, iii) a otimização, o controle e integração dos dados gerados.

Nesta edição, o WCAMA recebeu 22 submissões de artigos, dos quais 19 artigos completos e 3 artigos curtos. Destes, foram aceitos para publicação 7 artigos completos (~37% de aceitação) e 1 artigo curto (~33% de aceitação).

Gostaríamos de agradecer o apoio dos colegas pesquisadores do comitê de programa e das nossas instituições (FURG, UFPel e UFCG); e, saudamos os autores dos artigos e os participantes do WCAMA 2016.

Diana Francisca Adamatti (Universidade Federal do Rio Grande)

Marilton Sanchotene de Aguiar (Universidade Federal de Pelotas)

Cláudio Campelo (Universidade Federal de Campina Grande)

Comitê de Organização

Coordenação Geral

Diana Francisca Adamatti (Universidade Federal do Rio Grande)

Coordenação do Comitê de Programa

Marilton Sanchotene de Aguiar (Universidade Federal de Pelotas)

Cláudio Campelo (Universidade Federal de Campina Grande)

Comitê de Programa

Adenauer Yamin (Universidade Federal de Pelotas)

Alessandro Machado de Araújo (Universidade Federal de Goiás)

Aline Loreto (Universidade Federal de Santa Maria)

André Luiz Moura (Universidade Federal de Goiás)

Anselmo Paiva (Universidade Federal do Maranhão)

Antonio Carlos da Rocha Costa (Universidade Federal do Rio Grande)

Celia Ralha (Universidade de Brasília)

Claudio Baptista (Universidade Federal de Campina Grande)

Claudio Campelo (Universidade Federal de Campina Grande)

Clodoveu Davis (Universidade Federal de Minas Gerais)

Diana Adamatti (Universidade Federal do Rio Grande)

Eliane Araújo (Universidade Federal de Campina Grande)

Fabio Andrade (Instituto Federal da Paraíba)

Frederico Fonseca (Pennsylvania State University)

Hailton Lemos (Faculdade Nossa Senhora Aparecida)

Jean-Pierre Briot (Université Pierre et Marie Curie)

João Balsa (Universidade de Lisboa)

Marco Antônio Oliveira Domingues (Instituto Federal de Pernambuco)

Maria Isabel Sobral Escada (Instituto Nacional de Pesquisas Espaciais)

Marilton Aguiar (Universidade Federal de Pelotas)

Pedro Andrade (Instituto Nacional de Pesquisas Espaciais)

Valéria C. Times (Universidade Federal de Pernambuco)

Vinícius Sebba Patto (Universidade Federal de Goiás)

Vinicius Cunha M Borges (Universidade Federal de Goiás)

Vinicius Ribeiro (Universitário Ritter dos Reis)

Painéis / Palestras

Computação de Alto Desempenho Aplicada ao Entendimento da Dinâmica Oceânica: Observações e Simulações Numéricas

Prof. Dr. Fabrício Sanguinetti Cruz de Oliveira (Universidade Federal do Rio Grande)

A dinâmica oceânica integra um amplo espectro de movimentos em diferentes escalas, desde centímetros a milhares de quilômetros e de segundos a décadas. O real entendimento destes movimentos muitas vezes está associado ao nível de detalhes ao qual podemos observá-los, ou seja, a taxa de amostragem dos mesmos. Dados oceanográficos tradicionalmente coletados a bordo de cruzeiros de pesquisa são muitas vezes inadequados para estudar fenômenos espacialmente distribuídos e que se desenvolvem por um longo período, dada a natureza pontual das observações e o lento deslocamento das embarcações. Neste contexto, dados obtidos através de sensores remotos a bordo de satélites e modelos numéricos que resolvem as principais equações dinâmicas do oceano vêm sendo cada vez mais utilizados. O advento dos satélites de observação dos oceanos em alta resolução e a aumentos nos esforços em modelar o oceano resultaram em informações mais acuradas sobre os processos físicos oceânicos no espaço e no tempo. Todavia este aumento nos esforços de observar o oceano em alta resolução só foi possível devido ao grande desenvolvimento na tecnologia de computadores de alto desempenho. Estes promoveram significantes avanços no processamento e armazenamento de grandes volumes de informações, sejam obtidas por satélites ou por modelos numéricos. Satélites passaram a observar e distinguir entre áreas cada vez menores do oceano numa alta frequência. Diferentes tecnologias dentre as diferentes arquiteturas computacionais passam a permitir que oceano seja simulado utilizando-se as mais complexas e diferentes parametrizações matemáticas e em diversas resoluções espaciais e temporais. Quanto melhor os modelos numéricos são alimentados, parametrizados e inicializados, melhor será a reprodução da evolução espaço-temporal dos vários parâmetros e feições oceânicas.

Computação Aplicada ao Monitoramento das Massas de Gelo do Planeta

Prof. Dr. Jorge Arigony Neto (Universidade Federal do Rio Grande)

A palestra vai apresentar o histórico do uso no Brasil de métodos computacionais de laboratório e terreno, aplicados ao monitoramento de geleiras, gelo marinho e solos congelados.

Computação Aplicada aos Recursos Hídricos em Bacias Hidrográficas

Prof. Dr. Samuel Beskow (Universidade Federal de Pelotas)

Esta palestra abordará algumas aplicações da computação para o entendimento e resolução de problemas ligados a diferentes componentes do ciclo hidrológico visando à gestão de recursos hídricos em bacias hidrográficas. Serão contemplados aspectos referentes a técnicas computacionais aplicadas ao monitoramento e banco de dados hidrológicos, à representação e caracterização de bacias hidrográficas, à modelagem de

processos hidrológicos, à regionalização hidrológica, ao desenvolvimento e à otimização de modelos hidrológicos. Serão apresentados exemplos de aplicação da computação na área de recursos hídricos a fim de elucidar o grande potencial desta ciência para subsidiar a tomada de decisão em inúmeros problemas complexos de engenharia envolvendo, por exemplo, estiagens, cheias, erosão e transporte de sedimentos, reservatórios para diferentes finalidades, simulação do impacto de mudanças de uso do solo e mudanças climáticas, etc.

Trabalhos aceitos

Trabalhos Completos

Detecção do Molhamento Foliar por Meio de Análise de Imagens

Mauricio A. Z. Karrei (Universidade de Passo Fundo)
Renato W. Dallagasperina (Universidade de Passo Fundo)
Thiago Benvegnú (Universidade de Passo Fundo)
Rafael Rieder (Universidade de Passo Fundo)
Carlos A. Hölbíg (Universidade de Passo Fundo)
José Maurício C. Fernandes (Universidade de Passo Fundo)
Willingthon Pavan (Universidade de Passo Fundo)

Caracterização de Modelos para a Predição de Poluentes Atmosféricos

Maicon Nardino (Universidade Federal de Pelotas)
Jorge Ribes (Universidade Federal de Pelotas)
Elisia R. Correa (Universidade Federal de Pelotas)
Régis S. Quadros (Universidade Federal de Pelotas)
Willian S. Barros (Universidade Federal de Pelotas)
Daniela Buske (Universidade Federal de Pelotas)

Assessing the Risk of Extinction of Brazil's Flora: A Computational Approach Based on Microservices and Geospatial Analysis

Diogo Souza (Centro Nacional de Conservação da Flora)
Eline Martins (Centro Nacional de Conservação da Flora)
Eduardo C Dalcin (Instituto de Pesquisas Jardim Botânico do Rio de Janeiro)

Uma Ferramenta Computacional para Simulação de Espalhamento de Fluidos Baseada em Autômatos Celulares Bidimensionais Estocásticos

Marcos V. B. Lima (Instituto Federal do Triângulo Mineiro)
Cíntia C. Oliveira (Instituto Federal do Triângulo Mineiro)
Danielli A. Lima (Instituto Federal do Triângulo Mineiro)

SISDOC: Uma Biblioteca Digital Multimídia para Gestão de Documentos Ambientais

Hugo Feitosa de Figueirêdo (Instituto Federal da Paraíba)
Ruan Victor Amorim (Companhia Hidro Elétrica do São Francisco)
Cláudio de Souza Baptista (Companhia Hidro Elétrica do São Francisco)
Roberta Falcão de Cerqueira Paes (Universidade Federal de Campina Grande)
Brígida Duarte (Universidade Federal de Campina Grande)
Murilo Sérgio Lucena Pinto (Universidade Federal de Campina Grande)

Redes de Monitoramento Ambiental: Um Projeto Aplicado às Geleiras Patagônicas

Guilherme T. Netto (Universidade Federal do Rio Grande)
Jorge Arigony-Neto (Universidade Federal do Rio Grande)
Adenauer Correa Yamim (Universidade Federal de Pelotas)

Estratégia Evolutiva para Parametrização de Modelos de Previsão: Um Estudo de Caso com Níveis Máximos Mensais do Rio Xingu em Altamira/PA

Alen Costa Vieira (Universidade Federal do Pará e Centro Gestor e Operacional do Sistema de Proteção da Amazônia)

Gustavo Pessin (Universidade Federal do Pará e Instituto Tecnológico Vale)

Resumo

Mobile Application for Asian Soybean Rust Tracking in Brazil

José Henrique Andreis (Universidade de Passo Fundo)

Felipe Borella (Universidade de Passo Fundo)

Willington Pavan (Universidade de Passo Fundo)

Carlos Amaral Hölbig (Universidade de Passo Fundo)

Jaqson Dalbosco (Universidade de Passo Fundo)

Cláudia Vieira Godoy (Empresa Brasileira de Pesquisa Agropecuária)

José Maurício Fernandes (Universidade de Passo Fundo e Empresa Brasileira de Pesquisa Agropecuária)

Detecção do Molhamento Foliar por meio de Análise de Imagens

Mauricio A. Z. Karrei¹, Renato W. Dallagasperina¹, Thiago Benvegnú¹,
Rafael Rieder¹, Carlos A. Hölbig¹, José Maurício C. Fernandes¹, Willingthon Pavan¹

¹Instituto de Ciências Exatas e Geociências – Universidade de Passo Fundo (UPF)
99.052-900 – Passo Fundo – RS – Brasil

{122962, 71429, 105615, rieder, holbig, jmauricio, pavan}@upf.br

Abstract. *This paper presents the implementation of a leaf wetness detection system. Through the use of techniques for analysing and processing of digital images, we sought to identify and quantify the leaf wetness, as well as the leaf wetness duration. Moreover, a wireless sensor network was built to collect variables as temperature, relative humidity and precipitation. Preliminary tests demonstrated the viability and efficiency of the system, highlighting the use of NDVI images to enhance the quantification process. This will allow the correlation between the collected data in order to obtain results with higher accuracy.*

Resumo. *Este trabalho apresenta a implementação de um sistema capaz de detectar o molhamento foliar. Por meio da utilização de técnicas de análise e processamento de imagens digitais, buscou-se identificar e quantificar o molhamento foliar, bem como o tempo de duração deste. Além disso, foi construída uma rede de sensores sem fio capaz de coletar variáveis de temperatura, umidade relativa e precipitação. Testes preliminares comprovaram a viabilidade e a eficiência do sistema, destacando o uso de imagens NDVI para potencializar o processo de quantificação. Isso permitirá a correlação de informações, visando a obtenção de resultados com maior precisão.*

1. Introdução

A produção de alimentos é um dos grandes desafios mundiais na atualidade, e a agricultura é peça chave para a superação desses desafios. Desta forma, é de fundamental importância a implementação de novas tecnologias e ferramentas que otimizem o uso de insumos agrícolas e o manejo das culturas, contribuindo para a redução dos impactos ambientais, aumento da produtividade e, conseqüentemente, aumentando a lucratividade do produtor [Greenwade 1993].

A prática da agricultura automatizada e baseada em sensores, faz parte da agricultura de precisão, e está crescendo a cada dia, porém, seu avanço passa por questões como custo de implementação, resistência dos equipamentos em ambientes agressivos, manutenção e fornecimento de energia aos dispositivos eletrônicos em locais de difícil acesso [Rehman et al. 2014].

A germinação e a esporulação de fungos, bem como a infecção das plantas, são diretamente afetadas por fatores como a temperatura e a umidade do ar, a ação do vento e o molhamento foliar. Dentre estes parâmetros, o período do molhamento foliar, que

pode ser causado pela chuva, neblina, irrigação ou orvalho [Sentelhas et al. 2008], é um dos que mais impactam na relação entre o fungo e a planta, tornando esse um parâmetro de grande importância em modelos de simulação de doenças de plantas. Existem muitas formas de se realizar esta detecção, porém, algumas são falhas e outras são caras, além de existirem lacunas com relação a quantificação e normalização dos dados coletados [Rowlandson et al. 2015].

Nos últimos anos foram desenvolvidas diversas tecnologias na busca pelo monitoramento da duração do molhamento foliar (DMF) nas mais diferentes culturas [?]. Apesar dos resultados se mostrarem positivos, ainda há barreiras a serem superadas para ampliar a utilização deste artifício. Partindo do princípio de que um dos métodos mais utilizados e confiáveis para a determinação da DMF é a observação humana, é possível considerar que a visão computacional, por ser um método semelhante, é uma possibilidade interessante a ser explorada.

Entre os diversos dispositivos existentes no mercado, é possível destacar os sensores eletrônicos, que são amplamente utilizados. Estes sensores, simulam a folha e são capazes de detectar a presença de água em sua superfície, podendo ser resistivos [Inc. 2010] ou capacitivos [Devices 2010]. A implementação de sensores de umidade e temperatura, que possibilitam estimar o molhamento foliar por meio de modelos empíricos, também apresenta-se como uma das formas mais viáveis para a determinação da duração do molhamento foliar [Durigon and van Lier 2013, Leandro et al. 2003].

Estas tecnologias costumam ser suscetíveis a variações, de acordo com o tipo de planta analisado e as condições climáticas do local em que estão instalados. Os sensores de molhamento exigem cuidados como ângulo de instalação, posicionamento e diferença das características termodinâmicas entre as superfícies do sensor e da folha, enquanto os modelos empíricos necessitam de calibração e ajustes nos limites de parâmetros de interesse de acordo com as características do local de análise e a espécie de planta monitorada [Rowlandson 2011].

Nos últimos anos foram desenvolvidas diversas tecnologias na busca pelo monitoramento do molhamento foliar nas mais diferentes culturas [Sentelhas 2004]. Apesar dos resultados se mostrarem positivos, ainda existem barreiras a serem superadas para ampliar a utilização deste artifício. Partindo do princípio de que um dos métodos mais utilizados e confiáveis para a determinação do molhamento foliar é a observação humana, é possível considerar que a visão computacional, por ser um método semelhante, é uma possibilidade que deve ser explorada.

Devido aos constantes avanços tecnológicos, a captura e processamento remoto de imagens está se tornando cada vez mais acessível. O processamento de imagens apresenta vantagens em relação ao uso de sensores convencionais, como por exemplo o armazenamento das imagens capturadas para posterior análise e validação dos resultados, aumentando a confiabilidade nos dados gerados.

Assim, diante das informações expostas, é plausível considerar que o desenvolvimento de um sistema capaz de coletar imagens e variáveis ambientais, utilizando o processamento destas imagens e informações para determinar a presença de molhamento foliar oferece diversos benefícios a profissionais da área agrícola. Com isso, o objetivo deste trabalho é apresentar o desenvolvimento de um sensor capaz de detectar o molha-

mento foliar por meio de análise de imagens, além de uma rede de sensores sem fio de baixo custo, com o intuito de auxiliar no manejo de doenças de plantas.

2. Material e Métodos

Nesta sessão são abordados os conceitos fundamentais para a plena compreensão do projeto. São apresentadas as ferramentas de hardware utilizadas, bem como as bibliotecas de software que compuseram o desenvolvimento da solução.

2.1. Estrutura Geral do Sistema

Como peça principal para a realização deste trabalho, foi utilizado a placa de desenvolvimento Intel Galileo Gen2 (Figura 1(a)). Esta placa, é responsável por realizar as tarefas de coleta e processamento de imagens, bem como a aquisição dos dados provenientes da Rede de Sensores Sem Fio (Seção 2.4).

Para a captura das imagens utilizadas na identificação do molhamento foliar, é utilizada uma câmera com comunicação USB (Figura 1(d)), escolhida por sua compatibilidade com a placa Galileo.

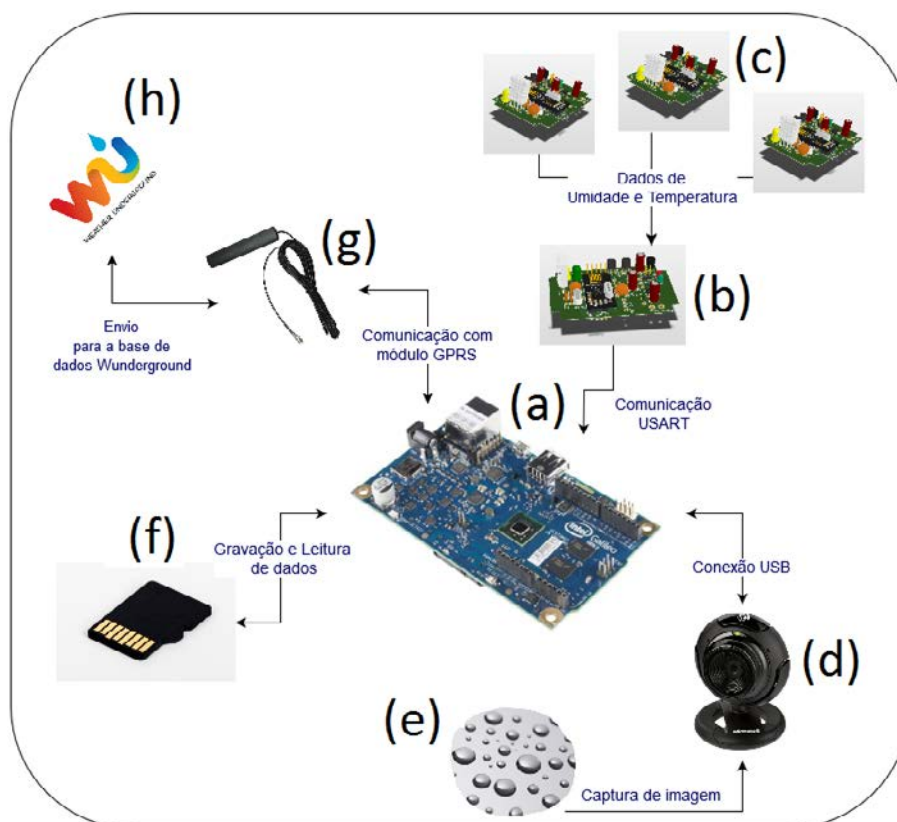


Figura 1. Diagrama em blocos do sistema proposto.

2.2. Captura de imagens

Para a realização da coleta das imagens, a câmera USB conectada à Galileo é ativada sempre que os dados coletados pela rede de sensores (Seção 2.4) indicam a probabilidade de existência do molhamento foliar.

O ambiente onde as imagens são capturadas, influencia diretamente no resultado do processamento digital destas. Muitas são as variáveis presentes no ambiente com potencial de interferência sobre o processo, algumas das quais podem ser citadas: iluminação, reflexo, cores de fundo, presença de movimento, ângulo da captura da imagem, entre outros. Para testes e validação inicial do sistema foi desenvolvido um ambiente controlado para minimizar a influência de algumas destas variáveis. Na Figura 2 é exibido o ambiente controlado construído.



Figura 2. Ambiente controlado construído para captura de imagens.

2.3. Superfície Alternativa

O ambiente natural possui grande influência sobre o resultado do processamento das imagens. Dessa forma, para eliminar o ruído causado pelo movimento da folha e pelos ruídos de fundo da imagem buscou-se uma superfície alternativa à folha da planta. Esta superfície deveria ser fixa, ser capaz de acumular água e ser suficientemente robusta para suportar condições adversas. Neste projeto, utilizou-se como superfície alternativa um fragmento de vidro jateado, onde sua opacidade com superfície seca, permite uma leitura mais próxima da cor branca, e no instante que a superfície possui acúmulo de água, o vidro jateado apresenta maior transparência.

Dessa forma, utilizou-se um material de cor preta colocado sob o vidro jateado. Assim, no momento em que ocorrer a presença de água livre sobre a superfície do vidro jateado, tornará visível o material posicionado sob o vidro. Caso a superfície do vidro não contenha água, a cor preta não terá tanto destaque e conseqüentemente não será identificada durante o processamento da imagem. Na Figura 3 é apresentado um fragmento do vidro utilizado. Nota-se que há um maior destaque da cor preta apenas na região com concentração de água.

2.4. Rede de Sensores sem Fio

Todos os nodos da rede de sensores sem fio possuem hardware idêntico, diferenciando-se apenas pela função exercida e pelo firmware embarcado. O nodo coordenador (Figura 1(b)) é responsável pelo recebimento dos dados coletados pelos demais nodos da



Figura 3. Superfície do vidro jateado, ressaltando a área molhada.

rede (Figura 1(b)), e transmissão destes dados para a Galileo. Cada nodo da rede possui sensores de temperatura e umidade relativa do ar.

Para a coleta de dados e também para aumentar a área de abrangência do sistema, foi desenvolvida uma rede de sensores sem fio. Todos os nodos são alimentados por baterias e células fotovoltaicas, garantindo sua autonomia energética. A rede utiliza a topologia estrela, sendo o nodo coordenador instalado junto a Galileo. O nodo coordenador possui, além do sensor de temperatura e umidade relativa do ar um sensor de precipitação.

2.5. Plataforma de recebimento de dados

Weather Underground ou Wunderground é um concentrador e fornecedor de dados meteorológicos, formado principalmente por estações meteorológicas particulares. Qualquer pessoa ou instituição pode enviar os dados coletados, em tempo real, para o Wunderground. Segundo informações do próprio serviço, atualmente, somente nos Estados Unidos existem mais de cem mil estações enviando informações, sendo mais de oitenta mil particulares, de "hobbyistas" ou entusiastas da área.

Assim, uma alternativa para disponibilizar as informações coletadas por este projeto ao público foi a utilização dos serviços oferecidos pelo Wunderground.

É utilizado um módulo GPRS, Figura 1(g), para o envio dos dados coletados à uma plataforma responsável pelo armazenamento, processamento e tratamento destes dados, também pode ser utilizado para enviar alertas via SMS. A plataforma usada para receber e disponibilizar os dados coletados ao público é a Weather Underground [und 2015], Figura 1(h). E ainda, todos os dados e imagens coletados são armazenados em um SD Card, Figura 1(f), ficando disponíveis para acesso quando necessário. A Figura 1 mostra o diagrama completo do sistema desenvolvido.

2.6. OpenCV

O processo de identificação do molhamento foliar, deu-se por meio da captura e processamento de imagens, utilizando-se de uma câmera USB conectada à plataforma Galileo. No que se refere ao método de análise e identificação da presença de água na estrutura montada (vidro jateado), foram realizados testes utilizando a biblioteca OpenCV.

Open Source Computer Vision Library (OpenCV) é uma biblioteca código fonte aberto que possui um conjunto de funções e recursos, possibilitando a construção de softwares de visão computacional. Nesta biblioteca estão incluídos um conjunto de algoritmos otimizados, que podem ser utilizados em aplicações de detecção e reconhecimento de faces/objetos e manipulação de imagens, tanto em 2D como em 3D [Cunha 2013].

O software foi implementado utilizando a linguagem de programação *Python* e o seu módulo "cv2", o qual é carregado com a utilização do pacote "python-opencv". Do módulo "cv2", foram utilizadas as funções *VideoCapture* e *inRange* para a manipulação das imagens. A primeira consiste em uma função responsável por realizar o interfaceamento do software desenvolvido com a câmera conectada à Galileo, oferecendo recursos para a captura da imagem da câmera. Já na função *inRange*, estão disponíveis os recursos necessários para analisar as imagens capturadas.

2.7. Armazenamento dos Dados Coletados

As informações provindas dos sensores de umidade relativa do ar, temperatura, precipitação, molhamento foliar, resultado do processamento da imagem e horário da coleta são transmitidas ao concentrador, e este, por sua vez, envia os dados para a Web por meio de uma conexão GPRS/3G ou por meio de uma conexão ethernet.

As imagens capturadas pelo sistema, além de serem processadas e analisadas, também são armazenadas em um cartão SD juntamente com os dados capturados pelos sensores, assim como informações do horário da coleta. Desta forma, quando necessário, o usuário pode acessar todas as informações diretamente do cartão SD.

3. Resultados

3.1. Rede de Sensores sem Fio

Os protótipos dos nodos da rede foram projetados e montados visando baixo custo e baixo consumo. Todos os nodos possuem hardware idêntico, diferenciando-se apenas pelo firmware embarcado. Para minimizar o consumo de energia, o sistema opera de forma que o nodo coordenador se mantém em modo de recepção constantemente, enquanto os demais nodos operam em modo *sleep*, despertando apenas para a coleta e envio de dados.

O sensor de temperatura e umidade relativa do ar precisou ser acomodado em uma estrutura especial, onde não ficasse exposto diretamente a luz solar, evitando que esta altere o valor lido pelo sensor. Na Figura 4 é apresentado o protótipo de um nodo da rede, com toda a estrutura necessária.

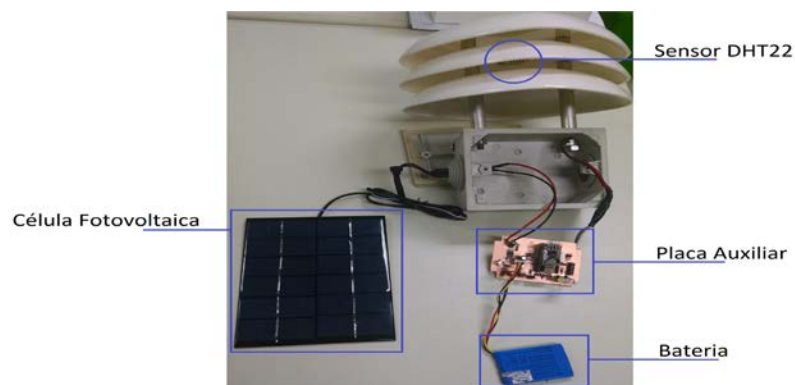


Figura 4. Nó da rede contendo todos os componentes utilizados.

3.2. Interpretação de Imagens

3.2.1. Detecção de Bordas

Inicialmente, realizaram-se testes utilizando técnicas de processamento de imagens para detecção de bordas, utilizando o método *Canny Edge Detector* da biblioteca OpenCv. O objetivo deste processo foi capturar a formação de água nas folhagens de plantas a partir da identificação do formato de sua borda na superfície da folha.

O uso desta técnica foi descartada para utilização neste projeto, visto que não apresentou bons resultados para a detecção da borda da gota d'água na superfície da folha. Devido a aspectos naturais do formato das folhas das plantas e também devido ao nível de transparência da água, nota-se que após a aplicação dos filtros de detecção de borda, o nível de falsa detecção obteve grande proporção. A Figura 5 apresenta o resultado da aplicação da técnica de detecção de borda.

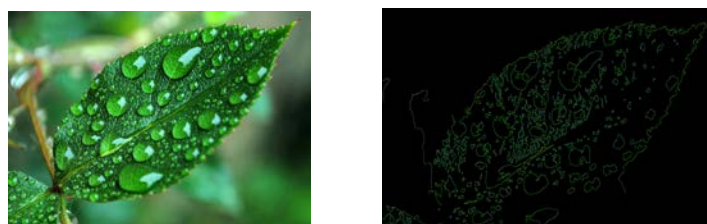


Figura 5. Identificação de formação de gotículas d'água utilizando técnicas de detecção de bordas.

3.2.2. Detecção de intervalos de cores

Em razão da dificuldade em realizar a detecção de água por meio da técnica de detecção de bordas, optou-se por utilizar a técnica de filtragem de cores. Para isso, a partir da obtenção da imagem da câmera USB no formato RGB, esta é convertida para o formato HSV com o auxílio da função da biblioteca OpenCV *cvtColor*. Esta conversão permite que os frames coletados sejam aplicados na função *inRange*, a qual recebe como parâmetro três valores: 1) *frame* coletado (imagem no formato HSV) e realiza o *thresholding* na imagem; 2) *array* com o intervalo baixo de cores (menor tonalidade da cor preta) a serem filtradas

na imagem; e 3) *array* contendo o intervalo máximo de cores (maior tonalidade da cor preta) a serem filtradas na imagem. O retorno desta função é uma imagem contendo a identificação dos pixels que contém a cor preta e também dos pixels que não contém a cor preta.

Este método de análise apresentou maior eficácia em comparação com o método de detecção de bordas, visto que as interferências na superfície da folha da planta, encontradas no primeiro método, ficaram inexistentes neste segundo método. Em consequência disso, foi possível identificar a presença de água na superfície do vidro, visto que no caso de existência de água, o vidro perde sua opacidade e ganha com um aspecto "transparente", tornando visível a cor preta que se encontra por detrás do vidro, permitindo assim, tornar viável a identificação de água sobre a superfície do vidro jateado. Esta identificação, é apresentada na Figura 6.



Figura 6. Processo de identificação da concentração de água no vidro jateado.

Além do exposto, o software desenvolvido realiza a quantificação de pixels, sendo possível, dessa forma, quantificar os pixels que são formados pela cor preta e os pixels formados por cores diferentes de preto.

3.3. Transmissão de dados

Para a transmissão dos dados coletados foram utilizados dois meios de comunicação: um via GPRS (telefonia móvel) e outro por meio de uma conexão Ethernet, sendo que ambos se mostraram eficazes. Os dados foram transmitidos para a plataforma *Weather Underground* [und 2015].

4. Discussões e Conclusões

A partir dos resultados apresentados, percebe-se que a identificação de molhamento foliar por meio de análise de imagens pode ser realizada. A superfície do vidro jateado, para esta situação, foi utilizada como alternativa à folha da planta e apresentou bons resultados.

Com a realização dos experimentos, constatou-se que fatores externos como intensidade de iluminação, ângulo da câmera e distância da câmera exercem forte influência sobre o funcionamento do sistema. Para tanto, nota-se que é preciso, primeiramente, dominar esses fatores em um ambiente controlado, para posterior aplicação e ajustes em ambiente aberto.

A quantificação do molhamento foliar por meio do processamento de imagens mostrou-se viável e eficiente. Além dos resultados obtidos, um novo trabalho está sendo

realizado, adaptando-se a câmara para funcionar com um filtro NDVI (*Normalized Difference Vegetation Index* [INSA 2008]). A Figura 7 apresenta alguns resultados preliminares obtidos utilizando-se o método do NDVI. Pode-se observar que a área das folhas da planta que possuem acúmulo de água ficaram destacadas em um tom de verde. As áreas onde existe acúmulo de água, mas não estão sobre a superfície das folhas da planta, obtiveram uma coloração na tonalidade azul. Além disso, nota-se que as áreas das folhas da planta onde não existe acúmulo de água, a cor vermelha e/ou amarela prevaleceu.

Este método também apresenta potencialidade de utilização para a identificação e quantificação do molhamento foliar. Com isso, destaca-se como fator positivo, a possibilidade da identificação da presença de água na superfície das plantas sendo feita diretamente no ambiente real, sem a utilização superfícies alternativas a folha das plantas.

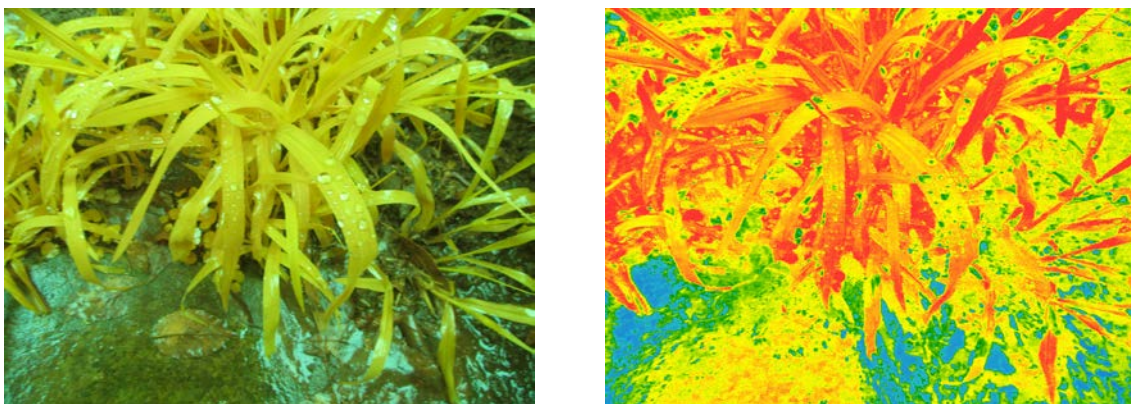


Figura 7. Imagens capturadas utilizando filtro NDVI.

Com o desenvolvimento deste trabalho, uma nova forma de identificar e quantificar o molhamento foliar foi determinada. Os resultados indicam que o uso do processamento digital de imagens é uma alternativa viável aos sensores de molhamento foliar presentes no mercado atualmente. Além disso, a utilização da rede de sensores sem fio permite extrair diferentes variáveis do ambiente, possibilitando a correlação de informações (como umidade do ar) a fins de validar e garantir maior precisão para a detecção do molhamento foliar.

Referências

- (2015). Weather underground.
- Cunha, A. L. B. N. d. (2013). *Sistema automático para obtenção de parâmetros do tráfego veicular a partir de imagens de vídeo usando OpenCV*. PhD thesis, Universidade de São Paulo.
- Devices, D. (2010). Leaf wetness dielectric sensor. Technical report.
- Durigon, A. and van Lier, Q. (2013). Duração do período de molhamento foliar: Medição e estimativa em feijão sob diferentes tratamentos hídricos. *Rev Bras Eng Agric Ambient*, 17:200–207.
- Greenwade, G. D. (1993). The Comprehensive Tex Archive Network (CTAN). *TUGBoat*, 14(3):342–351.

- Inc., C. S. (2010). Model 237 leaf wetness sensor. Technical report.
- INSA, I. N. d. S. (2008). Índice de vegetação por diferença normalizada.
- Leandro, L., Gleason, M., Nutter Jr, F., Wegulo, S., and Dixon, P. (2003). Influence of temperature and wetness duration on conidia and appressoria of *colletotrichum acutatum* on symptomless strawberry leaves. *Phytopathology*, 93(4):513–520.
- Rehman, A., Abbasi, A. Z., Zubair, N. I., and Shaikh, A. (2014). A review of wireless sensors and networks applications in agriculture. *Computer Standards and Interfaces*, 36:263–270.
- Rowlandson, T., Gleason, M., Sentelhas, P., Gillespie, T., Thomas, C., and Hornbuckle, B. (2015). Reconsidering leaf wetness duration determination for plant disease management. *Plant Disease*, 99(3):310–319.
- Rowlandson, T. L. (2011). *Leaf wetness: implications for agriculture and remote sensing*.
- Sentelhas, P. C. (2004). Duração do período de molhamento foliar: aspectos operacionais da sua medida, variabilidade espacial em diferentes culturas e sua estimativa a partir do modelo de penman-monteith. *Piracicaba: ESALQ/USP (Tese de Livre-Docência)*.
- Sentelhas, P. C., Marta, A. D., Orlandini, S., Santos, E. A., Gillespie, T. J., and Gleason, M. L. (2008). Suitability of relative humidity as an estimator of leaf wetness duration. *Agricultural and forest meteorology*, 148:392–400.

Caracterização de Modelos para a Predição de Poluentes Atmosféricos

**Maicon Nardino, Jorge Ribes, Elisia R. Correa, Régis S. Quadros,
Willian S. Barros, Daniela Buske**

Programa de Pós-Graduação em Modelagem Matemática (PPGMMat - UFPel)
Pelotas – RS – Brasil

{nardinomn, jlbribes, regisquadros, willianbarros,
danielabuske}@gmail.com, {elisiarc}@hotmail.com

***Abstract.** The objective of this study is based on statistical analysis of results obtained in the analytical development of the advection-diffusion equations using the combination of the Laplace transform and GILTT technique. The results obtained by four different models, using the Copenhagen experimental data, were analyzed. The performed statistical analyzes show that the proposed model is able to predict the concentration of pollutants using the Copenhagen experiment data set.*

***Resumo.** O objetivo do trabalho baseia-se na análise estatística dos resultados obtidos no desenvolvimento analítico das equações de advecção-difusão utilizando-se a combinação da Transformada de Laplace e da técnica GILTT. Os resultados obtidos por quatro modelos, utilizando os dados do experimento de Copenhagen, foram analisados. As análises estatísticas realizadas mostram que o modelo proposto é eficaz para a predição da concentração de poluentes utilizando o conjunto de dados do experimento de Copenhagen.*

1. Introdução

A qualidade do ar das cidades é um fator fundamental para a qualidade de vida da população e do meio ambiente. Devido a isso, diferentes grupos de pesquisa trabalham na modelagem de fenômenos atmosféricos e em sua simulação computacional. O grande desafio tem sido desenvolver ferramentas numéricas eficazes para estimar o nível de concentração de poluentes na atmosfera provenientes das inúmeras fontes de emissão, tanto antropogênicas como naturais. A partir da estimativa do nível de concentração próximo da fonte, pode-se realizar uma avaliação do impacto ambiental gerado e agir na tentativa de uma solução o mais conveniente possível para o problema.

A estatística descritiva compreende o conjunto dos métodos estatísticos para resumir e descrever as características proeminentes de conjuntos de dados referentes a observações de fenômenos particulares de uma classe de fenômenos. Nesse sentido, ela inclui a condensação de dados em forma de tabelas, sua representação gráfica e a determinação de indicadores numéricos de centro e de variabilidade [da Silva, 2004], [Pimentel-Gomes, 2009].

Este trabalho baseia-se na análise estatística dos resultados obtidos no desenvolvimento analítico das equações de advecção-difusão utilizando-se a combinação da Transformada de Laplace e da técnica GILTT (*Generalized Integral Laplace Transform Technique*). Esta técnica compreende os seguintes passos para a resolução de um problema bidimensional estacionário, por exemplo: solução de um problema associado de Sturm-Liouville, expansão da concentração de poluentes em uma série em termos das autofunções, substituição desta expansão na equação de difusão-advecção e finalmente tomar momentos. Esse procedimento nos leva a um conjunto de equações diferenciais ordinárias chamadas de problema transformado. O problema transformado é resolvido pela técnica da Transformada de Laplace e diagonalização [Wortmann et al., 2005]. A técnica GILTT é um método totalmente analítico, no sentido de que nenhuma aproximação é feita ao longo da sua derivação exceto pelo erro de truncamento.

A modelagem do problema se dá na Camada Limite Planetária (CLP) que é definida como a parte da troposfera que é diretamente influenciada pela superfície terrestre e responde aos forçantes superficiais com uma escala de tempo inferior a uma hora, efetuando-se uma análise estatística dos resultados. Aplicaremos alguns conceitos estatísticos de acordo com os resultados encontrados pré-conhecidos na literatura e encontrados em [Moreira et al., 2009] e [Buske et al., 2012].

Os resultados foram gerados pela solução analítica da equação de advecção-difusão do tipo: bidimensional estacionário, bidimensional transiente, tridimensional considerando uma Gaussiana na direção transversal e, tridimensional totalmente analítico, utilizando os dados do experimento de Copenhagen [Gryning and Lick, 1984] [Gryning et al., 1987].

Quanto às medidas de dispersão, o valor central do conjunto de dados, sendo uma estatística de posição, o desvio padrão e o coeficiente de variação. O desvio padrão (em dados brutos) e a variância são dados, respectivamente por:

$$S = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \text{ e } S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1},$$

sendo, que estes indicam, respectivamente, a distância que os valores se encontram da média e o quanto de variação o conjunto possui. Ainda, o coeficiente de variação, indica em termos percentuais, qual a variação dos dados, e é dado por:

$$CV = \frac{\sum_{i=1}^n S}{\bar{x}} \cdot 100$$

A partir dos dados obtidos, além da análise pela técnica GILTT, serão realizadas análises de *NMSE*, *COR*, *FB*, *FS*, além de média, variância, desvio padrão, correlação e distribuição de frequências, com a avaliação dos modelos com relação aos estimadores estatísticos analisados.

2. Modelagem Matemática do Problema

A equação de advecção-difusão da poluição do ar é uma representação da conservação de massa na atmosfera. Assumindo um fechamento Fickiano da turbulência, a equação de advecção-difusão pode ser escrita como [Buske, 2008]:

$$\frac{\partial c}{\partial t} + \bar{u} \frac{\partial c}{\partial x} + \bar{v} \frac{\partial c}{\partial y} + \bar{w} \frac{\partial c}{\partial z} = \frac{\partial}{\partial x} \left(K_x \frac{\partial c}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_y \frac{\partial c}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_z \frac{\partial c}{\partial z} \right) + S \quad (1)$$

onde, $c(x,y,z,t)$ representa a concentração média do poluente em g/m^3 , \bar{u} , \bar{v} , \bar{w} representam as componentes do vento médio em m/s nas direções x , y e z , respectivamente, K_x , K_y e K_z são as componentes cartesianas do coeficiente de difusão nas direções x , y e z , respectivamente e S é o termo fonte. Na primeira ordem de fechamento, todas as informações sobre a turbulência do processo estão contidas nos coeficientes de difusividade turbulenta.

A equação (1) está sujeita as condições de contorno de fluxo nulo nos contornos:

$$K_x \frac{\partial \bar{c}}{\partial x} = 0 \text{ em } x = 0 \text{ e } x = L_x \quad (2)$$

$$K_y \frac{\partial \bar{c}}{\partial y} = 0 \text{ em } y = 0 \text{ e } y = L_y \quad (3)$$

$$K_z \frac{\partial \bar{c}}{\partial z} = 0 \text{ em } z = 0 \text{ e } z = L_z \quad (4)$$

É considerado que a atmosfera está limpa no tempo inicial, ou seja, $c(x,y,z,0)=0$. A condição de fonte é representada por $\bar{c}(0,y,z,t)=Q\delta(y-y_0)\delta(z-H_s)$, onde Q é a taxa de emissão em g/s, z_i a altura da CLP em metros, H_s é a altura da fonte em metros, L_x e L_y são as distâncias longe da fonte e δ representa a função delta de Dirac.

A solução do problema é obtida aplicando-se a técnica da transformada integral na variável y , transformando o problema 3D em um problema 2D resolvido em [Buske et al., 2012]. Aplicando a transformada de Laplace na variável temporal, obtemos um problema estacionário, cuja solução é encontrada pelo método GILTT [Moreira et al., 2009].

3. Resultados

Os resultados preditos neste trabalho foram obtidos pelo método da GILTT à partir de dados coletados em Copenhagen e foi feita uma análise estatística quanto as medidas de posição e dispersão, tais como, média, mediana, desvio padrão, simetria, curtose, variância e coeficiente de variação e uma análise de correlação, além de uma análise de distribuição de frequência. Os detalhes do experimento de Copenhagen estão descritos em [Buske, 2008], [Gryning and Lick, 1984], [Gryning et al., 1987].

Na Tabela 1 podemos ver as concentrações observadas no experimento e preditas pelo modelo GILTT para os quatro casos estudados. Foram utilizados os dados de concentração integrados lateralmente para a equação bidimensional estacionária (aqui denominado GILTT-1), dados com uma maior resolução temporal para a equação bidimensional transiente (aqui denominado GILTT-2), dados tridimensionais para a equação tridimensional estacionária que aqui é resolvida assumindo-se que em y tem-se uma distribuição Gaussiana da pluma como em [7] (aqui denominado GILTT-3) e dados tridimensionais do modelo tridimensional analítico (aqui denominado GILTT-4). Podemos observar da Tabela 1 que o modelo GILTT descreve satisfatoriamente os dados observados no experimento de Copenhagen.

Tabela 1. Concentrações superficiais integradas lateralmente normalizadas pela taxa de emissão ($\cdot 10^{-4} \text{sm}^{-2}$), onde C_p são os dados preditos pelos modelos e C_o representa as concentrações observadas experimentalmente.

Dados C_{o1}	GILTT-1 C_{p1}	Dados C_{o2}	GILTT-2 C_{p2}	Dados C_{o3}	GILTT-3 C_{p3}	Dados C_{o4}	GILTT-3 C_{p4}
6,48	6,84	6,48	5,53	10,5	4,53	10,50	12,22
2,31	3,97	2,31	3,4	2,14	1,7	2,14	4,25
5,38	4,65	5,38	3,46	9,85	5,41	9,85	8,23
2,95	3,05	2,95	2,47	2,83	2,12	2,83	2,86
8,2	8,14	8,2	7,47	16,33	8,83	16,33	16,49
6,22	5,19	6,22	4,78	7,95	3,68	7,95	5,66
4,3	3,98	4,3	3,77	3,76	2,26	3,76	3,29
11,66	9,24	11,66	9,02	15,71	14,3	15,71	10,44
6,72	8,58	6,72	8,98	12,11	17,69	12,11	16,24
5,84	6,71	5,84	6,31	7,24	8,55	7,24	7,09
4,97	5,39	4,97	5,04	4,75	5,43	4,75	3,96
3,96	3,51	6,7	3,53	7,44	5,75	7,44	6,80
2,22	2,5	3,25	2,37	3,37	2,38	3,37	2,39
1,83	1,98	2,23	1,95	1,74	1,5	1,74	1,49
6,7	4,67	4,16	4,45	9,48	4,25	9,48	8,66
3,25	2,76	2,02	3,17	2,62	1,53	2,62	2,67
2,23	2,23	1,52	2,56	1,15	1,05	1,15	1,70
4,16	4,83	4,58	3,18	9,76	5,99	9,76	10,04
2,02	3,27	3,11	2,29	2,64	2,75	2,64	4,04
1,52	2,63	2,59	1,8	0,98	1,77	0,98	2,23
4,58	4,45			8,52	5,02	8,52	7,83
3,11	2,91			2,66	1,95	2,66	2,72
2,59	2,19			1,98	1,15	1,98	1,47

Na Tabela 2 apresentam-se os resultados dos índices estatísticos, descritos por [Hanna, 1989], [Pimentel-Gomes, 2009], e definidos como:

Erro quadrático médio normalizado (NMSE), dado por:

$$\frac{\overline{(C_o - C_p)^2}}{C_o C_p}$$

que informa sobre todos os desvios das concentrações do modelo e as concentrações observadas. É uma medida adimensional e quanto menor o valor, melhor o modelo.

Coefficiente de Correlação (COR):

$$\frac{\overline{(C_o - \overline{C_o})(C_p - \overline{C_p})}}{\sigma_o \sigma_p}$$

que descreve o grau de associação ou concordância entre as variáveis. O valor 1 garante uma boa performance.

Fator de dois (FA2), fração dos dados (% normalizados a 1), entre 0.5 e 2, ou seja, $0.5 \leq \frac{C_o}{C_p} \leq 2$. Quanto mais próximo de 1, mais próximos estão os dados.

Desvio fracional padrão (FS), é dado por:

$$\frac{\sigma_o - \sigma_p}{0.5(\sigma_o + \sigma_p)}$$

sendo, que valor ótimo é zero.

Fração de inclinação (FB), é dado por:

$$\frac{\overline{C_o} - \overline{C_p}}{0.5(\overline{C_o} + \overline{C_p})}$$

que indica a tendência do modelo de superestimar ou subestimar as concentrações observadas. O valor ótimo é zero.

Tabela 2. Avaliação estatística utilizando o experimento de Copenhagen para os quatro modelos GILTT -1, -2, -3 e -4.

	NMSE*	COR	FA2	FB	FS
GILTT -1	0,050	0,909	1,000	-0,005	0,142
GILTT -2	0,090	0,848	1,000	0,107	0,127
GILTT -3	0,330	0,802	0,870	0,281	0,088
GILTT -4	0,070	0,926	0,957	0,019	0,026

*NMSE: erro quadrático médio, COR: coeficiente de correlação, FA2: fator de 2, FB: fator de inclinação e FS: desvio padrão fracional.

Na tabela 3 apresentam-se as análises estatísticas descritivas realizadas, que são o objetivo principal deste artigo.

Tabela 3. Tabela de análises estatísticas dos dados de Copenhagen observados e preditos (Co₁, Cp₂) para experimento 1, 2, 3 e 4.

Parâmetros	Experimento 1		Experimento 2		Experimento 3		Experimento 4	
	Co ₁	Cp ₁	Co ₂	Cp ₂	Co ₃	Cp ₃	Co ₄	Cp ₄
Somatório	103,20	103,67	95,19	85,53	145,51	109,59	145,51	142,77
Média	4,48	4,50	4,75	4,27	6,32	4,76	6,33	6,21
Mediana	4,65	3,25	4,1	3,765	4,53	2,83	4,53	2,83
Var*	5,98	4,50	6,19	4,89	21,19	17,77	21,20	20,12
DP	2,39	2,07	2,42	2,14	4,50	4,12	4,50	4,39
Coef. Var (%)	42,51	67,48	48,67	64,31	84,56	81,45	84,56	70,68
Simetria	0,56	1,91	0,94	1,42	2,18	1,17	2,19	1,17
Curtose	-0,62	4,29	0,13	2,16	5,57	1,09	5,57	1,09
Quartis 1	3,27	2,31	2,56	2,31	2,38	2,14	2,38	2,14
Quartis 3	6,84	5,84	5,53	6,22	5,43	7,95	5,43	7,95
Correlação	0,9096		0,8016		0,8016		0,93	

*Var: variância, DP: desvio padrão e Coef. Var (%): coeficiente de variação.

4. Discussões

Analisando a Tabela 3, foram observadas algumas propriedades dos dados através resultados estatísticos. No primeiro conjunto de dados, no que diz respeito às medidas de tendência central, foram verificadas a média e a mediana. Calculando-se a média encontram-se $\bar{x} = 4.4870$ e $\bar{x} = 4.5074$ para os valores observados e preditos, respectivamente, identificados como Co_1 e Cp_1 para Co_2 e Cp_2 , da mesma forma, tem-se $\bar{x} = 4.7595$ e $\bar{x} = 4.2765$, $\bar{x} = 6.3265$ e $\bar{x} = 4.7648$, identificados como Co_3 e Cp_3 , e ainda $\bar{x} = 6.33$ e $\bar{x} = 6.21$, identificados como Co_4 e Cp_4 , estabelecendo-se assim o valor médio de cada conjunto de dados. Encontram-se também as medianas $Me = 4.65$ e $Me = 3.25$, $Me = 4.1$ e $Me = 3.765$, $Me = 4.53$ e $Me = 2.83$ e $Me = 4.53$ e $Me = 2.83$, respectivamente para $(Co_1$ e $Cp_1)$, $(Co_2$ e $Cp_2)$, $(Co_3$ e $Cp_3)$ e $(Co_4$ e $Cp_4)$, que determinam, em cada conjunto, o valor central.

Pelas medidas de dispersão (Tabela 3) observa-se que a variância para o experimento 1 foi de 5.9862 para os dados observados e 4.5087 para os preditos, no experimento 2, 6.1965 para os dados observados e 4.8942 para os preditos, no experimento 3, 21.1956 para os dados observados e 17.7735 para os preditos e no modelo 4, 21.20 para os dados observados e 20.12 para os preditos, o que é bastante alto nos quatro experimentos, mostrando que possuem realmente uma grande diferença entre os seus valores. Nota-se que a variância dos dados, observados e preditos nos dá a nítida percepção de que a variância dos primeiros é maior que a dos preditos em todas as observações. Quanto ao desvio padrão, tem-se para os dados observados e preditos, nos quatro experimentos, respectivamente, 2.3929 e 2.0767, 2.4263 e 2.1462, 4.5027 e 4.1232, 4.50 e 4.39, e na mesma ordem os valores percentuais aproximados respectivos ao coeficiente de variação, 42.51 e 67.48, 48.67 e 64.31, 84.56 e 81.45, 84.56 e 70.68, garantindo assim, que a técnica GILTT é eficiente, gerando bons resultados. Percebe-se também, que os dados observados e preditos possuem valores bastante semelhantes, o que nos remete a ideia da correlação, que demonstra a relação entre os dois conjuntos de dados (observados e preditos), ou seja, no presente trabalho, remete ao quão próximo estão os resultados do nosso modelo GILTT, com os valores reais observados, as magnitudes das correlações foram consideravelmente elevadas ($COR = 90,96\%$) no primeiro conjunto de dados e ($COR = 80,16\%$) no segundo e terceiro conjuntos e ($COR = 93\%$) no quarto conjunto de dados, que estão muito próximo de 100%, garantindo que os dados preditos estão perto dos observados, isto é, que o modelo GILTT de predição pode estimar com bastante precisão o que será observado em coletas futuras. Os valores para $NMSE$ são bastante pequenos, COR e $FA2$ tem valores próximos de 1 e FB e FS são próximos de zero, garantindo uma boa performance do modelo utilizado.

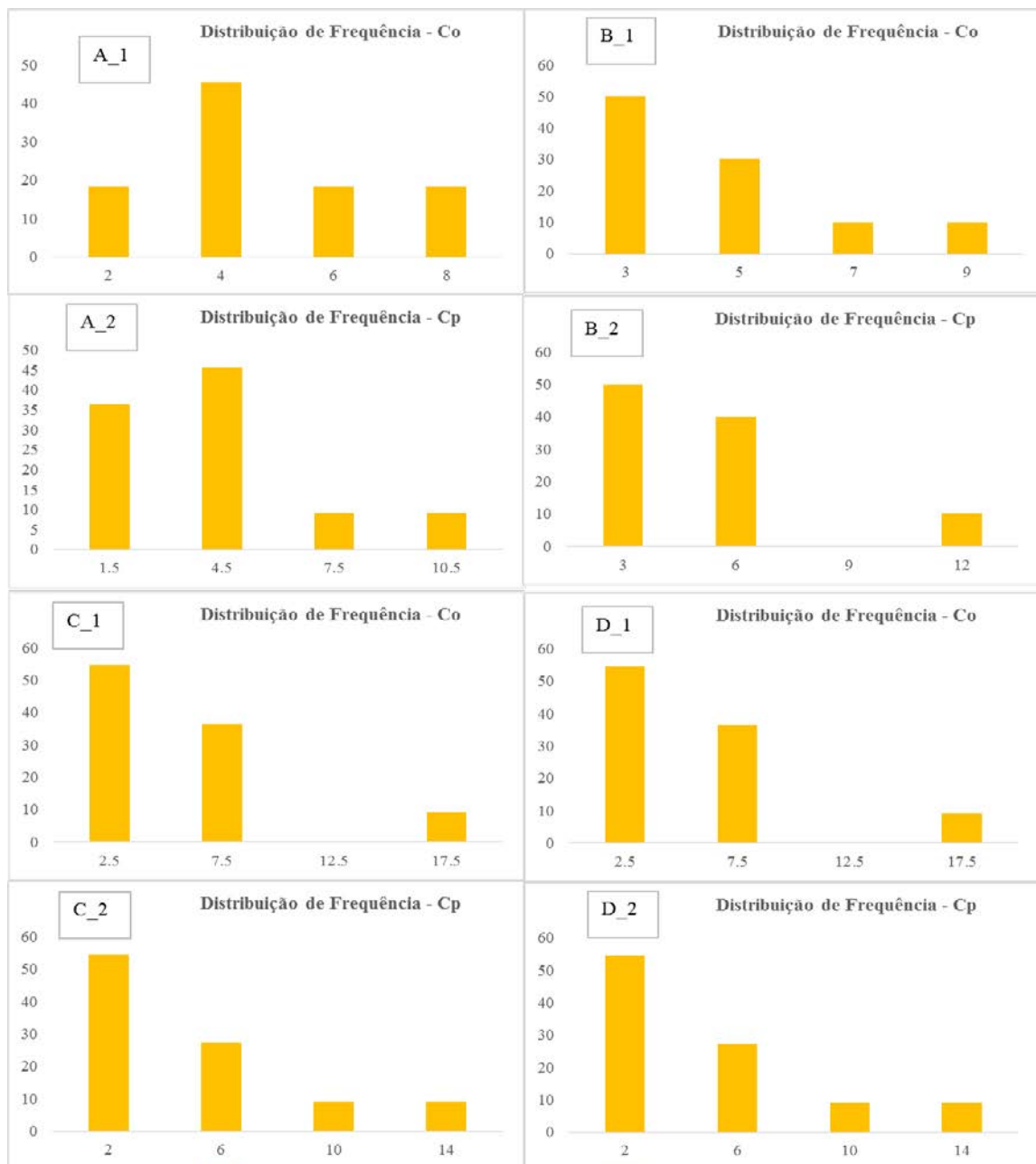


Figura 1. Análise da distribuição de frequência para os dados observados (A1 e preditos (A2) pelo modelo GILTT 1, GLITT 2 (B1 e B2), GLITT 3 (C1 e C2) e GLITT (D1 e D2).

Pode-se também observar os quartis, sendo o primeiro quartil (Q_1) o valor que divide o conjunto em duas partes, tais que um quarto ou vinte e cinco por cento (25%) dos valores sejam menores do que ele e três quartos ou setenta e cinco por cento (75%) dos restantes sejam maiores. O elemento que indica a ordem ou posição do primeiro quartil é determinado, para dados agrupados em classes, pela expressão $E_{Q_1} = n/4$, em que n é o número de valores do conjunto, ou número de observações. O terceiro quartil é o valor que divide o conjunto em duas partes tais que um quarto ou vinte e cinco por cento (25%) dos valores sejam maiores do que ele e três quartos ou setenta e cinco por cento

(75%) dos restantes sejam menores. O elemento que indica a ordem ou posição do primeiro quartil é determinado, para dados agrupados em classes, pela expressão $E_{Q_3} = 3n/4$, em que n é o número de valores do conjunto, ou número de observações.

Na Tabela 3, no primeiro conjunto de dados, verifica-se que o primeiro quartil tem valores 3.2700 para os valores observados e 2.3100 para os valores preditos, ou seja, 25% dos valores estão abaixo desses valores e que 75% deles estão acima, enquanto o terceiro quartil tem valores 6.8400 e 5.8400 para os valores observados e preditos, respectivamente, indicando assim que 75% dos valores estão abaixo e que 25% dos valores estão acima desses números. Indicando tal observação, tem-se no segundo conjunto de dados os valores 2.5600 e 2.3100 para o primeiro quartil para os dados observados e preditos, respectivamente e os valores 5.5300 e 6.2200 respectivamente, para os dados observados e preditos, para o terceiro quartil. Analogamente, no terceiro conjunto de dados tem-se 2.3800 e 2.1400 para o primeiro quartil e 5.4300 e 7.9500 para o terceiro quartil, bem como no quarto conjunto de dados tem-se 2.3800 e 2.1400 para o primeiro quartil e 5.4300 e 7.9500 para o terceiro quartil, na ordem dos conjuntos de dados anteriores.

Com relação a análise de distribuição de frequência, como pode ser visto na Figura 1, onde os dados obtidos através das coletas são indicados por $A1$, $B1$, $C1$ e $D1$ e os dados obtidos pelas técnicas são indicados por $A2$, $B2$, $C2$ e $D2$, observa-se que houve a formação de quatro classes para os dados observados e também preditos. A classe com maior percentual de observações para modelo GILTT-1 reúne aproximadamente 45% das informações, tanto observadas como preditas, para o modelo GILTT-2 a classe com maior percentual de observações reúne 30% das informações observadas e preditas, para o modelo GILTT-3 a classe com maior percentual de observações reúne em torno 35% das informações observadas e preditas e para o modelo GILTT-3 a classe com maior percentual de observações reúne aproximadamente 25% das informações observadas e preditas.

5. Conclusões

Utilizou-se neste trabalho a solução analítica para a equação de advecção-difusão obtida pelo método GILTT para quatro problemas, com modelagem do problema na CLP e efetuou-se uma análise estatística dos resultados. A estatística descritiva foi utilizada para resumir e descrever as características dos conjuntos de dados, fundamentando-se na análise estatística dos resultados obtidos no desenvolvimento analítico das equações de advecção-difusão combinando o método da Transformada de Laplace e da técnica GILTT com os dados do experimento de Copenhagen.

Nos quatro conjuntos de dados percebe-se que os valores de $NMSE$, COR , $FA2$, FS e FB levam-nos a acreditar que o modelo GILTT tem uma boa performance e mostrou-se confiável na predição da dispersão de poluentes atmosféricos, o que leva-nos a entender que se de grande importância e eficácia no auxílio à prevenção de futuros problemas ligados à dispersão de poluentes na atmosfera. Percebe-se também que os valores observados e preditos apresentaram grande variação para as diferentes coletas, indicando assim que tais coletas devem ser feitas em locais e horários diferenciados para uma melhor análise geral da dispersão de poluentes atmosféricos.

6. Referências

- Buske, D. (2008) Solução GILTT Bidimensional em Geometria Cartesiana: Simulação da dispersão de poluentes na atmosfera, Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul.
- Buske, D., Vilhena, M. T., Tirabassi, T. and Bodmann, B. (2012) “Air pollution steady-state advection-diffusion equation: the general three-dimensional solution”. *Journal of Envir. Protec*, 4:1-10.
- da Silva, J. G. C. (2004) “Estatística Básica”. Versão preliminar. Universidade Federal de Pelotas-UFPeL. Pelotas-RS.
- Gryning, S. E. and Lyck, E. (1984) “Atmospheric dispersion from elevated source in an urban area: comparison between tracer experiments and model calculations”. *Journal of Climate applied Meteorology*, 23: 651-654.
- Gryning, S. E., Holtslag, A.M.M., Irwin, J. and Sivertsen, B. (1987) “Applied dispersion modelling based on meteorological scaling parameters”. *Atmospheric Environment*, 21: 79-89.
- Hanna, S. R. (1989) “Condence limit for air quality models as estimated by bootstrap and jackknife resampling methods”. *Atmospheric Environment*, 23: 1385-1395.
- Moreira, D. M., Vilhena, M. T., Buske, D. and Tirabassi, T. (2009) “The state-of-art of the GILTT method to simulate pollutant dispersion in the atmosphere”. *Atmospheric Research* 92: 1-17.
- Pimentel-Gomes, F. (2009) “Curso de estatística experimental”. Piracicaba, v. 15: 451.
- Wortmann, S., Vilhena, M. T., Moreira, D. M. and Buske, D. (2005) “A new analytical approach to simulate the pollutant dispersion in the PBL”. *Atmospheric Environment*. 39 (12): 2187-2194.

Assessing the risk of extinction of Brazil's flora: A computational approach based on micro-services and geospatial analysis

Diogo Souza¹, Eline Martins¹, Eduardo C Dalcin²

¹Centro Nacional de Conservação da Flora - CNCFlora
Rua Pacheco Leão 915 - Rio de Janeiro - RJ - 22460-030

²Instituto de Pesquisas Jardim Botânico do Rio de Janeiro - JBRJ
Rua Pacheco Leão 915 - Rio de Janeiro - RJ - 22460-030

{diogo, eline}@cncflora.net, edalcin@jbrj.org

Abstract. *This paper describes a computational tool developed to assess the risk of extinction of flora according to the “B” of the IUCN Red List Categories and Criteria System. The tool consists of a set of systems arranged in a micro-services architecture and performs geospatial analysis in a significant set of data in an automated manner, with relatively low computational cost.*

Resumo. *Este trabalho descreve uma ferramenta computacional desenvolvida para avaliar o risco de extinção da flora brasileira segundo o critério “B” do sistema de critérios e categorias da União Internacional para a Conservação da Natureza e dos Recursos Naturais - IUCN. Esta ferramenta é formada por um conjunto de sistemas organizados em uma arquitetura de micro-serviços, e realiza análises geoespaciais em um conjunto significativo de dados, de forma automatizada, com relativo baixo custo computacional.*

1. Introduction

The extinction risk assessment is the first step in the conservation of a species, and should provide a scientific and objective assessment of the likelihood of a species becoming extinct at one time if the circumstances in which the species is found remains [Mace & Lande 1991]. In Brazil, to carry out the assessment of extinction risk, the system of criteria and categories of the International Union for Conservation of Nature was adopted - IUCN [IUCN 2001]. This system is composed of five quantitative criteria for a robust risk assessment and is scientifically based and can be applied consistently by different people, and also facilitates the comparison between assessments of a species or of different species [IUCN 2001].

The assessment of the risk of extinction of all known species is a global challenge agreed by signatories to the Convention on Biological Diversity (CBD) through the target 2 of the Global Strategy for Plant Conservation (GSPC): "an overall assessment of conservation status of all known plant to guide conservation actions" by 2020 [GSPC 2012]. Currently, Brazil has listed 46,113 species of flora [Flora of Brazil 2020 2016], but even with

a continuous and dedicated effort of the National Center for Plant Conservation (CNCFlora), since 2010, only about 11% of these were evaluated for the risk of extinction..

The remaining large knowledge gap about the conservation status of our flora, together with the challenge of evaluating it completely by 2020, shows that the use of technologies that allow rapid assessment of the risk of extinction supporting trained professionals to make decisions on the final categorization of species is essential.

In this context, this tool performs a risk assessment based on the criteria B of the IUCN [IUCN 2001], using the Extent of Occurrence (EOO), Area of Occupancy (AOO) and the number of subpopulations to categorize the risk of extinction. This criteria mainly uses the spatial information of the occurrence of the species and is the most widely used criteria for extinction risk assessments in Brazil [Martinelli and Moraes 2013].

Occurrence data of flora species comes mainly from herbarium collections, formed by specimens of dried plants mounted on a sheet of cardboard accompanied by a label with data relating to that sample. These data represent the occurrence of a biological specimen in time and space and are the primary source of data for studies on biodiversity and conservation.

In the last decade, significant investments were made to render the data from scientific collections digitally accessible [Beaman et al. 2012, Blagoderov et al. 2012]. More recently, in Brazil, an initiative of the Ministry of Science, Technology and Innovation (MCTI) has promoted the digitalization and provided access to data from collections from different national herbaria [Gadelha et al. 2014, SiBBR 2016]

The initiative of MCTI adopted the Darwin Core standard [Wieczorek et al. 2012] and the Integrated Publishing Toolkit (IPT) [Robertson et al. 2014] for data publication [Gadelha et al. 2014]. Likewise, the Rio de Janeiro Botanical Garden Research Institute (JBRJ) also offers its occurrence data and Species List Flora of Brazil using the same standards and tool.

The access to those data sets provided by the IPT publishing tool, added with the demand of the National Center for Conservation of Flora - CNCFlora to assess the risk of extinction of all flora, served as motivation and made possible the development of this tool.

2. Methodology

2.1. Data gathering

Two data sets are consumed by the tool in their risk assessment process: the *taxonomic set*, which is a list of accepted scientific names and their synonyms, and the *set of occurrences* of these names.

The taxonomic set is formed based on the resource "Lista de Espécies da Flora do Brasil", version 393.53 [Forzza 2014], made public by the IPT tool at <http://ipt.jbrj.gov.br/jbrj/>. In this version, the set recorded 39,506 accepted names and 18,376 synonyms for these names, totaling 57,882 scientific names.

The occurrences set is formed by made public instances of resources in the following IPTs:

- Inst. de Pesquisas Jardim Botânico do Rio de Janeiro, available at <http://ipt.jbrj.gov.br/jbrj/>
- Projeto REFLORA, available at <http://ipt.jbrj.gov.br/reflora/>

- Centro de Referência em Informação Ambiental - CRIA, available at <http://ipt1.cria.org.br/ipt/>

The occurrences set then totaled 3,632,660 records related to the names present in the taxonomic set mentioned above.

2.2. Analysis Performed

With the collected data sets - taxonomic and occurrences - the tool performs a series of calculations, explained below, for each species, represented then by its accepted name and associated synonyms.

In a first step, the tool selects, for each species, two subsets of occurrence data, based on the date of the sample. The historical subset (M) contains samples collected for over 50 years; and the recent subset (R) contains samples collected in the last 50 years. For each of these subsets, the tool selects for the following calculations only records that have geographic coordinate pairs (latitude and longitude).

In a second step, the tool performs the calculations of the geospatial Extent of Occurrence and Area of Occupancy, as described in IUCN 2012. The Extent of Occurrence (EOO) is the area contained within the shortest continuous imaginary boundary (*minimum convex polygon*) which can be drawn to encompass all known, inferred, or projected sites presently occupied by the taxon and the Area of Occupancy (AOO) is the area within the EOO which is *actually* occupied by the taxon (usually measured by overlaying a grid and counting number of occupied cells). In addition to these two calculations, the tool performs the calculation for subpopulations using the method of "circular buffer" [Rivers et al. 2010].

In its final stage, for each species, the tool confronts the results of calculations performed within these parameters (EOO, AOO and subpopulations) considered by the criterion "B", which is one of the five criteria to define the possible categories and justification of risk assessment in accordance of the methodology established by IUCN (Table 1):

Table 1. Possible results of risk assessments using the IUCN Red List Categories and Criteria - IUCN 2001

Category	Criterion	Justification
{CR,EN,VU}	Geographic range in the form of Extent of occurrence (EOO)- criteria B1	The size of EOO
{CR,EN,VU}	Geographic range in the form of Area of occupancy (AOO)- criteria B2	The size of AOO
{NT, LC}	N/A	The size of EOO and AOO
DD	N/A	Insufficient number of occurrences

CR, EN and VU are threatened categories mean critically endangered, endangered and vulnerable, respectively. NT and LC are categories of risk assessment, but not a threatened categoria, NT: not threatened and LC: least concern. DD: data deficient, means the specie has so few information and it is not possible to evaluate its.

3. Architecture and Technologies

With an architecture based on micro-services [Lewis & Fowler 2014], the tool is divided into different systems, which are independent and interchangeable. The main advantage of this approach is the possibility of independent development of the parties, which have loose coupling.

The systems are based on web services and communicate with each other through APIs using the HTTP standard transport protocol and the JSON format for encapsulating the information.

The Taxadata system, developed in PHP 5.7 programming language, and using as persistence the SQLite database, is responsible for importing and treating the taxonomic data set, as well as exposing the data as a Web service (Figure 1)..

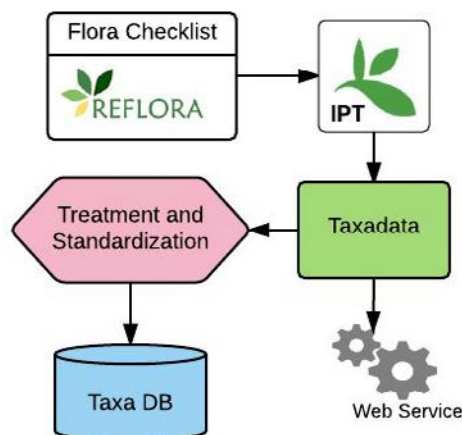


Figure 1. Taxadata system scheme

To aggregate the occurrence data, the Darwin Core Bot system was developed. This can import resources from IPTs, and also expose the set of occurrences as an API web service. This system was developed in Clojure programming language and uses the SQLite database to store the data (Figure 2).

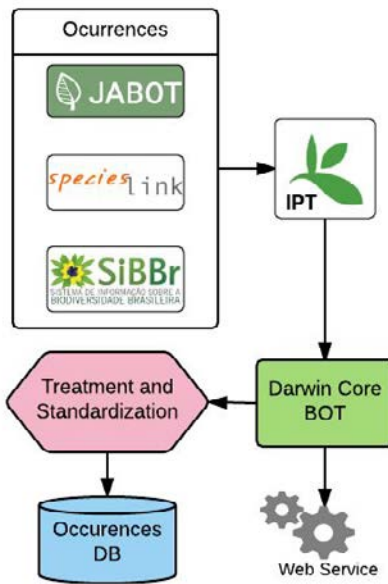


Figure 2. Darwin Core Bot system scheme

The spatial calculations and risk assessment are made through a web service without persistence (Stateless) of the Darwin Core Services tool, developed in Clojure. This system is responsible for receiving a set of events and perform the calculation and evaluation (Figure 3).

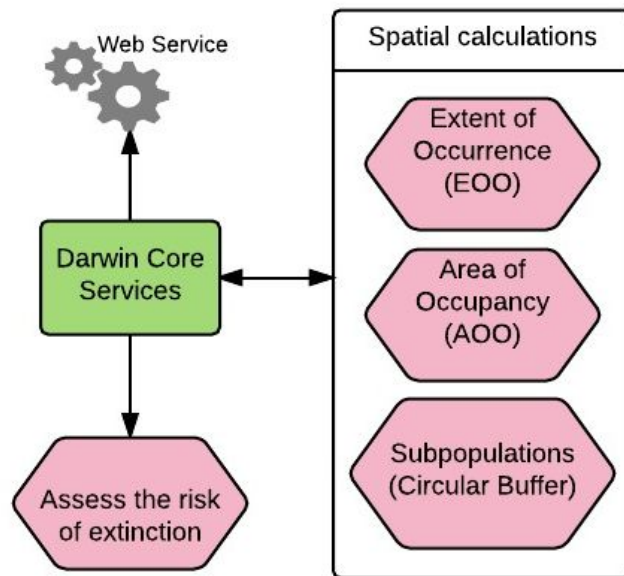


Figure 3. Darwin Core Services system scheme

Another system that is part of the tool is the Biodiv-idx. This system is responsible for consolidating the data of calculations and evaluations made possible by previous systems,

storing this data in a database and in the "ElasticSearch" search engine. This tool was also written using the Clojure programming language (Figure 4).

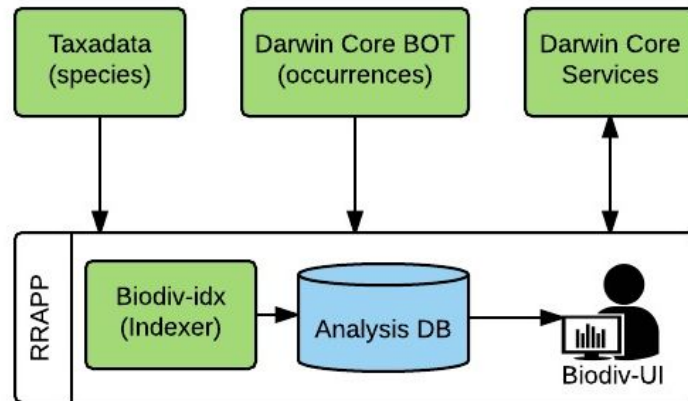


Figure 4. Figure showing the systems and the tool “Rapid Risk Assessment Application” - RRAPP

Finally, the Biodiv-UI system, developed with the PHP 7 language, provides a search interface and displays statistics on the calculations and risk assessments carried out (Figure 5).

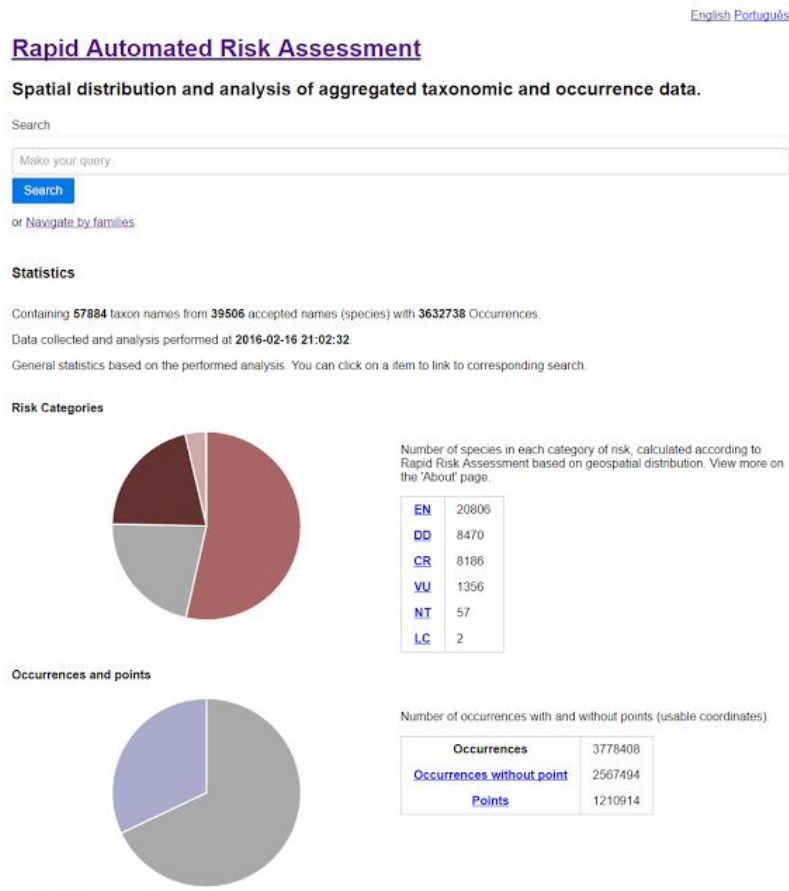


Figure 5. The web interface of the “RRAPP” Tool (partial)

All systems are deployed in a single virtual server with 4GB of RAM, 4 processing cores, and a 40GB SSD, where 30GB of which are allocated. The systems run in isolation and is kept in Docker containers.

3. Results

Regarding the performance of the systems, we recorded the following results:

- The creation of taxonomic set by Taxadata system totaling 57,882 names, was performed in 15 seconds;
- The composition of the occurrences set by the Darwin Core Bot system, totaling 3,632,660 occurrences was performed in 30 minutes;
- The calculation and assessment by Darwin Core Services and Biodiv-idx systems of a total of 39,506 species were conducted in 11 hours and 22 minutes.

Regarding the data analysis, the following results were recorded:

- The results of the analysis show that no occurrence records were found for 2,871 species (7.3%);
- 36,365 species (92%) have occurrence records;
- 3,907 (10.7%) species have up to 2 records;
- 8,578 species (21.7%) do not have occurrence records with valid coordinates;
- 6,363 species (16.1%) have only 1 or 2 records with valid coordinates.

These analyses also indicate that 8,470 species (21.4%) are "data deficient" (DD) and 30,348 species (76.8%) are in the category of "threat" (vulnerable, endangered or critically endangered).

The analyses also are available for the entire set of Brazilian flora species and also for each family and each species individually.

All technologies used are open source, free, and available for use, collaboration, and modification by third parties at <https://github.com/diogok/biodiv-compose>. The results are available on the portal at <http://rrapp.jbrj.org> or <http://rrapp.cncflora.net>.

4. Conclusion

The increasing availability of open data on biodiversity offered on internationally accepted standards enables the emergence of new analytical techniques that seek to explore and identify patterns that would not be apparent otherwise [Kelling et al. 2009]. These new techniques rely on implementing computing solutions capable of manipulating and analyzing a massive amount of data with a computational cost compatible with the scale of analysis proposed.

In this work, we demonstrate a tool that follows this new paradigm, to assist the achievement of the second goal of the Global Strategy for Plant Conservation - GSPC for 2020: An assessment of the conservation status of all known plant species, as the far as possible, to guide conservation action.

The analyses also show that the spatial and taxonomic qualification of scientific collections deposited in the herbaria is vital to the quality of the results of the analyses which use these sources of information.

As future work, we include: (i) the development of a "GBIF-bot", able to integrate the set of occurrences offered by the Global Biodiversity Information Facility - GBIF; (ii) the development of a data quality system in order to select more thoroughly the occurrences with geospatial coordinates; (iii) a comparison of the risk assessments made by the tool to the evaluations conducted by CNCFlora professionals, using the criterion B of the IUCN, so as to analyze the tool efficiency.

References

- Beaman, R. S. and Cellinese, N. (jan 2012). Mass digitization of scientific collections: New opportunities to transform the use of biological specimens and underwrite biodiversity science. *ZooKeys*, v. 17, n. 209, p. 7–17.
- Blagoderov, V., Kitching, I. J., Livermore, L., Simonsen, T. J. and Smith, V. S. (jan 2012). No specimen left behind: industrial scale digitization of natural history collections. *ZooKeys*, v. 146, n. 209, p. 133–46.
- Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Disponível em: < <http://floradobrasil.jbrj.gov.br/> >. Acesso em: 29 Fev. 2016
- Forzza R (2014): Brazilian Flora Checklist - Lista de Espécies da Flora do Brasil. v393.53. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro. Dataset/Checklist. doi:10.15468/1mtkaw
- Gadelha, L., Guimarães, P., Moura, A. M., et al. (2014). SiBBr: Uma Infraestrutura para Coleta, Integração e Análise de Dados sobre a Biodiversidade Brasileira. VIII Brazilian e-Science Workshop (BRESCI 2014). Proc. XXXIV Congress of the Brazilian Computer Society,
- GSPC (Global Strategy for Plant Conservation). (2012). Global Strategy for Plant Conservation. Botanic Gardens Conservation International, Richmond, Inglaterra. 38p.
- IUCN (International Union for Conservation of Nature and Natural Resources). (2001). The International Union for Conservation of Nature and Natural Resources. Guidelines for Application of IUCN Red List Criteria at Regional levels: Version 3.0.
- IUCN. (2012). IUCN Red List Categories and Criteria: Version 3.1. Second edition. Gland, Switzerland and Cambridge, UK: IUCN. iv + 32pp.
- Kelling, S., Hochachka, W. M., Fink, D., et al. (2009). Data-Intensive Science: A New Paradigm for Biodiversity Studies. *BioScience*, v. 59, n. 7, p. 613–620.
- J. Lewis and M. Fowler (2014). Microservices. Disponível em <http://martinfowler.com/articles/microservices.html>. Acessado em 01/03/2016.
- Mace G.M. & Lande, R. (1991). Assessing extinction threats: towards a re-evaluation of IUCN threatened species categories. *Conservation Biology*, n. 5, p. 148-157.
- Martinelli G. & Moraes M.A. (2013) Livro Vermelho da Flora do Brasil. Andrea Jakobsson Estúdio: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro.

- Rivers, M. C., Bachman, S. P., Meagher, T. R., Lughadha, E. N. e Brummitt, N. A. (2010). Subpopulations, locations and fragmentation: Applying IUCN red list criteria to herbarium specimen data. *Biodiversity and Conservation*, v. 19, n. 7, p. 2071–2085.
- Robertson, T., Döring, M., Guralnick, R., et al. (6 ago 2014). The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. *PLoS ONE*, v. 9, n. 8, p. e102623.
- SiBBr - Sistema De Informação Sobre a Biodiversidade Brasileira. Disponível em <http://www.sibbr.gov.br/>. Acessado em 26/02/2016.
- Wieczorek, J. R., Bloom, D., Guralnick, R. P., et al. (6 jan 2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, v. 7, n. 1, p. e29715.

Uma ferramenta computacional para simulação de espalhamento de fluidos baseada em autômatos celulares bidimensionais estocásticos

Marcos V. B. Lima¹, Cíntia C. Oliveira¹, Danielli A. Lima¹

¹Instituto Federal do Triângulo Mineiro (IFTM) – Campus Avançado Patrocínio
Av. Lúcia Terezinha Lassi Capuano nº 255 – 38.740-000 – Patrocínio – MG – Brazil

marcoslimaptc@gmail.com, cintiaoliveira@iftm.edu.br, danielli@iftm.edu.br

Abstract. *One of the main causes of water pollution are the mineral extraction and fossil fuels from soil or even oceans. The intense search for these natural resources for creating derivatives grown in large scale, increasing, consequently, their exploitation. However, its leak in nature can cause serious problems to the humans and animals lives. In this context, this paper presents a cellular automata study and its applications, and proposes a model which simulates a fluid slick model and then presents a prototype system to support these simulations. Experiments and tests were conducted to examine how the variables involved affect the fluid spreading in the simulated environment.*

Resumo. *Uma das principais causas de poluição da água está na extração de minerais e combustíveis fósseis do solo ou até mesmo de oceanos. A busca intensa por esses recursos naturais para criação de produtos derivados cresce em larga escala, aumentando, conseqüentemente, a sua exploração. Entretanto, o seu vazamento na natureza pode causar sérios problemas à vida dos humanos e animais. Neste contexto, o presente trabalho apresenta um estudo dos autômatos celulares e suas aplicações, e propõe um modelo para realizar simulações de espalhamento de fluidos, e posteriormente apresenta um protótipo do sistema para suportar essas simulações. Experimentos e testes foram realizados para analisar como as variáveis envolvidas afetam o espalhamento de fluidos dentro do ambiente simulado.*

1. Introdução

Estudos relacionados ao espalhamento de fluidos auxiliam principalmente as indústrias a quantificarem os níveis de espalhamento de fluidos tóxicos sobre uma determinada região, bem como o estudo de técnicas de remediação [Lehikoinen et al. 2015]. No caso do petróleo ou de outro rejeito tóxico, o espalhamento pode ocorrer através da elevação natural ou artificial devido à perfuração de poços artesianos [Mondol et al. 2015]. Assim, a excessiva extração desses materiais causam sérios prejuízos naturais na presença de um acidente, afetando, principalmente, a qualidade da água, causando a degradação do solo, prejudicando a vida dos animais.

O objetivo de um modelo para a previsão do espalhamento de fluidos é a determinação da evolução da poluição de um rejeito ou resíduo frente ao tempo em um ambiente simulado. Direção do vento ou correnteza, direção e velocidade das ondas são fatores determinantes num vazamento de grande porte. Esses modelos matemáticos

e computacionais podem ser parte de um sistema real que combinam um sistema de informação geográfica (GIS) e imagens de satélite com um modelo de espalhamento de fluidos [Mohamadi et al. 2015]. Entretanto, esta abordagem envolvendo processamento de imagens de satélite não será investigada no presente trabalho.

Uma ferramenta útil na elaboração de simulações gerenciadas por computador são estabelecidas por modelos matemáticos conhecidos como Autômatos Celulares (ACs). Existem diversas aplicações no uso de ACs, dentre elas, podemos citar a modelagem de fenômenos naturais [Duff et al. 2015], ambientais [Lima and Lima 2014], físicos [Jian-hua et al. 2015], biológicos [Wolfram 2002] ou sociais [Castro and Lima 2013], que seriam muito difíceis de serem modeladas pelas equações diferenciais, sendo estas as mais utilizadas nesse tipo de tarefa. Além disso, os ACs possuem alto nível de paralelismo, quando implementados em um hardware paralelo.

O presente trabalho tem como objetivo: (i) introduzir o conceito de autômatos celulares e apresentar suas características, (ii) implementar o método de simulação do espalhamento de fluidos baseado em autômatos celulares bidimensionais a partir de adaptações de modelos precursores [Ha et al. 2012, Karafyllidis 1997, Dilão 1993]. A fim de avaliar o modelo sugerido, (iii) testes em diferentes situações foram realizados. As simulações fornecem o tempo de espalhamento total e a partir de várias simulações é possível obter a média geral do tempo de espalhamento de resíduos no ambiente aquático. Além disso, um protótipo do sistema como um todo foi elaborado e novos parâmetros vêm sendo adicionados para aumentar o realismo e a precisão do modelo apresentado neste trabalho.

2. Fundamentação teórica e trabalhos relacionados

2.1. Autômatos celulares

Um AC é composto por um reticulado com uma dimensão d dividido em células ou unidades processadoras, sendo que, cada célula C é representada por um estado. As células modificam seus estados a cada passo de iteração de acordo com uma regra de transição. Podemos aplicar a regra de transição por T passos de tempo para obter a evolução espaço-temporal do reticulado do AC. A regra estabelecida por uma função de transição indica o novo símbolo a ser escrito na célula do reticulado de acordo com seu estado atual e dos estados de suas vizinhas (regra local). Em sua definição mais usual, a atualização dos estados se dá de forma síncrona e utiliza uma regra determinística, isto é, a cada passo de tempo todas as N células do reticulado são atualizadas. No modelo proposto neste trabalho, o sistema de regras é probabilístico, ou seja, é possível alterar o estado de uma célula a partir de uma função de probabilidade, isso significa que as regras de atualização são probabilísticas e, portanto, o estado de uma célula em um determinado instante depende probabilisticamente dos estados de suas células vizinhas no passo de tempo imediatamente anterior [Lima and Lima 2014].

A estrutura dos ACs unidimensionais (1D) é a mais estudada. Para um AC com regra de atualização determinística, a mudança de estado de uma célula depende de m vizinhas expressas por $m = (2r + 1)$, sendo r o raio do AC [Oliveira 2003]. Para ilustrar um AC unidimensional com regra de atualização determinista, considere a Figura 1 (a), que aborda uma modelagem conhecida como regra 30 [Wolfram 2002], contendo um reticulado de 6 células sendo que o estado inicial de cada célula é apresentado em $t = 0$. Uma regra binária de raio 1 é aplicada e a vizinhança de cada célula é formada

por três elementos: a própria célula e suas duas vizinhas adjacentes (à esquerda e à direita). Como esse AC é binário (2 estados possíveis), existem 8 diferentes vizinhanças, da 000 a 111. A regra em si é dada pelos 8 bits de saída associados a cada vizinhança possível: 01111000. Na Figura 1 (b) vemos a atualização do reticulado de $N = 6$ por 2 passos de tempo a partir de sua configuração inicial 101110 em $t = 0$. A cada passo, cada célula do reticulado é atualizada identificando-se sua vizinhança e seu novo estado é dado pelo bit de saída correspondente na regra de transição. Observe como exemplo, a célula de símbolo 1, destacada em $t = 0$, seu próximo estado será 0 em $t = 1$. O reticulado é submetido a condições periódicas de contorno, sendo que a primeira célula é vizinha imediata da última, e vice-versa. Aplicando-se esse procedimento para todas as células do reticulado de forma síncrona, tem-se a nova configuração do reticulado a cada passo de tempo [Castro and Lima 2013].

Os ACs bidimensionais (2D) são amplamente conhecidos na literatura

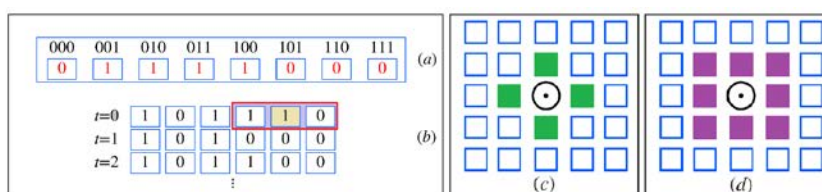


Figura 1. (a) Regra de transição de raio 1. (b) Evolução do AC por $T = 2$ passos de tempo. (c) Vizinhança de Von Neumann. (d) Vizinhança de Moore [Lima and Lima 2014].

[Oliveira 2003]. Essa dimensão contempla variadas formas para a consideração da vizinhança de uma célula, sendo que as duas principais estão apresentadas na Figura 1 (c) e (d), tal que o símbolo \odot representa a célula central e o símbolo \blacksquare identifica a célula vizinha. Dentre as aplicações de ACs 2D, o modelo mais conhecido é o jogo matemático proposto por Conway em 1970 denominado *Game of Life* apresentado na Figura 2.

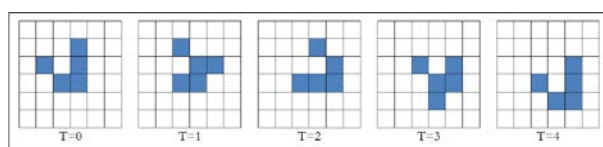


Figura 2. Evolução do AC “Game of Life” por $T = 4$ passos de tempo.

2.2. Modelos de espalhamento de fluidos baseados em Autômatos Celulares

Um dos trabalhos conhecidos nessa temática foi o estudo para modelagem de derramamento de fluidos através de um modelo microscópico proposto por [Dilão 1993]. Neste trabalho, o autor definiu o comportamento de espalhamento de fluidos, baseando-se no estudo de caso de derramamento no estuário Tejo. Para isso, o autor construiu um programa dedicado a partir de fotos capturadas por satélite. Para alcançar o realismo necessário, regras baseadas em Autômatos Celulares e Lattice Gas foram utilizadas num cenário microscópico. O AC apresentou um reticulado bidimensional com células no formato triangular e as partículas deslocam-se no decorrer das ligações entre essas células por um período de tempo constante. As colisões entre essas partículas foram resolvidas de

maneira não-determinística.

Um modelo para a predição de derramamento de óleo e sua movimentação e espalhamento foi proposto em [Karafyllidis 1997], utilizando-se Autômatos Celulares. O modelo proposto neste trabalho foi capaz de prever derramamento de óleo em regiões com limites complexos. Os efeitos de vento, correnteza e evaporação do óleo foram considerados. Um algoritmo para a simulação de derramamento de óleo, bem como regiões geográficas hipotéticas foi considerado. Os resultados da simulação ficaram em conformidade com o real movimento e espalhamento de derramamento de óleo.

No trabalho de [Ha et al. 2012] um modelo macroscópico foi proposto para analisar o fenômeno de espalhamento de óleos em oceanos quando o mesmo foi poluído por um navio de extração. O espaço ocupado pelo oceano foi convertido em células, que continham informações, tais como, quantidade de óleo, temperatura do oceano e a direção do vento. Dois estados denominados: limpo e poluído são definidos por cada célula. O óleo espalha nas células vizinhas através de regras de atualização. No trabalho foram empregadas 3 regras de atualização. A primeira considera que o óleo apenas espalha para uma célula x_i a partir da vizinhança dessa célula, se existe uma célula vizinha x_j com mais óleo que a célula x_i . A segunda considera que o óleo evapora numa proporção de acordo com a temperatura do oceano a cada passo de tempo. A terceira considera que a propriedade do óleo é afetada pela direção e velocidade do vento. A simulação mostrou que o modelo de cada célula é definido pela combinação de eventos discretos e que o tempo discreto da simulação foi capaz de avaliar um ambiente com 300×300 células com 10^3 kl em uma determinada corrente de vento.

Mais tarde, no trabalho de [Vourkas and Sirakoulis 2012], uma modelagem mais complexa sobre a utilização dos ACs, que são inerentemente paralelos, foram aplicadas com êxito para a modelagem do derramamento de resíduos tóxicos. Nesse trabalho, os autores definem a área do espalhamento como uma matriz de células de quadrados idênticos, e o tamanho das células do AC é dado entre a precisão de cálculo e os recursos computacionais necessários, para realização dos cálculos em tempo real. Além disso, uma implementação do modelo de autômatos celulares (ACs) foi apresentada na plataforma FPGA (Field Programmable Gate Arrays). Foram avaliadas as semelhanças entre os ACs e a plataforma FPGA. Posteriormente, foram discutidos os principais desafios da implementação, bem como o mapeamento apropriado e favorável do hardware para a exploração das características do ACs. Adicionalmente, foi realizada uma significativa avaliação de desempenho em comparação com as soluções de software e hardware correspondentes.

3. Modelo proposto

O modelo proposto é fortemente baseado nos modelos de [Dilão 1993], [Karafyllidis 1997] e [Ha et al. 2012]. Um modelo inicial básico foi criado, posteriormente mais parâmetros foram sendo agregados ao modelo a fim de proporcionar realismo às simulações realizadas com intuito experimental. As simulações apresentadas neste trabalho têm o intuito de mostrarem uma visualização sob o ponto de vista macroscópico. A seguir, o modelo será apresentado, e será detalhada cada modificação que foi realizada em relação aos modelos precursores.

No modelo aqui proposto considera-se uma rede bidimensional com simetria quadrada, vizinhança de Moore e cada célula da malha representa o estado de uma área/região do ambiente. No trabalho de [Ha et al. 2012] consideram-se dois estados

possíveis para uma célula: o estado associado à cor azul representa uma área de água intacta, à cor preta uma área que foi fortemente contaminada. Entretanto, no modelo aqui apresentado outros estados foram acrescentados. Esta modificação foi inspirada no modelo de [Hernandez et al. 2007] que também subdividiu o problema em outros estados a fim de dar mais realismo à simulação. A Figura 3 (a) e (b) apresenta a nova divisão do estado de poluição no modelo precursor de [Ha et al. 2012].

A regra de atualização proposta neste trabalho é probabilística e periódica e os estados da malha são atualizados a cada iteração de forma síncrona, sendo que a mesma é baseada do modelo de [Ha et al. 2012]. A ideia geral da regra de atualização será descrita a seguir. Se num certo instante uma célula está no estado azul (porção de água não contaminada) e uma das células vizinhas está no estado de contaminação, existe uma probabilidade não nula de, no instante seguinte, a célula mudar para o estado preto (Figura 3 (b) cor 1). A cada $t = 2$ o estado da célula é alterado incrementando-se a cor (ver Figura 3 (b) cor 2 até cor 4). Ou seja, quando uma célula está no estado de contaminação, ela vai manter-se nesse estado durante um tempo finito (neste caso, $t = 12$). Se uma célula da malha está no estado azul claro (porção de água levemente contaminada), então ficará nesse estado indefinidamente ($t = \infty$) ou até que sofra um processo de evaporação definido por uma constante de probabilidade de evaporação β . Ou seja, é possível que a partir de uma quantidade de evaporação (que varia de acordo com o fluido estudado), a porção contaminada volte para o estado intacto novamente. Uma vez que cada célula tem uma probabilidade $p_{ij} = \beta$ de evaporar. Ao evaporar, essa célula diminui a concentração da contaminação, ou seja, vai para o próximo estado da Figura 3. Se existir vento ou direção de correnteza numa certa direção, a probabilidade de transição azul \rightarrow para o estado de poluição é maior nessa direção. Assim, define-se o caráter não-determinista na evolução do autômato celular, similar ao que já foi abordado em [Dilão 1993]. No modelo aqui apresentado, a partir do momento que a célula entra no estado de contaminação, cada estado muda de acordo com o tempo para acontecer o espalhamento ($t = 2$), enquanto que no modelo de [Hernandez et al. 2007] essa mudança era dada por uma regra estocástica. Uma primeira adaptação que considera que uma célula (i, j) tem a probabilidade $p_{ij} = x$ de se contaminar, sendo que $x \in [0, 1]$. Ou seja, basta uma célula ter uma vizinha no estado contaminado que ela tem uma probabilidade não nula de no próximo instante de tempo passar do estado intacto para o estado contaminado, diferentemente do modelo de [Ha et al. 2012], onde a dispersão era realizada a partir de diferenças de quantidades entre células vizinhas.

Em relação ao fator vento abordado em [Karafyllidis 1997] regras estatísticas foram usadas. Neste trabalho, não será utilizado o parâmetro vento. Um parâmetro bem similar será usado que é a direção e velocidade da correnteza, para essa modelagem uma matriz de preferências de dimensão $m \times m$ foi usada. Essa matriz representa a vizinhança de Moore, adotada em modelos de ACs para dinâmica de partículas. De acordo com [Pereira et al. 2011], essa matriz indica uma maior preferência da partícula para uma determinada direção, mostrando-se muito eficaz para modelar o fator velocidade e direção da correnteza. Essa matriz poderá ser adaptada para modelar outros parâmetros do sistema, tais como, vento e densidade. A proposta considera uma matriz de preferências, tal que essa dimensão representa a vizinhança do AC com raio $r = 1$. Neste caso, a matriz de preferências indica que a velocidade da correnteza é de $v_{max} = 1$. Ou seja, $v_{max} = \sum_{i=1}^m x_{ij}$ é a soma dos pesos das células pertencentes à vizinhança m ($x_{ij} \in m$).

A matriz de preferências utilizada no modelo pode ser vista na Figura 3(c). Caso seja necessário alterar a direção do vento, basta rotacionar a matriz para a direção desejada, conforme proposto nos algoritmos de dinâmica de partículas [Pereira et al. 2011]. A probabilidade de infecção de uma área intacta é dada por $p_{ij} = \sum_{i=1}^m \sum_{j=1}^m mat_{ij} \times \beta$, sendo que mat_{ij} é a matriz de preferências e β é a constante de probabilidade de evaporação.

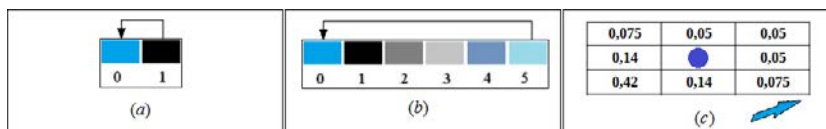


Figura 3. Cores correspondes aos estados das células: (a) modelo proposto por Ha (2012) adaptado para este problema. (b) modelo proposto neste trabalho. (c) Matriz de preferências para raio 1. A seta azul indica a direção da correnteza para a matriz de preferência e o círculo azul representa uma região intacta.

4. Resultados experimentais

Os testes apresentados nesta seção foram elaborados utilizando-se a linguagem de programação C++ e foram divididos em duas partes: a primeira trata os testes preliminares para validação do comportamento da regra de atualização e a segunda os testes de análise temporal, que analisa estatisticamente o tempo gasto para que a espalhamento se propagasse por todo o reticulado. Cada experimento avalia cada um dos parâmetros possíveis do modelo proposto neste trabalho.

4.1. Análise visual

Inicialmente foram realizados alguns testes para a verificação da formação de padrões na evolução do espalhamento de fluidos sem a utilização da matriz de preferências e foram contrastados com o modelo de [Ha et al. 2012]. Os experimentos representam um cenário onde um reticulado de dimensões 80 × 80 foi considerado. Além disso, um foco de contaminação inicial é colocado na célula central do reticulado no instante $t = 0$.

Primeiramente, uma simulação no tempo $t = 5$ é apresentada na Figura 4 (a) e mostra o espalhamento ainda em condições preliminares à sua difusão ao longo do reticulado. Na Figura 4 (b) tem-se a simulação do modelo de [Ha et al. 2012] em $t = 20$ sem a constante de evaporação β , e na Figura 4 (c) tem-se a simulação do mesmo modelo também em $t = 20$, mas considerando-se $\beta = 0.05$. Nesta primeira análise, a probabilidade de contaminação ($p_{ij} = 0.7$) é a mesma, quando tem-se uma ou mais células vizinhas contaminadas. Também é possível observar que com o passar do tempo, tem-se que ao centro, a tonalidade da água é reestabelecida e a poluição é espalhada radialmente, já que o modelo é o sem vento. Na Figura 4 (d) tem-se a alteração do aumento de estados da poluição em relação ao modelo precursor no instante de tempo $t = 20$, considerando-se $\beta = 0.05$ e $p_{ij} = 0.7$. Neste caso, o estado contaminado apresentou cinco subestados, sendo que a alteração para o estado seguinte ocorre a cada $t = 2$ passos de tempo. Conforme é observado, existe um atraso na recuperação da tonalidade da água, que é ocasionado devido ao tempo do estado “contaminado” em sua totalidade ser maior ($t = 12$) em relação ao modelo-base para a construção deste trabalho [Ha et al. 2012], que utilizou apenas dois estados. A Figura 4 (e) representa a simulação no instante $t = 35$, e a formação radial do espalhamento já encontra-se mais acelerada. Esse fato é ocasionado devido à constante

de evaporação $\beta = 0.05$ ser muito pequena. Além disso, uma formação radial pode ser observada em relação à proporção e formato da área contaminada, o que é garantido pela simulação do modelo com ausência de vento ou correnteza.

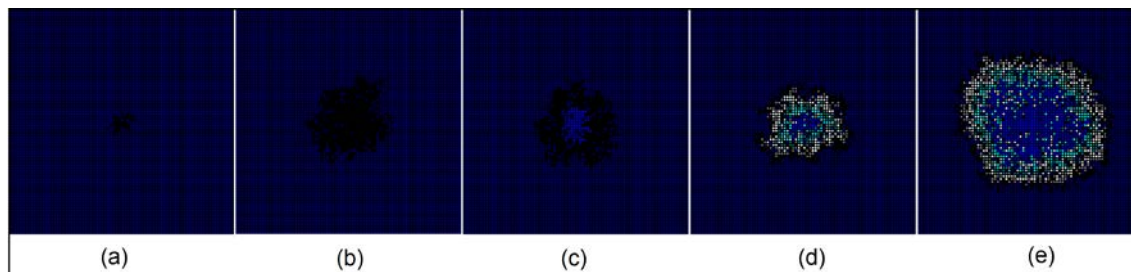


Figura 4. Observações da simulação do modelo baseado em ACs para o espalhamento de fluidos. (a) Configuração no instante $t = 5$. Configuração no instante $t = 20$: (b) Modelo baseado em [Ha et al. 2012] sem evaporação, (c) Modelo baseado em [Ha et al. 2012] com evaporação e (d) Modelo aqui proposto. (e) Configuração no instante $t = 35$ do modelo aqui proposto.

4.2. Análise estatística do tempo de propagação

Para validar o modelo sob o ponto de vista matemático, foram realizadas simulações com presença de correnteza, utilizando-se a matriz de preferências apresentada na seção anterior. Os experimentos avaliam reticulados de diferentes dimensões e formatos. Cada um desses reticulados apresenta um único foco de contaminação inicial ($t = 0$), na célula de posição central do reticulado. Foram realizadas 100 simulações para cada um dos parâmetros propostos e observou-se os tempos (T) de propagação gastos para que toda a área do reticulado fosse contaminada. Portanto, para que o critério de parada fosse atingido, a constante de evaporação β não foi utilizada. Histogramas de classe foram plotados para a validação dos tempos T de cada um dos modelos (Figura 5).

Os histogramas das Figuras 5 (a) e (b) referem-se ao experimentos com reticulados quadrados de dimensão 50×50 e 80×80 . Para o experimento referente ao histograma da Figura 5 (a), tem-se média $\bar{x} = 53.19$, mediana $\tilde{x} = 53.17$ e moda $mo = 52.43$. Para o experimento referente ao histograma da Figura 5 (b) os valores médios são maiores devido ao tamanho do reticulado que impacta diretamente neste resultado. Neste caso, tem-se média $\bar{x} = 69.36$, mediana $\tilde{x} = 68,83$ e moda $mo = 69.08$. Nestes dois primeiros gráficos é possível perceber que a distribuição dos dados referentes às simulações apresentam uma melhor distribuição, o que torna essa configuração de ambiente mais significativa.

Os histogramas das Figuras 5 (c) e (d) referem-se ao experimentos com reticulados não quadrados de dimensão 30×50 e 50×80 , respectivamente. Para o experimento referente ao histograma da Figura 5 (c), tem-se média $\bar{x} = 40.17$, mediana $\tilde{x} = 39.89$ e moda $mo = 39.02$. Para o experimento referente ao histograma da Figura 5 (d), os valores médios são maiores, devido ao tamanho do reticulado que impacta diretamente neste resultado. Neste caso, tem-se média $\bar{x} = 48.28$, mediana $\tilde{x} = 47.45$ e moda $mo = 46.94$. Em relação aos experimentos com reticulados quadrados, é possível observar que a distribuição naqueles é mais próximo à distribuição simétrica, quando comparados com os experimentos de reticulados não quadrados. Como é possível observar, os resultados referentes aos experimentos com reticulados não quadrados apresentam distribuições assimétricas positivas. As distribuições assimétricas são características de

resultados onde $\bar{x} > \tilde{x} > mo$. Além disso, os valores anotados de cada simulação são menores em relação aos dos experimentos anteriores, isso deve-se principalmente ao fato da dimensão do reticulado e formato. Reticulados com formato quadrado proporcionam maior velocidade ao espalhamento. Uma vez que, o reticulado utilizado neste trabalho não possui contorno periódico. Essa ausência de periodicidade no reticulado estimula a criação de barreiras por ausência de vizinhança para a aplicação da regra de transição, sendo esta o motor de toda a simulação.

Pode-se observar que a partir destes gráficos as distribuições das adaptações reali-

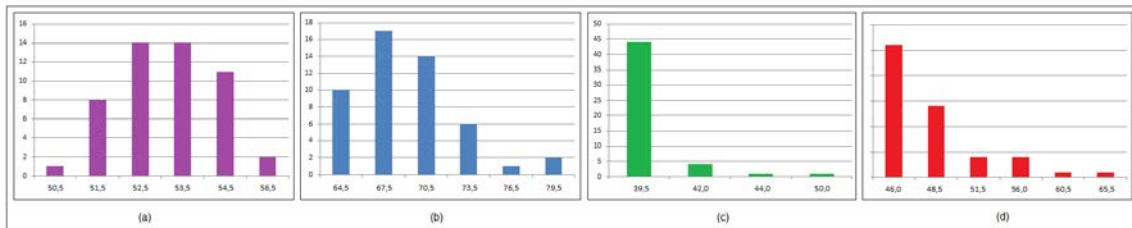


Figura 5. (a) Histograma de classes para os reticulados de dimensões (a) 50 × 50. (b) 80 × 80. (c) 30 × 50. (d) 50 × 80.

zadas para reticulados quadrados possuem maior confiança em relação aos experimentos realizados com reticulados não quadrados. Uma atenção especial para os gráficos de experimentos realizados com reticulados não quadrados é que estes apresentam mais valores que estão distantes da média. Esse resultado mostra que determinadas configurações apresentam mais dificuldade para que o ambiente fosse totalmente coberto pela contaminação.

4.3. Protótipo do sistema de simulação de espalhamento de fluidos

Esta seção apresenta o conjunto de interfaces propostas neste trabalho para a criação do sistema de simulação de espalhamento de fluidos, bem como a breve descrição de suas funcionalidades. A ferramenta inicialmente foi desenvolvida em linguagem C++ e será importada para as interfaces do Java juntamente com as opções de configuração do espalhamento. Na Figura 6 é possível observar a tela proposta para o sistema de simulação com as opções de configuração do espalhamento. No canto superior esquerdo, temos a representação do ambiente aquático ainda intacto (antes do espalhamento começar). No canto superior direito, um painel para escolha do tipo de fluido, bem como o símbolo de representação do mesmo. Cada fluido possui por padrão valores automáticos relativos à sua densidade, viscosidade, massa e peso, com exceção da opção “Outro” que permite ao usuário alterar o nome e os parâmetros de um novo fluido. Na parte inferior da tela, encontramos as demais opções de configuração do espalhamento, onde são reguladas as dimensões do ambiente, o tempo inicial e final do espalhamento (T), e a intensidade de correnteza e direção do vento. A opção “Estatísticas” exibe a tela apresentada à direita e nela valores dos dados inseridos são mostrados, tais como média, moda e mediana, assim como outros valores que serão agregados posteriormente.

5. Conclusões e trabalhos futuros

Neste trabalho foi investigado um modelo de autômatos celulares bidimensionais estocásticos fundamentado em modelos precursores aplicados à simulação de espalhamento de fluidos. Para a modelagem do AC detalhado no trabalho, considerou-se



Figura 6. Modelo de tela proposto para o ambiente da simulação em funcionamento e modelo de tela proposto para exibir os resultados estatísticos obtido pelas simulações.

a implementação descrita por [Ha et al. 2012], [Dilão 1993] e as complementações idealizadas por [Karafyllidis 1997]. Todas as adaptações lógicas foram concebidas com a análise desses modelos precursores. Além disso, uma análise das características do modelo desenvolvido foi abordada.

Com a execução do modelo descrito foi possível gerar dados visuais, onde os comportamentos padrões de espalhamento de fluidos puderam ser observados. Além disso, dados estatísticos foram obtidos relacionando-se a média das iterações com as adaptações de cada um dos parâmetros de dimensão e formato do reticulado. A grande vantagem de se usar o modelo aqui desenvolvido, reside no fato de que o mesmo é flexível para a entrada dos parâmetros já adotados e de novos que poderão ser inseridos. Entretanto, não é possível garantir que apenas os comportamentos inseridos por si o já comprovam o com exatidão a área contaminada a partir das simulações do modelo. Mas, elas iniciam uma análise que pode ser aprofundada em um trabalho de pesquisa posterior.

Como continuidade desse trabalho, seria interessante a adaptação do modelo 2D para um modelo 3D, considerando-se neste caso profundidade de contaminação, tornando a modelagem mais realista e precisa. Uma adaptação para o espalhamento de outros tipos de fluidos em diferentes meios também poderia ser incorporada ao trabalho. Para alcançar o paralelismo e simulações mais rápidas do modelo poderia ser utilizada uma implementação em hardware paralelo, uma vez que toda a motivação para o emprego de ACs na simulação reside na possibilidade de uma implementação eficiente, similar ao trabalho de [Vourkas and Sirakoulis 2012]. Além disso, um estudo de caso considerando-se um espalhamento de fluidos real poderia ser realizado, a exemplo, o espalhamento de rejeitos tóxicos na cidade de Mariana-MG em 2015.

Referências

Castro, A. P. and Lima, D. A. (2013). Autômatos celulares aplicados a modelagem de dinâmica populacional em situação de risco. In *4th Workshop of Applied Computing for the Management of the Environment and Natural Resources WCAMA 2013 CSBC*, Universidade Federal de Alagoas.

- Dilão, R. (1993). Autômatos celulares, máquinas de turing ou a natureza como máquina de cálculo. *Colóquio Ciências*, 12:3–20.
- Duff, T. J., Chong, D. M., and Tolhurst, K. G. (2015). Using discrete event simulation cellular automata models to determine multi-mode travel times and routes of terrestrial suppression resources to wildland fires. *European Journal of Operational Research*, 241(3):763–770.
- Ha, S., Cha, J.-H., Ku, N.-K., and Lee, K.-Y. (2012). Cell-based discrete event and discrete time simulation for the prediction of oil slick movement and spreading in ocean environment. *Transactions of the Society of CAD/CAM Engineers*, 17(1):45–53.
- Hernandez, L., Hoya, S., Martan, A., and Rodriguez, G. (2007). Modelling forest fire spread using hexagonal cellular automata. *Applied Mathematical Modelling*, 31(6):1213 – 1227.
- Jian-hua, Z., Tao, J., Sheng-an, W., and Jia-wei, M. (2015). Research of cellular automata traffic flow model for variable traffic flow density. *Traffic*, 10(20):26–25.
- Karafyllidis, I. (1997). A model for the prediction of oil slick movement and spreading using cellular automata. *Environment international*, 23(6):839–850.
- Lehikoinen, A., Hanninen, M., Storgard, J., Luoma, E., Mantyniemi, S., and Kuikka, S. (2015). A bayesian network for assessing the collision induced risk of an oil accident in the gulf of finland. *Environmental science and technology*, 49(9):5301–5309.
- Lima, H. A. and Lima, D. A. (2014). Autômatos celulares estocásticos bidimensionais aplicados à simulação de propagação de incêndios em florestas homogêneas. In *5th Workshop of Applied Computing for the Management of the Environment and Natural Resources WCAMA 2014 CSBC*, Universidade de Brasilia.
- Mohamadi, B., Xie, Z., and Liu, F. (2015). Gis based oil spill risk assessment model for the niger deltas vegetation. *Nature Environment and Pollution Technology*, 14(3):545.
- Mondol, M. R., Keshavmurthy, S., Lee, H.-J., Hong, H.-K., Park, H.-S., Park, S.-R., Kang, C.-K., and Choi, K.-S. (2015). Recovery of wild pacific oyster, *crassostrea gigas* in terms of reproduction and gametogenesis two-years after the hebei spirit oil spill accident off the west coast of korea. *Continental Shelf Research*, 111:333–341.
- Oliveira, G. M. B. (2003). Automatos celulares: aspectos dinâmicos e computacionais. *III Jornada de Mini-cursos em Inteligência Artificial (MCIA) - Sociedade Brasileira de Computação*, 8:297 – 345.
- Pereira, L. A., Duczmal, L. H., and Cruz, F. R. (2011). Simulação de evacuação emergencial via autômatos celulares: Uma proposta de modificação do modelo de schadschneider. In *XXXII Congresso Nacional de Matemática Aplicada e Computacional*, pages 692–698.
- Vourkas, I. and Sirakoulis, G. C. (2012). Fpga based cellular automata for environmental modeling. In *Electronics, Circuits and Systems (ICECS), 2012 19th IEEE International Conference on*, pages 93–96. IEEE.
- Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media - (1st edition): 1197 - 2006-09-19T07:35:05.000+0200.

SISDOC: Uma Biblioteca Digital Multimídia para Gestão de Documentos Ambientais

Hugo Feitosa de Figueirêdo¹, Ruan Victor Amorim², Cláudio de Souza Baptista²,
Roberta Falcão de Cerqueira Paes³, Brígida Duarte³, Murilo Sérgio Lucena Pinto³,

¹Instituto Federal de Ciência, Educação e Tecnologia da Paraíba (IFPB)
Esperança – PB – Brasil

²Companhia Hidro Elétrica do São Francisco (CHESF)
Recife - PE - Brasil.

³Universidade Federal de Campina Grande (UFCG)
Campina Grande- PB - Brasil

Abstract. *In this article, we propose a new method for managing documents generated by processes related to environmental management in the scope of a power generation company. The new method allows the optical character recognition, indexing the content, metadata extraction, annotation of attribute values created by a dynamic schema, image processing, highlighting, temporal filter and conversion to standard formats. The proposed method was implemented in the Environment Department of the Hydroelectric Company of São Francisco (Chesf) and improves the efficiency of document management, saving computer system resources (hardware, software and peopleware) adopted for management of multimedia documents, and expediting the environmental licensing procedures and supervision of reservoirs edges.*

Resumo. *Neste artigo, propõe-se um novo método para gerenciamento de documentos gerados por processos relacionados à gestão ambiental no escopo de uma empresa de geração de energia elétrica. O novo método permite o reconhecimento óptico de caracteres, a indexação pelo conteúdo, extração de metadados, anotação com valores de atributos criados por um esquema dinâmico, tratamento de imagem, destaque de termos encontrados em documentos, filtro temporal e conversão para formatos padrões. O método proposto foi implantado no Departamento de Meio Ambiente da Companhia Hidrelétrica do São Francisco (CHESF) e permite melhorar a eficiência na gestão de documentos, economizando recursos do sistema computacional (hardware, software e peopleware) adotados para gerenciamento de documentos multimídia, bem como agilizando os processos de licenciamento ambiental e fiscalização de bordas de reservatórios.*

1. Introdução

Bibliotecas digitais são um conjunto de mecanismos eletrônicos que facilitam a localização da demanda informacional, interligando recursos e usuários (Cunha, 2008). Estas bibliotecas têm sido propostas para gerenciar os recursos eletrônicos de forma distribuída. As bibliotecas digitais proveem dados e serviços e lidam com a complexidade de dados multimídias tais como imagens, textos, áudio e vídeo (de Vries,

Eberman, & Kovalcin, 1998).

Um dos grandes diferenciais das bibliotecas virtuais é o suporte a vários formatos de informações como textos, áudio, vídeo e imagens (Andrade, 2010). Este conjunto de dados heterogêneos requer mecanismos de indexação, consulta, apresentação e análise. Tais dados demandam um grande volume de repositório de dados e um dos maiores desafios enfrentados está no desenvolvimento de soluções para lidar com este volume de dados de forma eficiente, integrada, interoperável e distribuída. Por eficiência entende-se que o uso de informação multimídia envolve algoritmos paralelos, otimização de consultas e técnicas de indexação. Por integração entende-se o uso de modelos de dados que transpassam todos os tipos de objetos: simples e complexos. Por distribuição requer-se que o sistema possa prover de forma transparente para os usuários, um meio de acessar remotamente os dados, que devido aos custos de aquisição e manutenção, limita a replicação.

As bibliotecas digitais surgiram primeiro no meio acadêmico e depois foram difundidas na indústria, sendo as primeiras instituições a adotarem esse recurso no Brasil a Universidade de São Paulo (USP), Universidade de Campinas, Universidade de Brasília (UnB), entre outras (Andrade, 2010).

Uma solução que tem sido investigada é o uso de metadados para descrever a semântica, sintaxe e estrutura dos dados. Estes metadados ganharam aceitação em domínios de aplicação distintos tais como museus, bibliotecas, observatórios espaciais, dentre outros. Particularmente, algumas propostas surgiram para a gestão de metadados em bibliotecas digitais tais como: Dublin Core (DCMI, 2012), RDF (Lassila, 1998) e XML (Prescod, P. e Goldfarb, 1999).

Tais metadados descrevem a essência, os atributos e o contexto de um recurso eletrônico, além de modelar relacionamentos entre documentos, não se restringindo a dados que estejam contidos no texto, mas também podendo incluir dados da descrição física e o contexto de produção (Alvarenga, 2006).

Algumas vezes torna-se difícil separar dados de metadados (Huc, Levoir, & Nonon-Latapie, 1997). Por exemplo, alguém pode indagar se um *thumbnail* de uma fotografia pode ser considerado dado ou metadado. O mesmo se aplica ao resumo de um documento, um trailer de um filme. Neste sentido, Gunther e Voisard dizem: "Não existe uma distinção intrínseca entre dados e metadados, é uma questão de contexto indicar se um dado item representa um metadado ou não" (Gunther & Voisard, 1998).

Alvarenga (2006) destaca a importância que os metadados sejam elaborados por uma equipe técnica especializada que conheça as características do acervo, a realidade da biblioteca e de seus usuários, a fim de proporcionar uma melhor recuperação da informação. Por sua vez, documentos relacionados ao meio ambiente possuem uma grande variação nos metadados que podem ser utilizados em cada documento multimídia, por exemplo, com relação ao processo de licenciamento ambiental, cada órgão licenciador possui seu trâmite e informações distintas necessárias para a obtenção da licença. Dessa forma, uma biblioteca digital para área de meio ambiente deve permitir alteração do esquema dinamicamente por pessoas técnicas da área de meio ambiente e não da área de informática ou de ciência da informação.

Para a gestão ambiental e sociopatrimonial, uma empresa gera uma grande quantidade de documentos multimídia, sendo em sua grande maioria dos tipos textuais, áudios, vídeos e imagens. A agilidade na recuperação de um documento multimídia pode facilitar o trabalho dos funcionários da empresa na gestão ambiental e

sociopatrimonial. Outro ponto a destacar é o desperdício advindo por uma má gestão de documentos, que ocasiona aumento dos custos com material permanente (e.g., hardware), de consumo (e.g., papel, tinta) e recursos humanos (e.g., técnicos de informática, analistas). Com isso, o bom gerenciamento dos documentos multimídia torna-se essencial para o setor de gestão ambiental de uma empresa.

Neste artigo, propõe-se um novo processo para gestão de documentos na área de meio ambiente, como também a implementação de uma biblioteca digital multimídia que implementa este processo. O sistema foi testado junto ao departamento de meio ambiente da Companhia Hidrelétrica do São Francisco (CHESF).

O restante deste artigo está organizado como segue. Na seção 2, é apresentado o novo processo de gestão de documentos multimídia relacionados ao meio-ambiente. Na seção 3, é apresentada a biblioteca digital multimídia para área de meio-ambiente. Na seção 4, discute-se a avaliação do sistema proposto. Finalmente, na seção 5, são apresentadas as conclusões e trabalhos futuros.

2. Um novo Processo de Gestão de Documentos Multimídia Relacionados à Área de Meio-Ambiente

Na Figura 1, apresenta-se o modelo do processo de gestão de documentos multimídia proposto neste trabalho. Inicialmente, devem-se cadastrar os atributos e grupos de atributos para cada tipo de documento multimídia utilizado na gestão ambiental e sociopatrimonial. Em seguida, novos documentos podem ser inseridos por uma pessoa física ou por sistemas.

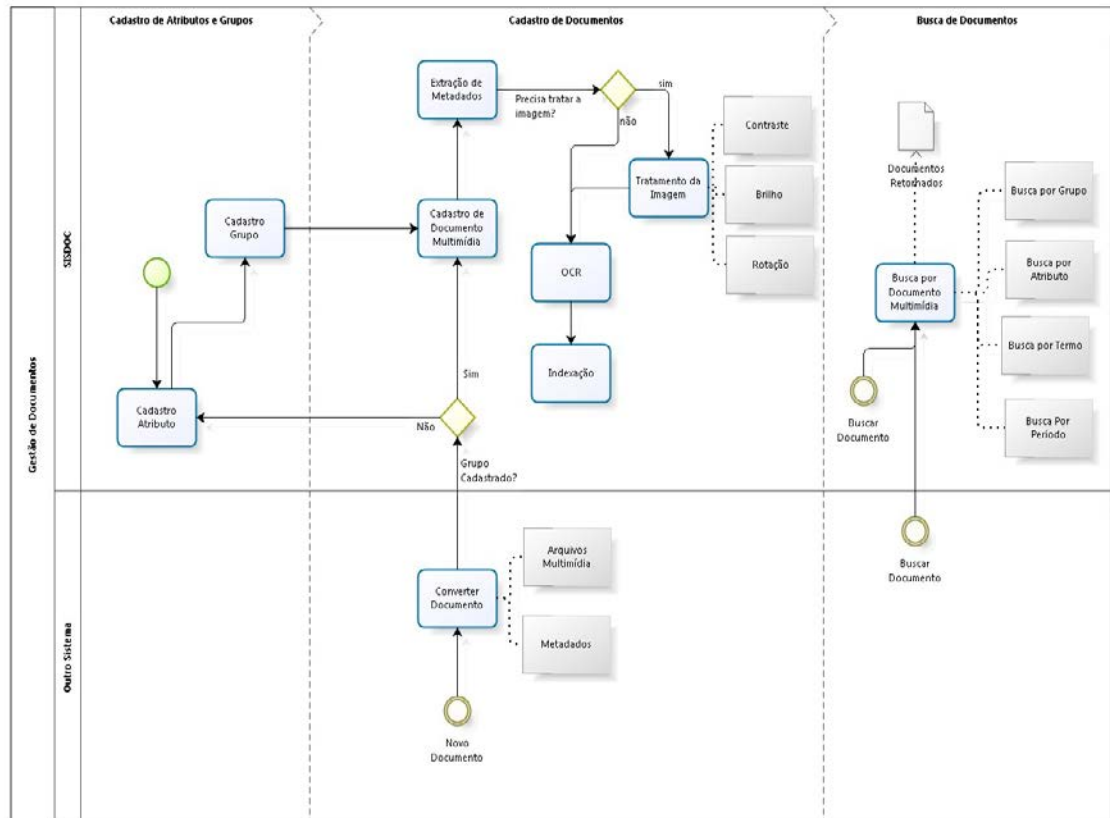


Figura 1. Modelagem do processo gestão de documentos multimídia.

A ideia central para uma boa gestão de documentos multimídia é a centralização do gerenciamento, sendo necessário que outros sistemas, que necessitem lidar com carga, geração e descarga de documentos, utilizem uma biblioteca digital multimídia central.

Quando um sistema precisar utilizar a gestão de documentos multimídia, deverá encapsular junto aos arquivos multimídia todas as informações importantes para anotação e enviar para a biblioteca digital multimídia central. Se o grupo para o qual o documento foi enviado não tiver sido cadastrado ainda, deverá ser criado na biblioteca um novo grupo com os atributos encapsulados passados. Em seguida, o documento será cadastrado na biblioteca.

O processo de cadastro pode necessitar de uma extração dos metadados dos arquivos, tratamento das imagens que podem ser melhoradas mediante alteração de contraste, brilho rotação e recorte, reconhecimento óptico dos caracteres (OCR) para extração de texto contido em imagens e, por fim, indexação. A busca de documentos poderá ocorrer por filtros de grupos, atributos, períodos de tempo e termos contidos nos documentos multimídia.

Para implantação deste processo de gestão de documentos multimídia foi implementado o SISDOC - Sistema de Gestão de Documentos Multimídia, a ser detalhado na próxima seção.

3. SISDOC – Sistema de Gestão de Documentos Multimídia

Na Figura 2, apresenta-se um diagrama com as funcionalidades e atores relacionados à biblioteca digital multimídia aqui proposta. Os atores do SISDOC podem ser classificados em três tipos: sistema externo, administrador e usuário pesquisador. Os sistemas externos representam outros aplicativos que utilizarão o SISDOC como repositório de arquivos ou enviarão os documentos criados e anexados para indexação. Como sistemas externos foram integrados aos SISDOC os seguintes sistemas: Sistema de Licenciamento Ambiental - SISLIC (Santana et al., 2015), Sistema de Fiscalização de Uso e Ocupação de APPs de Bordas de Reservatório - SISBORDAS (Paiva et al., 2015) e Sistema de Educação Sócio-Ambiental - SIPAS (ALVES, BAPTISTA, LEITE, CHAGAS, & PAIVA, 2014). Os administradores podem realizar cadastro de documentos e são os responsáveis pelo cadastro dos atributos e grupo de atributos necessários para o cadastro de documentos. O usuário pesquisador é o que realizará as buscas por documentos no SISDOC.

As principais funcionalidades providas pelo SISDOC são: cadastro de documentos, cadastro de atributos, cadastro de grupo de atributos e busca de documentos. Algumas funcionalidades secundárias são providas a partir da funcionalidade de cadastro de documentos: extração de conteúdo e extração de metadados.

Na Figura 3, apresenta-se a arquitetura do SISDOC, sendo baseada no padrão MVC e arquitetura N-Camadas. Uma implementação da especificação JavaServer Faces 2.0 (JSF) foi utilizada para a adoção do MVC e o Primefaces 3.5 (conjunto de componentes para interfaces gráficas de aplicações baseadas em JSF) foi utilizado para geração da interface gráfica com o usuário (GUI).

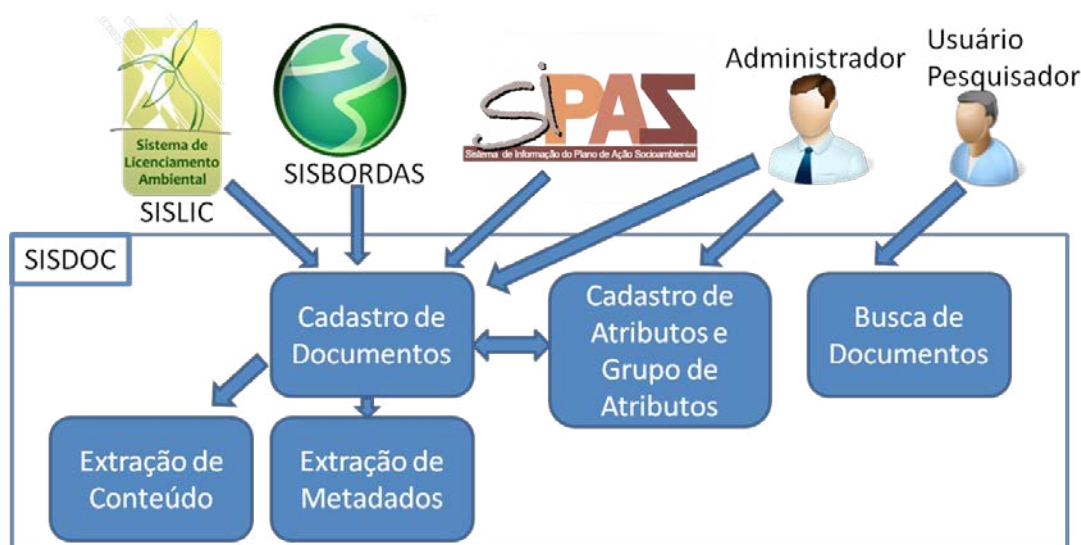


Figura 2 - Visão geral do SISDOC.

O SISDOC foi desenvolvido para ser executado em uma plataforma Java EE 5 e Java SE 6, sendo testado utilizando-se o servidor de aplicações Oracle Weblogic 10.3.5.

Na camada de dados, utiliza-se o Apache Solr 4.7 para indexação textual dos arquivos adicionados pelos usuários ou sistemas externos (SISLIC, SISBORDAS e SIPAS). O Solr é uma aplicação web que foi testada no Weblogic 10.3.5 que possui uma comunicação via REST com outros sistemas. Para comunicação entre o Solr e a camada de acesso aos dados (ver Figura 3), utiliza-se o Solrj 4.7 mediante a utilização de REST. Os arquivos adicionados e os convertidos são adicionados em uma pasta no sistema de arquivos do sistema operacional.

Para viabilizar a comunicação entre o SISDOC e as aplicações externas, foi desenvolvido um *Web Service*, permitindo a realização de intercâmbio de arquivos entre as ferramentas, seguindo o conceito de Cliente e Servidor. Para implementação do *Web Service* foi utilizado o conceito de *REST*. O framework Jersey 1.8 foi utilizado para a comunicação em REST, sendo esse uma implementação da especificação JAX-RS.

Inicialmente, um sistema externo, assumindo o papel de “cliente” na comunicação, envia um arquivo e seus metadados, por meio da Web, para o SISDOC. O SISDOC recebe o arquivo enviado e realiza todos os procedimentos necessários para indexação em seu sistema. Caso atenda todos os requisitos mínimos, o arquivo será persistido e um identificador único será gerado. Tal identificador será repassado ao cliente que realizou a requisição de envio de arquivo. Por sua vez, depois de recebida a confirmação satisfatória de envio, o sistema externo deve realizar o registro junto com o identificador recebido como resposta. Na Figura 4, é apresentado todo o processo de comunicação entre os sistemas utilizando o Sistema de Gestão de Licenciamento Ambiental (SISLIC) da Companhia Hidrelétrica do São Francisco (CHESF) como exemplo.

Alguns documentos adicionados no SISDOC possuem imagens com texto, e.g., licença digitalizada expedida por algum órgão de licenciamento ambiental. Para essas situações, a imagem passa por um processo de OCR (reconhecimento óptico de caracteres) antes da indexação, sendo esse processo realizado pelo Tesseract 3.02. Esta aplicação é externa à aplicação implantada no Weblogic. Para comunicação entre o

Tesseract e a aplicação no Weblogic foi utilizado o Tess4J 1.4.

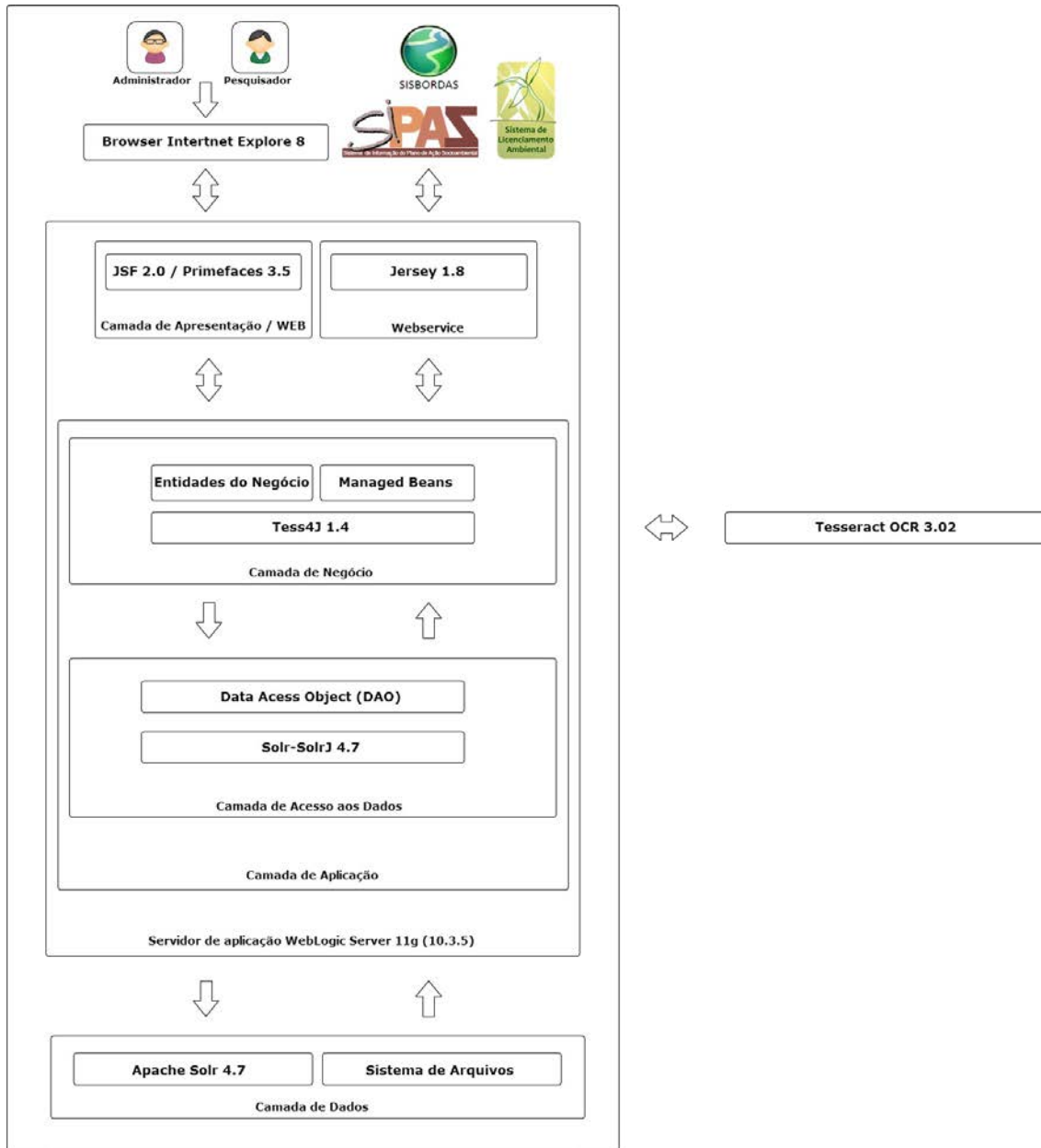


Figura 3. Arquitetura do SISDOC.



Figura 4: Comunicação entre Cliente e Servidor (SISDOC)

3.1. Mapeamento de conteúdo em documento PDF

Muitos dos documentos textuais que estarão armazenados no sistema são compostos por imagens e não possuem o reconhecimento óptico do conteúdo. Assim, foi elaborado um algoritmo que realiza o mapeamento do texto extraído do conteúdo das imagens em um arquivo PDF. Com este arquivo PDF, o conteúdo das imagens pode ser copiado para área de transferência e também pesquisado no próprio documento mediante os programas adequados, como o Adobe Reader via a tecla de atalho Ctrl+F.

O algoritmo utiliza reconhecimento óptico de caracteres para reconhecer e localizar as posições das palavras contidas nas imagens. Em seguida, um arquivo no formato PDF é gerado com a parte textual inserida de forma transparente nas posições adequadas.

Para a realização de consultas que retornem documentos que possuem um determinado texto em seu conteúdo ou nos metadados de forma eficiente, torna-se necessária a utilização de um mecanismo de indexação.

A indexação é um processamento realizado nos arquivos para extrair os metadados e conteúdo e criar uma estrutura de dados eficiente para a busca de documentos por palavras presentes no documento. Para um melhor gerenciamento dos documentos, torna-se necessário um servidor de indexação textual que possui as seguintes características: indexação e busca de documentos, destaques das informações em formato highlighting, exibição de resultados de uma busca em formato categorizado, integração com banco de dados, realização de buscas espaciais, entre outras.

4. Avaliação do SISDOC

O SISDOC foi implantado na Companhia Hidrelétrica do São Francisco no Departamento de Meio Ambiente com o objetivo de gerenciar os documentos multimídia desse setor, validando o processo proposto neste artigo. Além disso, o SISDOC foi integrado a outros sistemas da CHESF.

Outra avaliação realizada foi para determinar o quão eficiente é a técnica de busca de documentos por conteúdo. Nesta seção, apresentam-se a metodologia utilizada no processo de avaliação e os resultados obtidos a partir dos experimentos.

4.1. Metodologia

Tendo em vista a necessidade de avaliar a eficácia do método proposto, foi disponibilizada pela CHESF uma base de dados de documentos multimídia. Estes, oriundos de câmeras fotográficas e scanners, não possuem reconhecimento de caracteres e estão disponíveis em formato de imagem (JPG, PNG ou TIF) e PDF.

Para restringir o escopo da avaliação, a base de dados precisou ser particionada nas seguintes categorias: Autorização, Comprovante, Licença, Parecer e Protocolo. Os documentos são classificados de acordo com a representação do seu conteúdo e devem ser atribuídos a uma das categorias citadas.

Previamente, foi realizada uma análise textual dos arquivos cujo objetivo foi identificar padrões de texto que são comuns a uma ou mais categorias de documentos. Foi necessário construir uma tabela de resultados esperados, onde documentos e padrões são correlacionados. A utilidade desta tabela é auxiliar a verificação dos resultados da pesquisa a partir de um texto dado como entrada.

A Tabela 1 exemplifica como os resultados esperados estão dispostos. Os valores

da primeira coluna representam os padrões de texto que serão inseridos na entrada do método proposto. A segunda coluna representa os documentos multimídia que devem ser retornados na pesquisa quando o padrão relacionado for pesquisado.

Tabela 1. Exemplo dos resultados esperados para a categoria Autorização.

Padrões de Texto	Documentos
“Autorização Especial”	Doc-Aut-1.pdf
“Desenvolvimento Sustentável”	Doc-Aut-2.png, Doc-Aut-3.jpg, Doc-Aut-4.tif
“ <i>Termo de autorização</i> ”	Doc-Aut-1.pdf, Doc-Aut-2.png

Na Figura 5, demonstra-se a realização do experimento para um padrão selecionado. Na barra de pesquisa superior, define-se o padrão a ser pesquisado nos documentos indexados. No quadro inferior é demonstrado o resultado da pesquisa destacando o nome dos documentos e o conteúdo que foi relacionado ao padrão textual.

The screenshot shows a search interface. At the top, there is a search bar containing the text "Licença Ambiental", a "Buscar" button, and a "Limpar" button. Below the search bar, there are icons for "Acervo" and "Linha do tempo". The main area displays "Documentos resultantes da pesquisa" with a list of results. The first result is "Doc-Lic-14" with author "suporte" and ID "28". The highlighted content for this result is "Licença Ambiental Licença Prévia Nº 036/2011 Validade: 26.05.2013. O INSTITUTO DO MEIO AMBIENTE DO ESTADO ...". The second result is "Doc Lic: 9" with author "suporte" and ID "23". The highlighted content is "Licença Ambiental FMA Validade: 06.12.2013 Licença de Instalação Nº 081 ...". The third result is "Doc Lic: 10" with author "suporte" and ID "24". The highlighted content is "Licença Ambiental Validade: 06.12.2013 O INSTITUTO DO MEIO AMBIENTE DO ESTADO DE ALAGOAS - IMA/AL ...". The fourth result is "Doc Lic: 20" with author "suporte" and ID "34". The highlighted content is "Hídricos - SEMAR LICENÇA AMBIENTAL 1. Categoria 2. Número f Processo 3. Validade R novação de Licença de ...".

Figura 5. Exemplo de pesquisa e demonstração de resultados.

Durante a execução dos experimentos, deve-se coletar os resultados das pesquisas comparando-os com a tabela de resultados esperados. Dessa forma, é possível identificar quais documentos retornados são, de fato, relevantes. Com o número de documentos retornados e o número de documentos relevantes, as métricas de interesse: precisão, revocação e f-measure, podem ser calculadas, e assim possibilitará a avaliação da eficácia do método proposto.

4.2. Resultados

Na Tabela 2, são apresentados os resultados, agrupados por categorias, obtidos a partir dos experimentos. Os resultados são considerados excelentes para busca de documento por conteúdo, porém, os bons resultados do método proposto está condicionado à

qualidade visual dos documentos indexados. Documentos que possuem caracteres especiais (ex: °, ª, &), rasura, amasso, borrões ou outros defeitos, tendem a dificultar o processo de reconhecimento óptico de caracteres, prejudicando o reconhecimento de padrões textuais.

Tabela 2. Resultados da avaliação da técnica agrupado por categoria.

Categoria	Precisão	Revocação	F-Measure
Autorização	1	0,913	0,954
Comprovante	1	0,714	0,416
Licença	1	0,789	0,441

Em algumas categorias de documentos houve uma variação considerável nas métricas Revocação e F-Measure. O principal fator responsável por esta variação foi a qualidade visual dos documentos, que pode ser diferente dependendo da categoria a qual pertence. Os documentos das categorias Comprovante e Licença apresentaram defeitos como: caracteres especiais, amasso e textura irregular no background. Estes fatores contribuíram negativamente para o reconhecimento do conteúdo do documento, dessa forma, afetando as métricas supracitadas.

5. Conclusões

Neste artigo, foi proposto um processo para gerenciamento de documentos multimídia na área de meio ambiente.

O processo foi implementado por meio de um sistema intitulado SISDOC, o qual permite: (i) a evolução dinâmica do esquema, podendo ser modificado por especialistas da área de meio ambiente sem necessidade de conhecimentos avançados na área de informática; (ii) extração dos metadados contidos no documento multimídia; (iii) extração do conteúdo; (iv) busca por metadados; (v) busca por filtros temporais; (vi) mapeamento de conteúdo textual em imagens; (vii) visualização prévia de arquivos textuais, áudios, vídeos e imagens; e (viii) tratamento de imagens para melhoria de processo de reconhecimento óptico de caracteres.

O SISDOC foi implantado no departamento de meio ambiente da CHESF e foi integrado ao SISFAIXAS, SISBORDAS e SISLIC que são três soluções para a gestão do meio ambiente de uma empresa do setor elétrico que necessitam de um bom gerenciamento de arquivos.

Como pesquisas futuras, pretende-se avaliar as melhorias providas pelo novo processo no departamento de meio ambiente da CHESF para gestão de documentos, adoção da solução para outros setores e inclusão de suporte a consultas espaciais.

Agradecimentos

Os autores agradecem o suporte financeiro da ANEEL, sob o contrato de P&D+I N° ANEEL 0048-1119/2012.

Referências Bibliográficas

Alvarenga, L. (2006). Organização da informação nas bibliotecas digitais. *Organização Da Informação: Princípios E Tendências*, 76–98.

- Alves, A. L. F., Baptista, C. D. S., Leite, D. F. B., Chagas, M. I. A., & Paiva, A. C. (2014). Aplicação de Geoprocessamento em um Sistema de Gestão Ambiental para Diagnóstico de Ações Socioambientais. In *Encontro Nacional de Geoprocessamento do Setor Elétrico - ENGEIO*.
- Andrade, N. S. de. (2010). Biblioteca digital: repositório de informação e conhecimento. *Fonte*, 7(10). Retrieved from http://www.prodemge.mg.gov.br/images/com_arismartbook/download/11/revista_10.pdf#page=78
- Cunha, M. B. da. (2008). Das bibliotecas convencionais às digitais: diferenças e convergências. *Perspectivas Em Ciência Da Informação*, 13(1), 2–17. <http://doi.org/10.1590/S1413-99362008000100002>
- DCMI. (2012). Dublin Core Metadata Initiative Metadata Terms. Retrieved April 4, 2013, from <http://dublincore.org/documents/dcmi-terms/>
- de Vries, A., Eberman, B., & Kovalcin, D. (1998). The Design and Implementation of an Infrastructure for Multimedia Digital Libraries. In *Proceedings of the 1998 International Database Engineering and Applications Symposium* (pp. 103–120). IEEE.
- Gunther, O., & Voisard, A. (1998). *Metadata in Geographic and Environmental Data Management*. (W. Klas & A. Sheth, Eds.). McGraw Hill.
- Huc, C., Levoir, T., & Nonon-Latapie, M. (1997). Metadata: Models and Conceptual Limits. In *Proceedings of the Second IEEE Metadata Conference*. Maryland, USA: IEEE.
- Lassila, O. (1998). Web Metadata: A Matter of Semantics. *IEEE Internet Computing*, 30–37.
- Paiva, A. C. de, Campelo, C. E. C., Figueiredo, L. C. de, Rocha, J. H., Figueirêdo, H. F. de, & Baptista, C. de S. (2015). Management of Large Hydroelectric Reservoirs Surrounding Areas Using GIS and Remote Sensing. In A. Kó & E. Francesconi (Eds.), *Electronic Government and the Information Systems Perspective* (pp. 257–268). Cham: Springer. <http://doi.org/10.1007/978-3-319-22389-6>
- Prescod, P. e Goldfarb, C. (1999). *The XML Handbook*. Prentice Hall PTR.
- Santana, J. V., Figueirêdo, H. F. de, Baptista, C. de S., Paiva, A. C. de, Paes, R. F. de C., Pinto, M. S. L., & Duarte, B. (2015). SISLIC: Um metodo para gerenciamento do processo de licenciamento ambiental. In *VI Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais*. Retrieved from <http://www.lbd.dcc.ufmg.br/bdbcomp/servlet/Trabalho?id=23623>

Redes de Monitoramento Ambiental: Um projeto aplicado às Geleiras Patagônicas

Guilherme T. Netto¹, Jorge Arigony-Neto¹, Adenauer Correa Yamim²

¹Laboratório de Monitoramento da Criosfera – Universidade Federal do Rio Grande (FURG)
Rio Grande – RS – Brazil

²Laboratório de Computação Ubíqua e Pervasiva – Universidade Federal de Pelotas (UFPEL)
Pelotas – RS – Brasil

guilherme.netto@inf.ufpel.edu.br, jorgearigony@furg.br, adenauer@inf.ufpel.edu.br

Abstract. *The climate changes are an issue constantly debate today, and glaciers play an important role in this context as they are as natural sensors that react to these changes. Its monitoring is of great importance in order to know their behavior related to weather events and their influence on climate. This paper proposes the use of the sensor network model to carry out the monitoring of glaciers in an integrated manner, using a wide range of sensors. The first tests with automatic weather stations (AWS) were very promising even with severe weather conditions.*

Resumo. *As mudanças climáticas são um tema em constante debate atualmente, e as Geleiras têm um papel importante neste contexto, pois são como sensores naturais que reagem a estas mudanças. Seu monitoramento é de grande importância a fim de conhecermos seu comportamento perante a eventos climáticos e sua influência no clima. O presente trabalho propõe a utilização do modelo de redes de sensores para realizar o monitoramento de Geleiras de forma integrada, utilizando um conjunto variado de sensores. Os primeiros testes com estações meteorológicas automáticas (AWS) mostram-se promissores mesmo com condições climáticas severas.*

1. Introdução

Nas últimas décadas a discussão sobre a influência humana nas mudanças climáticas tem se intensificado. Programas de âmbito mundial para tratar do tema tem se formado, como exemplo concreto neste sentido, temos o Painel Intergovernamental sobre Mudanças Climáticas (IPCC). Órgão criado pelo Programa das Nações Unidas para o Meio Ambiente (PNUMA) e pela Organização Meteorológica Mundial (OMM) em 1988 para estudar o problema das mudanças climáticas. Reúne 2.500 cientistas de mais de 130 países. O IPCC tem emitido relatórios periódicos caracterizando as influências dos sistemas humanos sobre os ecossistemas naturais. O último relatório publicado evidencia a ação humana como grande responsável pelo aquecimento global detectado nos últimos 50 anos. Este aquecimento atinge vários aspectos climáticos, dentre os quais destaca-se a elevação da temperatura dos oceanos, variações extremas de temperatura do ar e os padrões de circulação atmosférica [IPCC 2014].

Outro aspecto tratado por este relatório é a influência destas variações climáticas no derretimento das massas de gelo do planeta, o que traria consequências negativas de

grandes proporções, como por exemplo o aumento do nível médio dos oceanos, alterações climáticas mais severas e a decorrente necessidade de migração populacional.

Estas massas de gelo que cobrem parte da superfície da terra, de forma perene ou sazonal, fazem parte da criosfera e cobrem 10% do total do planeta [CUFFEY and PATERSON 2010]. Segundo [Arigony-Neto 2006] houve uma grande aceleração no derretimento de parte destas massas, bem como desagregação de plataformas de gelo e a extinção de geleiras. Tais fenômenos podem influenciar o clima do planeta dada a importância da criosfera e do oceano austral nos processos de equilíbrio térmico do planeta [Mata et al. 2011, Simões 2014].

Geleiras são estruturas dinâmicas resultantes da ação do clima e da topografia. Formam-se por precipitação de neve, gelo ou chuva e perdem massa por derretimento e evaporação de neve, desprendimento de massas de gelo, erosão eólica e avalanches. Isto determina duas grandes áreas conceituais, a Área de acumulação e a Área de ablação [CUFFEY and PATERSON 2010]. O balanço entre acúmulo e perda de gelo determina a formação ou extinção de uma geleira.

Muitos esforços vem sendo empregados para monitorar o derretimento das geleiras [LaChapelle 1959, Müller and Keeler 1969, Bøggild et al. 2004, Braithwaite 2009, Hulth 2010, Van As 2011, Fausto et al. 2012] e diferentes estratégias com sensores específicos foram empregadas neste intuito, como a observação da variação das camadas de gelo por estacas [LaChapelle 1959, Müller and Keeler 1969], sensores de pressão [Bøggild et al. 2004, Van As 2011, Fausto et al. 2012], ultrassônicos [Hulth 2010, Van As 2011] e por fios de aço [Hulth 2010]. Cada estratégia mostrou pontos fortes e deficiências, demonstrando assim a necessidade de criarem-se sistemas híbridos, com mais de uma abordagem de aquisição dos dados e de forma distribuída.

Os parâmetros meteorológicos como vento (direção e intensidade), pressão atmosférica, temperatura do ar e do gelo, umidade e luminosidade são medidos por estações meteorológicas, que devido ao alto custo e à baixa portabilidade, são instaladas em número reduzido, limitando-se a poucas unidades. Esta resolução espacial não representa o que acontece em toda extensão das geleiras. Para monitorar o derretimento destas geleiras com uma maior precisão se faz necessária a utilização de mais de uma técnica de levantamento de dados, com uma maior resolução temporal para uma melhor calibração e validação nos estudos destes modelos de degelo, sobretudo em escalas de tempo menores em eventos específicos [Hulth 2010].

O emprego de mais de um ponto de coleta conectado a uma rede de sensores distribuídos pode contribuir significativamente para uma visão mais ampla destas dinâmicas (acúmulo e ablação). Desta maneira, comparando as informações a respeito da geleira, como temperatura e perda de massa, por exemplo, com dados meteorológicos, é possível verificar como os processos meteorológicos influenciam na perda de massa de gelo, sobretudo nos eventos extremos, não só nas zonas de ablação mas também nas zonas de acumulação, bem como em altitudes diversas e outras condições morfológicas.

A coleta automatizada de informações do meio por dispositivos computacionais, móveis ou não, com o intuito de qualificar o atendimento às necessidades dos usuários é chamada de computação ciente do contexto [Lopes et al. 2014].

Ampliando os conceitos de ciência do contexto, [Ye et al. 2012] propõe um mo-

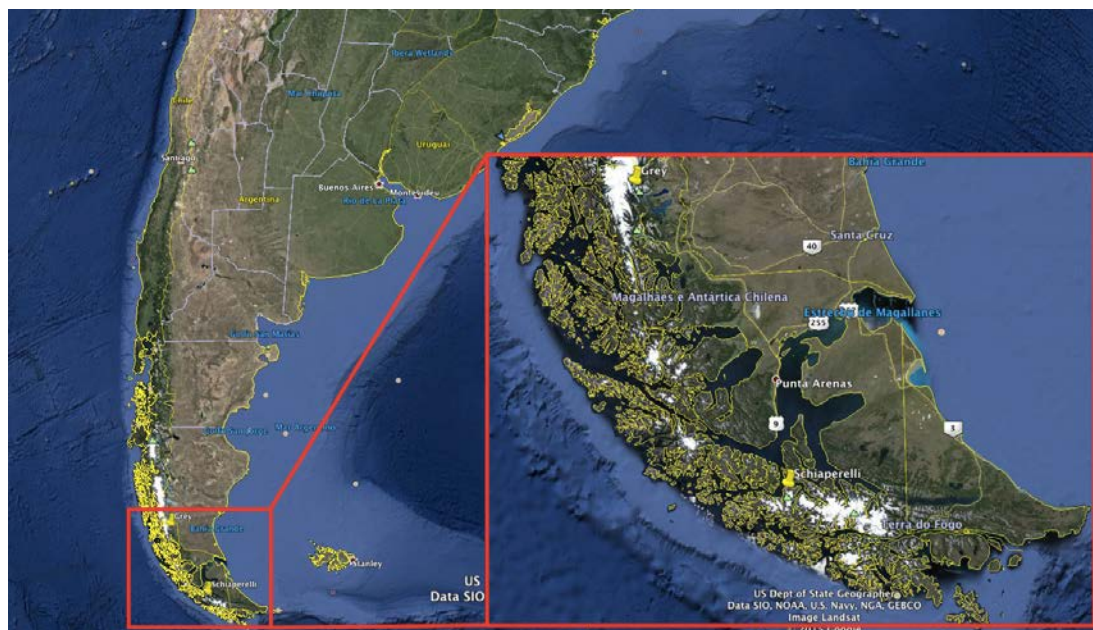


Figura 1. Mapa da Região de Estudo

delo de integração onde além da análise dos dados coletados do ambiente também é feito um acompanhamento da evolução dos mesmos, considerando mais uma dimensão, a análise temporal, introduzindo então o conceito de ciência de situação.

Neste sentido, este trabalho apresenta uma proposta para criação de uma rede de sensores meteorológicos e de ablação com o objetivo de monitorar um ambiente climático singular. As áreas de estudo situam-se ao sul da Patagônia Chilena, na Cordilheira Darwin, um pouco mais ao norte no Parque Nacional Torres del Paine, e chamam-se Geleira Schiaparelli e Geleira Grey (Vide Figura 1). As características geográficas, o isolamento e a baixa influência direta de fatores antrópicos tornam esta região propícia para este tipo de estudo.

2. Metodologia

A miniaturização de alguns sensores sem fio disponíveis atualmente e a distribuição integrada dos mesmos junto ao meio físico são indicativos de que as Redes de Sensores sem Fio são uma importante contribuição para consolidação da Computação Ubíqua.

Com os recentes avanços no desenvolvimento de sistemas microeletrônicos, as Redes de Sensores sem Fio podem ser formadas por um grande número de pequenos sensores, onde cada sensor é um elemento autônomo capaz de captar a informação do meio, tratar e enviar esta informação através de comunicação sem fio, consolidando assim a autonomia prevista na Computação Ubíqua [Lopes et al. 2014].

Para tanto, se faz necessário um transceptor para comunicação, uma unidade de sensoriamento, fonte de energia, memória e uma unidade de processamento. Tais elementos influenciaram várias aplicações desta tecnologia que agrega recursos antes intratáveis com modelos convencionais, por exemplo, no monitoramento de áreas remotas, com sensores sendo lançados sobre reservas ambientais, florestas, vulcões, rios [Ruiz 2003].

Por esse motivo, o presente trabalho busca agregar um conjunto de tecnologias, preferencialmente livres (i.e., projeto de domínio público), de baixo custo, mas perseguindo um nível de confiabilidade aceitável para medidas da dinâmica de geleiras. O desenvolvimento da rede de sensores está sendo feito de maneira incremental, pois existe a necessidade da validação dos sensores de borda para posteriormente interligá-los utilizando uma rede.

Na primeira etapa foram desenvolvidos dois conjuntos de sensores que ainda funcionam de maneira isolada, uma Estação Meteorológica Automática (AWS) e um sensor de Ablação utilizando *tags RFID*. Na segunda etapa estes sensores serão interligados utilizando uma rede sem fio e comunicação satelital com o servidor remoto responsável pelo registro e processamento das informações coletadas.

2.1. Ambiente Operacional

Na implementação da rede de sensores estão sendo utilizados os microcontroladores da família Arduino [Arduino 2016], que possuem as características técnicas necessárias ao projeto, e destacam-se pelo seu baixo custo e praticidade de desenvolvimento.

Para a comunicação de dados na rede projeta-se a utilização das placas ESP8266 versão 7, que possibilitam a conexão entre os nodos e o concentrador via protocolo TCP e possuem interface para antenas externas possibilitando um ganho de sinal maior comparado às outras versões. Já a comunicação remota da rede, concentrador e servidor deverá ser feita utilizando a tecnologia de Rádio-Pacote através de uma rede satelital de Rádio Amadores. Este modelo apresenta uma banda de comunicação reduzida, mas suficiente para as demandas de comunicação da proposta de monitoramento que está sendo proposta. O modelo de sensores de borda é composto por conjuntos de circuitos acoplados

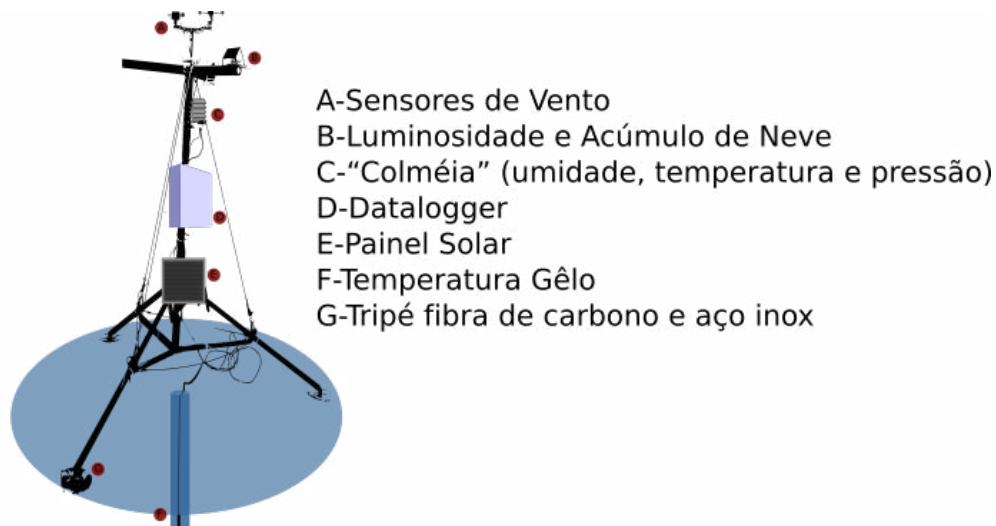


Figura 2. Estação Meteorológica Automática

pelos microcontroladores através de comunicação sem fio. Cada conjunto deverá possuir sensores para medir velocidade de fluxo das geleiras, pressão barométrica, temperatura do ar, umidade, temperatura da neve, quantidade de neve acumulada, ablação, radiação solar, direção e intensidade de vento, como mostra a Figura 2.

Um protótipo de estação meteorológica completa foi executado com tripé em fibra

de carbono, a um custo de R\$2.000,00, bastante inferior ao custo das estações comerciais adquiridas anteriormente com características semelhantes, e que têm um preço médio de R\$ 40.000,00. Pelo fato da estação ser construída com materiais leves o seu peso total é 11Kg, mostrando-se bastante portátil quando comparada com as estações anteriormente instaladas que pesam 40Kg aproximadamente.

Os sensores de ablação usualmente são desenvolvidos a partir do modelo de balizas, o qual consiste em inserir um conjunto de canos de PVC na geleira até 12 metros de profundidade, e a cada retorno da equipe ao campo são efetuadas medições de quanto estas balizas estão aparentes, apresentando assim uma baixa resolução temporal, em torno de seis meses.

No modelo proposto a detecção do descongelamento superficial é feito por sensores de reconhecimento por radiofrequência(RFID), Figura 3, com uma resolução espacial de 15 cm e um *datalogger*, assim será possível armazenar os dados de derretimento das geleiras nas regiões de ablação nos períodos exatos que ocorrem.

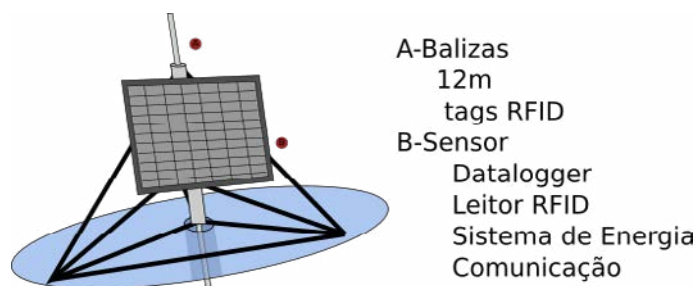


Figura 3. Sensor de Ablação

O modelo de integração será implementado utilizando-se uma arquitetura em camadas, com sensores de borda realizando a coleta das informações, uma camada intermediária chamada de *middleware*, responsável pela distribuição e transmissão das informações e os servidores de contexto que serão os responsáveis pelo armazenamento temporário e transmissão para a base remota. Tal abordagem torna o projeto bastante versátil quanto à customização da rede de sensores como demonstrou o trabalho de [Ruiz 2003] e [Neto 2009].

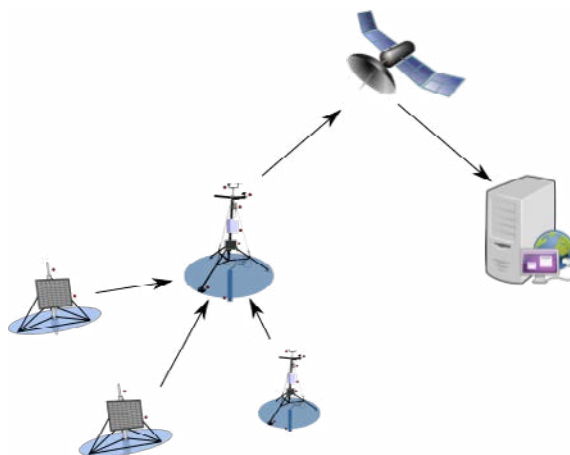


Figura 4. Topologia da rede

A topologia de rede a ser utilizada está ainda sob estudo, e o modelo mais adequado até este momento do estudo é o *multihop*, devido a possibilidade de cobrir distâncias maiores entre os nodos sensores e o concentrador.

3. Testes de Campo

Os primeiros testes de campo foram realizados com as Estações Meteorológicas Automáticas, e o objetivo destes experimentos era testar a robustez dos conjuntos. Testes de bancada foram efetuados durante o mês de julho de 2015 e as duas estações foram instaladas junto às estações meteorológicas da Furg, Figura 5, Campus Carreiros, para fins de aferição. Esta é uma típica linha de registro com dados gravados pela AWS:

25.08.15; 10 : 23 : 22; 491.00; 5.00; 19.69; 35.50; E; 62.40; 18.70; 18.40; 1019

Nesta linha os dados são separados por (;) e representam respectivamente, data e hora da leitura, luminosidade, acúmulo de neve, temperatura do gelo, velocidade do vento, direção do vento, umidade relativa do ar, temperatura do ar em dois sensores e pressão atmosférica. A frequência de leitura é de um registro por minuto.



Figura 5. Teste de Campo - Furg/Campus Carreiros

Com a experiência dos testes de bancada, ajustes foram feitos para adequar os equipamentos para o ambiente de estudo e em setembro de 2015 as duas AWS foram instaladas nas Geleiras Patagônicas. Em conjunto com o Instituto Antártico Chileno, Fundação CEQUA e Universidade de Magalhães, todos com sede em Punta Arenas no Chile, foram realizadas duas expedições, a primeira para instalação da estação na Geleira Schiaparelli (Figura 6), Cordilheira Darwin, e a segunda para instalação na Geleira Grey (Figura 7), Parque Nacional Torres del Paine. A Estação de Apoio Antártico (ESANTAR) também deu apoio fornecendo parte dos equipamentos de gelo.

Instalar e testar equipamentos em campo via de regra são tarefas bastante árduas, sobre tudo se as áreas de estudo estão em regiões remotas. É o caso das geleiras estudadas, principalmente a Schiaparelli, que exigiu uma logística bastante elaborada devido a localização geográfica, pela exposição a condições severas de clima (baixas temperaturas, ventos fortes e neve) e possuir um terreno bastante acidentado.

Para exemplificar, na primeira expedição houve um deslocamento de barco desde Punta Arenas, Chile, atravessando o Estreito de Magalhães, onde após o desembarque na praia a equipe precisou realizar o porteio dos equipamentos por dois dias até o campo

base, e somente então conseguir o acesso a geleira e começar os trabalhos no gelo propriamente dito. Isto exigiu alguns conhecimentos e treinamento específicos de escalada em gelo e deslocamento em geleiras, bem como uma preparação física prévia.



Figura 6. Geleira Schiaparelli (54 24'1.50"S 70 52'39.93"O) e AWS instalada

A escolha da Geleira Schiaparelli ocorreu por ser uma região bastante inóspita, com baixa influência antropogênica direta e por estar fortemente influenciada pelas forçantes climáticas provenientes da Antártida (vento, precipitação e umidade). A escolha também se deu por conveniência logística visto que os institutos chilenos citados anteriormente frequentemente realizam expedições nesta região, o que facilita enormemente as operações de instalação e manutenção.



Figura 7. Geleira Grey (54 24'1.50"S 70 52'39.93"O) e Montagem da AWS

A Geleira Grey também foi escolhida pela conveniência do apoio dos institutos chilenos, e pelo fato de ser um ambiente diverso, com um clima mais seco e apresentar dimensões maiores e um período de estudos pretéritos mais longo.

3.1. Primeiros Resultados

As duas estações estão em operação desde setembro de 2015, e em janeiro de 2016 foi trocado o cartão de memória da AWS Grey. Foi verificado seu funcionamento e condições físicas gerais. O sistema está operando de maneira satisfatória, tendo armazenado 172.625 leituras. Foi verificado apenas um incidente com um conjunto de sensores devido a uma falha em um conector que foi danificado devido ao congelamento. A estrutura

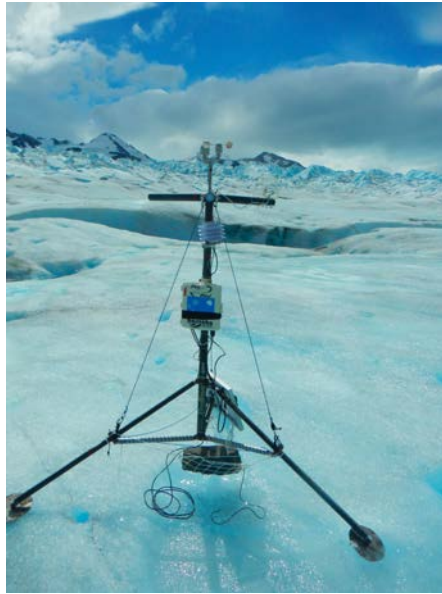


Figura 8. AWS Grey após 6 meses de funcionamento

física mostrou-se extremamente robusta, apesar do baixo peso. A Figura 8 mostra a AWS Grey após 6 meses de operação, demonstrando que as estratégias de ancoragem e escolha de materiais utilizados foram plenamente satisfatórias. No mês de abril será efetuada a troca do cartão de memória da AWS Schiaparelli.

4. Conclusões

Este trabalho tem como principal contribuição o desenvolvimento de um conjunto de tecnologias para o monitoramento de geleiras e o estudo de seu comportamento.

Os primeiros testes mostraram que a abordagem utilizada, aplicando plataformas livres e de baixo custo, foi bastante satisfatória. Com esta experiência será possível implantar o modelo de redes na próxima temporada, em setembro de 2016. Nesta nova etapa, já com os inéditos sensores de ablação com leitores RFID, será feito um acompanhamento com uma resolução temporal bem superior à utilizada atualmente.

É esperado que com este conjunto de dados, tendo uma resolução espacial e temporal maior que as utilizadas até então, seja possível correlacionar o comportamento destas geleiras com eventos climáticos, estimar de forma mais precisa as velocidades de deslocamento e seu cálculo de balanço de massa de gelo.

Referências

- Arduino (2016). Arduino Web Page. <https://www.arduino.cc/>. Accessed: 2016-01-15.
- Arigony-Neto, J. (2006). *Monitoring glacier parameters on the Antarctic Peninsula - a centerline approach combining satellite and GIS data*. PhD thesis.
- Bøggild, C. E., Olesen, O. B., Ahlstrøm, A. P., and Jørgensen, P. (2004). Automatic glacier ablation measurements using pressure transducers. *Journal of Glaciology*, 50(169):303–304.

- Braithwaite, R. J. (2009). Calculation of sensible-heat flux over a melting ice surface using simple climate data and daily measurements of ablation. *Annals of Glaciology*, 50(50):9–15.
- CUFFEY, K. and PATERSON, W. S. B. (2010). *The Physics of Glaciers*.
- Fausto, R. S., van As, D., Ahlstrom, A. P., Andersen, S. B., Andersen, M. L., Citterio, M., Edelvang, K., Larsen, S. H., Machguth, H., Nielsen, S., and Weidick, A. (2012). Ablation observations for 2008-2011 from the Programme for Monitoring of the Greenland Ice Sheet (PROMICE). *Geological Survey of Denmark and Greenland Bulletin*, (26):73–76.
- Hulth, J. (2010). Using a draw-wire sensor to continuously monitor glacier melt. *Journal Of Glaciology*, 56(199):922–924.
- IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report.
- LaChapelle, E. (1959). ERRORS IN ABLATION MEASUREMENTS. *Journal of Glaciology*, 3(26).
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014). A middleware architecture for dynamic adaptation in ubiquitous computing. *Journal of Universal Computer Science*, 20(9):1327–1351.
- Mata, M. M., José Goldemberg and Jefferson Cardia Simões and Heitor Evangelista and Lúcia de Siqueira Campos and Carlos Alberto Eiras Garcia, and Breme, U. F. (2011). *Antártica e as Mudanças Globais: um desafio para humanidade*. São Paulo, 1 edition.
- Müller, F. and Keeler, C. M. (1969). Errors in short-term ablation measurements on melting ice surfaces. *Journal of Glaciology*, 8(52):91–105.
- Neto, A. T. (2009). Rede de sensores sem fio e computação ubíqua na agropecuária. *Embrapa*, page 19.
- Ruiz, L. B. (2003). *Maná: uma arquitetura para gerenciamento de redes de sensores sem fio*. PhD thesis.
- Simões, J. C. (2014). Antartica, 2048: Mudanças Climáticas e Equilíbrio Global. pages 50–59.
- Van As, D. (2011). Warming, glacier melt and surface energy budget from weather station observations in the melville bay region of northwest greenland. *Journal of Glaciology*, 57(202):208–220.
- Ye, J., Dobson, S., and McKeever, S. (2012). Situation identification techniques in pervasive computing: A review. *Pervasive and Mobile Computing*, 8(1):36–66.

Estratégia Evolutiva para Parametrização de Modelos de Previsão: Um Estudo de Caso com Níveis Máximos Mensais do Rio Xingu em Altamira/PA

Alen Costa Vieira^{1,2}, Gustavo Pessin^{1,3}

¹Instituto de Ciências Exatas e Naturais
Universidade Federal do Pará (UFPA) – Belém, PA, Brasil

²Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM)
Centro Regional de Belém, Belém – PA, Brasil

³Laboratório de Computação Aplicada
Instituto Tecnológico Vale – Belém, PA, Brasil

alen.vieira@sipam.gov.br, gustavo.pessin@itv.org

Resumo. *O uso adequado de métodos de previsão pode auxiliar na prevenção, no gerenciamento e no planejamento de situações críticas. Métodos de previsão de variáveis de interesse podem ser endereçados como problemas de previsão de séries temporais. A previsão de séries temporais apresenta algumas questões em aberto, correntemente estudadas, entre estas questões estão (1) como definir o tamanho ótimo da janela de entrada do método de previsão e (2) como definir os conjuntos de variáveis (outras séries temporais) que impactam no modelo. Neste artigo, apresentamos como um algoritmo evolutivo pode ser empregado para escolher as variáveis climáticas e o tamanho dessas janelas para aumentar a precisão do modelo preditivo. O algoritmo evolutivo é empregado em um estudo de caso considerando níveis máximos do rio Xingu utilizando 18 diferentes séries temporais de variáveis climáticas. Mostramos também como configurações do algoritmo evolutivo podem levar a resultados mais consistentes.*

1. Introdução

De acordo com o IPCC [IPCC 2013], é perceptível o aquecimento do planeta, comprovado pelo aumento das temperaturas do ar e dos oceanos, e o aumento dos níveis dos mares. Por conta dessas mudanças climáticas, foram constatados que diversos sistemas naturais estão sendo impactados com a alteração no regime de precipitação e na mudança da temperatura. Essas mudanças possivelmente ocasionarão aumento no número e na severidade de eventos hidroclimatológicos, como cheias e secas mais prolongadas, afetando a disponibilidade hídrica para as atividades da população. O uso adequado de métodos de previsão pode auxiliar na prevenção, no gerenciamento e no planejamento de situações críticas provocadas pelo aumento ou diminuição dos níveis de rios, entre elas, com (1) remoção de pessoas de possíveis áreas afetadas antes que a calamidade ocorra, (2) atuação na construção de novas redes de distribuição de água em caso de seca grave, e (3) planejamento energético nas hidrelétricas aproveitando ao máximo a disponibilidade da água para a geração de energia.

Métodos de previsão de variáveis de interesse (no caso deste estudo: nível máximo, ou cotas fluviométricas, do rio Xingu) podem ser endereçados como problemas

de previsão de séries temporais. Usualmente, métodos de previsão em séries temporais empregam (1) dados passados da própria série ou (2) outras séries temporais de variáveis com certo grau de correlação. O item (1) descrito acima é também conhecido como “Tamanho da Janela de Entrada” para previsão. Por exemplo, para prever Valor_{t+1} , pode ser empregado um número de combinação de entradas que vão de Valor_{t_0} até Valor_{t-k} , sendo k o tamanho total da série. A definição do melhor tamanho de janela de entrada é um problema recorrente e constantemente estudado na área de séries temporais.

Em relação ao item (2) descrito anteriormente, além da questão de tamanho de janela ficar em aberto para cada variável adicional, ainda, outra questão que se abre é a escolha das outras séries; ou seja, quais das outras séries auxiliam a melhorar a precisão do modelo. Por exemplo, chuva no Ponto A pode ter relação com nível do rio no Ponto B, porém chuva no Ponto C pode não ter relação. Isso, em geral, pode ser facilmente resolvido por especialistas do domínio em sistemas com poucas variáveis, entretanto, ao se investigar sistemas com número muito grande de variáveis, essas observações não são triviais, e, quando feitas de forma empírica pode induzir a criação de sistemas de menor qualidade. A seleção dos parâmetros (janelas de tempos e variáveis climáticas) tem como objetivo melhorar o desempenho do método de previsão. Encontrar o melhor conjunto de parâmetros para um dado sistema é eventualmente intratável dependendo do número de variáveis. Problemas relacionados a seleção de variáveis são considerados de difícil solução [Blum and Langley 1997]. Nesse contexto, a utilização de estratégia evolutiva é amplamente difundida para diversos tipos de problemas de otimização. As estratégias evolutivas são bem aceitas por conta de duas características: (1) a capacidade de explorar um grande espaço de busca e (2) a capacidade de permitir ajustes finos explorando locais próximos do ótimo [Eiben and Schippers 1998].

Estudos como [Franco 2007] e [Rocha et al. 2007] definem as variáveis climáticas e suas janelas de tempo empregando correlações. [Chen and Yu 2007] realiza um trabalho que estima os níveis horários no Rio Lan-Yang no nordeste de Taiwan utilizando máquina de vetores de suporte escolhendo as variáveis e suas janelas de tempo por meio de correlação cruzada. Esses trabalhos desconsideram as interações entre as variáveis na busca do objetivo, pois a ocorrência de boa correlação não caracteriza uma relação de causa e efeito. Os trabalhos de [Dornelles et al. 2013] e [Rodrigues et al. 2015] testam diferentes combinações predefinidas de janelas de tempo nas suas previsões de níveis de rio, analisando o comportamento em diferentes janelas de tempo e comparando os resultados. É perceptível, nestes trabalhos, a dificuldade na escolha dos parâmetros para emprego no modelo de previsão.

Neste trabalho propomos e avaliamos um algoritmo evolutivo (notadamente um algoritmo genético – AG) a fim de identificar, dentre 18 diferentes séries temporais de variáveis climáticas, quais são as séries e os tamanhos de janelas que devem ser empregados para melhorar o desempenho de um método de previsão. A localidade escolhida para o desenvolvimento do trabalho tem um alto potencial energético, abrigando a Usina Hidrelétrica de Belo Monte – que deve ser a terceira maior hidrelétrica do mundo; Altamira fica nas margens do Rio Xingu onde reside uma população de mais de cem mil pessoas. O ambiente de previsão, o detalhamento das variáveis obtidas e o algoritmo genético desenvolvido são detalhados na Seção 2. A Seção 3 apresenta os resultados obtidos por meio do algoritmo genético, considerando variações de parâmetros internos do AG. O docu-

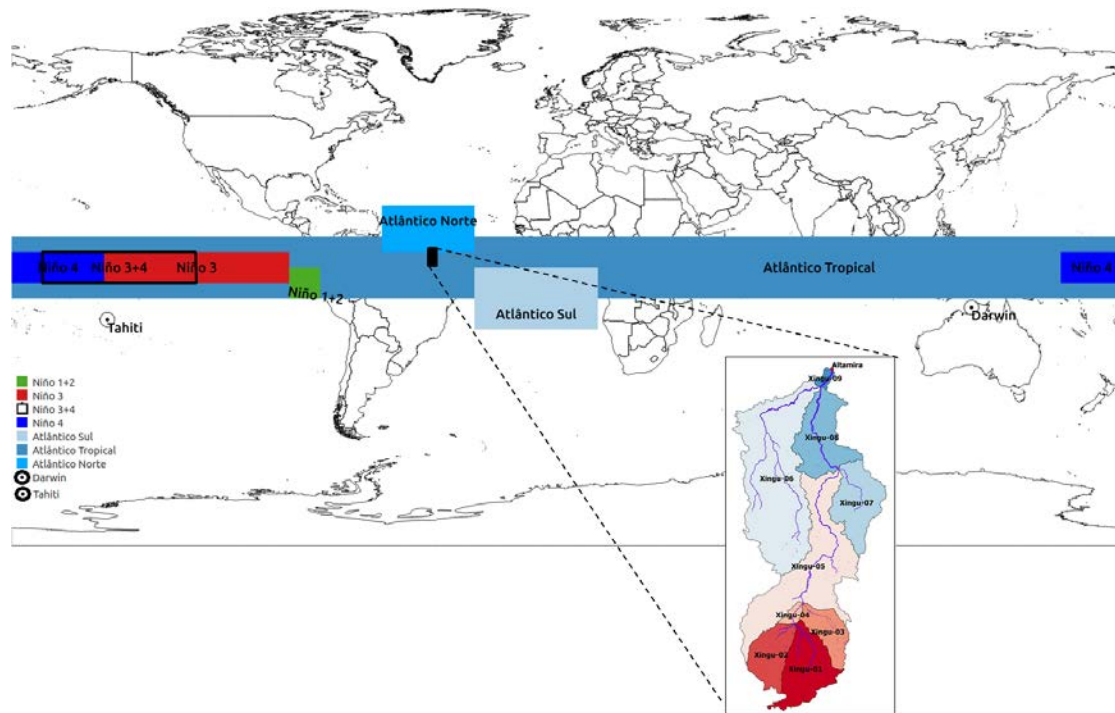


Figura 1. Mapa apresentando pontos de coleta das 18 variáveis de interesse com zoom nas sub-bacias do Rio Xingu. As 18 séries de variáveis climáticas potencialmente empregadas na previsão (atributos e tamanhos de janelas escolhidos pelo AG) são: {Niño1_2, Niño3, Niño4, Niño3_4, Atlântico norte, Atlântico sul, Atlântico tropical, Pressão Darwin, Pressão Taiti, Precipitação na bacia do Xingu - Estações (1..9)}.

mento é finalizado com a Seção 4 onde são apresentadas as conclusões e as sugestões de trabalhos futuros.

2. Métodos

2.1. O ambiente para previsão

A Fig. 1 apresenta o mapa com os pontos de coleta das 18 variáveis de interesse com zoom nas sub-bacias do Rio Xingu. São empregadas 18 diferentes séries temporais para a previsão do nível máximo do rio Xingu em Altamira. Estas séries temporais são: {Niño1_2, Niño3, Niño4, Niño3_4, Atlântico norte, Atlântico sul, Atlântico tropical, Pressão Darwin, Pressão Taiti, Precipitação na bacia do Xingu - Estações (1..9)}. Cinco exemplos de séries temporais (de 1979 até 2014) podem ser vistos na Fig. 2.

Foram utilizados dados de níveis máximos mensais do rio Xingu da estação de Altamira, disponíveis no Banco de Dados Hidrometeorológico da Agência Nacional de Águas (ANA). Os dados de médias mensais de temperatura e pressão dos oceanos Atlântico e Pacífico, e os dados de estimativas de precipitação são oriundos de observações de satélite e interpolados com dados de estações pela National Oceanic and Atmospheric Administration (NOAA). Todas as variáveis são relativas ao período de 1979 a 2014. Na atual implementação, separamos os cinco últimos anos para a previsão. Para utilização dos dados de precipitação, a bacia do Xingu (Fig. 1) foi dividida em sub-bacias de acordo com a proposta de [Pfafstetter 1989].

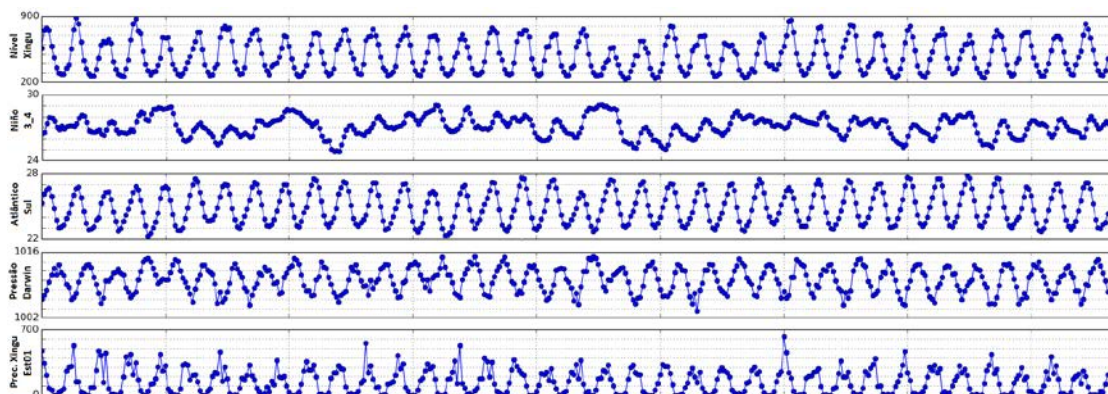


Figura 2. Cinco exemplos de séries temporais (de 1979 até 2014) empregadas neste trabalho. Nesta imagem, a série temporal mais acima é o nível do rio Xingu (a ser previsto). As demais séries (neste exemplo, Niño 3_4, Atlântico sul, Pressão Darwin, e precipitação Xingu Estação 09) são séries potencialmente empregadas na previsão (atributos e tamanhos de janelas escolhidos pelo AG).

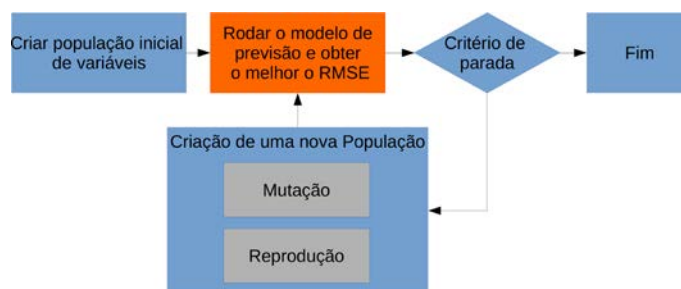


Figura 3. Fluxograma do AG desenvolvido. Para cada indivíduo (cromossomo) é avaliado o modelo de previsão usando regressão linear. O valor de RMSE é utilizado como *fitness*; neste caso, busca-se minimizar o RMSE.

2.2. Estratégia evolutiva para seleção de parâmetros

Por meio de uma estratégia evolutiva (algoritmo genético) é realizada a escolha das janelas e dos atributos das 18 séries temporais de variáveis climáticas a fim de prever as máximas mensais do rio Xingu. O AG seleciona as janelas e os atributos, e chama o módulo de previsão que na atual implementação conta com regressão linear. A Fig. 3 apresenta um fluxograma do AG desenvolvido, onde, para cada indivíduo (cromossomo) é avaliado o modelo de previsão usando regressão linear. A população inicial é criada de forma randômica, preenchendo os genes (lista de bits) com zeros ou uns. Cada indivíduo é um conjunto de parâmetros que deve ser avaliado por meio de regressão linear da base de treinamento e seu *fitness* é a Raiz do Erro Médio Quadrático (Root-Mean-Square Error – RMSE) da base de teste. Desta forma, o AG busca minimizar o erro (*fitness*) que é a reposta do método de previsão; isso ocorre até que o critério de parada seja satisfeito.

A Fig. 4 apresenta o cromossomo proposto neste trabalho. O cromossomo usa codificação binária (valores lógicos 0 ou 1) a fim de utilizar ou não um determinado valor de uma série temporal. Na atual implementação, são consideradas janelas de até 12 meses. Dessa forma temos um cromossomo com 216 genes. O critério de parada definido foi o número de gerações (1.000 gerações). O método de mutação escolhido foi a mutação binária simples, onde o valor do gene é invertido com probabilidade de 20%. O

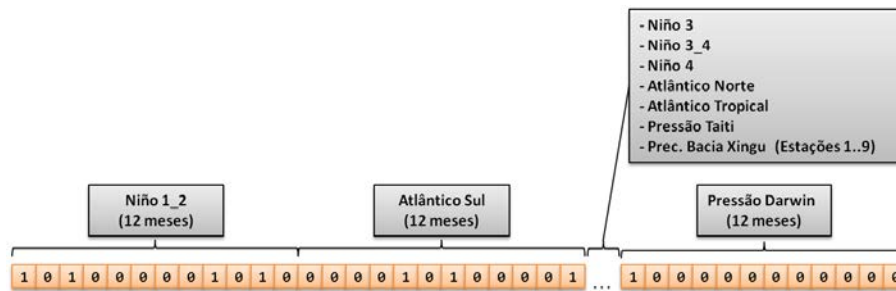


Figura 4. Cromossomo desenvolvido. Valores de 0 ou 1 ativam ou não o uso da informação das séries temporais. Na atual implementação, são consideradas janelas de até 12 meses. Dessa forma temos um cromossomo (indivíduo) com 216 genes (18×12).

método de seleção escolhido foi da roleta viciada, onde a probabilidade de um indivíduo ser selecionado é proporcional a sua aptidão. O método de reprodução foi o cruzamento de um ponto, sendo a taxa de cruzamento adotada em 80%. É aplicado elitismo (melhor indivíduo permanece de uma geração para outra). São realizadas avaliações com 25, 50, 100 e 200 indivíduos.

Um ponto que merece ser ressaltado é que os AGs são soluções interessantes como estratégia evolutiva devido a larga utilização pela comunidade científica, resultados promissores e grande flexibilidade. Algoritmos genéticos são técnicas evolutivas baseadas na evolução guiada por uma metáfora da seleção natural. Os indivíduos no AG são representações de uma solução do problema. Durante o processo evolutivo, são aplicados operadores de reprodução e mutação; para geração de descendentes, os indivíduos são avaliados por meio de uma função de aptidão (*fitness*). Esse processo evolutivo ocorre até que uma condição de parada seja atingida. Detalhes sobre algoritmos genéticos podem ser vistos em [Michalewicz 1996, Rezende 2003].

3. Resultados

Devido ao comportamento estocástico do AG, foram realizadas vinte execuções com tamanhos de população diferentes (25, 50, 100 e 200 indivíduos). A Fig. 5 apresenta a queda do erro (*fitness*) de acordo com o número de gerações, considerando diferentes quantidades de indivíduos. Podemos ver no gráfico que a queda é maior quanto maior o número de indivíduos. Sendo a diferença entre 25 indivíduos maior que a diferença entre os demais conjuntos (50, 100 e 200 indivíduos).

A Fig. 6 apresenta o resultado do melhor indivíduo (20 execuções), obtido de acordo com diferentes tamanhos de população. Sobre os resultados apresentados na Fig. 6 realizamos uma bateria de avaliações estatísticas fim de verificar a diferença entre os conjuntos. Inicialmente, empregamos o teste de normalidade de Shapiro-Wilk, a fim de averiguar a adequação a normalidade dos conjuntos de resultados. Para todos os casos, obtivemos p-valores superiores a 0,05, o que pode ser interpretado como aceitação a distribuição normal. Por ser aceita como adequada a distribuição normal (p-valor superior a 0,05) a comparação entre os conjuntos pode ser feita com o teste t por meio do teste Welch Two Sample t-test. O teste estatístico apresentou que os conjuntos com 100 e 200 indivíduos não tem diferença estatística entre si (p-valor = 0,14). Entretanto, os demais conjuntos (25, 50) apresentam diferença significativa (p-valor inferior a 0,05).

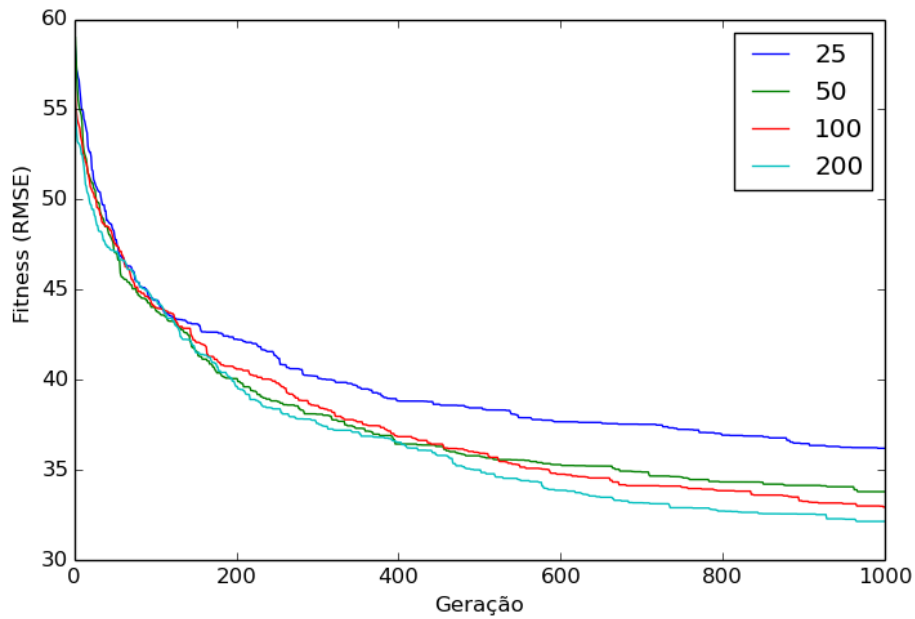


Figura 5. Queda do *fitness* (RMSE) de acordo com o número de gerações, considerando diferentes quantidades de indivíduos. Cada linha apresenta a média de 20 execuções.

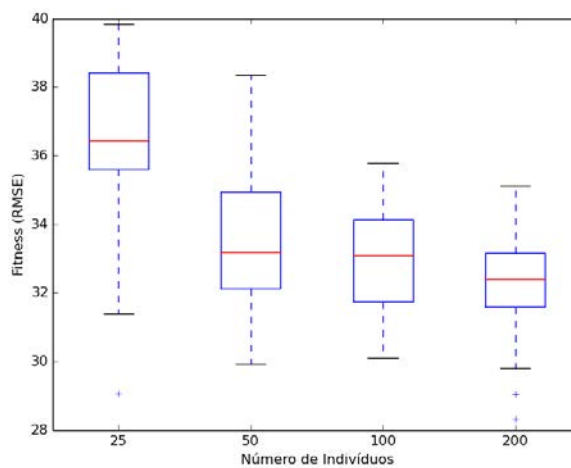


Figura 6. Resultados finais (*fitness*) para diferentes quantidades de indivíduos. Cada boxplot apresenta o resultado de 20 execuções.

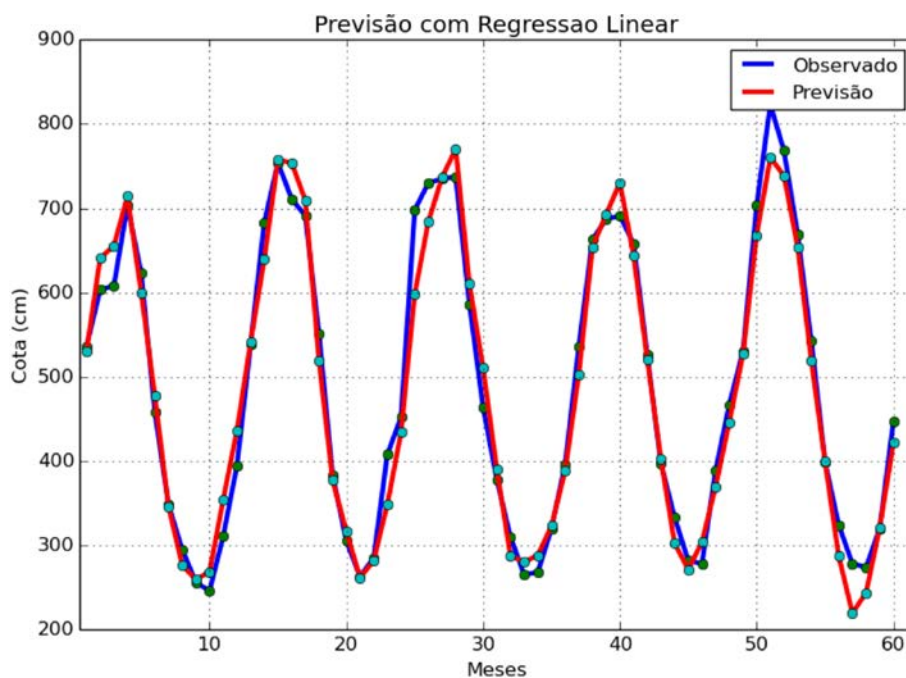


Figura 7. Valores observados e previstos utilizando regressão linear considerando os valores de entrada selecionados pelo algoritmo genético.

Dessa forma, sugere-se o emprego do conjunto com 100 indivíduos por apresentar resultados equivalentes ao conjunto com 200 indivíduos, necessitando metade do tempo computacional para execução.

A Fig. 7 apresenta um exemplo de série temporal de nível com os valores observados e previstos utilizando regressão linear considerando os valores de entrada selecionados pelo algoritmo genético. Podemos ver que as linhas são semelhantes, embora ocorram algumas discrepâncias em momentos de pico. Notadamente, próximos aos meses 10, 25 e 50.

4. Conclusões e trabalhos futuros

Neste artigo, apresentamos como um algoritmo evolutivo pode ser empregado para escolher as variáveis climáticas e o tamanho das janelas de tempo a fim de aumentar a precisão do modelo preditivo. O algoritmo evolutivo é empregado num estudo de caso considerando níveis máximos do rio Xingu, utilizando 18 diferentes séries temporais de variáveis climáticas. Podemos perceber que o AG é eficiente e pode diminuir o erro na previsão de ≈ 60 (RMSE) para ≈ 35 (RMSE). A análise estatística mostrou ainda que o aumento no número de indivíduos (tamanho da população) não necessariamente melhora o desempenho do sistema, neste caso, as soluções com 100 e 200 indivíduos se mostraram equivalentes, entretanto, é importante ressaltar que ambas foram significativamente melhores que empregando 25 ou 50 indivíduos.

Diversos trabalhos futuros são vislumbrados nesta pesquisa, entre eles: (1) Avaliação e comparação de outros métodos evolutivos, como Otimização por Enxame de Partículas e Evolução Diferencial, (2) Avaliação de diferentes modelos de previsão, como Redes Neurais Artificiais ou Máquinas de Vetores de Suporte, e (3) Avaliação de técnicas

de *Deep Learning* aplicadas em séries temporais (por meio de Redes Neurais Recorrentes).

Agradecimentos

Agradecemos ao Prof. Dr. Everaldo Barreiros de Souza e aos colegas do Laboratório de Computação Aplicada, do Instituto Tecnológico Vale, por sugestões no desenvolvimento deste trabalho. Os autores agradecem também ao Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM) pela disponibilização dos dados necessários e apoio na realização deste trabalho. O segundo autor agradece também ao apoio financeiro recebido através da Chamada 59/2013 MCTI/CT-Info/CNPq, processo 440880/2013-0.

Referências

- Blum, A. L. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2):245–271.
- Chen, S.-T. and Yu, P.-S. (2007). Pruning of support vector networks on flood forecasting. *Journal of Hydrology*, 347(1):67–78.
- Dornelles, F., Goldenfum, J. A., and Pedrollo, O. C. (2013). Artificial neural network methods applied to forecasting river levels. *Revista Brasileira de Recursos Hídricos*, 18:45–54.
- Eiben, A. E. and Schippers, C. A. (1998). On evolutionary exploration and exploitation. *Fundamenta Informaticae*, 35(1-4):35–50.
- Franco, V. S. (2007). *Previsão Hidrológica de Cheia Sazonal do Rio Xingu em Altamira-PA*. Dissertação de Mestrado (PPGCA/UFGA).
- IPCC (2013). *INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE - IPCC. Climate Change 2013: The physical science basis. Working Group I Contribution to the Fifth Assessment Report of the IPCC*. Stockholm.
- Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs (3rd Ed.)*. Springer.
- Pfafstetter, O. (1989). *Classificação de bacias hidrográficas - Metodologia de Classificação*. Departamento Nacional de Obras de Saneamento (RJ).
- Rezende, S. O. (2003). *Sistemas inteligentes: fundamentos e aplicações*. Malone.
- Rocha, E. J. P., Rolim, P. A. M., and Santos, D. M. (2007). Modelo estatístico hidroclimático para previsão de níveis em Altamira-PA. In *XVII Simpósio Brasileiro de Recursos Hídricos*.
- Rodrigues, M. M., Costa, M. G. F., and Filho, C. F. F. C. (2015). Proposta de um método para previsão de cheias sazonais utilizando redes neurais artificiais: Uma aplicação no rio Amazonas. In *Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais (WCAMA)*.

Mobile Application for Asian Soybean Rust Tracking in Brazil

José Henrique Andreis¹, Felipe Borella¹, Willingthon Pavan¹, Carlos Amaral Hölbig¹, Jaqson Dalbosco¹, Cláudia Vieira Godoy³, José Maurício Fernandes^{1,2}

¹Universidade de Passo Fundo (UPF)
99.052-900 – Passo Fundo – RS – Brazil

²Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
99.050-970 – Passo Fundo – RS – Brazil

³Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA)
86.001-970 – Londrina – PR – Brazil

{pavan,holbig,61779,jaqson}@upf.br, andreis@ufl.edu,
claudia@cnpso.embrapa.br, mauricio.fernandes@embrapa.br

Abstract. *Brazil is the second-largest soybean producer. With the arrival of Asian soybean rust in the Western Hemisphere in 2001, a considerable amount of resources has been allocated to understand and control this important yield-limiting disease. Due its rapidly dissemination, in 2004, the federal government led an effort to create the Asian soybean rust consortium, with the main goal of coordinating research activities involving public and private sector. This paper describes the development of a mobile application, designed to support the Asian Soybean Rust Consortium to monitor, in time and space, occurrences of Asian soybean rust in Brazil.*

Resumo. *O Brasil é o segundo maior produtor de soja do mundo. Com a chegada da ferrugem da soja no hemisfério ocidental em 2001, uma quantidade considerável de recursos foram alocados para entender e controlar esta importante doença. Devido a sua rápida disseminação, em 2004, o governo federal uniu esforços e criou o Consórcio Antiferrugem, cujo objetivo principal seria de coordenar atividades de pesquisa envolvendo empresas públicas e privadas. Este artigo apresenta o desenvolvimento de um aplicativo móvel, projetado para auxiliar este Consórcio no monitoramento em tempo e espaço, de ocorrências de ferrugem asiática de soja no Brasil.*

1. Introduction

Since the early twentieth century, when soybeans began to be produced and marketed in the United States, this crop has occupied a prominent place in agribusiness worldwide [Roberts 2006, Kim 2005]. In Brazil, this crop has been leveraged by various factors making it one of the most important crop established in the country. With the advance of soybean production in Brazil, there were also the emergence of serious diseases, causing significant damages, which can be avoided or minimized with efficient control and management of diseases [Esker 2007, Godoy 2015]. The Asian Soybean Rust, for example, if not early treated, can cause premature defoliation, preventing the complete

formation of the grains and generating sharp reductions in productivity. Reports of production losses due to the disease have reached 75% in some regions [Yorinori 2005].

In this sense, the CAF (Asian Soybean Rust Consortium – <http://www.consorcioantiferrugem.net>), consolidated by partnerships between universities, research institutions, laboratories and soybean producers, has strongly acted in Asian soybean rust monitoring spreading in Brazil, assisting producers to control the disease and reduce the risks.

2. The Application

Information technologies have been great allies of the CAF to obtain strategic information on disease dispersion and prevention. Considering the importance of fast identification of soybean rust to start an efficient control, minimize, or even avoid losses caused by it. Information technologies have also allowed the development of features that enable collaborative and online work among all actors involved in the management of the disease.

The development of a mobile version of the CAF has the goal of providing immediate access to the disease dispersion information to producers, directly from their smartphones. The CAF mobile application allows the easy visualization of the soybean rust dispersion in Brazil on disease dispersion maps. The map enable producers to become aware of the occurrence of the disease in their region and to anticipate the searching for infections on their fields, consequently minimizing the chances of great losses, which are very common in infections by *Phakopsora pachyrhizi*.

The application was developed using native programming languages and tools for the iOS platform (Swift, Xcode and iOS SDK) and Android platform (Android Studio and Android SDK). The app development follows the MVC (Model View Controller) software architecture patterns. Data are obtained by HTTP requests to a specific RESTful API written in Java and hosted on the CAF server.

Soybean rust infections are confirmed and registered in the system by 175 certified laboratories that make part of the Asian Soybean Rust Consortium and cover all soybean production regions in Brazil. The Figure 1 shows an interaction diagram between client (app), server, and certified laboratories, covering the entire system operation. JSON (JavaScript Object Notation) data-interchange format is used for the communication between the mobile app and the server REST API.

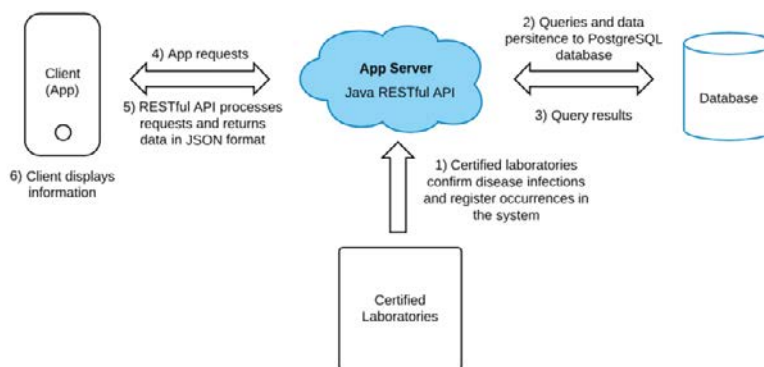


Figure 1. Interaction between client, server and certified laboratories

Besides the disease dispersion map, the application also presents the data through a list for easy access and search for specific cities and occurrences. Either dispersion map and list interface allows the users to have access to detailed information about each registered occurrence, what can be used to be reviewed and compared, aiding producers decision making on the field.

2.1 The Dispersion Map

The main screen of the app is composed of a dispersion map, which aims to provide a general visualization of the disease dispersion. As shown on Figure 2, the current selected season is presented in the navigation bar, on top of the screen. In the central area, red dots represent the cities where occurrences were registered. Additional information such as the name of the city and total of occurrences are presented when the user touch on a dot (Figure 3). The dots have different sizes what are proportional to the number of occurrences allowing an easy perception of the most affected regions. Select previous seasons are also allowed in the system. Users can tap the navigation bar title to have access to a new interface where the seasons are presented allowing its selection and updating e the information on the map (Figure 4).

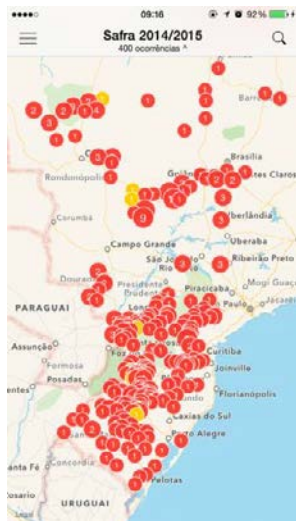


Figure 2. Map of dispersion

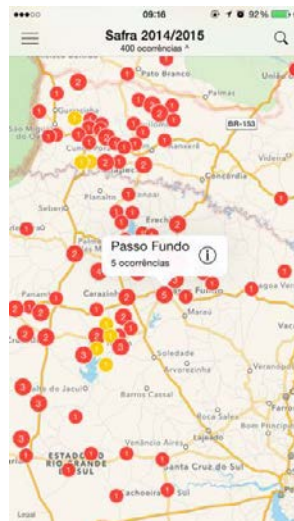


Figure 3. Selecting a city



Figure 4. Selecting a different season

2.2. Searching a City, Detailing of Observations and Notifications

Looking at facilitate the access for specific information, disease occurrences are also presented through a list, where users can easily search for a particular city. This option is available tapping in the search icon on the main screen. Users can enter the name of the city wanted to search or scroll through the list, comparing and checking the total amount of occurrences by city and state.

By the map interface as well as the list, users can access detailed information about each observation by selecting the desired city. After selecting a specific city, a new screen presents all observations registered to that place, bringing information such as: occurrence date; plant phenological stage; person or entity who made the observation; and the laboratory who confirmed the infection by the fungus *Phakopsora pachyrhizi*.

One of the great advantages of using mobile technologies on a disease dispersion system is the possibility to send alerts to users when an infection is registered in their region. The CAF mobile app sends push notifications using the Apple Push Notification Service (APNS) and Google Cloud Service (GCM) for users who opt-in to receive alerts.

2.5. Distribution

The application is available for download on the App Store and Play Store, which is accessible directly from any iOS and Android devices. Since the application was made available it has exceeded the mark of 4,200 unique downloads. The registered downloads come from different parts of the world and reveal the interest in this kind of information, not only by Brazilian producers and researchers, but also by the soybean market around the world.

3. Conclusions

The CAF mobile application keeps track of Asian soybean rust occurrences, in time and space, appearing as an important decision aid tool for the disease management. The application delivers crucial information by making use of real-time dispersion maps allowing the smooth tracking of the disease during the soybean growing season. Besides, it provides a valuable quantification of soybean rust intensity in different years and soybean producing areas in Brazil.

4. References

- EMBRAPA. Doenças e Medidas de Controle. Technical Report Sistema de Produção n. 1, Embrapa Soja, Londrina, Brazil, 2001. Available at: <http://www.cnpso.embrapa.br/producaosoja/doenca.htm>.
- Esquer, P.D., Del Ponte, E.M., Martins, M., Kato, J., Siqueri, F.V., Silva, L.H., and Godoy, C.V. Effects of location and year on the temporal disease progress of Asian soybean rust in Brazil. In: 2007 APS/SON Joint meeting, San Diego, *Phytopathology*, 97:S33 (2007).
- Godoy, C. V., Bueno, A. F., and Gazziero, D. L. P. Brazilian Soybean Pest Management and Threats to its Sustainability. *Outlooks on Pest Management*, 26:3 (2015), 113-117.
- Kim, K.S., Wang, T.C., and Yang, X.B. Simulation of apparent infection rate to predict severity of soybean rust using a fuzzy logic system. *Phytopathology*, 95 (2005), 1122-1131.
- Roberts, M. J., Schimmelpfennig, D., Ashley, E., and Ivingston, M. The value of plant disease early-warning systems: a case study of USDA's soybean rust coordinated framework. Technical Report Econ. Res. Rep. n.18. USDA-ERS, Washington, DC, 2006.
- Yorinori, J. T., Paiva, W. M., Frederick, R. D., Costamilan, L. M., Bertagnolli, P. F., Hartman, G. E., Godoy, C. V., and Nunes, J., JR. Epidemics of soybean rust (*Phakopsora pachyrhizi*) in Brazil and Paraguay from 2001 to 2003. *Plant Dis.* 89 (2005), 675-677.

WCN – Workshop on Cloud Networks

Apresentação

Welcome to the WCN 2016. The workshop will bring together researchers, engineers, and practitioners from academia, industry and government to discuss and advance the state-of-the-art research, advances and practice in the emerging area of cloud-based networks and applications. The WCN 2016 is sponsored by the EUBRASILCloudForum Project (<http://eubrasilcloudforum.eu>), with support of CTIC/RNP and the European Commission.

In this first edition, the TPC of the WCN 2016 selected five papers to be presented in its technical session. These papers present research work related to the latest challenges, technologies, solutions and techniques related to networking within the cloud. Moreover, the WCN 2016 is composed by two panels that will bring discussions about subjects related to cloud networks.

Priscila Solis Barreto (Universidade de Brasília)

Marco Vieira (Universidade de Coimbra)

Comitê de Organização

Coordenação Geral

Priscila Solis Barreto (Universidade de Brasília)
Marco Vieira (Universidade de Coimbra)

Coordenação Local

Tiago Ferreto (PUC/RS)

Comitê de Programa

Andrey Brito (UFCG)
Antonio Augusto Frohlich (UFSC)
Daniel Batista (IME – USP)
Daniel Figueiredo (UFRJ)
Edmundo Monteiro (University of Coimbra)
Eduardo Alchieri (UnB)
Flavio Wagner (UFRGS)
Giovani Gracioli (UFSC)
João Gondim (UnB)
Lisandro Zambenedetti Granville (UFRGS)
Moacyr Martucci (POLI-USP)
Nuno Laranjeiro (University of Coimbra)
Philippe Navaux (UFRGS)
Thais Vasconcelos Batista (UFRN)
Wagner Meira Jr. (UFMG)

Painéis / Palestras

Colaborative Research in Cloud Computing: Future and Challenges

Demi Getschko, Member of the Brazilian Internet Steering Committee (CGI.br)

Philippe O.A. Navaux, Professor at the Universidade Federal of Rio Grande do Sul and
Computer Science Area Representative at CAPES

Karin Breitman, General Manager of the Brazil R&D Center, EMC, Brazil

Interdisciplinary Research for Cloud Computing. Future and challenges

Wagner Meira, Federal University of Minas Gerais, Brazil

Claudio Caimi, Program Manager and Coordinator at Hewlett Packard Enterprise &
Cloud28+

Trabalhos aceitos

Realistic Networking in Generic Multi-Site Cloud Deployments

Carlos de Alfonso (Universidad Politécnica de València)

Ignacio Blanquer (Universidad Politécnica de Valencia)

Improving Hadoop Performance with Reliable Opportunistic Instances in OpenStack

Henrique Truta (Universidade Federal de Campina Grande)

Telles Nóbrega (Universidade Federal de Campina Grande)

Andrey Brito (Universidade Federal de Campina Grande)

Aspectos Práticos da Virtualização de Replicação Máquina de Estados

Ederson Torresini (Instituto Federal de Santa Catarina)

Luis Pacheco (Universidade de Brasília)

Eduardo Alchieri (Universidade de Brasília)

Marcos Caetano (Universidade de Brasília)

Um Mecanismo de Auto Elasticidade com base no Tempo de Resposta para Ambientes de Computação em Nuvem baseados em Containers

Marcelo Abranches (Universidade de Brasília)

Priscila Solis Barreto (Universidade de Brasília)

Segurança de Sistemas de Computação em Nuvem: estado, desafios e perspectivas

João Gondim (Universidade de Brasília)

Priscila Solis Barreto (Universidade de Brasília)

Luis Pacheco (Universidade de Brasília)

Realistic Networking in generic multi-site Cloud Deployments

Carlos de Alfonso¹, Ignacio Blanquer¹

¹ Instituto de Instrumentación para Imagen Molecular (I3M).
Centro mixto CSIC - Universitat Politècnica de València - CIEMAT
Camino de Vera s/n, 46022 Valencia, España

caralla@upv.es, iblanque@dsic.upv.es

Abstract. *This article analyses common problems and scenarios of federated Cloud sites with regards to the networking. It outlines the problems of creating private networks across sites and the migration of active virtual machines between different sites. The article also presents EPFloater, an open-source port forwarding tool to facilitate the creation of those scenarios on multiple, on-premise Clouds. EPfloater can associate specific ports on specific public IP addresses on internal ports and private IP addresses. This reduces the unnecessary consumption of public IP addresses and ports by Cloud resources and enables the reconfiguration of the network topology without reassigning IP addresses.*

Introduction

Cloud computing has become a powerful paradigm to facilitate the way users consume resources. The essential features of Cloud [Mell and Grance 2011]: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service enable building up different "as a service" offerings. Cloud computing has strongly based on virtualization. At the level of the infrastructure resources, virtual disks and Virtual Machines (VMs) enable the implementation of the above essential features. Virtualization reaches the level of application development, with Databases as a Service, Platforms as a Service and other multi-tenant Cloud-based applications and frameworks.

Cloud computing has also become a common resource provisioning mechanism for the private resources of organizations. On-premise Clouds are commonly deployed to facilitate the way users consume resources, isolate resources and data or deal with the highly variable software requirements of the users.

However, virtualization at the network IP level is not as straightforward. IP addresses must be shared and the geographic topology of Internet imposes some important restrictions. By definition, Cloud resources are accessed through a broadband network. Therefore, any externally accessible virtual machine in the Cloud will need a unique endpoint. With the advent of the Internet of Things (IoT), there is a deluge in the need of endpoints to support the communication among the trillions of devices becoming connected to the Internet. IPv4 is undoubtable insufficient to deal with this challenge, and IPv6 is not being adopted as expected. Figure 1 shows a the trends in google searches related to IPv6, quite stable since long ago. It is even more remarkable that one of the most searched terms related with IPv6 includes the term "disable". Disabling IPv6 is a term with increasing interest.

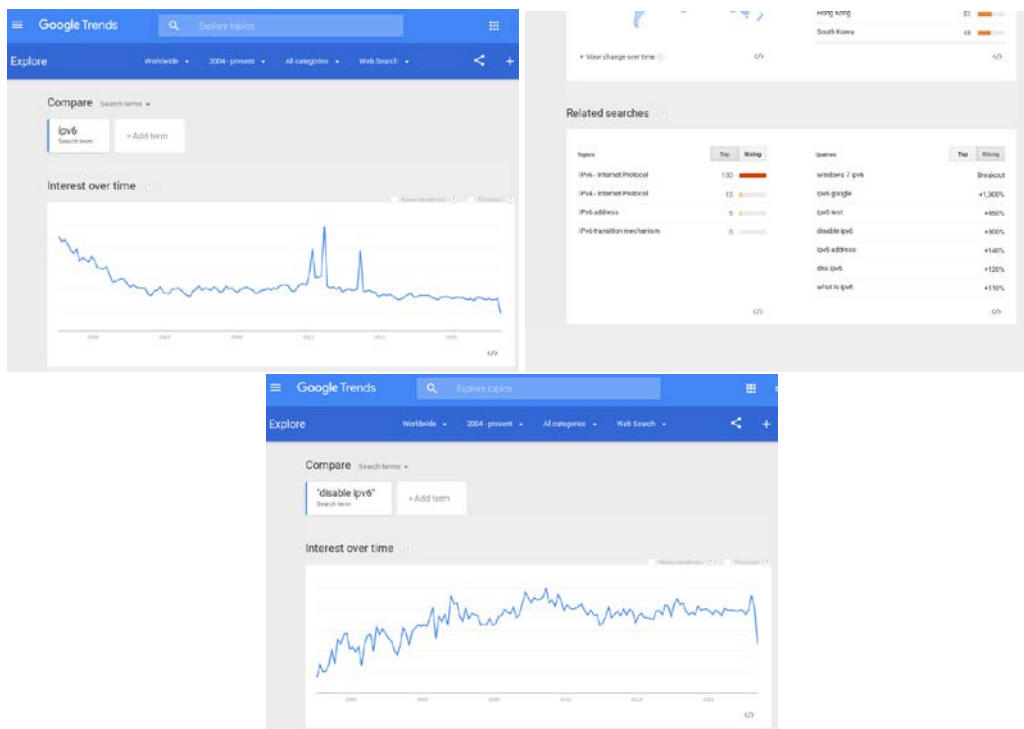


Figure 1. Trends in searches of IPv6 term in google. Top left: Interest has decreased in time; Top right: One of the most raising term related to “ipv6” is “disable”; and Bottom center: Disabling IPv6 is a term with increasing interest.

Actually, although all the VMs and virtual services require network access, not all of them will require direct Internet access from outside. Accessing Internet from resources in private networks is well addressed by Network Address Translation (NAT). Therefore, Cloud providers offer public IP addresses as a service, which can be attached to the Virtual Machines that need to be accessed from outside, and ranges of private IP addresses that can be freely used within the site network boundaries. The range of public IP addresses available (owned by the provider) is typically much more reduced than the number of network resources available.

Network Virtualization and SDN

A traditional switch or router consists of a high-speed data plane where the packets are routed, and a control plane that includes functions for setting up routing and forwarding, access control, quality of service, and so on. Separating the control plane from the data plane made it feasible to run the control plane in software on standard servers, and thus enable the creation of new virtualized controllers and custom-made services easily [Li and Liao 2013]

Such network architectures in which the control plane is decoupled from the data plane have been growing in popularity. This approach provides a more structured software environment for developing network-wide abstractions while potentially simplifying the data plane. We refer to this split architecture as Software-Defined Networking (SDN) [Tootoonchian et al. 2012].

According to the Open Networking Foundation¹, SDN is “is an emerging network architecture where network control is decoupled from forwarding and is directly programmable. This migration of control, formerly tightly bound in individual network devices, into accessible computing devices enables the underlying infrastructure to be abstracted for applications and network services, which can treat the network as a logical or virtual entity” [Open Networking Foundation 2012].

This enables to program the network, instead of having to make a manual configuration through low-level interfaces. In this way, it is possible to automate most of the configuration and management of the network by establishing a centralized controller that manages the global network policy and decides the admittance and routing of flows [Casado et al. 2007].

Network Virtualization (NV) can be defined as the ability to create logical, virtual networks that are decoupled from the underlying network hardware to ensure the network can better integrate with and support virtual environments [SDX Central 2016]. In a Cloud Computing environment, obtaining a resources such as a storage, Virtual Machines, etc. is usually a matter of seconds or minutes. In contrast, typical network management is manually performed by a human often using the command-line-interface (CLI) of each network element, and thus much slower [Tourrilhes et al. 2014]. And this is why SDN can be seen as an alternative to address NV.

A common approach to NV is to deploy a SDN Overlay network on top of a common network infrastructure that does not explicitly support SDN. That means to deploy pieces of software (e.g. virtual switches) that implement the functions made by traditional hardware (i.e. physical switches), and use the existing underlying networks to communicate these virtual components. In this way, the underlying network will act as a “transport network” in whose traffic will be encapsulated the information that is transmitted in the SDN-enabled network. An example of an overlay network can be seen in Figure 2. In this kind of networks, it exists a regular physical network (e.g. a TCP/IP based network). From the point of view of the end-points (i.e. the computers), the overlay network has a behaviour like the physical one (i.e. another TCP/IP based network). But there exist some devices that encapsulate the traffic from the overlay network inside the communication packages that are transmitted in the physical network. Once received these packages in other physical end-points, they are decapsulated, and injected back in the overlay network. As the traffic is encapsulated, it is not part from the physical network: the physical network consider that the encapsulated traffic are regular messages. In the case of the SDN network, the overlay network will be created by using virtual routers that are SDN-enabled. Then, we can dynamically reprogram its components to implement the Virtual Networks for each tenant, by using protocols such as OpenFlow [McKeown et al. 2008].

This is a very straightforward approach, because in this way we are virtualizing the level 2 and 3 components. Any other componet will continue working as usual: the network devices will be virtually connected to the virtual switches and they will be connected between them. Any other application can continue working without the need of any modification, by simply using the common protocols (e.g. TCP, IP, etc.). The other services in a network (e.g. DHCP, routers, load balancers, etc.) will work as usual.

¹<https://www.opennetworking.org>

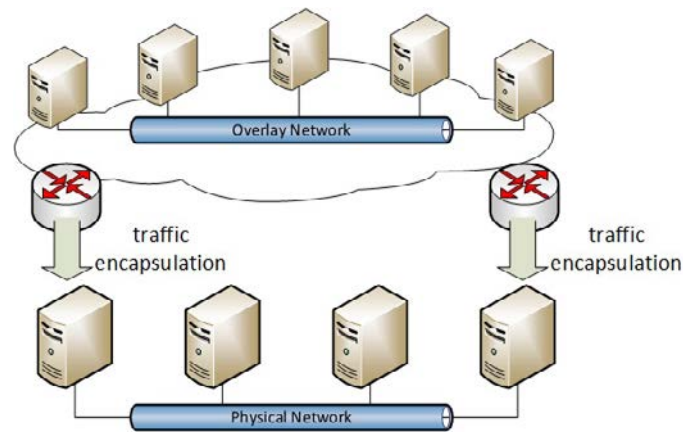


Figure 2. Overlay networks: An overlay network is a logical network that encapsulates the traffic in a regular physical network.

For the sake of completeness about the SDN hype, it is important to place the Network Function Virtualization (NFV) term in this crosswords. The NFV is a trend for virtualizing level 4 to 7 services and use them in a *XaaS* fashion. Some examples of NFV are Firewalls and Load Balancers, and the objective is to be able to dynamically create them and to be able to use them in the VN.

Requirements for a Multi-site Federated Cloud Deployment

One of the Cloud deployment models is the Community Cloud [Mell and Grance 2011]. Community Clouds imply the federation of resources from multiple organizations to provide similar services to any member coming from any of the organization in the federation.

Community Clouds are created to address multiple issues, such as the use of Cloud bursting techniques to forward workload exceeding a site capability to another site in the federation; the placement of processing close to data for efficiency or data protection issues; or to ensure high availability. Usually, federated Clouds coordinate geographically distributed sites that agree on a set of common rules. These rules will enable any user, regardless of the organization he or she comes from, to deploy equivalent VMs in any of the sites of the federation.

A minimal set of services that need to be agreed [EGI 2016] can be:

- A common mechanism to identify and authorise users, so users can seamlessly access any site using their credentials.
- Virtual Machine Image Distribution, to be able to deploy the same images of VMs in the different deployments of the federation.
- Accounting, to be able to account for the resources used at any moment by the users of the federation, and be able to implement balancing policies, and to have a coherent global view of the resource usage.
- Common interfaces, as VM templates (e.g. OpenStack flavours), common access portals, etc, that could provide final users with a global homogeneous view of the system and a single entry point.
- Global Scheduler ², which could provide a general view of the usage of the feder-

²This can be a added-value service

ation and to find the most suitable site for a specific request, based on the dynamic workload, Service Level Agreement and any other requirement.

One of the most interesting aspects of Federated Cloud is the Cloud bursting. A user requesting a reasonable high number of resources on a site could have his / her request fulfilled by a combination of other sites of the federation. This can be solved at the level of individual requests (e.g. a user demanding multiple individual and autonomous instances) [Vivas et al. 2016], but the complexity increases when we want to coordinate the usage of such resources as a whole.

In this sense, one of the most important issues is network access. By definition, VMs in a Cloud are accessed by means of network connections. In particular, VMs are associated to an IP address that uniquely identify them. However, sites normally lack from *public IP addresses*, and therefore the number of VMs that can be directly accessed from outside the site is limited to this number. Typically, Cloud providers offer *Floating IP addresses* that can be dynamically assigned to the VMs that require external access, keeping the VMs by default with only a private IP address. This is the approach that OpenStack³ uses through the NEUTRON⁴ network management system. However, there is no generic system that enables using floating IP addresses for any VM regardless of the Cloud Management Framework.

Even if all the site providers used the same Cloud Management Framework (e.g. OpenStack and NEUTRON), there are scenarios that cannot be addressed by NEUTRON as is. Examples of these use cases are:

- An infrastructure deployed using multiple providers. In this case, if the resources within a site are not enough to host the virtual infrastructure requested, one will expect the federation to deploy VMs across multiple sites. However, if we needed all the VMs to be located in the same private network (which is a common requirement for many scientific applications that share filesystems), it will not be possible to connect VMs across sites without using a public IP address.
- Virtual Machine Migration. In this case, a long-living VM needs to be moved to other provider. There may be multiple reasons to do so, such as maintenance interventions, access to protected data that cannot be accessed beyond a site or the requirement of specific device capabilities. Despite that VM migration can be addressed in common hypervisor and within CMFs resources, the migration across sites will affect the IP address, unless the IP ranges are closely coordinated and mechanisms to avoid assigning the same IP to multiple VMs in different sites, are put in place.

Finally, in most sites we face the issue that despite that a high amount of resources are available, the number of public IP addresses is limited.

EPFloater: an Endpoint Floater

When a user wants to expose standard services (such as SSH, HTTP, HTTPs, etc.) from a Virtual Machine running in the Cloud, he or she requests a persistent floating IP address

³<http://www.openstack.org/>

⁴<http://docs.openstack.org/developer/neutron/>

to be attached to that VM. Therefore, in most cases the VM only uses 3 or 4 TCP/UDP ports, but the provider is using the whole range of 65535 available at each IP address, preventing other resources to use them. This opens the opportunity of multiplexing the available public IP addresses exploiting the whole range of ports. This approach is extremely useful in centres that have a large amount of resources but a limited amount of public IP addresses, as the academic centres in Federated Clouds.

This way, we present *End-Point Floater* (EPFloater)⁵. It generalizes the concept of floating IP address up to the concept of *floating endpoint* (FEP). In the case of floating IP addresses, all the traffic that arrives to the public IP address is redirected to an internal private IP address. In the case of floating endpoints, all the traffic that arrives to a specific port of a public IP address is redirected to another port of an internal IP.

Using the endpoint mechanism, we can use the same IP address for the external access to a large number of VMs. In the context of scientific computation, most VMs are only accessed from the Internet through SSH protocols. Therefore, each floating IP could ideally serve up to 65535 VMs ssh ports. Obviously, all those VMs will have a full connectivity inside the private networks, so in most cases, only a few VMs would require such external connectivity. Considering the example of a virtual cluster, which is a major use case in scientific computing, one will only need such external connectivity for the front end node, which will use the private internal network to mount shared directories and batch queue communications. Moreover, internal nodes will also be able to have external connectivity if needed, using NAT techniques.

As OpenStack manages floating IP addresses, EPFloater works defining network rules for the well-known Linux Kernel firewall, included in most popular distributions: IPTables⁶. This way we do not need neither additional services nor systems that could add an overload, increased complexity or longer execution time. We use packet filtering of IPTables and we forward the TCP and UDP traffic that reaches the pair (Floating-IP, Floating-Port) to the pair (Internal-IP, Internal-port). We use the connection tracking features of IPTables to enrout the returning communication traffic.

As an example, to forward the traffic from the port 10022 of a floating public IP 111.22.33.44 to the port 22 of the private IP 10.10.2.2, EPFloater will introudce the next rules:

```
-A OUTPUT -d 111.22.33.44/32 -p tcp -m tcp \
    --dport 10022 -j DNAT --to-destination 10.10.2.2:22
-A POSTROUTING -s 10.10.2.2/32 -p tcp -m tcp \
    --dport 22 -j SNAT --to-source 111.22.33.44:10022
-A PREROUTING -d 111.22.33.44/32 -p tcp -m tcp \
    --dport 10022 -j DNAT --to-destination 10.10.2.2:22
```

EPFloater has been designed to be installed in a computer that will play the role of a gateway. Therefore, along with the floating endpoint management, it can manage firewalls or it can be placed behind a load balancer to provide high availability.

To ease the interaction with EPFloater, a REST API has been implemented, with the following semantics:

⁵EPFloater can be obtained from <https://github.com/grycap/ipfloater>

⁶<http://www.netfilter.org/>

- GET: It retrieves the information about the forwarding rules created (floating IP and port, internal IP and port, pool of IP addresses, etc.).
- PUT: It creates a forwarding rule from a pair (public IP, public port) to a pair (internal IP, internal port).
- DELETE: It removes an existing connection.

EPfloater also includes some degree of intelligence to request the forwarding end points. In this way, EPFloater can choose the rightmost floating IP address and port for an internal IP address. Then, EPFloater will use heuristics to try to provide the same public IP address for all the redirections of ports of an internal IP address, and it will try to keep the same internal and external port. EPFloater will get track of the used ports and IP addresses and if there are free IP addresses and ports that fulfill the previous criteria it will allocate those.

This behaviour is important from the end-user point of view. The end user could even not be aware of this forwarding in the cases where the desired public port is available. This could be important to the user to minimize the changes with respect to the way standard tools access the resources. For example, it will be desirable that *ssh* is located in port 22 to avoid an extra parameter in the command line, and that *http* traffic is located in port 80, so no changes are needed in the url of a browser.

Networking for Multi-site Federated Cloud Deployments

When planning the network for a multi-site federated Cloud deployment we must consider the heterogeneity of the underlying technologies that will build each site that will be part from the federation. While middlewares like OpenStack include networking components that could be somehow adapted to a multi-site deployment (including components and specific configurations), in a federation there is not any control over which middleware can be used. Each site deploys its Cloud using the middleware that better fits their requirements. As an example, FedCloud [EGI 2016] accepts sites based on OpenStack, OpenNebula or others.

In this sense, it is a good practice to let the sites deploy their Clouds according to their requirements and internal tools, and then establishing a mechanism to integrate the sites. According to this approach, from the point of view of the network, we will consider that each site has its own internal networks that will be isolated from the network traffic from the other sites. From now on, we'll call these networks *the site private network*.

But we need to communicate the sites between them, and that is why we need to introduce some new elements that will enable such communications. In this case, we will create links between the networks using SDN-enabled switches that from now on we will call *boundary switches*. It is important that the boundary switches are SDN-enabled, because we will need to control the flow of the traffic to avoid the overload of the network. The boundary switches will prevent that the traffic from one virtual network created in one site will flow to other site.

As they are switches, the boundary switches will also enable to create shared networks between the different sites, with a behaviour of *Local Area Networks* (LANs). That means that while being in geographically distributed locations, the protocols in a LAN can be still used. The caveat here is that the boundary switches must be able to

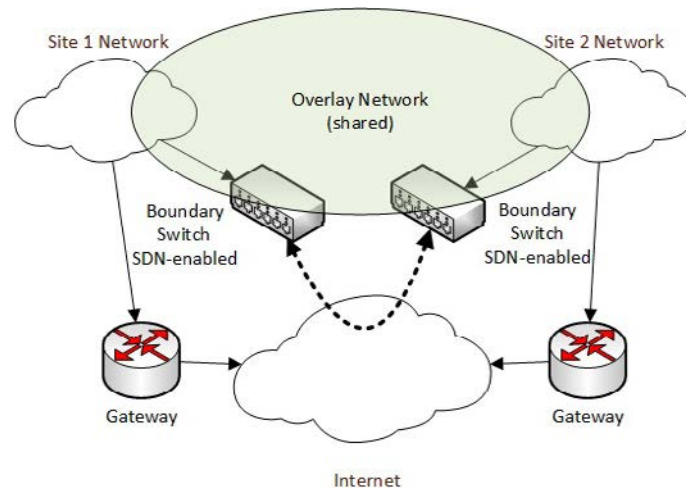


Figure 3. Boundary switches are SDN-enabled switches that enable to create overlay network shared between different sites.

distinguish whether they have to forward the traffic to the other location or to keep it inside the site. Here is where SDN is needed to enable to reprogram the boundary switches to limit the forwarded network traffic.

This scheme is shown in Figure 3 where each site has a gateway, that provides access to the Internet to the VMs that are created inside their networks. But the communication between the VMs using the internal IP addresses is possible because of the link created by the boundary switches. This is a simple yet powerful scheme that provides features such as:

- It is possible to create different networks which are private to each site, provided that the boundary switches are programmed to prevent forwarding the traffic related to that network to the other sites.
- It is possible to create different networks which are shared between different sites, provided that the boundary switches are programmed to forward the traffic whose destination is not in the site's network to the other sites.
- The different networks can coexist in a shared physical network, provided that there are included other techniques to differentiate between the networks. An example is the usage of VLAN or VXLAN techniques.
- It is possible to migrate one VM from one site to other, and the VM will keep its IP address. In this case, the boundary switches have to be programmed to forward the traffic, to reach to the site where the VM has been migrated.

Building the Federated Network

The boundary switches are the key components of this scheme, because they enable establishing a connection between the sites. They could be hardware devices which are SDN-enabled, from well known vendors, and one (or some) of their ports are connected via third party connections to the other sites. But there is a virtual alternative called OpenVSwitch [Open vSwitch 2016] that enables us to test this kind of connection prior to making an invest of money in expensive hardware. Moreover the virtual alternative may be enough depending on the networking requirements of the VMs.

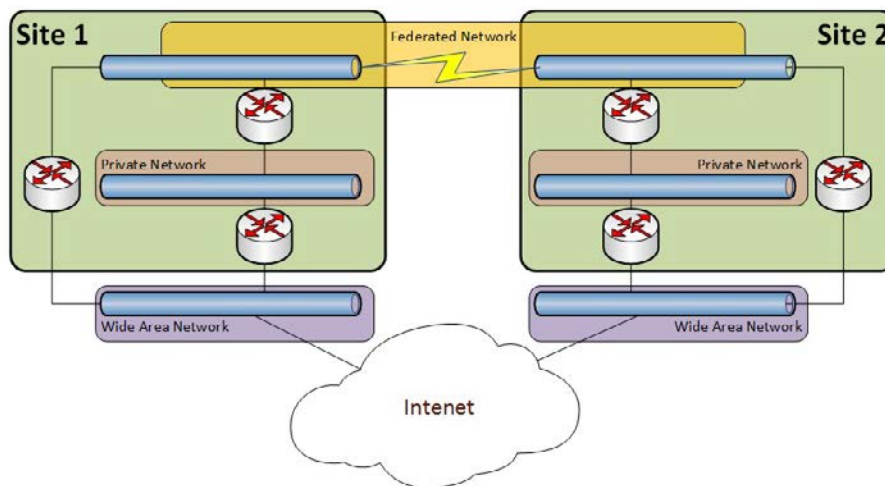


Figure 4. Differentiation of the physical networks to better isolate the traffic of the Virtual Networks.

OpenVSwitch (OVS) enables to create virtual bridges and to make connections between them, by simply establishing tunnels between virtual ports (e.g. GRE or VXLAN). These virtual ports are associated to IP addresses, and the configuration is very straightforward. Moreover it enables to establish complex connections (e.g. all to all), that can be used to enable failure tolerance. OpenVSwitch is SDN-enabled and it is programmable using the OpenFlow [McKeown et al. 2008] standard protocol. So it is a tool that fulfills all the requirements for our connection scheme.

In order to build the federated network the sites should consider three types of physical networks that should be isolated one from each other:

- The Wide Area Network (WAN), that will consist of publicly addressable IP addresses. The VMs in this network will have a public IP address and are potentially accessible from anywhere in the Internet.
- The Private Network (PN), that will be a network isolated from the other sites and from the Internet. The Virtual Networks (VN) created on this PN will be private for the site and the traffic will not be forwarded out. The VMs will have IP addresses to be able to communicate with the other VMs in these VN.
- The Federated Network (FN), that will be networks shared across the federation. The VN created in the FN will be private for the federation, and the traffic will not be forwarded out from the FN neither to the Internet nor the PN. The VMs will have IP addresses to be able to communicate with the other VMs in the VN, but the traffic will be forwarded to other sites, in case that any VM want to communicate with other VM that is not in the site, but is in the VN.

Although it is not mandatory, it is suggested that these networks are physically differentiated, in order to avoid that the traffic in one network will interfere with the performance of the other network. Anyway the traffic should be logically differentiated: in the case of using OVS, the traffic in the FN will use the WAN as a transport network but that traffic will not appear in the WAN, outside the communication established between the GRE or VXLAN tunnels.

The communication between the networks is needed, because the VMs that are

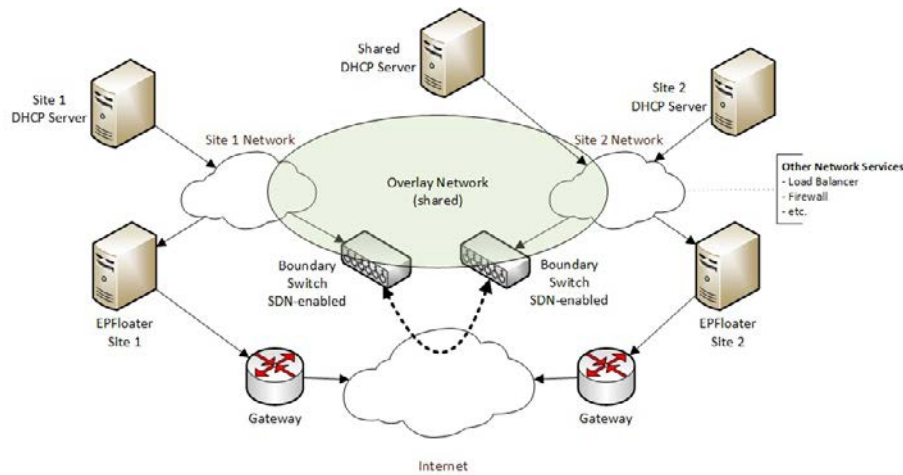


Figure 5. Multi site Cloud deployment with shared overlay networks and common network services.

inside the FN or the PN will probably need to access to the Internet. Moreover, it is possible that some VMs in the FN need to access to the PN. Such communication should be made using gateways, that could be either be provided by default or in a “as a Service” way (as it is made in OpenStack). Anyway it should be possible to create a VM that is connected to more than one network, and to configure it as a gateway for a particular VN of a user.

The differentiation of the networks can be seen in Figure 4. In terms of devices in a Linux operating system, we could think of mapping **eth0** to the WAN, **eth1** to the PN and **eth2** to the FN. Then it should be possible to create network bridges to which the devices of the VMs would be added.

Enhancing the federated Network

The network in a federated Cloud will probably need more than simple network connectivity between the VMs, even accross the sites. So the network should be enhanced by introducing some other services such as:

- Floating IP addresses or Floating End-Points. As described in section 4, public IP addresses are very limited for a site. So it is interesting to include a service that will rationalize the public IP addresses, and offer mechanism to enable VMs to be routable from the Internet.
- DHCP services. The VMs need IP addresses, whether private, public or federated. In any case it is possible to assign static IP addresses to the VMs and to include a mechanism for the VMs to guess such IP, or use the well-known DHCP mechanism. The DHCP servers can be included at any level of the network (FN, PN or WAN). The caveat here is to program the SDN devices to not to forward the DHCP traffic that is from the PN or from the WAN. The DHCP mechanisms could also be offered accross sites, by programming the SDN devices to forward the DHCP traffic from each site to the sites that have DHCP servers.
- Other services such as Load Balancers, Firewalls, etc. All these services would greatly enhance the network, because they offer common tasks for VMs in the network. Moreover, each of them could be offered at different points of the net-

works. E.g. a firewall can be set in front of the WAN, to filter out a part of the traffic, but it would be also interesting to have firewalls in the FN or in the PN. E.g. A Load Balancer could be offered as a whole IP load balancer or as End-Point load balancer with different pair (IP, Port) to balance.

The enhanced network is shown in Figure 5. This scheme provides some unique features up to now:

- Shared networks with DHCP service: It is possible to create VMs that are connected to the FN, and the VMs will ask for an IP address by using DHCP. The boundary switches will be programmed to forward the DHCP requests to the site where it is installed the federated DHCP server, and also to return back the responses.
- VM migration across sites keeping the publicly routable IP address: When a VM is migrated from a site to another, it will keep the federated IP address. The boundary switches are reprogramed to properly forward the traffic to reach to the placement where the VM has been migrated. If we assume that the public IP addresses are granted to the VMs by floating EP/IP mechanisms, the public IP address is translated into a federated IP address. As the traffic in the FN is programmed to be forwarded to the location of the VM, the traffic which came from the public endpoint will also reach to the migrated VM which had a floating EP. The same applies for Load Balancers or the Firewalls.

Conclusions and Future works

We have proposed some recommendations to enable realistic networking in federated multi-site Cloud deployments. By applying this scheme, it is possible to provide networking features of a LAN to a Federated Network that is shared across the sites. In our current proposal we have intentionally left apart the placement of a SDN-controller that would program the different devices across the sites. The current proposal relies on the features of the virtual switches based on OVS, that support protocols to discover IP and MAC addresses, such as the Spanning Tree. The out-of-the-box version of OVS is able to learn the ARP tables of the VMs and to properly forward the traffic. The problem here is that in some circumstances, it would be a slow procedure that could limit the start-up of the VMs and the migration of VMs across the sites. Including a SDN-controller in our scheme is a future work that will greatly enhance the performance of the network. It is also left for a future work making tests on the performance of the federated network in a real multi-site deployment. The main work will consist of identifying common problems of networking in VMs and to test them in a multi-site scenario.

References

- Casado, M., Freedman, M. J., Pettit, J., Luo, J., McKeown, N., and Shenker, S. (2007). Ethane: Taking control of the enterprise. *SIGCOMM Comput. Commun. Rev.*, 37(4):1–12.
- EGI (2016). The EGI federated cloud. <https://www.egi.eu/infrastructure/cloud/>.
- Li, C. and Liao, W. (2013). Software defined networks [guest editorial]. *IEEE Communications Magazine*, 51(2):113.

- McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., and Turner, J. (2008). Openflow: Enabling innovation in campus networks. *SIGCOMM Comput. Commun. Rev.*, 38(2):69–74.
- Mell, P. and Grance, T. (2011). The NIST Definition of Cloud Computing. NIST Special Publication 800-145 (Final). Technical report.
- Open Networking Foundation (2012). Software-defined networking: The new norm for networks. Technical report, Open Networking Foundation.
- Open vSwitch (2016). Production quality, multilayer open virtual switch. <http://openvswitch.org/>.
- SDX Central (2016). What is network virtualization? - definition. <https://www.sdxcentral.com/resources/network-virtualization/whats-network-virtualization/>.
- Tootoonchian, A., Gorbunov, S., Ganjali, Y., Casado, M., and Sherwood, R. (2012). On controller performance in software-defined networks. In *Proceedings of the 2Nd USENIX Conference on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services*, Hot-ICE'12, pages 10–10, Berkeley, CA, USA. USENIX Association.
- Tourrilhes, J., Sharma, P., Banerjee, S., and Pettit, J. (2014). Sdn and openflow evolution: A standards perspective. *Computer*, 47(11):22–29.
- Vivas, J. L., Brasileiro, F., Barros, A., da Silva, G. F., Jr, M. N., de Araújo Neto, F. G., Blanquer, I., Torres, E., Aloisio, G., Badia, R. M., Lezzi, D., Gomes, A. T. A., Cała, J., de Lima, M. J., and Ururahy, C. (2016). *EUBrazilCC Federated Cloud: A Transatlantic Multi-Cloud Infrastructure*. IGI Global.

Improving Hadoop performance with reliable opportunistic instances in OpenStack

Telles Nobrega, Henrique Truta, Andrey E. M. Brito

¹Laboratorio de Sistemas Distribuidos
Universidade Federal de Campina Grande
Campina Grande – PB – Brazil

{tellesmvn,henrique,andrey}@lsd.ufcg.edu.br

Abstract. *In order to be able to keep up with the eventual peaks in demand, cloud environments are most of the time underutilized. Therefore, there has been attempts to exploit these idle resources. One approach is the well-known “spot market” from Amazon Web Services. Nevertheless, spot instances may be terminated at any time, which make their usage not trivial. In this work, we describe a solution that considers the current and historical state of the cloud and predicts the presence of underutilized resources on the short term. With these resources, we are able to create opportunistic virtual machines and speed up the data processing. Since opportunistic instances have a lower priority and can be preempted at any time (as with the spot instances), processing time can be jeopardized. In this case, we use live migration of virtual machines, causing mispredictions to have minimum impact on the processing time. With this framework it is possible to speed up Hadoop executions with only minor side effects by combining opportunism and live migration, making the loss of VMs during the processing a rare event.*

1. Introduction

Hadoop¹ has been one of the major big data processing frameworks for a while now. It uses the MapReduce paradigm [Dean and Ghemawat 2008] to enable distributed processing of large data sets across clusters of computers and is designed to scale up from a single server to thousands of machines. In addition, rather than relying on hardware to deliver high-availability, the framework itself is designed to detect and handle failures at the application layer, delivering a highly-available service on top of a cluster of standard machines, each of which may be prone to failures. Hadoop is designed to run in a parallel manner, and a great amount of machines is necessary to take advantage of this parallelism. Large companies such as Google and Microsoft have huge data centers for big data processing. However, most end users do not have the means to buy servers and build a data center, specially considering that this processing may be needed only eventually, i.e., it might run just periodically (weekly, monthly and so on), or even a single time, making it unfeasible to maintain clusters exclusively for this task.

The lack of own servers, associated with the growth of cloud computing, motivated users to move their data processing jobs to the cloud. Public cloud providers offer resources over the internet in a pay-as-you-go model, enabling users to easily select

¹<https://hadoop.apache.org/>

memory, disk, and processing power to build his own cluster of virtual machines (VMs) without needing to acquire or manage the physical machines.

A common cloud model used on big data processing is the Platform-as-a-Service (PaaS), where, instead of requesting a set of clean virtual machines, the customer may request, for example a completely configured Hadoop cluster, with the requirements (amount, CPU, RAM and disk specs of machines, etc) that satisfy her needs. In this model, the cloud provider is responsible for creating and maintaining the platform, while the user only needs to upload his applications and input data. This model also provides an easy way for the user to scale the cluster. If some demand has a very strict deadline and more machines are needed to make it faster, the user can easily request extra ones, and the reconfiguration will be made by the provider in an automated way.

When buying and scaling clusters of virtual machines for MapReduce processing in Amazon Elastic MapReduce² (EMR), a popular public cloud platform for running MapReduce jobs, the user can choose between reliable or opportunistic VMs to build his cluster. Reliable VMs are machines that have a high Quality of Service (QoS) guarantee, ensuring availability around 99.5% of the time on most public cloud providers³⁴. On the other hand, there are the opportunistic VMs, named *Spot Instances* in EMR, that work on top of unused, but reserved, resources. These spot instances have lower or no QoS guarantees and, thus, can be preempted at any time. As an advantage, spot instances have a much lower per-hour cost when compared to the standard, reliable VMs, named *On-demand*.

Using opportunistic VMs is a way to improve Hadoop's processing at a lower cost in the public cloud scenario, where a user pays lower fees for the extra machines he intends to use. Similarly, one can picture that users in a private cloud could request opportunistic VMs for lower priority application (such as batch processing) in order to preserve available quota for higher priority tasks (such as user-facing applications). In both scenarios, opportunistic instances are meaningful and give the user an opportunity to go over the given capacity if resources are currently available. Moreover, there is no support for opportunistic instances in the two most popular cloud management platforms for private clouds⁵.

However, perhaps initially non-intuitive, the preemption of opportunistic instances may have a considerable impact on the job execution time when compared to using only standard instances [Chohan et al. 2010]. We conducted initial experiments to evaluate the impact of losing a single instance in a Hadoop cluster during an average-size job execution. This experiment, which was executed using Hadoop's default configurations – the most common for non-specialist users, showed that the time for completion can increase up to almost 4 times compared to the case where the machine is not lost. These initial results are depicted in Figure 1 and show that it is indeed not trivial to use opportunistic instances, even in a scenario where failures are supposed to be transparently handled.

Facing this problem of losing opportunistic instances, we propose an approach that

²<https://aws.amazon.com/elasticmapreduce/>

³https://azure.microsoft.com/en-us/support/legal/sla/virtual-machines/v1_0

⁴<https://aws.amazon.com/ec2/sla/>

⁵<https://wiki.openstack.org/wiki/Nova/APIFeatureComparison>

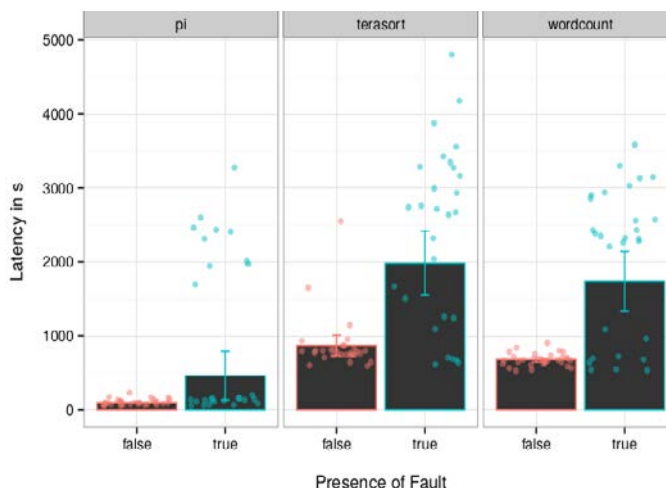


Figure 1. Mean Hadoop job execution time for different types of jobs in a cluster of 4 instances when a failure may occur during the execution. The jobs are the Hadoop canonical benchmarks: word count, terasort and pi calculation. The flag on the horizontal axis indicates the presence of a fault during the execution.

is aware of the short-term future utilization state of the cloud by using basic forecasting techniques, and with this information, allocates a conservative number of opportunistic machines, reducing the probability of preemptions. To make another improvement, instead of killing the opportunistic machine in case of a subestimated prediction, we try to move the virtual machine to a more spare host, using live migration features of the cloud platform. Our approach makes the usage of opportunistic instances in data processing batch jobs more reliable and introduces no considerable overhead in the environment.

The main contributions of this work were included in a framework targeted for users of a private cloud that provides the following features: (i) Predicts load to estimate the amount of available opportunistic resources; (ii) Abstracts the complexity of using opportunistic instances from users; (iii) Uses live migration of VMs to avoid losing the opportunistic machines and jeopardize the job execution makespan. Finally, we evaluate the use of migration in comparison to minimize the impact of lost opportunistic VMs.

The remaining of this paper is divided as follows. In Section 2 we provide relevant background, detailing opportunistic instances, live migration and OpenStack. Next, in Section 3 we present the architecture of our solution. In Section 4 we detail the results. After that, we discuss the state-of-the-art in Section 5. The future work and conclusions are presented in Section 6.

2. Background

In this section, we will discuss relevant background information on opportunistic instances, live migration and OpenStack.

2.1. Opportunistic Instances

The spreading of virtualization technologies arose from the necessity of improving data center utilization through consolidation of servers. Virtualization helps keeping an acceptable ratio between utilized and idle resources. Nevertheless, although it has considerably

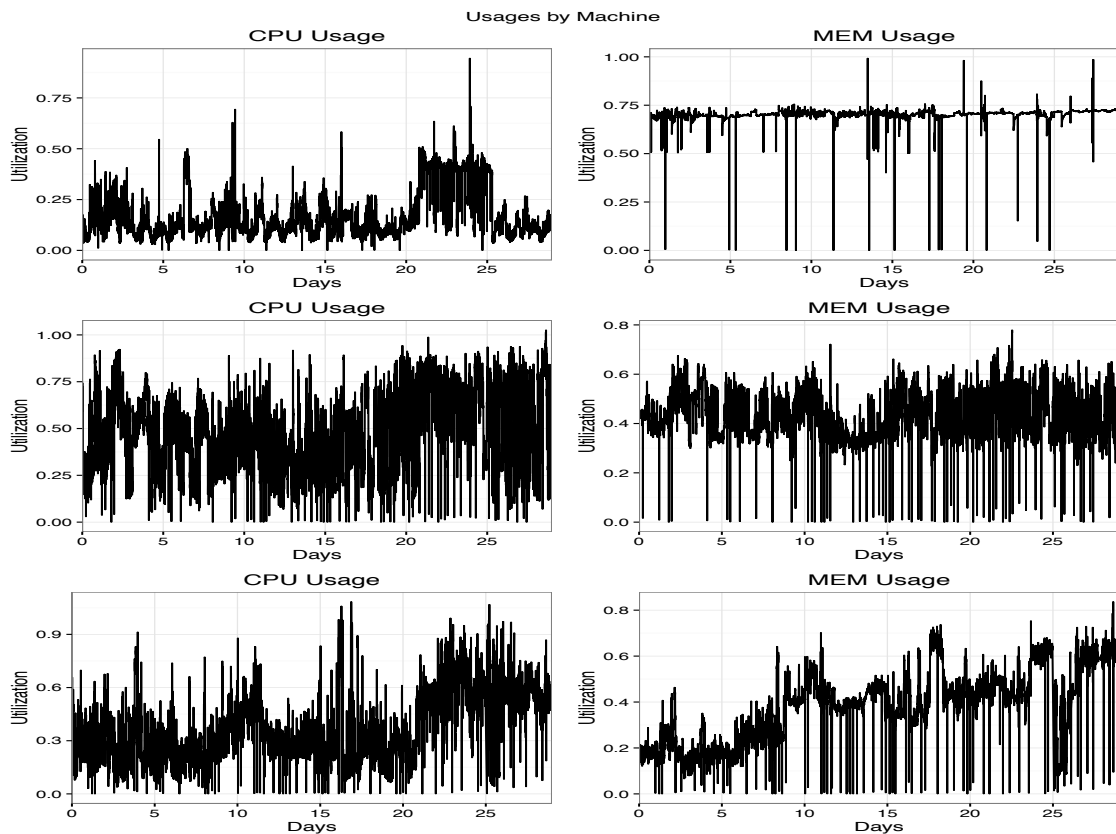


Figure 2. Utilization of CPU and RAM by 3 different servers from a Google cluster.

increased levels of node utilization inside the data center, resources still appear under-utilized due to time varying nature of modern applications. Especially in private clouds, where there is no payment for allocated resources, machines are often overdimensioned or left idle after they are no longer needed. To illustrate how much resources are typically underused, OpenStack, the most used platform for managing cloud computing, considers a default overcommitment ratio of 16 virtual CPUs for each physical CPU⁶. However, this approach exhibits considerable risks [Ghosh and Naik 2012] as performance can quickly deteriorate and, therefore, is not recommended for production environments.

Reiss et al. [Reiss et al. 2012] analyze traces from the Google data center and show that CPU utilization in the data-center do not usually exceed 60%, while RAM utilization does not exceed 50% on average. The full utilization of these resources is not common, as virtual machines remain considerably idle for a large portion of the time. If we take a deeper look into the machine level we can see long periods of low utilization for CPU and RAM along side with some peaks as show in Figure 2. It is also important to note that, as Figure 2 shows the average cluster utilization around 50% of the overall resources, it is common that some machines are using most of its resources whilst others may be at very low utilization.

In a scenario where *(i)* machines are at low utilization, *(ii)* machines are not energy proportional [Hsu et al. 2014, Ponciano et al. 2012] and *(iii)* energy costs are increas-

⁶http://docs.openstack.org/openstack-ops/content/compute_nodes.html

ingly important, opportunistic instances can be an opportunity to optimize the relation between cost and quality of experience in a private cloud. This low average utilization enables the spawning of VMs in hosts that are, in practice, considerably idle, increasing the mean utilization of the cloud.

The downside of the opportunistic instances is that they offer low or no QoS guarantees. As a consequence, opportunistic instances run only as long as they do not interfere the regular instances. If some variation on the utilization of a host happens, e.g. a peak on CPU utilization of previously idle VMs, it would represent a risk on the performance of regular, first-class instances. Consequently, the trivial solution would be to immediately terminate the spare, opportunistic instances.

Furthermore, we have shown that although Hadoop is designed to be fault tolerant and is able to recover from a certain amount of crashes, such as the case of having an instance killed, the job execution will have its performance considerably degraded when this type of fault happens. The penalty is due to the cluster waiting the timeouts, detecting the node failure, and redoing some of the work. To avoid losing these instances, our tool uses live migration of VMs, which will be described in the following subsection.

2.2. Live Migration versus Immediate Termination

In a virtualized world, a common approach to avoid overloaded hosts without needing to kill or stop virtual machines is moving them to underutilized hosts using live migration. This technique moves virtual machine contents from one physical host to another without turning the machines off at any time of the process, unlike the regular migration, where the VM is turned off, its data is transferred and it is turned on again. Live migration offers almost inconspicuous downtime, that range from a few milliseconds to a couple of seconds.

There more common types of live migration are:

- Shared storage live migration: Where the VM disk is stored at a shared storage back-end such as NFS, Ceph⁷ or even proprietary storage appliances. In this case, only the RAM contents are moved to the destination host, and as it is not necessary to move the disk, it causes less overhead in the cloud and takes less time to complete;
- Block Live Migration: Where the whole VM data is moved from its host to another host, having its disk and memory transferred through the network. Although this is simpler than the shared storage migration in terms of configuration and architecture, it has a higher overhead in both involved hosts and in the cloud network nodes.

Live Migration can be used in the case we want to avoid causing harm to a running application, such as a Hadoop job. When a change on the physical host occurs, e.g., a CPU utilization peak, instead of killing the extra, non-priority VM, the VM can be simply moved to another host.

To compare the effects of migration and termination, we have designed experiments, not considering opportunistic instances, in order to evaluate cases in which a load

⁷<http://ceph.com/>

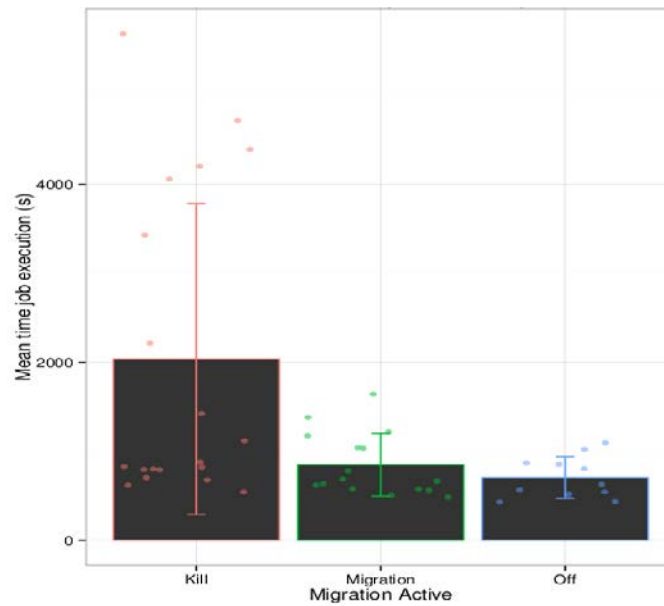


Figure 3. Mean job time execution with a cluster having an instance is abruptly terminated, is migrated or suffers no interference.

spike in the priority VMs would trigger a termination and cases where a load spike would trigger a migration. Figure 3 depicts the result of this experiment. In each of the three scenarios, we have ran a word count Hadoop job with a 3 GB input file in a 4 VMs cluster, with three slave nodes and one master node. Each scenario was executed 20 times. Each VM has 2 vCPUs, 2 GB of RAM and 30 GB of disk. The bars in the figure depict the mean time execution of the job and, from the left to right, the scenarios represent cases where one of the cluster's slave is terminated, migrated during the job, and suffers no intervention. The mean time of the job execution having an instance terminated was, on average, three times longer than what is observed when no interference takes place. Meanwhile, the case where the instance is migrated shows no statistic difference when compared with the no interference case.

2.3. OpenStack and Sahara

With the popularization of cloud computing, lots of tools gained popularity. Among them, there is OpenStack, an open source operational system for managing computational resources of public and private clouds. Among its big features, OpenStack has a big and active community, and a modular architecture with responsibilities divided among many services, each of them with very specific duties. Among the main ones, there are the Identity, Compute and Networking services, named Keystone, Nova and Neutron, respectively. This makes OpenStack's maintenance and new features development much easier, as well as allows the creation of new additional services for other specific tasks. Among these other services, there is Sahara, the Data-Processing-as-a-Service component for OpenStack.

Sahara eases the creation of data processing clusters, such as Hadoop and Spark⁸

⁸<https://spark.apache.org/>

for batch processing and Storm⁹ for real time processing. It enables the creation of cluster templates, defining the type of framework to run, and the resources capacities of slaves and master nodes. Once the template is created, it is possible to launch Hadoop VMs cluster using a single command. We used Sahara in our experiments to automate the management of Hadoop clusters and job launching.

Underneath the GUI it provides for most end users, OpenStack components run on top of well-defined REST APIs. Then, in addition to the REST API, each component of OpenStack maintains a Python library that can programmatically access resources. For example, Sahara, the Data-Processing-as-a-Service component, has a client named *python-saharaclient* that provides Python bindings for most Sahara operations and provide means to interact with other services, such as authentication.

Finally, the resource control at user level is made through a quota management driver. This driver considers static allocation of resources, organized by sets of users, named projects. A user is able to create, for example a Hadoop cluster of VMs in a project he belongs to, depending on whether this project has enough free VCPUs and RAM that matches those VMs. The downside side of this approach, as mentioned before, is checking only allocated resources, not considering actual usage of those resources, leading to low utilization. We present an alternative to this Sahara, using opportunistic VMs in OpenStack in a way that considers the real utilization of resources in the project, bypassing the usual quota management. This solution is described in the next section.

3. Framework Architecture

In this section, we discuss the main features of our framework, describing how it implements a solution to use underutilized resources of an OpenStack cloud to improve performance of a Hadoop job using reliable opportunistic instances.

In a summary, our solution overrides usual OpenStack quota process by using opportunistic VMs, that are aware of the cloud utilization. These over-quota VMs have less priority than the regular ones and could even be killed when other users use up their quota or when allocated regular VMs have usage spikes. At the same time, we minimize the chance of having a Hadoop job losing opportunistic instances by triggering live migrations when a given CPU threshold is reached and there is still room for these VMs in other hosts.

The framework consists of the following three components:

- A daemon that periodically collects information of CPU and RAM utilization of the hosts; in installations with infrastructure monitoring tools this component could be replaced by a simple API call to the monitoring system;
- Client tools that are executed by (human) users to submit jobs;
- A server that receives calls from client tools and interacts with the daemon and the underlying OpenStack services.

The workflow and interactions between the components is depicted in Figure 4. The components are further described in the next subsections.

⁹<https://storm.apache.org/>

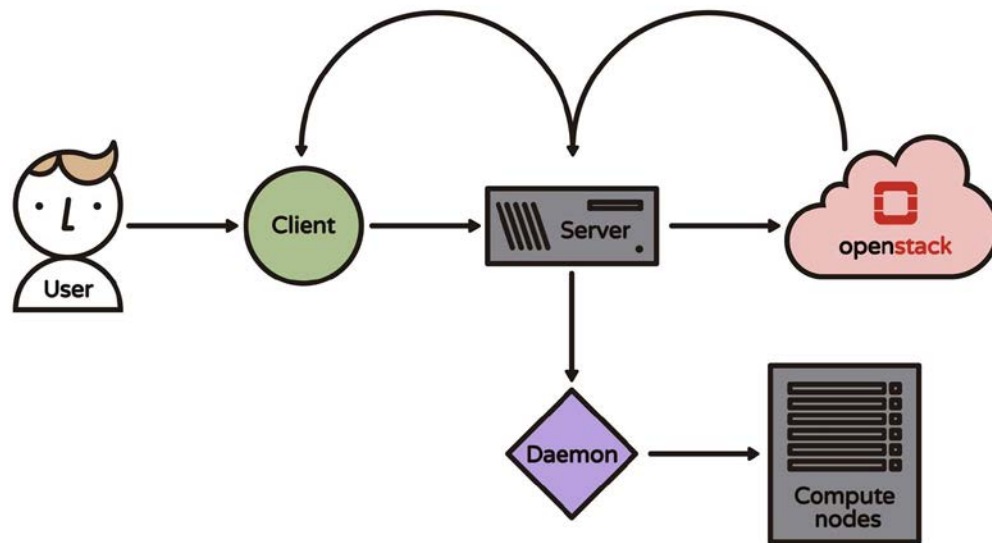


Figure 4. Architecture of the system

3.1. Components

The *daemon* is the component that runs in background and is responsible for collecting CPU and RAM utilization information from the hosts in the cloud. In its simplest implementation, the daemon collects utilization information of each host every 5 minutes through SSH connections. This information is then stored in a database used by the server component (see below). In installations that include an infrastructure monitoring system, the daemon can be replaced by configuring the *server* component to directly access this monitoring system.

The *client* is the entry point of the tool for the user. The user could be a human or an automation tool. The client is designed to be composed of an executable, and a configuration file. This configuration file must contain the required parameters, which include the address of the server, user credentials on the OpenStack cloud and data on the Sahara job specific parameters, like input and output data sources, job template and minimum cluster size to use in case opportunism is not available. These parameters are mandatory for executions of jobs using OpenStack Sahara. The client communicates with the main component, the server, using a simple HTTP POST call. The client will then wait for the server to return information about the job, such as completion status, time of execution, and amount of opportunism used.

The *server* is where the main flow of action happens. It receives a request from the client via HTTP and then starts the processing. It is responsible for connecting to OpenStack using the proper OpenStack Python service clients¹⁰, such as `python-saharaclient` and `python-novaclient`. As mentioned before, these clients provide a python API to facilitate OpenStack operations, making it easier to deal with python code, instead of making direct HTTP calls, abstracting some details like authentication and error handling. During its processing of a request, it executes the following steps: (i) it checks for quota availability; (ii) then it requests the creation of the processing cluster, considering the potential availability of additional resources; (iii) it launches the job to be executed; and, finally,

¹⁰<https://wiki.openstack.org/wiki/OpenStackClients>

(*iv*) it monitors the utilization of physical hosts to manipulate opportunistic instances when needed.

In order to execute step (*ii*) above, the server needs to predict future system load. Therefore, one of the main parts of this system is the load predictor, which will be explained in the next subsection.

3.2. Cloud workload predictor

The workload predictor lives inside the server and receives CPU and RAM utilization history of the hosts from the daemon and forecasts how occupied the host will be on the next prediction horizon. With this information, it computes the number of available machines on the next time window, and use this information to determine the available size of the cluster to be created for the user.

The predictor uses 4 different functions, *mean*, *rwf* without drifting, *rwf* drifting and *naive*. The *naive* and *rwf* methods repeat the previous value from the given input, the *rwf* with drift does a random walk with a drift model applied to the input and the *mean* method gets the input and applies to the mean of that as the predicted value.

The prediction works by receiving a history of workload data. These data will be passed to each of the prediction functions. The functions will return a prediction value and a confidence interval. We chose the conservative approach of considering the top of the confidence interval as the workload predicted value. Therefore, we can give a 95% probabilistic guarantee that we will use to decide about available resources. We are also conservative in the last stage to select which function gives the best result. Between the 4 predictors we select the one that gives the lowest number of available machines. Being conservative is important because even though we may not fully explore available capacity, we avoid providing resources that are too risky. The Figure 5 shows the predicted number of machines by prediction function in comparison with the actual value for the instances that would indeed be available.

Currently, we are using 30-minute prediction windows as most of Hadoop jobs tend to have such duration [Kavulya et al. 2010]. In the future, we will consider also a predictor for job duration that considers the history of submission for that user.

3.3. Framework implementation

Our framework was implemented in an OpenStack cloud in the Kilo version, using Ceph for instance and object storage¹¹, with Sahara, installed in the Kilo version, as well. Although, this is not essential to our approach it simplifies the management of Hadoop clusters. Using object storage as an input to Hadoop jobs not only eases the management of the data and produces more predictable performance, but, in our case, also rules out the need for managing the number of data replicas when killing instances. Object storage is easy to be made reliable and scalable and is, therefore, the primary storage option for big data.

¹¹The most recent OpenStack User Survey, from October 2015 [Ope 2015], shows that 62% of OpenStack installations have Ceph as a backend for block storage. Although the survey does not detail the percentage of installations that use Ceph for instance and object storage, we have anecdotal evidence that the percentages should be similar.

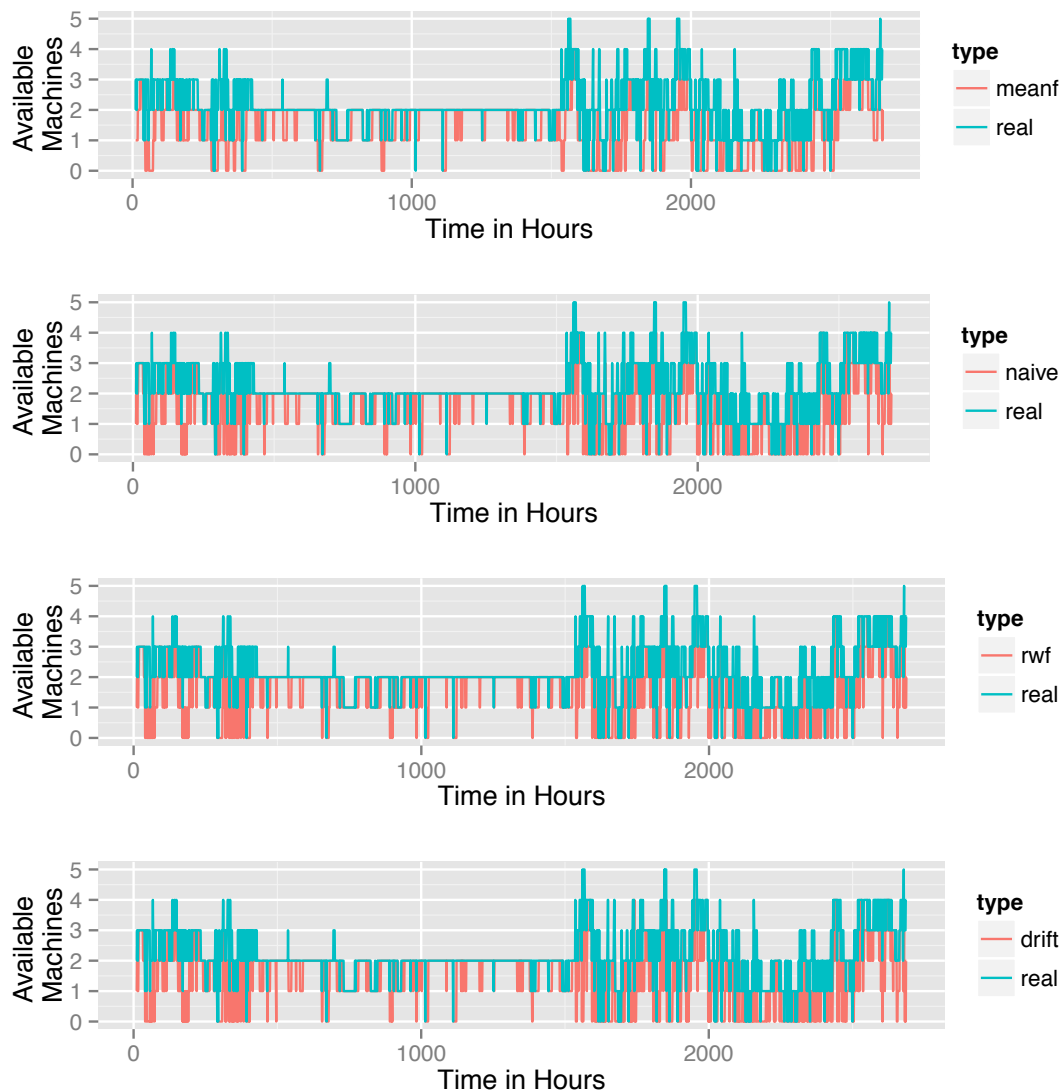


Figure 5. Number of machines predicted by the different prediction functions used in our workload predictor compared with the real number of available virtual machines based on a single server from the google traces.

The implemented tool is currently available to use. Code, documentation and installation instructions are available in our Git repository, at the following *URL*: <https://git.lsd.ufcg.edu.br/tellesnobrega/opportunistic-hadoop/>. In the root directory, there is a README file with major considerations, and in each module sub directory (client, daemon and server), there is a more detailed README showing which dependencies are required, how to setup the environment, and how to configure each component.

4. Results

In this work we found that the use of opportunistic instances improve Hadoop completion time. Nevertheless, the cost of losing an opportunistic instance while the job is running can be too high. By using a workload predictor we can give opportunistic instances a

higher QoS, that will enable the use of the cheaper instances with a probabilistic guarantee that the job duration will not be jeopardized.

Figure 6 shows the completion time of Hadoop wordcount job on three different scenarios. In the first, we are using opportunism and there is no fault. In the second, we introduce a fault but we migrate the instance instead of killing. In the last we do not use opportunistic instances. As depicted in the graphs, even by using a small amount of opportunistic instances the improvement can be seen. As discussed previously, killing opportunistic VMs is undesirable, and it can be seen that migration has no impact on the overall completion time, indicating that opportunistic instances are a viable option to improve the job execution time.

It is important to notice that by using the workload predictor the worst-case scenario of having an instance killed becomes rare, since it only happens in the combination of two situations: (i) The workload prediction, which is currently very conservative, is significantly wrong, i.e., unexpected sudden utilization peaks on the cloud; and (ii) there is no room in other hosts to migrate VMs, which is also uncommon, as depicted in the resource utilization graphics of production environments (see Figure 2). Therefore, we can say that with our approach the risk of using opportunistic instance to improve Hadoop processing is significantly decreased giving the user a safer and cheaper way to create data processing clusters in the cloud.

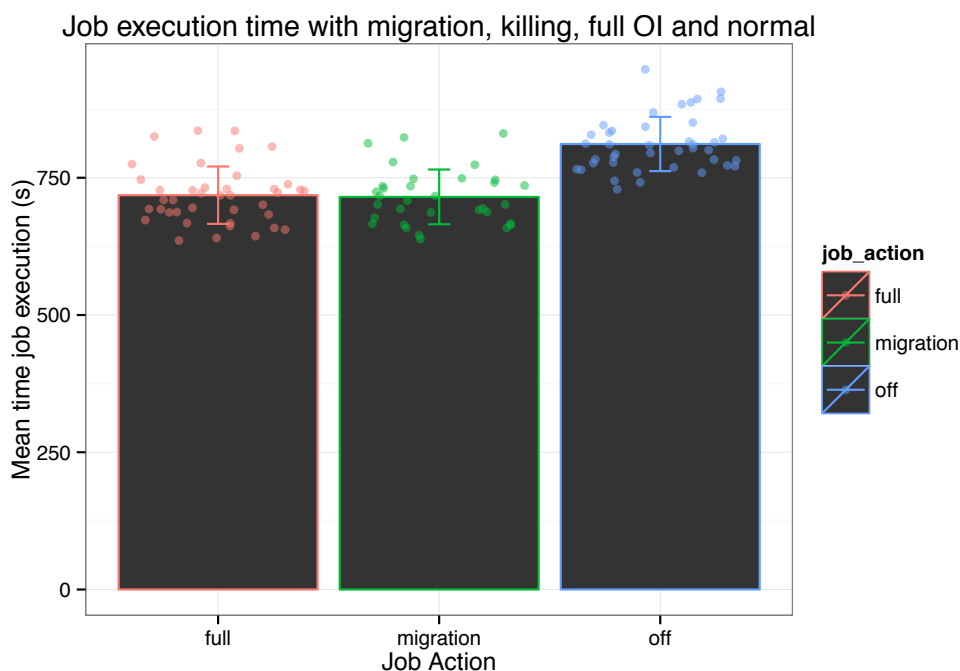


Figure 6. Mean Hadoop WordCount job completion time in the scenarios with opportunistic instances that are preempted and migrated, with opportunistic instances that are not preempted, and without opportunistic instances.

5. Related Work

Recent studies in Hadoop running on the cloud focus on public clouds cases where the main goal is to reduce cost when trying to speed up processing. Tian and Chen

[Tian and Chen 2011] present a model to estimate how long a Hadoop job will run in order to make better provisioning decisions. The model learns, from sample runs of a job, how complex some components are and estimate how long a job will take to finish for a large dataset.

Chohan [Chohan et al. 2010] experiments Hadoop's performance using Amazon Spot instances. He shows that a speed up on completion time is possible when no fault happens, but failure makes processing slower and more expensive. The authors propose a cost function in order to give a better estimate on how to bid for opportunistic instances trying to avoid losing them during processing run time. Our work uses parts of this paper as motivation differing on the type of cloud that we analyzed. Householder [Householder et al. 2014] et al. gives an overview of resource oversubscription on clouds, arguing that overload is a risk of the technique, giving us even more assurance that opportunistic instances are the right path to increase resource utilization at lower cost.

Carvalho et al. [Carvalho et al.] present the novel opportunistic instances named Economy. The instance is cheaper than on-demand and slightly more expensive than Spot instances. They achieve the goal of making an opportunistic instance with SLO by predicting cluster utilization and so guaranteeing, probabilistically, that resources will be available for a time window. The guarantees can be higher or lower depending on how much resource is being used for spawning opportunistic virtual machines. The authors also use time series to predict cluster behavior. They conclude that the new type of instance is better for both users and cloud providers, because it provides a cheaper alternative to the clients and the cloud provider has up to 60% higher profit. This work motivated us in improving the current paper with a prediction model of the cloud in smaller scale (minutes, hours or days) giving the user a QoS on opportunistic instances for data processing in the cloud.

The authors in [Voorsluys and Buyya 2012] propose a provisioning policy that aims to enable MapReduce to be run on spot instances by calculating the urgency of the job and bidding accordingly to guarantee the resources will be available until the end of the computation. They also present four techniques to achieve fault tolerance due to out-of-bid situations only.

In the area of the resource management through live migration, we highlight the work of Wood et al. [Wood et al. 2009], that created Sandpiper, a system that aims to eliminate points with high utilization of resources, called hotspots, that are considered prone to failures and with risk of violating SLAs. This system has both a black-box strategy, that is fully OS and application agnostic and a gray-box strategy, that considers application information. Once a hotspot is detected, live migration is triggered to move a set of virtual machines to under utilized hosts and eliminate the hotspots. As a result, they show that, even in the black-box approach, they are able to mitigate hotspots as far as necessary to not violate SLAs. We use a simpler but similar approach to detect whether a migration of an opportunistic instances needs to be triggered, based on resource utilization thresholds.

Ye et al. [Ye et al. 2012] proposed a technique to migrate groups of virtual clusters, that consist on a group of VMs that serve the same application. They investigate several approaches to make the migration of clusters faster and safer than what would

happen if the VMs were treated separately. They conclude that migrating a whole cluster concurrently is not recommended, specially if the method used is block live migration. In our work, we do not migrate an entire cluster, but this might be considered in the future if we intend to provide a new type of cluster that consists entirely of opportunistic instances.

Finally, there are studies that investigate how live migration interferes on the running instance during and after such migration. In [Xu et al. 2014], Xu et al. propose iAware, a VM migration strategy that analyzes VM metrics through running sets of workloads aiming to avoid the violation of SLAs. They show that it is possible to achieve some of the live migration goals (consolidation, load balancing) without degrade performance. This differs from our work because we aim at in not interfering the other instances on the compute node, as they have higher priority.

6. Conclusion

Intuitively, for most Hadoop jobs, increasing the size of the cluster will speed up its execution. Opportunistic instances emerged as a cost-efficient way for users to scale their clusters, using allocated but not utilized resources of data centers.

Because opportunistic instances have lower priority than the regular instances, they can be preempted. Loosing an instance of the cluster for any reason can drastically increase the job execution time. We then proposed a predictive approach to better estimate the availability of opportunistic resources. Additionally, for the cases where those predictions are not accurate enough, and the utilization is higher than expected, the system tries to move opportunistic instances away from busier machines, without having to kill or to turn them off. The combination of migration and workload prediction were seen to be a good alternative to speed-up the job processing and we observed that having the instance migrated showed no statistic difference on the job execution time when compared to the case where the opportunistic instance stays fixed and available during the complete execution.

As future work, we want to expand our evaluation for different types of data processing jobs and consider a job duration predictor. Using a predictor for the job duration as well as for the workload can help further minimize undesired consequences to the environment and to the executed jobs by better computing the amount of opportunistic resources to be used.

Acknowledgment

This work was partially funded by the EU-BR BigSea project (MCTI/RNP 3rd Coordinated Call).

References

- (2015). Openstack user survey. <https://www.openstack.org/assets/survey/Public-User-Survey-Report.pdf>.
- Carvalho, M., Brasileiro, F., and Wilkes, J. Long-term SLOs for reclaimed cloud computing resources.
- Chohan, N., Castillo, C., Spreitzer, M., Steinder, M., Tantawi, A., Krintz, C., and Barbara, S. (2010). See Spot Run: Using Spot Instances for MapReduce Workflows. In *2nd USENIX Workshop on Hot Topics in Cloud Computing, HotCloud*.

- Dean, J. and Ghemawat, S. (2008). Mapreduce: Simplified data processing on large clusters. *Commun. ACM*, 51(1):107–113.
- Ghosh, R. and Naik, V. K. (2012). Biting off Safely More than You Can Chew: Predictive analytics for resource over-commit in IaaS cloud. In *Proceedings - 2012 IEEE 5th International Conference on Cloud Computing, CLOUD 2012*, pages 25–32.
- Householder, R., Arnold, S., and Green, R. (2014). On Cloud-based Oversubscription. *International Journal of Engineering Trends and Technology*, 8(8):425–431.
- Hsu, C. H., Slagter, K. D., Chen, S. C., and Chung, Y. C. (2014). Optimizing energy consumption with task consolidation in clouds. *Information Sciences*, 258:452–462.
- Kavulya, S., Tan, J., Gandhi, R., and Narasimhan, P. (2010). An Analysis of Traces from a Production MapReduce Cluster. In *Proc. of the CCGrid - IEEE/ACM Int. Conf. on Cluster, Cloud and Grid Computing*, pages 94–103.
- Ponciano, L., Brito, A., Sampaio, L., and Brasileiro, F. (2012). Energy efficient computing through productivity-aware frequency scaling. In *Cloud and Green Computing (CGC), 2012 Second International Conference on*, pages 191–198.
- Reiss, C., Ganger, G. R., Katz, R. H., Kozuch, M. A., and Tumanov, A. (2012). Heterogeneity and Dynamicity of Clouds at Scale: Google Trace Analysis. In *IEEE SOC Conference*, page 13.
- Tian, F. and Chen, K. (2011). Towards optimal resource provisioning for Running MapReduce programs in public clouds. In *Proceedings - 2011 IEEE 4th International Conference on Cloud Computing, CLOUD 2011*, pages 155–162.
- Voorsluys, W. and Buyya, R. (2012). Reliable provisioning of spot instances for compute-intensive applications. In *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, pages 542–549.
- Wood, T., Shenoy, P., Venkataramani, A., and Yousif, M. (2009). Sandpiper: Black-box and gray-box resource management for virtual machines. *Computer Networks*, 53(17):2923–2938.
- Xu, F., Liu, F., Liu, L., Jin, H., Li, B., and Li, B. (2014). iAware: Making Live Migration of Virtual Machines Interference-Aware in the Cloud. *IEEE Transactions on Computers*, 63(12):3012–3025.
- Ye, K., Jiang, X., Ma, R., and Yan, F. (2012). VC-migration: Live migration of virtual clusters in the cloud. *Proceedings - IEEE/ACM International Workshop on Grid Computing*, pages 209–218.

Aspectos Práticos da Virtualização de Replicação Máquina de Estados

Ederson Torresini¹, Luiz A. Pacheco², Eduardo A.P. Alchieri² e Marcos F. Caetano²

¹Campus São José

Instituto Federal de Santa Catarina

²Departamento de Ciência da Computação

Universidade de Brasília

Abstract. *Virtualization could be seen as a technique that helps the development of fault-tolerant applications, since it maintains each virtual machine isolated from the others. State Machine Replication (SMR) is an approach widely used in the development of fault-tolerant applications. With the aim to improve the performance and utilization of computational resources available for the applications, recently some architectures and algorithms were proposed to SMR that uses virtualization, through both virtual machines and containers. This work aims to address practical aspects of the use of virtualization to implement SMRs. We developed an application on top of BFT-SMART, a library to SMR, and executed a set of experiments to assess practical implications about the use of virtualization to implement SMR.*

Resumo. *A virtualização pode ser considerada uma técnica que favorece o desenvolvimento de aplicações tolerantes a falhas, visto que cada máquina virtual fica isolada. No contexto de tolerância a falhas, a Replicação Máquina de Estados (RME) é uma abordagem amplamente utilizada. Com o objetivo de melhorar o desempenho e o aproveitamento dos recursos computacionais disponíveis para uma aplicação, recentemente foram propostos arquiteturas e algoritmos para Replicação de Máquina de Estados tolerante a faltas usando virtualização, tanto através do uso de máquinas virtuais quanto através do uso de containers. Este trabalho visa abordar aspectos práticos do uso de virtualização na implementação de RMEs. Através do desenvolvimento de uma aplicação sobre o BFT-SMART, biblioteca para RME, e da realização de um conjunto de experimentos, este trabalho busca analisar as implicações práticas da implementação de um RME usando tanto máquinas virtuais quanto containers.*

1. Introdução

A virtualização não surgiu recentemente, nas décadas de 60 e 70 a IBM já desenvolvia sistemas operacionais que suportavam virtualização [Goldberg 1973, Goldberg and Mager 1979]. Depois disso, o *hardware* ficou mais barato e a virtualização perdeu importância na década de 80, voltando a ganhar atenção a partir da década de 90 com o surgimento da linguagem de programação Java [Laureano and Maziero 2008].

A virtualização pode ser considerada uma técnica que favorece o desenvolvimento de aplicações tolerantes a falhas, visto que cada máquina virtual fica isolada [Dettoni et al. 2013]. De fato, com o objetivo de tolerar falhas, diversos trabalhos utilizaram virtualização [Garfinkel and Rosenblum 2003, Laureano et al. 2004, Jiang and Wang 2007, Dettoni et al. 2013, Netto et al. 2016].

No contexto de tolerância a falhas, técnicas de replicação são amplamente utilizadas para implementação de sistemas que toleram falhas em partes das réplicas. A Replicação Máquina de Estados (RME), por exemplo, é uma abordagem amplamente utilizada na implementação destes sistemas [Schneider 1990, Castro and Liskov 2002]. Esta abordagem consiste em replicar os servidores e coordenar as interações entre os clientes e as réplicas dos servidores, fazendo com que as várias réplicas apresentem a mesma evolução em seus estados.

Com o objetivo de melhorar o desempenho e o aproveitamento dos recursos computacionais disponíveis para uma aplicação, recentemente foram propostos arquiteturas e algoritmos para Replicação de Máquina de Estados tolerante a faltas usando virtualização, tanto através do uso de máquinas virtuais [Dettoni et al. 2013] quanto através do uso de *containers* [Netto et al. 2016].

Neste sentido, este trabalho visa abordar aspectos práticos do uso de virtualização, tanto através de máquinas virtuais quanto através de *containers*, na implementação de replicação Máquina de Estados. Para isso, utilizamos a biblioteca BFT-SMART para desenvolver uma aplicação tolerante a falhas através de RME e realizamos um conjunto de experimentos usando estas duas formas de virtualização, a fim de entender as implicações práticas da adoção de cada uma delas.

O restante deste texto está organizado da seguinte forma. A Seção 2 apresenta os conceitos de máquinas virtuais e *containers*, enquanto que a Seção 3 aborda os conceitos envolvendo uma RME e descreve brevemente o BFT-SMART. Os experimentos realizados, bem como a análise dos resultados obtidos, são apresentados na Seção 4. Finalmente, a Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Virtualização

Esta seção descreve as duas formas de virtualização que serão usadas para implementação de Replicação Máquina de Estados, para análise de seus aspectos práticos.

2.1. Máquinas Virtuais

Uma máquina física real é formada por vários componentes físicos que fornecem uma interface para os sistemas operacionais e suas aplicações. Já uma máquina virtual nada mais é do que uma camada de virtualização (*software*) construída sobre esta interface e que fornece uma nova interface para os demais componentes do sistema. Uma máquina virtual também pode ser definida como uma cópia isolada e protegida de uma máquina real [Goldberg and Mager 1979].

Uma máquina virtual é constituída de três partes [Laureano and Maziero 2008]:

- O sistema hospedeiro, que engloba os recursos reais de *hardware* e *software*.
- A camada de virtualização, ou o sistema supervisor/monitor, que fornece as interfaces virtuais a partir da interface real.
- E o sistema convidado, que é executado sobre o sistema virtualizado.

Os sistemas supervisores comumente apresentam as seguintes propriedades que podem ser usadas pelas aplicações para questões de segurança e tolerância a falhas [Laureano and Maziero 2008]: equivalência – o ambiente provido pelo supervisor

é quase idêntico ao da máquina real; controle de recursos – o supervisor faz o controle completo dos recursos da máquina real; eficiência – grande parte das instruções do processador virtual devem executar diretamente no processador real, sem a interferência do supervisor; isolamento – aquilo que é executado em uma máquina virtual fica isolado do resto do sistema; inspeção – o supervisor tem acesso e controle sobre o estado interno da máquina virtual; gerenciabilidade – o supervisor gerencia os recursos utilizados pelos sistemas convidados; encapsulamento – o supervisor pode fazer *checkpoints* das aplicações/sistemas convidadas a fim de permitir recuperação, migração e inspeção em caso de falhas; e recursividade – é possível executar um supervisor dentro de uma máquina virtual, produzindo uma nova camada de virtualização.

Para este trabalho é utilizado o sistema de virtualização em Software Livre conhecido como KVM (do inglês, *Kernel-Based Virtual Machine*) [KVM-Community 2016]. O KVM utiliza virtualização completa baseada na plataforma x86, a qual deve conter hardware compatível com modo de operação para virtualização, tais como: Intel’s VT (do inglês, *Virtualization Technology*) e AMD’s SVM (do inglês, *Secure Virtual Machine*). O sistema KVM é composto por dois componentes: módulo de núcleo (do inglês, *Kernel Module*) e espaço de usuário (do inglês, *user space*) [Che et al. 2008]. Módulo de núcleo é um *drive* de dispositivo responsável por gerenciar o hardware virtualizado e mapear a memória de virtualização para que o sistemas convidados tenham acesso. Já o espaço de usuário é responsável pelas operações virtualizadas de entrada e saída. Desta forma, todos os dispositivos, tais como: disco, adaptadores gráficos e dispositivos de rede podem ser acessados diretamente em espaço de usuário.

2.2. Containers

Containers é o nome dado a uma ambiente isolado, gerenciável e com arquivos em camadas que pode ser criado sobre um sistema operacional. A criação de *containers* é muito rápida e o provisionamento de recursos é mais eficiente quando comparado com máquinas virtuais. Neste trabalho é utilizado o sistema de virtualização baseado em *container* conhecido como *Docker* [Docker 2016].

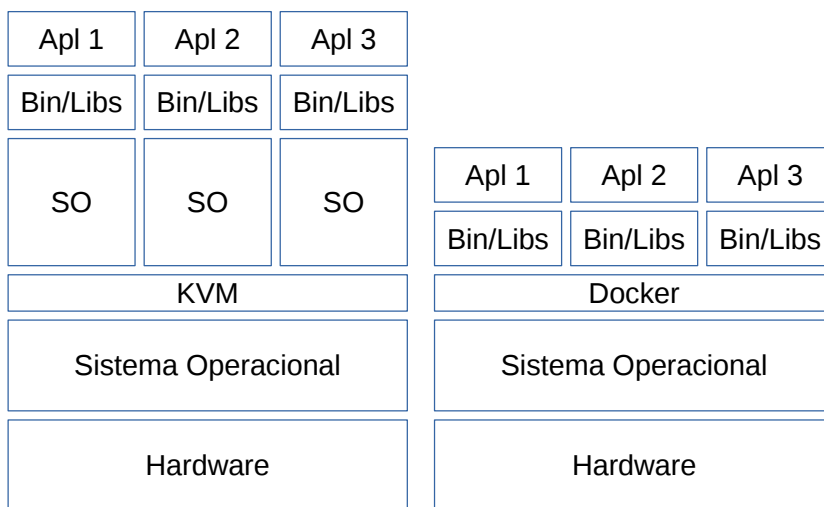


Figura 1. Diferenças entre arquitetura de Máquina Virtual e Containers [Docker 2016].

Conforme poder ser verificado na Figura 1, existem diferenças fundamentais na arquitetura de um sistema de virtualização completo e de um sistema de virtualização baseado em *containers*. Na arquitetura de virtualização completa, para que cada aplicação seja executada de forma isolada, faz-se necessário a instalação de um sistema operacional completo. Desta forma, os recursos serão desperdiçados quando comparados com abordagem de uso de *containers*. Nesta abordagem, o sistema gerenciador de *container* é o mediador no uso e acesso aos recursos computacionais, não havendo a necessidade da instalação de um sistema operacional completo para que cada aplicação possa ser executada de forma isolada [Merkel 2014].

3. Replicação Máquina de Estados

A Replicação Máquina de Estados (RME) consiste em replicar os servidores e coordenar as interações entre os clientes e estas réplicas, com o objetivo de que as mesmas apresentem a mesma evolução em seus estados. O modelo básico de programação de uma RME envolve duas primitivas:

- *invoke(operation)*: utilizada pelos clientes para invocar operações no serviço implementado pela RME.
- *execute(operation)*: implementada pelos servidores replicados, sendo utilizada sempre que uma operação deve ser executada pela RME.

A primitiva *invoke(operation)* é utilizada para enviar uma requisição *operation* para as réplicas, as quais executam a primitiva *execute(operation)* para processar *operation* e retornar a resposta a ser enviada para o cliente.

A implementação destas primitivas deve atender aos seguintes requisitos, que definem o determinismo de réplicas:

1. Estado inicial: todas as réplicas corretas devem iniciar a partir de um mesmo estado.
2. Determinismo: todas as réplicas corretas, que executam uma mesma operação sobre um mesmo estado, realizam a mesma mudança em seus estados e produzem a mesma resposta para o cliente.
3. Coordenação: todas as réplicas corretas executam a mesma sequência de operações.

Para garantir o item (1.) basta iniciar todas as réplicas com o mesmo estado (i.e., iniciar todas as variáveis que representam o estado com os mesmos valores nas diversas réplicas). Para prover o item (2.), é necessário que as operações executadas pelos servidores sejam determinísticas. Finalmente, para que as réplicas executem a mesma sequência de operações afim de garantir o item (3.), é necessária a utilização de um protocolo de difusão atômica [Hadzilacos and Toueg 1994]. Difusão atômica, ou difusão com ordem total, consiste em fazer com que todos os processos corretos, membros de um grupo, entreguem todas as mensagens difundidas neste grupo na mesma ordem.

Um resultado teórico fundamental em sistemas distribuídos é que a difusão atômica e o consenso são problemas equivalentes. O *problema do consenso* [Hadzilacos and Toueg 1994] consiste em fazer com que todos os processos corretos acabem por decidir o mesmo valor, o qual deve ter sido previamente proposto por algum dos

processos do sistema. Sendo assim, implementar difusão atômica através de um protocolo de consenso é trivial, pois basta que os processos utilizem o consenso para chegar a um acordo sobre a ordem de entrega das mensagens (operações).

Assumindo que os três requisitos acima sejam satisfeitos, uma implementação de RME deve satisfazer as seguintes propriedades de segurança e terminação:

- **Segurança:** todas as réplicas corretas executam a mesma sequência de operações.
- **Terminação:** todas as operações de clientes corretos são executadas.

A propriedade de segurança permite a implementação de consistência forte, conhecida como linearizabilidade (*linearizability*) [Herlihy and Wing 1990]. Já a propriedade de terminação garante que as operações invocadas por clientes corretos são executadas e acabarão terminando. Os protocolos propostos para RME limitam a quantidade de falhas que o sistema pode apresentar para que estas propriedades sejam atendidas, sendo necessário $n \geq 2f + 1$ (no modelo de falhas por parada - *crash*) [Lamport 1998] ou $n \geq 3f + 1$ (no modelo de falhas bizantinas) [Castro and Liskov 2002] réplicas para tolerar até f falhas. Além disso, outros requisitos necessários para a execução de uma RME incluem um sistema parcialmente [Dwork et al. 1988] síncrono para terminação [Schneider 1990, Lamport et al. 1982, Castro and Liskov 2002]. A ideia por trás destes modelos é de que o sistema trabalha de forma assíncrona (não respeitando nenhum limite de tempo) a maior parte do tempo. Porém, durante períodos de estabilidade, o tempo para transmissão de mensagens é limitado.

3.1. BFT-SMART: Implementação de Replicação Máquina de Estados

O BFT-SMART [Bessani et al. 2014] é uma biblioteca para implementação de aplicações através de replicação Máquina de Estados [Schneider 1990] que tolera falhas Bizantinas em algumas das réplicas (adicionalmente, o sistema pode ser configurado para tolerar apenas *crashes*). Esta biblioteca *open-source* de replicação foi desenvolvida em Java e implementa um protocolo para RME similar aos outros protocolos para tolerância a falhas Bizantinas (ex.: [Castro and Liskov 2002]). Além disso, são fornecidos protocolos para reconfiguração e para gerenciamento de estados (*checkpoints*, atualização e transferência de estados).

Uma descrição mais detalhada do BFT-SMART pode ser encontrada em [Bessani et al. 2014]. Resumidamente, cada réplica do BFT-SMART realiza as seguintes tarefas:

- **Recebimento de Requisições.** Os clientes enviam suas requisições para as réplicas, que as armazenam em filas diferentes para cada cliente. A autenticidade das requisições é garantida por meio de assinaturas digitais, i.e., os clientes assinam suas requisições (caso configurado). Desta forma, qualquer réplica consegue verificar a autenticidade das requisições e uma proposta para ordenação, que contém a requisição a ser ordenada, somente é aceita por uma réplica correta após a autenticidade desta requisição ser verificada.
- **Ordenamento de Requisições.** Sempre que existirem requisições para serem executadas, um protocolo de difusão atômica é executado onde uma instância de um protocolo de consenso é inicializada por uma réplica (chamada de líder) para definir uma ordem de entrega para um lote de requisições. No BFT-SMART a

ordenação é sequencial, i.e., uma nova instância do consenso só é inicializada após a instância anterior ter terminado. Porém, a ordenação de requisições em lotes visa aumentar o desempenho do sistema.

- **Execução de Requisições.** Quando a ordem de execução de um lote de requisições é definida, este lote é adicionado em uma fila para então ser entregue à aplicação. Após o processamento da cada requisição, uma resposta é enviada ao cliente que solicitou tal requisição.

3.1.1. Implementando Aplicações com o BFT-SMART

A forma de utilização do BFT-SMART, para programação de uma aplicação tolerante a falhas através de RME, é bastante simples. A API básica para clientes e servidores fornece métodos para o cliente invocar o serviço replicado e uma espécie de *callback* para servidores executarem estas requisições/operações.

Para acessar o serviço replicado, um cliente do BFT-SMART apenas deve instanciar uma classe *ServiceProxy* fornecendo seu identificador (inteiro) e um arquivo de configuração contendo o endereço (IP e porta) de cada um dos servidores. Após isso, sempre que o cliente desejar enviar alguma requisição para as réplicas (servidores), o mesmo deve invocar o método *invokeOrdered* especificando a requisição.

Por outro lado, para implementar o servidor, cada réplica deve estender a interface *Executable* e implementar o método abstrato *executeOrdered* que é invocado quando uma requisição deve ser executada. Além disso, é necessário instanciar uma *ServiceReplica* que representa propriamente a réplica, fornecendo o identificador (inteiro) da réplica que é mapeado para uma porta e endereço IP através de um arquivo de configuração.

Esta API é muito mais rica, apresentando também métodos para o gerenciamento do estado das réplicas (caso deseje-se empregar recuperação de réplicas). Uma descrição mais aprofundada sobre a API para programação de aplicações sobre o BFT-SMART pode ser encontrada em [Bessani et al. 2014].

4. Experimentos

Visando analisar o desempenho de uma Replicação Máquina de Estados sendo executada tanto em máquinas virtuais quanto em *containers*, bem como discutir aspectos práticos relacionados com estas duas abordagens, alguns experimentos foram realizados com um aplicação (Seção 4.1) desenvolvida sobre o BFT-SMART e os resultados são reportados e analisados nesta seção.

4.1. Aplicação

Uma aplicação de lista encadeada foi desenvolvida para a execução destes experimentos. Esta lista foi utilizada para armazenar inteiros e as seguintes operações foram implementadas para seu acesso:

- ADD – que adiciona um inteiro no final da lista e retorna 1 caso este inteiro ainda não esteja na lista, retorna 0 caso contrário;
- REMOVE – que remove um inteiro passado como parâmetro caso esteja na lista; retorna um código de erro caso contrário;

- GET – que retorna o elemento da posição passada como parâmetro ou um código de erro caso não exista um elemento na posição indicada (maior de que o tamanho da lista); e
- CONTAINS – que retorna verdadeiro caso o elemento passado como parâmetro está na lista, caso contrário retorna falso.

4.2. Metodologia

Para os nossos experimentos, foram criadas máquinas virtuais ou *containers* sobre uma máquina física com a seguinte configuração: Servidor DELL PowerEdge R610 com processador Intel Xeon Processor E5530 (8M Cache, 2.40 GHz, 5.86 GT/s Intel QPI), 16 núcleos e 12 GB de memória RAM. O sistema operacional base instalado é o Ubuntu 14.04 LTS. O gerenciador de máquinas virtuais utilizado foi o KVM (qemu-kvm 2.0.0+dfsg-2ubuntu1.24), enquanto que o gerenciador de *containers* foi o *Docker* (docker-engine 1.11.1-0-trusty).

O BFT-SMART foi configurado com $n = 3$ servidores para tolerar até 1 falha por parada (*crash*). Cada servidor executou em uma máquina virtual separada ou em um *container* separado, de acordo com cada experimento, enquanto que 10 clientes executaram em uma máquina separada.

A latência e o *throughput* para execução das operações da aplicação desenvolvida foram determinados, visando analisar o desempenho do sistema nos diversos cenários. A latência foi medida em um dos clientes e os valores apresentados representam a média de 25000 execuções, excluindo-se 10% dos valores com maior desvio. Já o *throughput* apresentado é o pico atingido pelos servidores, medido no servidor líder do consenso [Bessani et al. 2014] a cada 2000 requisições. Nestes experimentos, a lista foi inicializada com 100k entradas em cada réplica e os parâmetros para as operações ADD, GET, CONTAINS e REMOVE foram selecionados aleatoriamente seguindo uma distribuição uniforme. Além disso, do *workload* total em cada cliente, 25% corresponde a cada uma das operações.

Outras métricas de utilização do sistema também foram analisadas, como a utilização da CPU, da memória e da rede. Este consumo é importante principalmente no cenário de computação em nuvem, onde a tributação sobre o serviço ocorre de acordo com a utilização destes recursos.

4.3. Resultados e Análises

Esta seção apresenta os resultados e as análises referentes à comparação de execução do BFT-SMART executado em ambos ambientes de virtualização: KVM e *Docker*. Primeiramente são apresentados os resultados para o desempenho observado pela aplicação, seguindo para uma análise sobre o consumo de recursos do sistema.

4.3.1. Desempenho da Aplicação

Iniciamos nossa análise avaliando o desempenho da aplicação executada em ambas as arquiteturas. As Figuras 2 e 3 apresentam os dados para a latência e a taxa de transferência apresentados pela aplicação.

Conforme pode ser observado pela Figura 2, a aplicação ao ser executada no ambiente de virtualização completa (máquina virtual) apresentou o dobro da latência quando comparada com o ambiente de virtualização por *container*. Dentre os fatores que contribuem para a latência da aplicação, encontra-se a capacidade do hardware em processar e encaminhar as requisições a ele atribuída. Conforme apresentado na Figura 1, o ambiente de virtualização completo adiciona uma camada a mais de software na arquitetura de virtualização, o que se traduz em um *overhead* percebido nas medições feitas na aplicação. Neste sentido, os resultados apresentados pela Figura 3 estão condizentes com o esperado. A aplicação conseguiu atingir mais do que o dobro do *throughput* quando executada sobre a arquitetura *container*.

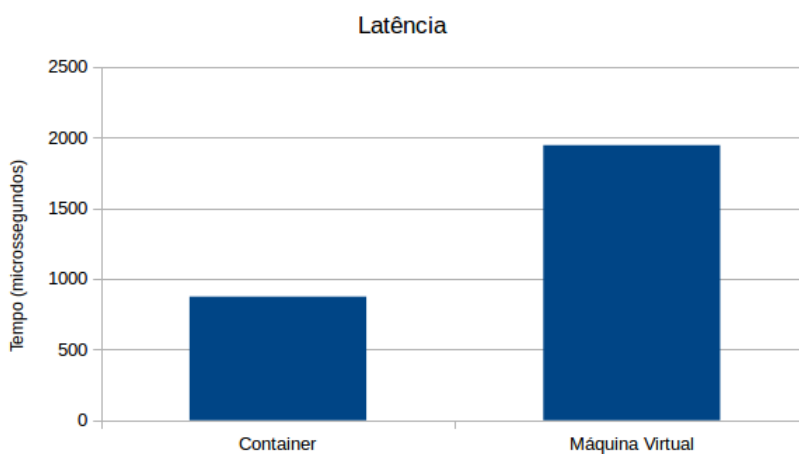


Figura 2. Latência da aplicação.

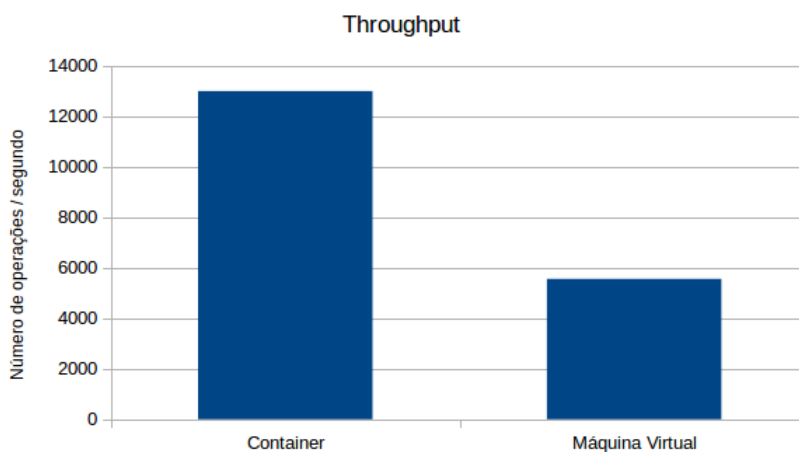


Figura 3. Throughput da aplicação.

4.3.2. Uso de Recursos do Sistema

Nesta subseção será avaliado o uso dos recursos computacionais utilizados para a execução da aplicação nas duas arquiteturas. Para realizar esta medição, no sistema operacional base foi utilizado o software de medição DSTAT (do inglês, *Versatile Resource*

Statistics Tool) [Dstat-Community 2016]. A medição foi iniciada alguns momentos antes dos ambientes virtualizados serem inicializados e a aplicação ser executada e foi interrompida logo após a finalização da execução da aplicação BFT-SMART.

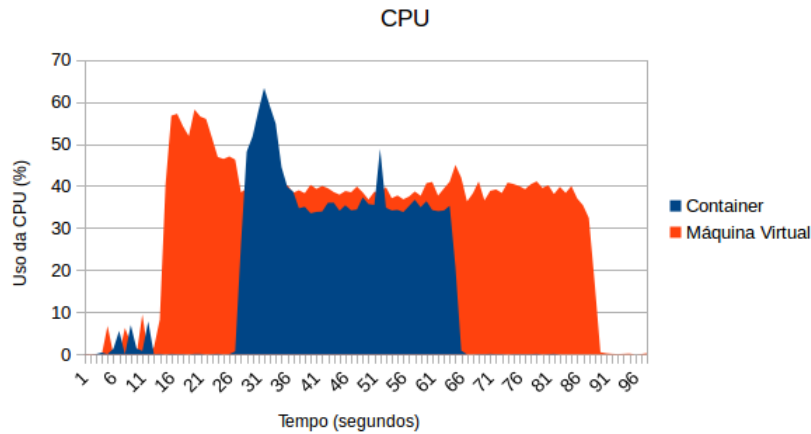


Figura 4. Uso do recurso de CPU.

Com relação ao uso dos 16 núcleos de CPU, a Figura 4 apresenta o comparativo medido. Apesar do ambiente *container* apresentar um pico de $\approx 5\%$ a mais no uso de CPU (segundo 33) do que o maior valor auferido para máquina virtual (segundo 23), analisando todos os 96 segundos de execução, claramente o ambiente com máquinas virtuais apresentou um maior uso deste recurso. A diferença no uso da CPU não foi tão acentuada, contudo, no modelo por máquina virtual o tempo de uso foi bem maior.

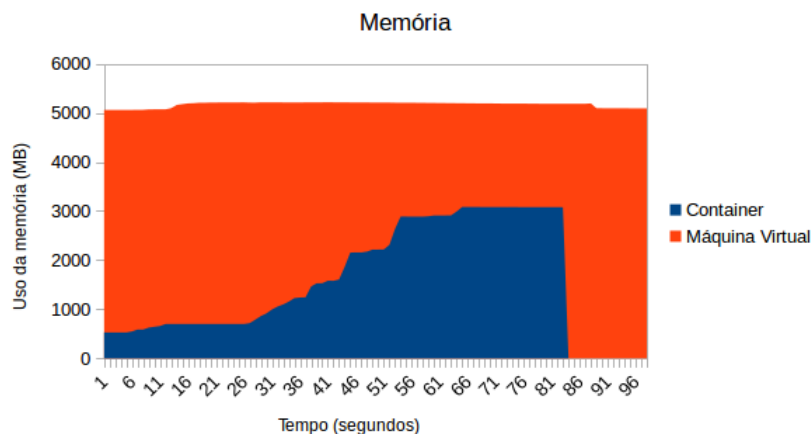


Figura 5. Uso do recurso de memória RAM.

A Figura 5 apresenta o uso de memória medido. É possível verificar que o uso de memória pelo ambiente de máquina virtual mostrou-se constante e em torno dos 5GB de memória RAM. Neste ponto é importante destacar que mesmo em momentos em que a aplicação não exigiu o uso de memória, a arquitetura de virtualização completa continua com memória alocada. Este tipo de comportamento não acontece com o ambiente de virtualização por *container*. Conforme pode ser verificado pela figura, o uso da

memória dar-se sob demanda, mantendo-se bem abaixo do modelo de virtualização completa. Neste sentido, o uso deste recurso de memória pelo modelo *container* torna-se mais eficiente.

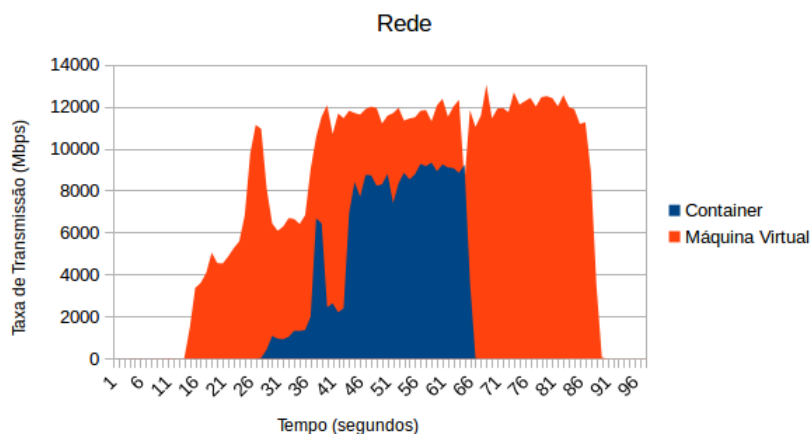


Figura 6. Uso do recurso de Rede.

Por fim, a Figura 6 apresenta o uso do recurso de rede. Conforme discutido, a medição realizada deu-se no sistema operacional base. Sobre o ponto de vista deste sistema, o acesso a interface de rede deu-se pelas três instâncias de virtualização. No caso do modelo de virtualização completo por máquina virtual, o uso da interface de rede deu-se por mais tempo e o gasto foi maior do que quando comparado com o modelo de virtualização por *container*. Contudo, é importante destacar que mesmo utilizando menos recurso de rede, a aplicação (Figura 3) conseguiu o dobro da taxa de transmissão quando executada no ambiente de virtualização por *container*.

5. Conclusões e Trabalhos Futuros

Este trabalho abordou aspectos práticos do uso de virtualização tanto através do uso de máquinas virtuais quanto através de *containers*, na implementação de replicação Máquina de Estados. Para isso, foi utilizada a biblioteca BFT-SMART para desenvolver uma aplicação tolerante a falhas através de RME e foi realizado um conjunto de experimentos usando estas duas formas de virtualização, a fim de entender as implicações práticas da adoção de cada uma delas. A implementação de máquinas virtuais deu-se através do uso do sistema de virtualização conhecido como KVM, já para o modelo utilizando *containers* optou-se pelo uso do sistema *Docker*.

Os resultados apresentados demonstram que a aplicação BFT-SMART atingiu o dobro da taxa de transferência com a metade da latência quando executada na arquitetura de virtualização por *container*. Estes valores estão condizentes com o esperado, uma vez que o ambiente de virtualização completo adiciona uma camada a mais de software na arquitetura de virtualização, o que se traduz em *overhead* para aplicação. Sob o ponto de vista do uso dos recursos computacionais, a arquitetura de *containers* mostrou-se com menor impacto sob todos os aspectos medidos. O uso de *containers* teve um menor impacto no uso de CPU, memória e rede. No quesito uso de memória, a arquitetura de virtualização completa utilizou de forma constante uma grande quantidade de

recurso sem se preocupar em promover a sua desalocação. Este comportamento gera um impacto negativo quando analisamos sob o prisma da escalabilidade da aplicação. Por fim, verificou-se que mesmo o ambiente de virtualização por *container* utilizando menos recursos de rede, a aplicação BFT-SMART conseguiu um melhor desempenho.

Como trabalhos futuros pretende-se analisar as implicações da elasticidade fornecida pela virtualização na implementação de replicação Máquinas de Estados Paralelas, que são replicações onde pelo menos algumas operações podem executar em paralelo nas réplicas. O objetivo será analisar como o provisionamento de mais recursos computacionais pode afetar dinamicamente o desempenho destas replicações, bem como o ajuste destes recursos disponíveis para a aplicação pode manter o serviço com um determinado nível de desempenho e, ao mesmo tempo, utilizar o mínimo possível de recursos computacionais.

Referências

- Bessani, A., Sousa, J., and Alchieri, E. (2014). State machine replication for the masses with BFT-SMaRt. In *Proceedings of the International Conference on Dependable Systems and Networks*.
- Castro, M. and Liskov, B. (2002). Practical Byzantine fault-tolerance and proactive recovery. *ACM Transactions on Computer Systems*, 20(4):398–461.
- Che, J., He, Q., Gao, Q., and Huang, D. (2008). Performance measuring and comparing of virtual machine monitors. In *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, pages 381–386.
- Dettoni, F., Lung, L. C., Correia, M., and Luiz, A. F. (2013). Replicação de máquina de estados tolerante a faltas bizantinas usando máquinas virtuais gêmeas. In *Anais do 31º Simpósio Brasileiro de Redes de Computadores - SBRC 2013*.
- Docker (2016). What is docker. <https://www.docker.com/what-docker>. Acessado em Maio de 2016.
- Dstat-Community (2016). Dstat: Versatile resource statistics tool. <http://dag.wiee.rs/home-made/dstat/>. Acessado em Maio de 2016.
- Dwork, C., Lynch, N. A., and Stockmeyer, L. (1988). Consensus in the presence of partial synchrony. *Journal of ACM*, 35(2):288–322.
- Garfinkel, T. and Rosenblum, M. (2003). A virtual machine introspection based architecture for intrusion detection. In *Proceedings of the Network and Distributed Systems Security Symposium*.
- Goldberg, R. P. (1973). Architecture of virtual machines. In *Proceedings of the Workshop on Virtual Computer Systems*.
- Goldberg, R. P. and Mager, P. S. (1979). Virtual machine technology: A bridge from large mainframes to networks of small computers. In *Compton Fall 79. Proceedings*.
- Hadzilacos, V. and Toueg, S. (1994). A modular approach to the specification and implementation of fault-tolerant broadcasts. Technical report, Department of Computer Science, Cornell.

- Herlihy, M. and Wing, J. M. (1990). Linearizability: A correctness condition for concurrent objects. *ACM Transactions on Programming Languages and Systems*, 12(3):463–492.
- Jiang, X. and Wang, X. (2007). "out-of-the-box" monitoring of vm-based high-interaction honeypots. In *Proceedings of the 10th International Conference on Recent Advances in Intrusion Detection*.
- KVM-Community (2016). Kernel virtual machine. http://www.linux-kvm.org/page/Main_Page. Acessado em Maio de 2016.
- Lamport, L., Shostak, R., and Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3):382–401.
- Lamport, L. (1998). The part-time parliament. *ACM Transactions Computer Systems*, 16(2):133–169.
- Laureano, M., Maziero, C., and Jamhour, E. (2004). Intrusion detection in virtual machine environments. In *Euromicro Conference, 2004. Proceedings. 30th*.
- Laureano, M. A. P. and Maziero, C. A. (2008). Virtualização: Conceitos e aplicações em segurança. In *VIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (Minicurso)*.
- Merkel, D. (2014). Docker: lightweight linux containers for consistent development and deployment. *ACM Linux Journal*, 2014(239):1–8.
- Netto, H., Lung, L. C., Correia, M., and Luiz, A. F. (2016). Replicação de máquinas de estado em containers no kubernetes: uma proposta de integração. In *Anais do XXXIV Simpósio Brasileiro de Redes de Computadores - SBRC 2016*.
- Schneider, F. B. (1990). Implementing fault-tolerant service using the state machine approach: A tutorial. *ACM Computing Surveys*, 22(4):299–319.

Um Mecanismo de Auto Elasticidade com base no Tempo de Resposta para Ambientes de Computação em Nuvem baseados em Containers

Marcelo Cerqueira de Abranches^{1,2}, Priscila Solis¹

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Prédio CIC/EST, Caixa Postal 4466 – 70.910-900 – Brasília – DF – Brazil

²Diretoria de Sistemas e Informação – Controladoria-Geral da União (CGU)
SAUS, Quadra 1, Bloco A – 70.050-904 – Brasília – DF – Brazil

marcelo.abranches@cgu.gov.br, pris@cic.unb.br

Abstract. *This paper proposes a cloud computing architecture based on containers and an algorithm for auto elasticity based on the response time requirements in a Web system. The proposed algorithm promotes an efficient allocation of containers to achieve expected response time in processing requests. This proposal was evaluated with a real time series obtained from a web system hosted by Controladoria-Geral da União (CGU). The results show that the proposed algorithm achieves the expected response times allocating a lower number of containers than other related proposals.*

Resumo. *Este artigo propõe uma arquitetura de computação em nuvem baseada em containers e um algoritmo de auto elasticidade com base no tempo de resposta de um sistema Web. O algoritmo proposto promove a otimização do processamento das requisições por meio da alocação eficiente do número de containers para alcançar tempos esperados de resposta no processamento de requisições. A proposta foi avaliada com uma carga de trabalho caracterizada com base em uma série temporal real resultante das requisições de um sistema Web da Controladoria Geral da União (CGU). Os resultados mostram que o algoritmo proposto consegue cumprir os requerimentos de desempenho alocando um número menor de containers que outras propostas relacionadas.*

1. Introdução

A pesquisa nos ambientes de computação em nuvem apresenta como um dos maiores pontos de interesse algoritmos de auto escalabilidade [Poddar et al. 2015] e balanceamento de carga. Como em qualquer processo de dimensionamento e planejamento, um dos pontos cruciais para esta tarefa é a correta caracterização da carga de trabalho e a otimização no uso de recursos, os quais devem atender um conjunto de requerimentos de desempenho das aplicações.

Diversos trabalhos recentes abordam o dimensionamento elástico e a auto escalabilidade em ambientes de computação em nuvem [Poddar et al. 2015], [Lorido-Bostrán et al. 2012], [Gong et al. 2010] e [Google 2016b]. As soluções propostas nestes trabalhos são na sua maioria baseadas em heurísticas, teoria de filas, análises de séries temporais, reação a aumento de consumo de recursos entre outros.

Este trabalho propõe um algoritmo para promover a auto elasticidade em um ambiente de nuvem baseado na alocação eficiente do número de *containers* para atender requisições de um ambiente *web*. A proposta foi avaliada em um ambiente simulado e os resultados mostram potencial para otimizar o processo de alocação de recursos de processamento em um ambiente de computação em nuvem.

O trabalho é estruturado da seguinte forma: a seção 2 apresenta os trabalhos relacionados. A seção 3 contém a revisão bibliográfica. A seção 4 descreve as principais características das ferramentas utilizadas, a arquitetura da solução e detalha o algoritmo de auto elasticidade desenvolvido. A seção 5 apresenta os resultados experimentais. Finalmente a seção 6 apresenta as conclusões e trabalhos futuros desta pesquisa.

2. Trabalhos Relacionados

O trabalho de [Lorido-Bostrán et al. 2012], realiza uma comparação entre os diversos métodos para obter escalabilidade e auto elasticidade em um ambiente de nuvem. Estes métodos são separados em duas categorias: reativos e preditivos. As técnicas utilizadas são baseadas em aprendizado de máquina, teoria de filas, teoria de controle, análises de séries temporais, entre outros. A proposta deste trabalho pode ser descrita como reativa, pois reage a variações no tempo médio de resposta de uma aplicação, alocando ou desalocando recursos, utilizando teoria de controle para isto.

Outros trabalhos também abordam o dimensionamento elástico (auto-escalável) para ambientes de nuvem. No trabalho de [Gong et al. 2010], os autores propõem um algoritmo chamado de PRESS para previsão de carga de CPU, que extrai padrões de consumo e ajusta a alocação de recursos. A abordagem dos autores utiliza dois métodos para realização de previsões em linha. O primeiro se baseia no uso de processamento de sinais e transformadas rápidas de Fourier para extração de frequências dominantes. Com essas frequências são geradas séries temporais e diversas janelas de tempo são comparadas. É gerado um índice de correlação de Pearson para as várias janelas comparadas. Caso se obtenha um índice de correlação maior que 0,85, o valor médio do uso de recursos dentro de cada posição da série temporal é utilizado para gerar uma previsão para a próxima janela e os recursos das máquinas virtuais são ajustados. Caso um padrão não seja identificado, os autores propõem uma abordagem que utiliza uma cadeia de Markov com número finito de estados para a realização das previsões.

Outro trabalho que propõe dimensionamento elástico (auto-escalável) para ambientes de nuvem é o Haven [Poddar et al. 2015]. Esta proposta utiliza ferramentas de um sistema operacional de nuvem para monitoramento de cargas de CPU e memória a que cada máquina virtual de um *pool* de balanceamento está submetido. A partir de limiares previamente estabelecidos para consumo de CPU e memória, o Haven pode instanciar novas máquinas virtuais e realizar a sua inserção em um *pool* de balanceamento de carga.

O [Google 2016b] propõe e disponibiliza uma ferramenta nativa de auto escalabilidade chamada de *Horizontal Pod Autoscaler* (HPA). Esta ferramenta trabalha escalando o ambiente a partir de limiares médios de consumo de CPU dos *containers* que atendem a determinada aplicação.

A proposta deste artigo se diferencia da abordagem PRESS na medida que não realiza previsão de carga, mas sim um dimensionamento de recursos, baseado na observação do tempo de resposta de uma aplicação atrás de um balanceador de carga. Outra

diferença é que o PRESS realiza escalabilidade vertical, ou seja, realiza aumento de recursos de memória e cpu, para se ajustar a carga de trabalho. Este trabalho propõe a escalabilidade horizontal, ou seja, novas instâncias capazes de atender a carga de trabalho, são alocadas atrás de um balanceador de carga, de modo a se ajustar a variações na demanda. A escalabilidade horizontal tem a vantagem de que os recursos alocados para atender a carga de trabalho, não são limitados aos recursos físicos de uma máquina. Além disso, esta abordagem facilita a alta disponibilidade, uma vez que as demandas são atendidas por um conjunto de instâncias em paralelo.

A diferença em relação à abordagem Haven é propor um método de dimensionamento do sistema a partir da observação do tempo de resposta das requisições. Isso permite uma visibilidade global do comportamento do sistema, pois nos casos em que não haja excessos de consumo de processamento e memória, o ambiente pode se beneficiar do aumento do paralelismo no atendimento das requisições. Assim como a proposta deste artigo, o Haven também realiza escalabilidade horizontal. Outra diferença na proposta deste artigo é a infraestrutura utilizada. Os trabalhos PRESS e HAVEN trabalham com a tecnologia de máquinas virtuais e a proposta deste artigo utiliza *containers* que são considerados nas publicações recentes [Docker 2016] como mais leves por não fazerem virtualização de hardware.

Este trabalho usa o *Kubernetes* como orquestrador de *containers*, porém propõe outro algoritmo de auto escalabilidade, baseado no tempo médio de resposta das aplicações. Outra diferença deste trabalho, é onde são feitas as medidas que decidem se o ambiente deve ser escalado. Enquanto o HPA realiza medidas de consumo de CPU nos próprios *containers*, a proposta deste trabalho realiza medidas externas aos *containers*, no caso são realizadas medidas do tempo médio de resposta das aplicações do ponto de vista do balanceador de carga. Esta abordagem traz a vantagem de poder observar a performance do ambiente, independente do nível de utilização dos recursos dos containers. Na Seção 5 é feita uma comparação entre o HPA e a solução proposta neste trabalho.

3. Revisão Bibliográfica

3.1. Containers

Container é uma tecnologia para a criação de instâncias de processamento separadas ou isoladas que permitem a virtualização no nível do sistema operacional ao disponibilizar porções protegidas de processamento. Dois *containers* rodando no mesmo sistema, não sabem que estão compartilhando recursos, pois tem sua própria abstração de camada de rede, memória e processos [Docker 2016].

Os *containers* apresentam uma maior portabilidade que as máquinas virtuais, ao serem configurados de forma genérica para qualquer sistema operacional baseado em Linux. A virtualização via *hypervisors* consome mais recursos do que a virtualização por *containers* dado que os últimos são executados em sistemas operacionais que rodam em espaços isolados entre si. Se um *container* não está executando nenhuma tarefa, ele não está consumindo recursos no servidor [Docker 2016]. Além disto os *containers* apresentam um grande dinamismo para serem criados e destruídos, dado que apenas tem que iniciar ou destruir processos em seu espaço isolado. Portanto verifica-se que existem vantagens em se utilizar a tecnologia de *containers*. Esta foi a tecnologia escolhida para atender o processamento das cargas de trabalho *http* nesta solução.

3.2. Serviços Web e Balanceadores de Carga

Os sistemas web podem sofrer com a alta variabilidade da carga demandada, muitas vezes de forma inesperada. Neste caso, a infraestrutura que hospeda este tipo de serviço, deve estar preparada para o atendimento da demanda com alto desempenho, escalabilidade e com alta disponibilidade. Uma abordagem comum para disponibilizar estas características é o uso de arquiteturas distribuídas que podem rotear as requisições para diversos servidores, de forma transparente ao usuário. Esta abordagem além de melhorar a performance e a disponibilidade dos serviços, pode aumentar sua escalabilidade ao se adicionar e remover membros no *cluster*. [F5 2012]

Porém vários desafios se impõem para que um *cluster* distribuído de servidores possam funcionar de forma eficiente como se fosse um único servidor. Esses desafios vão desde o roteamento das requisições para os membros do *cluster*, métodos para escolha do membro que receberá a demanda e manutenção do estado da conexão. [Gilly et al. 2011]

No balanceamento na camada de transporte, as requisições são distribuídas entre os membros do *cluster*. O balanceador distribui as conexões de clientes que conhecem o endereço IP do *cluster*, entre os diversos servidores que efetivamente respondem as requisições. Neste caso, como a distribuição é feita com base na camada 4, o balanceador escolhe um servidor, sem considerar o conteúdo ou o tipo da requisição. [Anicas 2014]

3.3. Controladores PID

Os controladores PID (Proporcional-Integral-Derivativo) são algoritmos de controle muito utilizados na indústria. Os controladores PID possuem três coeficientes: proporcional, integral e derivativo. Esses coeficientes são variados de forma a se obter a resposta de controle ideal desejada para um processo.

Um controlador PID trabalha dentro de um sistema em malha fechada, onde é possível a leitura do estado atual de determinada variável que se deseja controlar, e de acordo com seu valor, uma ação é executada de modo que a variável chegue, e tente permanecer no nível desejado (ainda considerando distúrbios externos), nas próximas iterações de tempo [Instruments 2015].

Sendo assim, o PID deve ler o estado atual da variável e calcular a resposta da saída do atuador, por meio do cálculo dos componentes proporcional, integral e derivativo e então somar os três para calcular a saída. O componente proporcional depende da diferença entre o valor desejado (*setpoint*) e o valor atual da variável. Esta diferença é referida como erro. O componente integral soma o termo de erro ao longo do tempo. A resposta derivativa é proporcional a taxa de variação da variável do processo.

Para que o controlador PID produza os ajustes necessários ao sistema, os parâmetros de ganho K_p , K_i e K_d devem ser ajustados. Existem diversos métodos de ajuste destes parâmetros, como por exemplo O método manual (*guess and check*) e o método Ziegler-Nichols [Instruments 2015].

No método manual, os ganhos de cada um dos componentes é ajustado usando tentativa e erro. Para isso, ao ajustar os parâmetros, devem ser conhecidos os efeitos que cada parâmetro provoca na saída do controlador. Neste método, os termos K_i e K_d são ajustados para zero, e o termo K_p é aumentado até que a saída do ciclo comece a oscilar. A partir daí aumenta-se lentamente o termo K_i para reduzir o erro estacionário. Neste

ponto inicia-se o incremento do termo K_d , de modo a diminuir as oscilações na saída do ciclo [Instruments 2015]. A discussão sobre outros métodos de ajuste dos parâmetros foge ao escopo deste trabalho, já que foi utilizado o método de ajuste manual.

4. Solução Proposta

O objetivo deste trabalho, é a provisão de um ambiente elástico de nuvem privada, auto escalável e com balanceamento de carga, para hospedagem de sistemas Web.

A solução propõe um ambiente de *cluster* baseado em *containers*, e um método de auto elasticidade que reage a aumentos no tempo de resposta do sistema, de modo a mantê-lo dentro de um limite. Para isso é utilizado um sistema de malha fechada com um controlador PID, que reage a variações no tempo de resposta do sistema aumentando ou diminuindo o número de *containers* no *cluster* de balanceamento de carga capaz de responder as requisições.

4.1. Ferramentas Utilizadas para Implementação

4.2. Docker

O *Docker* começou como projeto da empresa de PaaS (*Platform as a Service*) dotCloud em 2013 [Martin 2015], propondo ser um integrador e facilitador para adoção da tecnologia de *containers* em produção e em larga escala. O *Docker* utiliza o LXC (linux *containers*) para isolar os *containers* do servidor, criando isolamento de processos, rede e privilégios. A limitação e contabilização de recursos (CPU, memória, espaço em disco e E/S) é feita por meio dos *cgroups*. Além disso a utilização do sistema de arquivos é feita de forma eficiente pois se baseia em *copy-on-write*, que permite que as alterações em um *container* sejam simplesmente uma atualização diferencial da imagem anterior.

Uma das grandes vantagens do *Docker* é a habilidade de encontrar, baixar e iniciar imagens de *containers* que foram criados por outros programadores de forma muito rápida e prática. O *Docker* viabiliza o uso da tecnologia de *containers* de forma prática, razão pela qual é utilizado nesta solução.

4.3. Kubernetes

Kubernetes é um sistema desenvolvido pelo Google [Google 2016a], e disponibilizado para a comunidade, que visa gerenciar o ciclo de vida de *containers* nos nós de um *cluster*.

Desta forma, o *Kubernetes* é um orquestrador de *containers*, sendo capaz de agendar o lançamento de *containers* entre os nós do *cluster*, assim como fazer o controle de admissão dos *containers*, balanceamento de recursos e prover escalabilidade ao ambiente. O *Kubernetes* também provê funcionalidades como descoberta de serviços entre os *containers*, publicação do serviço para acessos a partir de entidades fora do *cluster* e balanceamento de carga entre os *containers* [Google 2016a]. Estas funcionalidades foram determinantes para a decisão de usar o *Kubernetes* na solução.

A infraestrutura de um *Kubernetes* é composta por nós do tipo *Master* que controlam os nós do tipo *Workers*, e estes rodam os *containers*. Todas as configurações do *cluster* ficam armazenadas de em um repositório de configurações distribuído, o *EtcD*. Os *PODs* são a unidade básica com a qual o *Kubernetes* trabalha. Os *containers* são agrupados em *PODs* e estes geralmente representam uma aplicação. Estes são criados por meio

dos Replication Controllers que são utilizados para definir PODs que podem ser escalados horizontalmente. Os Replication Controllers também são responsáveis por manter o número desejado de PODs ativos no *cluster*.

4.4. Apache Spark, Flume, HAproxy e Redis

O *Apache Spark* é uma ferramenta de processamento distribuído, ideal para processamento de grandes bases de dados. Foi desenvolvido pela AMPLab (UC Berkeley), e realiza processamento de dados em memória. Sua estrutura básica de abstração são os *RDDs* (*Resilient Distributed DataSets*), que são coleções de elementos que podem sofrer operações em paralelo, sendo possível a geração de novos *RDDs* a partir de transformações como *map*, *reduce*, *filter* e *join* em *RDDs*. [Spark 2015a] Esta ferramenta foi escolhida para integrar a solução pois com ela é possível realizar processamento de grandes bases em formato de texto, de forma escalável.

O *Apache Spark* oferece uma API chamada *Spark Streaming*, que permite o processamento em tempo real de dados, por meio da criação de estruturas chamadas *DStream* (*discretized stream*), que são representados como sequências de *RDDs*. A criação dos *DStreams* é feita por meio da classe *StreamingContext*, onde se configura a duração de cada janela de *DStreams* [Spark 2015b]. Neste trabalho, a duração de cada janela foi configurada como 5 segundos, o que é suficiente para a geração da série temporal em tempo real utilizada pela solução.

Neste contexto o *Spark Streaming* é utilizado para processar os logs do balanceador de carga, coletando o tempo de resposta de cada uma das requisições que o sistema atende, em tempo real. Esta informação do tempo de resposta é armazenada em formato de série temporal no servidor *Redis*, para uso no algoritmo de auto escalabilidade proposto neste artigo.

O recebimento das entradas de log em texto do balanceador de carga pelo *Spark*, ocorre por meio da ferramenta *Flume*. O *Flume* é um serviço que agrega, coleta e move grandes volumes de fluxo de dados. Para seu funcionamento é criada uma fonte que recebe os dados de interesse. Essa fonte (*Source*) é conectada a um canal (*Channel*), por onde os dados trafegarão em direção a um *Sink* [Flume 2016].

Portanto, a função desta ferramenta na solução, é realizar o envio, em tempo real, dos logs de acesso do balanceador carga, por meio da criação de uma fonte do tipo *Syslog*, que recebe os logs do balanceador de carga, e um canal de memória que carrega os dados da fonte em memória. A partir daí, o *Spark* consome este fluxo de dados por meio de um *Sink* do tipo *Avro*, que trafega pela rede.

O balanceador de carga utilizado na solução, é o *HAproxy*, que pode atuar como balanceador de camada 4 ou 7, terminador SSL, proxy reverso entre outros [Tarreau 2015]. Atualmente, este é o balanceador utilizado por *sites* como *Reddit*, *Stack Overflow/Server Fault*, *Instagram* entre outros, além de ter sido escolhido como balanceador da nuvem da Red Hat (*OpenShift*).

Esta adoção pela comunidade foi uma das razões de se ter escolhido esta ferramenta para compor a solução. Além disso, o *HAproxy* disponibiliza um *log* com informações viabilizam a implementação da proposta deste trabalho.

O log do *HAproxy* apresenta o seguinte formato:

```
May 18 06:24:25 10.125.7.229 haproxy[1078]: 10.125.8.252:43839
[18/May/2016:06:24:24.988] cherrypy cherrypy/10.125.7.227 0/0/2/26/28 200 169
-- -- 1/1/1/0/0 0/0 "GET /generate HTTP/1.0"
```

Nesta linha pode-se observar informações como o tempo de espera na fila do servidor de aplicação, método e código da resposta *http* entre outras. A parte com *0/0/2/26/28* contém as informações : *Tq ' / Tw ' / Tc ' / Tr ' / Tt*, onde O *Tr* é o tempo em milisegundos que o balanceador demora para receber uma resposta completa de uma requisição *http* ao servidor [Tarreau 2015]. Portanto este valor representa o tempo total de processamento da requisição pelo *container*.

No contexto deste trabalho o *SparkStreaming*, processa as entradas do *log*, separa o campo *Tr* e o armazena no banco *Redis*, no formato de uma série temporal. O *SparkStreaming* também é responsável por converter o formato da data de cada linha, para o número de segundos desde 0 hora de cada dia, para suportar a criação da série temporal.

O *Redis* é utilizado para armazenar a série temporal, pois consegue prover baixa latência tanto para escrita como para leitura, por manter os dados como estruturas em memória [Redis 2015].

A integração da solução com o *Redis*, ocorre por meio da biblioteca *Kairos*, que cria uma estrutura para armazenamento de séries temporais em bancos como *Redis*, *Mongo*, *SQL* ou *Cassandra* [Kairos 2015].

Esta biblioteca provê facilidades, como configuração do número de entradas a serem mantidas no banco e tamanho da unidade mínima de tempo de interesse da série. No caso deste trabalho são mantidos armazenados na série a cada momento os dados dos últimos 600 segundos, o que é suficiente para operação do algoritmo.

O *Kairos* também permite o cálculo de parâmetros estatísticos da série, utilizando tamanho de janelas configuráveis de tempo. Isso permite que seja calculada a média de tempo de resposta de uma aplicação nos últimos dois minutos por exemplo. A unidade mínima de tempo configurada nesta solução é 1 segundo. Isso permite boa flexibilidade para configuração dos tamanhos das janelas de tempo para os cálculos estatísticos, além de permitir a geração de gráficos com boa resolução de tempo para monitoramento e avaliação da solução.

4.5. Arquitetura da Solução

A arquitetura proposta é apresentada na Figura 1, em que é utilizado um controlador PID para manter o tempo médio de resposta de determinada aplicação próxima de um determinado limiar. A arquitetura proposta, chamada de PAS (PID based Autoscaler) opera com base na seguinte sequência:

1. É estabelecido um limiar (*setpoint*) de tempo médio de resposta desejado para as requisições. O monitor recebe o tempo de resposta das requisições que chegam no balanceador;
2. O monitor envia o tempo médio de resposta (ex: média dos últimos 200 segundos) para que o dimensionador PID calcule o número de *containers* necessários para atingir ou permanecer no *setpoint*. A média dos últimos 200 segundos foi definida, pois verificou-se experimentalmente que é uma quantidade suficiente para

que valores muito diferentes da média não exerçam influência negativa no dimensionamento, assim como permite que o sistema seja capaz de reagir rapidamente a mudanças no padrão da média do tempo de resposta.

3. O Dimensionador PID executa o algoritmo 1 e informa o número desejado de *containers* ao *Kubernetes*. O número atual de *containers* é pesquisado usando a ferramenta do *kubernetes* *kubectl*, e o tempo médio de resposta do *cluster* é determinado usando a série temporal presente no *Redis*.
4. O *Kubernetes* cria ou remove novos *containers*, além de garantir que o ambiente permanecerá com o número desejado de *containers* até a próxima rodada do algoritmo (ex: depois de 10 segundos). Este valor de 10 segundos foi definido pois verificou-se que é suficiente para que o *Kubernetes* inicie novos *containers* e estes passem a responder as requisições no *cluster* de balanceamento.

Algoritmo 1: PAS

Entrada: Tempo médio de resposta do cluster, Número atual de *containers*

Saída: Número desejado de Containers

1 **início**

2 Leia o limiar de tempo médio de resposta desejado para as requisições:
 $t_{ms_desejado}$

3 Leia o número atual de *containers*: $n_containers_atual$

4 Leia tempo de resposta médio do cluster em ms: t_{ms_atual}

5 Calcule o erro: $e(t) = t_{ms_desejado} - t_{ms_atual}$

6 Calcule a saída o controlador pid:

$$u(t) = K_p e(t) + K_i \int_0^t e(t) \delta t + K_d \frac{d}{dt} e(t) \quad (1)$$

$n_desejado_containers = n_containers_atual + u(t)$

7 **fim**

8 **retorna** $n_desejado_containers$

4.6. Operação da solução

Para viabilizar a operação do algoritmo que trabalha com dados dinâmicos, recebidos em tempo real, utiliza-se o fluxo de processamento mostrado na Figura 2. O *Haproxy* atua como balanceador de carga da solução. Seus *logs* são enviados ao *Flume* que envia os dados em um canal de memória e os envia ao *Spark Streaming*, desta forma disponibilizando as informações necessárias para operação do algoritmo de auto escalabilidade.

O *Haproxy* atua balanceando as requisições no modo *round robin* para os servidores *Docker/Kubernetes* que hospedam os *containers*. Quando o nó *Docker/Kubernetes* recebe a requisição, o serviço *Kubernetes proxy* é responsável por balancear a carga entre os *containers* de cada um dos servidores.

Desta forma, acontecem dois níveis de balanceamento, um entre os nós *Docker/Kubernetes* em que o responsável pelo balanceamento é o *Haproxy*, e outro internamente dentro do nó *Docker/Kubernetes*, onde o serviço *Kubernetes proxy* é o responsável pelo balanceamento.

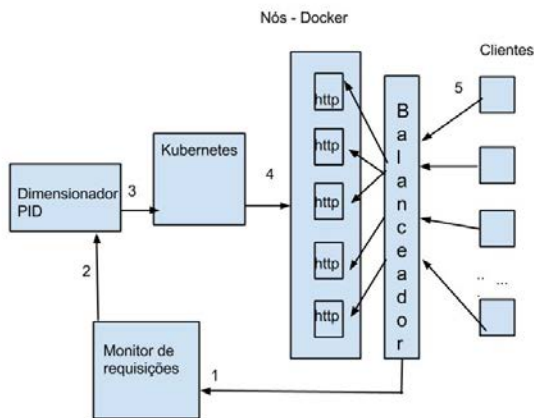


Figura 1. Arquitetura da Solução



Figura 2. Fluxo de processamento entre o conjunto de ferramentas utilizadas

5. Resultados Experimentais

5.1. Ambiente

Para avaliação da solução foi configurado um *cluster Kubernetes* v1.1.2 no sistema operacional *CoreOS* (899.6.0 (2016-02-02)), virtualizado em *VMWare ESXi* 5.5.0. Este ambiente foi configurado para validação da solução. Um ambiente de produção poderia se beneficiar mais se o sistema *CoreOS* fosse instalado diretamente em máquinas físicas, pois seria eliminada a camada de virtualização.

O *cluster* foi constituído com os seguintes componentes: 1 nó *master* (4 vCPUs, 6 GB de RAM), 1 nó *etcd* (4 vCPUs, 6 GB de RAM) e 4 nós *workers* (4 vCPUs, 6 GB de RAM). As máquinas virtuais (*VMWare ESXi* 5.5.0) foram instaladas no sistema operacional *Ubuntu* 14.04.3 LTS, com as seguintes configurações e ferramentas: 1 nó *haproxy* 1.5.4 (4 vCPUs, 4G GB de RAM), 1 nó *Spark* 1.5.2 + *Redis* 2.8.4 (2 vCPU, 10 GB de RAM) e 1 nó *Flume* 1.7.0 + dimensionador (2 vCPU, 4 GB de RAM).

Para validação do ambiente foi gerada uma imagem de *container* que roda o servidor web *cherryppy* 5.1.0. Foi configurado um *link* neste servidor. O *link* gera um *array* de tamanho aleatório entre 1.000 e 10.000 elementos a cada requisição.

A publicação do serviço no *Kubernetes* foi feita por meio da criação de um *Replication Controller* e a configuração de um serviço do tipo *NodePort*. O balanceador *Haproxy* foi configurado para balancear as requisições entre os nós *Docker/Kubernetes* usando os endereços IP dos nós e as portas publicadas pelo serviço do tipo *NodePort*. Cada *container* teve seus recursos de processamento e memória limitados a 18 MB de

memória RAM e 24 ms de CPU.

O algoritmo PID utilizou os seguintes parâmetros: $Kp=0.016$, $Ki=0.000012$ e $Kd=0.096$, que foram definidos após vários testes com a carga de trabalho, e aplicação do método de ajuste manual, ou *guess and check*, conforme descrito na seção 3.3

5.2. Carga de Trabalho

A geração de carga de trabalho foi feita usando a ferramenta *ab* (*apache bench*). Para geração de uma carga com um perfil de requisições por segundo realista, foi coletado um conjunto de acessos do portal da transparência (www.transparencia.gov.br), entre os dias 25/05/2015 a 25/06/2015. A série temporal capturada na escala de 1 segundo, representa o número de acessos naquele intervalo de tempo. A série foi caracterizada com o método Kettani-Gubner [Kettani et al. 2002]. A autossimilaridade e longa dependência da série foi confirmada com o parâmetro de Hurst $H=0,87$. nas escalas de 1 segundo, 100 segundos e 600 segundos.

Esta série é utilizada para gerar a cada segundo requisições simultâneas direcionadas ao endereço IP do balanceador de carga que distribui as requisições nos nós do *cluster Kubernetes*. A intensidade da carga foi ajustada em 3 níveis, multiplicando a série temporal por 1, 1.5 e 2 e preservando o mesmo índice de autossimilaridade. Estas cargas são referidas respectivamente nos experimentos como *carga_1*, *carga_1.5* e *carga_2*.

Uma amostra da série obtida utilizada para caracterizar a carga de trabalho, relativa a 2000 segundos pode ser vista na Figura 3.

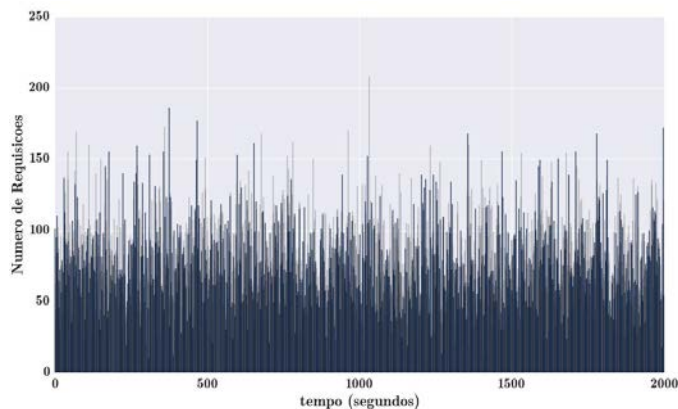


Figura 3. Amostra da carga utilizada nos testes, $H=0,87$

5.3. Cenários de avaliação

5.3.1. Cenário 1

No primeiro cenário o limiar de tempo de resposta no balanceador (*setpoint*) foi estabelecido em 50 ms e foram aplicadas as cargas *carga_1* e *carga_1.5*. A Figura 4 mostra o tempo de resposta do sistema ao ser submetido à *carga_1*. O gráfico mostra o ajuste provocado pelo PID e a estabilidade alcançada próximo do *setpoint* de 50 ms.

A Figura 4 mostra também a alocação de *containers*. Observa-se que o número de *containers* no início do experimento é igual a 2. O sistema alocou o número necessário

para que o tempo médio de resposta mostrado na Figura 4 atingisse o *setpoint*. No final da execução estavam alocados 26 *containers*.

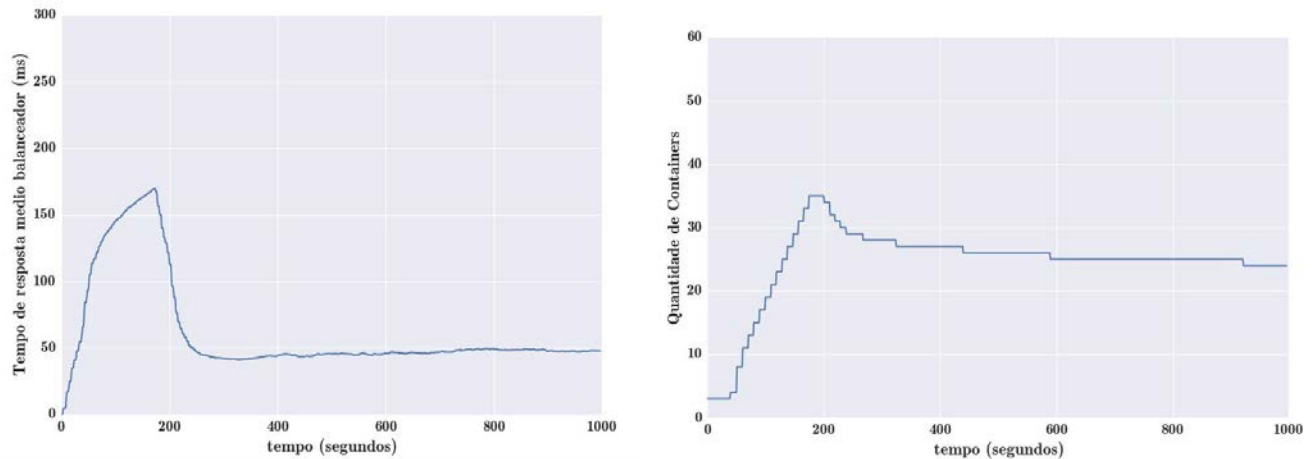


Figura 4. Carga_1, setpoint 50 ms

A Figura 5 mostra o tempo médio de resposta com o sistema sendo submetido a carga_1.5 e o *setpoint* mantido em 50 ms. A Figura 5 também mostra a alocação de *containers* durante este teste. Observa-se que a alocação de *containers* foi sendo ajustada de modo a manter o tempo de resposta médio do sistema próximo ao *setpoint*. No final da execução estavam alocados 52 *containers*. Neste caso, o sistema submetido a uma carga maior, alocou mais *containers* para manter o tempo de resposta dentro do limiar definido.

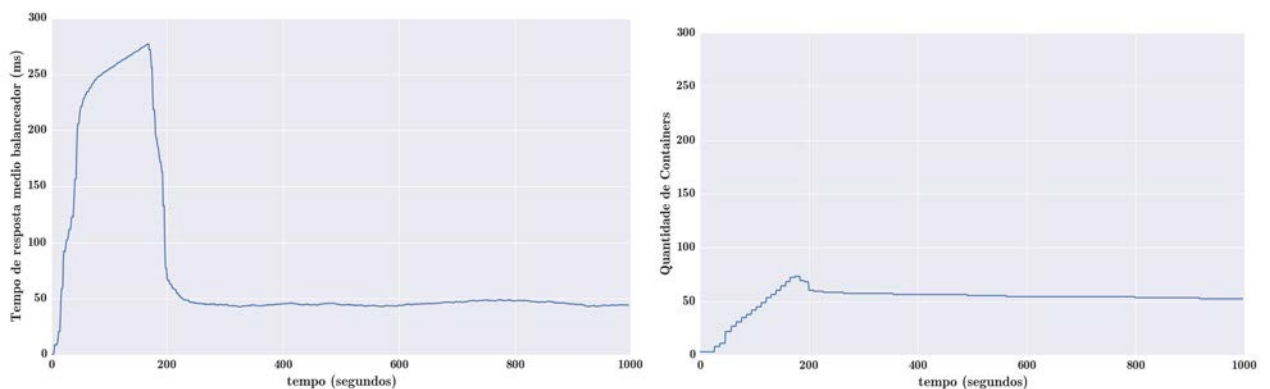


Figura 5. Carga_1.5, setpoint 50 ms

5.3.2. Cenário 2

O Cenário 2, compara o comportamento do algoritmo de escalabilidade proposto neste trabalho com o utilizado nativamente pelo *Kubernetes*, denominado de HPA [Google 2016b].

A comparação é feita da seguinte forma: o sistema configurado com o HPA_80 (configurado para escalar quando o consumo médio dos *containers* estiver acima de 80 %) será exposto às cargas carga_1, carga_1.5 e carga_2 durante 1000 segundos. No fim

da avaliação o tempo médio de espera das requisições na camada de aplicação (ponto de vista dos clientes) é observado.

A partir destes dados, foi definido um *setpoint* para ser utilizado com o algoritmo proposto (PAS) que permita comparar com o tempo de resposta próximo ao de referência do HPA. Os resultados serão comparados, verificando a quantidade média de *containers* utilizados durante os experimentos e os tempos de resposta obtidos na camada de aplicação dos clientes.

Como pode ser observado na Figura 6 os tempos médios de resposta na camada de aplicação com o algoritmo HPA para cada carga, foram: 66.18 ms para a carga_1, 110.12 ms para a carga_1.5 e 144.01 ms para a carga_2.

Para efeitos comparativos, os *setpoints* do PAS, foram configurados para entregar um tempo de resposta próximo aos obtidos com o HPA. Para isto foram configurados *setpoints* um pouco abaixo dos observados no HPA, dado que o *setpoint* é controlado no balanceador, portanto o tempo observado na camada de aplicação dos clientes deve ser maior. Os valores definidos para os *setpoints* são os seguintes: 50 ms (PAS_50) para a carga_1, 80 ms (PAS_80) para a carga_1.5 e 100 ms (PAS_100) para a carga_2.

As Figuras 6 e 7 mostram que para um tempo de resposta aproximado ao do HPA, o sistema PAS, alocou menos *containers* do que o HPA. O comparativo de alocação de *containers* mostra que: PAS_50 alocou 44.02 % do que foi alocado pelo HPA_80, PAS_80 alocou 36.07 % do que foi alocado pelo HPA_80, para a carga_1.5 e PAS_100 alocou 12.72 % do que foi alocado pelo HPA_80, para a carga_2. As médias da figura 7 foram calculadas realizando um somatório do número de *containers* a cada segundo e dividindo pelo total de tempo do teste (em segundos).

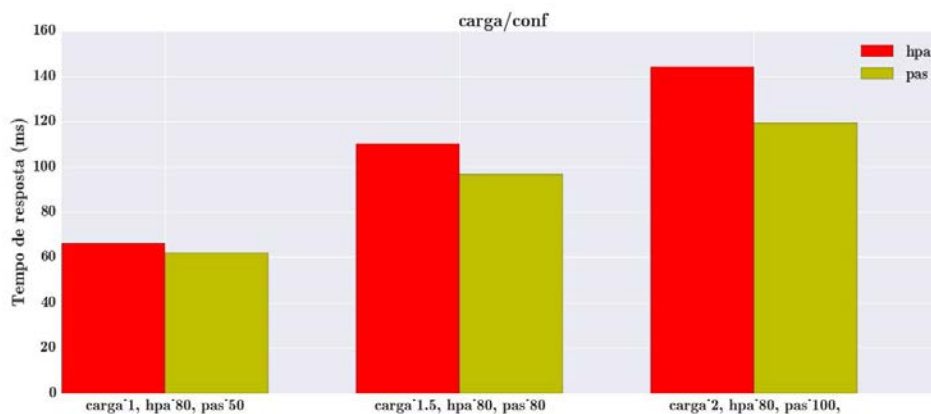


Figura 6. Comparativo entre HPA e PAS (tempo médio de resposta do sistema)

5.4. Análise dos Resultados

Os resultados experimentais mostram que nos cenários avaliados, o algoritmo PAS otimiza a alocação do número de *containers* para manter os valores de *setpoints* dentro de um limiar estabelecido. O HPA aloca um número maior de *containers* para alcançar limiares equivalentes de tempo de resposta para as requisições na camada de aplicação. Este resultado evidencia que a alocação de um número maior de *containers* pode aumen-

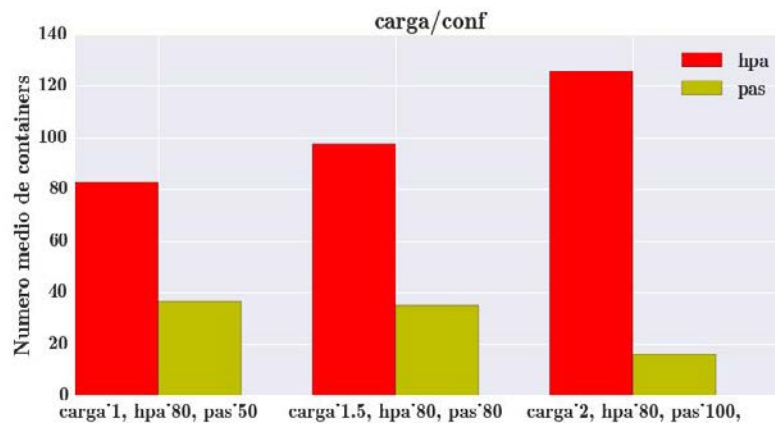


Figura 7. Comparativo entre HPA e PAS (número médio de containers)

tar a complexidade do tempo de balanceamento de carga e não necessariamente produzir melhores tempos de resposta.

Os resultados obtidos mostram que o algoritmo PAS, proposto neste trabalho apresenta potencial para promover a otimização de alocação do número de *containers* em um ambiente de computação em nuvem.

6. Conclusão e Trabalhos Futuros

Este trabalho apresentou o algoritmo PAS (*PID based Autoscaler*) e os resultados mostram um potencial para prover auto elasticidade a um ambiente de computação em nuvem baseado em *containers*. A comparação com a ferramenta de auto escalabilidade nativa do *Kubernetes*, mostra uma maior eficiência da solução proposta para alocação da quantidade de *containers* com o objetivo de cumprir limiares pré estabelecidos de tempo de resposta na execução de requisições para aplicações Web.

O uso de ferramentas como *spark* e *flume* possibilita que a arquitetura seja escalável, podendo vir a ser utilizada por *sites* com servidores com grande número de acessos, em que a série temporal resultante do número de requisições em diversas escalas de tempo tende a ter um alto grau de autossimilaridade.

A proposta possibilita um uso eficiente de recursos, o que pode permitir um menor custo de operação da infraestrutura instalada e dos serviços oferecidos em nuvem.

Os trabalhos futuros pretendem validar os resultados experimentais em um maior número de cenários e confronta-los com medições em ambientes de produção que permitam expandir a aplicação desta pesquisa para ambientes reais de computação em nuvem.

Referências

- Anicas, M. (2014). Mitchel Anicas an introduction to haproxy and load balancing concepts. <https://www.digitalocean.com/community/tutorials/an-introduction-to-haproxy-and-load-balancing-concepts>. [Acessado em 03/04/2016].
- Docker (2016). Docker the definitive guide to docker containers. <https://www.Docker.com/sites/default/files/WP-%20Definitive%20Guide%20To%20Containers.pdf>. [Acessado em 03/04/2016].

- F5 (2012). F5 load balancing 101 - nuts and bolts. <https://f5.com/resources/white-papers/load-balancing-101-nuts-and-bolts>. [Accessado em 03/04/2016].
- Flume (2016). Flume flume user guide. <https://flume.apache.org/FlumeUserGuide.html>. [Accessado em 03/04/2016].
- Gilly, K., Juiz, C., and Puigjaner, R. (2011). An up-to-date survey in web load balancing. *World Wide Web*, 14(2):105–131.
- Gong, Z., Gu, X., and Wilkes, J. (2010). Press: Predictive elastic resource scaling for cloud systems. In *Network and Service Management (CNSM), 2010 International Conference on*, pages 9–16. IEEE.
- Google (2016a). Google container cluster manager from google. <https://github.com/kubernetes/kubernetes>. [Accessado em 03/04/2016].
- Google (2016b). Google horizontal pod autoscaler. <https://github.com/kubernetes/kubernetes/blob/release-1.2/docs/design/horizontal-pod-autoscaler.md>. [Accessado em 03/04/2016].
- Instruments, N. (2015). National Instruments explicando a teoria pid. <http://www.ni.com/white-paper/3782/pt/>. [Accessado em 20/08/2015].
- Kairos (2015). Kairos time series data storage in redis, mongo, sql and cassandra. <https://pypi.python.org/pypi/kairos>. [Accessado em 03/04/2016].
- Kettani, H., Gubner, J., et al. (2002). A novel approach to the estimation of the hurst parameter in self-similar traffic. In *Local Computer Networks, 2002. Proceedings. LCN 2002. 27th Annual IEEE Conference on*, pages 160–165. IEEE.
- Lorido-Bostrán, T., Miguel-Alonso, J., and Lozano, J. A. (2012). Auto-scaling techniques for elastic applications in cloud environments. *Department of Computer Architecture and Technology, University of Basque Country, Tech. Rep. EHU-KAT-IK-09*, 12:2012.
- Martin, N. (2015). Nick Martin a brief history of docker containers' overnight success. <http://searchservervirtualization.techtarget.com/feature/A-brief-history-of-Docker-Containers-overnight-success>. [Accessado em 03/04/2016].
- Poddar, R., Vishnoi, A., and Mann, V. (2015). Haven: Holistic load balancing and auto scaling in the cloud.
- Redis (2015). Redis redis documentation. <http://redis.io/documentation>. [Accessado em 03/04/2016].
- Spark (2015a). Spark spark programming guide. <https://spark.apache.org/docs/1.5.2/programming-guide.html>. [Accessado em 03/04/2016].
- Spark (2015b). Spark spark streaming programming guide. <https://spark.apache.org/docs/1.5.2/streaming-programming-guide.html>. [Accessado em 03/04/2016].
- Tarreau, W. (2015). HAProxy haproxy configuration manual. <http://www.haproxy.org/download/1.5/doc/configuration.txt>. [Accessado em 03/04/2016].

Segurança de Sistemas de Computação em Nuvem: estado, desafios e perspectivas

João J. C. Gondim, Priscila A. Solis Barreto, Luis Alberto B. Pacheco

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Caixa Postal 70.910-900 – Brasília – DF – Brasil

gondim,pris@unb.br, luisbelem@gmail.com

Abstract. *Cloud computing is an emerging technology that is more and more popular. Cost, scalability, resiliency, on demand rightsizing, among other benefits, makes it irresistible. However, in terms of security those benefits come with many issues and open questions. This paper reviews some of the main issues involved and indicates how they have been addressed. In order to demonstrate some of the complexities that arise in cloud environments, prospective malicious use scenarios are presented to illustrate possible attack consequences.*

1. Introdução

A computação em nuvem – *cloud computing* – é uma tecnologia emergente que tem se tornado cada vez mais disseminada como um modelo computacional capaz de armazenar e processar grandes volumes de dados utilizando hardware que é suprido como *commodity*. A disponibilidade de serviços oferecidos por provedores já respeitados e a elegância técnica das soluções adotadas juntamente com um modelo conceitual de fácil compreensão são alguns dos fatores que influenciaram esta disseminação.

Em termos de benefícios, a computação em nuvem pode chegar a oferecer até a transferência total do ônus da operação da infraestrutura e serviços de tecnologia da informação – TI – a custo baixo, se comparado a outras soluções não tão abrangentes - como o *outsourcing* tradicional de serviços de TI.

Os benefícios citados, aos quais também podemos somar escalabilidade e *rightsizing* automático, são tão atrativos, tanto tecnicamente quanto financeiramente, que a adoção de uma solução de computação em nuvem se torna quase inevitável, principalmente por médias e pequenas empresas. Para estas, a computação em nuvem pode ser a única forma de se ter acesso a serviços de TI que elas não teriam como operar ou custear. Entretanto, todo este possível ganho traz consigo questões no mínimo preocupantes.

A transferência, possivelmente total, do ônus da operação, tanto do ponto de vista técnico quanto gerencial, da infraestrutura e serviços de TI, que sem dúvida é um dos maiores atrativos, é inerentemente uma fonte de risco. Este risco pode ser aceitável, se, por exemplo, estamos falando da caixa de correio de um usuário do Gmail. Entretanto, se o que se pulveriza na nuvem são os dados operacionais ou estratégicos de uma organização – como bancos de dados com informações de clientes – a situação do risco se torna bem mais sensível. Desta forma, o ambiente, o contexto, os requisitos e possíveis ameaças levam a necessidades de segurança bem mais complexas onde os conceitos e práticas atualmente disponíveis podem não ser adequados, suficientes ou aplicáveis.

Este artigo tem por objetivo abordar as principais preocupações e necessidades de segurança relacionadas aos ambientes de computação em nuvem, estando organizado em seções. A seção a seguir dedica-se a rever as várias formas como o conceito de computação em nuvem está definido na literatura, visando chegar a um denominador comum para os fins deste artigo. A seção 3 aborda as principais questões de segurança relacionadas aos ambientes de computação em nuvem, apontando como estas questões estão sendo atacadas. A seção 4 explora a utilização de computação em nuvem pelo governo brasileiro, comparando com a abordagem utilizada pelo Reino Unido e Estados Unidos. A seção 5 apresenta os possíveis cenários de uso malicioso de nuvens. A última seção conclui o artigo com comentários finais.

2. Computação em nuvem

O conceito de computação em nuvem está em evolução. Na literatura, temos várias definições, como em [Forrest and Barthold 2009], [Crowd 2012] e [Vaquero et al. 2008], onde são listadas mais de vinte definições. Para efeito de nossa discussão, vamos adotar a definição oferecida pelo NIST [Mell and Grance 2011] (tradução nossa):

Computação em nuvem é um modelo para habilitar, de forma conveniente e sob demanda, o acesso via rede de recursos computacionais configuráveis - como redes, servidores, armazenamento, aplicações e serviços - que podem ser rapidamente provisionados e liberados com mínimo esforço gerencial ou interação com o provedor. Este modelo promove a disponibilidade e é composto de cinco características básicas: auto serviço sob demanda, amplo acesso via rede, recursos organizados em *pool*, rápida elasticidade e serviço medido; três tipos de serviço: SaaS – software como serviço, PaaS – plataforma como serviço, e IaaS – infraestrutura como serviço; e quatro modos de emprego: público, privado, comunitário e híbrido.

Abaixo estão descritas as cinco características básicas necessárias em um serviço de nuvem:

- Auto serviço sob demanda: o consumidor tem condições de dimensionar os recursos computacionais como desejado, sem necessidade de interação com o provedor do serviço.
- Amplo acesso via rede: recursos computacionais estão disponíveis para acesso através da Internet, por meio de tecnologias padronizadas que possibilitam o acesso por diferentes tipos de dispositivos.
- Recursos organizados em *pool*: os recursos computacionais do provedor são utilizados de forma compartilhada pelos clientes. Esses recursos são atribuídos de forma dinâmica e o cliente não necessita de conhecimento sobre a localização exata do recurso.
- Rápida elasticidade: recursos devem ser escaláveis rapidamente, a capacidade deve aumentar e diminuir de acordo com a demanda necessária. As capacidades disponíveis devem parecer infinitas ao cliente.
- Serviço medido: sistemas em nuvem controlam e otimizam os recursos de acordo com métricas. Essas métricas devem ser monitoradas, controladas e relatadas, provendo transparência tanto para o provedor quanto para o cliente do serviço.

A computação em nuvem é implementada por meio de vários computadores individuais interconectados em grelha, que, combinando seus recursos em paralelo, oferece grandes capacidades de processamento e armazenamento. De acordo com [Veras 2012], uma arquitetura em nuvem pode ser implementada de quatro maneiras diferentes:

- Nuvem privada: o próprio cliente opera e gerencia a infraestrutura em nuvem. Essa arquitetura é apropriada para organizações que não podem divulgar seus dados publicamente.
- Nuvem pública: utilizada pelo público em geral, pode ser fornecida por organizações de qualquer espécie.
- Nuvem comunitária: infraestrutura utilizada por uma comunidade com interesses em comum, pode ser operada por uma ou mais entidades pertencentes à comunidade ou mesmo por terceiros.
- Nuvem híbrida: infraestrutura composta por duas ou mais dos outros tipos de nuvem, que continuam como entidades separadas mas propiciam a portabilidade de dados e aplicações.

Pode-se arguir que esta definição deixa de fora vários serviços que também podem ser disponibilizados em nuvem, mas ela é completa o suficiente para caracterizar os sistemas em nuvem sem se ater a tecnologias específicas. Entretanto, os aspectos relativos às tecnologias utilizadas para implementação de computação em nuvem são fundamentais para que se obtenham as características citadas acima. Além disso, como veremos adiante, estas tecnologias tem questões de segurança de per si, que se desdobram e potencializam no ambiente de nuvem.

Tecnicamente, a computação em nuvem está calcada em uma arquitetura voltada a serviços – SOA - e virtualização. SOA é uma abordagem arquitetural em que as aplicações são oferecidas como serviços, onde estas aplicações são tipicamente aplicações web. Já a virtualização, consiste na multiplexação dos recursos de um sistema computacional entre vários outros sistemas computacionais – as máquinas virtuais. Cada máquina virtual tem controle dos recursos a ela alocados, independente do sistema que os provê. Desta forma, cada máquina virtual tem a percepção de exclusividade no uso dos recursos, como um sistema real.

Revisitando a definição acima, temos o acesso via rede que permite que os recursos da nuvem estejam distribuídos, inclusive geograficamente, em diversas localidades. A distribuição aliada à redundância são fatores que levam a se obter elevada disponibilidade. Em contrapartida, a complexidade do controle de acesso aumenta, sendo necessários mecanismos apropriados de autenticação.

3. Segurança na nuvem

Um sistema seguro deve preservar a confidencialidade, integridade e disponibilidade da informação [ISO 2009]. De uma forma geral, os provedores de serviços de computação em nuvem, principalmente os maiores, utilizam soluções de segurança de acordo com as melhores práticas. Mas há uma grande diversidade [Rimal et al. 2009]. As soluções mais usadas incluem *firewalls*, VPNs, uso de certificados digitais, VLANs dedicadas e listas de controle de acesso. Entretanto, há provedores que deixam a cargo do usuário prover a segurança que este achar conveniente, sendo a postura mais difundida entre os provedores que oferecem IaaS – infraestrutura como serviço.

Diante deste cenário heterogêneo, vamos centrar nossa análise dos aspectos de segurança em tópicos conceituais do modelo. Sistemas de computação em nuvem são massivamente complexos, mas tem na sua base primitivas simples e unidades funcionais comuns [Mell 2009]. Assim, o problema da segurança pode ser tratado.

Algumas das principais questões são:

- **Confiança (*trust*):**

Como apontado anteriormente, a transferência, possivelmente total, do ônus da operação, tanto do ponto de vista técnico quanto gerencial, da infraestrutura e serviços de TI, que sem dúvida é muito atraente, é inerentemente uma fonte relevante de risco. Deve-se observar que a decisão pela citada transferência de responsabilidade requer que se tenha grande confiança no provedor que se venha a contratar. Normalmente, os provedores são avaliados quanto à reputação até o momento da decisão de usar a nuvem. Por maiores que sejam as precauções que orientem a decisão, e normalmente o custo é o principal fator na análise, não se pode garantir a total segurança da informação. A forma como normalmente se trata este tipo de situação é por incluir nos SLAs - acordos de nível de serviço - salvaguardas que garantam que junto com a transferência de responsabilidade também ocorra alguma transferência de risco. (Deve-se notar que mesmo desta forma, ainda não se tem a garantia desejada). Hoje, não há consenso quanto aos SLAs oferecendo este nível de compromisso. Outro aspecto em relação à confiança diz respeito a como os sistemas que controlam a nuvem tratam as interações entre diferentes processos e os dados que cada um deles podem acessar. Um critério comum para controle é baseado em *trust*. Normalmente estes sistemas inferem *trust* a partir de reputação. Apesar de haver muito esforço de pesquisa na área [Habib et al. 2012], as formalizações do conceito ainda precisam amadurecer e posteriormente vir a ser adotadas pelos provedores.

- **Múltiplos contratantes (*multi-tenancy*):**

As nuvens acomodam vários contratantes (*tenants*, literalmente inquilinos), sendo necessário garantir a necessária separação entre eles. Os mecanismos atuais tem por base a virtualização e os sistemas de confiança. Entretanto, é preciso que no nível mais básico esta separação se mantenha. Esta necessidade vem como uma forma de evitar que uma pane, incidente ou evento no contexto de um contratante afete outros. Entretanto, pode-se visualizar um cenário em que tivéssemos um incidente de segurança envolvendo um dos sistemas de um contratante, e que no processo de investigação forense fosse necessário “interditar” o contexto de outros contratantes. Situações como esta se não são inaceitáveis, são certamente indesejáveis, o que pode piorar se a distribuição geográfica for levada em conta. (Deve-se notar que situações deste tipo já podem ocorrer em grande sistemas de *storage*). Além do desafio do ponto de vista forense, há um desafio maior do ponto de vista do gerenciamento e alocação de recursos, não apenas de armazenamento, da nuvem.

- **Privacidade e Identidade:**

A privacidade e a identidade são duas questões complementares no contexto dos ambientes em nuvem. Inicialmente, a distribuição geográfica pode implicar na transnacionalidade, levando a diferentes requisitos legais para serem atendidos pelos provedores da nuvem. Com a pulverização dos dados na nuvem, pode ser

necessário atender aos diferentes requisitos globalmente, ainda que isso seja percebido com sendo um requisito local. A privacidade vai nortear quais dados podem ser expostos e a quem. Para que se tenha o controle efetivo, é necessário estabelecer, também de maneira eficaz, o “que” vem a ser quem. É aí que entram os aspectos relativos à identidade. Os serviços de identidade deverão ser capazes de suportar diferentes cenários [Cavoukian et al. 2008], tais como:

- fácil gerenciamento de identidades online;
- participar em atividades colaborativas online;
- sistemas de pagamento usando telefones celulares ou RFID com privacidade;
- uma infraestrutura de prontuários médicos.

Entre outros requisitos, os serviços de identidade deverão:

- ser independentes de dispositivo;
- *single sign-on*: permitir que uma única validação dê acesso a vários (preferencialmente todos) serviços;
- permitir pseudônimos e várias identidades discretas, para preservar privacidade;
- ser interoperáveis;
- permitir o gerenciamento federado da identidade;
- ser transparente e auditável.

Muito ainda precisa ser feito para que serviços atendendo estes requisitos sejam disponibilizados. Na verdade, como primeiro passo é necessária uma discussão entre os vários segmentos da sociedade para se chegar ao conjunto de requisitos a serem atendidos.

- **Uso de criptografia:**

O uso de criptografia na proteção dos dados, tanto quando armazenados quanto em trânsito possibilita se obter garantias mínimas quanto à confidencialidade e integridade [Kaufman 2009]. Entretanto, o *overhead* computacional decorrente da aplicação dos algoritmos de cifração e decifração, o esforço adicional do gerenciamento de chaves e o impacto sobre as necessidades de interoperabilidade colocam-se como alguns dos desafios para o uso eficiente da criptografia. (Deve-se registrar que nem todos os provedores implementam os recursos já disponíveis hoje). Avanços recentes no campo da criptografia completamente homomófica [Gentry 2009], que permitiria operações sobre os dados cifrados – sem necessidade de decifração, abrem possíveis caminhos para um nível bem mais elevado de segurança. Infelizmente, os sistemas disponíveis ainda são extremamente ineficientes, mas dada a importância das aplicações e o esforço de pesquisa que está ocorrendo, pode-se esperar bons resultados, que venham a ser absorvidos pelos provedores.

- **Conformidade:**

A conformidade hoje já constitui um desafio para os responsáveis pela gestão de TI. O quadro de normas e padrões atuais é complexo e de implementação onerosa sem que se garanta segurança efetiva. Estas normas e padrões tem sido aplicados e adaptados a ambientes em nuvem [Chen and Yoon 2010], mas ainda se faz necessário desenvolver um quadro normativo específico. Possivelmente, estas normas virão no devido tempo como resultado de boas práticas.

Além destas questões, há uma série de questões que são decorrentes das tecnologias utilizadas nos sistemas em nuvem. Uma discussão aprofundada das mesmas é muito ampla e está fora do escopo deste artigo. Entretanto, a seção a seguir ilustra como as propriedades da nuvem podem afetar a exploração de uma vulnerabilidade em uma das suas tecnologias de suporte.

4. Computação em nuvem: utilização no governo brasileiro

A utilização da computação em nuvem pode ser muito produtiva para entidades governamentais, a mesma redução de custos e agilidade na implantação observada em outras áreas também é válida neste caso. Um dos maiores desafios relacionados à adoção de computação em nuvem por entidades governamentais é a proteção de informações sensíveis.

Neste âmbito, são observadas duas características da computação em nuvem que causam preocupação:

1. A localidade da informação: em ambientes em nuvem muitas vezes a localidade da informação não é conhecida pelos usuários, o provedor pode possuir vários *data-centers* em diversas localidades e realizar um balanceamento de carga para melhorar o desempenho do seu serviço. Para entidades governamentais é imperativo que as informações sensíveis armazenadas na nuvem estejam em território nacional, pois de outra forma tais informações estariam sob jurisdição de outros países.
2. A proteção da informação: é preciso se assegurar de que o provedor da nuvem implementa os protocolos necessários para que informações sensíveis possuam a proteção necessária, essa pode ser uma tarefa de difícil execução para entidades governamentais que não possuam expertise nessa área.

No Brasil, a Lei de Acesso à Informação (LAI) [da República 2011] entrou em vigor em 2012, e assegurou o acesso à informação pública por pessoas físicas e jurídicas. De acordo com a LAI, qualquer informação produzida ou sob guarda do poder público é pública e acessível até o contrário seja estabelecido, os principais aspectos da LAI são:

- Toda informação gerada é pública até que seja tornada sigilosa;
- Não há exigência de motivação para o compartilhamento da informação;
- As hipóteses para se tornar uma informação sigilosa são limitadas e estabelecidas na lei;
- Não há cobrança para o fornecimento da informação, exceto algum custo de reprodução;
- Transparência ativa: informações de interesse público devem ser divulgadas de antemão;
- Transparência passiva: o acesso à informação deve ser facilitado por meio de procedimentos e prazos.

O artigo 24 da LAI define os graus de classificação de informação, são essas as informações que precisam de atenção à segurança quando armazenadas na nuvem:

- Reservado: a informação é sigilosa por 5 anos, sem possibilidade de prorrogação;
- Secreto: a informação é sigilosa por 15 anos, sem possibilidade de prorrogação;
- Ultrassecreto: a informação é sigilosa por 25 anos, e pode ser prorrogada apenas uma vez por mais 25 anos;

Recentemente o Ministério do Planejamento, Orçamento e Gestão divulgou um guia de boas práticas, orientações e vedações para contratação de Serviços de Computação em Nuvem [MPOG 2016], abordando os tempos de preocupação já citados. O documento visa incentivar o uso da computação em nuvem pelas entidades governamentais, a fim de reduzir os custos da implementação e manutenção de projetos de tecnologia da informação.

O guia traz a política de *cloud first*, já abordada em vários outros países [Office 2013] [Kundra 2011], limitando a contratação de salas-cofre à aprovação da Secretaria de Tecnologia da Informação (STI) do ministério. A recomendação estabelece que serviços que não comprometam a segurança nacional podem ser estabelecidos em nuvens híbridas, já serviços com informações classificadas devem utilizar provedores da Administração Pública Federal, dessa forma informações classificadas não ficam de posse de provedores privados. O guia não indica formas de auditoria aos recursos de segurança da informação implementados pela contratada, dessa forma o armazenamento de informações sensíveis pela mesma não seria indicado.

A recomendação também aborda a preocupação em relação à localidade da informação armazenada na nuvem, é indicado que deve-se estabelecer em contratado a obrigatoriedade do armazenamento exclusivo em território nacional, mesmo para informações não classificadas, incluindo replicação e cópias de segurança (*backups*). Por fim também são abordados itens como disponibilidade e portabilidade, a fim de garantir a qualidade do serviço a ser contratado.

A seguir são apresentados dois estudos de caso referentes à contratação de provedores de serviços de computação em nuvem por entidades governamentais, primeiramente é apresentada a solução escolhida pelo Reino Unido e em seguida pelos Estados Unidos da América. Os dois casos apresentados possuem um nível de maturidade avançado em relação ao Brasil e podem servir de exemplos para o avanço das normas brasileiras.

4.1. Reino Unido - G-Cloud

A iniciativa G-Cloud teve início em 2012 [BBC 2012], consiste em chamadas para credenciamento de serviços de nuvem oferecidos por empresas privadas. Após aprovação pelo governo, as empresas credenciadas entram para um catálogo para facilitar a contratação por entidades governamentais.

O processo de credenciamento não envolve nenhuma auditoria para certificar as empresas em relação à segurança ou qualquer outro aspecto, é composto apenas por um termo de serviço onde a empresa atesta estar de acordo com as leis de proteção de dados britânica, incluindo o armazenamento da informação apenas em território nacional. O *Digital MarketPlace* é uma loja virtual que contém as empresas privadas credenciadas pelo G-Cloud, entidades governamentais utilizam a loja para facilitar o processo de contratação.

No Reino Unido informações sensíveis possuem 3 classificações [Office 2014]:

- Official: a maior parte da informação criada é dessa categoria, informações oficiais podem ocasionar consequências danosas caso divulgadas mas não necessitam de um nível mais alto de proteção;
- Secret: informação muito sensível, necessita de medidas especiais para sua proteção, caso divulgada pode causar danos militares ou internacionais;

- Top secret: informação que caso divulgada pode causar ameaças à vida ou economia do Reino Unido.

Ao contrário da recomendação brasileira, informações classificadas do tipo *Official* podem ser armazenadas em nuvens de servidores privados. O governo britânico delega a responsabilidade de proteção da informação ao seu dono, dessa forma a entidade governamental contratante do serviço de nuvem que possui responsabilidade de aferir a devida proteção à informação. Para facilitar esse aferimento o governo britânico criou um guia com 14 princípios de segurança para serviços em nuvem [CESG 2014], esse guia fornece todas as ferramentas necessárias para que entidades governamentais assegurem que o provedor contratado gerencia a informação sensível de forma apropriada. A seguir os 14 princípios são apresentados, empresas contratadas precisam implementar todos eles:

1. Proteção dos dados em navegação
2. Proteção dos dados armazenados
3. Separação entre clientes
4. Governança da segurança da informação
5. Segurança operacional
6. Segurança de pessoal
7. Segurança do desenvolvimento
8. Segurança da cadeia logística
9. Segurança do gerenciamento de clientes
10. Autenticação e identificação
11. Proteção das interfaces externas
12. Segurança da administração do serviço
13. Auditoria do serviços
14. Uso seguro pelos clientes

4.2. Estados Unidos da América - FedRAMP

O FedRAMP [EUA 2011] (do inglês *The Federal Risk and Authorization Management Program*) é um programa dos Estados Unidos da América para credenciar empresas privadas à fornecer serviços de computação em nuvem para entidades governamentais americanas. O processo de credenciamento é realizado por empresas terceirizadas e, ao contrário da iniciativa britânica, envolve auditoria do serviço, verificando sua conformidade à normas preestabelecidas.

Após a realização da auditoria empresas recebem uma autorização de operação temporária e entram para uma loja virtual para facilitar a contratação por entidades governamentais. Ao contrário do processo britânico, as entidades governamentais não precisam aferir os mecanismos de segurança implementados pelas empresas contratadas, pois esse trabalho já foi realizado por empresas terceirizadas no processo de credenciamento.

A iniciativa também prevê a utilização de computação em nuvem para informações militares, a DISA, Agência de Defesa de Sistemas da Informação, estende o processo de credenciamento com normas mais rígidas para que os provedores privados possam participar dessa modalidade [Boyd 2015].

A abordagem escolhida pelos EUA pode ser considerada mais segura, pois os provedores a serem contratados passam por um processo rigoroso de certificação, já na abordagem utilizada pelo Reino Unido, as entidades governamentais que possuem a responsabilidade de certificar os provedores.

5. Cenários de uso malicioso da computação em nuvem

Como foi dito anteriormente, muito do que se implementa hoje em termos de segurança em ambientes em nuvem é adaptado. Uma análise das atuais principais ameaças de segurança à computação em nuvem é realizada por [Coppolino et al. 2016], um dos vetores de ataque explorados é relacionado à virtualização, uma das principais tecnologias de utilizadas na implementação de sistemas em nuvem.

Cabe ressaltar que a virtualização em si tem um bom nível de segurança – uma busca recente apresenta poucos resultados significativos ([Shree et al. 2015], [Dubrulle et al. 2015]). A vulnerabilidade em questão está descrita em [Rocha et al. 2013], e decorre da utilização combinada de mecanismos para a comunicação entre máquinas virtuais e a obtenção de informações de uma máquina virtual *guest* pelo *host* de forma automática. A utilização dessa combinação possibilita a obtenção de informações compartilhadas entre uma máquina virtual e o *host* por outra máquina virtual. Como consequência o atacante pode obter dados de monitoramento, tais como utilização da memória, eventos de *hardware* e acesso aos registradores do processador virtual.

Numa vertente mais especulativa, pode-se vislumbrar o cenário descrito como sendo o ponto de entrada para o comprometimento de toda uma nuvem, se seguido de uma escalção de privilégio. (Cabe dizer que não há relatos de que isto tenha ocorrido.)

Utilizando nuvens publicas, é possível a criação de *botclouds*, que funcionam como *botnets*, dessa forma a computação em nuvem também facilita a implementação de ataques massivos, como por exemplo ataques DDoS [Badis et al. 2014].

Continuando o exercício, uma grande nuvem comprometida, agindo como *botnet*, poderia viabilizar ataques capazes de afetar a estabilidade global do roteamento da Internet. Um ataque deste tipo foi simulado e seria viável com uma *botnet* de duzentos e cinquenta mil *bots* convenientemente distribuídos geograficamente.

A finalidade de expor estes cenários hipotéticos é tentar demonstrar a escala dos problemas que podem vir a ser enfrentados no caso de segurança inadequada nos ambientes de computação em nuvem.

6. Conclusão

Conforme vimos, a computação em nuvem tem benefícios tão atraentes que “*resistance is futile*”. Custo, escalabilidade, *on demand rightsizing*, entre outros benefícios são provavelmente a única forma de se ter acesso a serviços de TI para pequenas e médias empresas. Entretanto, há um preço do ponto de vista de segurança e as questões levantadas evoluem na medida em que o próprio conceito evolui.

De certa forma, as questões de segurança levantadas pelo ambiente em nuvem não são totalmente novas, e já vem sendo tratadas. Entretanto, as características dos sistemas em nuvem levam a novas considerações.

Conforme ilustramos, o comprometimento de uma grande nuvem poderia provocar na Internet efeitos devastadores. Estes custos não previstos certamente não seriam aceitáveis. O cenário descrito demonstra claramente como a nuvem pode potencializar incidentes que teriam consequências localizadas.

A computação em nuvem já é algo disseminado e amplamente usado. Resta-nos entender os riscos e trabalhar no sentido de reduzi-los.

Referências

- Badis, H., Doyen, G., and Khatoun, R. (2014). Understanding botclouds from a system perspective: a principal component analysis. In *Network Operations and Management Symposium (NOMS), 2014 IEEE*, pages 1–9. IEEE.
- BBC (2012). Cloudstore: Government launches public sector app store.
- Boyd, A. (2015). Disa security guide outlines future of dod cloud.
- Cavoukian, A. et al. (2008). Privacy and digital identity: Implications for the internet. In *Proceedings from Identity in the Information Society Workshop*.
- CESG (2014). Summary of cloud security principles.
- Chen, Z. and Yoon, J. (2010). It auditing to assure a secure cloud computing. In *Services (SERVICES-1), 2010 6th World Congress on*, pages 253–259. IEEE.
- Coppolino, L., D'Antonio, S., Mazzeo, G., and Romano, L. (2016). Cloud security: Emerging threats and current solutions. *Computers & Electrical Engineering*.
- Crowd, O. (2012). Open crowd cloud taxonomy.
- da República, P. (2011). Lei nº 12.527 - lei de acesso à informação.
http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm.
- Dubrulle, P., Sirdey, R., Dore, P., Aichouch, M., and Ohayon, E. (2015). Blind hypervision to protect virtual machine privacy against hypervisor escape vulnerabilities. In *Industrial Informatics (INDIN), 2015 IEEE 13th International Conference on*, pages 1394–1399. IEEE.
- EUA (2011). The federal risk and authorization management program, or fedramp.
- Forrest, W. and Barthold, C. (2009). Clearing the air on cloud computing. *Discussion Document from McKinsey and Company*.
- Gentry, C. (2009). *A fully homomorphic encryption scheme*. PhD thesis, Stanford University.
- Habib, S. M., Hauke, S., Ries, S., and Mühlhäuser, M. (2012). Trust as a facilitator in cloud computing: a survey. *Journal of Cloud Computing*, 1(1):1–18.
- ISO (2009). Iso/iec 27000:2009 information technology - security techniques - information security management systems - overview and vocabulary.
- Kaufman, L. M. (2009). Data security in the world of cloud computing. *Security & Privacy, IEEE*, 7(4):61–64.
- Kundra, V. (2011). Federal cloud computing strategy.
- Mell, P. (2009). Nist presentation on effectively and securely using the cloud computing paradigm v26.
- Mell, P. and Grance, T. (2011). The nist definition of cloud computing.

- MPOG (2016). Boas práticas, orientações e vedações para contratação de serviços de computação em nuvem.
- Office, C. (2013). Government adopts 'cloud first' policy for public sector it.
- Office, C. (2014). Government security classifications.
- Rimal, B. P., Choi, E., and Lumb, I. (2009). A taxonomy and survey of cloud computing systems. In *2009 Fifth International Joint Conference on INC, IMS and IDC*, pages 44–51. Ieee.
- Rocha, F., Gross, T., and Van Moorsel, A. (2013). Defense-in-depth against malicious insiders in the cloud. In *Cloud Engineering (IC2E), 2013 IEEE International Conference on*, pages 88–97. IEEE.
- Shree, T., Kumar, M., and Badal, N. (2015). Towards a hypervisor security-based service and its research challenges. *International Journal of Computer Applications*, 115(17).
- Vaquero, L. M., Rodero-Merino, L., Caceres, J., and Lindner, M. (2008). A break in the clouds: towards a cloud definition. *ACM SIGCOMM Computer Communication Review*, 39(1):50–55.
- Veras, M. (2012). *Cloud Computing: nova arquitetura da TI*. Brasport.

XXIV WEI - Workshop Sobre Educação em Computação

Apresentação

O WEI, Workshop sobre Educação em Computação, é o mais importante fórum brasileiro, promovido pela SBC, que traz uma oportunidade única de pesquisadores e professores discutirem problemas e temas relacionados à pesquisa, às iniciativas e avaliação da educação na área de Computação. O fórum discute diferentes temas relacionados à área, apresentando novos estudos, metodologias, ferramentas e discutindo experiências vividas em todo o Brasil.

Em sua 24ª edição, o WEI traz como principal tema “Ensino de computação: a interdisciplinaridade como prática”. A complexidade da sociedade moderna e dos sistemas sobre os quais ela se organiza exigem a formação de pessoas capazes de pensar de forma sistêmica. Para tanto, é necessário transpor não apenas a fragmentação dos saberes em disciplinas específicas de computação, mas também dialogar com outras áreas de conhecimento. Embora os educadores tenham consciência da importância do tema, ainda é preciso sistematizar práticas e métodos de ensino que efetivamente despertem a interdisciplinaridade.

As atividades previstas para o WEI têm o intuito de promover ampla discussão sobre o tema, compreendendo apresentação de artigos no contexto específico de cada trilha, além de palestras convidadas, painéis e grupos de trabalho (fóruns de discussão). Esses grupos têm atuado na discussão acerca dos currículos de referência da SBC, além de discutir outras questões relativas aos cursos de graduação e pós-graduação da área de Computação.

Nesta edição, foram recebidas 120 submissões de artigos para publicação, dentre artigos de pesquisa, ensaios, relatos de experiências e descrição de ferramentas ou recursos educacionais. Dentre estas, foram aceitos 47 artigos, sendo 13 artigos de pesquisa, 24 artigos de intervenção ou relato de experiências, 11 artigos de ferramentas ou recursos educacionais, e 02 ensaios. Para o Comitê de Programa, cinco destes trabalhos se destacaram e concorrerão, durante o XXXVI CSBC, ao prêmio de Best Paper WEI 2016.

Cada trabalho foi avaliado por pelo menos dois revisores. Quando a discrepância entre as avaliações foi considerada grande, um terceiro ou até mesmo um quarto revisor foi convidado para analisar o trabalho.

Aos autores, palestrantes, painelistas e demais membros da comunidade que contribuíram para a realização de mais uma edição do workshop, um agradecimento especial pelo reconhecimento e confiança no evento. Ao Comitê de Programa e aos revisores externos pelo excelente trabalho de avaliação de artigos, e pela disponibilidade dedicada ao WEI, nossa gratidão.

Esperamos que os trabalhos que compõe este XXIV WEI sejam inspiradores de novos estudos e intervenções no campo da educação em Computação. A todos os

participantes nossas boas vindas e o desejo de que todos possam aproveitar ao máximo os eventos que compõem esta edição do Workshop sobre Educação em Computação!

Ecivaldo Matos (UFBA)

Leandro Galvão (UFAM)

Coordenadores do WEI 2016

Comitê de Organização

Coordenação Geral

Ecivaldo Matos, UFBA
Leandro Galvão, UFAM

Coordenação Local

Lucia Giraffa, PUCRS

Comitê de Programa

Alberto Castro, UFAM
Alberto Signoretti, UERN
Alcides Calsavara, PUCPR
Alexandre Cidral, Univille
Alfio Martini, PUCRS
Ana Liz Araujo, UFPB
Andreia Malucelli, PUCPR
André Campos, UFRN
André Costa Drummond, UnB
André Lemos, IFTM
André Raabe, Universidade do Vale do Itajaí
Anna Friedericka Schwarzmüller, UFBA
Avelino Zorzo, PUCRS
Auri Vincenzi, UFSCar
Ayla Débora Rebouças, UFPB
Bruno Santana, UFRN
Carina Dorneles, UFSC
César França, UFRPE
Christina Chavez, UFBA
Clodis Boscaroli, UNIOESTE
Cristiano Maciel, UFMT
Daltro Nunes, UFRGS
Daniela Barreiro Claro, UFBA
Danielle Rousy, UFPE
David Moises Barreto dos Santos, UEFS
Débora Abdalla, UFBA
Debora Paiva, UFMS
Ecivaldo Matos, UFBA
Edmundo Spoto, UFG
Edson Pimentel, UFABC
Elaine Oliveira, UFAM

Eleandro Maschio, UTFPR
Ernesto Cuadros-Vargas, Universidad Catolica San Pablo, Perú
Fabíola Guerra Nakamura, UFAM
Fernanda Lima, UnB
Flávia Veloso Souza, UFPB
Igor Steinmacher, UTFPR
Isabel Nunes, UFRN
Ismar Frango Silveira, Makenzie
Jair Cavalcante Leite, UFRN
Jeane Melo, UFRPE
José Maria David, UFJF
José Palazzo M. de Oliveira, UFRGS
Juliana Diniz, UFRPE
Leandro Galvão, UFAM
Lúcia Giraffa, PUCRS
Luiz Merkle, UTFPR
Marcelo Duduchi, CEETEPS
Marco Winckler, Université Paul-Sabatier (França)
Marcos Borges, Unicamp
Maria Augusta Vieira Nelson, PUC-MG
Marilia Azevedo, CEETEPS
Mirella Moro, UFMG
Morganna Diniz, UNIRIO
Noemi Rodriguez, PUC-Rio
Patrícia Tedesco, UFPE
Pauleany Moraes, IFRN
Paulo Queiroz, UFERSA
Raquel Prates, UFMG
Renata Araújo, UNIRIO
Ricardo dos Santos Ferreira, UFV
Roberto Bittencourt, UEFS
Ronaldo Correia, UNESP
Ronaldo Mello, UFSC
Sandra Fabbri, UFSCar
Sílvia Amélia Bim, UTFPR
Simone Martins, UFF
Solange Rezende, USP
Sonia França, UFPE
Taciana Pontual, UFRPE
Taisy Weber, UFRGS
Tayana Conte, UFAM
Valéria Bastos, UFRJ
Vaninha Vieira, UFBA
Viviane Aureliano, IFPE
Wilson Henrique Veneziano, UnB
Yuska Paola Aguiar, UFPB

Revisores Convidados

Alba Lopes, IFRN
Ana Paula Chaves Steinmacher, UTFPR
Ana Paula Kuhn, UFMT
Christiane Santana, Unifacs
Debora Nascimento, UFS
Diego Marczal, UTFPR
Eduardo Alberto Felippsen, IFPR
Eliane Pereira, Unioeste
Elis Cristina Hernandez, UFSCar
Erik Antonio, UFSCar
Fábio Alves, UFMT
Francisco Mendes Neto, UFERSA
Jean Rosa, UFBA
Karla Susiane dos Santos Pereira, UFAM
Marcelo Quinta, UFG
Marco Aurélio Graciotto Silva, UTFPR
Michele Brandao, UFMG
Pedro Henrique Dias Valle, USP
Rômulo Silva, Unioeste
Rosemery Petter, UFMT
Sergio T. Carvalho, UFG
Willian Silva, Unioeste
Williamson Silva, UFAM

Painéis/Palestras

Computer Science for All (Ciência da Computação para Todos)

Prof. Dr. Jeffrey R. N. Forbes (Duke University / EUA)

Computer Science (CS) for All is a national effort announced by the Office of Science and Technology Policy (OSTP), led by the National Science Foundation (NSF) and the U.S. Department of Education (ED) in partnership with other federal agencies and private partners, to ensure Computer Science (CS) education is available to all students across the U.S. The effort accelerates ongoing efforts to build the knowledge base and capacity for rigorous and engaging CS education in schools across the nation. It will also bring together NSF and other federal agencies with private partners to support professional development for educators to teach CS.

Trabalhos aceitos

A game development-based strategy for teaching software design patterns through challenge-based learning under a flipped classroom approach

Ismar Frango Silveira, Universidade Cruzeiro do Sul, Brazil.

Avaliação de ferramentas de apoio ao ensino de técnicas de mineração de dados em cursos de graduação

José Viterbo, UFF, Brazil

Clodis Boscarioli, Universidade Estadual do Oeste do Paraná, Brazil

Flavia Bernardini, Universidade Federal Fluminense, Brazil

Mateus Felipe Teixeira, Universidade Estadual do Oeste do Paraná, Brazil.

Caracterização do público-alvo de jogos educacionais na área da computação

Paulo Eduardo Battistella, Christiane Gresse von Wangenheim, Universidade Federal de Santa Catarina, Brazil.

Computasseia: um jogo para o ensino de história da computação

Karen Figueiredo, Jean Carlos Santos, Universidade Federal de Mato Grosso, Brazil.

Construção e avaliação de um jogo educacional digital de modelagem de dados para o ambiente universitário

Stela Sampaio, Angelica Calazans, Bruno Magno Silva, Centro Universitario de Brasilia - Uniceub, Brazil.

Currículos de computação: por que permanecem assim?

Miguel Jonathan, Universidade Federal do Rio de Janeiro, Brazil.

Desafio jovem engenheiro: um torneio para incentivar o ingresso de estudantes de nível médio em carreiras de exatas

Poliana Vitti, Marcos Borges, Universidade Estadual de Campinas, Brazil.

Desafios e oportunidades ao processo de ensino e aprendizagem de programação para iniciantes

Viviane Aureliano, Universidade Federal de Pernambuco, Brazil

Patrícia Tedesco, Centro de Informática - UFPE, Brazil

Lucia Giraffa, PUC-RS, Brazil.

Desafios e perspectivas de um bacharelado interdisciplinar em ti

Daniel Araújo, Gustavo Girão, Jair Leite, Universidade Federal do Rio Grande do Norte, Brazil.

Duinoblocks for kids: um ambiente de programação em blocos para o ensino de conceitos básicos de programação a crianças do ensino fundamental i por meio da robótica educacional

Rubens Queiroz, Fábio Sampaio, Universidade Federal do Rio de Janeiro, Brazil.

Empregabilidade dos egressos tecnólogos do curso de análise e desenvolvimento de sistemas nos institutos federais mineiros

Giuliano Alkmim, IFNMG - Campus Januária, Brazil

Rosemary Dore, Universidade Federal de Minas Gerais, Brazil.

Ensino de linguagem de programação com ênfase na aprendizagem significativa

Gilberto Astolfi, Dejahyr Lopes Junior, Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul, Brazil.

Ensino de programação para futuros não-programadores: contextualizando os exercícios com as demais disciplinas de mesmo período letivo

Leandro Galvão, Bruno Gadelha, Fabíola Guerra Nakamura, David Fernandes, Elaine Oliveira, Universidade Federal do Amazonas, Brazil.

Estímulo à prática multidisciplinar no ensino de computação e design através de um evento de programação focado em problemas

Daniela Brauner, UFRGS - Universidade Federal do Rio Grande do Sul, Brazil

Pablo Margreff, Universidade Federal de Pelotas, Brazil

Alexandre Silva, Conrad Caine, Brazil

Vinicius Costa, IFSul, Brazil

Tatiana Tavares, Rafael Ferreira, Universidade Federal de Pelotas - UFPEL, Brazil.

Experiências de criação de aplicativos móveis com alunos do ensino fundamental e médio

Andrea Charao, Ana Luisa Solórzano, Patricia Pitthan Barcelos, Universidade Federal de Santa Maria, Brazil.

Experiências de ensino da computação desplugada em diferentes séries da educação fundamental maior

Alberto Alan Raiol, Ana Carina Lima, Breno Sousa, João Sarges, Rafael Santos, Fábio Bezerra, UFRA, Brazil.

Explorando a evolução do problema de pesquisa de doutorado: um estudo qualitativo no LER/UFPE

Enyo Gonçalves, Universidade Federal do Ceará, Brazil

Ana Carla Holanda, Dorgival Neto, Helena Bastos, Mayara Santos, Fabio Queda Bueno da Silva, Carla Silva, Jaelson Castro, Universidade Federal de Pernambuco, Brazil.

Hall of fame-shame: um padrão pedagógico para o ensino de programação

Andrea Charao, Alberto Kummer Neto, Benhur Stein, Patricia Pitthan Barcelos, Universidade Federal de Santa Maria - UFSM, Brazil.

Information organization via computational thinking: case study in a primary school classroom

Simone Cavalheiro, Universidade Federal de Pelotas, Brazil

Ana Marilza Pernas, UFPel, Brazil

Marilton Aguiar, UFPEL, Brazil

Luciana Foss, Andre Du Bois, Renata Reiser, Universidade Federal de Pelotas, Brazil

Clause Fátima de Brum, Piana, Brazil

Yuri Weissahn, Universidade Federal de Pelotas, Brazil.

L2PM: relato de uma experiência sobre o ensino integrado de lógica, programação e matemática para computação

Mateus Raeder, Universidade do Vale do Rio dos Sinos (UNISINOS), Brazil

Mônica Py, Sandro Rigo, Josaine Pinheiro, Universidade do Vale do Rio dos Sinos, Brazil.

LFApp: um aplicativo móvel para o ensino de linguagens formais e autômatos

Juventino Neto, Ricardo Terra, UFLA, Brazil.

Maratona para a popularização da ciência da computação na educação básica

Michel Brum, Marcia Kniphoff da Cruz, Universidade de Santa Cruz do Sul, Brazil.

Mão na massa: dinâmica vivencial para apoio ao ensino de gerenciamento de projetos de software

Pablo Schoeffel, Raul Wazlawick, Universidade do Estado de Santa Catarina - UFSC, Brazil.

MEGA GP: aplicando a gamificação no ensino de gerência de projetos

Sara Midori Tomisaki, Adler Diniz de Souza, Rodrigo Seabra, Universidade Federal de Itajubá, Brazil.

Melhora dos índices de aprovação pela formação de turmas com perfis de conhecimento semelhantes

Robinson Noronha, Universidade Tecnológica Federal do Paraná/Instituto Tecnológico de Aeronáutica, Brazil

Hilton Azevedo, Universidade Tecnológica Federal do Paraná-UTFPR, Brazil.

Modos contemporâneos de aprendizado e construção do conhecimento: reflexões sobre o ensino de teoria da computação para sistemas de informação

Isabel Cafezeiro, Leonardo da Costa, UNIVERSIDADE FEDERAL FLUMINENSE, Brazil.

O que explica o sucesso de jogos no ensino de engenharia de software? Uma teoria de motivação

Marcia Souza, CESAR.EDU, Brazil

Cesar França, Federal Rural University of Pernambuco, Brazil.

O uso de recursos lúdicos para o ensino de processos em engenharia de software

Ronney Moreira de Castro, Faculdade Metodista Granbery, Brazil

Gleison Santos, Unirio, Brazil.

O uso de robótica para aprendizado de programação integrando alunos de educação básica e ensino superior

Luemy Avila, Secretaria de Educação de Macaé, Brazil

Flavia Bernardini, Patrick Moratori, Universidade Federal Fluminense, Brazil.

Oficina do código: um projeto para o ensino e integração de alunos do ensino fundamental e médio na área de tecnologia

Stéphanie Alencar Braga, IFCE, Brazil

Caio Raveli Freitas Barbosa, Instituto Federal de Educação, Ensino e Tecnologia do Ceará, Brazil

Priscylla da Silva Tavares, Instituto Federal de Educação, Ciência e Tecnologia do Ceará, Brazil

Mikhail Jose Pires Pedrosa de Oliveira, Instituto Nacional de Pesquisas Espaciais (INPE), Brazil.

PBL e robótica no ensino de conceitos de lógica de programação

Flavia Nogueira e Silva, Faculdade de Tecnologia - Unicamp, Brazil

Marcos Borges, Universidade Estadual de Campinas, Brazil.

Pensamento computacional praticado com um jogo casual sério no ensino superior

Adilson Vahldick, Universidade do Estado de Santa Catarina, Brazil

António José Mendes, Maria José Marcelino, Universidade de Coimbra, Portugal

Paulo Farah, Universidade do Estado de Santa Catarina, Brazil.

PERTool Uma ferramenta educacional para manipulação de redes PERT/CPM

João Darcy, Diego Bevilaqua, Flaviane Vicente, Paulo Henrique, Universidade Federal de Goiás/Regional Jataí, Brazil

Heitor Costa, Paulo Parreira Júnior, Universidade Federal de Lavras, Brazil.

Plataforma arduino como apoio ao ensino de programação no curso de técnico em informática integrado

Otávio Soares Papparidis, Matheus Eloy Franco, IFSULDEMINAS Câmpus Machado, Brazil.

POOGame: um jogo sério para o ensino de programação orientada a objetos

Leuson Da Silva, Wellington Moura, Rogério Carvalho, Jefferson Barbosa, Brendo Colaço, Carla Ilane Moreira Bezerra, Paulyne Jucá, Universidade Federal do Ceará, Brazil.

PortuCol: uma pseudolinguagem inspirada em C ANSI para o ensino de lógica de programação e algoritmos

Lucas Lemos Barbosa, Christian Marlon Souza Couto, Ricardo Terra, UFLA, Brazil.

Práticas interdisciplinares apoiadas por problemas na formação do técnico em informática

Olavo José Luiz Junior, Instituto Federal do Paraná, Brazil

Luiz Fernando Delboni Lomba, Instituto Federal de Mato Grosso do Sul, Brazil.

Processo de desenvolvimento de software: uma atividade prática supervisionada baseada nos papéis de adquirente e fornecedor

Andre Ortoncelli, Marisângela Pacheco Brittes, UTFPR, Brazil.

Recursos computacionais para suporte ao ensino de teoria da computação, linguagens formais e autômatos

Ícaro Souza, Ecivaldo Matos, Débora Abdalla, Ranansamir da Silva, Universidade Federal da Bahia - UFBA, Brazil.

Sistema de apoio à avaliação de atividades de programação por reconhecimento automático de modelos de soluções

Marcia Oliveira, Instituto Federal do Espírito Santo, Brazil

Leonardo Reblin, Universidade Federal do Espírito Santo, Brazil

Elias Oliveira, UFES, Brazil.

The teaching of functions as the first step to learn imperative programming

Carla Delgado, Universidade Federal do Rio de Janeiro, Brazil

Joao Carlos da Silva, UFRJ, Brazil

Ana Luisa Duboc, Colégio Pedro II, Brazil

Fabio Mascarenhas, PUC-Rio, Brazil.

Torneios baseados em robocode para incentivar jovens a aprender programação

Matheus Meira, Marri Lima, Marcos Borges, UNICAMP - Universidade Estadual de Campinas, Brazil.

Uma análise da evasão escolar nos cursos de tecnologia da informação: um estudo de caso em Floresta/PE

Everton Silva, Instituto Federal de Educação Ciência e Tecnologia do Sertão Pernambucano, Brazil

Josilaine Souza Silva, Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, Brazil

Cassiano Henrique de Albuquerque, Instituto Federal de Educação Ciência e Tecnologia do Sertão Pernambucano, Brazil.

Uma ferramenta para ensinar introdução à teoria da computação

Candido Ferreira Xavier Mendonça Neto, Luciano Digiampietri, Universidade de São Paulo, Brazil.

Uma plataforma ubíqua e pervasiva para distribuição de conteúdo em ambientes educacionais

Laura Assis, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ, Brazil

Andre Felipe Monteiro, CEFET/RJ Campus Petrópolis, Brazil

Dalbert Mascarenhas, CEFET-RJ, Brazil.

Utilização do mastery learning em sala de aula: uma abordagem integrada aos mundos virtuais

Felipe Becker Nunes, Fabrício Herpich, Manuel Zunguze, José Valdeni De Lima, Universidade Federal do Rio Grande do Sul (UFRGS), Brazil.

A Game Development-based strategy for Teaching Software Design Patterns through Challenge-Based Learning under a Flipped Classroom approach

Ismar Frango Silveira

Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie
Rua da Consolação, 920 – 01302-907 – São Paulo – SP – Brazil

ismar@mackenzie.br

***Abstract.** Software Design Patterns, namely those belonging to GoF catalog, are gradually being incorporated into undergraduate Computer Sciences-related curricula worldwide. However, this is a subject that often requires students to have some maturity on Software Development issues, which is a scenario that is barely found on actual classrooms. Besides, the comprehensive taught of many patterns' aspects depends upon professors' expertise on a large range of Software Engineering and Object Orientation-related topics. Both aspects are possibly caused by the complexity and abstract nature of the subject, whose concrete, real-world applications not always are successfully reached in an undergraduate course. In this sense, this paper presents a strategy for Teaching GoF Design Patterns through a Game Development Process, under a Flipped Classroom pedagogical context. The results of some case studies are also shown, taken from two classes where Challenge-based Learning techniques were applied.*

1. Introduction

Design Patterns are a better way to think about software. Since they encourage modularity and reuse, they would naturally support and be supported by the principles of Object-Oriented paradigm. In this sense, it sounds natural that such a subject would be a must have on undergraduate curricula for courses in Computer Sciences (CS) and similar ones. However, the teaching and learning of this subject in undergraduate levels has been theme for academic discussion by more than one decade. Since the effective popularity of some pattern proposals— more specifically the GoF catalog [Gamma et al, 1995], many teaching strategies have been presented in the literature to deal with the necessity of learning a highly complex subject like Software Design Patterns in an undergraduate course of Computer Science and correlated ones.

Few years ago, the teaching of Design Patterns still played a discrete role in some Software Engineering syllabi, being barely visited in an undergraduate course and led to be covered in elective courses or at a intermediate level, as already defended by Johnson and Barnes (2005). Recently, this subject has been absorbed by CS-related undergraduate careers, in a movement that appears to be a primary need of business contractors, which have

been pushing Universities to include Design Pattern in their curriculum as soon as possible, giving the high probability of students having to pursue such knowledge in their internships.

This discussion, in fact, is not a new one: Pecinovsky et al. (2006) already proposed a Design Pattern approach for teaching Object Orientation good principles as early as possible. Their arguments rely on the fact that letting this subject to be learned too late would “calcify bad programming habits in students”. Clearly, early adoption of Design Patterns on curricula would require more suitable approaches for younger publics.

Ludic elements have been used by some authors as strategic components for teaching Design Patterns along last years: Dukovich (2008) presents a ludic proposal of showing Design Patterns through multimedia examples and many other authors rely on games to implement their strategies. For instance, Mustaro et al. (2009) present an experience of reverse engineering of games to teach principles of Design Patterns in a undergraduate level; Wick (2005) proposes a sequence of laboratory activities using the Game of Life – an approach similar to Gómez-Martín et al. (2009), who used a family of games instead; on the other hand, Gestwicki and Fu (2008) proposed to teach Design Patterns through the development of a game, strategy that is being used in the context of this paper.

Besides the ludic application to motivate students, some specific teaching strategies were chosen for this experiment, namely the Flipped Classroom and Challenge-Based Learning (CBL). These strategies include cognitive mobilization activities that can be undertaken through the support of technological resources, which includes videos, digital games and simulations, among other elements [Davies et al., 2013

Often simplistically defined as “homework done at school and school tasks done at home” the set of techniques that make up what is now known widely as Flipped Classroom comprises a set of strategies and diverse approaches. Bishop and Verleger (2013) bring a comprehensive review of the literature, which will serve as a guiding principle for the experiment presented here.

Together with Flipped Classroom, it was adopted a CBL (Challenge-Based Learning) approach. O’Mahony et al. (2012) showed that this strategy has been providing greater learning performance gains over a lecture format As in Baloian et al. (2013), the CBL approach here presented consists on an extended form of PBL (Problem-Based Learning), in which the problems are of realistic, open-ended nature, leading to more experiential, project-based and decision-based learning perspectives.

Both belong to the context of Active Learning. Methodologies, whose fundamentals were mapped by an early paper of Prince (2004); a more recent, despite not up-to-date review can be found in Settles (2010). Felder and Brent (2009), define them under a broader approach, categorizing as part of Active learning as all that is related to a course that students are encouraged to do than just watch, listen and annotate. In this sense, such methodologies would embrace a wide spectrum of pedagogical techniques that are mainly based on the introduction of activities to be performed by students in the context of traditional education and the promotion of student engagement in the learning process, as opposed to teaching strategies that do not require students to play an active role in the process, considering them passive

receptors of pieces of information, not always connected among themselves or even related to students’ cognitive framework.

Under these perspectives, next section will show the pedagogical framework being proposed for teaching Design Patterns.

2. The SuperRoamer Game

The experiment consists on the following steps: a GDD (Game Design Document) about a game proposal, here named “The SuperRoamer Game” is to be presented to students in a gradual manner, along four weeks. Students are meant to understand the requirements of each week and construct a class diagram that gradually could model these requirements. Depending on the context, an implementation could also be considered.

All main principles of flipped classroom and CBL are to be applied in all classroom encounters during this experiment, which must never rely on explanatory, pre-expositive classroom moments.

2.1 Week 1: Creational GoF Patterns

In Week 1, the basic game mechanics are stated: SuperRoamer has to go through two jungles: the Amazon Rainforest in Brazil and the African Savannah. He must try to get as many diamonds and pots of gold, besides some jackfruits and *açaí* berries to maintain his energy. He should take care of some beasts scattered along the jungle, which can hurt or even kill him, depending on his accumulated energy.

Table 1 shows the main game requirements for the first week, together with an icon to represent the situation in GDD and the GoF Pattern that is expected to be discovered and applied by students to fulfill the requirement.

Table 1. Creational GoF Patterns



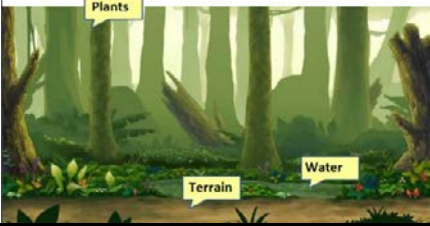



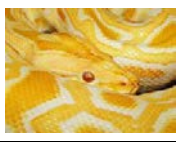
GDD Icons	Game requirement	Expected Pattern
	There is one and only one SuperRoamer along the entire game	Singleton
	There are four collectable items, and the pots of gold and diamonds represent wealth. The consumption of one jackfruit gives two minutes of energy to the player, whilst another item to be sought is the <i>açaí</i> , which should be consumed in great number. Each <i>açaí</i> berry gives 5 second of power.	Factory Method and Prototype.

Table 1, continued. Creational GoF Patterns

		<p>The levels themselves must be constructed following a pre-determined order and combination of elements like terrain, water and plants.</p>	<p>Builder</p>
<p><i>Amazon Rainforest</i></p>	<p><i>African Savannah</i></p>	<p>Depending on the jungle, there will be different animals from three types: snakes, raptors and felines. For the Amazon Forest, we have Anacondas, Caracaras and Jaguars. For the African Savannah, Pythons, Martial Eagles and Tigers. The attack of a raptor takes 30 seconds of power; the attack of a snake, 1 minute; the attack of a feline 5 minutes.</p>	<p>Abstract Factory</p>
			
			

Clearly, all requirements are about situations that required the creation some object. Singleton is the most straightforward pattern, followed by two factory patterns, namely, Factory Method and Abstract Factory. Students must identify two different situations for creating objects, being able to identify Abstract Factory as the most adequate for creating families of objects, instead of Factory Method. As an example, Figure 1 shows an extract of a class diagram made by students that correctly models the last requirement using the Abstract Factory pattern. All classes' names, attributes and methods are in Portuguese.

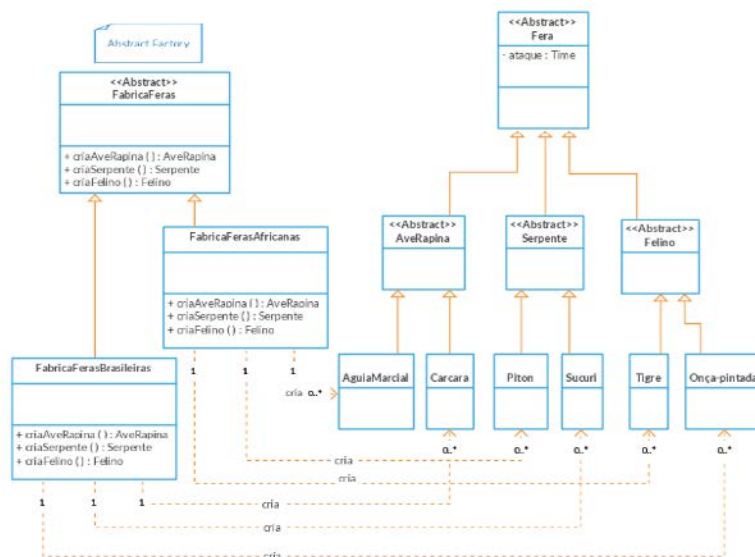


Figure 1. Abstract Pattern applied to Week 1's fourth requirement





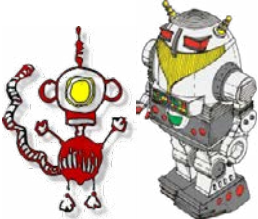
Considerations about Singleton being an anti-pattern and preoccupations about “class explosion” in the context of Factory Method could be explored here by the professor. The natural complexity of Builder pattern could also be explored in this application, as well as considering the use of Prototype to model *açaí* berries because of their expressive quantities (and not for their complexity) could lead to some interesting discussions about performance of dynamic instantiation versus cloning.

It must be noted that no suggestion of pattern should be previously given to students, since the reflection and discovery of which pattern would fit better in each situation is necessary to be taken by students.

2.2 Week 2: Structural GoF Patterns

Activities in Week 2 start by discussing the results of last week and reinforcing theoretically the main concepts behind the five creational patterns. Thus, new requirements could be presented, in which game mechanics slightly change in order to provoke some reflection about the structural aspects of game project. Table 2 shows these new requirements.

Table 2. Structural GoF Patterns

GDD Icons	Game requirement	Expected Pattern
	Users could choose between a male or female avatar.	Bridge
	Every avatar could have any combination of the following items (or not, depends on the user's choice): Cover, Mask, Gloves, Glasses and a magic Utility Belt	Decorator
	The Magic utility belt is an element apart: it has objects and countless pockets; each pocket may contain other objects or even other pockets within them and so on.	Composite
	Many objects in the belt are weapons or shields that can be used to defend or face the challenges of the jungle, as the beasts. There are, however, two types of objects: ordinary objects and magical objects. Common objects can be handled directly by the character to attack or defend. However, magical objects have to be handled by means of wearing magic gloves, which allow the use of objects making the magic affecting less the character.	Adapter
	When the character enters the game, he wins two "beaters", which are robots-emissaries who go ahead to check the operating conditions. There are Water Emissaries to see if the character can follow the jungle rivers and Land Emissaries, who walk certain dry part of the jungle to predict the presence of beasts. With the evolution of the game, the character can win or lose emissaries.	Proxy

As an example, Figure 2 shows a diagram (in Portuguese), made by students, using Decorator to model the requirement related to Cover, Mask, Gloves, Glasses and Belt.

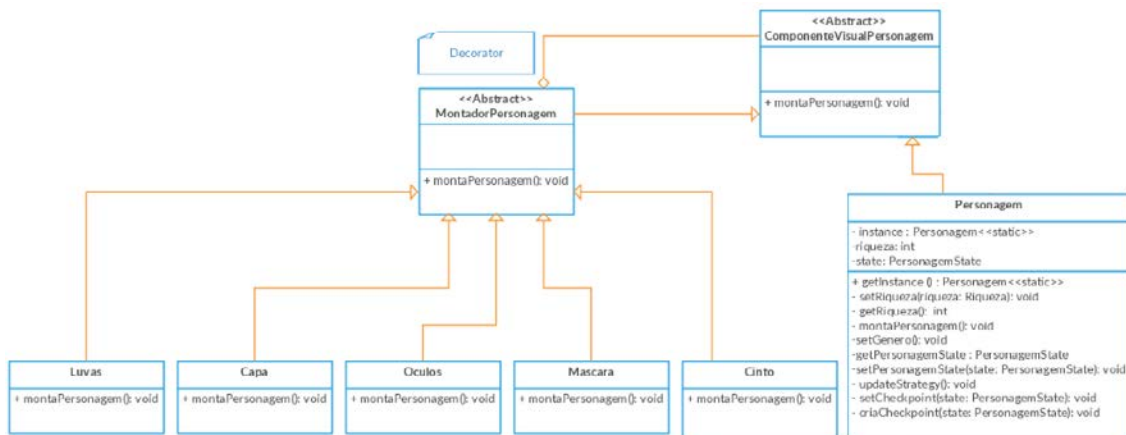


Figure 2. Decorator Pattern applied to Week 2's second requirement

The application of many of the structural patterns does not appear to be so straightforward as the creational ones. While the requirements for Decorator and Composite – and maybe Bridge – are less complex to understand and discover, Proxy and Adapter are two patterns that, in spite of being widely used, requires a higher degree of abstraction. In this point, professor could intervene and bring some “real-world” examples for these two patterns, like the use of stubs (as Proxies) and skeletons (as Adapters) in many distributed computing standards, like CORBA and RMI. Depending on the maturity of the class, more details could be given about different implementations of Adapter (based on multiple inheritance versus delegation) and variations of Proxy (Virtual, Remote, Protection, Smart) [Geary, 2002].

It must be noted that two structural patterns were not covered by these requirements: Flyweight and Façade. It could be asked to students for creating new requirements for showing the application of these patterns.

2.3 Week 3: Behavioral GoF Patterns

As in last week, the discussions about structural pattern must precede the activities of Week 3, which deals with the most complex part of the catalog, the behavioral patterns. Besides being more numerous – there are eleven of these patterns, their understanding relies on a higher capacity of abstraction. For this week, only 3 new requirements are to be presented, as shown in Table 3 and 4.

Table 3. Structural GoF Patterns


GDD Icons	Game requirement	Expected Pattern
	<p>The belt has a number of useful elements. Some of them are specific to defend the attacks of wild beasts:</p> <ul style="list-style-type: none"> • Magic Sword: it must be drawn when the character is being attacked by a snake; • Giant cage: it must be used to trap a raptor; • Hose: used to throw water on felines 	Strategy

Figure 3 shows a students' application of Strategy Pattern to meet this requirement (diagram in Portuguese).

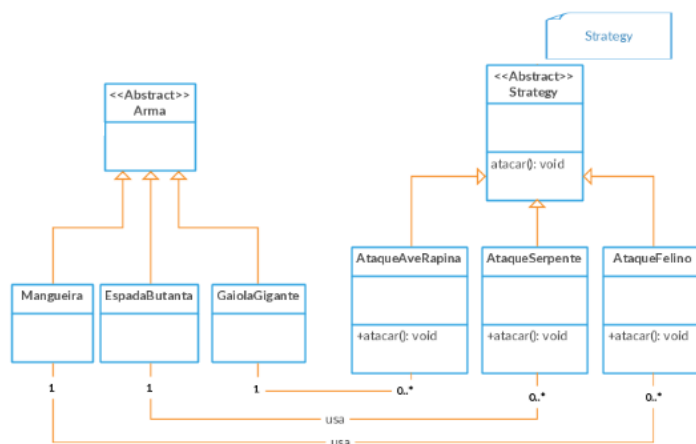


Figure 3. Strategy Pattern applied to Week 3's second requirement

Two additional requirements were presented to students, as shown in Table 4.

Table 4. Behavioral GoF Patterns (continued)

	<p>All items on the belt to fight the beasts have their magical versions. However, it is recommended to use these magic versions only if the character is in a healthy state, namely more than 30 seconds of remaining power. As the use of magic items, even with gloves, consumes some lifetime of the character, if the character is in a moribund state, (the remaining lifetime of the character is less than 30s), the magic weapon can destroy the character even before attacking the beast. In this case (moribund state), normal weapons should be used.</p>	<p>State</p>
	<p>When walking in a jungle, the character can spread three types of electronic organs (the "watchers" - which are somewhere in the belt) along the environment. These watchers are connected (by long-distance brain vibrations) to the character and warn him of the following events:</p> <p>Ear watchers: they warn when there is a suspicious noise in the woods, which can be a birdsong, or the sound snake approaching (felines use to be more silent);</p> <p>Eye watches: they warn when approaching any of the beasts of their field of vision (they are not very useful against snakes);</p> <p>Nose watchers: they warn when they feel a sharp odor in the jungle (felines and raptors usually exhale very strong smells).</p>	<p>Observer</p>

At this point, the instructor could suggest students to create up to eight situations where behavioral patterns could be applied. Some of these patterns would require extremely specific situations (like Interpreter, for instance), but depending on classroom's maturity, even

complex behavioral patterns like Visitor, could be covered (in this case, the combination State/Strategy could be replaced by a Visitor). Experience has shown that students could easily find applications for those more straightforward patterns, like Memento. Some other patterns, like Template Method, Iterator, Mediator and Chain of Responsibility would require a more direct instructor intervention.

2.4 Week 4: Final Pitch and evaluation

In the last week, it would be advisable to discuss the results each student have reached. A “pitch” – a brief presentation of the project – could be scheduled, which would motivate students to discuss their result with colleagues, which would allow them to learn from others in a collaborative way.

3. Experimentation and Analysis

An experimental application of the technique related in this paper was performed with two classes of fourth-semester undergraduate students of a Brazilian University. The study comprised 42 students that had previous contact with only some very basic GoF Design Patterns, as part of a previous course they have taken.

The methodological approach was based on the concept of Perception of Learning [Kuhn and Rundle-Thiele, 2009]. Since it follows qualitative basis, no procedure involving control groups could be taken, so that the same experiments were taken with both classes. Since they had theoretically the same formal pre-requisites, they were considered as a unique group for analysis.

The experiment took four weeks, as described before. A survey was performed in order to try to capture students’ self-awareness of Learning. Using a five-tiered, labeled Likert scale, it was asked to students to self-evaluate their learning of each one of the 23 patterns. Figure 4 shows the results for the creational ones (values in %).

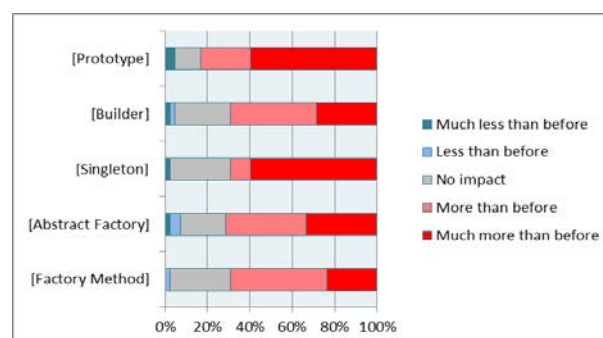


Figure 4. Students’ self-evaluation of Perception of Learning of Creational Patterns

It must be noted that all patterns showed a positive perception of learning of more than 60% (the sum of those who answered that now, they perceive that they know “more than before” and “much more than before”). The Prototype pattern was the best learned, according to students’ perception. The Singleton and Factory Method patterns presents a very good perception of learning, but also presents a considerable amount of neutral responses (those who answered “no impact”), which denotes that part of students already have built the

correspondent knowledge at the previous contact with these patterns, which was before mentioned.

Figure 5 shows the results for the structural and behavioral patterns.

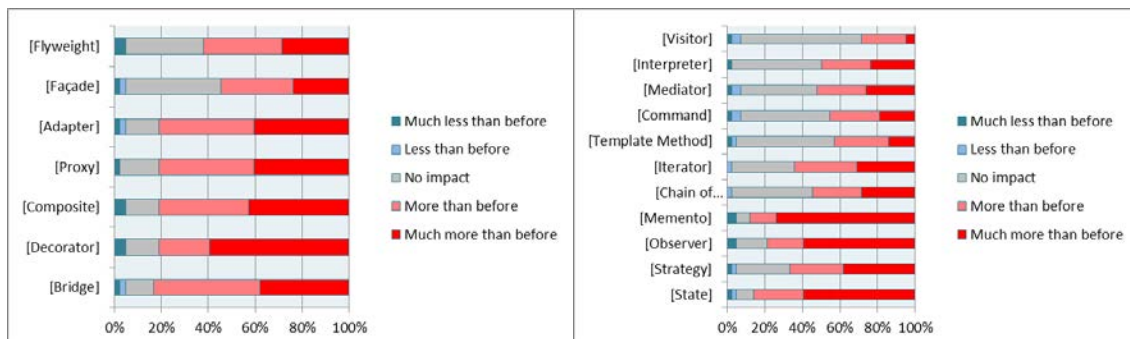


Figure 5. Perception of Learning of Structural (left) and Behavioral (right) Patterns

As expected, Flyweight and Façade were the worst learned structural patterns, since no game requirements were elaborated for them. All the others had excellent perceptions of learning (around 80%), being the Decorator pattern the most understood by students. For behavioral ones, the same scenario is perceived: the three patterns that were covered by game requirements (Strategy, State and Observer) had good perceptions of learning, and the same (or even better) with Memento, which was discovered by students themselves. Iterator and Chain of Responsibility (both covered by situations created by students) had also good values. All other patterns had no good perceptions, in special Visitor, which was considered by students as the worst one.

4. Conclusions and further work

This paper presented a teaching strategy based on the fundamentals of Flipped Classroom and CBL. At each encounter, students were motivated to discuss the diagram they prepared along the past week, with instructor presenting some theoretical discussions about that family of patterns. If an expositive, traditional approach were chosen, probably the contact of students with this abstract subject barely would be converted into meaningful learning. A game was used as the object of study, but eventually any kind of software project could be proposed, according to the audience.

Further works will deal with some improvements on this technique, in order to better cover the behavioral patterns, probably extending the numbers of weeks and augmenting the proposed requirements in order to cover more patterns. However, the stimulus for students to create new requirements is important for developing their creativity. New applications of this technique are currently being performed, together with a variation of it, applied to Architectural Patterns, whose results will be shown in other occasion.

References

Baloian, N.; Breuer, H.; Hoeksema, K.; Hoppe, U.; Milrad, M. (2004) Implementing the Challenge Based Learning in Classroom Scenarios. Proceedings of the Symposium on Advanced Technologies in Education. Argostoli, Greece.

- Davies, R. S.; Dean, D. L.; Ball, N. (2013) Flipping the classroom and instructional technology integration in a college-level information systems spreadsheet course. In: *Educational Technology Research and Development*, v. 61, 4, p. 563-580.
- Dukovitch, A. (2008) Design Patterns go to Hollywood: teaching Patterns with Multimedia. MSc. Thesis, Faculty of California Polytechnic State University, San Luis Obispo, USA.
- Gamma, E. et al. (1995) Design patterns: Elements of reusable object-oriented software.. Reading, Massachusetts, USA: Addison-Wesley, 1995.
- Geary, D. (2002) Take control with the Proxy design pattern. Available on the Internet: <http://www.javaworld.com/article/2074068/learn-java/take-control-with-the-proxy-design-pattern.html?page=2> retrieved March 19, 2016.
- Gestwicki, P; Sun, F. S. (2008) Teaching Design Patterns through Computer Game Development. *Journal on Educational Resources in Computing*, vol. 8 (1), article 2.
- Gómez-Martín, M. A.; Jiménez-Díaz, G.; Arroyo, J. (2008) Teaching design patterns using a family of games. *ACM SIGCSE Bulletin*, 08/25/2009, vol. 41(3), 268.
- Johnson, C. W.; Barnes, I. (2005) Resigning the intermediate course in software design. In *ACE '05: Proceedings of the 7th Australasian conference on Computing education*, pages 249-258, Darlinghurst, Australia.
- Kuhn, K. A.; Rundle-Thiele, S. R. (2009) Curriculum Alignment: Exploring Student Perception of Learning Achievement Measures. *International Journal of Teaching and Learning in Higher Education*, vol. 21 (3), 351-361.
- Mustaro, P. N.; Silva, L.; Silveira, I. F. (2009). Using Games to Teach Design Patterns and Computer Graphics. In R. Ferdig (Ed.), *Handbook of Research on Effective Electronic Gaming in Education*, 525-545. Hershey, PA: Information Science Reference. doi:10.4018/978-1-59904-808-6.ch030.
- O'Mahoney, T. K.; Vye, N. J.; Bransford, J.; Sanders, E. A.; Stevens, R.; Stephens, S. D.; Richey, M. C.; Lin, K. Y.; Soleiman, M. K. (2012) A Comparison of Lecture-Based and Challenge-Based Learning in a Workplace Setting: Course Designs, Patterns of Interactivity, and Learning Outcomes. *The Journal of the Learning Sciences*, vol. 21,1, p182- 206
- Pecinovský,R; Pavlíčková, J.; Pavlíček, L. (2006). Let's modify the objects-first approach into design-patterns-first. *SIGCSE Bull.*, 38(3):188-192, 2006
- Prince, M. (2004) Does Active Learning Work? A Review of the Research. (2004) In: *Journal of Engineering Education*. In: *Journal of Engineering Education*, v.93, 3, p. 223–231
- Settles, B. (2010) Active learning literature survey. Computer Sciences Technical Report, 1648. University of Wisconsin-Madison.
- Wick, M. R. (2005) Teaching design patterns in CS1: a closed laboratory sequence based on the game of life. *Proceedings of the 36th SIGCSE technical symposium on Computer science education*, 487-491.

Avaliação de Ferramentas de Apoio ao Ensino de Técnicas de Mineração de Dados em Cursos de Graduação

José Viterbo¹, Clodis Boscarioli², Flavia Cristina Bernardini³,
Mateus Felipe Teixeira²

¹Instituto de Computação
Universidade Federal Fluminense (UFF)
Niterói – RJ – Brasil

²Colegiado de Ciência da Computação
Universidade Estadual do Oeste do Paraná (UNIOESTE)
Cascavel – PR – Brasil

³Instituto de Ciência e Tecnologia
Universidade Federal Fluminense
Rio das Ostras – RJ – Brasil

viterbo@ic.uff.br, clodis.boscarioli@unioeste.br,
fcbernardini@id.uff.br, mateusteixeira_@hotmail.com

Abstract. *The labor market requires more and more professionals with analytical sense to handle large data sets to produce useful knowledge for decision making. The data mining process (DM) is among the key concepts of knowledge discovery and involves the use of different machine learning algorithms applied to extract new standards in databases, commonly supported by tools that implement these different algorithms. The use of DM tools by undergraduate students is essential for them to acquire practical experience. It is done in courses such as Artificial Intelligence, Data Mining or other, addressing machine learning techniques. This study evaluates different tools used in DM teaching, in order to offer teachers a guide in the selection and use of these tools from the point of view of usability, measured by undergraduate students in the discovery and understanding of associated knowledge.*

Resumo. *O mercado de trabalho requer, cada vez mais, profissionais com senso analítico para lidar com volumosos conjuntos de dados de forma a produzir conhecimento útil à tomada de decisão. O processo de Mineração de Dados (MD) está entre os principais conceitos de descoberta de conhecimento e envolve o uso de diferentes algoritmos de aprendizado de máquina, aplicados para extrair novos padrões em bases de dados, comumente apoiado por ferramentas que implementam esses diferentes algoritmos. O uso de ferramentas para MD por alunos de graduação em disciplinas como Inteligência Artificial, Mineração de Dados ou outras que abordem técnicas de aprendizado de máquina, é fundamental para que eles adquiram experiência prática. Este trabalho avalia diferentes ferramentas utilizadas no ensino de MD, no intuito de oferecer aos docentes um guia na escolha e uso*

dessas ferramentas, do ponto de vista da usabilidade, aferida por alunos de graduação na descoberta e compreensão dos saberes associados.

1. Introdução

O mundo dos negócios vem passando por uma revolução impulsionada pelo uso analítico dos dados para orientar a tomada de decisões. O processo de análise eficiente de grandes massas de dados exige esforço multidisciplinar, uma vez que há desafios em diversas áreas que devem ser tratados. Neste cenário, surgem novas oportunidades de empregos, contudo, encontrar profissional qualificado à área de análise de dados passa a ser um grande desafio. A ciência de dados é um campo emergente e o cientista de dados é um profissional que precisa ter raciocínio lógico apurado, conhecimento profundo de estatística e modelagem matemática, além de domínio de sistemas computacionais e expertise do negócio e do mercado em que atuará [BRETERNITZ; SILVA 2013].

Dentre os conhecimentos fundamentais que esse profissional deve ter, se destacam diferentes tarefas e técnicas para apoiar as diversas etapas do processo de Mineração de Dados (MD) [REZENDE *et al.* 2005], [WITTEN 2011]. A MD é amparada por ferramentas que permitem a aplicação dos mais diversos algoritmos de aprendizado de máquina na análise de dados, sem que o usuário tenha que implementá-los. Nesse contexto, há a oportunidade de transformar o papel acadêmico de programador de algoritmos de mineração de dados para analista de inteligência de negócios, a partir do uso de ferramentas que deem suporte ao processo de descoberta de conhecimento em bases de dados, tornando-os capazes de analisar diferentes conjuntos de dados e de produzir relatórios que possam auxiliar na tomada de decisões.

Por outro lado, a MD apresenta novos desafios aos docentes que lecionam disciplinas nesta área. Um deles é a identificação de qual ferramenta computacional usar para introduzir o assunto em um curso de graduação. É importante que alunos de graduação em disciplinas como Inteligência Artificial, Aprendizado de Máquina, Mineração de Dados – ou outras cujo foco inclua suporte analítico para tomada de decisão, tomem contato prático com essas ferramentas, para vivenciar, de forma antecipada, processos de MD durante sua formação acadêmica.

Neste trabalho, investigamos como a usabilidade e a experiência do usuário com tais ferramentas pode melhorar a compreensão do processo de MD e do conhecimento descoberto e assim, melhorar a qualificação desse profissional. Foram avaliadas quatro ferramentas: Knime, Orange Canvas, Rapidminer Studio e Weka, e o resultado dessa avaliação pode auxiliar como um guia na escolha de ferramentas para ensino de MD.

A próxima seção traz os conceitos básicos relacionados à MD, de forma a contextualizar o leitor sobre o estudo realizado. Na Seção 3, são apresentados trabalhos relacionados a este estudo. Na Seção 4, são descritas as ferramentas avaliadas neste estudo. Na Seção 5, é discutida a metodologia empregada na pesquisa. Na Seção 6, é apresentada uma análise dos resultados obtidos com a avaliação das ferramentas de MD por estudantes de graduação em Computação. E, por fim, na Seção 7 são apresentadas as conclusões e perspectivas da pesquisa.

2. Mineração de Dados

O objetivo de um processo de MD é extrair conhecimento a partir de informações “ocultas” nos dados, que sejam úteis nas tomadas de decisões, utilizando métodos de diferentes áreas científicas, como Estatística, Banco de Dados, Inteligência Artificial, Aprendizado de Máquina e Reconhecimento de Padrões, caracterizando esse processo como interdisciplinar [WITTEN 2011]. As etapas desse processo são compostas pelo pré-processamento, construção de padrões e modelos e pós-processamento de dados, que envolvem, respectivamente, o tratamento dos dados de forma a qualificá-los para a entrada dos algoritmos de mineração de dados, a execução dos algoritmos para extração de padrões, e a compreensão das saídas para geração de conhecimento.

Na MD, a principal atividade – iterativa e interativa – envolve algoritmos de extração de padrões a partir da definição de uma tarefa de análise. As tarefas podem ser preditivas ou descritivas. As tarefas preditivas consistem na generalização de exemplos ou experiências passadas com respostas conhecidas, relacionadas às classes dos exemplos. Se a classe dos exemplos pertence a um conjunto discreto e finito de valores, a tarefa é denominada classificação; e se pertence a um conjunto numérico de valores, é denominada regressão. Já as tarefas descritivas consistem na identificação de comportamentos intrínsecos do conjunto de dados, sem que haja classes definidas dos exemplos. As tarefas descritivas podem ser de construção de regras de associação ou de agrupamento de dados [REZENDE *et al.* 2005], [WITTEN 2011].

Considerando que o processo de MD é interdisciplinar para compreender os algoritmos de aprendizado, seus ajustes de parâmetros e para avaliar os modelos construídos, ferramentas que possam auxiliar no entendimento de todo esse processo podem ser adotadas no processo de ensino e aprendizagem. Existem várias ferramentas disponíveis. Algumas delas são bastante difundidas e adotadas por docentes que ministram disciplinas correlatas a Mineração de Dados, por serem as mais antigas, a exemplo da Weka. Tal fato foi observado empiricamente pelos autores com diversos colegas docentes e observando ementas de diferentes disciplinas disponíveis na web. Essa adoção, porém, se dá mais por tradição que por uma decisão pautada em estudos de interação, preocupados com descobrir qual dessas ferramentas é a mais adequada para fins didáticos. O foco deste estudo parte da premissa que o uso de ferramentas de computação para ensinar disciplinas de computação deve ser pautado na avaliação de aspectos de interação com o usuário final, no caso em questão, aluno de graduação.

3. Trabalhos Relacionados

Vários são os autores que têm relatado a maneira que diferentes conteúdos de Inteligência Artificial e de Mineração de Dados têm sido abordados em disciplinas de graduação, bem como a importância dos conteúdos na formação. Em [FERREIRA *et al.* 2012] os autores analisam a abordagem de diferentes conceitos de Inteligência Artificial, incluindo conceitos aplicáveis à MD, apontando a importância em abordar tais conteúdos devido ao crescente uso de ferramentas de mineração de dados. Há uma análise da abordagem baseada em ferramentas para o ensino do processo de MD em [JAFAR 2010], que recomendam fortemente estender o currículo de cursos de Sistemas de Informação para incluir o estudo de MD em pelo menos uma disciplina. Para esses autores, as ferramentas permitem concentrar-se em ensinar os aspectos analíticos e o uso de algoritmos por meio de práticas, e concluem que os alunos ganham em

compreensão conceitual com essa abordagem. Em [MANASEER; MALIBARI 2012] também é reforçada a questão da prática, e apresentada uma proposta para o ensino de MD, permitindo aos alunos selecionar uma aplicação em dados reais. Esses autores afirmam ainda que a maioria da informação teórica ensinada é perdida após a graduação se não estiver relacionada a aspectos práticos e aplicações. As aulas práticas em cursos da área de Computação têm neste contexto, portanto, fundamental importância.

Em relação ao uso de ferramentas, em [KING e SATYANARAYANA 2013] são discutidas questões relevantes no ensino de Mineração de Dados em cursos de graduação, incluindo os principais tópicos que uma disciplina deve abordar. Os autores recomendam o uso de ferramentas de mineração de dados tradicionais (Weka, R e Rattle, sendo Rattle uma interface gráfica para uso da ferramenta R), por conterem uma ampla oferta de métodos para pré-processamento e extração de padrões. Em [CHAWLA 2005] é reportada a experiência do ensino de MD utilizando a ferramenta Weka, com destaque à importância da inserção dessa disciplina nos currículos de graduação. Os alunos do curso oferecido afirmaram que as aplicações práticas foram significativas no processo de aprendizado. Em [SATYANARAYANA 2013] o autor compara o uso de três ferramentas proprietárias – SAS, IBM SPSS Modeler e Matlab – e três ferramentas de código aberto – R, Weka e RapidMiner – no ensino de MD e conclui que, em cursos introdutórios de MD, o uso de ferramentas de código aberto é indicado, e em cursos avançados, o uso de ferramentas proprietárias pode ser também considerado.

É interessante observar que os diversos autores mencionam o uso de ferramentas, abertas ou não. No entanto, nenhum deles avalia, como neste trabalho, a questão da usabilidade dessas ferramentas no contexto da aprendizagem.

4. Ferramentas Avaliadas

A escolha das ferramentas analisadas, abaixo descritas, deu-se não apenas pelo fato de serem bastante conhecidas na área, mas também por terem uma variedade de funcionalidades em comum e possuírem versões de acesso livre, conforme já abordado em [BOSCARIOLI *et al.* 2015].

Knime [KNIME 2013]: ferramenta para uso na MD, Estatística e outras áreas. Possui diversos métodos que possibilitam extração de conhecimento de uma determinada base de dados. O funcionamento da Knime é todo baseado na ideia de adição de nodos de métodos a um fluxo de execução. A ferramenta apresenta algumas funcionalidades ainda não detalhadas, o que pode dificultar o seu uso.

Orange Canvas [CANVAS 2013]: ferramenta de código aberto de MD para classificação, regressão e tarefas descritivas de dados, e para mineração visual de dados. O fluxo de execução da ferramenta é dado por uma estrutura de nodos, em que cada nodo adicionado ao campo de fluxo executa uma determinada tarefa. Também apresenta um sistema de retorno ao usuário (*feedback*) para cada método, retornando-lhe os dados de entrada e de saída de cada método.

Rapidminer Studio [RAPIDMINER 2013]: ferramenta proprietária voltada à Estatística, Banco de Dados e processos de análises de dados que possui também uma versão para testes. Tem como foco disponibilizar ambiente de trabalho com presença de elementos gráficos que significam uma operação em questão, por exemplo, um método de MD. A ferramenta possui um fluxo de execução intuitivo e baseado em nodos que representam um determinado processo.

Weka [HALL *et al.* 2009]: ferramenta de aprendizado de máquina aberta usada para tarefas de MD, que contém técnicas para pré-processamento de dados, classificação, regressão, agrupamento, regras de associação e visualização.

5. Metodologia

Na avaliação das ferramentas, primeiramente realizou-se uma inspeção por meio do Percurso Cognitivo, um método de avaliação de IHC (Interação Humano-Computador) por inspeção cujo principal objetivo é avaliar a facilidade de aprendizado de um sistema interativo por meio da exploração de sua interface [WHARTON 1994]. Nesta abordagem, não há a necessidade da presença de usuários, e o(s) avaliador(es) devem associar objetivos específicos às ações e identificar problemas de interação de diferentes tipos de usuários e contextos reais.

Embora os percursos para realizar as ações sejam diferentes em cada ferramenta, pode-se concluir que o usuário vai tentar atingir o resultado correto desde que saiba o funcionamento teórico dos métodos implementados para a realização de ajustes dos parâmetros, bem como o conteúdo da base de dados a ser analisada para a verificação semântica dos resultados gerados. Os avaliadores deveriam identificar botões que possibilitam carregar a base de dados e especificar o caminho; botões que executam o algoritmo solicitado e configurar os parâmetros de entrada; e botões que possibilitam inserir métodos de visualização e, por fim, interpretar a saída dos algoritmos.

Há particularidades de interação entre essas ferramentas. Por exemplo, em todas as ferramentas o usuário necessita saber em que categoria um determinado item é classificado para poder encontrá-lo. No Weka, por exemplo, para a ação de abrir/carregar a base de dados podem acontecer erros na identificação do componente da ferramenta. Em alguns pontos, o componente chama-se “ARFF Reader” e noutros, é disponibilizado como “FILE”, ficando implícito que este componente também poderá abrir/carregar um arquivo de extensão “.arff”. Caso não saiba *a priori*, o usuário deverá aprender, enquanto realiza a tarefa, qual o fluxo necessário para realização das atividades de MD almeçadas, em cada uma das ferramentas.

Compreendido o percurso cognitivo de uso, um teste de usabilidade foi então planejado e realizado com um grupo de 30 estudantes de Ciência da Computação da Universidade Federal Fluminense (UFF), aplicado imediatamente após terem cursado a disciplina de Aprendizado de Máquina, na qual foram ensinados os conceitos introdutórios de MD, suas tarefas e as principais técnicas de agrupamento e classificação de dados, usadas durante o teste de usabilidade com as ferramentas. Um estudo piloto foi previamente realizado em uma disciplina optativa chamada Introdução a Mineração de Dados no curso de Ciência da Computação da Universidade Estadual do Oeste do Paraná (UNIOESTE), campus de Cascavel.

Cabe destacar que, como parte das atividades dessa disciplina, estes alunos foram divididos em grupos de tal forma que cada 1/5 dos alunos ficou responsável pelo estudo prático da execução das tarefas de MD em uma das seguintes ferramentas: Knime, Orange Canvas, RapidMiner, Weka e CMSR Data Miner. Dessa maneira, cada uma das ferramentas avaliadas era conhecida por apenas 20% dos alunos (usuários), porém, todos deviam fazer as atividades em todas elas e, assim, responder a todas as questões propostas.

O teste de usabilidade consistiu na realização, em um laboratório de ensino de informática, de um conjunto de atividades propostas, relacionadas tanto a agrupamento quanto à classificação de dados, com tempo de duração suficiente para execução das tarefas e para responderem a um questionário contendo questões construídas pelos autores deste estudo sobre o perfil do usuário e a usabilidade da ferramenta.

Para o estudo das ferramentas com relação a agrupamento de dados, como exemplo, as seguintes tarefas deveriam ser realizadas pelos usuários: (1) Carregamento do arquivo de dados “Iris.arff”; (2) Execução do método *k-means* com $k = 3$ para o de agrupamento de dados (e demais parâmetros com valores *default*); (3) Interpretação dos resultados gerados observando todos os elementos de apresentação dos grupos na ferramenta; (4) Emprego de técnicas de visualização de dados sobre o resultado produzidos para observar se a compreensão dos resultados ficou mais fácil ou clara. Para a tarefa de Classificação, a variação desta metodologia está no item (2) onde o solicitado foi a execução do algoritmo *k-nn* (*k-nearest neighbors* vizinhos mais próximos), com parâmetros *default* das ferramentas, apenas variando o número de k .

6. Análise dos Resultados

A parte inicial do questionário procurou identificar aspectos relacionados ao perfil do usuário. A Figura 1 ilustra as respostas à Questão 1 – “Como você classifica seu conhecimento sobre Mineração de Dados?”, à qual os alunos puderam escolher entre as opções “Básico/Introdutório”, “Intermediário” ou “Avançado”. Como era de se esperar, a maioria dos alunos se avaliou com nível de conhecimento “Básico/Introdutório”, enquanto outros se classificaram em nível “Intermediário” ou “Avançado”.

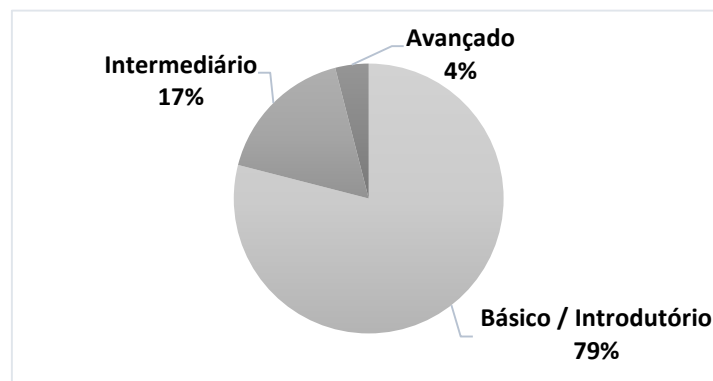


Figura 1. Nível de *expertise* de MD com o qual os estudantes se auto avaliam após terem cursado uma disciplina na área

A segunda parte do questionário procurou identificar extensivamente os diversos aspectos relacionados à usabilidade de cada uma das ferramentas. Neste trabalho, uma vez que o foco é a adequação das ferramentas para apoiar o ensino de MD, deu-se ênfase a questões com foco na facilidade de uso, no uso das técnicas de visualização para interpretação dos resultados e no quanto a navegação das ferramentas está relacionada aos conceitos da área estudada.

Os estudantes avaliaram (I) a facilidade da utilização de cada ferramenta respondendo à Questão 2 – “Esse sistema é fácil de utilizar?”; (II) a importância da visualização para interpretação dos resultados percebida em cada ferramenta respondendo às questões “As técnicas de visualização ajudaram na interpretação dos

resultados dos agrupamentos [Questão 3] ou da classificação [Questão 4]?”; (III) a disposição das tarefas na interface estaria coerente com os aspectos teóricos aprendidos, respondendo à Questão 5 – “A maneira em que os métodos aplicados para a execução das tarefas estão dispostos na interface da ferramenta permite concluir que esta execução segue, na ordem, as três principais etapas do processo de descoberta de conhecimento (pré-processamento, mineração de dados e pós-processamento)?”; e (IV) a categorização dos métodos disponíveis em cada ferramenta e a associação desses métodos às tarefas de mineração de dados, respondendo às Questões 6 – “Os menus e o esquema de navegação pela interface da ferramenta categorizam os métodos disponíveis de acordo com as tarefas mais conhecidas de mineração de dados?” e 7 – “A interface da ferramenta faz corretamente a associação dos métodos disponíveis com cada uma das tarefas de mineração de dados para as quais podem ser empregados?”.

Para todas as perguntas, os usuários deviam atribuir notas de 1 a 5, onde 1 significava “Concordo Totalmente” e 5 significava “Discordo Totalmente”. As médias dos resultados obtidos para a avaliação (I) – Questão 2 – para cada ferramenta constam na Figura 2. De acordo com os estudantes, a Orange Canvas é a que apresenta a interface onde é mais fácil encontrar as informações e elementos desejados, e Weka foi a pior avaliada, embora seja uma das ferramentas mais comumente adotadas. As médias dos resultados obtidos para a avaliação (II) – Questões 3 e 4 – para cada ferramenta podem ser observadas na Figura 3. Verifica-se que o Knime obteve uma avaliação melhor que o Orange Canvas, mas o Weka mais uma vez obteve um mau resultado.

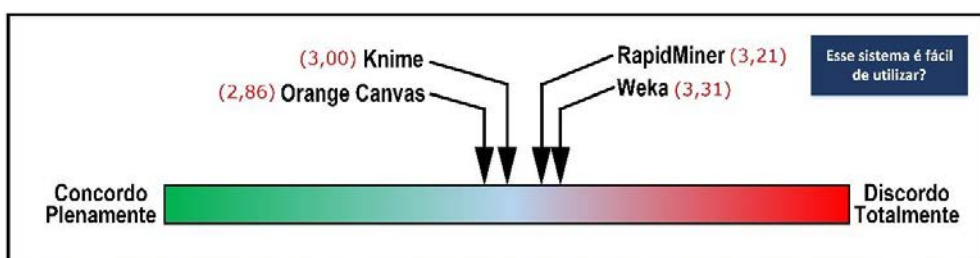


Figura 2. Avaliação dos estudantes em relação à facilidade de uso de cada ferramenta

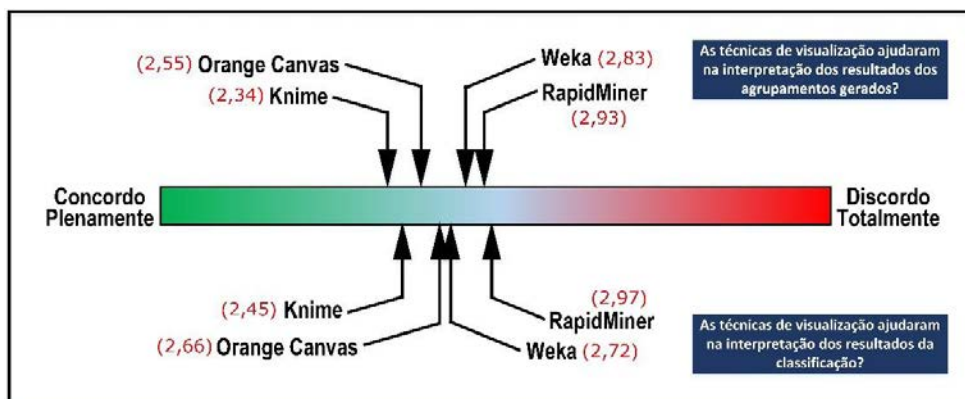


Figura 3. Avaliação dos acadêmicos em relação à importância da visualização para interpretação dos resultados

As médias dos resultados obtidos para a avaliação (III) – Questão 5 – para cada ferramenta podem ser observadas na Figura 4. A Knime se destacou mais neste aspecto,

seguida por RapidMinder. Weka novamente não foi bem avaliada. Já a ferramenta mais distante foi a Orange Canvas. No entanto, é curioso observar que essa ferramenta foi citada como a mais fácil de utilizar, como pode ser visualizado na Figura 2. Embora isso possa parecer contraditório – a disposição dos métodos na interface estar, dentre as ferramentas avaliadas, menos associada às três etapas do processo de MD e, ainda assim, ser a mais fácil de usar, compreende-se que, mesmo que os artefatos de interação não estejam à disposição como um fluxo de processo bem definido na interface, a forma como esses elementos estão dispostos ainda assim facilitam a interação do usuário, por prover-lhes subsídios de metacomunicação que os orientam na execução da tarefa de mineração escolhida. Por fim, as médias dos resultados obtidos para a avaliação (IV) – Questões 6 e 7 – para cada ferramenta podem ser observadas na Figura 5. Observa-se que em ambos os aspectos Knime obteve a melhor avaliação, seguida por Orange Canvas e Weka no primeiro aspecto e, Weka no segundo. Orange Canvas e RapidMiner obtiveram o pior resultado no primeiro aspecto, e RapidMiner, no segundo.

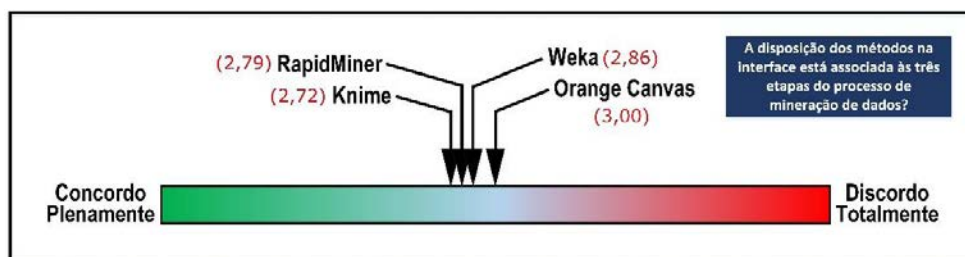


Figura 4. Avaliação dos estudantes em relação a se a disposição das tarefas na interface está coerente com os aspectos teóricos aprendidos

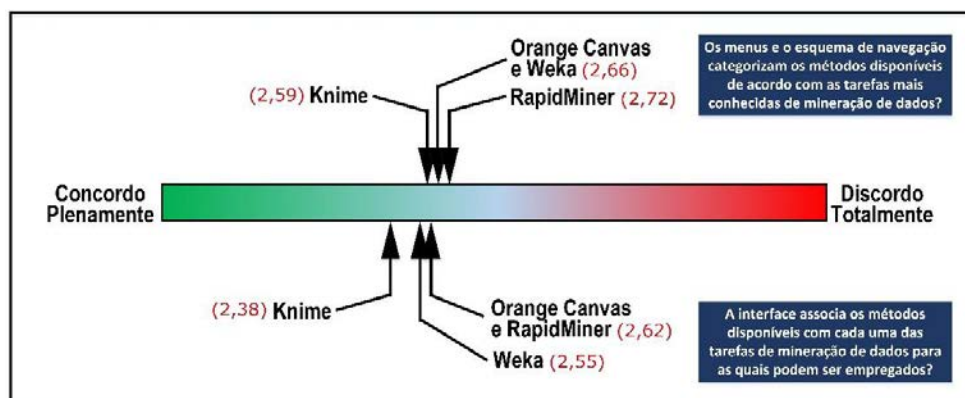


Figura 5. Avaliação dos estudantes em relação à categorização dos métodos e à associação desses às tarefas de mineração de dados

Muito embora todas as ferramentas sejam bastante maduras e abrangentes nas tarefas e técnicas de mineração de dados, nota-se que há diferenças significativas entre elas quando comparadas pelo prisma da interação com o usuário, o que pode ser bastante restritivo para usuários iniciantes, como os estudantes que as utilizam para consolidação de conhecimentos teóricos vistos em sala de aula.

7. Conclusões e Perspectivas

Neste trabalho foram avaliadas sob o ponto de vista de usabilidade, quatro ferramentas que apoiam o processo de MD, do ponto de vista de alunos que estão aprendendo o

referido processo. Pôde-se perceber que, em geral, as ferramentas analisadas dão forte ênfase à visualização de dados, permitindo ao usuário, por exemplo, visualizar os grupos gerados pelos métodos de agrupamento, favorecendo sua compreensão, interpretação e extração de conhecimento sobre o conjunto de padrões. Essa característica é de grande valia aos estudantes para aquisição/consolidação de conhecimentos na área, de forma aplicada e prática.

O uso de ferramentas de MD necessita, em primeira instância, de conhecimento técnico sobre o assunto. Quão maior for o domínio teórico e o conhecimento dos jargões técnicos, estima-se que menor será o tempo de aprendizado de uma ferramenta de MD. Porém, a interface desempenha papel fundamental para o domínio e produtividade em tais ferramentas, dado que influencia diretamente na usabilidade, em aspectos como experiência do usuário e memorização. Portanto, a interface dessas ferramentas passa a ter importância maior, em se tratando de primeiro uso, para fins de aprendizado e aplicação de MD em todos os seus conceitos e nuances.

De modo geral, a usabilidade e a facilidade de operação dos métodos de cada ferramenta são boas, o que as permite abranger públicos com qualquer nível de conhecimento delas, ou até mesmo sobre os métodos que disponibilizam, já que para a execução todas apresentam fluxos de execução que embora diferentes, são bastante claros como processo, tendo início, meio e fim.

A ferramenta Weka, uma das mais utilizadas em cursos de graduação e pós-graduação no Brasil, muito pelo seu pioneirismo, foi, no geral, a que obteve uma das piores avaliações pelos estudantes quanto à sua usabilidade. Outras ferramentas, menos adotadas, como Knime e Orange Canvas, tiveram melhores avaliações gerais. Esse resultado é bastante interessante para mostrar que a preocupação com a escolha de uma ferramenta, mesmo para o ensino de estudantes de cursos da área de Computação, deve estar também pautada em aspectos de interação que podem ser facilitadores ao processo de apropriação do conhecimento de MD.

A avaliação da usabilidade de ferramentas para MD pode aumentar o potencial dessas ferramentas, utilizando o conhecimento do usuário e ajudando-o a obter resultados cada vez mais expressivos das bases de dados. Por outro lado, a avaliação realizada mostra que outros aspectos de IHC necessitam ser abordados mais cuidadosamente para extrair diretrizes que orientem no *design* de interação de ferramentas para auxiliar o processo de MD.

Como trabalhos futuros os autores pretendem ampliar a avaliação ora apresentada: (i) incluindo novas tarefas de mineração de dados, abrangendo ao menos um método de cada uma das principais tarefas de MD (Classificação, Associação, Agrupamento e Regressão); e (ii) avançando na indicação de qual dessas ferramentas seria mais indicada para o ensino dessa disciplina, a partir da avaliação da interação dos estudantes. Com isso, há também a intenção de avaliar mais profundamente essas ferramentas e o quão experiente o usuário tem que ser para poder usar seus algoritmos. Finalmente, pretende-se propor uma ontologia que disponha de todos os métodos de MD, de forma que melhore a interação do usuário (estudante) com tais ferramentas.

Referências

- BOSCARIOLI, C.; VITERBO, J.; TEIXEIRA, M. F.; ROHSIG, V. Analyzing HCI Issues in Data Clustering Tools. (2014) In: Proceedings of the 16th International Conference on Human-Computer Interaction. London: Springer, v. 8521. p. 22-33.
- BRETERNITZ, V. J.; SILVA, L. A. “Big Data: Um Novo Conceito Gerando Oportunidades e Desafios” (2013), Revista Eletrônica de Tecnologia e Cultura – RETC. 13ª Ed., FATEC, ISSN: 2177-0425, p. 106-113.
- CANVAS. Orange Canvas: Data Mining – Fruitful and Fun. (2013). Disponível em <http://orange.biolab.si/>. Acessado em 05 de março de 2016.
- CHAWLA, N.V. Teaching Data Mining by Coalescing Theory and Applications. In: Proc. 35th ASEE/IEEE Frontiers in Education Conference.
- FERREIRA, P. C.; BERNARDINI, F. C.; SILVA, C. R. C.; VITERBO, J. How Computer Science Undergraduate Programs in Brazilian Public Universities Address Artificial Intelligence and its Topics: A Preliminary Report (2012). Learning and Nonlinear Models, v. 12, pp. 73-85.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update (2009); SIGKDD Explorations, Volume 11, Issue 1.
- JAFAR, J. M. A Tools-Based Approach to Teaching Data Mining Methods (2010). Journal of Information Technology Education: Innovations in Practice. Beachboard, J (Editor). Volume 9, p. 1-24.
- KING, B.R.; SATYANARAYANA, A. Teaching Data Mining in The Era of Big Data (2013). In: Proc. 120th ASEE Annual Conference & Exposition.
- KNIME. Knime: Open for Innovation (2013). Disponível em <https://www.knime.org/>. Acessado em 29 de maio de 2015.
- MANASEER, S. A.; MALIBARI, A. Improve Teaching Method of Data Mining Course (2012). I. J. Modern Education and Computer Science, Volume 2, p. 15-22.
- RAPIDMINER. Rapidminer: Analytics for Anyon (2013). Disponível em <https://rapidminer.com/>. Acessado em 08 de março de 2016.
- REZENDE, R. O.; PUGLIESI, J. B.; MELANDA, E. A.; DE PAULA, M. F. (2005). Mineração de Dados. Cap 12. In: REZENDE, S. O. (Org.). Sistemas Inteligentes: Fundamentos e Aplicações. Barueri-SP: Ed. Manole, p. 307-335.
- SATYANARAYANA, A. Software Tools for Teaching Undergraduate Data Mining Course (2013). In: Proc. ASEE MidAtlantic Conference.
- WHARTON, C., RIEMAN, J., LEWIS, C., and POISON, P. The cognitive walkthrough method: A practitioner's guide (1994).
- WITTEN, I.; FRANK, E.; HALL, M. Data Mining: Practical Machine Learning Tools and Techniques (2011). Morgan Kaufmann Publishers.

Caracterização do Público-Alvo de Jogos Educacionais na área da Computação

Paulo E. Battistella¹, Christiane G. von Wangenheim¹

¹Grupo de Qualidade de Software (GQS) - Instituto Nacional para Convergência Digital (INCoD)–Departamento de Informática e Estatística (INE)– Universidade Federal de Santa Catarina (UFSC)
Florianópolis –SC – Brasil

paulo@incod.ufsc.br, c.wangenheim@ufsc.br

Abstract. *Educational games can bring several benefits also for teaching computing. However, in order to obtain these benefits it is important to design these games taken into consideration the characteristics of this specific target audience. Thus, the objective of this paper is to characterize the computing students with respect to the application of educational games. Data has been collected via an online survey with the participation of 406 students. The results show that the target audience are mostly male, frequently play digital games frequently on their notebook or smartphone preferring RPG and action-adventure games. The result of our survey can help developers of games for teaching computing as part of the context analysis in game design.*

Resumo. *Jogos educacionais podem trazer diversos benefícios, incluindo no ensino de computação. No entanto, para obter esses benefícios é importante conceber jogos que levam em consideração a caracterização do público-alvo. Assim, o objetivo deste artigo é caracterizar os estudantes de computação com relação à aplicação de jogos educacionais. Os dados foram coletados por meio de uma pesquisa on-line com 406 participantes. Os dados mostram que o público-alvo é principalmente masculino, frequentemente jogam jogos digitais em seus notebooks ou smartphones, preferem jogos de RPG e ação-aventura. O resultado da pesquisa pode ajudar os desenvolvedores de jogos educacionais, como uma parte da análise de contexto no design de jogos.*

1. Introdução

Jogos são considerados poderosos métodos de ensino e acredita-se resultar em uma ampla gama de benefícios. Com isso, aumentando a eficácia da aprendizagem, interesse, motivação e persistência dos alunos [Kirriemuir e McFarlane, 2004; Freias, 2006; Mitchell e Savill-Smith, 2004; Prensky, 2007].

Jogos podem promover a “*aprendizagem ativa*” para alcançar um aprendizado profundo em um tempo de ensino aceitável e sem sobrecarga para o instrutor [Wangenheim e Shull, 2009]. Além disso, servem como meio de entretenimento para aprofundar e praticar conceitos, bem como ilustrar dinâmicas ou princípios abstratos [Hays e Singer, 1988]. Devido a suas características intrínsecas, tais como, competição, desafio e interação, podem transformar o aprendizado em uma experiência envolvente e divertida [Kafai, 2001].

Portanto, aprendizagem baseada em jogos também é uma alternativa promissora para ensinar computação no ensino superior. A utilização de jogos educacionais no

ensino de computação está se tornando cada vez mais popular [Poulova e Klimova, 20015]. Existe uma grande diversidade de jogos para ensinar computação [Boyle et al., 2016; Battistella e Wangenheim, 2016], p.ex. para estrutura de dados (SORTIA [Battistella, Wangenheim, e Wangenheim, 2012], DEG4Trees [Barbosa et al., 2015], RPG4Sorting [Nunes e Parreira, 2015]), engenharia de software (UsabiliCity [Ferreira, 2014], Simulador para Ensinar SCRUM [Gestal e Barros, 2014]) e programação (Abstração Game [Oliveira e Barros, 2013]).

Para obter os benefícios de aprendizagem esperados, os jogos precisam ser desenvolvidos de forma sistemática [Fullerton, 2008; Branch, 2009]. Seguindo abordagens de desenvolvimento de jogos educacionais (p.ex. *The ELEKTRA Project* [Kickmeier-Rust et al., 2006], *Game-Based Learning Framework* [Staalduinen e Freitas, 2011] e também ENgAGED (*EducatioNAl GamEs Development*) [Battistella e Wangenheim, 2015]), um dos primeiros passos é a análise do público-alvo para possibilitar o alinhamento do jogo as características, necessidades e preferência do público-alvo [Branch, 2009]. No entanto, é possível verificar que os jogos educacionais disponíveis na literatura não apresenta um levantamento sistemático do público-alvo [Battistella e Wangenheim 2016; Boyle et al. 2016].

Neste contexto, o presente artigo apresenta os resultados de uma pesquisa entre alunos de cursos de computação no ensino superior. Os resultados da pesquisa podem ajudar os desenvolvedores de jogos educacionais, como uma parte da análise de contexto no design de jogos.

2. Definição da pesquisa

O objetivo desta pesquisa é caracterizar o público-alvo de jogos educacionais na área de computação. Para alcançar o objetivo foi utilizado o método GQM (*Goal, Question and Metric*) [Basili, Caldiera e Rombach 1994]. A partir deste objetivo foram derivadas as perguntas de análise e as métricas. Para operacionalizar a coleta de dados foi projetado um questionário. A tabela 1 apresenta as questões propostas para alcançar os objetivos da pesquisa.

Tabela 1. Questões e escalas de resposta do questionário.

Caracterização dos Jogadores da Área da Computação	Escala de resposta
1. Qual curso você está cursando ou cursou?	Bacharelado em Ciência da Computação; Bacharelado em Sistemas de Informação; Bacharelado em Engenharia da Computação; Bacharelado em Engenharia de Software; Bacharelado em Informática; Licenciatura em Computação ou Informática; Tecnologia em Informática; Pós-graduação em Computação.
2. Qual é a sua idade?	Menor de 15 anos; Dos 15 aos 18 anos; Dos 19 aos 22 anos; Dos 23 aos 25 anos; Dos 26 aos 30 anos; Dos 31 aos 40 anos; Maior de 40 anos.
3. Qual é o seu gênero?	Masculino ou Feminino.
4. Qual é a sua preferência de gêneros de jogo?	Ação-aventura; RPG, Estratégia; Corrida, Simulação; Puzzle; Quiz; Adivinhação; Roll-and-Move; [Battistella e Wangenheim, 2015].

5. Você tipicamente prefere jogar jogos em qual plataforma?	Jogo digital de computador stand-alone p. ex. computador pessoal sem acesso a internet. Jogo digital de computador online p. ex. computador pessoal com acesso a internet. Jogo digital de console p. ex. Xbox 360, Playstation, Wii - Nintendo. Jogo digital mobile - smartphone p. ex. iPhone, Samsung Galaxy. Jogo digital mobile - tablet p. ex. iPad, Samsung Galaxy Note. Jogo não-digital de tabuleiro. Jogo não-digital de cartas. Jogo não-digital papel & lápis (para ser jogado apenas com papel e lápis.) p. ex. jogo imagem e ação, "caça-palavras". Jogo não-digital acessórios (para ser jogado com acessórios ou adereços (objetos portáteis)) p. ex. Lego.
6. Qual é o modo de interação preferido de jogo?	Single-player; Multiplayers competitivo; Multiplayers cooperativo; Multigrupos competitivo; Multigrupos cooperativos. [Battistella e Wangenheim, 2015]
7. Com que frequência você joga jogos digitais?	Nunca; Raramente; Mensalmente; Semanalmente; Diariamente.
8. Com que frequência você joga jogos não-digitais?	Nunca; Raramente; Mensalmente; Semanalmente; Diariamente.
9. Liste os nomes dos seus três jogos favoritos?	Campo de texto aberto.
10. Quais são os seus cenários de jogos favoritos?	Esporte; Medieval; Dungeons&Dragons; Ficção científica; Animais; Outro. [Battistella e Wangenheim, 2015]
11. Quais dispositivos eletrônicos você utiliza em geral, não apenas para jogo?	Computador; Tablet; Smartphone, celular; outro.
12. Onde você joga jogos?	Casa; <i>Lan house</i> ; Universidade; outro.
13. Qual sistema operacional para computador, você utiliza em casa?	Windows; Linux; MacOS; outro.
14. Qual é a velocidade de banda da sua conexão de internet em sua casa?	1 Mb; 2 Mb; 4 Mb; 5 Mb; 8 Mb; 10 Mb; 20 Mb; 30 Mb; 50 Mb; 100 Mb; outro.
15. Quais navegadores de internet você mais usa?	Internet Explorer; Firefox; Chrome; Safari; Opera; outro.
16. Qual a resolução do monitor que você utiliza?	176 x 144 (px); 320x200 (px); 352 x 288 (px); 480 x 320 (px); 720 x 240 (px); 720 x 480 (px); 768 x 576 (px); 854 x 480 (px); 1920 x 1080 (px); 2560 x 2048 (px); 3840 x 2400 (px); outro.
17. Você possui algum comentário a mais para fazer sobre suas preferências de jogo?	Campo de texto aberto.

3. Execução da Pesquisa

A pesquisa foi realizada de modo online, entre novembro de 2014 até abril de 2015. O convite de participação foi realizado virtualmente, utilizando a lista de e-mails de cursos de computação da UFSC, lista de e-mails da Sociedade Brasileira de Computação (SBC), grupos de redes sociais como *Facebook* e *LinkedIn*. Ao total participaram da pesquisa 406 alunos e egressos de cursos de computação, sendo que 304 participantes responderam a pesquisa completamente.

A maior parte dos participantes são alunos ou egressos do curso de Ciência da Computação, com percentual de 40%. Também participaram da pesquisa 25% dos alunos dos cursos de Sistema de Informação. A pesquisa teve a participação de 12% de alunos de cursos de pós-graduação e 8% de alunos do curso de Engenharia de Computação. Nos demais cursos, a participação não passou de 5% dos alunos ou egressos.

4. Resultados

Grande parte dos participantes (56%) da pesquisa está entre a faixa de 19 e 30 anos de idade, o que evidencia que o público-alvo é jovem. A pesquisa evidenciou que a maioria dos alunos (84%) é do gênero masculino.

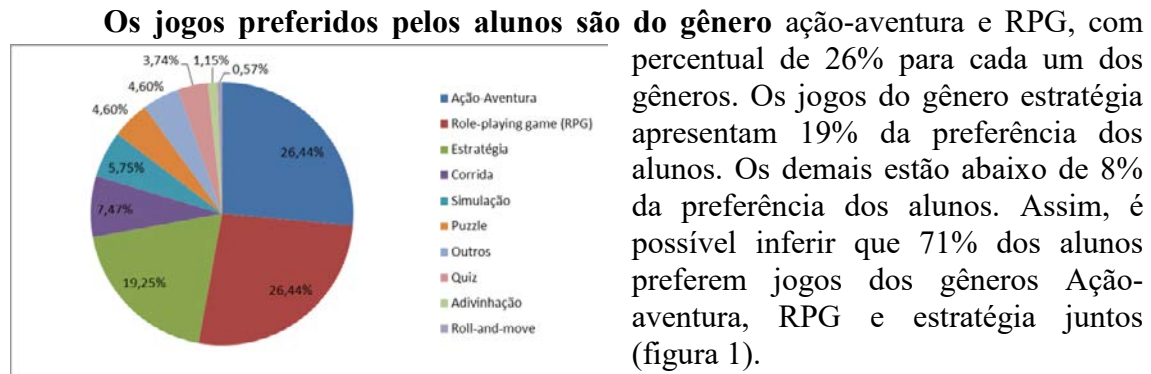


Figura 1. Preferências de gênero de jogo, entre 348 respostas de alunos.

de computador online (37%), console (23%), computador *stand-alone* (19%) e *mobile – smartphones* (11%). As demais plataformas não apresentaram mais do que 3% da preferência dos alunos.

Considerando que entre os 93% dos alunos que preferem jogos digitais, destaca-se que mais de um terço dos alunos (37%) preferem jogos online. Tipicamente estes jogos são *multiplayers* e do gênero RPG. Analisando as preferências dos alunos por jogos digitais, principalmente pelos jogos de computadores e console é possível inferir que o desenvolvimento de novos jogos para este público-alvo terá mais receptividade se for um jogo de computador online ou de console (Figura 2).

O modo de interação de jogos mais utilizado pelos alunos é o *singleplayer* (41%). No entanto, os alunos apresentam 40% de preferência pelo modo de interação *multiplayer*. Ele teve 23% para *multiplayers competitivos*, os quais os jogadores competem entre si, e 17% *multiplayers cooperativo*, os quais os jogadores colaboram entre si (figura 3).

Em questão de múltiplas respostas, fica evidente que os participantes preferem jogar sozinhos (84%), normalmente em suas casas. Quando se trata de jogar com colegas, existe pouca diferença entre jogar juntos com os colegas (40%) ou a distância (online) (47%). Vale destacar que a opção *lanhouse* não houve resposta.

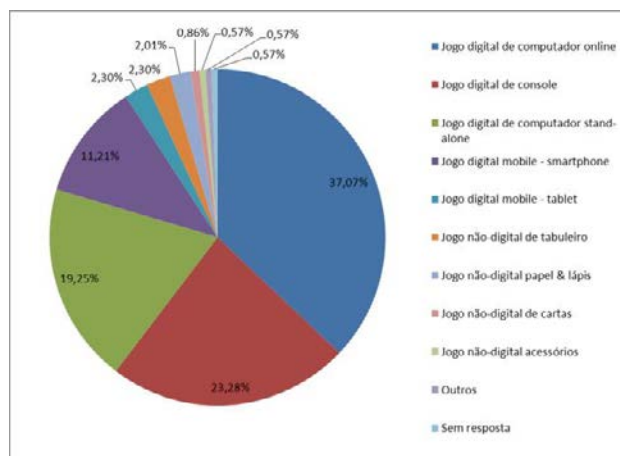


Figura 2. Plataformas de jogo de interesse entre 346 respostas de alunos.

A frequência dos alunos que jogam jogos digitais evidencia a preferência em jogar semanalmente (33%), diariamente (25%) e mensalmente (15%). O que evidencia a preferência de 73% dos alunos em jogar jogos digitais. No entanto, 24% dos alunos raramente jogam jogos digitais, como apresenta a figura 4.

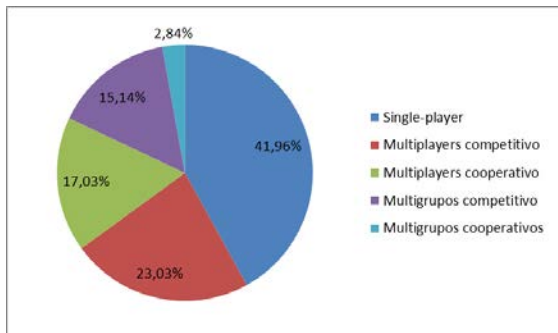


Figura 3. Modo de interação de preferência de 317alunos.

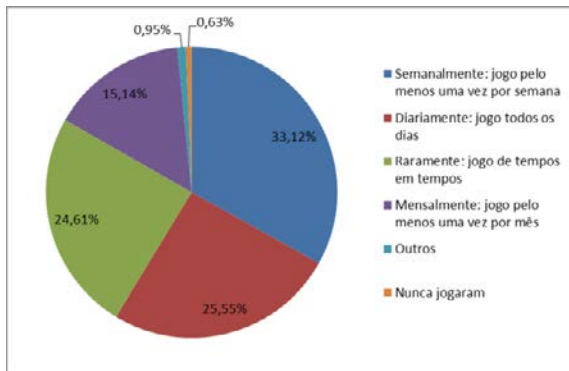


Figura 4. Frequências de jogo para jogos digitais, entre 317 respostas dos alunos.

Em relação a jogar jogos não-digitais é possível verificar que 93% dos alunos se interessam por estes jogos. Diferentemente dos digitais, os alunos jogam raramente ou mensalmente. Isso evidencia que mesmo com muitos alunos interessados, a frequência de uso é baixa.

A frequência dos alunos que jogam jogos não-digitais é de 65% raramente, 19% mensalmente e 9% semanalmente. Apenas 4% dos alunos nunca jogam jogos não-digitais.

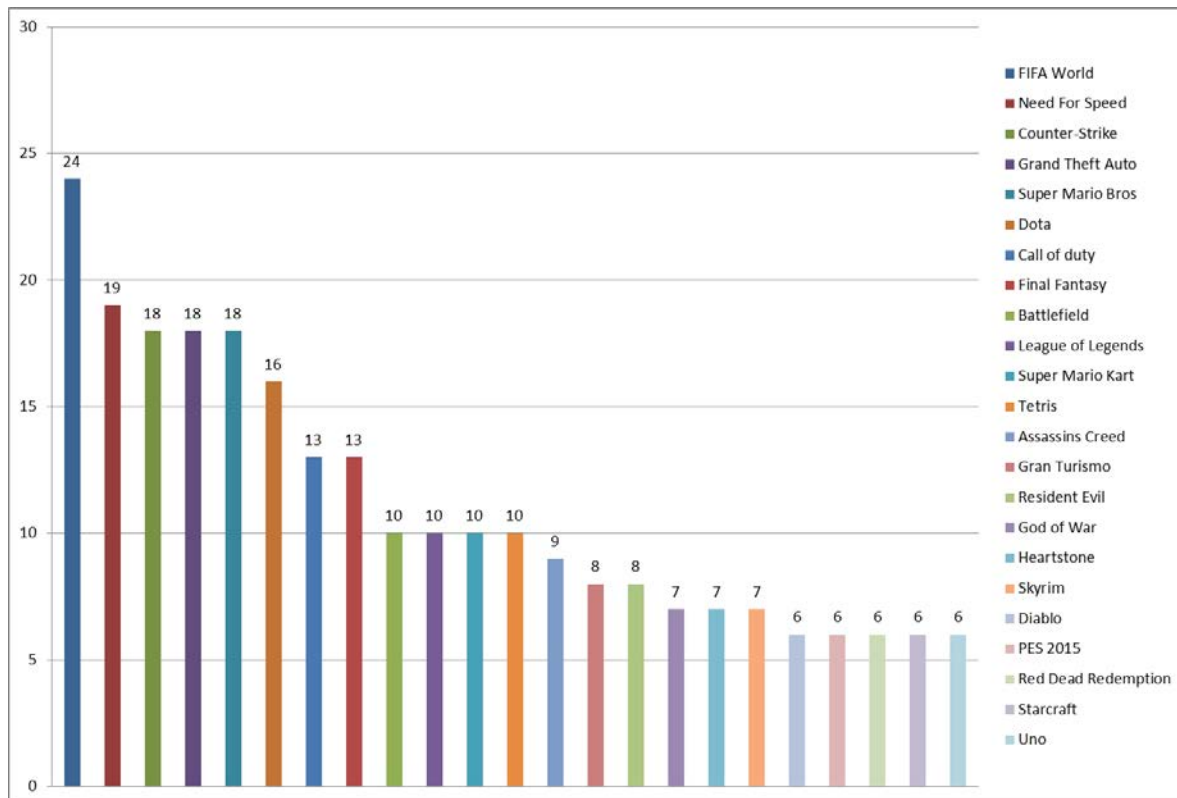


Figura 5. Ranking dos 23 jogos mais jogados, entre 310 respostas dos alunos.

O ranking dos jogos mais jogados apresenta 234 jogos, sendo a grande maioria digital. Os 5 jogos mais citados foram *FIFA World* (24 citações), *Need For Speed* (19 citações) e *Super Mario Bros*, *Grand Theft Auto* (GTA) e *Counter-strike* (18 citações). Entre os 23 jogos mais citados (figura 5), destaca-se que apenas 1 é não-digital.

Os cenários preferidos pelos alunos são medieval (51%), *Dungeons & Dragons* (48%) e ficção-científica (48%). Estes três cenários estão ligados também a preferência pelos gêneros de ação-aventura e RPG. Estes cenários tipicamente apresentam uma narrativa interessante e que leva o jogador à uma maior imersão ao jogo e apresenta gráficos de qualidade (figura 6).

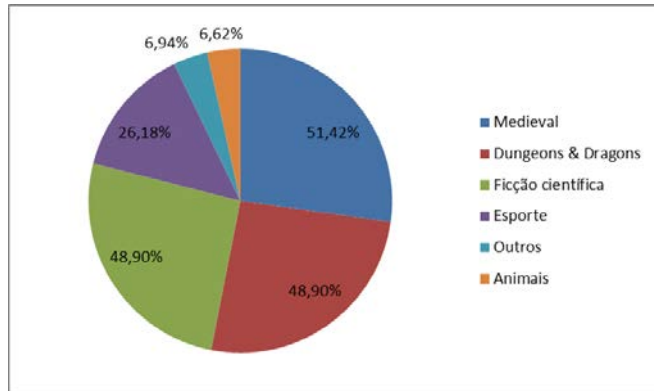


Figura 6. Cenários dos jogos preferidos por 406 alunos.

Em relação aos equipamentos eletrônicos foi coletada mais de uma resposta por aluno. Deste modo, os jogadores optaram por mais de um equipamento utilizado para jogar. Ficou evidente que a maior parte deles são *notebook*, *smartphone* e *desktop*. Vale destacar que *smartphone* está em segundo lugar, perdendo apenas para os notebooks. Assim, fica evidente que os jogadores preferem equipamentos que permitam maior mobilidade. Por exemplo, utilizando o *notebook* o jogador pode jogar em vários locais, p. ex. em casa, faculdade ou reunidos em grupos de amigos (figura 7).

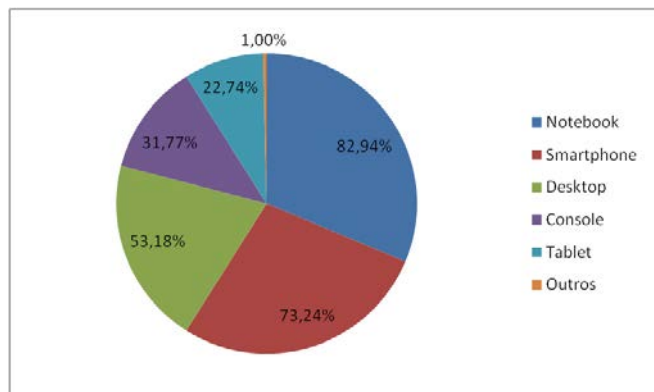


Figura 7. Dispositivos eletrônicos utilizados por 406 alunos para jogar jogos digitais.

O sistema operacional mais utilizado para jogar é o Windows (86%). Este fato fica evidente no Ranking dos jogos mais jogados (figura 5) por apresentar principalmente os jogos deste tipo de sistema operacional. Considerando que a questão apresenta opções de múltiplas respostas, o Linux apresentou 36% do interesse dos alunos para jogar jogos, o MacOS 9% e os demais sistemas operacionais 1%.

A velocidade da conexão de banda larga foi informada por 292 alunos. Entre eles, 120 possuem internet com velocidade de 10 Mb, 82 com velocidade 1 a 5 Mb. Também 23 alunos possuem internet com velocidade de 15 Mb e 22 participantes possuem internet entre 25 e 30 Mb. A partir destes dados, é possível inferir que menos da metade dos participantes não possuem conexão de banda larga.

Os navegadores de internet mais utilizados pelos alunos em questão de múltipla escolha evidenciou que 84% dos alunos utilizam o navegador *Chrome*, seguido do *Firefox* com 55%. Os navegadores Safari, Opera e Internet Explorer apresentaram

menos de 6% das preferências dos alunos. Outros navegadores apresentaram pouco mais de 1%.

Entre a preferência dos alunos pela resolução de monitor fica evidente que não existe uma resolução padrão. No entanto, houve um predomínio de 22% para monitores de 22 polegadas e 19% para monitores de 17 polegadas.

Nos comentários adicionais dos alunos, apenas 15% informaram comentários sobre suas preferências de jogos. Entre essas informações é possível destacar: (a) preferência por jogos que melhorem a cognição, como *Neuro Nation* e *Fitbrains*; (b) interesse por jogos cooperativos por serem divertidos, no entanto, são raros de encontrar; (c) preferência por jogos que ofereçam liberdade para o jogador fazer o que desejar, sem ter que seguir linearmente uma sequência de passos; (d) interesse por jogos de tabuleiro, não-digitais, porém apresenta dificuldade de encontrar pessoas para jogar; (e) preferência por jogos versáteis, tendo recursos que podem ser otimizados, como no RPG.

5. Discussão

A partir dos dados coletados é possível concluir que atualmente o público-alvo de jogos educacionais na área da computação é na grande maioria do gênero masculino e formado por jovens. Eles têm o hábito de frequentemente jogar jogos digitais online nos seus *notebooks* e/ou *smartphones*. Em relação ao modo de interação, os alunos ficaram relativamente divididos entre *singleplayer* e *multiplayers*. Na figura 5 é possível verificar o interesse dos alunos por estes dois tipos de modo de interação, principalmente porque muitos jogos citados apresentam versões *singleplayer* e *multiplayers*.

Em geral eles preferem jogos de RPG e ação-aventura, preferindo cenários do tipo *Dungeons & Dragons* e ficção-científica. Entre os jogos de RPG mais citados destaca-se *Diablo*, *Dota*, *Final Fantasy*, *League of Legends*, *Starcraft*. Entre os jogos mais citados de Ação-aventura destaca-se *Super Mário Bros* e *Grand Theft Auto* (GTA).

Com base nas respostas dos alunos, é possível verificar que grande parte deles apresentam interesse em jogos não-digitais. Porém, mesmo com grande interesse dos alunos é possível verificar que estes jogos são jogados raramente. Entre os 23 jogos listados no Ranking (figura 5), apenas um jogo é não-digital (*Uno*).

Em relação ao equipamento tecnológico ficou evidente a preferência dos alunos pelo Windows como sistema operacional. E também as respostas mostraram que mais da metade dos alunos possuem banda larga para acessar a internet de suas casas.

De modo geral é possível perceber um grande interesse dos alunos de computação por jogos. Deste modo, é possível utilizar este interesse para estimular o ensino de conteúdos computacionais por meio de jogos educacionais, tornando assim o aprendizado mais atrativo e divertido. Com base nesta pesquisa, o professor também pode estimular o aprendizado de gêneros ou plataformas de jogos que não são jogados com frequência, mas que são de interesse dos alunos. Por exemplo, no caso de jogos de não-digitais que podem promover cooperação ou interação entre os alunos. Ou então, a utilização de jogos educacionais que simulam uma atividade real, como nos jogos para área de Engenharia de Software *Lidar com Pessoas Difíceis*, *SimSE* e *Problems and*

Programmers apresentados na revisão sistemática da literatura [Battistella e Wangenheim, 2016].

5.1 Ameaças à validade da pesquisa

Os resultados obtidos nesta pesquisa precisam ser interpretados com cuidado, levando em consideração as ameaças à sua validade. Entre os fatores que a ameaçam, destaca-se o impacto na validade dos resultados, principalmente pela generalização dos resultados. Outra ameaça está relacionada ao tamanho relativamente pequeno da amostra, considerando a grande quantidade de alunos de cursos de computação no Brasil. No entanto, mesmo com uma amostra relativamente pequena é possível considerar os dados analisados relevantes.

Outro fator que pode ser considerado ameaça, são as respostas dos alunos que não interpretaram as perguntas de modo adequado ou ao preencher a pesquisa esconderam informações. Outra ameaça identificada é em relação aos alunos que responderam a pesquisa, mas desconheciam o tema de pesquisa, por exemplo, no caso de alunos que não conhecem jogos, mas participaram da pesquisa.

Outra ameaça à validade é o fato de poucos alunos de cursos como Bacharelado em Informática, Licenciatura em Computação ou Informática e Bacharelado em Engenharia de Software terem participado da pesquisa, mostrando que a caracterização do público-alvo realizada está voltada principalmente para os alunos dos cursos de Ciência da Computação e Sistema de Informação.

Outra ameaça identificada é em relação aos gêneros de jogos educacionais. Atualmente os gêneros não apresentam uma taxonomia padrão que seja utilizada amplamente na literatura. Muitas vezes estas taxonomias apresentam os jogos educacionais como um gênero, assim como ação-aventura, estratégia, *puzzle*, o que torna difícil a caracterização deste gênero. Para minimizar esta ameaça, foi utilizada a taxonomia criada para o processo ENgAGED [Battistella e Wangenheim 2015], na qual foi criada uma taxonomia de gêneros de jogos (digitais e não-digitais) para classificar os jogos educacionais desenvolvidos para área da computação com base nos gêneros disponíveis na literatura e sites de jogos online.

6 Conclusão

O artigo apresentou uma pesquisa para caracterizar os perfis de jogadores da área da computação. Participaram da pesquisa 406 alunos e egressos dos cursos de computação, porém, apenas 304 alunos responderam completamente. Os resultados evidenciam que a maioria dos alunos são do gênero masculino e jovens entre 19 e 30 anos de idade.

A preferência por jogos mostra um predomínio dos interesses dos alunos por jogos online e dos gêneros Ação-aventura e RPG. O *ranking* com os 23 jogos mais citados também evidencia esta preferência. Além disso, a preferência também mostra que os jogadores jogam com grande frequência os jogos digitais e com pouca frequência os não-digitais.

Em relação a preferência por equipamentos eletrônicos, a pesquisa mostra que o Windows é o sistema operacional mais utilizado. E também, mais da metade dos alunos possuem acesso a conexão de banda larga e utilizam com frequência seus *notebooks* e

smartphones para jogar. Assim, fica evidente o interesse dos alunos por jogos online que possam ser jogados por equipamentos que permitam mobilidade.

A pesquisa realizada neste artigo permitiu a caracterização dos perfis de jogadores, fase essencial do processo de Design Instrucional e Design de Jogos. Deste modo, os resultados apresentados podem ser aplicados no desenvolvimento de novos jogos educacionais para área de computação.

Agradecimentos

Este trabalho foi apoiado pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), uma entidade do governo brasileiro focada no desenvolvimento científico e tecnológico e pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

Referências

- Barbosa, W. A. et al. (2015) “DEG4Trees: Um Jogo Educacional Digital de Apoio ao Ensino de Estruturas de Dados”, Workshop sobre Educação em Computação, Recife, Pernambuco, Brasil.
- Basili, V., Caldiera, G. e Rombach, H. D. (1994) “Goal Question Metric Approach”, Encyclopedia of Software Engineering, John Wiley & Sons, Inc., p. 528-532.
- Battistella, P. E., Wangenheim, A. e Wangenheim, C. G. von. (2012) “SORTIA - Um Jogo para Ensino de Algoritmo de Ordenação: Estudo de caso na Disciplina de Estrutura de Dados”, Simpósio Brasileiro de Informática na Educação, Rio de Janeiro, RJ, Brasil.
- Battistella, P. E. e Wangenheim, C. G. von. (2016) “Games for Teaching Computing in Higher Education – A Systematic Review”, IEEE Technology and Engineering Education.
- Battistella, P. E. e Wangenheim, C. G. von, (2015) “ENgAGED: Processo de Desenvolvimento de Jogos para Ensino em Computação”, Relatório Técnico INCoD/GQS.01.2015.P, Florianópolis, Brasil.
- Boyle, E. A. et al. (2016) “An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games”, Computers & Education, 94, pp. 178-192.
- Branch, R. M. (2009) “Instructional Design: The ADDIE Approach”, Springer, New York, EUA, p. 203.
- Ferreira, B. et al. (2014) “UsabiliCity: Um Jogo de Apoio ao Ensino de Propriedades de Usabilidade de Software Através de Analogias”, Simpósio Brasileiro de Informática na Educação, Dourados, MS, Brasil.
- Freitas, S. I. de (2006) “Using games and simulations for supporting learning”, Learning, Media and Technology, Digital Games and Learning, 31(4), p. 343-358.
- Fullerton, T. (2008) “Game design workshop: A playcentric approach to creating innovatives games”, 2ª ed., Elsevier.

- Gestal, P. R. E. e Barros, R. M. de. (2014) “Proposta de Um Simulador para Auxiliar no Processo de Ensino do Scrum”, Simpósio Brasileiro de Sistemas de Informação, Londrina, PR, Brasil.
- Hays, R. T. e Singer, M. J. (1988) “Simulation fidelity in training system design: Bridging the gap between reality and training”, *Recent Research in Psychology*, Springer, p. 415.
- Kafai, Y. B. (2001) “The educational potential of electronic games: From games-to-teach to games-to-learn”, *Conference on playing by the rules: the cultural policy challenges of video games*. Chicago, IL, EUA.
- Kickmeier-Rust, M. D. et al. (2006). “The ELEKTRA project: Towards a new learning experience”, In M. Pohl, A. Holzinger, R. Motschnig, & C. Swertz: *M3 – Interdisciplinary Aspects on Digital Media & Education*, Vienna, p. 19-48.
- Kirriemuir, J. e McFarlane, C. A. (2004) “Literature Review in Games and Learning Literature Review in Games and Learning”, Bristol: FutureLab Series.
- Mitchell, A. e Savill-Smith, C. (2004) “The use of computer and video games for learning: a review of the literature”, Learning and Skills Development Agency, Ultralab, London, UK.
- Nunes, I. de F. e Parreira Junior, P. A. (2015) “RPG4Sorting - Um Jogo Educacional para Auxílio ao Ensino de Métodos de Ordenação”, *Workshop sobre Educação em Computação*, Recife, Pernambuco, Brasil.
- Oliveira, E. D. de. e Barros, R. M. de. (2013) “ABstração Game: Um Jogo Para Facilitar o Processo de Ensino e Aprendizagem da Orientação a Objetos”, *Workshop sobre Educação em Computação*, Maceió, AL, Brasil.
- Poulova, P. e Klimova, B. (2015) “Education in Computational Sciences”, *International Conference on Computational Science*. Reykjavík, Iceland, p. 1996-2005.
- Prensky, M. (2007) “Digital game-based learning”, Paragon House, p.464.
- Staalduinen, J. v. e Freitas, S. de. (2011) “A Game-Based Learning Framework: Linking Game Design and Learning Outcomes”, Nova York: Peter Lang, p. 29-45.
- Wangenheim, C. G. von e Shull, F. (2009) “To game or not to game?”, *IEEE Software*, 26(2), p. 92-94.

Computasseia: Um Jogo para o Ensino de História da Computação

Jean Carlos Oliveira Santos, Karen da Silva Figueiredo

Instituto de Computação – Universidade Federal do Mato Grosso (UFMT)
Cuiabá, Mato Grosso, Brasil

jcoliveira93@gmail.com, karen@ic.ufmt.br

***Abstract.** The study of History of Computing is necessary to develop the knowledge about the essentials of the area; to understand the evolution of the resources, current techniques and methods; and to expand the critical thinking regarding the present and future of Computing. However, the process of learning the History of Computing can become very boring because of the many names, facts and dates involved. The use of games as educational resources to support computer education is a strategy which can change a boring experience into an engaging experience. This paper proposes a strategy card game about History of Computing that can be used at different levels of education.*

***Resumo.** Estudar a história da Computação é essencial para conhecer os fundamentos da área, entender a evolução dos recursos, técnicas e métodos atuais e ampliar o senso crítico com relação ao desenvolvimento da Computação no presente e futuro. Entretanto, o ensino de História da Computação pode tornar-se enfadonho, com a presença de muitos nomes, fatos e datas. A utilização de jogos como recursos educacionais no apoio do ensino de Computação é uma estratégia capaz de transformar uma experiência entediante em uma experiência engajadora. Este trabalho propõe um jogo de cartas sobre diversos conteúdos relacionados à história da Computação que pode ser utilizado em diferentes níveis de ensino.*

1. Introdução

É crescente o reconhecimento nas universidades de que estudantes de Computação (Ciência da Computação, Sistemas de Informação, entre outros cursos) devem compreender, não somente os aspectos técnicos, mas também os aspectos culturais da sua área para a evolução dos seus estudos acadêmicos [Berry *et al.* 2011].

Uma grande contribuição neste sentido é o estudo da história da Computação. Diversas universidades possuem em seus currículos dos cursos de Computação alguma disciplina que discute os aspectos históricos da área como parte total ou parcial da ementa. Segundo Fonseca Filho (2007), a Universidade de Stanford e o Instituto Charles Babbage da Universidade de Manchester podem ser citados como algumas das instituições pioneiras na criação das disciplinas de História da Computação na década de 1990.

Estudar a história da Computação é essencial para conhecer os fundamentos da

área, entender a evolução dos recursos, técnicas e métodos utilizados atualmente e ampliar o senso crítico com relação ao desenvolvimento da Computação no presente e suas perspectivas futuras. Entretanto, o ensino de história da Computação possui, em certos aspectos, as mesmas características do ensino de uma disciplina típica de História. Dependendo da estratégia pedagógica utilizada, a disciplina pode ser considerada entediante e de propósito duvidoso ou relevante, inspiradora e recompensadora para os alunos [Haydn *et al.* 2014].

Entre as estratégias contemporâneas para um ensino mais atraente e engajador encontra-se a utilização de jogos sérios, do inglês “*serious games*”, produtos desenvolvidos com princípios de *design* de jogo interativo, com o objetivo de transmitir um conteúdo educacional ou de treinamento ao usuário que vai além do entretenimento, *i.e.* que tenham uma mensagem a transmitir [Michael e Chen 2005]. Os jogos sérios estão sendo empregados com sucesso tanto no ensino de disciplinas de História *e.g.* [Baptista e Carvalho 2013, Froschauer *et al.* 2011, Neves 2011], como em diversas iniciativas em disciplinas da área de Computação, tais como Algoritmos e Estrutura de Dados [Nunes e Parreira Júnior 2015, Vahldick *et al.* 2015], Redes [Krassmann *et al.* 2015] e Teoria da Computação [Leite *et al.* 2014].

O objetivo deste trabalho é propor um jogo sério para o ensino de História da Computação chamado Computasseia. O Computasseia é um jogo de cartas de estratégia sobre diversos conteúdos relacionados à história da Computação que pode ser utilizado em diferentes níveis de ensino (técnico, graduação e pós-graduação) a fim de auxiliar na retenção de conteúdos e no engajamento com a disciplina. A partir desta introdução, este trabalho está organizado da seguinte forma: a importância do ensino da História da Computação é discutida na Seção 2; a Seção 3 descreve alguns trabalhos relacionados; o *design* do jogo Computasseia, seus elementos e dinâmica são apresentados na Seção 4; e por fim, as conclusões e trabalhos futuros são traçados na Seção 5.

2. Por que aprender História da Computação?

Fonseca Filho (2007) descreve no único livro de língua portuguesa brasileira (disponível até o presente momento) totalmente dedicado ao assunto de História da Computação, cinco motivações para se estudar História da Computação, são essas:

- **Necessidade de discernir fundamentos:** o ensino de História da Computação é fundamental para compreender que os eventos não estão individualizados, que por muitas vezes são complexos e que influenciam significativamente o saber atual. É preciso possuir um quadro que os reúna para ter-se uma ideia geral, coerente e significativa.
- **Incentivar à educação para a qualidade do software:** o estudo da História da Computação sob a ótica das ideias e conceitos que formaram a base do desenvolvimento da Computação pode contribuir para sensibilizar e entusiasmar o aluno sobre a importância dos fundamentos teóricos, e o desenvolvimento de software e tecnologia de qualidade.
- **Tornar claros e ligar os fatos:** é possível compreender a História da Computação como uma forma de olhar para trás com o intuito de descobrir paralelismos e analogias com a tecnologia moderna e assim, proporcionar uma

base potencial para uma atividade futura ou atual.

- **Acompanhar novas tendências:** o resgate humano e o destaque da História da Computação tem sido valorizados por instituições governamentais e iniciativas privadas em todo o mundo, por meio de homenagens, criação de museus, etc.
- **Revalorizar o fator humano:** toda tecnologia disponível foi desenvolvida por pessoas reais, cada uma acrescentando sua pequena ou grande contribuição. Sem a compreensão do fator humano da Computação, não é possível entender a trajetória histórica da área, nem o seu possível desenrolar futuro.

É possível ainda, acrescentar a possibilidade de atração de novos talentos para a Computação, motivados pelo fator inspirador que o ensino das histórias dos grandes feitos computacionais e personalidades podem promover.

3. Trabalhos Relacionados

Alguns recursos educacionais contemporâneos que vem sendo explorados por educadores como apoio no processo de ensino-aprendizagem de História da Computação são, principalmente, os museus virtuais e as linhas do tempo digitais.

Dalasso *et al.* (2014) apresentam um museu virtual 3D brasileiro sobre História da Computação, desenvolvido através da plataforma de mundos virtuais *OpenSimulator* com galerias que incluem temáticas como Gerações dos Computadores, Personalidades, Linguagens e Pré-História da Computação. Outra iniciativa nacional interessante é o Museu do Computador da Universidade Estadual de Maringá¹ que possui um acervo online sobre personalidades e dispositivos. Alguns exemplos de museus virtuais de História da Computação internacionais são o Museu Virtual da Informática (MVI) [Almeida 2006] e o *Computer History Museum*, que apesar de ser um museu físico localizado em Boston, também disponibiliza um *tour* virtual pelo *site* do museu² e possui uma linha do tempo digital³ com marcos da história da Computação. Outras linhas do tempo digitais interessantes a serem destacadas são as do *Centre of Computing History*⁴ e as do Museu da Computação e Informática (MCI)⁵, que apresenta um acervo de linhas do tempo em diferentes temáticas.

Já na área de jogos sérios para o ensino de Computação, existem diversos trabalhos que utilizam jogos baseados em cartas, tais como: o jogo Simules-W [Monsalve *et al.* 2010] para o ensino de Engenharia de Software; o jogo Baralho das Variáveis para o ensino de princípios de programação [Kahwage *et al.* 2013]; o jogo JEES [Figueiredo *et al.* 2010] para o ensino de Gerência de Configuração, entre outros. Destaca-se também o trabalho de Curzon e McOwan (2008) para ensinar conteúdos de programação e interação humano-computador utilizando jogos e truques com baralhos. O livro clássico “*Computer Science Unplugged*” [Bell *et al.* 1998], que é constantemente atualizado e empregado no ensino, também utiliza cartas em diversas atividades lúdicas para o ensino de conceitos de Computação. Por fim, o *Computing*

¹ <http://www.din.uem.br/museu/museuvirtual.htm>

² <http://www.computerhistory.org/>

³ <http://www.computerhistory.org/timeline/>

⁴ <http://www.computinghistory.org.uk/cgi/computing-timeline.pl>

⁵ <http://www.mci.org.br/>

Research Association juntamente com o *Anita Borg Institute* disponibilizam uma página⁶ de relatos de educadores e pesquisadores sobre o uso de cartas no ensino de Computação muito interessante para consulta.

4. Computasseia

O Computasseia é um jogo de cartas de estratégia desenvolvido, prioritariamente, para dar suporte ao ensino da história da Computação em diferentes níveis de ensino (técnico, graduação e pós-graduação), tendo em vista que os conteúdos abordados e a linguagem trabalhada são abrangentes o suficiente para a inclusão de tal estratégia.

Este jogo foi elaborado como recurso educacional de apoio ao ensino de História da Computação pensando na dificuldade encontrada algumas vezes por educadores ao trabalharem de forma atrativa para os alunos tais temáticas históricas repletas de nomes, fatos e datas importantes. O jogo entra no processo de ensino-aprendizagem como ferramenta lúdica, auxiliando na aprendizagem e retenção de conteúdos pelos alunos.

O nome “Computasseia” foi escolhido para o jogo em referência ao poema grego clássico “Odisseia” de Homero, disponível em edição de língua portuguesa [Loureço 2003], que narra a jornada do herói Odisseu (ou Ulisses para os romanos) na ilha de Ítaca, durante o final da guerra de Tróia. O nome foi escolhido não somente devido à importância histórica da obra para a Literatura Ocidental [Lopes 2013], mas também pelo seu conteúdo, mostrando a longa e desafiadora jornada do herói, aludindo assim ao caminho da Computação para firmar-se como ciência e da trajetória do jogador explorando este caminho.

O Computasseia é composto por um conjunto inicial de 60 cartas distintas, cada uma representando uma parte significativa da história da Computação. Cada carta do jogo pertence a uma Categoria e a uma ou mais Áreas do Conhecimento, a fim de melhor organizar os conteúdos e trazer mais um nível de informação ao jogador, respectivamente. As Categorias e Áreas do Conhecimento são descritas nas subseções a seguir.

O conjunto inicial de cartas do jogo pode ser estendido, para incluir novas cartas no futuro, seguindo a padronização do jogo. É possível, inclusive, desenvolver novos *decks* de cartas temáticos dentro de determinada subárea de uma Área do Conhecimento, para que a temática escolhida possa ser trabalhada de forma mais aprofundada. Todas as cartas e demais informações sobre o jogo Computasseia estão disponíveis no site do jogo⁷.

4.1. As cartas do jogo

4.1.1. Categorias das cartas

Cada carta em Computasseia é categorizada individualmente em uma das seguintes três categorias:

- **Personalidades:** representam pessoas que contribuíram de forma significativa com teorias, invenções, ideias ou princípios para pelo menos uma das Áreas do

⁶ <http://www.cs.duke.edu/csed/wikipedia/cardsusing.html>

⁷ <http://gameinclass.ic.ufmt.br/computassy/>

Conhecimento definidas na seção 4.1.2.

- **Instituições:** representam empresas, associações, sociedades e outras instituições com contribuições significativas para pelo menos uma das Áreas do Conhecimento definidas na seção 4.1.2.
- **Eventos:** representam acontecimentos em geral, tais como invenções, descobertas, reuniões, seminários, entre outros associados a, pelo menos, uma das Áreas do Conhecimento definidas na seção 4.1.2.

As Categorias podem ser observadas na parte superior da visão da frente das cartas (ver item 4 na Figura 1). A Figura 2 apresenta um exemplo de carta do jogo de cada uma das três categorias citadas.



Figura 1. Exemplo de carta do jogo Computasseia e seus elementos indicados

4.1.2. Áreas do Conhecimento

Cada carta em Computasseia é classificada em uma ou mais Áreas do Conhecimento, definidas para melhor organização das cartas do jogo e para incrementar a carta com mais um nível de informação para o jogador. As áreas do Conhecimento foram selecionadas a partir da análise das subáreas da computação por AMC e IEEE (2013) e da organização das comissões especiais da Sociedade Brasileira de Computação⁸.

As Áreas do Conhecimento do jogo Computasseia são:

- **Hardware:** Compreende as tecnologias de desenvolvimento de equipamentos, materiais e demais tecnologias físicas que impactam a Computação. Exemplo: A

⁸ <http://sbc.org.br/403-comissoes-especiais>

invenção dos transistores.

- **Software:** Compreende avanços de tecnologias e teorias de sistemas relacionados ao processo de desenvolvimento de software. Exemplo: Modelos de Análise de Cooper.
- **Teoria Computacional e Matemática:** Avanços teóricos na Computação e em áreas matemáticas de grande influência. Exemplo: Teoria dos Grafos.
- **Computação e Sociedade:** Compreende impactos sociais, culturais e legais causados, alterados ou acelerados pela adoção de novas tecnologias, bem como o contrário, em que particularidades de uma sociedade definiram os rumos da Computação. Exemplos: Primavera Árabe e o Marco Civil da Internet Brasileira.
- **Gestão de Dados:** Compreende os impactos de tecnologias e teorias para gestão de dados e seus processos de armazenamento, transmissão e visualização. Exemplo: A invenção do IBM 350, primeiro disco rígido comercializado.

As áreas designadas são correlacionadas e não se restringem apenas aos casos citados. As cartas podem ser encaixadas em mais de uma Área do Conhecimento, *e.g.* as contribuições de Alan Turing (1912-1954) podem se encaixar nas categorias de Hardware, Software e Teoria Computacional e Matemática. As Áreas do Conhecimento podem ser observadas na parte inferior da visão da frente e de verso das cartas (ver item 3 na Figura 1). A Figura 2 apresenta exemplos de carta do jogo de cada uma das áreas.

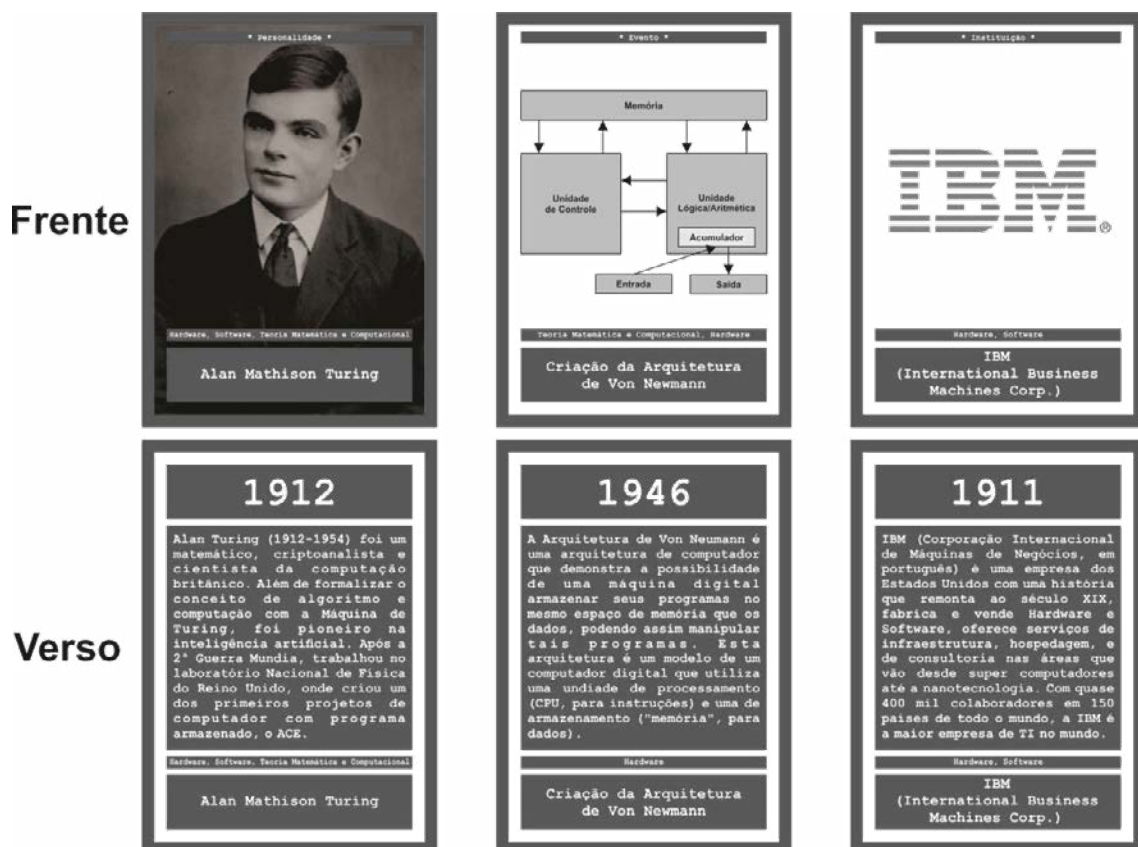


Figura 2. Exemplos de cartas do jogo Computasseia das Categorias Personalidade, Evento e Instituição.

4.1.3. Elementos das cartas e sua organização

As cartas em Computasseia são compostas dos seguintes elementos gráficos que podem ser visualizados na Figura 1:

- **Título (1):** O nome da carta, indicando de forma direta o que a carta representa conforme sua Categoria: para Personalidades, o nome da pessoa representada (também pode conter um apelido ou epíteto se necessário); para Instituições, o nome e a sigla da instituição (se existir); e para Eventos, o nome como o acontecimento é popularmente conhecido.
- **Imagem ilustrativa (2):** Representação imagética da Personalidade, Instituição ou Evento da carta. Esse elemento encontra-se apenas na visão da frente da carta.
- **Área do Conhecimento (3):** Representada por texto, indicando a área ou áreas conforme a subseção 4.1.2.
- **Categoria (4):** Representada por texto, indicando a categoria da carta conforme a subseção 4.1.1. Esse elemento encontra-se apenas na visão da frente da carta.
- **Descrição (5):** Descrição textual que introduz o conceito da carta, complementando a informação do título e expandindo sua contextualização histórica e relevância computacional. Esse elemento encontra-se apenas na visão do verso da carta.
- **Data (6):** Marco temporal que está associado ao fato histórico que a carta representa, de acordo com a sua Categoria: para Personalidades, a data representa o ano de nascimento da pessoa; para Instituições, o ano de sua fundação; e para os eventos, a data exata ou aproximada em que foi realizado. Esse elemento encontra-se apenas na visão do verso da carta.

4.2. Dinâmica do jogo

A dinâmica do jogo Computasseia consiste na elaboração conjunta pelos jogadores de uma Linha do Tempo de marcos da história da Computação, formada pela disposição sequencial em ordem crescente da Data das cartas, constituindo uma cronologia. Essa dinâmica é inspirada no jogo infanto-juvenil Timeline⁹ da Galapagos Jogos.

Cada jogador recebe no início do jogo um conjunto de 6 cartas atribuídas de forma aleatória. As demais cartas restantes do jogo são empilhadas em um montante denominado Montante de Compra com a visão da frente voltada para cima. O Computasseia comporta de 2 a 6 jogadores devido à sua quantidade de cartas atual.

Ao receber as cartas, o jogador que as recebe deve continuar observando suas cartas apenas a partir da visão frontal das mesmas, sem observar o lado das cartas que contém a Data e a Descrição.

Os jogadores se posicionam em volta do local em que será montada a Linha do Tempo, preferencialmente sobre uma mesa. Um por um, em sentido horário ou anti-horário (a ser definido pelos jogadores), o jogador da vez deverá escolher uma das cartas em sua mão para utilizar em sua jogada. Assim que fizer a escolha, deverá tentar inseri-

⁹ <https://www.galapagosjogos.com.br/jogos/timeline>

la na posição que julga ser a correta na Linha do Tempo. Feito isto, a face da carta jogada é virada, revelando suas informações de Data e Descrição contidas no verso. Logo, deve ser verificado o acerto ou erro da jogada.

Caso o jogador tenha errado a posição da carta, essa carta é mantida na mesa, porém a mesma deve ser recolocada na sua posição apropriada na Linha do Tempo e o jogador deve retirar uma nova carta no Montante de Compra para as suas cartas da mão. A vez de jogada é passada ao próximo jogador.

O jogo acaba quando um dos jogadores consegue descartar de forma apropriada todas as cartas de sua mão para a Linha do Tempo do jogo, sendo o mesmo declarado vencedor da partida, ou quando o Montante de Compra se esgota (neste caso, o vencedor da partida é o jogador (ou jogadores) com o menor número de cartas em mão).

4.2.1. Exemplos de jogadas

A Figura 3 ilustra uma mesa com um grupo de jogadores jogando uma partida de Computasseia. O Jogador 1 é escolhido dentre os participantes da mesa para fazer a distribuição das cartas. O Jogador 1 distribuiu as cartas de modo que nenhum jogador (incluindo ele mesmo) possa observar a visão do verso das cartas. A área *a*) da Figura 3 mostra a distribuição de cartas já realizada e sua disposição em relação aos jogadores.

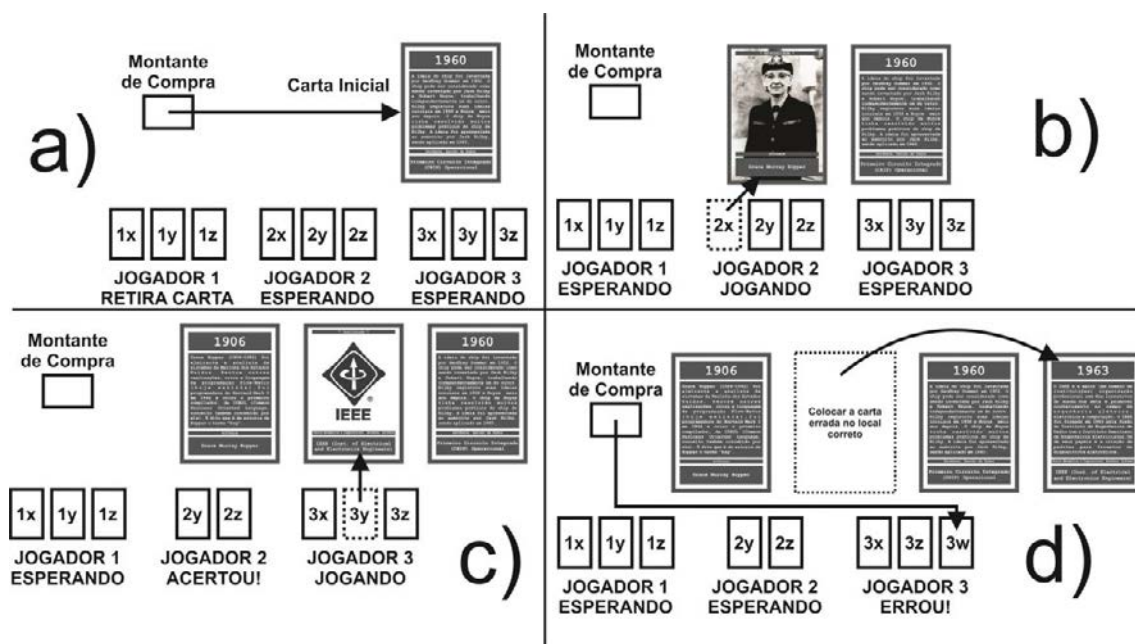


Figura 3. Passos de uma demonstração de jogada

Em seguida o Jogador 1, responsável pela distribuição das cartas, retira uma carta do Montante de Compra para dar início ao jogo, colocando esta Carta Inicial na Linha do Tempo com seu verso virado para cima, exibindo sua Data e Descrição. Isso é feito de maneira a nenhum jogador iniciar o jogo com vantagem, pois a carta inicial sempre estará na posição correta.

Em seguida, o Jogador 2 (o próximo jogador da sequência) será o primeiro a realizar uma jogada de fato. O Jogador 2 escolhe a carta 2x e resolve incluí-la no lado esquerdo da Carta Inicial, como pode ser observado na área *b*) da Figura 3. Depois que

ele fez este movimento, o próprio jogador vira a carta e revela para todos a Data e a Descrição da carta. É possível observar na área *c)* da Figura 3 que o Jogador 2 acertou sua jogada. Assim, sua carta permanece na posição correta e o mesmo possui uma carta a menos em mãos, ou seja, está mais próximo da vitória.

Agora, é a vez do Jogador 3 fazer sua jogada. O jogador 3 escolhe a carta *3y* e decide por colocá-la entre as duas cartas já dispostas na Linha do Tempo. Ao virar a carta a fim de revelar a Data, conforme ilustrado na Figura 3 área *d)*, é possível notar que ele errou a posição correta temporal dos acontecimentos. A carta jogada é, por sua vez, reposicionada na posição correta na Linha do Tempo, e o Jogador 3 compra uma nova carta do Montante de Compra. A vez do jogo é passada para o Jogador 1.

Devido ao seu erro, o Jogador 3 está com o mesmo número de cartas, ou seja, mais longe da vitória do que o Jogador 2 que acertou na sua vez de jogar. Conforme a Linha do Tempo cresce, a dificuldade do jogo aumenta, tendo em vista que o risco de erro é maior devido ao número de locais que os jogadores podem escolher para introduzir a carta. O jogo segue até que um dos jogadores fique sem cartas na mão (esse jogador é declarado o vencedor da partida) ou quando o Montante de Compra se esgota.

5. Conclusões

Este trabalho apresentou um jogo sério para o ensino de História da Computação chamado *Computasseia* que aborda fatos importantes de diferentes áreas do conhecimento da Computação e que pode ser utilizado em diferentes níveis de ensino. A sua dinâmica estratégica estimula o jogador na retenção dos conteúdos das cartas e a ludicidade do jogo incentiva o seu envolvimento com a disciplina. Os autores do presente trabalho acreditam que o jogo de cartas proposto é um recurso acessível aos educadores (podendo ser facilmente impresso no site do jogo e utilizado) e pretendem estender o jogo no futuro adicionando novos elementos de estratégia e criando novos *decks* de cartas temáticas de subáreas específicas para incrementar o jogo atual.

Referências

- ACM e IEEE (2013) “Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science”, ACM, New York, 518 p.
- Baptista, R., e Carvalho, C. V. D. (2013) “TimeMesh—a serious game for european citizenship”, *EAI Endorsed Transactions on Game Based Learning*, 13(01-12), p. 1-7.
- Bell, T. C., Witten, I. H., Fellows, M. (1998) “Computer Science Unplugged: Off-line activities and games for all ages”, *Computer Science Unplugged*.
- Berry, G., Sheard, J., Quartly, M. (2011) “A Virtual Museum of Computing History: an educational resource bringing the relationship between people and computers to life”, In: *Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114*, Australian Computer Society, Inc., p. 79-86.
- Curzon, P. e McOwan P.W. (2008) “Engaging with Computer Science through Magic Shows”, In: *ACM. ITiCSE 2008, The 13th Annual Conference on Innovation and Technology in Computer Science Education*, p.179-183.
- Dalsasso, P. Bos, A. S., Bernardi, G. Müller, F. M. Cordenonsi, A. Z. (2014)

- “Resgatando A História da Computação através de um Museu Virtual 3D Interativo”, In: XXII Workshop sobre Educação em Informática (WEI 2014), Brasília, DF.
- Figueiredo, K., Ferreira, J., Murta, L., Clua, E. (2010) “Jogo de Estratégia de Gerência de Configuração”, In: III Fórum em Educação de Engenharia de Software (FEES), Simpósio Brasileiro de Engenharia de Software (SBES), Salvador.
- Fonseca Filho, C. (2007) “História da computação: O Caminho do Pensamento e da Tecnologia”, EDIPUCRS, Porto Alegre, 205 p.
- Froschauer, J., Arends, M., Goldfarb, D., Merkl, D. (2011) “Towards an online multiplayer serious game providing a joyful experience in learning art history”, In: Games and Virtual Worlds for Serious Applications, IEEE Conference, p. 160-163.
- Haydn, T., Stephen, A., Arthur, J., Hunt, M. (2014) “Learning to teach history in the secondary school: A companion to school experience”, Routledge, New York, 360 p.
- Kahwage, C., de França, E., Nunes, R., Carvalho, R., Souza, D. (2013) “Jogo Baralho das Variáveis”, In: XXI Workshop sobre Educação em Informática (WEI 2013).
- Krassmann, A. L., Falcade, A., Silva, L. E. G., Medina, R. D. (2015) “Serious Games to Computer Networks Learning With CyberCIEGE: A Case Study in Brazilian Higher Education”, In: XXIII Workshop sobre Educação em Informática (WEI 2015).
- Leite, L. S., Sibaldo, M. A. A., Carvalho, T. B. A., Souza, R. (2014) “Montanha de Chomsky: jogo tutor para auxílio no ensino de Teoria da Computação”, In: XXII Workshop sobre Educação em Informática (WEI 2014), Brasília, DF.
- Lopes, J. J. (2013). “A Ilíada e a Odisseia - Dois Pilares da Civilização Grega e Legado Para a Posteridade”, In: Revista Memento, 4(1), p. 118-127.
- Lourenço, F. (2003) “Homero, Odisseia”. Lisboa, Livros Cotovia.
- Michael, D. R., e Chen, S. L. (2005) “Serious games: Games that educate, train, and inform”, Muska & Lipman/Premier-Trade.
- Monsalve, E., Werneck, V., Leite, J. C. S. P. (2010) “Simules-w: Um jogo para o ensino de engenharia de software”, In: III Fórum em Educação de Engenharia de Software (FEES), Simpósio Brasileiro de Engenharia de Software (SBES), Salvador, p. 17-26.
- Neves, I. B. D. C. (2011). “Jogos digitais e potencialidades para o ensino de história: um estudo de caso sobre o history game Tríade–liberdade, igualdade e fraternidade”, Dissertações de Mestrado, Programa de Pós-Graduação em Educação (PPGE), Universidade Federal da Bahia, 243 p.
- Nunes, I. F. e Parreira Júnior, P. A. (2015) “RPG4Sorting - Um Jogo Educacional para Auxílio ao Ensino de Métodos de Ordenação”, In: XXIII Workshop sobre Educação em Informática (WEI 2015), Recife, PE.
- Vahldick, A., Mendes, A. J., Marcelino, M. J., Hogenn, M., Schoeffel, P. (2015) “Testando a Diversão em um Jogo Sérioso para o Aprendizado Introdutório de Programação”, In: XXIII Workshop sobre Educação em Informática (WEI 2015).

Construção e avaliação de um jogo educacional digital de Modelagem de Dados para o ambiente universitário

Stela Sampaio Silva¹, Angelica Toffano Seidel Calazans¹, Bruno Magno da Silva¹

Curso de Análise e Desenvolvimento de Sistemas – Centro Universitário de Brasília
(UniCEUB) – Brasília¹
70790-075 – Brasília– DF– Brasil

stelasampaios@gmail.com,
angelica_toffano@yahoo.com.br, brunno.magnno@gmail.com

Abstract. *The purpose of this research was to develop and evaluate an educational game for the Data Modelling discipline in System Analysis and Development at UniCEUB, to add diversity to the discipline and improve the skills and knowledge of the students. The game was developed in 4 stages: pre-production, production, testing and post-production. Java, JavaScript with Bootstrap, C#, PHP and MySQL database were used. The game evaluation followed the Savi et al. model, and the Goal-Question-Metrics approach to metrics. The game was evaluated by 22 students and results show that: it increased motivation, was regarded as a positive experience and contributed to learning.*

Resumo. *O objetivo foi desenvolver e avaliar um jogo digital educativo de Modelagem de Dados para o curso de Análise e Desenvolvimento de Sistemas do UniCEUB, visando agregar diversidade à disciplina e aprimorar as habilidades e conhecimentos dos alunos. Para produção do jogo foram seguidas 4 fases: pré-produção, produção, testes e finalização. Foram utilizadas as ferramentas JAVA, JavaScript e Bootstrap, C# e PHP e banco de dados MySQL. A avaliação do Jogo foi realizada com base no modelo de Savi et AL, e para construção das métricas utilizou-se o Goal-Question-Metrics. O jogo foi avaliado por 22 alunos e foi identificado que: proporcionou motivação, foi uma experiência positiva e contribuiu para aprendizagem.*

1. Introdução

Os jogos digitais tornaram-se uma forma de entretenimento muito usada tanto por crianças, quanto por jovens e adultos. Esses jogos normalmente criam desafios aos jogadores, os quais devem desenvolver uma maneira de solucioná-los, fazendo com que passem pelo ciclo de observação, questionamento, hipótese e teste (DE FREITAS, 2006). Esse ciclo os estimulam a desenvolverem habilidades como raciocínio lógico, raciocínio estratégico, planejamento, comunicação e capacidade para tomada de decisões (KIRRIEMUIR & MCFARLANE, 2004).

Teorias modernas sobre aprendizado sugerem que o aprendizado é mais efetivo quando envolve uma atividade experimental, com resolução de problema e com um feedback imediato (BOYLE, CONNOLLY & HAINEY, 2011). Os jogos podem proporcionar essa capacidade de aprendizado, além de serem um meio de

entretenimento, por isso, passaram a ser usados em ambientes educacionais, como escolas, universidades e em grandes empresas (BOZSIK, 2013).

Os jogos digitais são usados no ambiente acadêmico por serem mais simples e fáceis de serem produzidos do que os jogos eletrônicos, apesar de sua jogabilidade e realismo ser semelhante (ANTONIETTI, 2002; MIGLINO, 2007). Considerando esses aspectos, alguns questionamentos foram elaborados: os jogos digitais podem facilitar o aprendizado de estudantes que cursam disciplinas muito teóricas com assuntos de alta abstração? Em cursos de graduação como Análise e Desenvolvimento de Sistemas de Informação e Ciências da Computação, um jogo digital pode melhorar o aprendizado, a habilidade dos alunos e tornar o ensino mais agradável?

Para responder a essas questões, a disciplina escolhida para o desenvolvimento do jogo foi a modelagem de dados. As pesquisas realizadas não identificaram a existência de um jogo digital para esse contexto. E foi definido o objetivo geral dessa pesquisa que é construir e avaliar um jogo digital de modelagem de dados para o ambiente universitário, visando agregar mais diversidade ao ensino da disciplina e aprimorar as habilidades e conhecimentos dos alunos.

E são os seguintes objetivos específicos:

1. Identificar e analisar as metodologias para o desenvolvimento de jogos digitais educativos;
2. Pesquisar e analisar as ferramentas para construção de um jogo digital educativo;
3. Construir um jogo digital para o ambiente universitário;
4. Avaliar os aspectos positivos e pontos a melhorar da utilização desse jogo no ambiente universitário;

Para atender a esses objetivos apresenta-se, na seção 2, uma breve descrição dos jogos educacionais focados no ensino superior. Na seção 3, apresenta-se a Metodologia da pesquisa e da produção do jogo, e a análise dos resultados obtidos são apresentados na seção 4. Finalmente, as conclusões são apresentadas na seção 5.

2. Jogos digitais educacionais para o ensino superior

O jogo digital educativo possui uma função lúdica e educativa, sendo que ambas devem coexistir em equilíbrio. O envolvimento ativo do estudante em um jogo educativo acontece através de ações utilizadas nos jogos que provoquem o aluno a interagir com elementos novos, possibilitando que este elabore respostas aos problemas suscitados, superando as contradições entre a sua representação e a realidade (VASCONCELLOS, 2002).

No contexto de jogos universitários focados na área de análise e desenvolvimento de sistemas e/ou Ciência da computação, identificam-se algumas propostas, tais como: Ilha de Requisitos (THIRY, ZOUCAS & GONÇALVES, 2010), Re-O-Poly (RENEL & GOTEL, 2008), U-Test (SILVA, 2010), SPI City (SILVEIRA, THIRY & ZOUCAS, 2013).

Não foram encontrados jogos focados na disciplina de modelagem de dados para cursos de Engenharia de software e Análise e desenvolvimento de sistemas. A disciplina

de Modelagem de dados possui carga horária de 75 horas e tem como objetivo “Introduzir os conceitos fundamentais da natureza e do propósito da modelagem de dados na construção de sistemas de informação, dando ênfase em modelagem relacional”. A seguir apresenta-se a metodologia da pesquisa.

3. Metodologia da pesquisa

Essa investigação utilizou a abordagem qualitativa e quantitativa. No que tange ao alcance temporal, a pesquisa se caracterizou como uma pesquisa interseccional, uma vez que investigou as metodologias existentes para desenvolvimento de jogos digitais educativos e as ferramentas para construção de um jogo digital educativo. Além disso, foi construído um jogo e realizada sua avaliação em determinado período.

O estudo foi baseado no levantamento documental sobre jogos digitais, metodologia, ferramentas, sua aplicação e avaliação no contexto universitário. Foram seguidos os passos definidos por Chandler (2012) para construção do jogo: pré-produção, produção, testes e finalização.

O método adotado para avaliação das vantagens e desvantagens da aplicação desse jogo no ambiente universitário foi o estudo de caso. Segundo Wangenheim et al. (2009), é necessário avaliar a efetividade dos jogos como ferramentas de aprendizagem. A avaliação do Jogo de Modelagem de Dados (JMD) foi realizada com base no modelo específico para jogos educacionais proposto por Savi et al. (2011). Esse modelo é baseado no modelo de avaliação de programas de treinamento de Kirkpatrick (1994) e nas estratégias motivacionais do modelo ARCS (KELLER, 1987).

Segundo Savi et al. (2011), o modelo possui o foco em jogos que possam ser utilizados como material educacional para apoiar o processo de ensino e aprendizagem da Engenharia de Software, avaliando se o jogo motiva os estudantes a utilizarem a ferramenta como material de apoio, se proporciona experiência positiva com os jogadores e principalmente se o jogo educativo proporciona o conhecimento para o conceito relacionado. Os itens da escala de avaliação foram divididos em três sub-escalas: Motivação, Experiência do Usuário e Aprendizagem. O modelo propõe o uso de questionário para avaliar as subescalas.

O questionário foi dividido em duas partes: A primeira parte utiliza o formato de resposta dos itens em uma escala que vai de -2 até +2. Uma nota -2 indica forte discordância com um item e +2 forte concordância. A segunda parte avalia mais pontualmente a aprendizagem, através de itens relacionados aos objetivos de aprendizagem. Nessa avaliação considera-se uma escala de 1 até 5. E o usuário avalia a sua percepção sobre o objetivo de aprendizagem antes e após o jogo. Para essa avaliação foram considerados dois objetivos de aprendizagem: Aprendizagem com relação aos conceitos de entidades, atributos, identificadores e Aprendizagem com relação aos conceitos de relacionamentos e cardinalidades.

Para construir as métricas e indicadores para avaliação do jogo e dos objetivos de aprendizagem, utilizou-se o GQM (Goal-Question-Metrics) (BASILI et al., 1994), conforme a Tabela 1. A abordagem GQM é utilizada sobre a suposição de que para medir de forma objetiva, deve-se especificar quais os objetivos a serem alcançados com as medições estabelecidas. Tais objetivos direcionam a elaboração de questões que,

depois de refinadas, resultam em métricas, cuja aplicação responderá as questões estabelecidas e, conseqüentemente, os objetivos de medição identificados (BASILI et al., 1994). A Tabela 1 apresenta o Objetivo do estudo segundo o GQM e a Tabela 2 apresenta parte das questões avaliadas com as medidas e indicadores criados.

Tabela 1: Objetivo do estudo segundo o GQM

Objetivo 1: Analisar o Jogo de Modelagem de Dados
Propósito: caracterizar
Em relação a: Efetividade de Aprendizagem, Motivação e Experiência do Usuário
Do ponto de vista: dos alunos
No contexto da: disciplina de Modelagem de dados do Curso de Análise e desenvolvimento de sistemas

Tabela 2: Questões avaliadas de efetividade de aprendizagem dos conceitos de entidade, atributos, identificadores e cardinalidades

Questão 1.1: Qual a efetividade da aprendizagem dos conceitos de entidades, atributos, identificadores, relacionamentos e cardinalidades?	
Métrica 1.1a	O objetivo dessa métrica é mensurar a efetividade da aprendizagem com relação aos conceitos de entidades, atributos, identificadores, relacionamentos e cardinalidades
Classificação	Objetiva e quantitativa
Base de medição	Relação entre a nota dada para cada questão e o quantitativo de alunos que avaliaram - Questões do questionário 25, 26 e 27 – Percepção dos alunos
Medidas	Para cada uma das questões 25,26 e 27 $M1.1a_1 = (\sum \text{qtd forte discordância} / \sum \text{qtd respostas}) * 100$ $M1.1a_2 = (\sum \text{qtd discordância} / \sum \text{qtd respostas}) * 100$ $M1.1a_3 = (\sum \text{qtd concordância} / \sum \text{qtd respostas}) * 100$ $M1.1a_4 = (\sum \text{qtd forte concordância} / \sum \text{qtd respostas}) * 100$
Indicador	(M1. 1a_3 + M1.1a_4) para cada questão (25,26 e 27) - de 50% a 70% (o jogo contribuiu para a aprendizagem) - acima de 70% (o jogo contribuiu muito para aprendizagem)

Para avaliar os objetivos de aprendizagem foi utilizado o teste t de *Student* para amostras pareadas. O objetivo foi comparar as MÉDIAS DE DUAS distribuições normais, supondo que se trata da MESMA população, mas em dois momentos diferentes: antes e após a execução do jogo. Segundo Barbetta (2002), “o teste t é apropriado para comparar dois conjuntos de dados quantitativos, em termos de seus valores médios”. Esse teste permite identificar se a tendência identificada poderia ser explicada apenas por efeitos causais ou não.

3. Produção do jogo e análise dos resultados

Para produção do jogo, foi utilizado o ciclo básico proposto por Chandler composto de 4 fases: pré-produção, produção, testes e finalização. Assim, na fase de pré-produção foram definidos objetivos, mecânica do jogo, áreas, entre outros. A história do jogo se inicia com a personagem principal indo para um mundo dos sonhos quando adormece em uma aula de modelagem de dados, neste mundo ela deve enfrentar

desafios que surgem em seu caminho e adquirir conhecimento. A arte conceitual do jogo é baseada em jogos dos anos 80, como Super Mario Bros e Alex Kidd.

A mecânica do jogo foi baseada na curva de aprendizado do jogador e no decorrer do jogo, as questões aumentam de nível assim como a dificuldade do próprio cenário. O jogador é recompensado sempre que acerta uma questão, demonstrando seu aprendizado durante o jogo, enquanto encontra itens de conhecimento e questões que deve responder sobre a disciplina de modelagem de dados. O conceito do jogo é baseado na aquisição de pontos, ou seja, a missão do jogador é adquirir o maior número possível de pontos acertando as questões relacionadas à disciplina de modelagem de dados e finalizar todas as fases do jogo sem ser derrotado pelos inimigos.

O jogo possui três áreas: Área do Administrador que tem o *layout* completo do sistema, com todas as opções de botões. Foi desenvolvido utilizando a biblioteca Bootstrap; Área do professor composta pelas opções de manutenção das questões do jogo e da geração de relatórios que não existem na área do aluno; Área do aluno que é composta apenas pelas opções de ação de jogar o jogo e alterar senha. Na manutenção das questões do jogo o professor pode inserir questões tipo exercícios (objetivas) ou questões tipo estudo de caso, onde cabe ao aluno montar o Modelo Entidade Relacionamento – MER derivado do estudo de caso proposto.

O jogo foi construído em duas interfaces: a primeira interface engloba o cadastro de questões e a administração do ambiente (área do professor e do administrador). Essa interface foi desenvolvida em JAVA, JavaScript e Bootstrap. A segunda engloba o jogo (área do aluno), onde foram implementados tanto as questões tipo exercício quanto as de estudo de caso, e foi desenvolvida em C# e PHP. As duas interfaces utilizam o banco de dados MySQL. A Figura 1 apresenta uma das telas do jogo e a Figura 2 um exemplo do tipo de exercício estudo de caso.

Após o jogo ter sido considerado finalizado, foi implementado em duas turmas do primeiro semestre de 2015 nas disciplinas de modelagem de dados e banco de dados do curso de Análise e desenvolvimento de sistemas do UniCEUB. Apesar dos testes, o jogo apresentou algumas pequenas falhas encontradas pelos alunos enquanto jogavam durante a implantação, o que pode ter causado uma desmotivação de alguns jogadores de continuarem o jogo. No entanto, 22 alunos finalizaram o jogo. Foi possível observar os seguintes resultados de acordo com as subescalas avaliadas:

Subescala Motivação: De modo geral, observa-se que o jogo teve um efeito positivo na motivação dos alunos em grande parte dos itens, como pode ser observado no gráfico de frequências da Figura 3. O jogo foi considerado motivador por 87,02% dos alunos, que atribuíram notas +1 ou +2 aos itens desta subescala. Sendo que 100% dos alunos consideram que foi por se esforçarem que conseguiram avançar no jogo, que atribuíram notas +1 ou +2 a este item, e 95,24% consideram que o conteúdo do jogo é relevante para os interesses deles. No entanto, o item que avalia o *design* do jogo não teve um nível de concordância tão alto, apenas 61,9%, indicando que há espaço para melhorias nesse aspecto do jogo.

Figura 1. Ex. da Interface do jogo JMD



Figura 2. Ex. questão estudo de caso

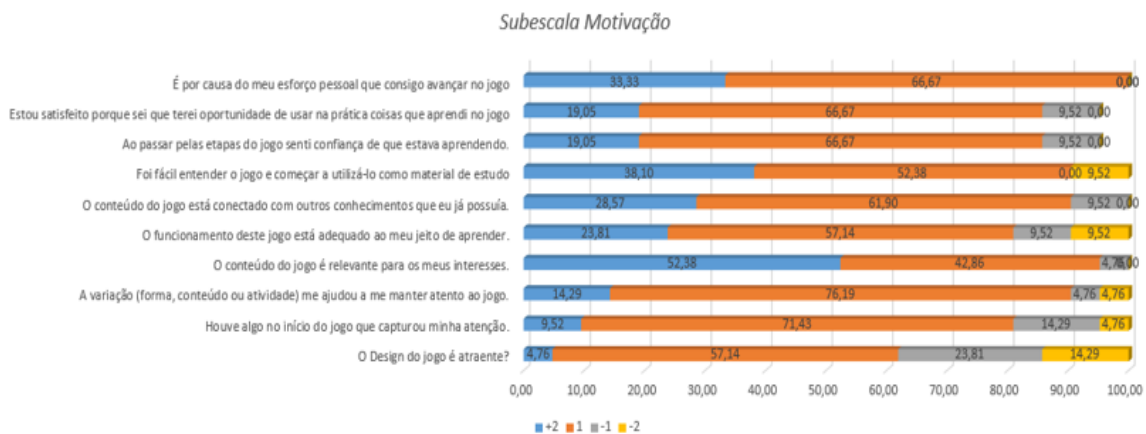


Figura 3. Gráfico de frequências da Subescala Motivação

Subescala Experiência - Observa-se que o jogo proporcionou uma experiência positiva aos alunos, conforme Figura 4. Foi considerado que o jogo proporciona uma experiência ao usuário por 54,39% dos alunos, que atribuíram notas +1 ou +2 aos itens desta subescala. O item melhor avaliado foi o item de dificuldade do jogo, 85,71% dos alunos consideraram que o jogo era adequadamente desafiador sem tarefas muito fáceis ou muito difíceis, atribuindo notas +1 ou +2 a este item, e 71,43% consideram que o jogo evolui em um ritmo adequado, não ficando monótono. Já o item que pergunta se os alunos se sentiram mais no ambiente do jogo do que no mundo real não teve o mesmo grau de concordância, 14,29% dos alunos deram notas +1 ou +2 e a faixa de notas abaixo de zero ficou em 80,96%. Assim como este item, os outros itens sobre imersão no jogo tiveram notas mais baixas, sendo que apenas 41,85% não perceberam o tempo passar enquanto jogavam e 42,86% não conseguiram esquecer suas preocupações do dia-a-dia durante o jogo.

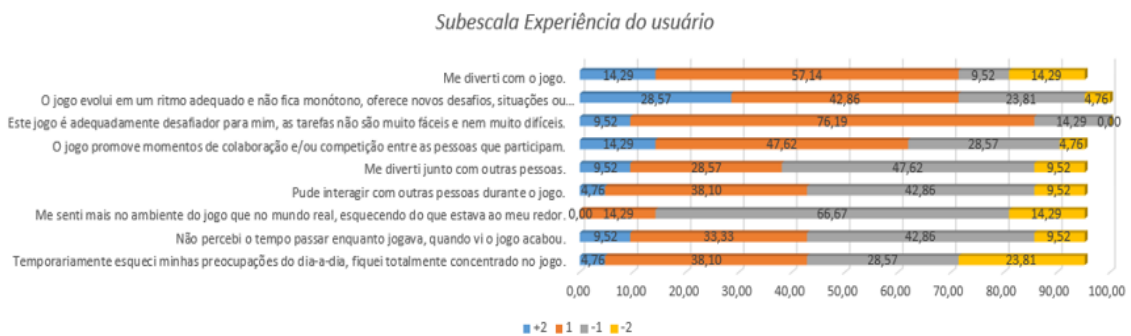


Figura 4. Gráfico de frequências da Subescala Experiência

Subescala Aprendizagem - Ao observar as respostas dos itens indicadores de efetividade da aprendizagem, o jogo contribuiu para a aprendizagem dos alunos, como pode ser observado no gráfico da Figura 5. Sendo que 85,24% dos alunos atribuíram notas +1 ou +2 aos itens desta subescala, consideraram que o jogo contribuiu para sua aprendizagem em relação aos conceitos de entidade, atributo, relacionamento, cardinalidade e identificadores da disciplina de modelagem de dados. Houve um nível de 85,72% de forte concordância de que o jogo foi eficiente para a aprendizagem, quando comparado com outras atividades da disciplina, com 71,43% de notas +1 e 14,29% de notas +2, e 85,71% dos alunos consideraram que o jogo contribuiu para a aprendizagem da disciplina. O item com a menor nota teve uma concordância de 76,19% dos alunos que consideraram que a experiência com o jogo vai contribuir com seu desempenho na vida profissional.

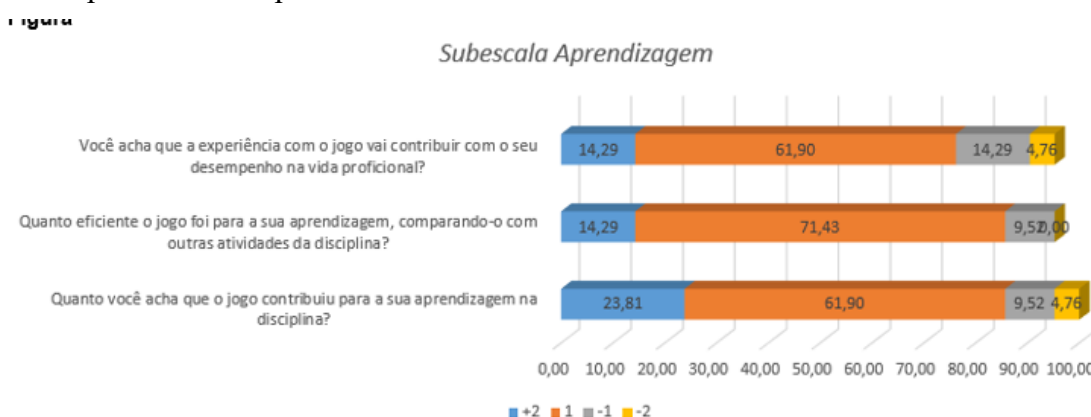


Figura 5. Gráfico de frequências da Subescala Aprendizagem

A Tabela 3 e a Figura 6 apresentam as médias da auto-avaliação dos alunos em relação aos objetivos de aprendizagem nos momentos antes e depois do jogo com base no modelo de Savi et al. (2011). Os resultados foram obtidos por teste t de *Student* para amostras pareadas. O objetivo foi comparar as MÉDIAS DE DUAS distribuições normais, supondo que se trata da MESMA população, mas em dois momentos diferentes: antes e após a execução do jogo. É possível constatar que, os alunos perceberam um aumento de nível de conhecimento em todos os objetivos de aprendizagem do jogo. E os maiores saltos ocorreram nos itens:

- Lembrar Conceitos modelo conceitual (entidade, atributos, identificadores)
- Compreender Conceitos modelo conceitual (entidade, atributos, identificadores)
- Aplicar Conceitos modelo conceitual (entidade, atributos, identificadores)
- Compreender Conceitos modelo conceitual (relacionamento e cardinalidades)

Em todos os tipos de testes t é necessário relatar o valor de "t" chamado de razão crítica assim como do p-valor, pois assim pode-se identificar a veracidade da sua probabilidade (p). Com essas informações pode-se afirmar se a diferença ocorre na realidade ($p < 0,05$) ou se existe apenas ao acaso ($p > 0,05$). Na população examinada foi identificado $p < 0,05$ nos itens listados a seguir, demonstrando dessa forma que, para esses itens, a diferença entre a aprendizagem nos momentos antes e depois do jogo ocorreu na realidade, considerando a população estudada:

2. Compreender Conceitos modelo conceitual (entidade, atributos, identificadores) - 0,0373925000135277

3. Aplicar Conceitos modelo conceitual (entidade, atributos, identificadores) - 0,0261449195292569

Tabela 3. Médias da auto-avaliação dos alunos em relação aos objetivos de aprendizagem nos momentos antes e depois

Perguntas	Momento	Média	Variancia	Diferença antes-depois
Lembrar Conceitos modelo conceitual (entidade, atributos, identificadores)	Antes	3,4	0,98947	0,45
	Depois	3,85	0,87105	
Compreender Conceitos modelo conceitual (entidade, atributos, identificadores)	Antes	3,3	0,85263	0,55
	Depois	3,85	0,87105	
Aplicar Conceitos modelo conceitual (entidade, atributos, identificadores)	Antes	3,47	0,98642	0,475
	Depois	3,945	0,99734	
Lembrar Conceitos modelo conceitual (relacionamento e cardinalidades)	Antes	3,7325	0,7206	0,1625
	Depois	3,895	0,93629	
Compreender Conceitos modelo conceitual (relacionamento e cardinalidades)	Antes	3,55	0,68158	0,35
	Depois	3,9	0,83158	
Aplicar Conceitos modelo conceitual (relacionamento e cardinalidades)	Antes	3,3	0,95789	0,3
	Depois	3,6	1,2	

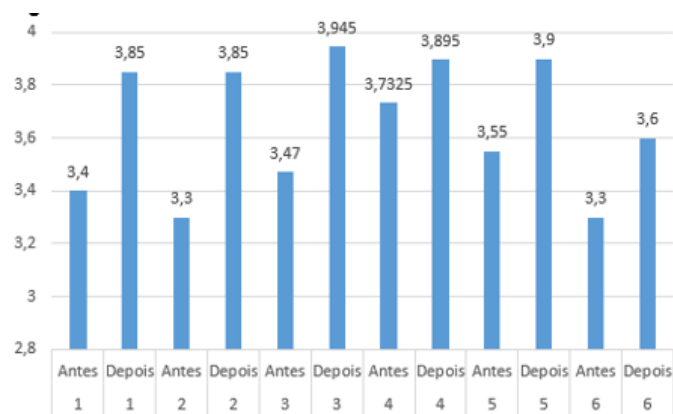


Figura 6. Médias da auto-avaliação dos alunos em relação aos objetivos de aprendizagem nos momentos antes e depois do jogo

Legenda:

1. Lembrar Conceitos modelo conceitual (entidade, atributos, identificadores)
2. Compreender Conceitos modelo conceitual (entidade, atributos, identificadores)
3. Aplicar Conceitos modelo conceitual (entidade, atributos, identificadores)

4. Lembrar Conceitos modelo conceitual (relacionamento e cardinalidades)
5. Compreender Conceitos modelo conceitual (relacionamento e cardinalidades)

4. Conclusões e estudos futuros

Na educação, os docentes encontram cada vez mais alternativas na tecnologia para melhorar a forma como repassar conhecimentos e desenvolver as habilidades de seus alunos. Os jogos digitais surgiram como uma forma de entretenimento entre jovens no início da popularização dos computadores e videogames, mas esta realidade tem mudado nos últimos anos e os jogos passaram a serem vistos como uma forma de auxiliar professores em sala de aula.

Este trabalho teve como objetivo desenvolver e avaliar um jogo digital educativo de Modelagem de Dados para o curso de graduação em Tecnologia de Análise e Desenvolvimento de Sistemas visando agregar mais diversidade ao ensino da disciplina e aprimorar as habilidades e conhecimentos dos alunos.

Para atender aos objetivos geral e específicos foram pesquisados os jogos educacionais para ambiente universitário e a aplicabilidade do jogo em disciplinas do Curso de Análise e Desenvolvimento de Sistemas. Foram pesquisadas as metodologias e ferramentas existentes para a criação de um jogo. Após essas pesquisas, foi definido o foco do jogo como sendo a disciplina de Modelagem de Dados e a utilização do ciclo básico proposto por Chandler para produção do jogo, composto de 4 fases: pré-produção, produção, testes e finalização. Foram utilizadas as ferramentas já citadas e o jogo foi avaliado na percepção de 22 alunos da disciplina considerando as dimensões de Experiência, Motivação e Aprendizagem. Os resultados foram promissores e foram identificados alguns aspectos de melhorias.

Como trabalhos futuros, é interessante que novas turmas de modelagem de dados utilizem o jogo e o avaliem novamente, uma vez que o jogo, foi evoluído no 2o. semestre de 2015 para uma nova *release* com implementações de melhorias em suas funcionalidades e na interface. Além disso, será interessante também realizar avaliações de desempenho (com a ferramenta e sem a ferramenta) para tentar mensurar os ganhos com a mesma.

Referências

- Antonietti, A., Rasi, C. and Underwood, J.. I videogiochi: una palestra per il pensiero strategico? “Rivista: Ricerche di Psicologia”. Fascicolo: 1, p.125-144, 2002.
- Barbetta, P. A. “Estatística aplicada as ciências sociais”. 5ª. Ed. –Florianópolis: Ed. Da UFSC, 2002.
- Basili. V., Caldiera, G.and Rombach, H. Goal Question Metric paradigm. In: “Encyclopédia of Software Engineering”. V. 2, 1994. p. 527 –532.
- Boyle, E. A., Connolly T. M. and Hailey, T. The role of psychology in understanding the impact of computer games. “Entertainment Computing”, 2, 69–74, 2011.
- Bozsik S. Business simulation game as a tool of practice-oriented education. “Practice and Theory in Systems of Education”, Volume 8, Number 2, p. 126-134, 2013.

- Chandler, H. M. “Manual de Produção de Jogos Digitais”, 2^a edição. Bookman, VitalBook file, 2012.
- De Freitas, S. Learning in immersive worlds. “Joint Information Systems Committee”, 2006.
- Kirriemuir, J. and Mcfarlane, A. Literature Review in Games and Learning. “A NESTA Futurelab Research report” - report 8,. 2004.
- Keller, J.M. Development and use of the ARCS model of instructional design. “Journal of instructional development”. Volume 10, Issue 3, pp 2-10, September, 1987
- Kirkpatrick, D. L.,. “Evaluating Training Programs - The Four Levels”. Berrett-Koehler Publishers, 1994.
- Miglino, O.; Ferdinando A. D. and Rega, A. Le nuove macchine per apprendere: simulazioni al computer, robot e videogiochi multi-utente. “Rivista Sistemi intelligenti”. Fascicolo 1, aprile p. 113-135, 2007.
- Renel, S. and Gotel, O. Gameplay to Introduce and Reinforce Requirements Engineering Practices, 16th IEEE “International Requirements Engineering Conference”, Barcelona, Spain, IEEE Computer Society Press. 2008.
- Savi, R.; Wangenheim, C. and Borgatto, A. Análise de um modelo de avaliação de jogos educacionais. Disponível em: <https://sites.google.com/site/savisites/avaliacao-de-jogos-educacionais>, 2011.
- Silva, A. C. “Jogo educacional para apoiar o ensino de técnicas para elaboração de testes de unidade”. Dissertação (Mestrado) - Curso de Computação Aplicada, Universidade do Vale do Itajaí, São José, 195 f. , 2010.
- Silveira, J. L, Thiry, M. and Zoucas, A. SPI City: Jogo Educacional para Apoiar o Ensino de Melhoria de Processo de Software. In: “XII Simpósio brasileiro de qualidade de software”. Salvador, Bahia – Brasil. 01 a 05 de Julho, 2013.
- Thiry, M., Zoucas, A and Gonçalves, R. Q. Promovendo a Aprendizagem de Engenharia de Requisitos de Software Através de um Jogo Educativo. In: “XXI SBIE - Simpósio Brasileiro de Informática na Educação”, 2010, João Pessoa. 2010.
- Vasconcellos, C.S. “Construção do conhecimento em sala de aula”. São Paulo: Libertad, 2004.
- Wangenheim, C., Kochanski, D. and Savi, R. Revisão Sistemática sobre Avaliação de Jogos Voltados para Aprendizagem de Engenharia de Software no Brasil. In: “Anais do FEES 2009/ XXIII SBES”, pp. 41-48, Fortaleza, Brasil. 2009.

Currículos de Computação: porque permanecem assim?

Miguel Jonathan

Departamento de Ciência da Computação /IMAT e Programa de História das Ciências e das Técnicas e Epistemologia (HCTE) – Universidade Federal do Rio de Janeiro (UFRJ)
Rua Sambaíba, 478 ap.401 22450-140 – Rio de Janeiro – RJ – Brasil

migjonathan@gmail.com

***Abstract.** Undergraduate Computer Science programs in Brazilian universities have been historically focused in mathematics, physics and computing science and technology subjects. Mandatory content of inter or multidisciplinary nature designed to give the graduates an awareness on the interactions of these technologies and systems with society represent less than 5% of the total. In contrast, computing job opportunities in Brazil are turning increasingly to build applications that directly or indirectly affect the lives of people, individually and in society. This article analyzes the reasons for the persistence of this training model 40 years after its introduction.*

***Resumo.** A formação universitária em Ciência da Computação no Brasil tem sido focada historicamente em matemática, física e ciência e tecnologia da computação. Conteúdos obrigatórios de natureza inter ou multidisciplinar destinados a preparar o egresso para refletir sobre a interação dessas tecnologias e sistemas com a sociedade representam menos de 5% do total. Em contraste, as oportunidades de trabalho em computação no Brasil se voltam cada vez mais para a construção de aplicações que afetam direta ou indiretamente a vida das pessoas, individualmente e em sociedade. O artigo analisa os motivos da persistência desse modelo de formação 40 anos após a sua introdução.*

1. Introdução

Neste ensaio, abordamos os currículos dos cursos de graduação em ciência da computação, como surgiram e evoluíram, com atenção especial para as redes de atores que contribuíram para a conservação, ao longo dos últimos 40 anos, de um formato e conteúdo aproximadamente semelhantes nas diversas universidades brasileiras, apesar das profundas transformações que ocorreram nas relações da computação com quase todos os ramos da atividade humana nesse período.

O texto é uma reflexão crítica sobre a continuidade da absoluta predominância de conteúdos técnicos e científicos nos currículos em face das crescentes responsabilidades dos egressos com as transformações organizacionais, culturais e econômicas que a sua atividade vem promovendo ao longo das últimas décadas.

Seguindo uma metodologia inspirada nos estudos de ciência, tecnologia e sociedade (CTS), em particular a teoria ator-rede [Latour 2007], procuramos seguir os passos dos atores que produziram os currículos, as circunstâncias que contribuíram para as decisões,

as alianças que se formaram para produzir consenso em torno de uma forma final, e as condições que permitem uma relativa estabilização dos conteúdos.

Nas seções a seguir apresentamos um histórico resumido da formação dos profissionais de computação no Brasil, a partir da introdução dos primeiros computadores, a criação de cursos de tecnologia pela iniciativa da Capre, a entrada em cena dos programas de pós-graduação das grandes universidades, a criação dos cursos de graduação plena, a formação da SBC e seu papel decisivo na consolidação de um modelo curricular hegemônico com a edição de currículos de referência e uma aliança com o Ministério da Educação.

Em outra vertente complementar, o texto reflete criticamente sobre os efeitos da organização universitária, especialmente as públicas, onde a estrutura departamental contribui para a preservação de currículos uni-disciplinares e onde a burocracia e os critérios de avaliação inibem uma maior dedicação à melhoria do ensino de graduação.

2. Seguindo o fio da meada

Computadores foram introduzidos no Brasil no início da década de 1960. Tratava-se de conhecimento novo, ainda incipiente mesmo nas melhores universidades. Dado o grande interesse dos fabricantes internacionais em comercializar estes equipamentos no Brasil, estes tomaram a si a tarefa de formar os primeiros profissionais, através da criação de centros de treinamento próprios.

Os novos profissionais precisavam dominar o entendimento e a operação de máquinas extremamente complexas antes que fosse possível desenvolver qualquer aplicação útil. Isso levou a que a profissão fosse constituída principalmente por pessoas com formação em ciências naturais e engenharia, muito embora a maioria dos problemas tratados por computadores fossem tipicamente da área de humanas (administração, finanças, faturamento) e suas soluções afetassem o trabalho e o relacionamento de pessoas e organizações. O treinamento oferecido pelos fabricantes enfatizava a compreensão do funcionamento da máquina e a capacidade de dominar linguagens de programação baseadas em lógica.

No início da década de 1970 já existiam no Brasil organizações de grande porte com tradição na utilização de computadores, como Serpro, Petrobrás e as concessionárias de serviços públicos. Possuíam corpo técnico experiente, formado pelos fabricantes, mas já se sentiam capazes de formar seus próprios quadros, através de programas internos de treinamento. Típicos desse período foram os programas como os do Centro Nacional de Treinamento do SERPRO que recrutavam engenheiros, matemáticos, físicos e até administradores recém-formados, para transformá-los em analistas de sistemas ao longo de um curso intensivo de um ano [Lobato 1982,p.61].

Com a expansão acelerada do uso de computadores para atividades de processamento de dados, o governo cria em 1972, no Ministério do Planejamento, a CAPRE - Coordenação das Atividades de Processamento Eletrônico. Entre as suas atribuições está a de formular um Programa Nacional de Ensino de Computação, visando suprir em prazo relativamente curto a forte demanda prevista de profissionais capacitados para suprir o mercado de trabalho em processamento de dados administrativos nos próximos anos. Com o apoio do MEC e de várias universidades, a CAPRE organiza com bastante

sucesso o Projeto 19 para a criação de cursos de curta duração (máximo de 3 anos), denominados Cursos Superiores de Tecnologia em Processamento de Dados, também conhecidos como cursos de graduação curta. Cursos estes, por definição, destinados a formar profissionais capacitados a desenvolver aplicações de computadores de complexidade pequena e média, que respondiam então pela grande maioria da demanda de pessoal.

É interessante notar que o currículo e materiais didáticos dos cursos de tecnologia foram preparados por equipes convocadas nas universidades, sob a coordenação da CAPRE, mas procurando ouvir as sugestões e críticas dos empregadores dos futuros egressos, como mostra esse extrato de uma reunião realizada em 28/4/1974 em São Paulo, que envolveu 45 organizações entre a CAPRE, empresas e instituições de ensino, públicas e privadas (grifos nossos):

"A agenda dessa reunião constava de duas fases: ... a segunda, pela tarde ... com a finalidade de detetar os desejos das empresas com relação ao perfil do profissional a ser formado pelos cursos do Projeto 19 em processamento de dados... Para auxiliar e objetivar a formação do profissional voltada para as necessidades do mercado, o Prof. Ziviani aponta algumas atitudes que são da maior importância: ...- debates e sugestões das empresas sobre o currículo adotado." [CAPRE 1974b]"

O currículo resultante que foi aprovado pelo Conselho Federal de Educação, inclui matérias de tecnologia (formação e especialização em programação e uso de computadores) mas também dá espaço razoável para matérias que preparam o egresso para atuar profissionalmente, incluindo Economia e Finanças, Noções de Direito, Inglês e outras “visando o aprimoramento sócio-cultural do futuro profissional em processamento de dados” [CAPRE, 1974a].

3. As universidades entram em cena – nasce a SBC

As universidades, no começo apenas as maiores, começaram a atuar em computação principalmente pela via da pós-graduação. Após a implantação do regime militar em 1964, uma forte ênfase foi dada ao desenvolvimento em ciência e tecnologia. Os primeiros programas foram criados na UFRJ (COPPE 1968), PUC-Rio (1968), USP, Unicamp, UFMG, UFRGS (1973) e outras. O governo fomentou a formação de doutores no exterior com bolsas de estudo custeadas pelo Fundo de Desenvolvimento Tecnológico do Banco Nacional de Desenvolvimento Econômico (FUNTEC/BNDE). Em meados da década de 1970 já retornavam ao Brasil vários mestres e doutores em ciência e engenharia de computação.

Com o retorno dos primeiros pesquisadores, os programas de pós-graduação ganham impulso e em poucos anos esse contingente relativamente pequeno começa a se articular como um grupo com interesses comuns, voltados para a pesquisa científica, o ensino, e o desenvolvimento de computadores e da computação no Brasil. Na política, há um movimento em busca do apoio do governo para a implantação de computadores nas universidades e para uma indústria de computadores genuinamente brasileira. Para promover essas ideias, a comunidade de pesquisa organiza a realização dos Seminários de Computação nas Universidades (SECOMU), cuja primeira edição se dá em 1971, no

Rio Grande do Sul, organizado por professores da PUC-Rio e com o apoio do Conselho de Reitores (CRUB) [Dantas 1988,p.88]. Na pesquisa e desenvolvimento, os pesquisadores organizam em janeiro de 1974 o primeiro SEMISH - Seminário sobre o Desenvolvimento Integrado de Software e Hardware, sediado pelo Curso de Pós-graduação em Ciência da Computação da UFRGS [Oliveira 2016]. É possível inferir que, nesse momento, a prioridade seria a tecnologia, e não o que os atores sociais fariam com ela.

Data desse período os projetos de criação de cursos de graduação plena em computação nas grandes universidades. A UFRJ, UFMG, PUC-Rio e UFRGS criam seus cursos de graduação, a exemplo das iniciativas pioneiras da Unicamp e da UFBA em 1969. Esses cursos, com denominações diversas como Ciência da Computação, Engenharia da Computação e Informática, duração de 4 a 5 anos, e currículos semelhantes, foram idealizados por pesquisadores e buscavam formar egressos com forte conteúdo matemático, científico e tecnológico. Houve desde o início um distanciamento claro entre os objetivos desses novos cursos e aqueles dos cursos de Tecnologia em Processamento de Dados. Os cursos de graduação plena enfatizavam o estudo dos fundamentos da ciência da computação e das tecnologias dos computadores, e abriam possibilidades para seus alunos prosseguirem tanto em estudos avançados de pós-graduação, como para atuarem em projeto e desenvolvimento de aplicações e na nascente indústria brasileira de computadores dos anos 70.

Em 1978, em uma reunião histórica do SEMISH realizada na UFRJ, a comunidade acadêmica de computação considerou-se madura o suficiente para propor a criação de uma nova sociedade científica. Nascia assim a Sociedade Brasileira de Computação. [Cabral 2008,p.17]. O SEMISH e o SECOMU, antes eventos independentes, passam a integrar o Congresso Anual da SBC.

4. A SBC atua na formação

A SBC logo se preocupou com a rápida proliferação de cursos de graduação em computação que ocorria em todo o país nos anos 80, incluindo os de curta duração, motivada pela crescente procura de profissionais. Havia uma percepção de que muitos poderiam ser de qualidade duvidosa. Como relata Maria Izabel Cabral, os congressos anuais passaram a incluir uma reunião geral de coordenadores de cursos de graduação a partir de 1986, visando fomentar o debate sobre a formação e os currículos, e obter um consenso sobre o que deveria ser uma formação em computação. Em 1987, a sociedade organizou a edição de um catálogo dos cursos em todo o país, e lançou a iniciativa de elaborar um currículo de referência. Em 1990, no congresso de Vitória, a SBC constituiu a sua primeira Comissão de Ensino, embrião da atual Diretoria de Educação. A comissão, após diversas consultas à comunidade, e com apoio financeiro da Secretaria de Ensino Superior do MEC (SESu/MEC) produziu o Currículo de Referência que obteve sua aprovação na reunião de coordenadores de cursos realizada no congresso de 1991, referendada pela Assembleia Geral, em Santos, SP. [Cabral 2008, p.37].

O CR91, como ficou conhecido, reflete o pensamento dominante da comunidade acadêmica de computação das grandes universidades públicas e confessionais que constituía o núcleo da SBC. Desenvolvido por uma comissão composta por especialistas doutores em computação, e sem consulta a representantes do mercado empregador,

como havia feito a Capre, não estranha que o currículo se concentre nas matérias de ciência e tecnologia, e dedique atenção marginal a questões humanas e sociais, sejam culturais, políticas, legais, organizacionais, psicológicas, econômicas ou filosóficas, que certamente fariam parte do universo dos egressos desses cursos. Para atender a essas questões, o CR91 apenas menciona, na seção referente aos Objetivos dos Currículos, que “*Matérias de cultura geral devem ser incorporadas para prover uma formação humanística*” e “*Matérias de outras áreas podem ser introduzidas com vistas a formar pessoal especializado para participar na solução computacional de problemas que afetam cada uma dessas áreas*”. Direito, Economia e Administração apenas são sugeridas como matérias complementares [SBC 1991].

De fato, uma análise das matérias recomendadas pelo CR91 para cursos de Ciência da Computação e Engenharia de Computação mostra que aquelas consideradas essenciais incluíam: Álgebra, Lógica Matemática, Cálculo Diferencial e Integral, Álgebra Linear, Probabilidade e Estatística, Análise Numérica, Teoria da Computação, Linguagens Formais e Autômatos, Análise de Algoritmos, Máquinas Sequenciais e Combinacionais, Arquitetura e Organização de Computadores, Software Básico, Algoritmos e Estruturas de Dados, Paradigmas de Linguagens de Programação, Paralelismo, Banco de Dados, Sistemas Operacionais, Compiladores, Redes e Transmissão de Dados, Sistemas Distribuídos, Engenharia de Software e Projeto de Sistemas Digitais [SBC 1991].

Fica claro que neste primeiro currículo de referência a SBC postulava uma formação sem conexão com as ciências humanas e sociais para os futuros profissionais que teriam a missão de idealizar e produzir os sistemas computacionais que iriam influenciar profundamente as relações humanas e organizacionais do país nas próximas décadas.

O CR91 entra em cena como um forte aliado da visão tecnocientífica da formação em computação. Amplamente distribuído para todas as universidades com a chancela do porta-voz da comunidade científica, o CR91 tornou-se efetivamente um “*ponto de passagem obrigatória*” no sentido atribuído por Bruno Latour: “*o fato [ou artefato] construído se torna indispensável*” [Latour 2000, p.218]. De fato, a partir da aprovação, o currículo de referência passou a ser aceito sem maiores discussões pelos proponentes de novos cursos, o que pode ter contribuído para inibir visões alternativas de como formar os profissionais de computação.

A SBC continuou a influenciar fortemente a formação universitária em computação ao longo da década de 90. Em 1993, o Prof. Daltro Nunes, presidente da Comissão de Ensino, propôs a realização de um workshop para estender a toda a comunidade os debates que já ocorriam nas reuniões de coordenadores de cursos de computação durante o congresso anual. O Workshop sobre Educação em Computação / Informática foi realizado nos dias 7 e 8 de setembro de 1993, em Florianópolis, marcando o início de uma série que continua até hoje [Cabral 2008, p.37]. Na pauta, discussão sobre currículo de referência e infra-estrutura necessária para cursos de graduação.

Uma aliança tácita entre a SBC e a Secretaria de Ensino Superior do MEC (SESu) levou à formação da Coordenação de Especialistas de Ensino de Computação e Informática (CEEInf) da Diretoria de Ensino Superior da SESu/MEC, em 1994, que foi desde logo constituída por membros atuantes da SBC sob a liderança de Daltro Nunes [CEEInf, s/data]. À CEEInf coube diversas missões que impactaram na evolução dos cursos de graduação em todo o país, especialmente os novos que iam sendo criados em todas as

regiões: conceituar e produzir os primeiros Padrões de Qualidade para avaliação dos cursos da área de Computação e Informática, organizar as comissões de avaliação dos cursos, homologar as avaliações e, mais adiante, coordenar os trabalhos para elaboração das primeiras Diretrizes Curriculares para a área em 1999, sempre em sintonia com a SBC. Paralelamente, a Comissão de Ensino, e sua sucessora, a Diretoria de Educação, continuaram a rever as recomendações curriculares, produzindo os currículos de referência de 1996 [SBC 1996], de 1999 [SBC 1999], e de 2005 [SBC 2005], as Diretrizes Curriculares de 1999 [BRASIL 1999] e, mais recentemente, as Diretrizes Curriculares de 2012 [BRASIL 2012].

Os Padrões de Qualidade embutiram a essência do CR91 no quesito currículo. As comissões nomeadas pela SESu/MEC para verificação e avaliação dos cursos de graduação eram orientadas pela CEEInf para exigir a conformidade com a essência do currículo de referência. A CEEInf, em parceria com a SBC, contribuiu decisivamente para a consolidação dos WEI's, e estimulou a criação dos Cursos de Qualidade a partir de 1999, que passaram a ser realizados desde então em conjunto com os WEI's.

O CR96, aprovado no Congresso de agosto de 1996, em Recife, introduz as matérias de Interfaces Homem-Máquina, no tópico Tecnologias da Computação, reconhece que o egresso deve possuir, entre as suas características, uma *“formação humanística permitindo a compreensão do mundo e da sociedade”* e inclui no tópico Contexto Social e Profissional as matérias de Computadores e Sociedade e de Formação de Empreendedores de Informática, as quais, no entanto, não são incluídas entre as rotuladas de “Importantes”.

O CR99, para cursos de Computação e Informática, procurou ficar “em sintonia” com as Diretrizes Curriculares de 1999, ambas abrangendo tanto cursos como atividade fim e como atividade meio. Enquanto o CR99 foi elaborado no âmbito da Diretoria da Educação da SBC, e aprovadas no WEI de julho de 1999, as Diretrizes, elaboradas pela CEEInf, foram finalizadas em março do mesmo ano. Ambas tiveram a colaboração intensa da comunidade acadêmica de computação, com muita interseção. No item “Formação Humanística” das diretrizes aparecem pela primeira vez as matérias de História da Ciência da Computação (porque não da computação em geral?), Ética, Sociologia e Filosofia, além de Computadores e Sociedade e Empreendedorismo. Não há sugestão de um percentual de horas para essa área de formação, nem como devem se relacionar com o restante do curso, apenas a recomendação que *“As disciplinas devem cobrir essa matéria de forma geral”*.

O CR2005, que se ateu aos cursos de Ciência e Engenharia (atividade fim), indica que o perfil dos egressos, em seus “Aspectos Gerais”, deve incluir uma *“formação humanística permitindo a compreensão do mundo e da sociedade”*, além de *“habilidades de trabalho em grupo e de comunicação e expressão”*. E em seus “Aspectos Ético-Sociais” assinala que os egressos devem *“Respeitar os princípios éticos da área de computação”* e *“Ter uma visão humanística crítica e consistente sobre o impacto de sua atuação profissional na sociedade”*. O currículo inclui um núcleo de matérias de *Contexto Social e Profissional* que lista 14 matérias. É interessante observar que há uma diretiva de recomendar 30 créditos para esse núcleo, de um total de 160 a 200 créditos previstos para todo o curso, ou cerca de 15%.

Finalmente, as novas Diretrizes de 2011/12 resultaram de uma extensa revisão das Diretrizes de 1999 (nunca homologadas formalmente) realizada por um grupo de trabalho formado por membros tradicionais da SBC e da Diretoria de Educação, coordenado pelo incansável Daltro Nunes. Elas expressam uma conscientização bem maior sobre a necessidade de se ampliar a formação nos cursos de computação para além dos conteúdos técnicos e científicos. Em um extenso trecho dedicado à *Formação Humanística e Social* o texto reconhece que “*A Computação permeia praticamente todas as atividades humanas, incluindo trabalho, lazer, saúde e comunicação, cabendo aos profissionais da Área a responsabilidade pelo desenvolvimento de soluções, ferramentas e processos coerentes com a moral, bons costumes, valores éticos e interesse social, e que também busquem o bem-estar do homem e o avanço tecnológico*”. E mais, considera *indispensável* os estudos de História da Computação, do empreendedorismo, das questões éticas, dos impactos da automação na sociedade, da Filosofia e da Sociologia, incluindo mudanças na organização do trabalho provocadas pelas inovações tecnológicas e a “*compreensão e o questionamento do mundo tecnológico e do mundo sociocultural que o circunda*”. E conclui enfatizando que “*O estudo integral da Computação transcende as questões meramente técnicas, exigindo a compreensão do processo de construção do conhecimento*”. [BRASIL 2012]

5. Obstáculos a uma formação interdisciplinar

Nas seções anteriores apresentamos a emergência da formação típica em ciência e engenharia da computação nas nossas universidades voltada quase que totalmente para os estudos de ciência e tecnologia. Apesar da grande maioria dos egressos desses cursos se tornarem profissionais responsáveis por desenvolver artefatos e sistemas com potencial de afetar diretamente as atividades humanas, muitos currículos atuais ainda permanecem essencialmente semelhantes em conteúdo aos primeiros currículos de décadas atrás. Vimos como as diretrizes e recomendações curriculares foram aos poucos abrindo espaço para uma formação mais interdisciplinar ao reconhecer o caráter multifacetado do profissional de computação.

Existem, no entanto, outros obstáculos para que transformações nesse sentido sejam de fato realizadas.

A Reforma Universitária de 1968 [BRASIL 1968] instituiu a descentralização das competências acadêmicas, ao criar os institutos e departamentos especializados nas universidades públicas [Martins 2009]. Essa forma de organização acadêmica horizontalizada procurou reunir os pares em estruturas homogêneas de conhecimento, reunindo pesquisa e ensino, visando uma maior interação e produtividade na produção acadêmica. Na forma anterior verticalizada, as unidades se organizavam pelas suas finalidades de formação: cada escola reunia no seu corpo docente especialistas de diversas áreas para organizar o ensino dos seus alunos. As escolas de engenharia, por exemplo, contratavam matemáticos, físicos, químicos, professores de línguas, filosofia, e o que mais necessitassem. Com a reforma, cada área de conhecimento ficou mais compartimentalizada em seu departamento.

O caso específico da computação é emblemático, pois os primeiros cursos de computação já foram criados na nova estrutura, que permanece até os nossos dias.

Docentes contratados por critérios de qualificação para a pesquisa acadêmica em determinadas especialidades são reunidos em departamentos de computação, e passam a responder na prática pelos currículos dos cursos. Muitos passam diretamente de uma formação puramente acadêmica em computação para a atividade docente, sem terem tido experiência profissional e conhecimento do mercado de trabalho onde seus alunos irão atuar. É nesse ambiente que são discutidos e modificados os currículos dos cursos de graduação. Ambiente que é naturalmente pouco permeável a promover uma formação integral do aluno para além das especialidades do grupo, levando os currículos a continuarem mais ou menos estáveis.

Esses fatores contribuem para inibir mudanças em profundidade nos currículos, onde o ensino das disciplinas reflete uma visão idealizada da ciência e da tecnologia como conhecimento universal, neutro e desinteressado, que são apresentadas sem perspectiva histórica ou local, e sem relacionamento com questões culturais, econômicas e sociais. Não há preocupação em estimular o pensamento crítico dos alunos e formar juízos de valor sobre as consequências do uso das tecnologias de computação nas diversas esferas da sociedade. Disciplinas como Computadores e Sociedade são uma exceção, mas representam uma fração muito pequena da formação e são ministradas em geral sem relação com as demais disciplinas do currículo.

Uma mudança dessa natureza requer tempo e esforço. A falta de autoridade real dos coordenadores de curso, aliada ao baixo incentivo dessa função para o progresso na carreira docente, dificultam e inibem iniciativas mais inovadoras de reforma curricular, que envolvem gerir conflitos de interesse, obter o apoio dos pares, articulações com outras unidades e dedicação e esforço continuado por vários meses. Por outro lado, as agências de fomento e os critérios internos de progressão na carreira contribuem para reduzir o interesse dos docentes pelo aperfeiçoamento do ensino de graduação, ao valorizar muito mais os resultados da pesquisa acadêmica.

6. Conclusões

A contribuição dos estudos de ciência, tecnologia e sociedade (CTS) neste debate se dá pela exposição das circunstâncias e das redes de relacionamentos que levaram à atual situação do ensino de computação nas universidades. Os Estudos CTS permitem colocar em questão práticas estabelecidas que, de tão repetidas, são percebidas como naturalizadas, como resultantes de uma evolução natural que não poderia ter sido diferente.

Os Estudos CTS não produzem diretamente, e nem necessariamente, propostas de caminhos alternativos. O propósito é oferecer uma narrativa que apresente um olhar descompromissado, que estimule a reflexão e permita perceber outros caminhos possíveis a partir do entendimento das circunstâncias e contingências, envolvendo pessoas, interesses, regulamentos, instituições e vários outros atores, que levaram à presente situação.

Referências

- BRASIL (1968) Lei nº 5.540, de 28 de Novembro de 1968 (Reforma Universitária), disponível em <<http://www2.camara.leg.br/legin/fed/lei/1960-1969/lei-5540-28-novembro-1968-359201-publicacaooriginal-1-pl.html>>. Acesso em 03/09/2013.
- BRASIL. (1999) MEC/ Secretaria de Educação Superior, Departamento de Políticas do Ensino Superior. Comissão de Especialistas de Ensino de Computação e Informática. Diretrizes curriculares de cursos da área de computação e informática. Brasília: MEC, Disponível em: < <http://www.inf.ufrgs.br/ecp/docs/diretriz.pdf>>. Acesso em: 10/5/2016.
- BRASIL (2012) MEC/CNE/CES Diretrizes Curriculares Nacionais para os cursos de graduação em Computação. Parecer CNE/CES no. 136/2012 Disponível em http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=11205-pces136-11-pdf&category_slug=julho-2012-pdf&Itemid=30192, acesso em 10/12/2015.
- Cabral, M.I.C. et al. (2008) A Trajetória dos Cursos de Graduação da Área de Computação e Informática: 1969-2006. Rio de Janeiro: Editora SBC.
- Capre (1974a) Curso Técnico de Nível Superior em Processamento de Dados, Boletim Informativo, v.2(2), 42-50, abr/jun 1974.
- Capre (1974b) Reunião do Projeto 19 com as Empresas, Boletim Informativo, v.2(3), 20-29, jul/set 1974.
- CEEInf (s/data) DPES/SESu/MEC Comissão de Especialistas de Ensino de Computação e Informática. Disponível em: <http://www.inf.ufrgs.br/mec/ceeinf.comissao.html>, acesso em 2/2/2016
- Dantas, V. (1988) Guerrilha Tecnológica - A Verdadeira História da Política Nacional de Informática. Rio de Janeiro: LTC-Livros Técnicos e Científicos Editora Ltda., 1988.
- Latour, B. (2000) Ciência em Ação: como seguir cientistas e engenheiros sociedade afora. São Paulo: Editora UNESP.
- Latour, B. (2007) Reassembling the Social-An Introduction to Actor-Network-Theory. New York: Oxford University Press.
- Lobato, W.S. (1982) SERPRO -Uma crônica de 18 anos, 1982. Publicação avulsa, disponível em https://www.serpro.gov.br/sala-de-imprensa/conteudo/p_publicacoes/serpro-18-anos.pdf/@@download/file/livro-18-anos.pdf, acesso em 19/05/2016.
- Martins, C.B. (2009) A reforma universitária de 1968 e a abertura para o ensino superior privado no Brasil. Educação e Sociedade, Campinas, vol. 30, n. 106, p. 15-35, jan./abr. 2009. Disponível em <http://www.scielo.br/pdf/es/v30n106/v30n106a02>.

- Oliveira, J.P.M. (2016) I SEMISH -1974. Site do Prof. Palazzo. Disponível em <http://www.palazzo.pro.br/Wordpress/i-semish-1974/>, acesso em 19/05/2016.
- SBC (1991) Currículo de Referência da SBC para Cursos de Graduação Plena em Computação 1991. Diretoria de Educação, Sociedade Brasileira de Computação, junho de 1991. Disponível em: <http://homepages.dcc.ufmg.br/~bigonha/Cr/cr91.html>, acesso em 17/2/2015.
- SBC (1996) Currículo de referência da SBC para cursos de graduação em computação, versão 1996 – Parte 1, Diretoria de Educação, Sociedade Brasileira de Computação. Disponível em: <http://homepages.dcc.ufmg.br/~bigonha/Cr/crf96.html>, acesso em 24/5/2016.
- SBC (1999) Currículo de referência da SBC para cursos de graduação em computação, versão 1999 – Parte 1, Diretoria de Educação, Sociedade Brasileira de Computação. Disponível em: <http://lad.dsc.ufcg.edu.br/ec/cr99.pdf>, acesso em 24/5/2016.
- SBC (2005) Currículo de Referência da SBC para Cursos de Graduação em Bacharelado em Ciência da Computação e Engenharia de Computação., Proposta versão 2005. Diretoria de Educação, Sociedade Brasileira de Computação. Disponível em <http://www.sbc.org.br/documentos-da-sbc/summary/131-curriculos-de-referencia/760-curriculo-de-referencia-cc-ec-versao2005>, acesso em 24/5/2016.

Desafio Jovem Engenheiro: um torneio para incentivar o ingresso de estudantes de nível médio em carreiras de exatas

Poliana Vitti¹, Marcos Augusto Francisco Borges¹

¹Faculdade de Tecnologia – Universidade Estadual de Campinas (UNICAMP)
Limeira – SP – Brasil

p137316@dac.unicamp.br, marcosborges@ft.unicamp.br

Abstract. *Brazil has a deficit of 150,000 engineers but data show that Brazilians prefer humanities. Thus, the country needs to arousing of youth's interest about Science, Technology, Engineering and Maths (STEM). Desafio Jovem Engenheiro (DJE) is an online tournament that was proposed to encourage high school students' applications for courses in the STEM fields. The tournament occurs on a project-based learning approach through the proposition of interesting and fun challenges about engineering concepts. In its second edition, DJE involved 326 Brazilian students of four regions of the country, who have engaged and have collaborated to solve each challenge.*

Resumo. *O Brasil apresenta um déficit de milhares de engenheiros e profissionais de Tecnologia da Informação (TI). Dados demonstram que o brasileiro apresenta preferência pelas áreas de humanas, o que justifica o baixo ingresso de jovens em áreas das ciências exatas. O Desafio Jovem Engenheiro (DJE) é um torneio criado para motivar alunos do ensino médio a se candidatarem para cursos na área de exatas, como engenharia e TI. Todo o torneio ocorre através da internet, a partir de projetos desafiadores que usam a abordagem de aprendizagem baseada em projetos. Este trabalho apresenta o DJE, descrevendo sua condução e alguns resultados observados ao longo das duas edições já realizadas.*

1. Introdução

O Brasil tem enfrentado um período de crescimento econômico pouco expressivo, acarretado, dentre diversos motivos, pela baixa produção de alta tecnologia e pela pouca variedade de produtos de exportação. A exportação brasileira é baseada em *commodities* que possuem pouco valor agregado. A tecnologia é importada de países como China, Estados Unidos da América (EUA), Alemanha e Japão, que possuem uma grande fatia desse segmento no mercado mundial. Segundo dados do Ministério do Desenvolvimento, Indústria e Comércio Exterior, os bens primários corresponderam a cerca de 50% das exportações do Brasil em 2014, enquanto os bens de capital e de consumo representaram 66% das importações [MDIC 2014].

Sabe-se que para um país crescer são necessários investimentos em educação e em pesquisa para gerar novas tecnologias e tornar o mercado mais competitivo [JIN e JIN 2014, KORABLEVA e LITUN 2014]. Para gerar tecnologia de ponta é necessária mão-de-obra especializada [WINTERS 2014]. Entre as principais categorias de profissionais responsáveis pela geração de novas tecnologias estão os engenheiros e

profissionais da área de Tecnologia da Informação (TI). O Brasil apresenta um déficit de 150 mil engenheiros [CONFEA 2012]. Entretanto o interesse do brasileiro pela área de engenharia e ciências exatas é baixo. De acordo com dados de 2013 do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) apenas 20% dos jovens ingressaram em cursos de exatas enquanto 61% preferiram o campo das ciências humanas. Comparando o Brasil com os países associados à Organização para a Cooperação e Desenvolvimento Econômico¹ (OCDE), há 17% de ingressantes a mais nos cursos de Ciências Sociais, Negócios e Direito. Enquanto isso, o ingresso em engenharia, produção e construção é 3% maior nesses países [INEP 2013].

O objetivo deste artigo é apresentar a aplicação de uma abordagem de aprendizagem baseada em projetos que tem como finalidade incentivar o ingresso de jovens em carreiras de exatas. Essa aplicação se deu por meio de um torneio educacional chamado Desafio Jovem Engenheiro (DJE), que envolveu estudantes do ensino médio de diversas regiões do Brasil.

Este documento está dividido em sete seções, incluindo esta. A Seção 2 discute a falta de profissionais de exatas no Brasil e a baixa procura por cursos de Engenharia. A Seção 3 apresenta trabalhos correlatos a partir de exemplos de iniciativas semelhantes ao DJE. A Seção 4 é dedicada à apresentação do DJE como ferramenta para atrair jovens para cursos de exatas. A Seção 5 descreve os materiais e métodos. A Seção 6 é destinada a apresentar os resultados e dados do torneio. A Seção 7 apresenta as considerações finais.

2. A Falta de profissionais de exatas no Brasil

A área das ciências exatas tem baixa procura pelos estudantes brasileiros. Por conta de dificuldades encontradas nas disciplinas de matemática, física e química no ensino básico, muitos estudantes estabelecem uma espécie de bloqueio para aprender e dominar este conteúdo [RASK 2010]. Uma consequência desse bloqueio diz respeito ao crescente desinteresse pelos jovens por programação e pela área de Tecnologia da Informação (TI). Além disso, esse bloqueio acaba por tornar as humanidades mais atrativas do que os cursos de exatas, o que demanda estratégias para torná-los mais interessantes [MENESES et. al 2015].

A Figura 1 é baseada no Censo da Educação Superior de 2013 elaborado pelo INEP sobre o ingresso de alunos em cursos de graduação a cada 10.000 habitantes. Nota-se que o ingresso de estudantes é maior em cursos cuja área está envolvida com as ciências humanas, por exemplo, Ciências Sociais, Negócios, Direito e Educação. A Figura 2 apresenta uma comparação com os países associados à OCDE, no qual se pode observar que o Brasil obteve 17% de ingressantes a mais nas áreas de Ciências Sociais, Negócios e Direito. Além disso, o interesse pela área da Educação é 15% maior no Brasil, enquanto o ingresso em engenharia, produção e construção é 3% menor em relação aos países membros da OCDE.

¹ Organisation for Economic Co-operation and Development (OECD) é uma organização mundial sediada na França que conta com 34 países associados, cuja missão é promover políticas que melhorem o bem-estar econômico e social de pessoas do mundo todo [OECD 2016].

Há também alta evasão de alunos nos cursos de engenharia. Nos EUA, entre os anos de 2003 e 2009, 69% abandonaram seus cursos até 2009 [CHEN e SOLDNER 2013]. Na Figura 3, é possível perceber que o cenário brasileiro é ainda pior. No ano de 2013, 80% dos ingressantes abandonaram seus cursos de engenharia e aproximadamente 70% abandonaram seus cursos nas áreas de ciências, matemática e computação, as taxas de abandono mais altas entre todos os cursos.

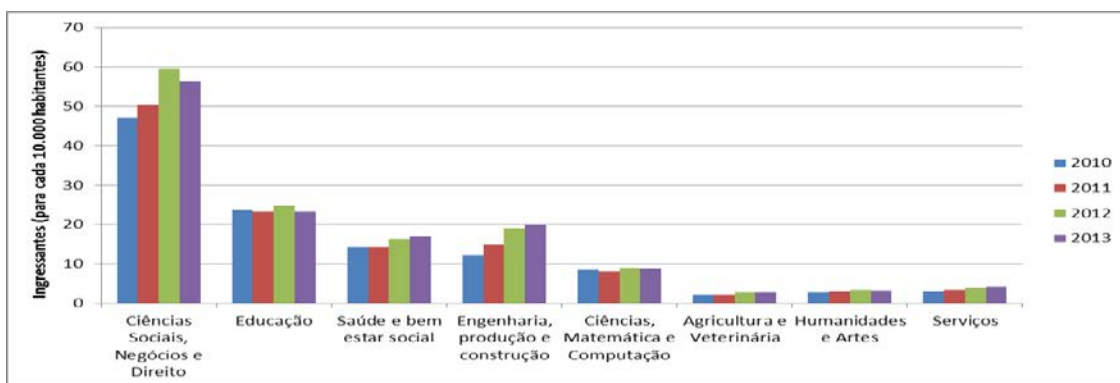


Figura 1. Número de Ingressos em Cursos de Graduação no Brasil para cada 10 mil habitantes, segundo Área Geral do Curso.

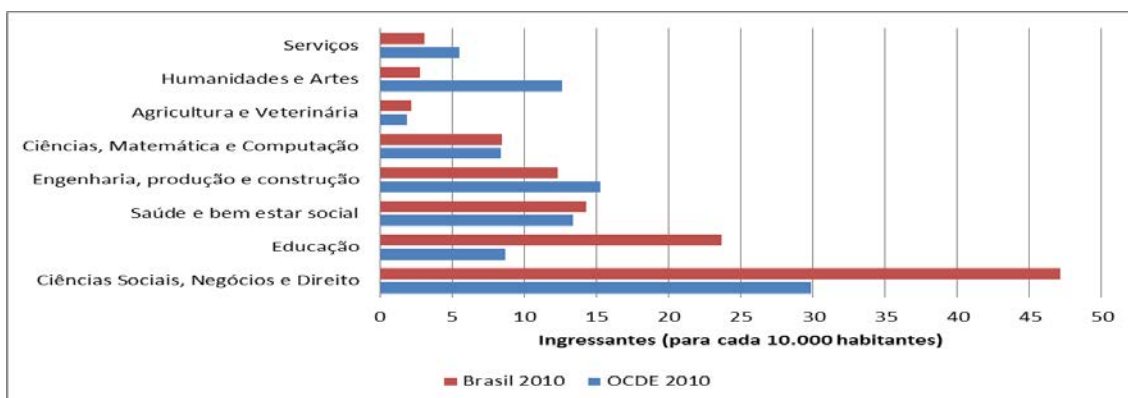


Figura 2. Comparação de ingressantes em cursos superiores entre o Brasil e a OCDE em 2010.

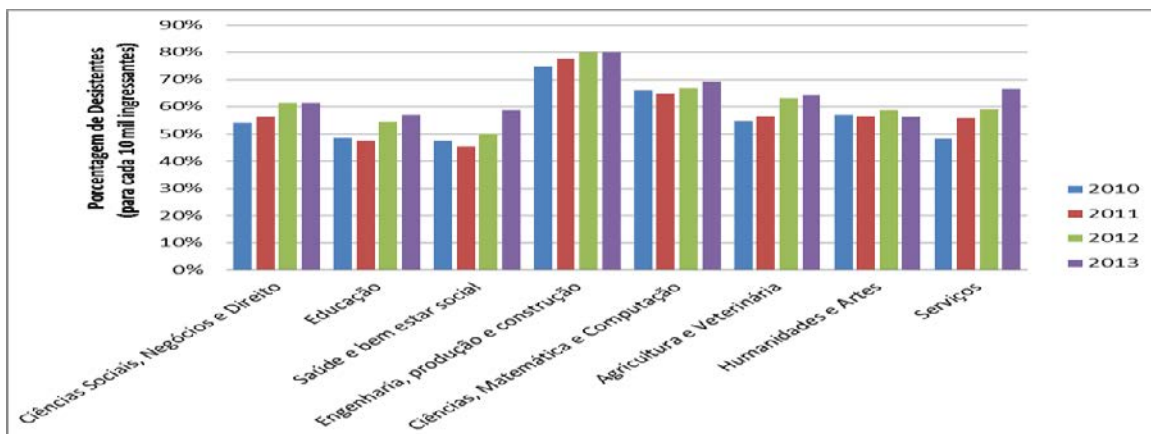


Figura 3. Porcentagem de desistentes por áreas de conhecimento

3. Trabalhos Correlatos

Diante do cenário mundial e nacional, é importante estudar meios para superar a aversão dos jovens pelas áreas de exatas. As subseções a seguir apresentam três exemplos de projetos que visam motivar a candidatura de alunos em carreiras de exatas. A subseção 3.1 trata do programa híbrido de inovação para a diversificação e reforço das capacidades em ciências exatas (ISTEM). A subseção 3.2 aborda o programa de oferecimento de experiências tecnológicas inovadoras para estudantes e professores (ITEST). Na subseção 3.3 é apresentada uma abordagem que utiliza uma gincana para ensino de conceitos computacionais para alunos do ensino médio.

3.1. ISTEM

ISTEM é um projeto do Centro de Pesquisa e Aprendizagem de Ciências Exatas (STELAR) dos EUA para desenvolver e determinar estratégias eficazes para envolver alunos no campo das ciências exatas. O projeto se foca em estudantes da terceira à oitava séries do ensino fundamental, que são atingidos por meio de mentoria nas escolas e experiências externas ao ensino formal. Essa iniciativa recruta, treina e mantém um mínimo de sessenta mentores durante um período de três anos. O mentor pode ser um aluno de faculdade, um profissional ou um voluntário que realiza duas reuniões por mês com grupos de alunos que são engajados em, no mínimo, cinco experiências informais associadas à ciência. As atividades do programa são baseadas em quatro grandes projetos em engenharia determinados pela Academia Nacional de Engenharia: Energia e Meio Ambiente; Saúde; Segurança e Aprendizagem e Computação. A mentoria busca facilitar a aprendizagem, identificando teorias e métodos relacionados ao cotidiano dos alunos de maneira a ajudá-los a contextualizar os problemas [STELAR 2012].

3.2. ITEST

O ITEST é um programa da Fundação Nacional de Ciência da Diretoria para Educação e Recursos Humanos dos EUA que tem como objetivo financiar as melhores práticas para motivar a participação de alunos do ensino fundamental e médio no núcleo de disciplinas de ciência, tecnologia, engenharia e matemática. Essa iniciativa procura implementar e difundir estratégias, ferramentas e modelos inovadores que propiciem o interesse dos alunos por meio de experiências formais e informais, na escola e fora dela. O ITEST financia projetos que divulgam a área de ciências exatas, que orientam os alunos a buscarem trajetórias educacionais apropriadas e que proporcionam o desenvolvimento de habilidades práticas demandadas pelos empregadores do setor. Experiências que envolvam ativamente a indústria e os negócios ligados à área de exatas são fortemente encorajadas [NSF 2014b].

O ITEST financia, por exemplo, o projeto FUSE Studios, organizado a partir de uma sequência de desafios envolvendo atividades práticas. Esses desafios são orientados pelo interesse de grupos de alunos que são apoiados por uma tutoria on-line de especialistas do setor. A competição envolve o julgamento dos artefatos produzidos pelos alunos que são premiados com base nas realizações coletivas em cada desafio. O projeto pretende investigar como a temática dos desafios afeta o interesse de jovens [NSF 2014a].

3.3. Aplicação de gincanas no ensino da computação

Barbosa et al (2015) propuseram uma gincana a fim de levar conceitos da computação como, por exemplo, habilidades relacionadas ao raciocínio lógico e à resolução de problemas, para alunos de 1º ano do ensino médio. A gincana ocorreu em duas etapas, com três encontros presenciais semanais. Em cada encontro, monitores distribuíram os participantes em grupos de, em média, seis alunos, para os quais um tópico da computação necessário para aprendizagem da programação foi apresentado. Inicialmente era realizada uma explicação conceitual através de artefatos familiares aos participantes e, em seguida, os alunos resolviam uma atividade relacionada com o conceito aprendido. Segundo os autores, o objetivo de ensinar conceitos computacionais foi atingido satisfatoriamente pelo estímulo aos alunos para raciocinar e desenvolver sua própria estratégia de resolução de problemas.

4. A proposta do DJE

O Desafio Jovem Engenheiro (DJE) é um torneio criado para incentivar alunos do ensino médio a se candidatarem a cursos superiores na área de exatas e tecnológicas. O DJE se utiliza de projetos desafiadores direcionados a equipes que, dependendo de seu desempenho, podem ser premiadas na etapa final. As equipes são compostas por até quatro integrantes, sendo quatro estudantes do ensino médio, ou ainda, três estudantes e um mentor que auxilia a equipe na resolução do projeto. O mentor pode ser professor de uma disciplina de exatas do ensino médio, um professor universitário ou um aluno de graduação em exatas.

O DJE é estruturado em duas etapas classificatórias e uma etapa final. As etapas classificatórias selecionam as equipes com as melhores pontuações para continuar no torneio na etapa subsequente. A primeira etapa possui três projetos desafiadores, enquanto a segunda conta com dois. As dez equipes melhores colocadas ao final das etapas classificatórias são selecionadas para competirem na etapa final. O tema de cada projeto é um problema colocado em um determinado contexto, no qual os alunos precisam explorar habilidades de uma área das exatas para encontrar a solução. Os projetos desafiadores são propostos por professores universitários ou por empresas patrocinadoras do torneio [DJE 2014].

77% dos adolescentes brasileiros são usuários da internet [CGI.BR 2014]. Por este motivo, a Internet é utilizada como principal ferramenta de Tecnologia Digital de Informação e Comunicação (TDIC) para realização do DJE, tornando possível a participação de estudantes de diferentes perfis e residentes em todos os estados do Brasil. Como o DJE ocorre totalmente online, é necessário o uso de uma ferramenta que simplifique a comunicação entre os participantes e a organização. Para isso é utilizada a ferramenta TelEduc. As equipes recebem o projeto através do Teleduc e é estabelecido um prazo para que resolvam o problema e enviem a solução.

5. Materiais e métodos

A falta de profissionais de tecnologia no Brasil implica a necessidade de iniciativas como o DJE, que visa à atração de jovens para as carreiras de exatas. Os trabalhos correlatos — especialmente o ITEST — oferecem exemplos da utilização de projetos desafiadores apoiados por mentoria e ferramentas *online* para motivar equipes de alunos

e desenvolver suas habilidades. Nesse sentido, a implementação do DJE se deu a partir da utilização dos materiais e métodos que são descritos a seguir. A Subseção 5.1 trata da aprendizagem baseada em projetos. As subseções 5.2 e 5.3 apresentam as ferramentas de TDIC utilizadas pelo DJE: o TelEduc e as redes sociais, respectivamente.

5.1. Aprendizagem baseada em projetos

A Aprendizagem Baseada em Projetos (PBL) é destinada a reforçar as capacidades dos alunos em projetos, trabalho e comunicação interpessoal através de experiências desenvolvidas na solução de problemas como uma equipe. PBL tornou-se muito utilizada no ensino da engenharia, pois a indústria requer graduados com habilidades voltadas para a execução de projetos [HAN, CAPRARO e CAPRARO 2014, YUKAWA et. al 2012]. Nessa abordagem, os estudantes utilizam etapas de planejamento e execução de um projeto para solucionar um problema. A aprendizagem autodirigida ocorre quando a equipe se une em um processo que envolve procurar, conciliar e selecionar as soluções para resolver o problema [HONG et. al 2012].

PBL busca capturar o interesse dos alunos através de uma situação do mundo real e provocar reflexões a respeito de como eles adquirem e aplicam o novo conhecimento em um contexto de resolução de problemas. O papel do professor é de facilitar a resolução, formulando perguntas que auxiliem os alunos a encontrarem a solução. Uma das vantagens dessa abordagem é que pode ser aplicada dentro e fora da sala de aula, uma vez que o mentor, ao contrário de um instrutor, deve avaliar principalmente como os estudantes aprenderam com essa experiência [EFSTRATIA 2014].

5.2. TelEduc (EaD)

A comunicação entre os organizadores do DJE e as equipes é feita através do TelEduc, uma ferramenta de Ensino a Distância (EaD) desenvolvida pelo Núcleo de Informática Aplicada à Educação (NIED) da Universidade Estadual de Campinas (UNICAMP). Essa ferramenta integra todo o torneio em um único ambiente e permite a publicação dos enunciados dos projetos. Isso facilita a comunicação entre as equipes e os organizadores que podem interagir através do correio eletrônico. Além disso, o TelEduc dispõe de portfólios para as equipes publicarem as resoluções de cada projeto [NIED 2016]. Uma característica importante do TelEduc, além de suas inúmeras funcionalidades, é o fato de ser uma ferramenta livre com código aberto [TELEDUC 2016].

5.3. Redes Sociais

Sackey, Nguyen e Grabill (2015) apontam a necessidade de utilização de duas plataformas, usadas de formas distintas, a fim de permitir o engajamento dos participantes. No TelEduc, a interação é restrita apenas ao conteúdo do DJE e ocorre somente quando o aluno está conectado à plataforma. Por outro lado, as redes sociais permitem um envolvimento com o participante para além do momento da entrega dos projetos, através da publicação de curiosidades e notícias, em uma plataforma que ele utiliza frequentemente. Estudos relatam que cerca de 90% dos adolescentes entre 15 e 17 anos possuem um perfil nas redes sociais [CGI.BR 2014]. A opção por um veículo mais familiar ao cotidiano do jovem tem o intuito de dar maior visibilidade e divulgar informações pertinentes ao torneio. O DJE utiliza uma página do [Facebook](#) para

publicar informações e esclarecimentos sobre os projetos, o período de inscrição, além de diversas notícias e curiosidades sobre a área de tecnologia. O *Facebook* se apresenta como importante ferramenta para auxiliar a atração de participantes para o torneio.

6. Resultados e discussão

O DJE ocorre anualmente desde 2013. Em duas edições, reuniu aproximadamente 180 equipes oriundas de diversas regiões do Brasil. A primeira edição teve caráter experimental no que diz respeito à implantação da proposta. Apesar disso, o número de equipes inscritas (90 equipes) e a repercussão positiva motivaram a realização de uma nova edição. No ano de 2014 ocorreu a segunda edição do DJE, que contou com a participação de 326 estudantes distribuídos em 89 equipes.

A Figura 4 apresenta informações referentes ao total de equipes inscritas na competição de 2014. O gráfico (A) mostra que o envolvimento de indivíduos do sexo feminino é 44% menor que o de indivíduos do sexo masculino. Esse resultado pode indicar que o interesse de mulheres em cursos de exatas tende a ser menor. O gráfico (B) apresenta a distribuição das equipes inscritas de acordo com as regiões geográficas do país. A região Nordeste teve metade das equipes inscritas, um índice aproximadamente 20% maior do que o da região Sudeste, segunda colocada em número de participantes. Nessa edição, a região Sul não foi representada, já que não contou com equipes inscritas. O gráfico (C) contém a distribuição dos inscritos pelo tipo de instituição de ensino. Aproximadamente metade dos participantes do DJE é oriunda do setor público de ensino composto por escolas estaduais, escolas técnicas estaduais (ETECs) e institutos federais. O setor privado de ensino respondeu por 32% do total de inscritos.

A Figura 5 apresenta informações referentes ao total de equipes que se inscreveram, mas deixaram de participar do torneio ao longo das etapas do DJE 2014. As desistências representaram 65% do total de inscritos. O gráfico (A) mostra que a taxa de desistência de indivíduos do sexo masculino é cerca de 40% maior do que a de indivíduos do sexo feminino. O gráfico (B) apresenta a distribuição das equipes desistentes de acordo com as regiões geográficas do país. Nesse aspecto, a distribuição das taxas de abandono foi proporcional às taxas de inscrição. A região Nordeste teve metade das equipes que abandonaram a competição, seguida pela região Sudeste com cerca de 30% de desistentes. O gráfico (C) permite notar que o setor privado de ensino respondeu pelo maior número de equipes desistentes dentre os inscritos no DJE. Das equipes que abandonaram o torneio, aproximadamente 30% vieram de colégios particulares.

A Figura 6 apresenta informações referentes às dez equipes que obtiveram as melhores colocações nas duas primeiras etapas e, por isso, foram selecionadas para a fase final do DJE 2014. O gráfico (A) mostra que a presença de indivíduos do sexo feminino e masculino foi proporcional às taxas de inscrição. O gráfico (B) apresenta a distribuição das equipes finalistas de acordo com as regiões geográficas do país. Somente equipes do Nordeste e do Sudeste se classificaram para a final do torneio, contando com 60% e 40% dos finalistas, respectivamente. O gráfico (C) mostra que, dentre as equipes que se classificaram para a final do DJE, houve apenas representantes de escolas estaduais, ETECs e institutos federais. Estes responderam por 70% dos participantes melhor classificados. O campeão da segunda edição do DJE foi uma

equipe do Instituto Federal do Espírito Santo (Sudeste) composta por quatro indivíduos, um do sexo feminino e três do sexo masculino.

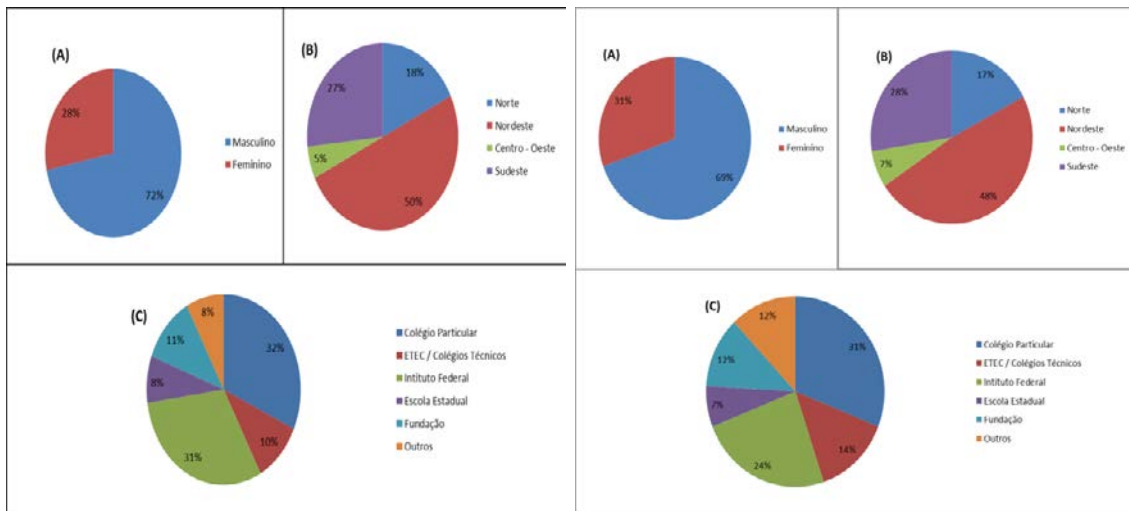


Figura 4. Inscritos edição de 2014.
(A) % por gênero,
(B) % de equipes por região,
(C) equipes por tipo de inst. de ensino.

Figura 5. Desistentes edição de 2014.
(A) % por gênero,
(B) % de equipes desistentes por região,
(C) desistentes por tipo de inst. de ensino.

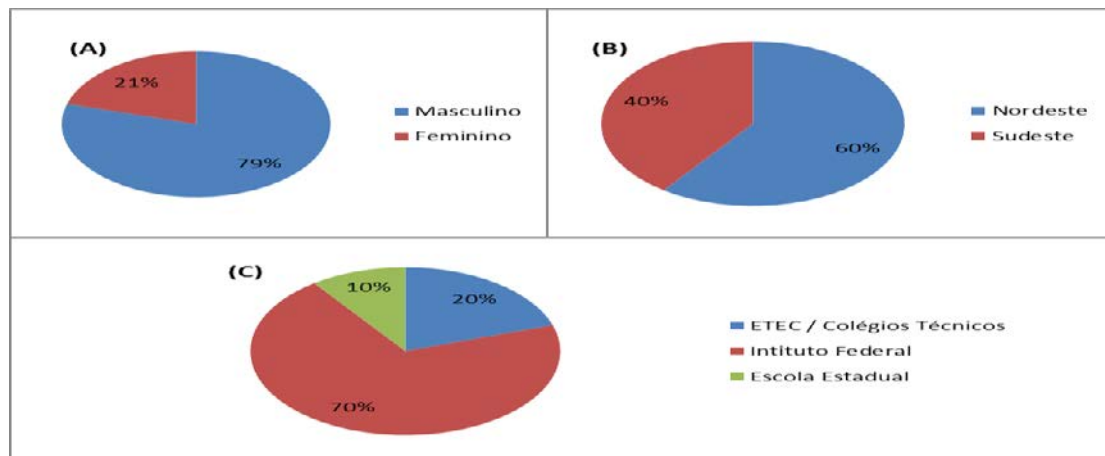


Figura 6. % finalistas edição de 2014 (A) por gênero, (B) por região, (C) por tipo inst.

7. Conclusão

Este trabalho apresenta o Desafio Jovem Engenheiro (DJE), uma proposta alternativa para incentivar jovens a se candidatarem para cursos de tecnologia e engenharia. Essa proposta se estabelece sobre um torneio para alunos de ensino médio, que é composto por projetos desafiadores distribuídos em etapas que visam desenvolver o interesse desses jovens pelas ciências exatas. A competição é inspirada por outras iniciativas semelhantes pelo mundo, entretanto utiliza conceitos de aprendizagem baseada em projetos. Assim, a pontuação de cada etapa depende da elaboração de projetos vinculados à atuação de profissionais da área de exatas, como engenheiros e analistas de sistemas. No ano de 2014, em sua segunda edição, o DJE contou com 326 jovens, que se engajaram, trabalharam em equipe e colaboraram entre si, desenvolveram novas

habilidades, adquiriram e exercitaram conhecimentos e resolveram problemas a partir de conceitos da área de tecnologia.

A ausência de equipes inscritas da região Sul demanda que nos próximos anos sejam desenvolvidas novas estratégias na divulgação do torneio nos estados do Paraná, Santa Catarina e Rio Grande do Sul. A falta de representantes do Sul do Brasil pode indicar que a utilização das redes sociais (em especial, o *Facebook*) como ferramenta mobilizadora não surtiu o mesmo nível de efeito nessa região. Também devem ser feitas considerações acerca da baixa participação de indivíduos do sexo feminino. As próximas edições do DJE devem, portanto, preocupar-se com esse público por meio de iniciativas focadas na atração das jovens estudantes para as ciências exatas. Além disso, deve-se investigar a alta taxa de desistência de modo a compreender os motivos que levaram os participantes a abandonar o torneio.

Outra questão a ser abordada trata do impacto do DJE para a escolha da carreira de cada um dos participantes. Para tal, deve-se realizar, futuramente, um acompanhamento dos inscritos de cada edição, a fim de mensurar a real contribuição do torneio para atração e formação dos futuros engenheiros.

Referências

- Barbosa, A., Pereira Neto, A., Oliveira, R., Costa, T., Araújo, A., Costa, F. (2015). O ensino de conceitos computacionais para alunos do ensino médio: relato de experiência de uma gincana e das estratégias utilizadas pelos alunos na resolução das atividades desplugadas. *In: Anais do 23º Workshop sobre Educação em Informática (WEI 2015)*, Recife, PE.
- CGI.BR. (2014). TIC KIDS ONLINE BRASIL 2013: Pesquisa sobre o uso da Internet por crianças no Brasil. Comitê Gestor da Internet no Brasil, São Paulo.
- Chen, X. e Soldner, M. (2013). STEM Attrition: College Students' Paths Into and Out of STEM Fields. National Center for Education Statistics, Washington.
- CONFEA. (2012). Faltam 150 mil engenheiros no País. Acesso em 04 de 03 de 2016, disponível em CONFEA - Confederação Federal de Engenharia e Agronomia: <http://www.confea.org.br/cgi/cgilua.exe/sys/start.htm?inoid=16978&sid=1206>
- DJE. (2014). Desafio Jovem Engenheiro - Sobre o desafio. (Faculdade de Tecnologia) Fonte: Laboratorio de Informática, Aprendizagem e Gestão: <http://www.ft.unicamp.br/liag/wp/jovemengenheiro/>
- Efstratia, D. (2014). Experiential education through project based learning. *Procedia - Social and Behavioral Sciences*, 152, 1256 – 1260.
- Han, S., Capraro, R. e Capraro, M. M. (2014). How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement. *International Journal of Science and Mathematics Education*.
- Hong, J.-C., Chen, M.-Y., Wong, A., Hsu, T.-F. e Peng, C.-C. (November de 2012). Developing physics concepts through hands-on problem solving: a perspective on a technological project design. *International Journal of Technology and Design Education*, 22, 473-487.

- INEP. (2013). Censo da Educação Superior 2013. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília.
- Jin, J. C. e Jin, L. (2014). On the relationship between university education and economic growth: the role of professors' publication. *Education Economics*, 22, 635-651.
- Korableva, O. e Litun, V. (2014). The potential of transitive economies' growth based on innovative strategy. *WSEAS Transactions on Business and Economics*, 11, 725-736.
- MDIC. (2014). Balança Comercial Brasileira: Dados consolidados. Ministério do Desenvolvimento, Indústria e Comércio Exterior, Brasília.
- Meneses, L., Mai, L., Rosario, J., Oliveira, E., Gomes, R. (2015). IntroComp: Atraindo Alunos do Ensino Médio para uma Instigante Experiência com a Programação. In: *Anais do 23º Workshop sobre Educação em Informática (WEI 2015), Recife, PE*.
- NIED. (s.d.). TelEduc. Acesso em 20 de 02 de 2016, disponível em Núcleo de Informática Aplicada à Educação: <http://www.nied.unicamp.br/?q=content/teleduc>
- NSF. (2014). FUSE Studios: A New, Interest-Driven Model for Engaging Youth In STEM and Career Development Through Challenges and Partnership with Industry. Acesso em 28 de 02 de 2016, disponível em National Science Foundation: http://www.nsf.gov/awardsearch/showAward?AWD_ID=1433724&HistoricalAward_s=false
- NSF. (2014). Innovative Technology Experiences for Students and Teachers (ITEST) . Acesso em 28 de 02 de 2016, disponível em National Science Foundation: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5467&org=DRL
- OECD. (s.d.). OECD About. Acesso em 25 de 02 de 2016, disponível em OECD: <http://www.oecd.org/about/>
- Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review*, 29, 892–900.
- Sackey, D. J., Nguyen, M. T. e Grabill, J. T. (2015). Constructing learning spaces: What we can learn from studies of informal learning online. *Computers and Composition*, 35, 112–124.
- STELAR. (2012). Projects. Acesso em 28 de 02 de 2016, disponível em STEM LEARNING AND RESEARCH CENTER: <http://goo.gl/fyzLoy>
- TELEDUC. (s.d.). Histórico. Acesso em 20 de 02 de 2016, disponível em Teleduc: <http://www.teleduc.org.br/?q=historico>
- Winters, J. V. (2014). STEM graduates, human capital externalities, and wages in the U.S. *Regional Science and Urban Economics*, 48, 190–198.
- Yukawa, T., Iwazaki, T., Ishida, K., Nishigaki, Y., Fukumura, Y., Yamazaki, M. e Miura, H. (2012). Online Collaboration Support Tools for Blended Project-Based Learning on Embedded Software Development — Final Report —. *Intelligent Interactive Multimedia: Systems & Services*, 14, 419-428.

Desafios e oportunidades aos processos de ensino e de aprendizagem de programação para iniciantes

Viviane Cristina Oliveira Aureliano^{1,2}
Patricia Cabral de Azevedo Restelli Tedesco²
Lúcia Maria Martins Giraffa³

¹Instituto Federal de Pernambuco – Campus Belo Jardim
Belo Jardim, PE – Brasil

² Universidade Federal de Pernambuco – Centro de Informática
Recife, PE – Brasil

³ Pontifícia Universidade Católica do Rio Grande do Sul – Escola de Humanidades
Porto Alegre, RS – Brasil

[vcoa, pcart]@cin.ufpe.br; giraffa@pucrs.br

Abstract. *Introductory programming courses are intrinsically difficult, and are generally associated to high failure and evasion rates. It is up to the programming teacher to deal with which such difficulties through the organization of teaching activities that help learners to acquire the necessary knowledge. In this light, this paper discusses challenges faced by teachers when planning and executing such activities, and presents research opportunities in the area. Moreover, we propose an approach to minimize some of the problems faced by programming students. The use of this approach in a digital games course has contributed to the learning of novice programming students.*

Resumo. *Disciplinas introdutórias de programação são intrinsecamente difíceis, comumente associadas a altos índices de reprovação e de evasão. Para lidar com essas dificuldades, cabe ao professor de programação organizar atividades de ensino e de aprendizagem que auxiliem os estudantes na aquisição do conhecimento necessário para a disciplina. Neste âmbito, este artigo discute desafios enfrentados pelos professores no planejamento e execução destas atividades, apresentando oportunidades de pesquisa na área. Além disso, apresentamos a proposta de uma abordagem para minimizar alguns dos problemas vivenciados pelos estudantes. A aplicação desta abordagem em um minicurso de jogos digitais demonstrou contribuir para a aprendizagem de estudantes iniciantes em programação.*

1. Introdução

A literatura na área de ensino e de aprendizagem em programação é unânime ao afirmar que começar a programar é considerado difícil pela maioria dos estudantes iniciantes nesta disciplina [CASPERSEN, 2007; DU BOULAY, 1989; GOMES; MENDES, 2007; GOMES; SANTOS; MENDES, 2012; MILNE; ROWE, 2002; ROBINS;

ROUNTREE; ROUNTREE, 2003]. Normalmente, as disciplinas introdutórias de programação estão relacionadas a altas taxas de reprovação, que podem chegar a um terço dos estudantes matriculados nestas disciplinas no mundo todo [BENNEDSEN; CASPERSEN, 2007; WATSON; LI, 2014]. Estas disciplinas também levam a altas taxas de evasão nos cursos de tecnologia [KINNUNEN; MALMI, 2006; SILVA FILHO et al., 2007]. Programar é considerada uma tarefa tão complexa que estudos apontam que, mesmo após passar por uma disciplina introdutória de programação, estudantes ainda apresentam sérios problemas ao aplicar os conceitos da disciplina [KURLAND et al., 1986; MCCRACKEN et al., 2001]. Pesquisas também apontam que a maioria dos estudantes têm seus conhecimentos de programação consolidados apenas ao final da segunda disciplina de programação [TEW; MCCRACKEN; GUZDIAL, 2005]. Por este motivo, ensinar programação é considerado um dos grandes desafios da área de educação em computação [CASPERSEN 2007].

Os obstáculos vivenciados pelos iniciantes em programação incluem problemas com a sintaxe de uma linguagem de programação (LP) [DU BOULAY, 1989], com os diferentes construtos de uma LP, tais como variáveis [PUTNAM et al., 1989; SAMURÇAY, 1989], estruturas condicionais e de repetição [CHERENKOVA; ZINGARO; PETERSEN, 2014; SPOHRER; SOLOWAY, 1989] e *arrays* [DU BOULAY, 1989]. Além disso, eles também encontram dificuldades para resolver os erros em seus programas e para dividir o programa em funções [LAHTINEN et al., 2005]. Por fim, eles encontram problemas ao lidar com conceitos mais abstratos de uma LP, tais como recursão e ponteiros [MILNE; ROWE, 2002].

Aliado a isso, os estudantes de programação adotam metodologias de estudo que são ineficientes. Eles decoram soluções para os problemas ao invés de entendê-las [GOMES; MENDES, 2007], estudam de maneira passiva ou superficial [BIGGS; TANG, 2011; CLARK; NGUYEN; SWELLER, 2005] e fazem poucos exercícios [GOMES; MENDES, 2007]. Além disso, os estudantes de programação possuem um tipo de conhecimento dito frágil, que é aquele conhecimento que eles possuem, mas não conseguem aplicar em novas situações ou problemas encontrados [PERKINS; MARTIN, 1986].

Contudo, o maior dos problemas vivenciados pelos iniciantes em programação não parece ser o entendimento dos conceitos básicos de uma LP, mas a combinação e a utilização adequada destes conceitos na construção de um determinado programa [LAHTINEN et al., 2005; ROBINS et al., 2003; SPOHRER; SOLOWAY, 1989; WINSLOW, 1996]. Sendo assim, a maioria dos erros cometidos pelos estudantes não resulta de falhas no conhecimento sobre os construtos da LP, mas resulta de dificuldades em combiná-los corretamente, ou seja, “*em colocar as peças do programa juntas*” [SPOHRER; SOLOWAY, 1989 p. 401].

Ser professor de uma disciplina introdutória de programação é uma tarefa tão complexa quanto ser estudante desta disciplina. Ou seja, o desafio de ensinar é proporcional ao de aprender. Razão pela qual muitas universidades escolhem como docentes destas disciplinas os professores mais experientes e com mais vivência nas questões relacionadas ao ensino de programação. Programar vai requerer habilidades que são diferentes daquelas necessárias para ensinar a programar. A literatura relata algumas decisões que precisam ser tomadas pelo professor, tais como a LP e as

ferramentas associadas e as práticas pedagógicas do professor [PEARS et al., 2007]. No que diz respeito às práticas pedagógicas, as preocupações do professor devem ir além da apresentação do conteúdo aos estudantes em sala de aula ou da seleção de exercícios relacionados ao conteúdo. Estas preocupações devem incluir também a proposição de atividades que auxiliem e orientem os estudantes durante os seus processos de aprendizagem. Sendo assim, a atuação docente deve acontecer em pelo menos dois momentos, na definição das atividades de ensino e de aprendizagem dos estudantes que estejam alinhadas com os objetivos de aprendizagem definidos para a disciplina introdutória de programação [BIGGS; TANG, 2011].

Dentro deste contexto, o presente trabalho tem por objetivo discutir alguns dos desafios enfrentados pelos professores no planejamento e na execução de atividades de ensino e de aprendizagem e apresentar oportunidades de pesquisa na área de educação em programação para iniciantes. Além disso, apresentamos a proposta de uma abordagem para minimizar alguns dos desafios elencados. Para alcançar estes objetivos, o restante deste trabalho encontra-se organizado da maneira descrita a seguir. Na Seção 2 apresentamos os desafios e oportunidades elencados. Na Seção 3 descrevemos uma proposta para minimizar alguns dos desafios discutidos e apresentamos os resultados obtidos com um experimento realizado com a proposta. Por fim, na Seção 5 apresentamos as considerações finais deste trabalho.

2. Alguns desafios e oportunidades

Como mencionamos anteriormente, o papel do professor de programação deve buscar orientar o processo de aprendizagem dos estudantes na disciplina [BIGGS; TANG, 2011]. Sendo assim, discutimos nesta seção alguns dos desafios enfrentados pelos professores no planejamento e na execução destas atividades e apresentamos oportunidades de pesquisa na área de educação em programação para iniciantes.

2.1. Professor deve propor atividades de ensino

Uma das principais questões a serem consideradas pelo professor no ensino de programação diz respeito às suas práticas pedagógicas. A primeira coisa que um professor de programação pensa é a respeito de quais atividades de ensino adotar. Neste contexto, uma possibilidade é a utilização de exemplos trabalhados como parte do material instrucional proposto pelo professor.

Os exemplos trabalhados¹ (*worked examples*) são uma sequência de passos que demonstram a realização de uma tarefa em particular ou a solução de um problema específico [CLARK et al., 2005]. Os exemplos trabalhados auxiliam os estudantes iniciantes na construção de novos esquemas nas fases iniciais do processo de aquisição da habilidade cognitiva [ATKINSON et al., 2000; RENKL, 2002; VANLEHN, 1996]. Exemplos são considerados, tanto por estudantes quanto por professores, o tipo de material instrucional mais útil para a área de educação em LP [LAHTINEN et al., 2005].

¹ Utilizaremos os termos *exemplos trabalhados* e *exemplos* indistintamente no decorrer deste trabalho.

Dada a sua importância como ferramenta de ensino, é imprescindível que a escolha dos exemplos leve em consideração questões, tais como a sua qualidade e legibilidade [BÖRSTLER; CASPERSEN; NORDSTRÖM, 2015], a sua estruturação [CLARK et al., 2005], o formato por meio do qual ele é apresentado aos estudantes [BENNEDSEN; CASPERSEN, 2008] e o seu alinhamento com os objetivos de aprendizagem que se desejam alcançar [BIGGS; TANG, 2011].

2.2. Professor também deve propor atividades de aprendizagem

Para fazer com que os estudantes aprendam de maneira eficaz, as preocupações do professor devem ir além da apresentação do conteúdo aos estudantes em sala de aula ou da seleção de exercícios relacionados ao conteúdo. Neste sentido, é preciso que o professor faça com que os estudantes adotem metodologias de estudo mais eficazes. Portanto, é necessário que o professor de programação inclua em suas práticas pedagógicas atividades de aprendizagem que estimulem os seus estudantes a processarem o material instrucional mais profundamente.

Neste sentido, atividades de aprendizagem que promovam a reflexão [CHI et al., 1989], que façam com que os estudantes sejam sujeitos ativos e participantes do seu próprio processo de aprendizado [BRANSFORD; BROWN; COCKING, 2000] e que os encoraje a ser menos superficiais [BIGGS; TANG, 2011] têm demonstrado resultados positivos para o processo de aprendizado dos estudantes.

Dada a sua importância como ferramenta de aprendizado do estudante, é imprescindível que as atividades propostas estejam de acordo com os objetivos de aprendizagem que se deseja alcançar [BIGGS; TANG, 2011].

3. Uma proposta de solução para os desafios identificados

Tendo em vista os desafios relacionados aos processos de ensino e de aprendizagem de programação para iniciantes, uma abordagem é apresentada para minimizar os problemas elencados. Trata-se da abordagem *Stepwise self-explanation* [AURELIANO; TEDESCO, 2015] para o aprendizado de programação, proposta com o objetivo de orientar os estudantes iniciantes enquanto eles estudam para uma disciplina de programação.

3.1. *Stepwise self-explanation*

A abordagem *Stepwise self-explanation* lança mão de dois referenciais teóricos principais, o framework do *Stepwise Improvement* (FSI) [CASPERSEN, 2007; CASPERSEN; KÖLLING, 2009] e a teoria das auto-explicações (AE) [CHI et al., 1989]. O FSI é um framework conceitual que descreve a atividade de programação como um processo sistemático e incremental que engloba três atividades diferentes, extensão, refinamento e reestruturação [CASPERSEN, 2007; CASPERSEN; KÖLLING, 2009]. A atividade de extensão ocorre quando a especificação é estendida de maneira a cobrir mais casos de uso. A atividade de refinamento ocorre quando o código abstrato é modificado de maneira a se construir código executável que implemente a especificação correspondente. A atividade de reestruturação ocorre quando é realizada uma melhoria dos aspectos não funcionais do programa, no entanto essa modificação não envolve uma mudança no comportamento aparente do programa. O FSI oferece uma contribuição

significante à área de educação em programação fornecendo orientação nas três atividades que engloba, oferecendo orientação na maneira que materiais instrucionais (e.g., livro-texto, exercícios, aulas e exemplos) podem ser estruturados e na construção de programas propriamente dita.

Apesar do FSI fornecer orientação na estruturação e organização de materiais instrucionais, ele não oferece nenhum tipo de orientação na maneira que os estudantes podem estudar e aprender programação a partir destes materiais instrucionais. Em virtude disso, propomos que este tipo de orientação aconteça através da AE [CHI et al., 1989]. A AE consiste em *“um diálogo mental que aprendizes possuem enquanto estudam um exemplo trabalhado e que os ajuda a entender o exemplo e a construir um esquema a partir deste exemplo”* [CLARK et al., 2005]. De acordo com Chiu e Chi (2014), a atividade de auto-explicar, ou explicar para si mesmo, promove o aprendizado através da elaboração da informação que está sendo estudada, da associação dessa nova informação com a conhecimento prévio que o aprendiz possui, da construção de inferências e da conexão dos diferentes pedaços de informação.

Para esta abordagem, escolhemos exemplos como parte das atividades de ensino. Os exemplos trabalhados são apresentados normalmente de maneira estática, mas decidimos apresentá-los no formato de vídeo. Adotamos os exemplos em vídeo, pois eles possuem o formato ideal para exibição de programas para iniciantes em uma disciplina de programação, podendo combinar a apresentação dos programas com todo o seu processo de construção [BENNEDSEN; CASPERSEN, 2008]. Como atividades de aprendizagem, escolhemos estimular e orientar o processo de reflexão dos estudantes através da inclusão de questões para que eles auto-expliquem os exemplos que estão estudando. Para que pudéssemos incluir as questões para AE nos exemplos em vídeo, estruturamo-os de forma que eles tivessem duas partes distintas. A primeira delas é a apresentação do enunciado do problema seguido por exemplos de execução do programa com entradas específicas e as saídas correspondentes (atividade “Estende especificação” na Figura 1). Em seguida, a segunda parte consiste na apresentação de uma possível sequência de passos que soluciona o problema apresentado. Para facilitar o processo de aprendizado dos iniciantes em programação, decidimos construir vídeos que apresentam somente sequências de atividades de extensões e refinamentos. Assim, a cada novo passo da solução executado no vídeo, teremos uma extensão intermediária (atividade “Estabelece especificação intermediária” na Figura 1) e um refinamento (atividade “Refina implementação” na Figura 1).

As questões, por sua vez, foram adicionadas aos vídeos em quatro momentos distintos de sua execução: (i) um após a apresentação do enunciado do problema e de seus exemplos de entrada e saída (atividade “Auto-explica extensão principal” na Figura 1); dois a cada novo passo da solução apresentada: (ii) um após se informar a funcionalidade que será implementada no passo e antes de se apresentar a implementação para o referido passo (atividade “Antecipa refinamento” na Figura 1) e (iii) outro após a implementação para o passo ser apresentada (atividade “Auto-explica refinamento” na Figura 1); e (iv) o último após apresentar a solução completa do problema (atividade “Avalia código entradas específicas” na Figura 1).

As questões apresentadas no momento (i) tem por objetivo fazer com que o estudante auto-explique o propósito geral do programa apresentado no exemplo. As

questões apresentadas no momento (ii) tem por objetivo fazer com os estudantes reflitam a respeito de uma possível solução para o passo de resolução do problema antes da solução do professor ser apresentada e, depois comparem a solução pensada por eles com a solução apresentada pelo professor. As questões apresentadas no momento (iii) tem por objetivo fazer com que os estudantes façam inferências sobre a implementação apresentada pelo professor através da proposição de modificações no código apresentado. As questões apresentadas no momento (iv) tem por objetivo fazer com os estudantes calculem mentalmente a saída do código a partir de uma determinada entrada, de forma que eles reflitam sobre a execução do programa apresentado.

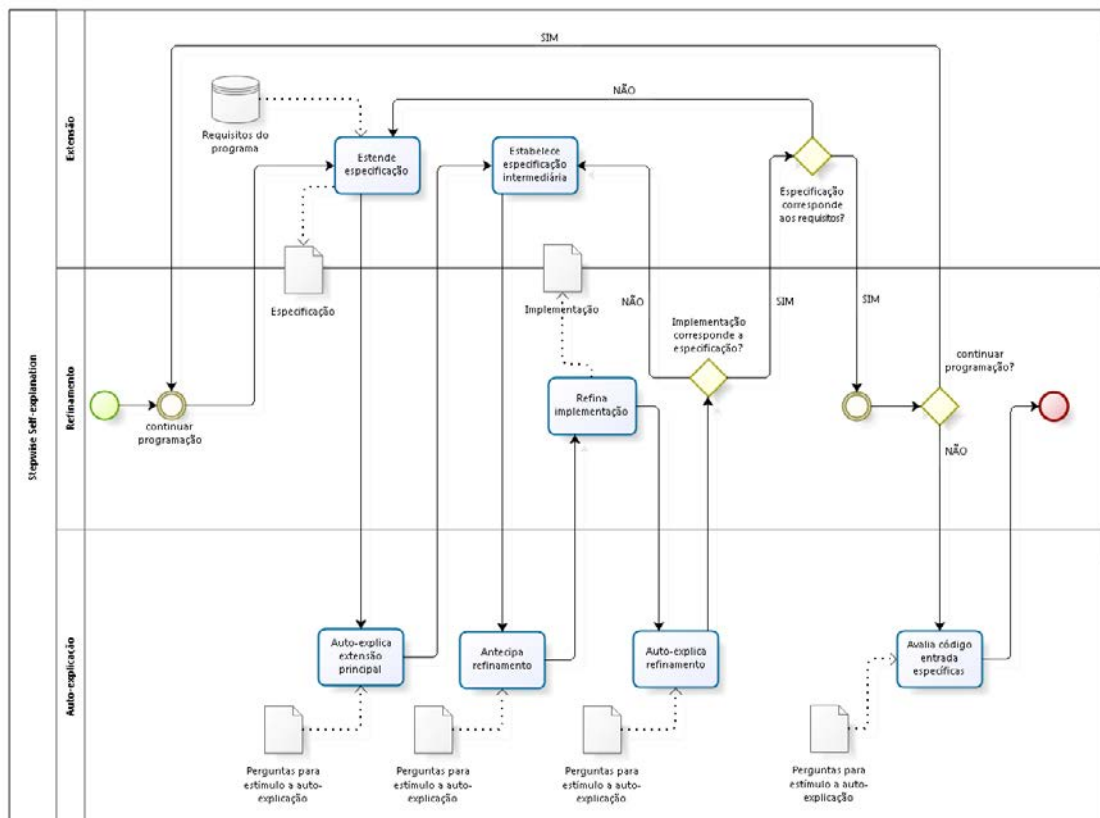


Figura 1. Abordagem Stepwise self-explanation.

Para ilustrar este procedimento, apresentamos na Figura 2 um trecho de um dos exemplos em vídeo produzidos segundo a abordagem, onde utilizamos a ferramenta *Scratch* para a construção de um jogo. Este trecho apresenta o passo “fazer com que o gato movimente-se para cima quando a seta para cima for pressionada” que é parte do problema “fazer com que o gato se movimente através das quatro setas direcionais”. Juntamente com o trecho de vídeo (Figura 2 (c)), apresentamos também as questões para os momentos anterior (Figura 2 (a)) e posterior (Figura 2 (b)) à implementação do passo de resolução do problema. A cada questão apresentada, o vídeo pausa; em seguida, a questão é carregada e o estudante continua o vídeo após a submissão de sua resposta.

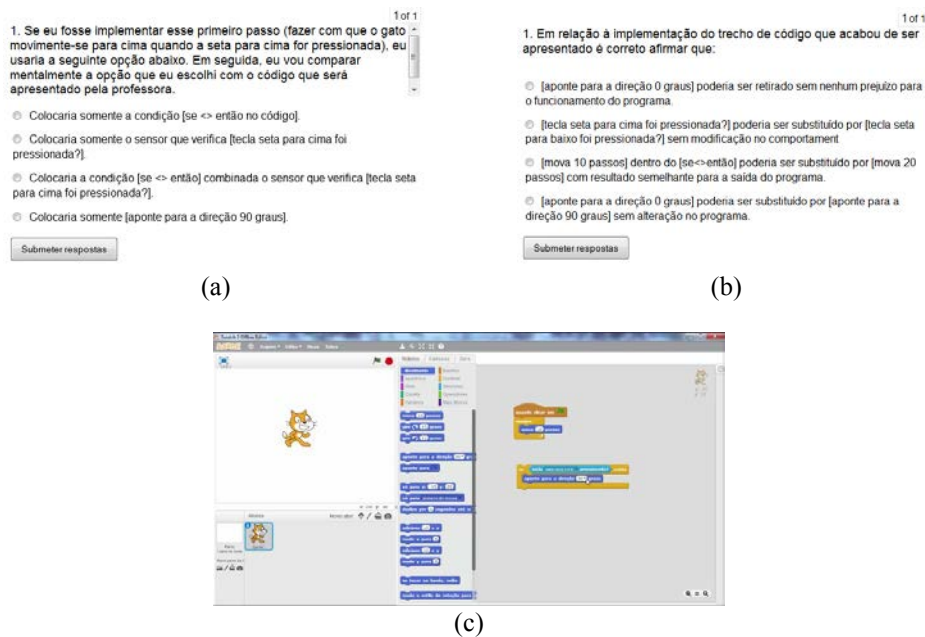


Figura 2. Trecho de exemplo em vídeo, apresentando passo da implementação do problema (c) juntamente com as questões anterior (a) e posterior (b) à implementação do passo do problema.

3.2. Experimentando a abordagem *Stepwise self-explanation*

Como forma de avaliar o potencial da abordagem *Stepwise self-explanation*, realizamos um experimento com estudantes de uma instituição de ensino que atua na educação de nível médio, técnico e superior no Estado de Pernambuco, no primeiro semestre de 2015. Este experimento foi realizado como um minicurso de introdução à programação através da construção de jogos digitais através da ferramenta *Scratch*. Portanto, o minicurso teve como objetivo de aprendizagem:

[OA1] Aplicar os conceitos de variáveis, estruturas de repetição e de decisão para a construção de um jogo usando a ferramenta *Scratch*.

Para alcançar este objetivo, a atividade de ensino do minicurso combinou períodos de aula expositiva seguidos por períodos de estudo individual dos exemplos em vídeo. As atividades de aprendizagem, por sua vez, consistiram em um conjunto de questões para promover a reflexão dos estudantes enquanto eles a partir dos exemplos em vídeo apresentados. Sendo assim, os materiais instrucionais produzidos para este experimento foram: (i) apresentação utilizada para guiar a aula expositiva; (ii) um conjunto de quatro vídeos que exibiam a aplicação dos conceitos apresentados na aula expositiva; (iii) pré e pós-testes para avaliar o desempenho dos estudantes antes e após o minicurso ministrado; (iv) um conjunto de questões para a AE.

Durante a execução do experimento, dois grupos foram organizados, o grupo controle (GC) e o grupo instrucional (GI). O GC fez uso de exemplos trabalhados em vídeos, enquanto que o GI fez uso dos mesmos exemplos em vídeo em conjunto com as questões de AE desenvolvidas para estes vídeos. Os resultados obtidos por ambos os grupos no pós-testes podem ser observados na Tabela 1. Com base nestes resultados, obtidos a partir das estatísticas descritivas e no âmbito do experimento realizado,

podemos dizer que aprender programação a partir da abordagem *Stepwise self-explanation* trouxe resultados superiores quando comparados à condição controle.

Tabela 1. Médias e desvios padrão (entre parênteses) dos resultados dos pré e pós-testes dos GC e GI.

	GC (n=5)	GI (n=7)
Pós-teste	8.2 (5.12)	10.57 (2.94)

Como complemento aos resultados positivos obtidos, os estudantes do GI e do GC tiveram uma opinião positiva a respeito da qualidade e utilidade dos exemplos e das explicações fornecidas pela professora nos exemplos em vídeo. Além disso, os estudantes do GI tiveram uma opinião positiva a respeito da qualidade e utilidade das questões empregadas para promover a AE dos vídeos.

4. Considerações finais

Neste trabalho, discutimos alguns desafios enfrentados pelo professor no planejamento e execução de atividades relacionadas aos processos de ensino e de aprendizagem de disciplinas introdutórias de programação, bem como apresentamos algumas oportunidades de pesquisa na área de Educação em Computação. Apresentamos a abordagem *Stepwise self-explanation* como uma proposta alternativa auxiliar a compor o projeto pedagógico e a definição das práticas pedagógicas para minimizar os desafios elencados. Além disso, demonstramos o potencial da abordagem através da realização de um experimento realizado como um minicurso de Introdução à Programação por meio da construção de jogos digitais e do feedback positivo fornecido pelos estudantes integrantes do experimento.

Cientes das restrições inerentes do estudo e, resgatando as dificuldades associadas ao trabalho de campo envolvendo docentes e discentes em ambiente escolar, pretende-se replicar a experiência em outras escolas e buscar trabalhar com segmentos específicos de alunos, organizando por faixa etária, tipo de curso e formação prévia. Entendemos que o trabalho realizado nos permitiu elencar indicadores de que o uso da auto-explicação pode colaborar para o aluno organizar seu pensamento e processos cognitivos, hábitos e atitudes. Entretanto, tudo isto funcionará se estiver devidamente contextualizado no planejamento do docente. Esta é uma das técnicas e atividades a serem consideradas no conjunto de possibilidades que temos para organizar o ensino de programação para iniciantes.

Referências

ATKINSON, R. K. et al. Learning from Examples: Instructional Principles from the Worked Examples Research. **Review of Educational Research**, v. 70, n. 2, p. 181-214, 2000.

AURELIANO, V. C. O.; TEDESCO, P. C. D. A. R. **Aprendendo linguagem de programação através da auto-explicação de exemplos em vídeo**. Proceedings of the IV Congresso Brasileiro de Informática na Educação. Maceió - Alagoas - Brasil 2015.

BENNEDSEN, J.; CASPERSEN, M. E. Failure rates in introductory programming. **SIGCSE Bull.**, v. 39, n. 2, p. 32-36, 2007.

- _____. Exposing the Programming Process. In: BENNEDSEN, J.; CASPERSEN, M., *et al* (Ed.). **Reflections on the Teaching of Programming**: Springer Berlin Heidelberg, v.4821, 2008. cap. 2, p.6-16. (Lecture Notes in Computer Science). ISBN 978-3-540-77933-9.
- BIGGS, J.; TANG, C. **Teaching For Quality Learning At University**. McGraw-Hill Education, 2011. ISBN 9780335242757.
- BÖRSTLER, J.; CASPERSEN, M.; NORDSTRÖM, M. Beauty and the Beast: on the readability of object-oriented example programs. **Software Quality Journal**, p. 1-16, 2015/02/14 2015.
- BRANSFORD, J. D.; BROWN, A. L.; COCKING, R. R. **How People Learn: Brain, Mind, Experience and School**. Washington, DC: {N.R. Council} National Academy Press, 2000.
- CASPERSEN, M. E. **Educating Novices in the Skills of Programming**. 2007. 323 DAIMI Aarhus University
- CASPERSEN, M. E.; KÖLLING, M. STREAM: A First Programming Process. **Trans. Comput. Educ.**, v. 9, n. 1, p. 1-29, 2009.
- CHERENKOVA, Y.; ZINGARO, D.; PETERSEN, A. **Identifying challenging CS1 concepts in a large problem dataset**. Proceedings of the 45th ACM technical symposium on Computer science education. Atlanta, Georgia, USA: ACM: 695-700 p. 2014.
- CHI, M. T. H. et al. Self-explanations: How students study and use examples in learning to solve problems. **Cognitive Science**, v. 13, p. 145-182+, 1989.
- CLARK, R. C.; NGUYEN, F.; SWELLER, J. **Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load**. Wiley, 2005. ISBN 978-0-7879-7728-3.
- DU BOULAY, B. Some difficulties of learning to program. In: (Ed.). **Studying the Novice Programmer**, 1989. p.431-446. ISBN 1317786203.
- GOMES, A.; MENDES, A. Learning to program-difficulties and solutions. **International Conference on Engineering Education–ICEE**, v. 2007, 2007.
- GOMES, A. J.; SANTOS, A. N.; MENDES, A. J. A study on students' behaviours and attitudes towards learning to program. Proceedings of the 17th ACM annual conference on Innovation and technology in computer science education, 2012. ACM. p.132-137.
- KINNUNEN, P.; MALMI, L. Why students drop out CS1 course? , Proceedings of the second international workshop on Computing education research, 2006. ACM. p.97-108.
- KURLAND, D. M. et al. A study of the development of programming ability and thinking skills in high school students. **Journal of Educational Computing Research**, v. 2, n. 4, p. 429-458, 1986.
- LAHTINEN, E. et al. **A study of the difficulties of novice programmers**. Proceedings of the 10th annual SIGCSE conference on Innovation and technology in computer science education. Caparica, Portugal: ACM: 14-18 p. 2005.

- MCCRACKEN, M. et al. **A multi-national, multi-institutional study of assessment of programming skills of first-year CS students.** Working group reports from ITiCSE on Innovation and technology in computer science education. Canterbury, UK: ACM: 125-180 p. 2001.
- MILNE, I.; ROWE, G. Difficulties in learning and teaching programming—views of students and tutors. **Education and Information technologies**, v. 7, n. 1, p. 55-66, 2002.
- PEARS, A. et al. A survey of literature on the teaching of introductory programming. **SIGCSE Bull.**, v. 39, n. 4, p. 204-223, 2007.
- PERKINS, D. N.; MARTIN, F. **Fragile knowledge and neglected strategies in novice programmers.** Papers presented at the first workshop on empirical studies of programmers on Empirical studies of programmers. Washington, D.C., USA: Ablex Publishing Corp.: 213-229 p. 1986.
- PUTNAM, R. T. et al. A summary of misconceptions of high school Basic programmers. In: (Ed.). **Studying the novice programmer**, 1989. p.301-314. ISBN 1317786203.
- RENKL, A. Worked-out examples: instructional explanations support learning by self-explanations. **Learning and Instruction**, v. 12, n. 5, p. 529-556, 10// 2002.
- ROBINS, A. V.; ROUNTREE, J.; ROUNTREE, N. Learning and Teaching Programming: A Review and Discussion. **Computer Science Education**, v. 13, n. 2, p. 137-172, / 2003.
- SAMURÇAY, R. The concept of variable in programming: Its meaning and use in problem-solving by novice programmers. **Studying the novice programmer**, v. 9, p. 161-178, 1989.
- SILVA FILHO, R. L. L. E. et al. A evasão no ensino superior brasileiro. **Cadernos de Pesquisa**, v. 37, p. 641-659, 2007.
- SPOHRER, J. C.; SOLOWAY, E. Novice mistakes: are the folk wisdoms correct? In: (Ed.). **Studying the novice programmer**, 1989. p.401-416. ISBN 0805800034.
- TEW, A. E.; MCCRACKEN, W. M.; GUZDIAL, M. Impact of alternative introductory courses on programming concept understanding. Proceedings of the first international workshop on Computing education research, 2005. ACM. p.25-35.
- VANLEHN, K. Cognitive skill acquisition. **Annu Rev Psychol**, v. 47, p. 513-39, 1996.
- WATSON, C.; LI, F. W. B. **Failure rates in introductory programming revisited.** Proceedings of the 2014 conference on Innovation & technology in computer science education. Uppsala, Sweden: ACM: 39-44 p. 2014.
- WINSLOW, L. E. Programming pedagogy—a psychological overview. **SIGCSE Bull.**, v. 28, n. 3, p. 17-22, 1996.

Desafios e Perspectivas de Um Bacharelado Interdisciplinar em TI

Daniel S. A. Araújo, Gustavo Girão, Jair Leite

Universidade Federal do Rio Grande do Norte (UFRN)
Av. Senador Salgado Filho, 3000 – Lagoa Nova, CEP: 59.078-970 – Natal/RN
{daniel,girao}@imd.ufrn.br, jair@dimap.ufrn.br

***Resumo.** Em 2013, a Universidade Federal do Rio Grande do Norte (UFRN) criou o Bacharelado em Tecnologia da Informação (TI), cujo objetivo é congrega os cursos e pesquisadores das subáreas da TI em torno de um modelo de curso em dois ciclos altamente ligado ao mercado de trabalho. A criação do curso seguiu a tendência dos bacharelados interdisciplinares promovida, principalmente, pelo REUNI e iniciada na UFABC. Os bacharelados interdisciplinares exigem desafios que se concentram em quesitos relacionados à flexibilidade curricular, à interdisciplinaridade e a insuficiência de conhecimento dos estudantes em áreas básicas. Com base nisso, este artigo tem como objetivo descrever algumas experiências adotadas para contornar os desafios dos bacharelados interdisciplinares no âmbito do Bacharelado em Tecnologia da Informação da UFRN. Como resultado da nossa experiência, foi possível avaliar a viabilidade das políticas e abordagens necessárias para implementar os requisitos de flexibilidade curricular e interdisciplinaridade, bem como reduzir os efeitos da deficiência em áreas como Matemática básica no Ensino Superior.*

1. Introdução

A interdisciplinaridade em cursos superiores no Brasil vem sendo discutida e implantada desde meados dos anos 2000. O projeto pedagógico da UFABC (UFABC, 2006) e o anteprojeto preliminar de reestruturação da arquitetura curricular da UFBA (UFBA, 2007), estimulados pelo programa REUNI do governo federal brasileiro, introduziram a proposta de bacharelados interdisciplinares com a ideia de fornecer uma formação generalista em uma determinada área e retardar a escolha dos conhecimentos específicos de um discente. Desde então, este modelo vem sendo implantado no Brasil nas mais diversas áreas.

Em 2013, a UFRN criou o Bacharelado em Tecnologia da Informação (BTI)¹, cujo objetivo é congrega os cursos e pesquisadores das subáreas da TI e de áreas externas à TI em torno de um modelo de curso em dois ciclos com o objetivo de favorecer flexibilidade curricular, interdisciplinaridade e maior aproximação do mercado de trabalho (LEITE, 2013). O modelo adotado segue a proposta de um

¹Mais detalhes sobre o curso podem ser vistos em: <http://portal.imd.ufrn.br/ensino/bti/>

bacharelado interdisciplinar integrando cursos já existentes há anos na universidade e disciplinas de outras áreas.

Após três anos desde a sua implantação, vários desafios surgiram durante a execução deste modelo curricular. O primeiro deles está associado às insuficiências de conhecimentos básicos em matemática nos alunos ingressantes. O segundo desafio é a gestão e manutenção da estrutura flexível e interdisciplinar de um curso com oito ênfases e dois ciclos de formação. Por fim, a formação voltada para a inovação e o empreendedorismo tem se mostrado um desafio que requer uma mudança de cultura de todos os integrantes.

Com base nisso, este artigo relata a experiência de implantação, gestão e evolução do curso de Bacharelado em Tecnologia da Informação. Para isso, vamos apresentar o contexto no qual o curso ocorre e como ele contribui de forma decisiva para a formação básica e avançada em TI de forma ampla, interdisciplinar, integrada e flexível. Esperamos como isso apresentar os acertos que esse modelo de curso vem proporcionando aos alunos, bem como ajustes que vem sendo feitos nestes três anos.

O artigo está estruturado da seguinte forma: a Seção 2 descreve o contexto no qual o curso está inserido. A Seção 3 aborda com detalhes o Bacharelado em Tecnologia da Informação e todas as suas características. Os desafios enfrentados, que são o foco deste artigo, são tratados na Seção 4. Algumas perspectivas para o futuro de iniciativas que podem ajudar ao BTI são apresentadas na Seção 5. Por fim, algumas conclusões e considerações finais são feitas na Seção 6.

2. O IMD e seus Percursos de Formação

O Instituto Metr pole Digital (IMD)   uma unidade acad mica especializada na UFRN que tem por miss o realizar a forma o de pessoas na  rea de Tecnologia da Informa o e promover o desenvolvimento de um ecossistema de inova o. Nesse sentido, o IMD (i) oferece forma o integrada, interdisciplinar, flex vel e de grande amplitude em TI, nos n veis t cnico, de gradua o e de p s-gradua o; (ii) realiza pesquisa cient fica e tecnol gica, bem como a oes de extens o que promovam a inova o por meio da transfer ncia de tecnologias para a ind stria de TI do pa s; e (iii) promove e apoia o surgimento, o crescimento e a inova o da ind stria de TI na regi o, atrav s de programas de est mulo ao empreendedorismo e de incubaq o de empresas emergentes.

Para cumprir sua miss o, o IMD oferece seus cursos baseados no conceito de percursos de forma o nos quais os estudantes podem optar por diversos caminhos que integram cursos t cnicos, de gradua o e de p s-gradua o. Todo esse percurso pode ser visto na Figura 1, onde o processo se inicia com busca por talentos em TI para ingresso nos cursos t cnicos e pode terminar em uma forma o altamente especializada ou com a gera o de um neg cio.

Os cursos t cnicos tem o objetivo de prospectar talentos para a  rea de TI e oferecer uma forma o b sica em computa o. O curso   oferecido em quatro m dulos: b sico, intermedi rio, avan ado e integrador. Atualmente, o curso t cnico conta com cinco habilita oes: T cnico em Inform tica para Internet, T cnico em Redes de Computadores, T cnico em Eletr nica, T cnico em Automa o Industrial e T cnico em Programa o de Jogos Digitais.

O Bacharelado em Tecnologia da Informação aparece como passo natural a ser seguido pelos discentes que concluem o nível técnico e aos demais alunos concluintes do ensino médio que se interessam por tecnologia. O BTI, por ser parte do tema principal deste artigo, será abordado com mais detalhes na próxima seção.

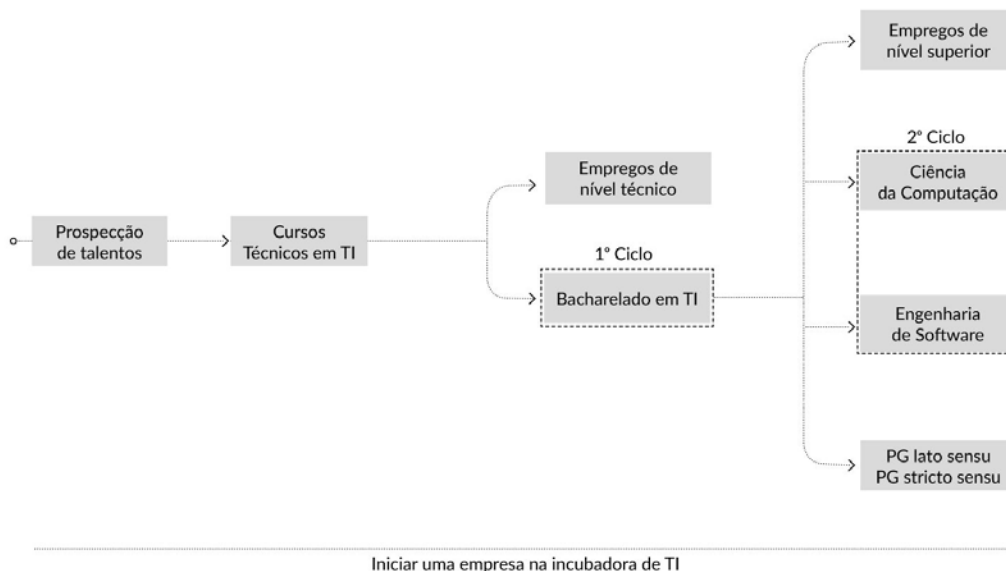


Figura 1. Percurso que um aluno pode percorrer dentro do IMD.

O BTI provê uma formação generalista e flexível na área de TI suficiente para o egresso atuar em uma alta variedade de funções no mercado de trabalho. No entanto, caso o concluinte queira aprofundar conhecimentos em uma área mais tradicional da computação, ele pode ainda ingressar em um dos cursos de segundo ciclo, ofertados em parceria com outro departamento da UFRN: o Bacharelado em Ciência da Computação ou Bacharelado em Engenharia de Software (MEC, 2012).

Ao concluir um curso de graduação (primeiro ou segundo ciclo), o aluno pode permanecer no IMD e cursar um curso de pós-graduação. Atualmente, o IMD conta com dois programas de pós-graduação: o Programa de Pós-graduação em Engenharia de Software (PPgSW) e o Programa de Pós-graduação em Bioinformática (PPgBio).

É importante salientar que a criação do programa PPgBio está completamente alinhada a criação da ênfase de Bioinformática dentro do BTI. Dessa forma, a cooperação de esforços de vários departamentos diferentes da Universidade, dada pela natureza altamente interdisciplinar da área, tende a construir um grande potencial na formação de recursos humanos a nível de graduação e pós-graduação.

Por fim, durante o processo de formação, o aluno tem contato direto com a Inova Metrópole, incubadora de empresas de TI do IMD, cuja missão é incentivar e promover o empreendedorismo e a inovação dessa área através da interação entre universidade, governo, empresa e sociedade em geral. O processo de incubação consiste em uma série de etapas durante as quais os envolvidos que detenham ideias ou produtos inovadores possam transformá-los em negócios socialmente relevantes, sustentáveis ou rentáveis. Esse apoio inclui salas individualizadas e compartilhadas, diversos tipos de laboratórios, capacitações, consultorias, assessorias e monitoramento da evolução da

empresa. A Inova possui forte colaboração com o SEBRAE e com a ANPROTEC, que são parceiros para estabelecimento de contato com a indústria de TI.

3. O Curso de TI

O Bacharelado em Tecnologia da Informação da UFRN segue as premissas apresentadas para um Bacharelado Interdisciplinar (UFBA, 2007), tendo como área de conhecimento principal a Tecnologia da Informação. A formação superior através dos bacharelados interdisciplinares caracteriza-se, principalmente, pela flexibilidade na sua matriz curricular. A partir do cumprimento de componentes curriculares obrigatórios, os discentes podem criar uma combinação própria de componentes curriculares optativos de acordo com as suas necessidades, interesses e méritos acadêmicos. Para tanto, deverão considerar as competências e habilidades profissionais que desejam desenvolver e aprimorar.

Sendo um bacharelado de caráter interdisciplinar, o BTI assume fortemente a interdisciplinaridade de duas formas. A primeira delas ocorre em seu processo de ensino-aprendizado. A segunda forma ocorre no processo de formação com 8 ênfases de livre escolha ao discente: Bioinformática, Ciência da Computação, Engenharia de Software, Informática Educacional, Produção de Jogos Digitais, Redes de Computadores, Sistemas de Informação de Gestão e Sistemas Embarcados.

Na interdisciplinaridade no ensino-aprendizado, busca-se uma formação humana baseada na superação da fragmentação do conhecimento de forma que os educandos possam articular outros saberes e práticas abordadas nos diferentes componentes curriculares do curso. A interdisciplinaridade rompe com o modelo de formação disciplinar existente na Educação Superior Brasileira e colabora para a resolução de problemas complexos que se apresentam no mundo do trabalho. Daí a importância da constante revisão das práticas educativas e do currículo, considerando também a fugacidade do conhecimento e o caráter inovador dos campos científicos, tecnológicos e sociais.

Sendo assim, o Bacharelado em Tecnologia da Informação assume quatro princípios basilares para sua proposta metodológica, os quais estão em consonância com os Referenciais Orientadores para Bacharelados Interdisciplinares (MEC, 2010). São eles: (i) formação baseada na interdisciplinaridade e no diálogo entre as áreas de conhecimento e os componentes curriculares; (ii) foco nas dinâmicas de inovação científica, tecnológica, artística, social e cultural, associadas ao caráter interdisciplinar dos desafios e avanços do conhecimento; (iii) permanente revisão das práticas educativas tendo em vista o caráter dinâmico e interdisciplinar da produção de conhecimentos; (iv) prática integrada da pesquisa e extensão articuladas ao currículo.

Em suma, aliado a práticas voltadas para o diálogo crítico, reflexivo e participativo dos discentes, bem como a adoção pelos docentes de um planejamento coletivo de conteúdos, elaboração de materiais didáticos, atividades e instrumentos avaliativos, todas essas iniciativas metodológicas visam uma concepção de ensino em que haja a reflexão sobre “o que, como, porque, para que e para quem” estamos desenvolvendo as Tecnologias da Informação e Comunicação.

O percurso de formação dentro do BTI pode ser visto na Figura 2. Nele, os discentes possuem um núcleo de componentes curriculares básicos e comuns à sua formação em TI que permanece durante os quatro primeiros semestres de sua formação. A partir daí, cada aluno pode escolher uma das ênfases ofertadas pelo curso ou, até mesmo, pode montar seu próprio percurso particular escolhendo componentes de diferentes áreas. Note que, após a conclusão do BTI, o aluno ainda pode ingressar em um curso de segundo ciclo e com mais dois ou três semestres, ele obterá o diploma de Bacharel em Engenharia de Software ou Bacharel em Ciência da Computação.



Figura 2. O curso de Bacharelado em Tecnologia da Informação.

A natureza de escolha dos componentes que fazem parte da formação do aluno do BTI está totalmente de acordo com os Referenciais para Bacharelados Interdisciplinares (MEC, 2012), que prezam pela flexibilidade e autonomia na escolha do estudante no que se refere ao seu perfil de formação.

4. Desafios

O Bacharelado em Tecnologia da Informação conta, atualmente, com aproximadamente 900 alunos em seu quarto ano de funcionamento. Devido à sua dimensão e característica interdisciplinar, vários desafios precisam ser superados. Nesta seção, serão apresentados os desafios mais relevantes e as ações realizadas no sentido de garantir um ensino de qualidade em um bacharelado interdisciplinar em TI.

4.1 Nivelamento

No ano de 2013, a estrutura curricular do BTI encontrava-se em sua primeira versão. Neste modelo, o aluno tinha uma expectativa de conclusão de 6 semestres no turno diurno e 8 semestre no turno noturno, sendo os 3 primeiros considerados um núcleo comum de formação. Após a experiência com a primeira turma (ano de 2013), percebeu-se que os alunos tinham uma deficiência em conhecimentos matemáticos do ensino médio. Além disso, a pouca capacidade de aplicar o formalismo matemático exigido no nível superior fez com que o Núcleo Docente Estruturante do BTI tivesse que repensar a estrutura curricular existente.

Com o objetivo de superar este desafio de capacitar os estudantes com as necessidades matemáticas no ensino superior, foi criada uma disciplina chamada *Resolução de Problemas Matemáticos para Tecnologia da Informação*. Esta é uma

disciplina de 180 horas, das quais 150 são direcionadas ao ensino de matemática e 30 horas para uma intervenção de TI. Nesta intervenção, um professor da área periodicamente demonstra aos alunos como os conhecimentos matemáticos adquiridos são utilizados na tecnologia da informação.

Passados dois anos desde a implantação desta disciplina, foi feito um breve estudo quantitativo sobre o impacto no desempenho dos alunos. Para isso, levou-se em consideração o desempenho dos alunos na disciplina de programação do segundo semestre (primeira vista pelos alunos durante o curso). O gráfico da Figura 3 apresenta o percentual de alunos aprovados nesta disciplina de programação. Os anos de 2014 e 2015 representam resultados após a implantação da disciplina de nivelamento e, no caso de 2014, percebe-se uma melhora no desempenho dos alunos considerando-se um aumento de 4% no número de alunos aprovados. Entretanto, em 2015 verifica-se uma piora nestes números. Esta variação será investigada em maiores detalhes pela coordenação do curso juntamente com a equipe pedagógica ao longo dos próximos anos para que se tenha uma melhor noção de como esta disciplina de nivelamento impacta no desempenho dos alunos.

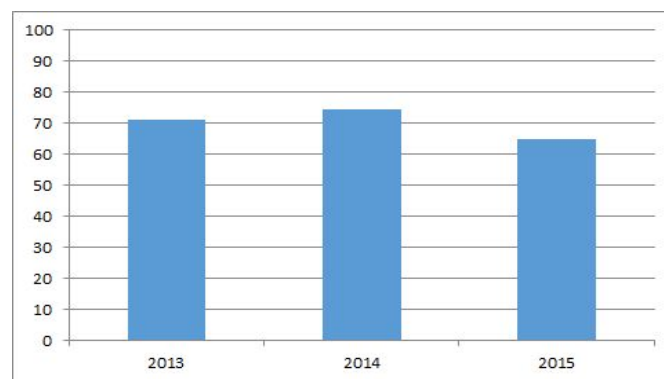


Figura 3. Percentual de alunos aprovados em disciplina de programação.

4.2 Interdisciplinaridade

Para acomodar esta disciplina na estrutura curricular do curso, foi criado um novo semestre inicial (aumentando o número de semestres para 7 no turno diurno e 9 no turno noturno) que comporta, além desta, outras disciplinas básicas que correlacionam o estudo de TI com outras áreas de conhecimento, tais como: Língua Portuguesa, Língua Inglesa e Tecnologia da Informação e Sociedade. Essas disciplinas são planejadas de maneira a integrar as atividades desenvolvidas em sala. Os alunos trabalham um tema central a ser desenvolvido no âmbito de Língua Portuguesa, Língua Inglesa e Tecnologia da Informação e Sociedade, exercitando as correlações entre as áreas com o objetivo de gerar um novo conhecimento onde não existe desassociação entre elas, gerando assim a interdisciplinaridade. Além de criar um degrau intermediário de nivelamento, este primeiro semestre estimula uma cultura de interdisciplinaridade tanto nos discentes quanto nos docentes.

Dado este esforço interdisciplinar no início do curso, existe a expectativa de que este tipo de atividade se dissemine ao longo do mesmo. Nos semestres seguintes, o aluno está apto a cursar disciplinas mais avançadas de matemática e disciplinas iniciais de programação. Esta é uma ótima oportunidade para aplicar parte das ideias discutidas

na intervenção de Tecnologia da Informação do semestre anterior. No intuito de catalisar este processo, a coordenação Bacharelado em Tecnologia da Informação juntamente com a equipe pedagógica promove anualmente um evento chamado SeIntegraBTI que utiliza uma dinâmica de Grupos de Trabalho para reunir professores de disciplinas de mesmo semestre para que estes discutam alternativas interdisciplinares no ensino de seus componentes curriculares.

4.3 Gerenciamento de ênfases

O BTI conta atualmente com 8 ênfases (conforme apresentado na Seção 3), as quais o aluno pode escolher seguir ao longo da sua graduação. Como mencionado anteriormente, esta escolha tardia tem o intuito de evitar escolhas erradas por parte do aluno e diminuir assim o índice de evasão. Além disso, com as ênfases, é possível oferecer disciplinas especializadas em diferentes subáreas da TI e em áreas afins e complementares. A necessidade de cursos flexíveis que permitam a formação em diversos perfis da TI já é objeto de estudos e proposta há quase 10 anos (Ferreira e Silva, 2007).

A oferta de um número grande de ênfases (oito, no caso do BTI) é um desafio que exige a capacidade de gerenciar recursos físicos (salas de aula) e humanos (professores) em grande escala. Todas estas ênfases são gerenciadas de forma tal que cada uma delas tem sua própria estrutura curricular, contendo um conjunto de disciplinas obrigatórias e optativas. Objetivando dar oportunidade para que o aluno obtenha conhecimento de outras ênfases, adotou-se por padrão fazer com que todas as disciplinas obrigatórias de todas as ênfases pudessem ser cursadas como disciplinas optativas de uma outra ênfase. Assim, acredita-se que há um melhor aproveitamento das disciplinas ministradas considerando a relação de número de alunos por turma.

A oferta das disciplinas obrigatórias das ênfases pode gerar um número consideravelmente grande (em torno de 125 turmas de disciplinas obrigatórias ofertadas em 2016.1). Para que esta oferta possa ser feita de maneira eficiente, é necessário que o corpo docente tenha sido formado considerando os conhecimentos específicos necessários para cada uma das estruturas curriculares. Informalmente, dentro do grupo docente, os professores são divididos em grupos destinados a lecionar disciplinas de uma dessas 8 ênfases. Periodicamente, os grupos se reúnem para decidir qual será a oferta da ênfase e que professores lecionaram quais turmas. Esta etapa é feita com uma participação mínima da coordenação. De maneira hierárquica, a coordenação recebe as sugestões de cada ênfase e trabalha na elaboração da oferta final considerando uma visão macro da oferta das disciplinas.

Do ponto de vista da oferta das ênfases, a coordenação do curso, juntamente com o setor pedagógico, organiza semestralmente o Seminário de Orientação às Ênfases do BTI. O objetivo deste seminário é apresentar em detalhes para os alunos quais são as características de cada ênfase tais como: disciplinas que fazem parte da estrutura curricular, perfil do egresso e oportunidades de atuação no mercado. Neste seminário são convidados professores das ênfases e profissionais locais que atuam na área em questão para falar de suas experiências. Com isso, espera-se que o aluno tenha mais segurança ao escolher uma das ênfases em um determinado momento do curso.

4.4 Formação para Inovação

Como forma de incentivar uma formação de pessoas que buscam inovar na sua atuação profissional, o IMD tem estimulado ações de extensão e pesquisa.

Ao compartilhar um mesmo ambiente com a incubadora Inova, percebeu-se que o processo de interação entre a comunidade acadêmica (docentes e discentes) e empresas incubadas tornou-se mais fluido. Através de atividades expositivas e demonstrativas, foi possível encontrar trabalhos em comum que tem potencial de gerar produtos e patentes. Duas destas atividades são a Mostra de Profissões e o Workshop do IMD, realizados anualmente. Na Mostra de Profissões os cursos são convidados a realizar palestras e as empresas incubadas são convidadas a realizar demonstrações de seus produtos e atividades. Por outro lado, no Workshop, os laboratórios de pesquisa do IMD são convidados a promoverem palestras e minicursos para a comunidade. Este processo despertou o interesse de empresas incubadas que manifestaram interesse em parcerias no desenvolvimento de tecnologias inovadoras.

Como forma de estimular a pesquisa, o IMD escolheu um tema de integrador e de amplo espectro com o objetivo de agregar iniciativas de projetos de desenvolvimento tecnológico e inovação: Cidades Inteligentes e Humanas. Esse tema, além de ser bastante contemporâneo por buscar resolver problemas reais existentes em cidades de todo o mundo, utiliza tecnologias de última geração como Internet das Coisas, *Big Data*, Computação em Nuvem, Sistemas de Sistemas e outras. Com isso, o IMD estimula pesquisa que envolvem os docentes do instituto e dos departamentos associados como também oferece possibilidades de bolsas para alunos.

Nesse sentido, o IMD está executando desde Janeiro de 2016 o projeto *SmartMetropolis*. Neste projeto, 18 estudantes do curso de TI e 12 egressos dos outros cursos de computação já estão atuando como bolsistas pesquisadores no desenvolvimento de tecnologias e soluções inovadoras. Espera-se ainda que as soluções sejam objetos de transferência tecnológica para empresas parceiras ou que gerem novos empreendimentos como *spin-offs* na incubadora INOVA.

A ideia de um tema comum como estímulo a integração curricular já ocorre no curso de Engenharia de Computação da UEFS (Bittencourt e Figueiredo, 2003). No entanto, a nossa proposta vai além por tratar o tema integrador como um desafio de pesquisa concretizados com work packages de pesquisa sob a responsabilidade de grupos de professores e empresas incubadas.

A integração com a incubadora de empresas e a vertente de cidades inteligentes desenvolvida pelo IMD trazem para os discentes do BTI uma visão muito mais realista do desenvolvimento de tecnologia e fornece incentivos para a permanência do aluno no curso e atraem a atenção para novas demandas de alunos. Dessa forma, o BTI busca maximizar a relação existente entre os conteúdos dos componentes curriculares trazendo os conceitos de forma transversal durante todo o curso.

5. Perspectivas

O curso de TI da UFRN não foi criado como um curso isolado. Descrevemos anteriormente que ele, além de integrar os antigos cursos existentes de Ciência da Computação e Engenharia de Software, está também integrado aos cursos técnicos e de pós-graduação, bem como ao processo de incubação de empresas da Inova Metrôpole.

Além disso, outras ações já estão em andamento para ampliar ainda mais esta integração. Nesta seção, vamos citar a implantação de um Parque Tecnológico que visa ampliar oportunidades de emprego e empreendedorismo e a integração do cursos de Engenharia de Computação da UFRN que ainda não está como opção de segundo ciclo.

5.1 O Parque Tecnológico

O IMD apresentou à prefeitura da cidade, uma proposta de parque tecnológico urbano, nos arredores do campus universitário onde ele está situado, para empresas privadas com atividade preponderante nos segmentos de software e hardware que contarão com benefícios fiscais, como reduções de IPTU, ISS e ITIV e taxa de localização. A proposta é, com o apoio do governo municipal, fomentar um ecossistema de empresas de TI que possam colaborar entre si e com os pesquisadores da universidade buscando o desenvolvimento econômico e a inovação tecnológica. Tal ecossistema deverá aumentar a oferta de vagas para os profissionais formados no curso de TI, ao mesmo tempo que também permite a estes novos profissionais a possibilidade de empreender e fazer parte desta cadeia produtiva.

5.2 Novas ênfases e Integração com a Engenharia de Computação

Uma nova proposta de estrutura curricular esta sendo estudada de modo a permitir que a inclusão de novas ênfases possa ser conciliada com o espaço físico disponível e a capacidade do corpo docente. A ideia principal da proposta é fazer com que as próprias ênfases tornem-se flexíveis no que diz respeito a oferta de seus componentes curriculares. Dessa forma, não existiriam disciplinas obrigatórias e sim um grupo das quais os alunos precisam cumprir uma determinada porcentagem sem se estabelecer precisamente quais seriam elas.

Um desafio importante que precisamos mencionar é a não integração do curso de Engenharia de Computação (EC) com o BTI até o momento. O curso de EC surgiu em 1995 como uma alternativa ao já existente curso de Ciência da Computação, com um foco em Sistemas e em Controle e Automação. Com a adoção do modelo de bacharelado interdisciplinar, a partir de 2009, no Bachelarelado em Ciência e Tecnologia (BCT), o curso de EC passou a ser um curso de segundo ciclo junto com vários outros cursos de engenharia. Esse arranjo foi favorecido porque o BCT possui um conjunto de disciplinas comuns a estas engenharias, tais como os cálculos, as físicas, fenômenos de transporte e outras. Esse compromisso inicial da EC com o BCT e a falta das disciplinas de engenharia no núcleo comum no nosso curso de TI vem dificultado esta integração. No entanto, já há entendimentos para se buscar alternativas e esperamos apresentar uma proposta em breve.

6. Conclusão

A área de TI exige profissionais com diferentes perfis profissionais e um grande desafio nas universidades é definir o perfil dos seus cursos. Além disso, diversas novas áreas estão surgindo integrando diretamente a TI como parte fundamental e essencial da formação. Sistemas de Informação de Gestão, Sistemas Embarcados e Bioinformática são exemplos destas áreas.

Este artigo apresentou um modelo de curso inspirado nas propostas de bacharelado interdisciplinar que visa a integração de diferentes cursos de computação e de outras áreas de conhecimento. O Bacharelado em TI foi criado utilizando um modelo flexível que permitiu integrar os cursos existentes de Ciência da Computação e Engenharia de Software a novas ênfases em áreas estratégicas. Esse modelo de curso inovador mostra-se com um promissor caminho para a formação de profissionais em uma área em constante evolução e ampliação.

Com as experiências descritas, foi possível notar que o conceito de interdisciplinaridade pôde ser muito bem aplicado dentro de um mesmo componente curricular ou unindo vários componentes do curso para um esforço conjunto com o objetivo de dar ao aluno a oportunidade de construir seu conhecimento de uma forma mais ampla e crítica. Também foi possível constatar que as deficiências relativas ao domínio de conteúdos básicos, já conhecidas nos cursos superiores, podem ser trabalhadas nas Universidades de forma não só a suprir essas deficiências, mas a aprimorar a iniciação do discente no nível superior e motivar a sua permanência nele.

Por fim, o ambiente integrado de vários níveis de ensino, pesquisa e mercado de trabalho, aliado as novas perspectivas de projetos do IMD, mostram que é possível prospectar talentos de jovens interessados na área e garantir um futuro profissional que atende aos requisitos de excelência, tanto para o meio acadêmico como para esse mercado de trabalho exigente e dinâmico.

Referências

- Bittencourt, R. A., and O. A. Figueiredo. "O currículo do curso de engenharia de computação da UEFS: Flexibilização e integração curricular." *XI Workshop sobre Educação em Computação 2003*.
- Ferreira, A. P. L. and Silva, D. B. "Projetos Flexíveis de Cursos de Graduação" *XV Workshop sobre Educação em Computação 2007*.
- LEITE, J. C. ; GOMES, A. V. ; SILVA, B. S. ; Rego, J. I. . , 2013, Maceió. Anais do XXI Workshop sobre Educação em Computação. Porto Alegre, RS : Sociedade Brasileira de Computação, 2013. v. 1. p. 701-710.
- MINISTÉRIO DA EDUCAÇÃO CONSELHO NACIONAL DE EDUCAÇÃO PROCESSO Nº: 23001.000026/2012-95 Diretrizes Curriculares Nacionais para os cursos de graduação na área da Computação (2012).
- _____. Referenciais Orientadores para os Bacharelados Interdisciplinares e Similares: documento elaborado pelo Grupo de Trabalho instituído pela Portaria SESu/MEC No. 383 de 12 de abril de 2010.
- UFABC Universidade Federal do ABC Projeto Pedagógico. (2006) <http://www.ufabc.edu.br/images/stories/pdfs/institucional/projetopedagogico.pdf> (acessado em 07 de março de 2016)
- UFBA Plano de Expansão e Reestruturação da Arquitetura Curricular na Universidade Federal da Bahia - Termo de Referência (2007). <http://www.cpa.ufba.br/node/37> (acessado em 07 de março de 2016)

DuinoBlocks for Kids: um ambiente de programação em blocos para o ensino de conceitos básicos de programação a crianças do Ensino Fundamental I por meio da Robótica Educacional

Rubens Lacerda Queiroz¹, Fábio Ferrentini Sampaio^{1,2}

¹Programa de Pós-graduação em Informática (PPGI/UFRJ)

²Instituto Tércio Pacitti de Pesquisas e Aplicações Computacionais.– Universidade Federal do Rio de Janeiro (UFRJ)
Cidade Universitária - Ilha do Fundão - CEP 20001-970, Rio de Janeiro, RJ, Brasil

rubensqueiroz@outlook.com, ffs@nce.ufrj.br

***Abstract.** Today there are different relevant initiatives to promote code learning by young children. One of them is based on the association of two successful strategies: Educational Robotics and Visual Programming Environments. This paper presents “DuinoBlocks for Kids”, a Visual Programming Language based on free technology focusing on teaching some basic concepts of computer programming for elementary school children, allowing them to perform robotic experiments both in a remote laboratory - through a web application – and locally, using an Arduino board connected to the user’s computer.*

***Resumo.** Existem hoje algumas iniciativas bastante relevantes no sentido de viabilizar o ensino de programação de computadores já nos primeiros anos do ensino fundamental, sendo uma delas baseada na associação de duas estratégias muito bem sucedidas: a Robótica Educacional e a utilização de Ambientes de Programação Visual baseados em blocos de encaixar. Este trabalho apresenta o DuinoBlocks for Kids, um Ambiente de Programação em Blocos baseado em tecnologia livre e voltado ao ensino de conceitos básicos de programação a crianças do Ensino Fundamental I, permitindo a manipulação de experimentos em Robótica tanto em um laboratório remoto - via aplicação web - quanto localmente.*

1. Introdução

Na sociedade informatizada em que vivemos as crianças já nascem imersas no mundo digital, sendo conhecidas, devido à sua aparente fluência com as novas tecnologias, como “nativos digitais” [Prensky 2001].

Entretanto, Mitchel Resnick (2014) argumenta que, para se adquirir fluência em relação às novas tecnologias, não é suficiente saber usar com naturalidade as Tecnologias de Informação e Comunicação (TIC), é necessário que se saiba “projetar, criar e se expressar através dessas tecnologias” e, para tanto, é imprescindível saber programar.

Um maior domínio das novas tecnologias que é, como mencionado anteriormente, potencializado por meio da aquisição da habilidade de saber programar

computadores, permite aos alunos utilizarem com mais propriedade e eficiência o computador como ferramenta de apoio ao aprendizado das disciplinas do currículo escolar.

Além disso, a aprendizagem de programação trabalha habilidades e competências ligadas ao raciocínio lógico e matemático, úteis a muitos outros aprendizados [Resnick et al. 2009; Buechley, Eisenberg & Elumeze 2007; Barbero & Demo 2011] e ao desenvolvimento do Pensamento Computacional - uma forma de pensamento característica dos cientistas da computação, mas, universalmente aplicável, que envolve um conjunto de atitudes e habilidades tais como o uso da recursividade, abstração e decomposição na solução de problemas tanto técnico-científicos quanto da vida cotidiana [Wing 2006].

Tais características tornam bastante nítida a relevância da inclusão do ensino de programação de computadores como parte integrante do currículo básico desde as primeiras séries do ensino fundamental. E, para que esse aprendizado possa ocorrer de forma eficiente, eficaz e prazerosa, faz-se necessário o desenvolvimento de ferramentas que deem o suporte adequado a este processo. O DuinoBlocks for Kids vem exatamente ao encontro dessa demanda.

2. Fundamentação Teórica

2.1. Robótica Educacional

A ideia do uso da Robótica em educação se baseia fortemente na Teoria Construcionista de Seymour Papert, que une a Teoria Construtivista de Jean Piaget - com quem trabalhou entre 1958 e 1963 na Universidade de Geneva, Suíça - ao uso do computador na educação [Pinto 2011].

Papert (1980) adicionou à teoria de Piaget (1974) a ideia de que a construção do conhecimento acontece de maneira mais efetiva quando o aprendiz se engaja, de forma consciente, na construção de objetos, não importando se físicos ou virtuais, desde que possam ser vistos e analisados e façam parte do universo de interesses da criança [Alimisis 2013], como é o caso dos robôs que, naturalmente, instigam a sua curiosidade [Perez et al. 2013]

A Robótica Educacional permite à criança manipular e controlar objetos concretos e, através destes objetos, observar a materialização dos comandos dados por ela ao computador, processo a partir do qual se estabelece a construção do seu conhecimento.

Por tratar-se de uma disciplina de caráter multidisciplinar, a Robótica permite que os alunos trabalhem uma grande diversidade de competências e habilidades à medida que engloba, em um único objeto de estudo, diversas áreas do conhecimento, como matemática, eletrônica, mecânica, inteligência artificial, artes e programação, sendo esta o foco mais específico deste trabalho.

Além disso, a aprendizagem por meio da Robótica estimula a colaboração, a habilidade de trabalhar em grupo, a elaboração de estratégias para a solução de problemas e o desenvolvimento do pensamento formal [Souza & Duarte 2015, López-Rodríguez & Cuesta 2015; Perez et al. 2013; Rus 2006; Denis & Hubert 2001].

2.2. Linguagens de Programação Visual para Crianças

Junto com os primeiros computadores pessoais (PCs), no fim dos anos 70, surgiu o interesse de se utilizar esses equipamentos nas escolas para o aprendizado de programação.

No entanto, a dificuldade de compreensão, por parte das crianças, da sintaxe das linguagens de programação existentes na época, bem como a não conexão dos programas desenvolvidos com os interesses delas [Resnick et al. 2009] acabaram contribuindo para o insucesso de muitas dessas iniciativas.

Desde então, novas iniciativas surgiram no sentido de se tentar vencer estas dificuldades. No que diz respeito ao entendimento da sintaxe das linguagens de programação textuais, uma alternativa encontrada foi o uso de Linguagens de Programação Visuais (Visual Programming Language - VLP), ou seja, linguagens nas quais “a sintaxe (semanticamente significativa) inclui expressões visuais” [Burnett 1999, p.1].

Com base nesse paradigma, o MIT Media Lab (<https://llk.media.mit.edu/>), iniciou, em 2003, o desenvolvimento do Scratch (<https://scratch.mit.edu/>), uma Linguagem de Programação Visual baseada em “blocos de encaixar” que tinha por objetivo permitir que qualquer pessoa, de qualquer idade, pudesse programar. O sucesso do projeto foi tal que, em 2009, apenas dois anos depois do lançamento do site, usuários de todas as partes do mundo, na sua maioria crianças e jovens de 8 a 16 anos, já faziam upload de mais de 1500 projetos por dia.

Além do Scratch, outros projetos de grande vulto, voltados ao ensino de programação para crianças, também adotam o conceito da Programação Visual por meio de blocos de encaixar. Dentre eles podemos citar o Code.org (<https://code.org/>), que possui suas próprias ferramentas de ensino de programação por blocos e o Programaê (<http://programae.org.br/>) e Code Club Brasil (<http://codeclubbrasil.org/>), que utilizam o Scratch em seus programas de aprendizado de programação.

A adoção, por esses projetos, de Ambientes de Programação Visual baseada em blocos, cujas combinações são bastante intuitivas (uma vez que as crianças comumente possuem brinquedos que trazem esta ideia) [Jr. & Guedes 2015], sugerem ser esta uma abordagem bastante acertada no que diz respeito ao ensino de programação para crianças, o que veio a motivar o uso deste mesmo paradigma no desenvolvimento do DuinoBlocks for Kids.

2.3. Software e Hardware Livres

O Movimento Software Livre é "baseado no princípio do compartilhamento do conhecimento e na solidariedade praticada pela inteligência coletiva" [da Silveira 2003, p.438].

Existem atualmente disponíveis em repositórios de Software Livre como o GitHub (<https://github.com/>), um número incalculável de softwares, cujos códigos podem ser livremente baixados, utilizados, modificados e redistribuídos, permitindo a qualquer pessoa auxiliar, em qualquer tempo, na melhoria dos softwares lá disponibilizados, bem como desenvolver, individual ou coletivamente, novos programas baseados nos códigos existentes.

Outro movimento relacionado à democratização do acesso à tecnologia, é o Movimento Hardware Livre. Hardware Livre é um hardware eletrônico projetado e oferecido da mesma maneira que um Software Livre. Pode-se citar como exemplo de projetos desse tipo o Arduino (<http://www.arduino.cc/>), utilizado neste trabalho, e o Raspberry Pi (<https://www.raspberrypi.org>).

Um aspecto bastante importante em relação ao desenvolvimento de aplicações a partir do uso de Software Livre e da associação destes ao uso de hardware de baixo custo, em contraponto ao uso de soluções proprietárias, é a democratização do acesso à tecnologia.

2.4. Arduino

Arduino (<http://www.arduino.cc/>) é um projeto de Hardware Livre desenvolvido por um grupo de 5 pesquisadores de diferentes nacionalidades (Massimo Banzi, David Cuartielles, Tom Igoe, Gianluca Martino e David Mellis) no Interaction Design Institute Ivrea (<https://interactionivrea.org/>), que tinha por objetivo a criação de uma plataforma de prototipagem eletrônica de código aberto, simples de ser utilizada por qualquer estudante, inclusive crianças, de maneira que elas pudessem facilmente criar “coisas” que funcionassem sem a necessidade de frequentar um curso completo de eletrônica.

O seu baixo custo, qualidade, flexibilidade e facilidade de uso, têm feito das placas de prototipagem eletrônica Arduino uma excelente opção para o desenvolvimento de projetos na área de Robótica Educacional.

3. Trabalhos Relacionados

Não foi encontrado na literatura um Ambiente de Programação Visual voltado à programação de placas de prototipagem eletrônica Arduino (tecnologia adotada neste trabalho), pensado e desenvolvido especificamente para o ensino de programação a crianças do Ensino Fundamental I. Ou seja, um Ambiente de Programação Visual composto por blocos graficamente mais atrativos e dotados de uma semântica menos abstrata que aquelas empregadas nos ambientes de programação que trabalham com o mesmo hardware.

Como exemplos de Ambientes de Programação Visual para placas Arduino atualmente existentes, podemos citar o DuinoBlocks [Alves 2013], que serviu de inspiração para o desenvolvimento do DB4K; O S4A - Scratch for Arduino (<http://s4a.cat/>), um fork do Scratch que permite a programação de placas Arduino; e o Ardublockly (<https://github.com/carlosperate/ardublockly>), ambiente sobre o qual foi desenvolvido o DuinoBlocks for Kids.

Algumas características do DuinoBlocks for Kids que o diferenciam dos ambientes acima listados e que buscam tornar o seu uso mais acessível e aprazível a crianças do Ensino Fundamental I são:

- Blocos de programação desenhados de modo a apresentarem uma semântica diretamente relacionada com os dispositivos sendo manipulados e com os efeitos por eles causados sobre estes dispositivos.
- Supressão de detalhes relacionados ao hardware, como pinagens e valores de níveis de tensão.

- Uso de linguagem icônica.
- Conjunto "enxuto" de blocos.
- Simplificação dos valores dos parâmetros utilizados nos blocos, como por exemplo: Temperatura: alta/baixa; Luz: muita/pouca; Velocidade: rápida/média/devagar.

4. DuinoBlocks for Kids (DB4K)

O Grupo de Informática Aplicada à Educação (GINAPE) da UFRJ vem, desde 2008, desenvolvendo o projeto LabVad - Laboratório Virtual de Atividades Didáticas em Ciências e Robótica (<http://labvad.nce.ufrj.br/>), cujo objetivo principal é o de, a partir de pesquisas acerca da utilização das TIC na educação, fomentar práticas pedagógicas que tirem proveito da ideia de "aprender fazendo", desenvolvendo para tanto atividades experimentais [Pinto 2011].

A estratégia utilizada pelo LabVad para auxiliar na viabilização da prática de atividades experimentais, tão importantes na Educação Básica, sobretudo nas disciplinas de ciências naturais (Física, Química e Biologia), foi a utilização de Laboratórios Remotos, ou seja, experimentos reais sendo executados a distância por meio de computadores ligados à internet [Pinto 2011].

A utilização dos laboratórios remotos é uma solução que possibilita a execução de experimentos reais sem que para isso seja necessária a aquisição de equipamentos especiais ou alocação de espaço para a sua realização.

Dentre outros experimentos, o ambiente LabVad possui um laboratório de Robótica Educacional denominado LabVad Robótica [Souza et al. 2014], que busca fornecer ao usuário uma experiência de programação para Robótica muito próxima daquela que ele teria se estivesse realizando um experimento localmente.

A figura 1 apresenta a imagem vista pelo usuário, por meio de um vídeo em tempo real apresentado na interface do LabVad Robótica, durante a realização dos experimentos.



Figura 1. Imagem vista pelo usuário por meio do vídeo apresentado pelo LabVad Robótica

O labVad Robótica foi desenvolvido inicialmente visando os públicos jovem e adulto, de maneira que não dispunha de ferramentas que viabilizassem o uso do mesmo por crianças. O desenvolvimento do DuinoBlok for Kids surge exatamente do desejo de expandir o alcance do referido projeto a crianças do Ensino Fundamental I, tendo o mesmo sido inspirado no Ambiente DuinoBlocks [Alves 2013], implementado pela mesma equipe.

4.1. O Ambiente

O DB4K foi projetado de modo a permitir a sua integração com o LabVad, tanto no que se refere às tecnologias utilizadas quanto aos experimentos a serem manipulados por meio do ambiente.

No entanto, além da utilização do DB4K para controle, via internet, dos experimentos disponibilizados pelo LabVad Robótica, buscou-se possibilitar a utilização do ambiente também em uma versão “client-side”, ou seja, que fosse capaz de ser executada no computador do usuário, sem a necessidade de conexão com a internet, e capaz de controlar, diretamente em uma placa Arduino conectada ao computador, outros experimentos além daqueles disponibilizados pelo LabVad.

A interface do DB4K presente no ambiente LabVad permite ao usuário, via programação, a manipulação dos seguintes dispositivos: Um conjunto de LEDs, um LED RGB, um Display de 7 Segmentos, um Display LCD um Servo Motor e um Motor DC.

Já o Ambiente DB4K “client-side”, permite também a utilização de um conjunto de sensores e a manipulação de um carrinho robô e de um braço robótico (Figura 2), o que vem a possibilitar o desenvolvimento de uma gama bastante variada de atividades relacionadas ao ensino de conceitos básicos de programação.

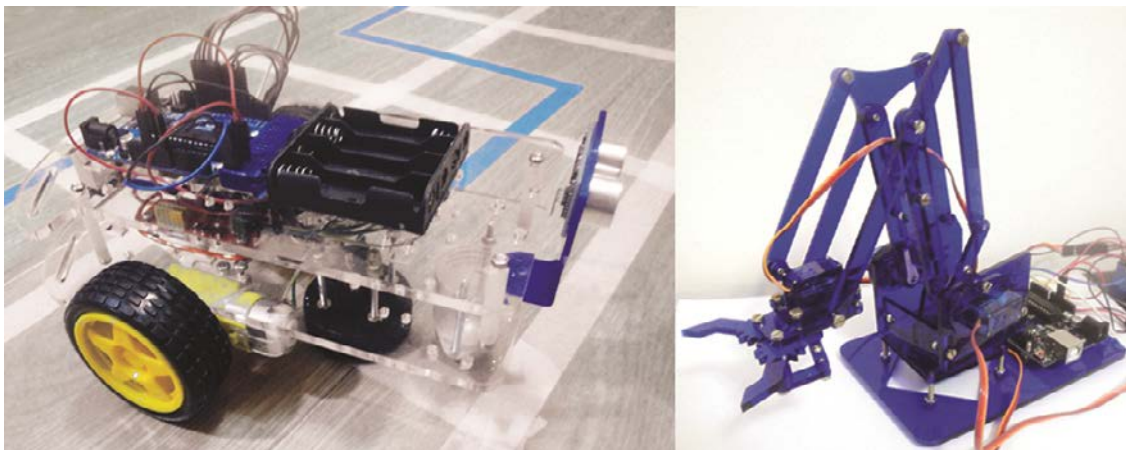


Figura 2. Carrinho e Braço Robóticos programáveis via DB4K client-side

Os blocos de programação presentes no ambiente foram desenhados de modo a apresentarem uma semântica diretamente relacionada com os dispositivos sendo manipulados e com os efeitos por eles causados sobre estes dispositivos.

Por exemplo, para acender um LED, desenhou-se o bloco “acender LED”; para escrever no display de LCD, desenhou-se o bloco “escrever ‘texto’ na linha ‘1’ do LCD”. Alguns desses blocos podem ser observados na Figura 3.

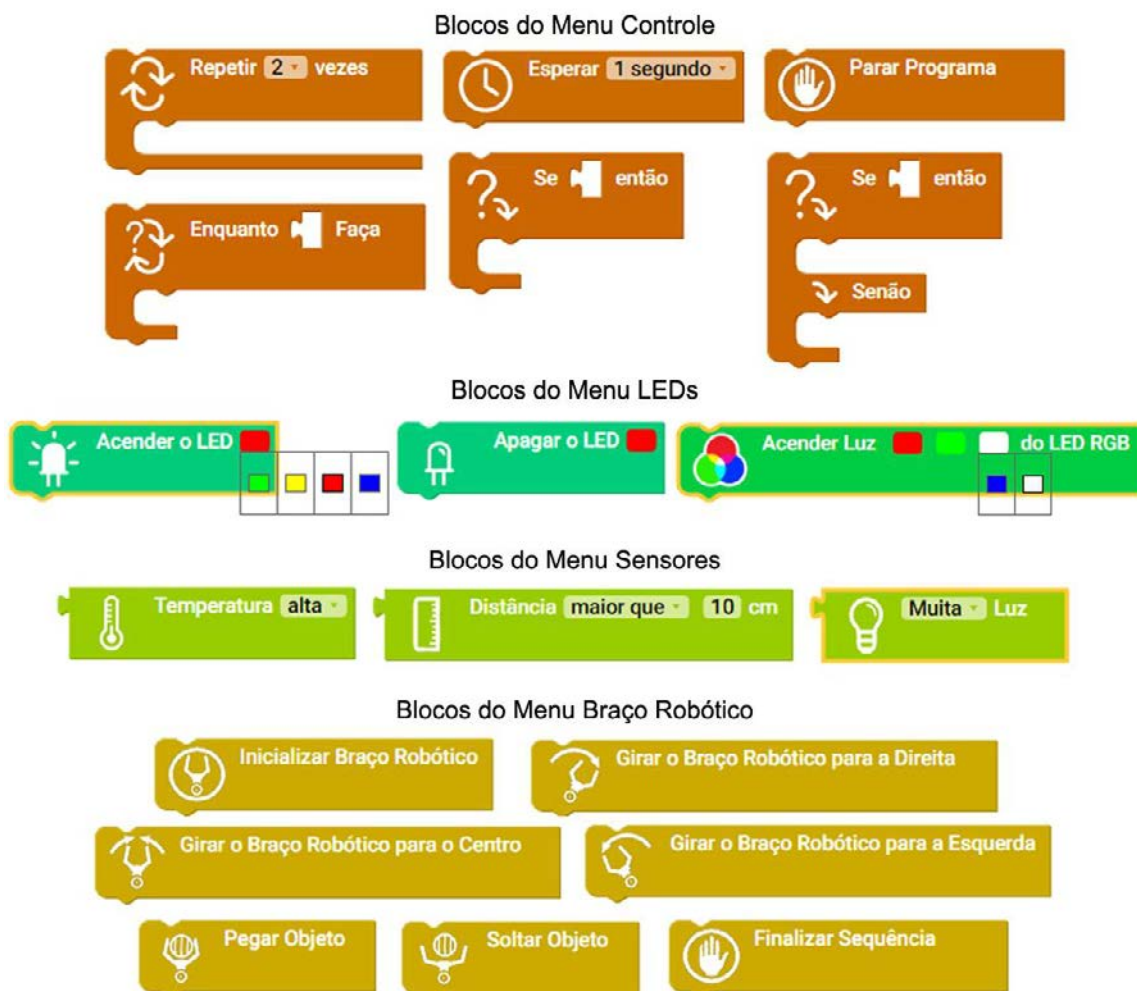


Figura 3. Amostra de blocos presentes no ambiente DB4K

Além dos blocos diretamente relacionados com os dispositivos a serem manipulados, foram desenhados também blocos para as estruturas de controle utilizadas em programação, como *repetição* e *decisão*.

A exemplo do que ocorre com o ScratchJr (<http://www.scratchjr.org/>) e com o Lego Mindstorms (<http://mindstorms.lego.com/>), optou-se por utilizar no DB4K, tanto uma representação textual quanto icônica para descrever as "funções" dos blocos. A ideia é a de facilitar, por parte dos usuários, o entendimento do fluxo de informação [Jr. & Guedes 2015].

Em relação às cores utilizadas no ambiente, decidiu-se por adotar cores vivas para os blocos e seus respectivos menus, direcionando a atenção do usuário exatamente para esses que são os elementos fundamentais do DB4K. Para o plano de fundo e barras de menus utilizou-se cores suaves de modo a não tornar visualmente cansativa a permanência possivelmente prolongada no ambiente.

Além da área de programação em blocos, para que o usuário possa, caso deseje, ter acesso à linguagem textual Wiring¹ associada a cada bloco do DuinoBlocks for Kids, decidiu-se por disponibilizar, na interface do DB4K, uma área para a exibição do programa textual correspondente ao programa criado por meio de seus blocos.

O DuinoBlocks for Kids (Figura 4) foi desenvolvido a partir do Ardublockly (<https://github.com/carlosperate/ardublockly>), um ambiente de programação em blocos para placas de prototipagem eletrônica Arduino direcionada ao público jovem e adulto e desenvolvido a partir da Biblioteca Blockly (<https://developers.google.com/blockly/>), uma biblioteca de código aberto criada pela Google Developers (<https://developers.google.com/>) e voltada para a construção de Ambientes de Programação Visual.

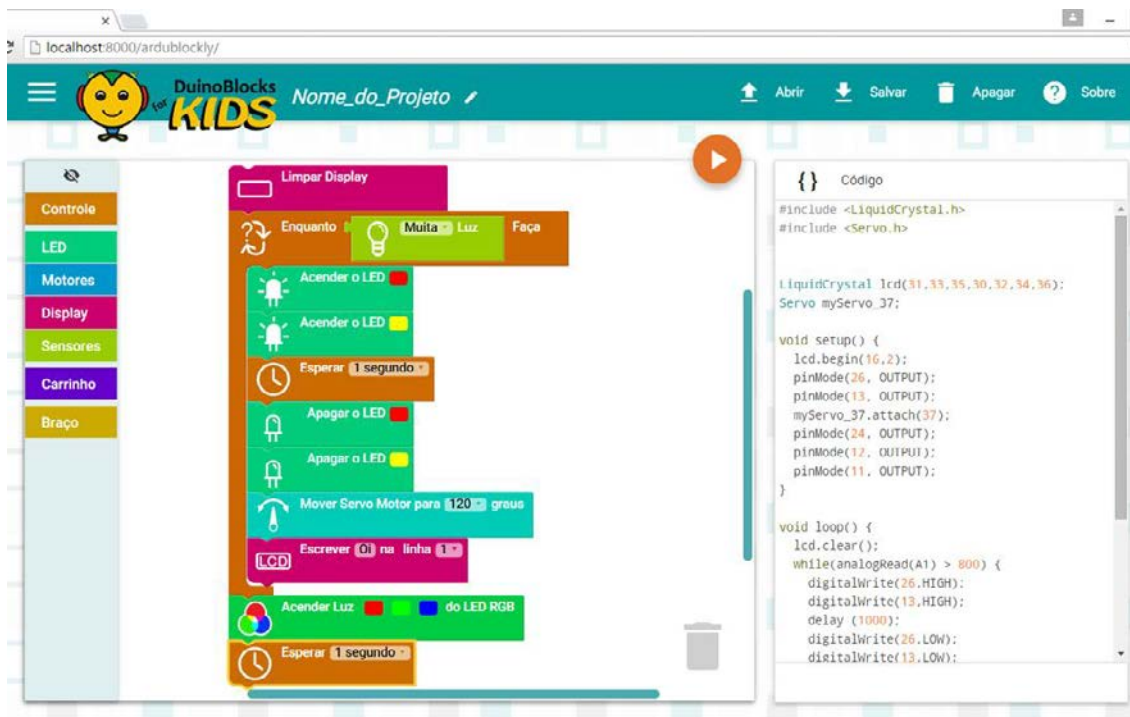


Figura 4. Visão Geral da Interface do DB4k client-side

O desenho da interface do DB4K aproveitou muitos dos elementos da interface do Ardublockly, já os blocos utilizados na programação foram totalmente redesenhados e programados. Alguns desses blocos estão presentes na Figura 3.

Desde o início do seu desenvolvimento, o ambiente vem sendo testado e avaliado pela equipe de especialistas do GINAPE/UFRJ, estando sua versão online, disponível para o público desde outubro de 2015, sendo atualmente utilizada em um curso de Robótica Educacional no Colégio Estadual José Leite Lopes, no município do Rio de Janeiro.

¹ Linguagem nativa para a programação das placas de prototipagem eletrônica Arduino, derivada da linguagem C/C++.

5. Considerações Finais

Os resultados aqui apresentados nos permitem concluir ser perfeitamente viável - lançando-se mão de bibliotecas e aplicações de código aberto disponíveis na Web, e da associação dessas a hardware de baixo custo - desenvolver ferramentas que facilitem aos docentes e discentes o uso das TICs para o aprendizado de programação por meio da Robótica.

Embora o DuinoBlocks for Kids tenha sido projetado para o ensino de conceitos básicos de programação a crianças do Ensino Fundamental I, acredita-se que o ambiente possa ser utilizado também para auxiliar na introdução ao aprendizado de programação por meio da Robótica Educacional a qualquer pessoa com pouco ou nenhum conhecimento nessas áreas.

Ressalta-se o fato de que, por se tratar de um ambiente construído com base em tecnologia livre, o DB4K pode servir como base para o desenvolvimento de outros ambientes adequados a determinadas necessidades específicas, assim como este utilizou o Ardublockly para o seu desenvolvimento, que por sua vez foi influenciado pelo ambiente BlocklyDuino (<https://github.com/BlocklyDuino/BlocklyDuino/>), tendo em todos eles sido utilizada a biblioteca Blockly.

5.1. Trabalhos Futuros

Para dar suporte ao uso adequado do DuinoBlocks for Kids e para que sejam feitas possíveis melhorias no ambiente de modo a torná-lo tão eficaz e eficiente quanto possível, é necessária ainda a realização de um conjunto de ações, entre elas:

- Criação de um site onde estarão presentes: uma descrição completa do ambiente, um "help" para o mesmo, tutoriais para a montagem dos experimentos controlados pelo DB4K, sugestões de atividades, a versão "client-side" do ambiente para download e um link para o LabVad.
- A realização de experimentos que busquem apontar os reais benefícios da utilização do DuinoBlocks for Kids no ensino de programação por meio da Robótica Educacional nas séries iniciais do Ensino Fundamental, bem como o aprimoramento do ambiente a partir dos resultados obtidos através desses experimentos.

Agradecimentos

Este trabalho foi desenvolvido com o financiamento parcial da RNP (Rede Nacional de Ensino e Pesquisa) através do Edital para os Grupos de Trabalho (GTs) da referida organização para o período 2014-2015.

Referências

- Alimisis, Dimitris. (2013) "Educational roboTIC: Open questions and new challenges". Themes in Science and Technology Education, v. 6, n. 1, p. pp-63.
- Alves, Raphael M. (2013) "Duinoblocks: Desenho e Implementação de um Ambiente de Programação Visual para Robótica Educacional". Dissertação (Mestrado em Informática) - Programa de Pós Graduação em Informática, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

- Barbero, A.; Demo, G. B. (2011) “The art of programming in a Technical Institute after the Italian secondary school reform.” Proceedings ISSEP 2011.
- Buechley, L.; Eisenberg M.; Elumeze, N. (2007) “Towards a curriculum for electronic textiles in the high school classroom”. In: ACM SIGCSE Bulletin-2007, p. 28-32.
- Burnett, Margaret M. (1999) “Visual programming”. Wiley Encyclopedia of Electrical and Electronics Engineering.
- Da Silveira, Sérgio A. (2003) “Inclusão digital, software livre e globalização contra-hegemônica. Software Livre e Inclusão Digital”. Sao Paulo: Conrad, v. 7, p. 11.
- Denis, Brigitte; Hubert, Sylviane. (2001) “Collaborative learning in an educational roboTIC environment”. Computers in Human Behavior, v. 17, n. 5, p. 465–480.
- Jr., Almir de O.C.; Guedes, Ello´a B. (2015) " Uma Análise Comparativa de Kits para a Robótica Educacional". WEI - XXIII Workshop sobre Educação em Informática.
- López-Rodríguez, Francisco M.; Cuesta, Federico. (2015) “Andruino-A1: Low-Cost Educational Mobile Robot Based on Android and Arduino”. Journal of Intelligent & Robotic Systems, 3 maio.
- Papert, Seymour. (1980) “Mindstorms: Computers, Children and Powerful Ideas”. NY: Basic Books. 255p.
- Perez, Anderson L. F.et al. (2013) “Uso da Plataforma Arduino para o Ensino e o Aprendizado de Robótica”.[S.l: s.n.].
- Piaget, Jean. (1974)“To Understand is to Invent”. N.Y.: Basic Books.
- Pinto, Marcos de C. (2011) “Aplicação de Arquitetura Pedagógica em Curso de Robótica Educacional com Hardware Livre”. Dissertação (Mestrado em Informática) - Programa de Pós Graduação em Informática, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Prensky, M. (2001) “Digital Natives, Digital Immigrants Part 1”, On the Horizon, Vol. 9 Iss: 5, p. 1–6.
- Resnick, Mitchel et al. (2009) “Scratch: programming for all”. Communications of the ACM, v. 52, n. 11, p. 60-67.
- Resnick, Mithel. (2014) “Palestra no evento Transformar 2014”. 46'30". Disponível em: <<https://www.youtube.com/watch?v=hRGJUc2opQ4>>. Acesso em: 4 mar.2016.
- Rus, Daniela. (2006) “Teaching roboTIC everywhere”. IEEE RoboTIC & Automation Magazine, v. 13, n. 1, p. 15–94.
- Souza, Marcos A.M; Duarte, José R.R. (2015) “Low-cost educational roboTIC applied to physics teaching in Brazil”. Physics Education, v. 50, n. 4, p. 482.
- Souza, Paulo Roberto A. et al. (2014) “LabVad: Laboratório Remoto para o Desenvolvimento de Atividades Didáticas com Robótica”. TISE : XX Congresso Internacional de Informática Educativa.
- Wing, Jeannette,M. (2006) "Computational thinking". Communications of the ACM, v. 49, n. 3, p. 33-35.

EMPREGABILIDADE DOS EGRESSOS TECNÓLOGOS DO CURSO DE ANÁLISE E DESENVOLVIMENTO DE SISTEMAS NOS INSTITUTOS FEDERAIS MINEIROS

Giuliano Viana de Alkmim¹, Rosemary Dore Heijmans²

¹Instituto Federal do Norte de Minas Gerais - *Campus* Januária Fazenda São Geraldo, s/n, Km 6 - CEP 39480-000 - Januária / MG

²Universidade Federal de Minas Gerais - Faculdade de Educação, Depto. de Ciências Aplicadas à Educação - Sala 1606, Av. Antonio Carlos 6627 – BH-MG CEP 31270-901

giuviana@gmail.com, rosemarydore@gmail.com

Abstract: *This research aimed to analyze the process of inserting and maintaining of the graduates in the Higher Technology Course in Analysis and System Development (CST in ADS) in the labor market. This research was conducted with the technologists graduates of that course in the Federal Institutes (IFs) of Minas Gerais where there were graduates. It was performed through the electronic questionnaire and semi-structured interviews. The results showed that the technologist formed in CST in ADS, in (IFs) miners, who were part of the study, get a large insertion/maintenance in the labor market, as it was proved that the employment rate of this technologists is high.*

Resumo: *Esta pesquisa teve como objetivo analisar o processo de inserção e manutenção do egresso do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas (CST em ADS) no mundo do trabalho. Esta investigação foi realizada com os egressos Tecnólogos do referido curso nos Institutos Federais (IFs) Mineiros onde existiam alunos formados. A pesquisa foi realizada com aplicação de questionário eletrônico e de entrevista semiestruturada. Os resultados obtidos evidenciaram que o tecnólogo formado no CST em ADS, nos (IFs) mineiros que fizeram parte do estudo, consegue grande inserção/manutenção no mundo do trabalho, pois, ficou provado que, o índice de empregabilidade desse tecnólogo é alto.*

1. INTRODUÇÃO

Atualmente, a formação de tecnólogo é muito enfatizada no Brasil, como meio para preparar os jovens para uma inserção mais rápida no mercado de trabalho. No entanto, não existem pesquisas, de tecnólogos, nos Institutos Federais, na área de informática, voltadas para identificar como os egressos desses cursos percebem suas experiências e quais são as tensões vivenciadas na sua prática profissional, indicando como se concretizaram suas expectativas quanto ao relacionamento da educação profissional com o mercado de trabalho, assim como, os principais motivos de escolha dessa modalidade de ensino nos IFs mineiros, o local de sua atuação e os principais fatores que fazem diferença para que o esse tecnólogo consiga os maiores salários.

O CST em ADS, segundo o SISTEC, 2013, é o curso que representa a maior demanda de matrícula entre os cursos superiores de tecnologia nos IFs brasileiros. A diferença na quantidade de matrículas entre o CST em ADS e o segundo curso da modalidade, o Curso de Gestão Ambiental, é de praticamente 100%. Apesar do grande

interesse dos IFs no CST em ADS, não existe um estudo e praticamente nenhum acompanhamento sobre o seus egressos.

O artigo busca responder se os CST são planejados conforme as demandas: do mundo do trabalho; das particularidades da região onde a instituição de ensino, que oferta tais cursos, está inserida. Visa, ainda, saber se o seus egressos estão conseguindo um boa empregabilidade. Em linhas gerais, espera-se que este estudo possa oferecer melhor compreensão sobre os CST e o perfil do profissional por eles formado.

2. A EMPREGABILIDADE

Existem diferentes concepções sobre o termo empregabilidade, porém, há, entretanto, um elemento comum: a educação. Esta é destacada como fator fundamental para a aquisição de conhecimentos de que o indivíduo necessita para se manter ativo no mundo do trabalho.

A relação mais forte entre educação, trabalho e indivíduo contribuiu para direcionar as políticas públicas da educação nacional para uma proposta de formação influenciada pela noção de empregabilidade, bem como para a retomada do debate sobre a Teoria do Capital Humano; qual teve como principal representante Theodore Shultz.

O principal aspecto da teoria de Schultz (1973), um dos maiores expoentes da Teoria do Capital Humano, é o de que o indivíduo que investe em si mesmo (em capital humano) consegue ampliar as suas escolhas por um posto de trabalho. Entretanto, a teoria do capital humano, segundo o mesmo autor, valoriza o processo de aprendizagem e treinamento que também acontece fora do ambiente escolar, através da experiência de trabalho e da migração (quantidade de serviços).

Além do Capital Humano, existem outros fatores que são tomados como paradigmas para analisar a empregabilidade do indivíduo: o Capital Cultural e o Capital Social [HELAL, 2005]. Este formulado por Pierre Bourdieu e por James Coleman; aquele, somente por Bourdieu.

O Capital Cultural, caracteriza, por exemplo, a interferência do ambiente familiar em seus gostos e até mesmo em seu processo de socialização, bem como em sua trajetória de aprendizagem escolar [BOURDIEU, 1979]. ([Helal 2005], p. 7) afirma que "indivíduos socializados em lares cujos pais têm níveis educacionais/culturais mais elevados e posições ocupacionais de maior destaque levam vantagem no mercado de trabalho em relação àqueles cujos pais possuem menor nível de escolaridade e posição ocupacional inferior".

Diante das concepções que envolvem o Capital Social, Helal (2005) o conceitua pelos seus principais representantes (Bourdieu e Coleman). Utilizando um elemento comum associado por eles a esse tipo de capital: a rede de relacionamentos dos indivíduos. Essa rede é muito utilizada para mostrar a interferência do Capital Social na empregabilidade do indivíduo, uma vez que os seus relacionamentos e sua "participação em grupos e organizações" o mantêm atualizado sobre informações e conhecimentos relativos ao mundo do trabalho, afetando "positivamente a probabilidade desse indivíduo estar empregado" ([HELAL, 2005], p. 9).

A probabilidade de um indivíduo estar empregado contempla elementos, que vão além de credenciais escolares, que estão associadas à posse do diploma. Assim, somam-se à escolaridade do indivíduo e suas origens familiares, envolvendo nível

educacional, cultural e econômico, além da rede de relacionamentos e experiências profissionais. O conjunto dessas variáveis pode determinar a situação de emprego ou desemprego do indivíduo.

3. METODOLOGIA

A natureza da presente pesquisa caracterizou-se como mista (quantitativa e qualitativa), definida por possuir resultados mais abrangentes, além de poder neutralizar vieses dos métodos separados e poder gerar a triangulação dos dados por meio da convergência dos métodos qualitativos e quantitativos [CRESWELL, 2007].

A principal estratégia de investigação utilizada no presente estudo foi a pesquisa de *survey* (levantamento), que através de uma amostra significativa de uma população pode-se ter estimativas sobre a população total [BABBIE, 1999].

Os participantes do estudo foram os egressos do CST em ADS dos Institutos Federais no Estado de Minas Gerais, nos anos de 2006 até 2013. O universo da pesquisa configurou-se num total de 311 egressos. Destes, 122 são do IFNMG – *Campus* Januária; outros 99 do IFTM – *Campus* Uberaba e 90 do IFMG – *Campus* Bambuí. Vale ressaltar que a justificativa para a escolha dos referidos *campi* foi devido ao fato de serem os únicos em Minas Gerais que tinham egressos de CST em ADS no período estudado e por pertencerem a Institutos Federais distintos.

Na pesquisa, o intervalo de coleta de dados ocorreu de outubro a dezembro de 2014. Ou seja, como os egressos são de 2006 até o final do ano de 2013, esses ex-alunos possuem no mínimo um ano de formação.

Para o cálculo da amostra, foi utilizado o método para estimação de proporções para populações finitas, conforme [Bolfarine and Bussab 2005].

Como o CST em ADS nos Institutos Federais Mineiros, que possui egressos, encontra-se em três *campi* diferentes, com um quantitativo de egresso diferente em cada lugar, foi realizada, em seguida, uma alocação proporcional da amostra por *campus*. Ou seja, quanto maior o quantitativo de egressos no *campus* maior o percentual desse *campus* na amostra.

A amostra utilizada foi de 193 egressos, respeitando a proporção de cada *campus*, definido a partir da margem de erro 5% e o nível de significância de 5% .

Vale ressaltar que o método de seleção dos participantes para responderem ao questionário foi o aleatório, todos tiveram a mesma probabilidade de responder ao questionário [BABBIE, 1999]. Já a quantidade de entrevistas semiestruturadas para a pesquisa qualitativa foi de nove egressos (três em cada instituição) escolhidos por conveniência. Foram eles: os ex-alunos que trabalhavam como empregados com carteira assinada; funcionários públicos e autônomos/proprietários de empresas.

4. ANÁLISE DOS DADOS E RESULTADO

Analisando o índice de empregabilidade dos tecnólogos do CST em ADS de todos os Institutos Federais de Minas Gerais que possui egresso do referido curso, por meio do gráfico 1, observa-se que, no *campus* de Bambuí, 89,1% dos egressos do referido curso encontram-se trabalhando ou trabalhando e estudando, no *campus* de Januária, esse percentual sobe para 94,8%, já no *campus* de Uberaba, encontra-se o maior percentual,

98,3% de analistas de sistemas exercendo suas atividades profissionais dentro do mercado de trabalho.

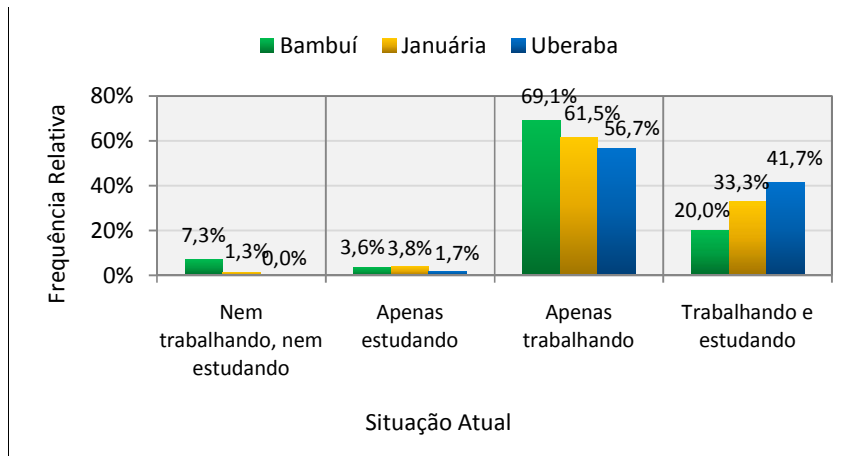


Gráfico 1: Distribuição da situação atual dos entrevistados por campus

Fonte: dados da pesquisa.

Ressalte-se, ainda, que, de forma geral, o índice de ocupação dos tecnólogos também foi satisfatório, dos 193 egressos participantes da pesquisa, 182 deles estavam trabalhando, ou seja, 94,3% da amostra total.

Procurou-se evidenciar, também, neste estudo, por meio do gráfico 2 a seguir, se as atribuições direcionadas ao tecnólogo, em seu local de trabalho, estão relacionadas à sua formação.

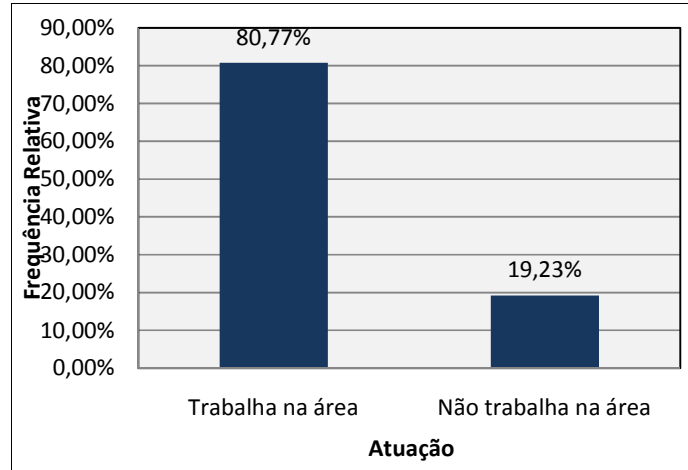


Gráfico 2: Proporção de egressos que afirmaram trabalhar ou não na área de informática

Fonte: dados da pesquisa.

Os dados quantitativos encontrados permitem dizer que, do conjunto total da amostra de ex-alunos dos três institutos federais estudados, quase 81% exercem suas atividades profissionais em área relacionada com a sua formação, ou seja, informática. Embora muitos deles, realizem, em seu local de trabalho, outras funções paralelas às do tecnólogo, como as pertinentes ao técnico de informática.

Essa realidade torna-se perceptível no relato dos egressos, durante a entrevista qualitativa.

[...] minhas atribuições estão relacionadas com a área de informática, só que o meu cargo lá é de técnico em informática, [...] eles me colocaram para trabalhar com a parte de redes. Então no curso de Análises teve a disciplina de redes, mas assim, no geral, o meu trabalho mesmo é de técnico. (Entrevistado U1).

No meu cargo especificamente, acaba que eu exerço outras funções, além das próprias do tecnólogo. [...] eu tenho que mexer com contas, por exemplo, coisas que não são específicas do curso, mas que a própria gestão do setor exige. (Entrevistado J1).

Eu mesmo tenho o CST, mas minha carteira é assinada como técnico, [...]. Mas a maioria das coisas que a gente aprende no curso de tecnólogo dá pra colocar em prática. (Entrevistado J2).

Outra informação interessante, obtida na pesquisa, é que a maioria dos egressos não migrou para outros estados após a conclusão da graduação tecnológica, ou seja, 88,71% exercem suas atividades profissionais em Minas Gerais, o estado de origem de seus Institutos Federais.

A tabela 1 especifica, dentro do estado de Minas Gerais, as mesorregiões onde os egressos realizam seus trabalhos.

Tabela 1: Locais, por Mesorregião do estado de Minas Gerais, onde o egresso realiza o trabalho principal

Mesorregião de Minas Gerais	IFMG - <i>Campus</i> Bambuí	IFNMG <i>Campus</i> Januária	IFTM - <i>Campus</i> Uberaba	Total
Noroeste de Minas	0,00%	5,33%	0,00%	2,15%
Norte de Minas	0,00%	69,33%	0,00%	27,96%
Vale do Mucuri	0,00%	0,00%	0,00%	0,00%
Jequitinhonha	0,00%	4,00%	0,00%	1,61%
Triângulo Mineiro / Alto Paranaíba	15,39%	2,67%	89,83%	33,87%
Central Mineira	7,69%	0,00%	0,00%	2,15%
Metropolitana de Belo Horizonte	9,62%	8,00%	0,00%	5,92%
Vale do Rio Doce	0,00%	0,00%	0,00%	0,00%
Oeste de Minas	49,99%	0,00%	0,00%	13,98%
Sul/Sudeste de Minas	0,00%	0,00%	0,00%	0,00%
Campos das Vertentes	3,84%	0,00%	0,00%	1,08%
Zona da Mata	0,00%	0,00%	0,00%	0,00%
TOTAL	86,54%	89,33%	89,83%	88,71%

Fonte: dados da pesquisa. Nota: grifos nossos.

Assim, fica evidente que parte significativa dos tecnólogos está trabalhando na mesorregião onde fez o curso. Praticamente 50,0% dos tecnólogos egressos do *Campus* de Bambuí ocupam postos de trabalho na região Oeste de Minas, onde se localiza o referido *campus*. Em Januária, mais de 69,0% atuam também em torno da origem do instituto, ou seja, no Norte de Minas. Ressalta-se que a Mesorregião do Triângulo Mineiro/Alto Paranaíba foi a que mais reteve seus graduados; todos os egressos do *Campus* de Uberaba que permaneceram em Minas Gerais exercem suas atividades profissionais nessa localidade.

Verifica-se, conforme dados da pesquisa, que a média salarial entre os três institutos é semelhante. Há discrepância, todavia, quando se compara a renda mensal dos egressos com a média de remuneração da mesorregião onde eles estão inseridos. Conforme mostra a tabela 2, percebe-se que, apesar de o *campus* de Uberaba ter se destacado pelo seu baixo índice de desemprego, apenas 1,7%, por ser a mesorregião que mais emprega os seus egressos (quase 90%) e a que apresenta a maior média salarial para o nível superior de ensino (R\$ 3.696,48), em comparação à de outros institutos incluídos na pesquisa, representou o menor percentual, 55%, de egressos que recebem dentro da média salarial da mesorregião ou acima dessa. Em seguida, aparece Bambuí com 58,1% e Januária com 60,3% de egressos, recebendo dentro ou acima da média de sua mesorregião.

Tabela 2: Mesorregiões de Minas Gerais - Remuneração média, de dezembro de 2013, segundo grau de instrução superior completo e gênero - 2013

Mesorregião	Ensino Médio Completo			Superior Completo		
	Masculino	Feminino	Total	Masculino	Feminino	Total
Norte de Minas	1.255,81	982,22	1.131,16	3.667,16	2.376,03	2.810,49
Triângulo						
Mineiro/Alto	1.789,90	1.200,65	1.521,00	4.882,04	2.936,64	3.696,48
Paranaíba						
Oeste de Minas	1.488,48	1.078,35	1.282,73	3.658,26	2.315,01	2.775,72
Total - MG	1.882,73	1.220,56	1.579,31	5.900,47	3.423,69	4.364,94

Fonte: RAIS - Dec. 76.900/75.

Nota: elaborado por CGET/DES/SPPE/MTE

Considerando o fato de que parte significativa dos tecnólogos desta pesquisa estava trabalhando, principalmente em sua área de formação, conforme mostrado anteriormente, e que atuam na região de abrangência dos seus respectivos institutos federais, afirma-se que o CST em ADS, dos *Campi* que fazem parte desta investigação, foram implantados seguindo um critério de planejamento, organização e funcionamento voltado para o "atendimento às demandas do cidadão, do mercado de trabalho e da sociedade", bem como às demandas peculiares de suas regiões, de acordo com o que foi instituído pelas Diretrizes Nacionais Gerais para a organização e funcionamento dos Cursos Superiores de Tecnologia por meio da Resolução CNE/CP nº 3, de 2002 e por [Bastos 1991].

Como estudos indicam a influência do capital econômico na empregabilidade, como relatado anteriormente, pesquisamos se o referido capital poderia ter influência na escolha do CST em ADS nos IFs mineiros, através de um levantamento dos principais motivos de escolha.

A gratuidade foi, entre os fatores de justificativa de escolha, o que mais se destacou na entrevista qualitativa.

Na época que eu fiz o curso, a primeira opção foi porque era um curso público, não era particular, eu não tinha condições financeiras de bancar um curso. (Entrevistado B1).

Como eu estava apertado de dinheiro [...], fui obrigado a escolher uma coisa mais perto que no caso era aqui no IFMG e entre os cursos que tinham no instituto, na época, esse [Análise e Desenvolvimento de Sistemas] era o que tinha mais sintonia com o meu perfil. (Entrevistado B3).

O motivo que mais influenciou a minha entrada tanto no curso técnico como no superior foi a oferta gratuita dos cursos; segundo por causa da minha dificuldade de sair da cidade. (Entrevistado J2).

O principal explicador de demanda para a escolha do CST, também, na pesquisa quantitativa, se relaciona-se com as condições econômicas dos egressos, na época do ingresso no ensino superior. Como apresentado no gráfico 3, que nos mostra a ordem de classificação desses fatores

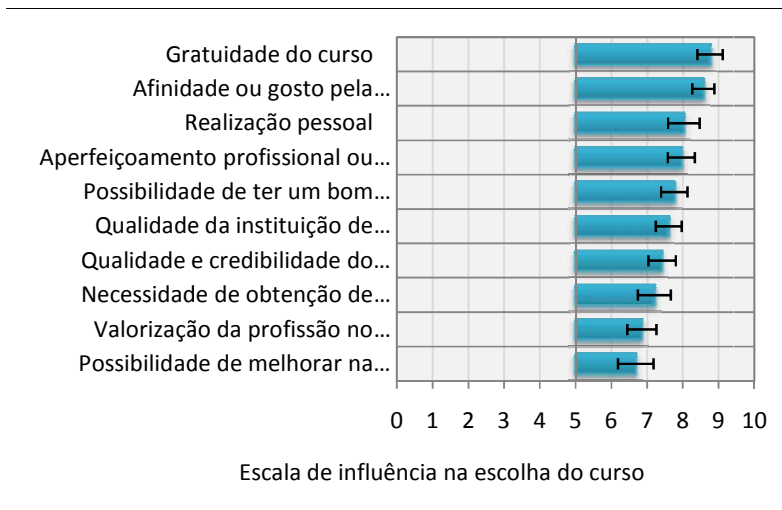


Gráfico 3: Média da escala de influência para os dez itens mais importantes para a escolha do CST

Fonte: dados da pesquisa.

No que concerne à afinidade ou gosto pela área/profissão (segundo principal motivo de escolha), conforme apresenta o gráfico 4, observa-se que cursar o técnico em informática no nível médio de ensino pode ser considerado um dos fatores responsáveis por ter despertado nos egressos o interesse pelo CST em ADS, principalmente, por meio da descoberta da afinidade pela área do curso.

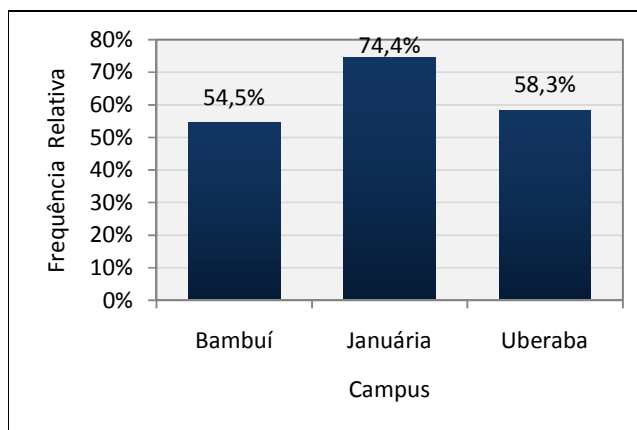


Gráfico 4: Proporção de egressos que afirmaram ter realizado curso técnico de nível médio na área de informática

Fonte: dados da pesquisa.

Nota-se, também, na entrevista qualitativa, que os egressos reafirmam que essa afinidade pela área surgiu com a conclusão do curso técnico como mostrado nos depoimentos a seguir:

Eu escolhi o curso porque, primeiramente, eu iniciei aqui fazendo o curso técnico de informática,[...] gostei da área e decidi fazer o curso superior aqui mesmo. (Entrevistado B2).

[...] já tinha feito curso técnico no IF do Triângulo Mineiro – *Campus* Uberaba. Gostei do curso técnico e quis dar prosseguimento com o curso superior. (Entrevistado U1).

O que influenciou na verdade foi eu ter feito o curso técnico em informática. Ainda durante o curso, eu comecei a trabalhar com informática, comecei a gostar da área, então, queria dar prosseguimento na carreira com o curso superior. (Entrevistado J1).

Esses achados corroboram com os do estudo de Andrade (2009), que deixa subtendida a tendência dos técnicos de buscar uma diplomação que se relacione com a área na qual já tenham desenvolvido alguma competência.

Para analisar a influência do capital humano, cultural e social nas maiores remunerações dos egressos, selecionaram-se todos os egressos que possuíam algum trabalho remunerado na época do levantamento. Em seguida, utilizou-se uma regressão logística e o método de Stepwise [AGRESTI, 2002]. Visto que, para o procedimento de Foward (critério de entrada das variáveis na análise de regressão), foram selecionadas as variáveis que poderiam ter alguma influência nessa renda. Assim, utilizou-se as variáveis referentes ao: a) perfil do egresso envolvendo: gênero, localização do trabalho, cor/raça, estado civil, possuir filho, tipo de ensino fundamental e médio, idade na época em que concluiu o curso; b) capital cultural: foram utilizadas as variáveis relativas ao *background* familiar como: escolaridade do pai e da mãe e a renda média familiar; c) capital social referindo-se às vias utilizadas para conseguir emprego (como conseguiu o emprego); d) capital humano, sondagem se o egresso possuía ou não o curso técnico, tempo de formado, tempo de experiência, se continuou ou não os estudos.

Após ajustar a regressão logística múltipla e aplicar o método Backward, verificou-se que o gênero, a escolaridade do pai, a renda média familiar, a idade, o tempo de formado e o tempo de experiência exerceram influência significativa no que se refere ao egresso ter um rendimento maior que 6 salários mínimos. Dessa forma, ao analisar cada uma das variáveis citadas anteriormente, em separado e, controlar as demais, conclui-se que existe uma preponderância do egresso do gênero masculino receber os maiores salários. Assim, as chances de um homem receber acima de 6 salários é superior a das mulheres 33,09 vezes. Esse dado vai ao encontro ao apresentado por Gentili (2002), em que o autor destaca outros fatores associados à empregabilidade, que podem representar um diferencial nos processos seletivos fornecedores de emprego e renda, ou seja, “ser branco, ser negro, ser imigrante, ser gordo, ser surdo, ser nordestino...” ([GENTILI, 2002], p. 55). O caso particular ‘ser mulher’, como demonstra os dados da regressão, ainda implica ter menos chances de receber os maiores salários. Vale ressaltar, que os maiores salários não necessariamente são para aqueles que atuam na informática (apesar da maioria atuarem nessa área).

Do mesmo modo, os tecnólogos egressos cujos pais concluíram o ensino médio apresentam quatorze vezes mais chances de receber acima de 6 salários em relação aos tecnólogos cujos pais não concluíram o ensino médio. Essa constatação mostra-nos a importância do capital cultural na empregabilidade.

Um achado interessante relaciona-se ao capital cultural, no que se refere à renda familiar. Este estudo mostrou que os egressos provenientes de famílias com renda mensal superior a 6 salários mínimos têm mais chances, cerca de 126 vezes, de receberem uma renda igual ou superior a essa em relação àqueles cujos pais têm renda menor que a especificada, ou seja, quanto maior o nível salarial de sua família de origem, maiores são as chances do egresso adquirir melhor remuneração no mercado de trabalho.

Outro fator de importante destaque na investigação é que o capital humano, referente ao investimento nos estudos, não causou influência significativa quanto ao nível salarial. Vale ressaltar que os egressos do curso estudado, mesmo associando outros títulos à graduação, como uma pós-graduação, não influenciou para a obtenção de um rendimento mensal superior a 6 salários mínimos.

As variáveis tempo de formado, tempo de experiência e migração (quantidade de empregos), consideradas derivações do capital humano, exerceram influência positiva para o aumento de chances dos egressos receberem acima de 6 salários: em 2,88 vezes para cada ano que se aumenta no tempo em que este adquiriu a titulação; 1,43 vezes para cada ano a mais de experiência; 1,92 vezes para cada emprego a mais que eles têm.

O capital humano continua a influenciar para as maiores remunerações, no caso estudado de 6 salários mínimos, porém, essa influência para os que já possuem o curso superior, como na presente pesquisa, não aconteceu com o investimento na continuação dos estudos e sim nos outros fatores relacionados ao capital humano, como demonstrado: tempo de formado, tempo de experiência e quantidade de empregos.

As variáveis cor/raça e o fato de ser branco não mostraram diferença no estudo, nem no ingresso no mercado, nem nas maiores remunerações. Em contrapartida, conforme demonstrado, os tipos capitais, principalmente o capital humano e cultural, fazem toda diferença na obtenção dos maiores salários.

5. CONSIDERAÇÕES FINAIS

Realizadas a investigação e a comparação da situação profissional dos egressos do CST em ADS, afirma-se que o índice de empregabilidade desses egressos, nos três institutos federais mineiros que participaram da pesquisa, é alto. Principalmente quando os resultados quantitativos revelaram que 94,3% da amostra total estavam trabalhando no momento da aplicação do questionário.

A maioria dos egressos está atuando no mercado de trabalho das mesorregiões que envolvem sua origem institucional. Tal situação deixa subentender que houve, por parte das instituições, uma aproximação com o setor de produção, no sentido de conhecer os arranjos produtivos locais, antes de implantarem o curso analisado, pois pelo que já foi apresentado, no decorrer da pesquisa, as demandas do mercado de trabalho regional, em particular do aluno e da sociedade, foram atendidas com a oferta do CST em questão.

Outra constatação é a de expressivo quantitativo de participantes deste estudo, que associou a escolha do CST em ADS às suas condições financeiras, reafirmando que a Educação Profissional e Tecnológica se configurou na principal via de expansão das possibilidades de acesso à educação superior, para os filhos de trabalhadores desprovidos de nível econômico de grandes representações, que também escreveram sua

trajetória educacional em escolas públicas e, em sua maioria, não tiveram a oportunidade de ir além do nível médio de ensino.

Por fim, pode-se afirmar a grande influência que o capital humano, cultural e social tem sobre a empregabilidade. Tendo em vista, por exemplo, que para os egressos do CST em ADS nos IFs estudados, quanto maior o ensino e a renda dos pais (capital cultural), maiores são as chances dos egressos receberem os melhores salários, o mesmo acontece em relação ao capital humano, quanto maior o tempo de formado e o número de empregos, maiores são as chances do egresso receber os maiores salários.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- AGRESTI, A. *Categorical data analysis*. New York: Wiley, 2002.
- ANDRADE, Andréa de Farias Barros. *Cursos superiores de tecnologia: um estudo de sua demanda sob a ótica dos estudantes*. 2009. 152f. Dissertação (Mestrado em Educação) — Faculdade de Educação Programa de Pós-Graduação em Educação. Universidade de Brasília, Distrito Federal, 2009.
- BABBIE, E. *Métodos de pesquisa de survey*. Belo Horizonte: Editora UFMG, 1999.
- BASTOS, João Augusto de Souza Leão de Almeida. *Cursos superiores de tecnologia: avaliação e perspectivas de um modelo de educação técnico profissional*. Brasília: SENET, 1991.
- BOLFARINE, H.; BUSSAB, W. O. *Elementos de amostragem*. São Paulo: Blucher, 2005.
- BOURDIEU, P. Os três estados do capital cultural. In: NOGUEIRA, M. A.; CATANI, A. (Org.). *Escritos de educação*. Petrópolis: Vozes, 1979. p. 73-79. (3. ed. 2001).
- BRASIL. Parecer CNE/CP nº 03, de 03 de dezembro de 2002. Institui as Diretrizes Curriculares Nacionais Gerais para a organização e o funcionamento dos cursos superiores de tecnologia. *Diário Oficial da República Federativa do Brasil*, Brasília, DF.
- CRESWELL, J. W. *Projeto de pesquisa: métodos qualitativo, quantitativo e misto*. 2. ed. Porto Alegre: Artmed, 2007.
- GENTILI, P. Três teses sobre a relação trabalho e educação em tempos neoliberais. In: LOMBARDI, José Claudinei; SAVIANI, D.; SANFELICE, José Luís (Org.). *Capitalismo, trabalho e educação*. Campinas: Autores Associados; HISTEDBR, 2002. Parte II, Estudo Três. p. 45-59.
- HELAL, D. H.. Flexibilização organizacional e empregabilidade individual: proposição de um modelo explicativo. *Cad. EBAPE.BR*, Rio de Janeiro, v. 3, n. 1, mar.2005.
- RAIS, Relação Anual de Informações sociais. www.acessoinformacao.gov.br
- SCHULTZ, T. W.. *O capital humano: investimento em educação e pesquisa*. Rio de Janeiro: Zahar Editores, 1973.

Ensino de Linguagem de Programação com Ênfase na Aprendizagem Significativa

Gilberto Astolfi¹, Dejahyr Lopes Junior¹

¹Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul. (IFMS)
Campus Campo Grande
79.021-000 – Campo Grande – MS – Brasil

{gilberto.astolfi,dejahyr.lopes}@ifms.edu.br

Abstract. *This paper describes an intervention, with emphasis on teaching practice, which prioritizes an educational approach led by the bias of meaningful learning in which the student develops the process of learning based on their previous knowledge. In the approach it is proposed to identify previous knowledge of the students, analyze them, minimize the conceptual deficiencies and finally introduce the new knowledge through the development of a software project throughout the course. The results presented suggest a constant reflection and resumption of the pedagogical practices and their implications, in which teachers and students share the central role of teaching and learning process.*

Resumo. *Este artigo relata uma intervenção, com ênfase na prática docente, que prioriza uma abordagem de ensino conduzida pelo viés da aprendizagem significativa, em que o estudante desenvolve o processo de aprendizagem tendo como base seu conhecimento prévio. Na abordagem é proposto identificar os conhecimentos prévios dos estudantes, analisá-los, amenizar as deficiências conceituais e por fim introduzir o novo conhecimento por meio de desenvolvimento de um projeto de software ao longo da unidade curricular. Os resultados apresentados sugerem uma constante reflexão e retomadas das práticas pedagógicas e suas implicações, as quais professores e alunos compartilham o protagonismo do processo de ensino e aprendizagem.*

1. Introdução

Aprender a partir daquilo que já sabemos de forma que o novo conhecimento tenha significado e se relacione com o conhecimento existente é a essência da teoria da Aprendizagem Significativa proposta por Ausebel (2003). Ensinar com o objetivo de favorecer tal aprendizagem não é uma tarefa simples, pois corresponde em proporcionar condições para que o estudante perceba e interprete o significado que lhe foi apresentado e, por meio da interação cognitiva desta nova idéia, com algum conhecimento prévio relevante que já possua, (re) organize sua estrutura cognitiva absorvendo assim o novo conhecimento [Mendonça 2012].

Os conhecimentos prévios de um estudante podem ter sido adquiridos em atividades cotidianas, na escola, no entretenimento, ou seja, em diversas condições e assuntos dos mais variados temas, formando assim a sua estrutura cognitiva.

No contexto do ensino de linguagem de programação, principalmente nos últimos semestres de um curso, os conhecimentos prévios de um estudante são fomentados por disciplinas pré-requisitos de semestres anteriores. Em muitos casos, é percebido que os estudantes chegam com deficiências conceituais e despreparados para adquirir o novo conhecimento que lhe será transmitido.

Não identificar os conhecimentos prévios desses estudantes, principalmente as suas deficiências, e ensiná-los sem saber o que eles já conhecem, faz com que o docente pratique meramente o ensino mecânico ou repetitivo. Dessa forma, a aprendizagem pode não fazer sentido ao estudante e o novo conhecimento pode não se relacionar com o conhecimento existente em sua estrutura cognitiva. Nesse tipo de aprendizagem o novo conhecimento passa a ser armazenado isoladamente ou por meio de associações arbitrárias na estrutura cognitiva [Pelizzari *et al* 2002, Moreira 2012].

Nesse contexto o presente artigo propõe relatar uma intervenção que objetiva identificar os conhecimentos prévios e deficiências conceituais dos estudantes do sétimo semestre do Ensino Médio Integrado em Informática, no contexto da unidade curricular “Linguagem de Programação IV”, a fim de orientá-los na construção de seus conhecimentos com base na teoria da aprendizagem significativa, ou seja, ensiná-los a partir daquilo que já sabem. Dessa forma, de acordo com Novak e Gowin (1999), os estudantes se tornam mais capazes de aprender a aprender, uma habilidade fundamental para o estudante de informática devido às constantes mudanças tecnológicas.

Nesse sentido, entendemos como objetivos deste artigo, contribuir para identificação de alguns pontos de dificuldade a serem superados no processo de ensino de linguagem de programação, bem como motivar a prática docente em iniciativas de ensino que possam fazer sentido no aprender do estudante.

O método utilizado na intervenção foi composto por três fases. Na primeira fase foi aplicada uma avaliação diagnóstica [Luckesi 2003] com o objetivo de identificar os conhecimentos prévios dos estudantes em relação ao tema “Linguagem de Programação de Computadores”, como também suas deficiências conceituais. Na segunda fase, depois de uma análise do questionário aplicado na fase anterior por especialistas da área, foram preparados materiais compostos principalmente por vídeoaulas e postagens em blogs na internet, que consideramos serem organizadores prévios [Tavares 2010], com o objetivo de amenizar as deficiências conceituais dos estudantes em relação ao assunto que será ensinado, como também, ajustar os conhecimentos prévios dos estudantes para absorção do novo conhecimento. Finalmente, na terceira fase, os estudantes definiram seus projetos de softwares para serem desenvolvidos ao longo da unidade curricular.

2. Aprendizagem Significativa

Na concepção de Ausubel (2003), a aprendizagem significativa caracteriza-se pela interação entre um novo conhecimento e o conhecimento já existente na estrutura cognitiva do aprendiz de maneira substantiva (não literal) e não arbitrária, a um aspecto relevante da estrutura cognitiva do indivíduo. Nesse processo de aprendizagem, o conhecimento existente fica mais sólido, diferenciado e elaborado em termos de significados, adquirindo mais estabilidade e se internalizando por mais tempo [Moreira 2000].

Desse modo, os novos conhecimentos trabalhados pelo educador relacionam-se com o conhecimento que o estudante traz consigo; de experiências anteriores. Juntos eles formam uma estrutura de conhecimento que faz sentido ao estudante. Pode ser citado, como exemplo, a associação realizada pelo estudante, sobre conteúdos de diferentes disciplinas. Assim, um estudante que cursa linguagem de programação consegue relacionar o conteúdo de matrizes, conhecimento adquirido na disciplina de matemática, com vetores bidimensionais da informática, vislumbrando o potencial uso de matrizes matemáticas.

Tendo em vista o processo de aprendizagem, Ausubel (2003) afirma que só podemos aprender a partir daquilo que já conhecemos, destacando o conhecimento prévio como a variável mais importante no processo de aprendizagem. Assim, o aprendiz deve fazer uso do conhecimento internalizado, usando-o como âncora para absorção de novo conhecimento, tornando a aprendizagem progressiva, ou seja, os significados vão sendo captados e conectados e/ou fundidos progressivamente ao conhecimento já existente.

Cabe ao educador promover mecanismos para que o indivíduo consiga progressivamente diferenciar sua estrutura cognitiva, como também, reorganizar seus conhecimentos a fim relacioná-los. Dessa forma, é desejável que o novo conhecimento seja transmitido de maneira organizada e bem planejada, respeitando a sua estrutura hierárquica, além de procurar caminhar do mais genérico para o mais específico, sempre buscando estratégias para consolidação que seja atraente ao aluno.

Por outro lado, pouco se consegue se não houver condições para que a aprendizagem significativa ocorra. Ausubel (2003) afirma que para haver aprendizagem significativa são necessárias duas condições: primeira, o aprendiz precisa estar disposto a aprender. Se o aprendiz não estiver disposto a relacionar o novo conhecimento ao existente, optando por memorizá-lo, então a aprendizagem passa a ser mecânica. Isso mostra que o indivíduo não é só um receptor passivo e sim o protagonista no processo ensino-aprendizagem [Moreira 1980]; segunda, o novo conhecimento deve despertar o interesse do aprendiz, possuindo potencialmente significado, ou seja, o novo conhecimento deve ser lógico ao aprendiz e psicologicamente significativo.

3. Metodologia

A intervenção ocorreu em uma sala com 33 estudantes do sétimo semestre do Curso Técnico de Nível Médio Integrado em Informática do Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul, campus Campo Grande, na unidade curricular “Linguagem de Programação IV”, ministrada por um dos pesquisadores. O período de intervenção foi de vinte semanas, sendo que as aulas ocorreram no período vespertino, totalizando 40 horas-aula de 45 minutos cada.

O conteúdo abordado foi o de Desenvolvimento Web utilizando a Linguagem de Programação Java. No planejamento da unidade curricular foi definido que o conteúdo seria abordado por meio de aulas objetivas, atividades práticas e pesquisa, objetivando que cada estudante tenha condições de desenvolver uma aplicação voltada para web utilizando a Linguagem de Programação Java.

No início da unidade curricular o estudante deve definir um projeto de software a fim de resolver um problema real. Tal projeto é desenvolvido individualmente no decorrer de toda unidade. Espera-se com essa abordagem que o estudante tenha condições de aplicar os conceitos aprendidos de forma gradual, ou seja, à medida que o projeto é desenvolvido ao longo da unidade. Além disso, esperamos que resolvendo um problema cotidiano por meio de um projeto de software, o assunto a ser ensinado possa fazer sentido ao estudante e conseqüentemente despertar seu interesse, indo ao encontro das condições necessárias para que a aprendizagem significativa possa ocorrer.

Para tentar promover o princípio da aprendizagem significativa, ou seja, ensinar o estudante a partir daquilo que ele já sabe, optou-se por realizar uma avaliação diagnóstica a fim de levantar os conhecimentos prévios dos estudantes em relação ao tema que será ensinado. Esse mecanismo permite ao docente conhecer melhor seu grupo de alunos, ou seja, quais seus conhecimentos em relação ao tema e, principalmente, suas deficiências conceituais. Para atender os objetivos da unidade curricular, onde o estudante deve ser capaz de desenvolver uma aplicação web utilizando Java, é necessário que o estudante tenha noção, principalmente, em relação ao tema comunicação entre cliente e servidor, design de páginas utilizando HTML e folhas de estilo.

Nesse sentido foram aplicadas 11 questões para cada estudante sobre o tema que eles deveriam possuir certo conhecimento para dar início à unidade curricular, das quais suas respostas foram avaliadas por dois especialistas da área de desenvolvimento de software web. As respostas foram categorizadas em: Satisfatória (o estudante possuía certo conhecimento prévio sobre o tema); Parcialmente Satisfatória (o estudante poderia ter vestígios de conhecimento prévio sobre o tema) e; Insatisfatória (o estudante não demonstra possuir conhecimento prévio sobre o tema).

Para ser realizada esta classificação, destacou-se o fato de que as respostas não necessariamente deveriam estar corretas em sua plenitude para que sejam consideradas Satisfatórias, mas sim, o estudante deveria demonstrar que possuía alguns conceitos sobre o conceito abordado nas questões. O especialista deveria identificar na resposta do estudante algum conhecimento que pudesse ser capaz de servir como âncora ou requisito para a absorção do novo conteúdo que seria ensinado. Vale ressaltar, que antes dos especialistas avaliarem as respostas, foi explanada a eles a teoria da Aprendizagem Significativa com enfoque principalmente no conceito de conhecimentos prévios.

Após análise das repostas foram preparados organizadores prévios, utilizando vídeoaulas e postagens em blogs da internet, para servirem de âncora para a absorção do novo conhecimento que seria ensinado, ou seja, foram introduzidos conteúdos sobre os conceitos que os estudantes não demonstraram conhecimento prévio aceitável para absorver o novo conhecimento.

Organizadores prévios são conceitos introdutórios, apresentados a fim de manipular a estrutura cognitiva dos estudantes, mediando e/ou interligando conceitos que o estudante sabe com o que ele pretende aprender [Ausebel 2003]. Assim, os organizadores prévios funcionam como conectores cognitivos, fornecendo um suporte para a incorporação e retenção sólida de novos conceitos, atuando como uma estratégia facilitadora para promoção da aprendizagem significativa [Tavares 2010].

Após a aplicação dos organizadores prévios, o mesmo questionário de avaliação diagnóstica foi submetido aos estudantes, e suas respostas foram avaliadas pelos mesmos dois especialistas, realizando comparações entre os primeiros resultados, a fim de verificar se houve avanço ou não em relação aos conhecimentos prévios dos estudantes.

Finalmente, após analisar os conhecimentos prévios dos estudantes e promover o reparo das possíveis arestas conceituais que possuíam, o conteúdo novo passou a ser ensinado de forma gradual, sendo que cada novo conceito ensinado deveria ser aplicado pelo estudante em seu projeto de software a fim de construí-lo de maneira incremental até conseguir resolver um problema do cotidiano.

4. Análise dos dados da intervenção

A análise se concentra nas três fases da intervenção, a avaliação diagnóstica, a aplicação dos organizadores prévios e na observação do aproveitamento dos estudantes, como também, na prática docente no decorrer da unidade curricular.

4.1. Avaliação diagnóstica

Na avaliação diagnóstica foram aplicadas questões abertas relacionadas ao tema Internet e Desenvolvimento Web, como apresentadas na Tabela 1.

Tabela 1. Questões utilizadas na avaliação diagnóstica

Questões	
Q1	O que você entende por aplicação cliente-servidor?
Q2	Em uma aplicação cliente-servidor, como o cliente e servidor se comunicam?
Q3	O que é HTTP?
Q4	Quais são os métodos usados no HTTP?
Q5	O que você entende por páginas estáticas e páginas dinâmicas no contexto da web?
Q6	O que você entende por servidor de aplicação?
Q7	Em uma aplicação cliente-servidor, como são transferidas as informações do cliente para o servidor?
Q8	O que é HTML?
Q9	Para que você acha que é utilizado um formulário HTML?
Q10	O que você entende por requisição e resposta no contexto de uma aplicação cliente-servidor?
Q11	O que você entende por aplicação web?

O Gráfico na Figura 1 apresenta os resultados da avaliação diagnóstica. Os resultados de cada questão (Q1 a Q11) são apresentados em formato de porcentagem, sendo as respostas classificadas como: Satisfatório (SA), Parcialmente Satisfatória (PS) e Insatisfatória (IN). No gráfico é apresentada a média simples da análise dos dois especialistas para cada questão. Por exemplo, o especialista 1 considerou que 65% das repostas da Q1 eram consideradas insatisfatórias (IN), já o especialista 2 considerou 67% (IN), no cálculo da média simples conseguiu-se 66% de respostas insatisfatórias (IN) para Q1. Esse critério foi aplicado para todas as questões para definir o percentual de AS, PA e IN para cada uma delas.

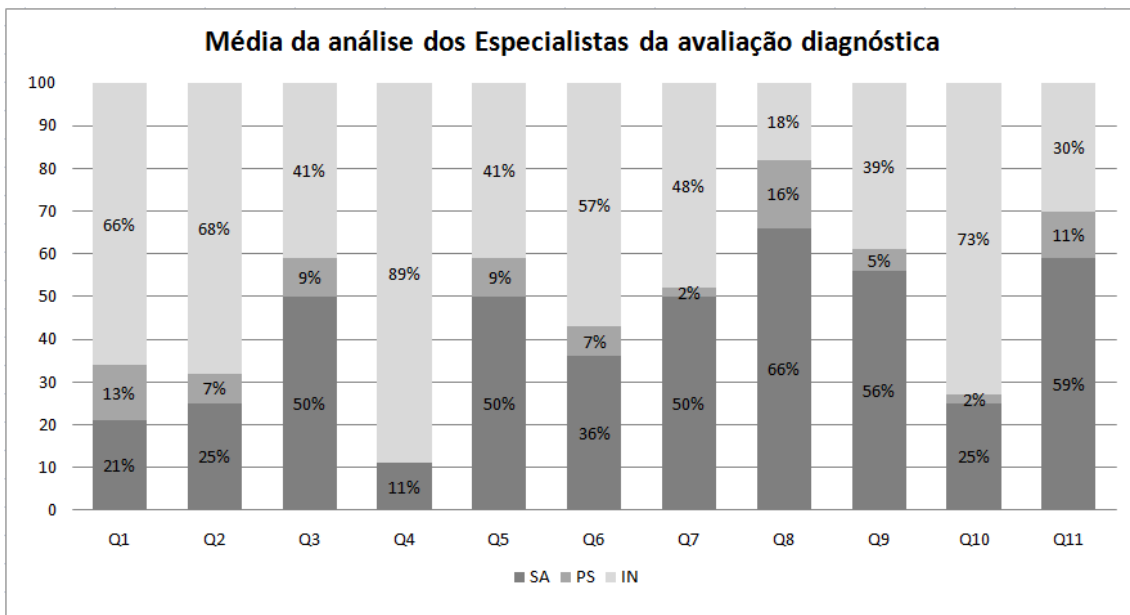


Figura 1. Resultado da análise da avaliação diagnóstica antes da introdução dos organizadores prévios.

Observou-se que os estudantes possuíam deficiências em relação ao conhecimento prévio indispensável para aprender o novo conteúdo em relação à temática. No resultado da questão Q4, por exemplo, que tinha como objetivo verificar se os estudantes possuíam conhecimento sobre “*os métodos de comunicação entre um cliente e servidor*”, um conhecimento base para se entender como um cliente se comunica com um servidor, verificou-se que apenas 11% das respostas foram satisfatórias. Isso demonstra indícios de que o grupo de estudantes possuía pouco conhecimento prévio sobre o conteúdo.

Notou-se também em respostas às questões Q1, Q2, Q6 e Q10, que respectivamente objetivavam identificar se os estudantes possuíam conhecimentos prévios sobre “*aplicações cliente-servidor*”, “*comunicação entre cliente e servidor*”, “*servidor de aplicação*” e “*requisição e resposta entre cliente e servidor*”, os resultados foram abaixo do desejável. Pois, esse tipo de conhecimento, mesmo que o estudante traga consigo de forma não consolidada, é imprescindível para o entendimento do conteúdo que seria ensinado.

As respostas às questões Q3, Q5, Q7, Q8, Q9 e Q11 atingiram um nível de satisfação (SA), em relação aos conhecimentos prévios exigidos para a compreensão do novo conteúdo, igual ou superior a 50%. Este fato não pode ser considerado ótimo, uma vez que, nenhuma das questões aproximou-se de 100%. Por outro lado, por se tratar de um conteúdo complexo, até mesmo para estudantes de graduação, os conhecimentos prévios observados dos estudantes para essas questões podem ser considerados razoáveis para a introdução do novo conteúdo.

4.2. Aplicação dos organizadores prévios

Os organizadores prévios, baseados em materiais expositivos (postagens em blogs da Internet e vídeoaulas), trataram principalmente de conceitos que os estudantes

apresentaram maior dificuldade, observadas a partir da análise de suas repostas na avaliação diagnóstica.

Durante a exposição das vídeosaulas o docente sempre pausava e apresentava os conceitos de forma diferente, a fim de consolidar o conteúdo. Essa abordagem foi adotada, principalmente quando eram expostos os conceitos referentes às questões Q1, Q2, Q4, Q6 e Q10, onde foi percebido que os estudantes possuíam pouco conhecimento prévio sobre os conceitos necessários para o entendimento do novo conteúdo. O Gráfico da Figura 2 apresenta os resultados após a análise da avaliação diagnóstica e introdução de organizadores prévios.

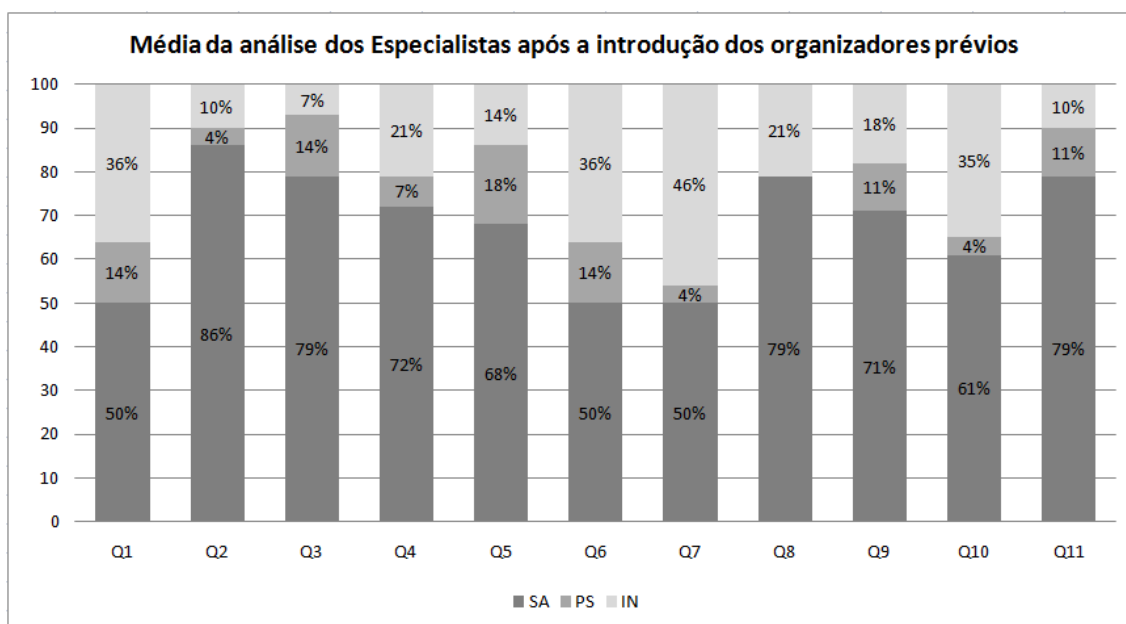


Figura 2. Resultado da análise das repostas dos estudantes após a introdução dos organizadores prévios.

As questões Q1, Q2, Q4, Q6 e Q10, após a introdução dos organizadores prévios, tiveram um avanço em relação à avaliação diagnóstica. Todas atingiram um nível de satisfação (SA) acima de 50%. A questão Q4, que possuía o pior desempenho, obteve um avanço considerado satisfatório, não somente em números, mas na qualidade das repostas observada durante a análise textual. As demais questões Q1, Q2, Q6 e Q10 as repostas dos estudantes também demonstraram maior qualidade. Esse avanço demonstra que há indícios que houve um efeito satisfatório em relação à abordagem adotada pelo professor, que sempre pausava a exposição dos vídeos e reforçava a explicação dos conceitos.

As outras questões também demonstraram uma evolução quanto ao índice de satisfação (SA), como também na qualidade das repostas. Os dois Gráficos da Figura 3 mostram a evolução dos conhecimentos prévios dos estudantes na avaliação diagnóstica (Antes) e após a aplicação dos organizadores prévios (Depois). Os gráficos são baseados nas médias dos índices Satisfatório (SA), Parcialmente Satisfatório (PS) e Insatisfatório (IN).

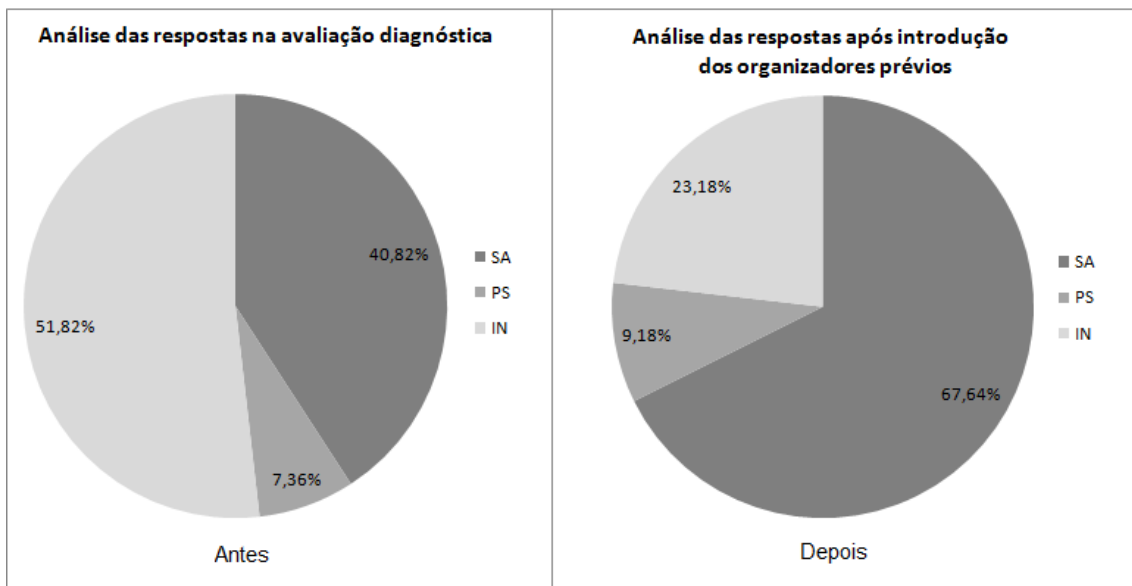


Figura 1. Resultado da análise das questões antes e após a introdução dos organizadores prévios.

Comparando os gráficos na Figura 3, observou-se que houve uma evolução em relação aos conhecimentos prévios dos estudantes necessários para receber o novo conhecimento. O que pode significar indícios de que a aplicação de organizadores prévios, a fim de aparar as arestas conceituais dos estudantes, apresentou resultados que, em certa medida, refletem um avanço do estado anterior de conhecimento.

Além disso, com os dados levantados sobre os conhecimentos dos estudantes, o docente poderá contemplar em sua prática pedagógica ações que estejam alinhadas com aquilo que os estudantes já sabem, promovendo assim a aprendizagem significativa, ou seja, o docente pode utilizar artifícios para que os estudantes consigam relacionar o novo conhecimento ao conhecimento prévio existente.

4.3. Aproveitamento dos estudantes

O desempenho individual de cada estudante foi analisado com base em sua média final na unidade. Para compor tal média, não foram aplicadas avaliações consideradas tradicionais, onde os estudantes respondem questões dissertativas ou escrevem trechos curtos de código para demonstrarem que aprenderam determinado conteúdo.

A avaliação da aprendizagem foi composta por três etapas durante o processo de desenvolvimento do projeto de software. Em cada etapa o estudante tinha que apresentar o projeto com as funcionalidades previamente definidas pelo docente. Por meio de uma entrevista o docente verificava se o estudante atingiu o objetivo daquela etapa, ou seja, se as funcionalidades estavam todas implementadas e se de fato foi o estudante que as fez. Para a segunda verificação o docente solicitava ao estudante que explicasse em detalhes como foi implementada uma determinada funcionalidade.

Para o estudante conseguir aprovação na unidade curricular sua média final deve ser superior ou igual a 7 e a frequência deve ser de no mínimo 75%. As médias finais ficaram da seguinte forma:

- 43% dos estudantes obtiveram notas entre 9 e 10, esses estudantes conseguiram implementar o projeto de software integralmente, aplicando todos os conceitos aprendidos durante a unidade curricular.
- 24% obtiveram notas entre 8 e 9, alguns desses estudantes implementaram todo o projeto, mas durante as entrevistas eles não souberam responder algumas perguntas ao docente. Além disso, alguns estudantes, na primeira ou segunda etapa, não tinham implementado todas as funcionalidades definidas para aquela fase do projeto;
- 21% obtiveram entre 7 e 8, esses estudantes não entregaram todas as funcionalidade definidas para seus respectivos projetos em cada etapa, como também, alguns não souberam responder perguntas feitas pelo docente durante as entrevistas.
- 9% de reprova, ou seja, nota final inferior a 7 e 3% de desistência. Um fato observado entre esses estudantes é que todos tinham dependências de semestres anteriores e não concluiriam o curso naquele semestre.

5. Análise e considerações finais

Em um contexto onde se discute modelos tradicionais de educação, em que o ensino tipicamente é tratado de forma sequencial e segmentado, propor novos meios para ensinar a fim de despertar o interesse dos estudantes pode ser considerado um desafio. Conduzir o ensino por meio de projetos reais a serem desenvolvidos pelos próprios estudantes demonstrou ser uma alternativa para estimular um maior engajamento dos mesmos uma vez que conseguem atingir o objetivo da unidade curricular de uma maneira mais positiva. Dito de outro modo, fazendo algo que faça sentido a eles. Nessa perspectiva a aprendizagem pode ser considerada significativa ao estudante, visto que o desenvolvimento de projetos de certa forma obriga-o a relacionar o seu conhecimento existente com o novo conhecimento a fim de construir algo palpável e percebível.

Para o docente, conhecer o que o estudante já sabe é importante, porque dá a oportunidade, durante a sua prática, de fazer relações mais seguras entre o conhecimento prévio do estudante com o novo conhecimento que será ensinado. Isso pode promover a consolidação do conhecimento prévio do estudante fazendo com que ele tenha sentido após a interação com o novo conhecimento.

O desenvolvimento de projetos ao longo da unidade curricular dá a oportunidade ao docente de avaliar a aprendizagem de forma não tradicional, como ocorreu durante a intervenção. Percebe-se que nessa abordagem, o estudante não demonstra tanta angústia e insegurança no momento de ser avaliado, pois ele está apresentando seu trabalho que foi realizado ao longo de algumas semanas. Isso pode evitar que o estudante aprenda mecanicamente com o objetivo de ser aprovado em um único teste. Por outro lado o estudante passa a ser estimulado a buscar novos conhecimentos fora do escopo da unidade curricular a fim de atingir seu objetivo.

Finalmente, acreditamos que o trabalho a partir de projetos resulte num constante movimento de reflexão e retomadas das práticas pedagógicas e suas implicações, onde professores e alunos compartilham o protagonismo do processo de ensino e aprendizagem de Linguagem de Programação e, principalmente, decidem por

um trabalho fora de suas zonas de conforto, inaugurando assim um espaço diferenciado de aprendizagem dos conceitos aqui apresentados.

Agradecimentos

Nós agradecemos ao Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul (IFMS) pelo apoio financeiro recebido por meio do Edital nº 009/2015 - PROPI/IFMS.

Referências

- Ausubel, D. P. (2003). *Aprender a aprender*, Plátano Edições Técnicas, 1ª edição.
- Luckesi, C. C. (2003). *Avaliação da aprendizagem na escola: reelaborando conceitos e recriando a prática*, Malabares Comunicação e Eventos.
- Mendonça, C. A. S. (2012). Investigando conhecimentos dos licenciandos em biologia sobre aprendizagem significativa e mapas conceituais. In *Aprendizagem Significativa em Revista*, p.14-24.
- Moreira, M. A. (1980). Mapas conceituais como instrumentos para promover a diferenciação conceitual progressiva e a reconciliação integrativa. *Ciência e Cultura*, p. 474-479.
- Moreira, M. A. (2000). Aprendizagem significativa crítica. In: III Encontro Internacional sobre Aprendizagem Significativa, 2000, Peniche. Anais do III Encontro Internacional sobre Aprendizagem Significativa, 2000.
- Moreira, M. A. (2012). Al final, que és aprendizaje significativo? *Qurrriculum (La Laguna)*, p. 29-56.
- Novak, J. P. and Gowin, D. B. (1999). *Aquisição e retenção de conhecimentos: uma perspectiva cognitiva*, Plátano Edições Técnicas, 1ª edição.
- Pelizzari, A., Kriegl, M. L., Baron, M. P., Finck, N. T. L. and Dorocinski, S. I. (2002). Teoria da Aprendizagem Significativa Segundo Ausubel. *Revista do Programa de Educação Corporativa (PEC)*, p. 37-42.
- Tavares, R. (2010). Aprendizagem significativa codificação dual e objetos de aprendizagem. *Revista Brasileira de informática na Educação*, p. 4-16.

Ensino de Programação para Futuros Não-Programadores: Contextualizando os Exercícios com as Demais Disciplinas de mesmo Período Letivo

**Leandro S. G. Carvalho, Bruno F. Gadelha, Fabíola G. Nakamura,
David B. F. Oliveira, Elaine H. T. Oliveira**

Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Av. General Rodrigo Octávio, 6200 – Coroado I – CEP: 69077-000 – Manaus – AM – Brasil

{galvao,bruno,fabiola,david,elaine}@icomput.ufam.edu.br

***Resumo.** Disciplinas de introdução à programação são usualmente ministradas para cursos de graduação em Engenharia e Ciências Exatas com o propósito de desenvolver nos aprendizes algumas habilidades e competências, tais como: raciocínio lógico, capacidade de resolver problemas utilizando associação, generalização, modularização, entre outras. Muitos dos alunos desses cursos costumam não ter motivação para se aplicar à disciplina, já que a programação não será sua atividade profissional fim. Para trabalhar essa dificuldade, uma equipe de professores concebeu uma metodologia de ensino-aprendizagem que considera o conjunto de disciplinas que estão sendo ministradas no mesmo período letivo. A metodologia foi aplicada em quatro turmas, de cursos e professores diferentes, no primeiro período letivo de 2015. Os resultados preliminares indicam um índice de aprovação maior do que em turmas lecionadas no mesmo período onde a metodologia não foi aplicada. A metodologia foi avaliada pelos discentes que, na maioria, aprovaram a dinâmica de aplicação e as atividades propostas.*

***Abstract.** Introductory programming courses are part of undergraduate degree programs in engineering and exact sciences with the goal of developing in the students some skills and competencies, such as logical reasoning, problem-solving by means of association, generalization, modularization, among others. Many students of these programs usually have no motivation to engage in the course, since programming is not directly related to their future professional career. To deal with this issue, a group of professors developed a learning methodology that considers the set of courses taken at the same term. The methodology was applied in four classes in different programs and by different professors, in the first term of the 2015 academic year. The preliminary results indicate that the approval rating among students was higher than that of classes in the same term where the methodology was not applied. The methodology was evaluated by the students that, in most cases, approved the dynamic and the proposed activities.*

1. Introdução

Disciplinas de introdução à programação costumam fazer parte do ciclo básico de formação nos currículos de cursos de graduação em Engenharia e Ciências Exatas. Pela falta de maturidade em perceber a importância da disciplina no exercício da sua

atividade profissional, alunos desses cursos normalmente não têm motivação suficiente para se aplicar à disciplina, o que resulta em altos índices de reprovação na mesma.

Para trabalhar essa dificuldade, uma equipe de professores da Universidade Federal do Amazonas concebeu uma metodologia de ensino-aprendizagem dinâmica, que não se limita à exposição de conteúdo, mas que trabalha com diferentes formas de exercícios. Além disso, na elaboração de tais exercícios, os domínios de problemas trabalhados em cada curso são considerados, bem como as disciplinas que estão sendo ministradas no mesmo período letivo. Essa tarefa de elaborar exercícios relevantes e contextualizados é essencial para o sucesso da disciplina e engajamento dos estudantes, ainda que não seja uma tarefa trivial [Feldman e Zelenski 1996, Stevenson e Wagner 2006, e Hundley e Britt 2009]. O objetivo é estimular os alunos a perceberem a importância da disciplina em sua formação integral.

Neste artigo, relatamos a metodologia de ensino-aprendizagem concebida, em grande parte baseada nos trabalhos de Píccolo et al. (2010) e Zanini e Raabe (2012), bem como os resultados. Na próxima seção, contextualizamos a disciplina na universidade em que foi lecionada e na Seção 3 apresentamos os trabalhos relacionados. Em seguida, detalhamos a metodologia de ensino-aprendizagem, aplicada em quatro turmas, de cursos e professores diferentes, no primeiro período letivo de 2015. Na penúltima seção, analisamos os resultados obtidos em termos de índice de aprovação, índice de facilidade e poder de discriminação das questões das avaliações, bem como o *feedback* qualitativo dos discentes. Por fim, propomos trabalhos futuros.

2. Contextualização

A disciplina de Introdução à Programação de Computadores (IPC) é ministrada para cerca de 500 alunos do primeiro período de onze cursos de Engenharia e de Ciências Exatas na Universidade Federal do Amazonas. Cerca de metade dos ingressantes é selecionada por meio do Sistema de Seleção Unificada (Sisu) e a outra metade por meio do Processo Seletivo Contínuo (PSC), realizado pela própria universidade ao longo dos três anos do ensino médio.

Como acontece em outras universidades [Píccolo et al. 2010], por se tratar de uma disciplina de serviço, poucos professores se sentem encorajados a ministrar IPC, de modo que ela é normalmente atribuída a professores substitutos. Entretanto, ensinar programação para alunos iniciantes não é uma tarefa trivial, pois eles vêm com uma base matemática insipiente e estão em cursos em que a Computação não é uma atividade fim.

O gráfico da Figura 1 mostra a evolução da taxa de aprovação na disciplina de IPC, isto é, a razão entre o número de alunos aprovados e o número de alunos matriculados, considerando todas as onze turmas ofertadas em 2015. Os decrescentes índices obtidos de 2011 a 2014, bem como o aumento no número de matrículas, motivaram um grupo de professores a reformular a metodologia de ensino, com o propósito de reduzir a taxa de reprovação por motivo de nota.



Figura 1 – Evolução da taxa de aprovação na disciplina IPC.

O gráfico apresenta dois pontos destoantes na tendência apontada: os anos de 2010 e de 2015. Em 2010, a Universidade adotou, pela primeira vez, o Sisu como processo seletivo. No Sisu o aluno pode selecionar até 02 (duas) opções de curso, indicando qual a sua ordem de preferência. Tem-se observado que, quando o candidato não possui nota suficiente para ser selecionado em sua primeira opção, ele costuma trocar a ordem de preferência. Só mais tarde, depois de matriculado, ele percebe que o curso não corresponde à sua vocação e o abandona, o que se reflete na taxa de reprovação por frequência. Já no ano de 2015, o aumento na taxa de aprovação pode ser creditado provavelmente à adoção da metodologia aqui apresentada em quatro das onze turmas da disciplina, como veremos nas próximas seções.

O regimento da Universidade não permite que o aluno tranque disciplinas ofertadas para os dois primeiros períodos letivos. Conseqüentemente, quando o aluno não é aprovado em IPC, ou ele reprova por motivo de nota (quando não atinge média 5, de 0 a 10 pontos), ou por motivo de frequência (quando não frequenta pelo menos 75% das aulas previstas, independentemente do desempenho).

Parte da reprovação pode ser explicada pelo descontentamento do aluno com o curso, o que uma disciplina isolada como IPC não contribui substancialmente. Porém, parte da reprovação pode ser explicada pela falta de uma metodologia de ensino-aprendizagem que os estimulasse a se empenhar.

Para tratar tal problema, fomos buscar na literatura relatos e técnicas que nos fornecessem embasamento para reformular a metodologia de ensino-aprendizagem de formar eficiente e eficaz.

3. Trabalhos relacionados

Segundo Araujo et al. (2013), “no aprendizado de programação, é essencial que os alunos direcionem o seu estudo à codificação e à resolução de exercícios”. Os autores identificaram que o estudo regular (diário) é um fator que contribui para o sucesso do aluno na disciplina, mais do que a quantidade de tempo total dedicada ao estudo.

Nesse mesmo sentido, Ihantola et al. (2015) afirmam que “a avaliação contínua durante um curso de programação garante que os estudantes pratiquem bastante, bem como obtenham *feedback* sobre a qualidade de suas soluções”. Isso acontece porque a avaliação orienta a aprendizagem dos alunos e serve de *feedback* tanto para o aluno

quanto para o professor sobre o processo de aprendizagem, seja de um tema em específico, seja do curso como um todo.

Contudo, a elaboração de muitos exercícios de programação requer dois tipos de tarefas [França et al. 2011]:

- Gerenciamento de atividades: compreende a elaboração e disponibilização de exercícios, controle da submissão de trabalhos e registro de notas.
- Acompanhamento do aprendiz: compreende a análise de corretude das soluções submetidas e o provimento de *feedback* ao aluno para que este entenda seus erros.

Ambas as tarefas despendem muito tempo do professor e monitores, e não há garantia de que sejam eficientes. Um ambiente virtual de aprendizagem (AVA) típico resolve o problema de gerenciamento das atividades, mas não oferece suporte ao acompanhamento do aprendiz. Para contornar esse problema, muitos cursos de programação lançam mão de sistemas conhecidos como “juízes online”.

Juízes online são sistemas que compilam, executam e testam códigos-fonte com base em dados padronizados para julgar se estão corretos. Ihantola et al. (2015) apresentam uma revisão sobre sistemas de avaliação automática – que neste artigo tomamos como sinônimo de juízes online. De acordo com Pieterse (2013), o uso de ferramentas automáticas aumenta a necessidade de uma cuidadosa concepção pedagógica das configurações do exercício e de sua avaliação. Quando os exercícios são avaliados manualmente, é possível contornar problemas causados por enunciados confusos, pontuando-se pela criatividade das soluções. No entanto, esse nem sempre é o caso quando juízes online são utilizados como ferramenta de ensino e avaliação.

Em contrapartida, como apontam Hundley e Britt (2009) e Pieterse (2013), desenvolver exercícios de programação contextualizados e automatizados requer uma atenção especial, a fim de evitar problemas que possam comprometer o êxito dos estudantes. Como apontado por Zanini e Raabe (2012), esse tipo de exercício não é comum, pelo menos em língua portuguesa, o que dificulta a reutilização. Portanto, o docente que se propõe a aderir à risca às premissas aqui colocadas deve dispor de um tempo para planejar questões de enunciado claro, de contextos variados, em diversas versões de mesmo grau de dificuldade para permitir aleatorização, e com exemplos de entradas e saídas que deem um indício do processo de solução. Também requer consulta a livros de Cálculo, Álgebra Linear, Física, Estatística, entre outros, para relembrar conteúdos que já não domina mais, a fim de aplicá-los nos exercícios de programação.

Essa carga de trabalho extra pode ser equivalente ou até mesmo suplantar a carga de trabalho envolvida na correção de atividades no modelo tradicional de ensino de programação [Pieterse 2013]. Para contornar esse efeito colateral, é essencial que uma equipe de professores e tutores divida entre si as tarefas, de modo que, na média, o trabalho seja menor para todos.

4. Descrição da Metodologia de Ensino-Aprendizagem

Com base nos trabalhos discutidos na Seção 3, os autores deste artigo decidiram reformular a metodologia de ensino da disciplina de IPC. Tomou-se como ponto de partida a metodologia proposta por Píccolo et al. (2010), cujos resultados foram bem-

sucedidos em um cenário muito semelhante à da instituição de ensino dos autores. A metodologia adotada apresenta as características descritas nos próximos parágrafos.

A disciplina – que prevê 60 horas de aulas presenciais, sendo 30h de carga teórica e outras 30h de prática – foi remodelada para sete módulos de oito horas (quatro encontros de duas horas), mais um módulo inicial de quatro horas (dois encontros) de motivação e apresentação da disciplina. O conteúdo foi distribuído da seguinte maneira: (1) variáveis; (2) estruturas condicionais; (3) estruturas condicionais aninhadas; (4) estrutura de repetição por condição; (5) vetores; (6) estrutura de repetição por contagem; e (7) matrizes. O oitavo módulo sobre *strings* teve o conteúdo ministrado, mas não foi avaliado. O tópico de “funções” foi ministrado de forma diluída entre os módulos de 5 a 7. Cada módulo tinha quatro aulas de duas horas de duração, na seguinte sequência: aula teórica, laboratório de codificação, laboratório de exercícios, e avaliação parcial. Todas as aulas eram realizadas em laboratório com computadores.

Os exemplos e exercícios trabalhados em cada módulo eram cumulativos. Por exemplo, embora o tópico principal do módulo 6 fosse “estruturas de repetição por contagem”, nele eram trabalhados conceitos de vetores e de estruturas condicionais. Por conta disso, as avaliações parciais dos módulos mais avançados tinham maior peso no cômputo da média final na disciplina.

A primeira aula de cada módulo, a aula teórica, era de natureza presencial. Nela, o professor apresentava o conteúdo do módulo à classe, que era estimulada a reproduzir alguns exemplos nos computadores à sua disposição. A segunda e a terceira aula, chamadas de “laboratórios”, eram de natureza prática e nelas a presença era facultativa aos alunos. Elas eram acompanhadas apenas pelo tutor, a fim de sanar dúvidas. O desempenho dos alunos nessas atividades contribuía, em menor peso, para a média final na disciplina.

O **laboratório de codificação** era destinado à codificação de problemas cadastrados no juiz online *CodeBench*¹. Nessa ferramenta, as questões eram corrigidas automaticamente e o aluno recebia *feedback* imediato. Tais questões tinham como objetivo exercitar a habilidade de resolução de problemas, abstração e codificação. O **laboratório de exercícios** era destinado à resolução de questões cadastradas em um AVA semelhante ao *Moodle*². Tais questões eram de cunho mais conceitual ou trabalhavam a habilidade de rastrear código.

A divisão entre laboratório de codificação e laboratório de exercícios tinha finalidade apenas em termos de planejamento. Na prática, os exercícios de ambas as atividades eram liberados para visualização e resolução a partir do dia da aula teórica, com validade até a véspera da avaliação parcial. O aluno ficava à vontade para tirar dúvidas de qualquer uma das atividades durante o tempo de aula com o tutor.

A quarta e última aula do módulo era destinada à avaliação do conteúdo ministrado. As questões da avaliação normalmente eram adaptações de questões trabalhadas no laboratório de codificação. O desempenho dos alunos na avaliação parcial contribuía em maior peso para a média final na disciplina.

¹ Disponível em <http://codebench.icomp.ufam.edu.br>

² Disponível em <https://moodle.org/>

Adotou-se a linguagem Python, pois sua sintaxe é de simples compreensão, exige que o programador organize o código respeitando indentação, possui material didático em português, além de outras vantagens [Menezes 2014]. Adotou-se a IDE Spyder, pelas seguintes características: tem versões para vários sistemas operacionais; possui destacador de sintaxe; incorpora as bibliotecas Numpy e MathPlotLib em seu instalador; e inclui um visualizador de estado das variáveis (critério decisivo quando se lida com aprendizes com baixa capacidade de abstração).

5. Resultados

Nesta seção, apresentamos a taxa de aprovação nas turmas em que a metodologia foi aplicada e nas turmas em que não foi; o índice de facilidade e o índice discriminação das questões elaboradas para as avaliações; e uma análise qualitativa das opiniões dos alunos sobre a metodologia.

A Tabela 1 apresenta os cursos para os quais a disciplina IPC foi ministrada em 2015, destacando-se as quatro turmas que participaram da metodologia. Dos 172 alunos matriculados nelas, 61,6% foram aprovados. Dos 319 alunos matriculados nas outras seis turmas, 40,8% foram aprovados. Essa diferença de mais de 20% parece indicar que a metodologia foi bem-sucedida, pelo menos na sua primeira aplicação, em 2015.

Em seguida, analisamos se essa diferença poderia ser explicada por um enviesamento na atribuição de turmas com alunos melhor preparados no Ensino Médio aos professores que adotaram a metodologia. Verificou-se, no entanto, que a correlação tau de Kendall (pois trabalhou-se com um pequeno conjunto de dados) entre a nota de corte no Sisu e a taxa de aprovação em IPC, independente da metodologia aplicada, resultou em um valor em torno de 0,6 (em uma escala de -1 a +1). Esse resultado sugere, portanto, que, há uma forte correlação positiva entre a nota de corte no Sisu e a taxa de aprovação em IPC que independe da metodologia aplicada. Desse modo, a maior taxa de aprovação pode ser creditada à adoção da metodologia, pelo menos no ano de 2015.

Tabela 1 – Cursos de graduação para os quais foi ministrada a disciplina de Introdução à Programação de Computadores em 2015.

Curso	Nota de corte SISU	Nº de matrículas	Nº de aprovados	Percentual de aprovados
Engenharia Química	703,6	61	43	70,5%
Engenharia Mecânica	694,0	54	43	79,6%
Matemática Bacharelado Diurno	645,3	17	7	41,2%
Matemática Licenciatura Matutino	625,6	40	13	32,5%
Engenharia de Petróleo e Gás	703,8	50	28	56,0%
Engenharia de Produção	682,9	52	33	63,5%
Engenharia de Materiais	666,8	47	23	48,9%
Física Bacharelado Diurno	657,6	25	3	12,0%
Estatística	630,9	56	20	35,7%
Matemática Licenciatura Noturno	626,4	57	18	31,6%
Física Licenciatura Diurno	615,1	32	5	15,6%

Após o fechamento das turmas, analisamos a adequabilidade das questões com relação ao índice de facilidade e o poder de discriminação. Esses critérios são adotados pelo INEP no Exame Nacional de Desempenho dos Estudantes (ENADE) para avaliar as questões propostas em cada exame.

Para o INEP, o índice de facilidade de uma questão está diretamente relacionado com seu percentual de acerto. Questões com baixo percentual de acerto são consideradas difíceis, enquanto as que possuem alto índice de acerto são consideradas fáceis. A Tabela 2 mostra os índices de facilidade adotados pelo comitê do ENADE.

Tabela 2 – Classificação de questões segundo o índice de facilidade.
Fonte: INEP (2014)

Índice de Acerto	Classificação
Acima de 86%	Muito fácil
61% a 85%	Fácil
41% a 60%	Médio
16% a 40%	Difícil
Abaixo de 15%	Muito difícil

A Tabela 3 apresenta algumas características das questões elaboradas para a avaliação dos alunos, incluindo o contexto explorado pelo enunciado de cada uma delas. A quarta coluna mostra os índices de facilidade das questões. Relacionando a Tabela 3 com a Tabela 2, percebemos que a maioria das questões foi classificada como fácil ou mediana. Esse alto nível de facilidade é um reflexo da contextualização dos enunciados com as experiências e vivências do aluno. Apenas a questão 6 foi classificada como uma questão difícil, pois envolvia conceitos mais avançados de geometria, que muitos alunos tiveram dificuldade de resolver.

Tabela 3 – Questões usadas para avaliação dos alunos.

Avaliação	Questão	Contexto	Índice de Acerto	Valor de r_{pb}
1	A	Comercial: percentual de gorjeta	77%	0,38
	B	Geometria: volume de recipiente e custo	64%	0,38
2	A	Escolar: aprovação em disciplina	79%	0,37
	B	Matemática: aritmética	67%	0,33
3	A	Físico-química: estado físico de elementos	81%	0,34
	B	Geometria: volume de um sólido não trivial	21%	0,34
4	A	Cálculo: aproximação de séries	77%	0,43
	B	Financeira: aplicação bancária	56%	0,46
5	A	Cálculo: derivada de polinômios	68%	0,40
	B	Estatística: identificação de padrão em vetor	74%	0,39
6	A	Estatística: contagem de ocorrências em vetor	57%	0,45
	B	Frações contínuas	71%	0,43
7	A	Álgebra Linear: sistema de equações lineares	55%	0,39
	B	Álgebra Linear: manipulação de matrizes	51%	0,42

Outra métrica adotada pelo INEP é o poder de discriminação. Dizemos que uma questão possui alto poder de discriminação quando ela é capaz de diferenciar alunos de bom desempenho dos alunos de mau desempenho. A capacidade de discriminação de uma dada questão pode ser quantificada através do coeficiente de correlação ponto-bisserial, que mede a correlação entre uma variável dicotômica (acerto/erro da questão) e uma variável numérica (nota final do aluno). Esse coeficiente é dado por:

$$r_{pb} = \frac{M_Q - M_T}{DP_T} \sqrt{\frac{p}{q}}$$

onde M_Q é a média das notas finais dos alunos que acertaram a questão Q, M_T é a média das notas finais de todos os alunos, DP_T é o desvio padrão das notas finais de todos os alunos, p é o percentual de alunos que acertaram a questão Q, e q é a proporção de alunos que erraram a questão Q ($q = 1-p$).

Tabela 4 – Classificação de questões segundo o poder de discriminação.
Fonte: INEP (2014)

Valor de r_{pb}	Classificação
Acima de 0,40	Muito bom
0,30 a 0,39	Bom
0,20 a 0,29	Médio
Abaixo de 0,19	Fraco

A Tabela 4 apresenta as classes de poder discriminatório adotadas pelo INEP. Relacionando-a com a última coluna da Tabela 3, que apresenta os valores de correlação ponto-bisserial alcançadas pelas questões das avaliações em IPC, podemos observar que todas as questões obtiveram um índice de discriminação bom ou muito bom. Dessa forma, podemos concluir que a probabilidade de um aluno ter acertado uma questão esteve diretamente relacionada com seu domínio sobre o tópico de programação explorado na questão, e não com um efeito aleatório.

Ao final da disciplina, foi solicitado que os alunos participassem de uma pesquisa para avaliar a metodologia utilizada através do preenchimento de um questionário. Esse questionário consistia de perguntas sobre o entendimento e a dificuldade na realização das diversas atividades propostas, o grau de interação com professores e tutores, e a opinião do aluno acerca da metodologia utilizada. Neste artigo, serão apresentados somente os resultados relacionados às atividades e à metodologia.

Para analisar os comentários dos participantes nas perguntas abertas, foi usada a técnica descrita no Método de Explicitação do Discurso Subjacente (MEDS) [Nicolaci-da-Costa 2007]. Foi realizada uma análise inter-participantes onde suas respostas foram sistematicamente comparadas em busca de recorrências, emergindo assim as categorias. Parte delas é apresentada na Tabela 5.

Sobre as atividades realizadas (categorias 1, 2, 3 e 4), os alunos perceberam a relação entre o conteúdo ministrado nas aulas presenciais com as atividades cobradas na avaliação. Destacaram também o nível de exigência de cada atividade, como por exemplo, na citação “Estavam na dificuldade adequada: não eram extremamente difíceis, mas podiam ser respondidos com estudo e conhecimento”. A maioria dos alunos também destacou a clareza nos enunciados das atividades e a importância das

dicas oferecidas para exemplificar o que se pretendia com a atividade. Porém, alguns alunos discordaram e afirmaram ter problemas com a interpretação de algumas atividades, como ilustram as citações: “Algumas questões eram claras, mas as mais difíceis não eram claras” e “O objetivo da questão era nos confundir, e o objetivo foi realizado com sucesso”. Ainda com respeito às atividades, os alunos perceberam que ao avançar no curso, as atividades ficavam mais complexas como pode se observar na citação: “... apesar das aulas terem sido ministradas com clareza, achei as questões do lab em um nível de dificuldade muito acima do nível da aula” e “... foram ficando cada vez complexos...”.

Tabela 5 – Categorias que emergem da análise inter-participantes do MEDS.

#	Categoria	Ocorrências
1	Coerência das atividades com conteúdo ministrado	24
2	Clareza nos enunciados das atividades	70
3	Dicas nos enunciados das atividades	7
4	Complexidade das atividades	33
5	Qualidade do material de apoio	12
6	Disponibilidade dos sistemas e material de apoio	11
7	Quantidade de exemplos de teste	9

Sobre a qualidade do material de apoio (categoria 5) disponibilizado aos alunos, foi relatado que eram satisfatórios, de linguagem de fácil entendimento e que realmente ajudavam na realização das atividades como relatado na citação: “Havia tipo um "tutorial" explicando detalhadamente questões semelhantes”. A disponibilidade tanto do material de apoio quanto das ferramentas de suporte na disciplina (categoria 6) foi outro ponto de destaque nos relatos dos alunos, que afirmavam: “Porque foi possível acessar os arquivos de apoio no AVA” e “porque podia consultar a apostila e o Spyder”.

Com respeito às atividades de codificação, os alunos destacaram a quantidade de exemplos para testes dos scripts propostos (categoria 7). Alguns alunos relataram dificuldades no entendimento dos enunciados e apontaram a falta de exemplos para os testes, como pode-se observar na citação: “Achei que faltavam mais exemplos para testar o programa”. Outros não relataram a mesma dificuldade, como observa-se na citação: “Por que eram bem explicativos e no material disponibilizado no AVA continha muitos exemplos”.

6. Considerações finais

Aqui foi apresentado um relato de experiência da aplicação de uma metodologia de ensino-aprendizagem em turmas de Introdução à Programação de Computadores. Os resultados preliminares indicaram um índice de aprovação maior que em períodos anteriores. Além disso, as questões interdisciplinares elaboradas se mostraram efetivas tanto pelo seu poder discriminatório quanto pela avaliação dos próprios discentes.

Como trabalhos futuros, desejamos fazer uso de Inventários de Conceito [Caceffo et al. 2016] para aferir o ganho de aprendizagem antes e depois da disciplina. Além disso realizaremos o cruzamento mais detalhado dos dados deste estudo com os de disciplinas concomitantes, a fim de diferenciar entre alunos que se afastam informalmente de todas as disciplinas do curso daqueles que se afastaram apenas da disciplina de IPC.

Referências

- Araujo, E. C.; Gaudencio, M.; Menezes, A.; Ferreira, I.; Ribeiro, I.; Fagner, A.; Ponciano, L.; Morais, F.; Guerrero, D. S.; Figueiredo, J. A. (2013). O papel do hábito de estudo no desempenho do aluno de programação. In *Workshop sobre Educação em Computação (WEI 2013)*. Maceió.
- Caceffo, R.; Wolfman, S.; Booth, K. S.; Azevedo, R. (2016). Developing a Computer Science Concept Inventory for Introductory Programming. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16)*. ACM, New York, NY, USA, 364-369.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Relatório Síntese da Área de Ciência da Computação – Enade (2014). Disponível em <http://portal.inep.gov.br/enade/relatorio-sintese-2014>
- Feldman, T. J.; Zelenski, J. D. (1996). The quest for excellence in designing CS1/CS2 assignments. In *ACM SIGCSE Bulletin* (Vol. 28, No. 1, pp. 319-323). ACM.
- França, A. B.; Soares, J. M.; Gomes, D. G.; Barroso, G. C. (2011). Um sistema orientado a serviços para suporte a atividades de laboratório em disciplinas de técnicas de programação com integração ao ambiente Moodle. *Revista Novas Tecnologias na Educação – RENOTE*, 9(1).
- Hundley, J.; Britt, W. (2009). Engaging students in software development course projects. In *Proceedings of the Fifth Richard Tapia Celebration of Diversity in Computing Conference: Intellect, Initiatives, Insight, and Innovations* (pp. 87-92). ACM.
- Ihantola, P.; Ahoniemi, T.; Karavirta, V.; Seppälä, O. (2010). Review of recent systems for automatic assessment of programming assignments. In *Proceedings of the 10th Koli Calling International Conference on Computing Education Research* (pp. 86-93).
- Menezes, N. N. C. (2014). *Introdução à Programação com Python*. Editora Novatec, 2ª edição.
- Nicolaci-da-Costa, A. M. (2007). O Campo da Pesquisa Qualitativa e o Método da Explicitação do Discurso Subjacente (MEDS). In *Psicologia: Reflexão e Crítica*. vol.20 no.1. ISSN: 0102-7972. RS, Porto Alegre.
- Píccolo, H. L.; Sena, V. F.; Nogueira, K. B.; Silva, M. O.; Maia; Y. A. N. (2010). Ambiente Interativo e Adaptável para ensino de Programação. In *Workshop sobre Educação em Computação*, pp. 555–566.
- Pieterse, V. (2013). Automated assessment of programming assignments. In *Proceedings of the 3rd Computer Science Education Research Conference* (pp. 45-56). Open Universiteit, Heerlen.
- Stevenson, D. E.; Wagner, P. J. (2006). Developing real-world programming assignments for CS1. In *ACM SIGCSE Bulletin* (Vol. 38, No. 3, pp. 158-162).
- Zanini, A. S.; Raabe, A. L. A. (2012). Análise dos enunciados utilizados nos problemas de programação. In *Workshop sobre Educação em Computação*, pp. 555–566. Curitiba, PR.

Estímulo à prática multidisciplinar no ensino de Computação e Design através de um evento de programação focado em problemas

**Daniela Brauner¹, Pablo Margreff², Tatiana Tavares²,
Vinicius Kruger da Costa³, Alexandre L. Silva⁴**

¹Escola de Administração – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil.

²CDTec - Centro de Desenvolvimento Tecnológico - Universidade Federal de Pelotas
(UFPel)
Pelotas - RS - Brasil.

³Coordenadoria de Design - Instituto Federal de Educação, Ciência e Tecnologia
Sul-rio-grandense (IFSul)
Pelotas - RS - Brasil.

⁴Conrad Caine - Possible
Pelotas - RS - Brasil.

¹daniela.brauner@ufrgs.br, ²{pmargreff,tati...}@inf.ufpel.edu.br,
³viniciusdacosta@pelotas.ifsul.edu.br, ⁴alexandre.silva@possible.com

Abstract. *Hackathons are programming events focused on problem solving through a collaborative and competitive dynamics. The informal nature of hackathons make them excellent platforms for learning, able to attract diverse audience, especially those who would like to try something different from the traditional model of the classroom. This paper discusses an experience of a hackathon with students from Computing and Design. The experiment was held in an IT company and it has university and industry mentors. Despite the competitive nature, the hackathon provide an important stimulus tool for learning, collaborative and multidisciplinary work.*

Resumo. *Hackathons são eventos de programação focados na resolução de problemas através de uma dinâmica colaborativa e competitiva. A natureza informal dos hackathons o tornam excelentes plataformas de aprendizagem, capazes de atrair público diversificado, especialmente, àqueles que gostariam de experimentar algo diferente do dia-a-dia de sala de aula. Neste artigo é discutido um relato de experiência de um hackathon com estudantes de Computação e Design. A experiência realizada dentro de uma empresa contou com mentores da universidade e indústria. Apesar da natureza competitiva, o hackathon se mostrou uma importante ferramenta de estímulo ao aprendizado e ao trabalho colaborativo e realmente multidisciplinar.*

1. Introdução

Segundo Heikki Topi e Allen Tucker (2014) um *hackathon* é um evento de programação em Computação focado em problemas. Briscoe e Mulligan (2014) acrescentam que *hackathons* são instrumentos importantes para encorajar a experimentação e criatividade, aspectos essenciais à inovação. Dessa forma, o *hackathon* pode ser visto como uma ação conjunta entre programadores e outros *stakeholders*¹ que atuam colaborativamente no desenvolvimento de software engajados de forma intensa e desafiadora durante um curto período de tempo.

Com o intuito de propor esse desafio para a comunidade da parte sul do Rio Grande do Sul, em novembro de 2015 a Universidade Federal de Pelotas (UFPEL) e o Instituto Federal Sul-Riograndense (IFSul) organizaram e executaram o CodeArena. O CodeArena oportunizou uma experiência multidisciplinar voltada para problemas práticos através de um evento no estilo *hackathon*. O objetivo deste evento foi oferecer para alunos dos cursos de Computação e de Design, a interação entre áreas, que é pouco abordada durante a graduação. Considerando que atualmente a interdisciplinaridade é praticamente um requisito para o trabalho em TI, o CodeArena ofereceu essa oportunidade para a troca de experiência entre alunos de diferentes instituições de ensino, fazendo com que a troca bilateral de diferentes olhares contribuísse para motivar os alunos de ambas as instituições envolvidas.

Este artigo descreve a concepção e execução do hackathon CodeArena no intuito de motivar a organização de práticas multidisciplinares como estímulo ao ensino e aprendizado em Computação e Design. Para tanto, primeiramente, é apresentada uma visão geral da dinâmica dos *hackathons* destacando os principais pontos para organizar um evento com essas características. E, então, é discutida a abordagem utilizada no hackathon CodeArena através do detalhamento das atividades realizadas. Por fim, são abordadas as lições aprendidas com a realização de um evento deste tipo, bem como as vantagens para a universidade e as empresas, nas diferentes perspectivas de envolvimento.

2. A Dinâmica dos *Hackathons*

A origem da palavra *hackathon* é a combinação de *hack* e *marathon*. A palavra "*hack*" remete ao senso de investigação e exploração inerente à atividade de programação. Já a palavra "*marathon*" está relacionada a modalidade esportiva maratona que é uma corrida ininterrupta de longa distância que geralmente ocorre nas ruas, desafio, exaustão e competição são aspectos intrínsecos da maratona. Dessa forma, um *hackathon* envolve a investigação de soluções que envolvam Computação até a exaustão através de uma competição.

Briscoe e Mulligan (2014) destacam que *hackathon* é uma atividade que pode possuir muitas versões diferentes e, como tal, não se limita a uma linha particular. No entanto, existem alguns elementos que se destacam como "universalmente necessários", como é o caso da colaboração. Embora seja difícil estabelecer um conjunto de melhores

¹ Stakeholder é o termo utilizado para qualquer pessoa ou entidade envolvida com o processo de software, por exemplo, *designers*, testadores, cliente.

práticas, os autores abordam princípios gerais que podem/devem ser observados: (i) Hackathons são orientados a produção de saídas; (ii) Hackathons devem ser inclusivos; (iii) Hackathons devem oportunizar o aprendizado e o compartilhamento de conhecimentos; (iv) Hackathons devem incentivar o aprendizado através da tentativa-erro; (v) Hackathons necessitam de conectividade; (vi) Hackathons são beneficiados por um local inspirador, acolhedor e confortável; (vii) Hackathons são atividades colaborativas e em grupos. Líderes dos grupos devem ser incentivados, mas não devem ser obrigatórios.

A Fig.1 apresenta uma visão da dinâmica dos hackathons baseada na proposta da Clever Together² para uma ação chamada #ChangeChallenge. Nessa visão observados os pontos essenciais para a realização de um hackathon: (i) o entendimento do problema através da compreensão do seu contexto através da delimitação das suas fronteiras; (ii) a pesquisa de soluções relacionadas que possam agregar na proposta a ser desenvolvida; (iii) o estabelecimento de prioridades para identificar atividades e entregáveis o que compõe o planejamento do grupo; (iv) a etapa de "hands-on", isto é, a realização das atividades planejadas para o desenvolvimento da solução.

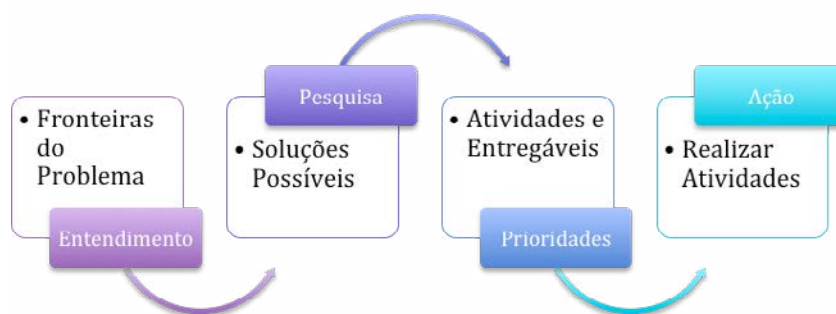


Figura 1. Visão geral da metodologia estilo hackathon.

De forma complementar a dinâmica apresentada por Briscoe e Mulligan (2014) destaca que normalmente os *hackathons* começam com uma ou mais apresentações sobre o evento, incluindo a apresentação do formato geral do evento, seus desafios e os prêmios, se disponível. O desafio é parte primordial da dinâmica do *hackathon*, na qual esta decisão pode ser feita em conjunto com os participantes ou de forma orientada pela organização.

A formação das equipes pode considerar interesses individuais e habilidades. Uma vez formada a equipe, é a hora da "largada" quando o trabalho principal do *hackathon* começa, que pode durar de algumas horas a vários dias. Nesse caso, se alimentar e dormir são atividades que ocorrem geralmente no local do evento. Ao final, cada grupo apresenta seus resultados que muitas vezes envolvem protótipos de software. Nesse momento, uma banca de juízes seleciona as equipes vencedoras e, então, os prêmios são entregues. Em muitos *hackathons*, os juízes são os próprios organizadores, bem como os patrocinadores do evento.

Segundo Fassbinder et al. (2012) o desenvolvimento de competições de conhecimento é uma forma de estimular os alunos a aprimorarem seus conhecimentos

² <http://clevertogether.com/public/healthcare/change-challenge-methodology/>

em determinada área. Os autores acrescentam que ao ingressarem numa disputa os participantes se vêem engajados na missão de provar o seu valor e o seu conhecimento. Além disso, guiados pelo instinto competitivo, não medem esforços para buscar ainda mais conhecimento e transpor qualquer barreira que surja em seu caminho. Adicionalmente, Nandi e Mandernach (2016) destacam a importância do aspecto informal de aprendizado inerente aos hackathons. Dessa forma, a dinâmica dos *hackathons* pode estimular a experiência multidisciplinar e colaborativa para a resolução de problemas, especialmente, envolvendo a programação de soluções. A competição e a imersão inerentes ao formato do evento impulsionam a busca por soluções possíveis para os desafios propostos.

3. Atividades Realizadas

A primeira etapa foi a **chamada de ideias**, para tanto a organização definiu um período de um mês para que qualquer pessoa, sendo parte ou não da comunidade acadêmica, submetesse um problema e sua respectiva sugestão de solução através do site³ do evento como ilustrado na Fig.2 (a). No total foram submetidas 19 problemas, destes foram escolhidos 6 para o desenvolvimento durante o evento, por uma comissão formada por pesquisadores da universidade, através de uma análise que contemplou os seguintes critérios: originalidade, viabilidade do produto, viabilidade técnica e impacto social.

A **formação das equipes (times)** foi realizada em duas etapas online: cadastramento dos participantes e candidatura do participantes para as ideias selecionadas conforme ilustrado na Fig.2 (b). Cada time deveria agrupar 3 participantes, sendo 2 (duas) vagas para a área de Computação e 1 (uma) para a área de Design. Nesta primeira edição do CodeArena foram oferecidas 24 vagas no total, sendo preenchidas, destas, 19 vagas. Quanto à formação, 5 alunos eram do Design e 14 da Computação sendo, quanto ao gênero, apenas 3 pessoas do sexo feminino e 16 do sexo masculino.

Um resultado que já podemos destacar foi o envolvimento de diferentes cursos e instituições nesta etapa de formação das equipes. Dentre os cursos da área da Computação foram envolvidos os cursos de: Bacharelado em Ciência da Computação (UFPel), Bacharelado em Engenharia de Computação (UFPel), Pós Graduação em Computação (UFPel), Técnico em Eletrônica (IFSul), Licenciatura para Computação (IFSul) e Tecnólogo em Sistemas para Internet (IFSul). Já para os cursos da área de Design destacamos: Bacharelado em Design (IFSul), Bacharelado em Design Gráfico (UFPel) e Bacharelado em Design Digital (UFPel). Dessa forma, observamos o engajamento de alunos de nove cursos: 6 da área de Computação e 3 de Design.

Uma vez formadas as equipes pela página Web foi dado o início das atividades efetivas do *hackathon*. Para proporcionar o primeiro encontro dos times realizamos uma reunião na noite anterior que antecedeu o início do *hackathon* propriamente dito. Nessa **reunião de aquecimento** as equipes tiveram o primeiro contato e puderam interagir com os membros da organização. Também foram detalhadas as propostas para todos os times permitindo que todos soubessem as ideias e também para possibilitar a troca de participantes entre os times, caso houvessem interessados. Na mesma ocasião, as equipes tiveram a oportunidade de alinharem as funções dos membros dos times e

³ <http://computacao.cc/codearena/>

projetar um MVP (*Minimal Viable Product*) com as funcionalidades básicas que cada projeto precisaria contemplar. Esse também foi o primeiro contato com os mentores que puderam auxiliar os grupos a definir quais seriam as prioridades para o produto.

Sugestão de problema/tema

Faça sua proposta de problema/solução.

*Obrigatório

Descrição *
Descreva abaixo o funcionamento da ferramenta proposta.
Sua resposta

Motivo *
Descreva por que/quando você viu a necessidade de uma solução para esse problema.
Sua resposta

Prioridade *
Caso sua ideia seja selecionada, deseja fazer parte da equipe que irá desenvolver a ideia?

Sim
 Não

Tarefa 1 - Computação

Proposta: Criação de um serviço web que agregaria informações dos sites de imobiliárias da cidade (ou de cidades escolhidas) e mostraria informações como histórico de preços, mapa com preços médios por bairro/região e outras informações relevantes.

*Obrigatório

Nome *
Digite seu nome completo:

Termo de aceite *
Para efetivar o cadastro por favor, aceite um dos termos abaixo.

Eu por meio deste atesto que sou aluno devidamente matriculado da Universidade Federal de Pelotas, nos cursos de Ciência da Computação ou Engenharia de Computação.

Eu por meio deste atesto que sou aluno devidamente matriculado do IF-SUL dos cursos de Tecnologia em Sistemas para Internet ou Engenharia Elétrica ou ainda do curso Técnico Integrado em Eletrônica.

RG *
Digite seu RG:

E-mail *
Digite seu e-mail:

(a)

(b)

Figura 2. Em (a) screenshot da página de submissão de ideias e em (b) screenshots da página de candidaturas dos participantes do CodeArena.

O *hackathon* CodeArena iniciou efetivamente suas atividades na empresa Conrad-Caine no sábado às 08:30 com a etapa de “*hands-on*” onde os participantes deram continuidade ao desenvolvimento dos projetos esboçados na noite anterior. Depois de ter um *sketch* dos projetos em mãos e de suas funcionalidades básicas, os participantes deram início ao ciclo de 36 horas de desenvolvimento dos projetos. Um observação interessante foi a integração das equipes, onde cada uma ficou livre para definir o seu processo de desenvolvimento. A Fig. 3 mostra uma das estratégias adotadas por uma das equipes que descreveu sua dinâmica através de uma quadro de *kanban*. As responsabilidades foram identificadas pelo perfil: design e computação (através de backend e frontend) conforme Fig.3 (a) e as atividades detalhadas em post-its (vide Fig.3(b)), bem como seu andamento na Fig.3 (c). Pela imagem é possível notar que o grupo implementou, através de cores, a separação dos papéis dos integrantes dos times.

A dinâmica dessas 36 horas de maratona, além das atividades inerentes a programação e ao design dos projetos, envolveu também atividades de entretenimento, bem como espaços para alimentação e integração entre as equipes. Jogos, desafios (vide Fig.4 (a)), instrumentos musicais, pizza, lanche, bombons, salada de frutas, águas e café ficaram constantemente a disposição dos participantes. Além disso, os participantes poderiam levar colchonetes e uma sala foi disponibilizada para o seu descanso, caso necessário. Os participantes também podiam deixar o local de evento para descansar no momento em que desejassem e combinassem com sua respectiva equipe. Também ao longo do final de semana foram feitas diversas atividades, relacionadas ou não com as

áreas envolvidas no desenvolvimento do produto, para que os participantes tivessem a oportunidade de se integrarem com os outros grupos e ter um momento mais descontraído durante o evento, além de render alguns brindes aos seus vencedores.

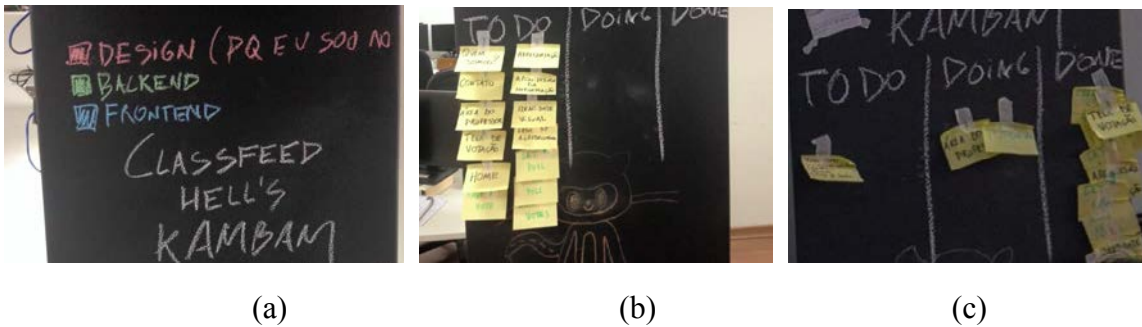


Figura 3. Quadro Kanban desenvolvido por uma das equipes. Em (a) podemos observar a divisão integrada das responsabilidades (design e computação), em (b) o detalhamento das atividades e em (c) o andamento.

As **apresentações dos produtos** foram divididas em duas partes, a primeira foi o *pitch* onde cada equipe teve 5 minutos para apresentar sua solução para os jurados e as outras equipes conforme ilustrado na Fig.4 (b). A segunda parte foi o teste dos protótipos dos projetos conforme Fig.4. (c). Durante os testes os mentores visitaram todas as equipes e utilizaram os protótipos, tiraram eventuais dúvidas e discutiram possíveis melhorias a serem implementadas em cada projeto. Como resultados tivemos quatro aplicativos em fase de protótipos funcionais e um em fase de ideia e concepção.

Por fim, a **etapa de premiação** quando todos os participantes foram envolvidos. Após a deliberação da comissão julgadora que contou com a presença de representantes da empresa anfitriã do evento, a equipe vencedora foi premiada. Os vencedores receberam um micro-estágio na empresa sede do evento, que será realizado de forma complementar as atividades acadêmicas. Os alunos poderão atuar em atividades relacionadas aos seus interesses acadêmicos (*front-end*, *back-end*, *design* ou projetos), relacionando ou não aos trabalhos conduzidos na Universidade. Os demais envolvidos, incluindo a organização do evento e representantes dos patrocinadores, receberam medalhas de participação conforme pode ser visto na foto da Fig.5.



Figura 4. Desafio de programação realizado (a), apresentação dos produtos, em (b) pitches de 5 minutos e em (c) demonstrações.



Figura 5. Premiação.

5. Resultados Pós-Evento

Como etapa pós-evento foi lançado um questionário de participação voluntária para que pudéssemos avaliar a recepção do evento pela perspectiva dos participantes. Questões a respeito da apreciação geral, organização e comida/bebida obtiveram 100% de respostas "excelente". A avaliação também possibilitou analisar a premissa deste artigo: hackathons como ferramentas de estímulo a aprendizagem. As figuras Fig.6 (a) e (b) destacam o estágio do curso e a opinião dos participantes quanto ao objetivo e formato do evento, respectivamente. Observamos que a maioria dos participantes está no meio do curso e todos classificaram como Bom (12,5%) e Excelente (87,5%) o formato do evento. Fortalecendo essa tendência dos dados obtidos, destacamos o seguinte depoimento de um dos participantes: *"Aprendi muito nesse evento graças aos colegas e professores. Desde o básico a coisas avançadas do design! Também vi o que funciona ou não na hora da programação. Essas 36 horas me ensinaram coisas que eu demoraria muito mais tempo pra aprender. E achei maravilhoso ter medalha de prata do último ao segundo lugar. Não tenho nada de ruim pra falar sobre o evento, me diverti e aprendi muito! Só estou esperando o próximo evento :)"*.

Também chamou atenção nos resultados obtidos neste questionário a impressão dos participantes do local e da duração do evento. Quanto ao local a grande maioria classificou como Excelente (87,5%) e 12,5% como Bom (vide Fig.6 (d)). A imersão em um ambiente empresarial foi muito importante como ferramenta de motivação. Já a avaliação do tempo de duração do evento demonstrou que a maioria aprovou as 36 horas (87,5%) conforme Fig.6 (c).

Além disso, percebeu-se como resultado do evento o interesse dos acadêmicos na aplicação dessa experiência de integração entre a Computação e o Design em períodos maiores e presente no calendário das instituições. Dessa forma, conseguimos ofertar aos alunos de Engenharia da Computação, Ciências da Computação, Design Gráfico e Design Digital da UFPel uma disciplina em conjunto, com 32 vagas divididas

entre os 4 cursos, que ocorrerá durante o primeiro semestre de 2016. Com essa iniciativa, as experiências proporcionadas pelo evento poderão ser contempladas por alunos matriculados nos respectivos cursos e que não tiveram, por algum motivo, a oportunidade em participar do evento. A metodologia da disciplina seguirá os mesmos padrões do CodeArena, exceto pelo tempo de desenvolvimento, e será ministrada por dois dos professores organizadores do evento.

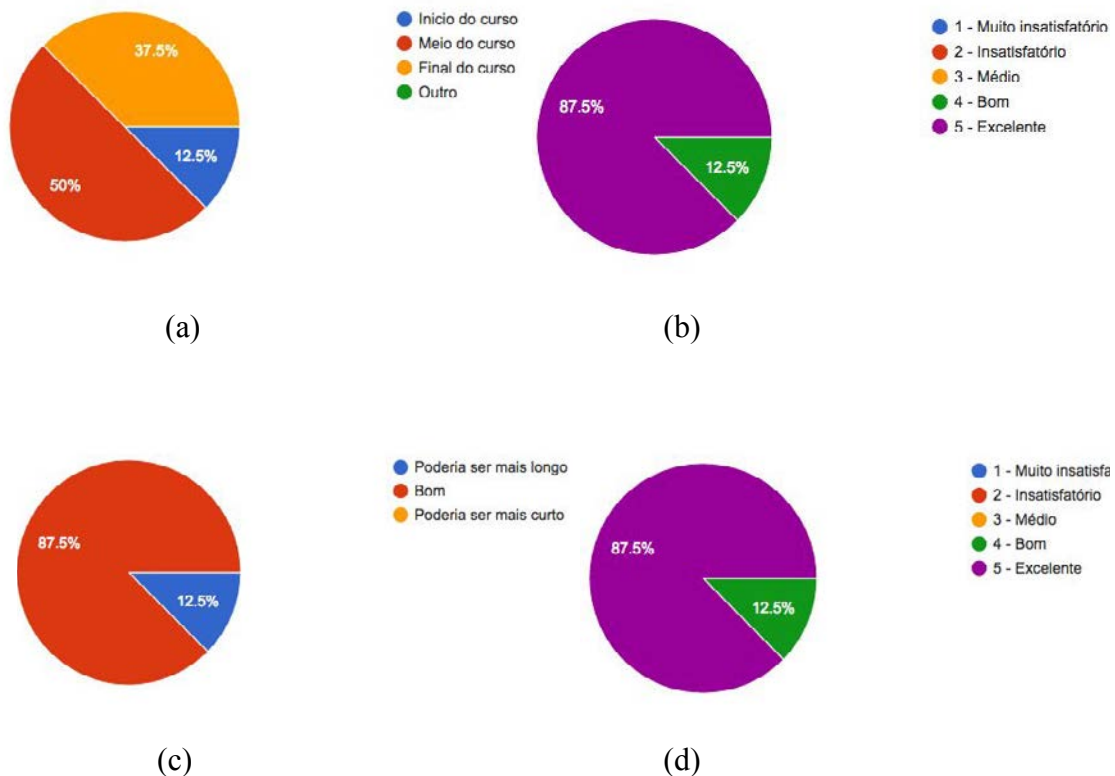


Figura 6. Algumas respostas do formulário de avaliação do evento.

6. Conclusões e Trabalhos Futuros

Além dos resultados tecnológicos e de caráter empreendedor, a realização do evento trouxe ganhos intangíveis para os envolvidos. Os ganhos serão descritos nos próximos parágrafos sob as perspectivas da universidade e das empresas envolvidas, relatando as vantagens percebidas pelos envolvidos de ambas perspectivas.

Da perspectiva da universidade, foram inúmeros os aspectos observados como vantajosos na realização de um evento deste tipo, segundo opiniões coletadas em conversas informais com os professores e alunos que participaram do evento. Do ponto de vista dos alunos, tivemos os próprios integrantes dos times, que participaram ativamente na concepção dos protótipos resultantes. Os participantes gostaram do formato do evento e do desafio de desenvolverem, em curto espaço de tempo, projetos aplicados para resolver algum problema real da sociedade. Relataram que tais desafios não são comuns nos trabalhos acadêmicos e que os conhecimentos necessários para desenvolvê-los vem para complementar os ensinamentos teóricos vividos em sala de aula. Estes mesmos alunos, relataram que ficaram muito empolgados por estarem dentro do ambiente da empresa. Segundo eles, o ambiente proporcionou uma sensação de

motivação e vivência de um dia como um profissional. Aliado ao desafio dado pelo curto prazo do evento, que simulava uma demanda real do mercado de trabalho, eles se sentiram como se tivessem uma entrega de uma *release* do software para avaliação do cliente.

Outro fator importante, citado pelos participantes, foi a presença de mentores. Segundo eles, os mentores tiraram dúvidas e auxiliaram em questões práticas que permitiram aos times ter mais agilidade e qualidade na entrega dos protótipos. Como mentores, participaram alunos, integrantes da empresa Júnior da UFPel (Hut8), profissionais das empresas hospedeira e patrocinadoras, e professores da UFPel e IFSul, dos cursos de Computação e Design. Os mentores relataram que a experiência proporcionou a aproximação com os alunos e a troca de experiências além de apoiar os participantes com dicas de programação, design e apoio técnico nas ferramentas e técnicas utilizadas pelos times.

Para os mentores, em especial os professores, o evento proporcionou conhecer melhor as habilidades de alguns alunos, como por exemplo identificar talentos em frameworks, linguagens de programação e tecnologias. O evento proporcionou mapear competências para futuras bolsas ou projetos de seus laboratórios. Além é claro da integração entre alunos e professores. Os alunos da empresa júnior também aproveitaram o momento para conhecer novos talentos e fazer networking com profissionais da empresa que estavam no evento.

Essa integração foi auxiliada por mentores com vivência e domínio desde a parte de negócios, passando pela experiência de usuário e também com pleno domínio em desenvolvimento. Ainda para tentar dar maior similaridade ao dia-a-dia do mercado de trabalho, além de afastar um pouco do ambiente acadêmico com o qual os alunos estão acostumados e confortáveis, toda essa experiência aconteceu dentro de uma empresa alemã que possui uma de suas filiais na cidade de Pelotas. Também foi estimulado o empreendedorismo com foco na resolução de problemas locais e até mesmo outros mais amplos que com a globalização ultrapassam barreiras territoriais e atingem uma parte significativa da sociedade.

Sob a perspectiva da empresa sede do evento, as vantagens foram inúmeras. Além de permitir que alunos e professores conhecessem suas instalações, o empresário e sua equipe puderam se aproximar dos alunos e professores de forma a mapear talentos para futura captação para seus projetos. Como a empresa sede foi uma empresa de design e marketing digital, eles puderam sugerir aos alunos tanto de computação quanto de design, algumas técnicas e ferramentas utilizadas na empresa, através dos mentores disponibilizados por eles. Os patrocinadores, puderam enviar mentores e participar de todo o evento. Nesta edição, nenhum dos patrocinadores, exceto a empresa sede, pôde enviar mentores para participar na data do evento, portanto não foi possível coletar a percepção destas empresas com relação aos ganhos com a mentoria no evento. Mas em conversas com os empresários, foi possível perceber, que de posse dos dados sobre os participantes, eles farão aproximações com o intuito de identificação de talentos, além de aproximação entre as próprias empresas envolvidas, visando futuros acordos comerciais. Ainda na perspectiva da universidade, uma vantagem importante foi a aproximação entre os cursos de design e computação. Tal demanda, inicialmente identificada pelos alunos de design e computação, através de seus diretórios

acadêmicos, foi a grande motivadora da criação deste evento com esse formato integrador. A aproximação dos dois cursos permitiu que os grupos fossem formados de forma interdisciplinar, similar a formação dos times nas empresas. Tal interdisciplinaridade, trouxe a complementaridade na condução dos projetos de desenvolvimento dos aplicativos. Enquanto uns programavam front e backend, o designer atuava na concepção do layout e telas da aplicação, inclusive, em alguns times, desenhando telas que não seriam implementadas a tempo do evento, mas que poderiam contribuir para a apresentação de funcionalidades de futuras releases. A integração dos times proporcionou uma troca, em tempo real, de demandas de mudanças que impactaram tanto o design quanto a implementação de front e backend, proporcionando que os integrantes dos times tivessem que agir de forma rápida às mudanças. O uso de técnicas de gerenciamento ágil de projetos, proporcionou aos grupos que aplicaram, um rápido planejamento e visualização das atividades necessárias e o andamento do desenvolvimento. O grupo responsável pelo desenvolvimento do projeto ClassFeed usou ciclos rápidos de desenvolvimento e teste do aplicativo, através da aplicação de Scrum. A implementação do Kanban aliado às cores, segundo eles, trouxe agilidade e garantiu a produtividade do grupo no tempo previsto do evento.

Como trabalhos futuros, já estamos trabalhando na concepção do Code Arena de 2016. Trabalharemos com uma metodologia similar, aplicando as lições aprendidas e melhorando a interação com todos os envolvidos.

Referências

- Heikki Topi and Allen Tucker. Computing Handbook, Third Edition: Information Systems and Information Technology. CRC Press, 2014.
- Briscoe, Gerard, and Catherine Mulligan. "Digital innovation: The hackathon phenomenon." London: Creativeworks London Work Paper 6 (2014). (Disponível em: <http://www.creativeworkslondon.org.uk/wp-content/uploads/2013/11/Digital-Innovation-The-Hackathon-Phenomenon1.pdf>)
- Fassbinder, Aracele Garcia de O., Lílian Cristina de Paula, e João Cláudio D. Araújo. "Experiências no estímulo à prática de Programação através do desenvolvimento de atividades extracurriculares relacionadas com as competições de conhecimentos". *In*: Anais do Workshop sobre Educação em Computação (WEI), Congresso da Sociedade Brasileira de Computação (CSBC). 2012.
- de Barros, Louise G., and Janne Yukiko Y. Oeiras. "Extensão universitária em Computação por meio da Olimpíada Brasileira de Informática." *In*: Anais do Workshop sobre Educação em Computação (WEI), Congresso da Sociedade Brasileira de Computação (CSBC). (2010) Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/wei/2010/009.pdf>
- Arnab Nandi and Meris Mandernach. 2016. Hackathons as an Informal Learning Platform. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education (SIGCSE '16). ACM, New York, NY, USA, 346-351. DOI=<http://dx.doi.org/10.1145/2839509.2844590>

Experiências de Criação de Aplicativos Móveis com Alunos do Ensino Fundamental e Médio

Ana Luísa V. Solórzano¹, Andrea S. Charão²
Patrícia P. Barcelos²

¹ Curso de Ciência da Computação

²Departamento de Linguagens e Sistemas de Computação
Universidade Federal de Santa Maria
Santa Maria, RS, Brasil

{alsolorzano, andrea, pitthan}@inf.ufsm.br

Resumo. *A disseminação de fundamentos de programação a crianças e jovens é alvo de diversas iniciativas atualmente, inspiradas em seus potenciais benefícios para esta geração. Neste trabalho, apresenta-se as experiências de um projeto de extensão universitária que incentivou alunos de Ensino Fundamental e Médio a produzirem aplicativos móveis. No projeto, utilizou-se a ferramenta MIT App Inventor como introdução à programação, através de encontros semanais para a produção de um aplicativo de perguntas e respostas. A atividade colocou em evidência obstáculos a serem resolvidos, como maneiras de melhor adequar a didática utilizada e de manter os encontros atrativos. De forma geral, as experiências mostraram-se estimulantes tanto para o público-alvo como para os graduandos que se envolveram no projeto.*

Abstract. *Disseminating programming fundamentals to children and youth is the goal of several initiatives inspired by its potential benefits for this generation. In this work, we present the experience of a university extension project that encouraged students of primary and secondary education to create mobile applications. In the project, we used the MIT App Inventor tool as a programming introduction, through weekly meetings to produce a questions and answers mobile application. The activity has highlighted obstacles to be overcome, such as ways to better tailor the didactic used and to keep the meetings attractive. Overall, the experience proved to be stimulating both for the audience and for undergraduate students who were involved in the project.*

1. Introdução

Atualmente, pode-se observar uma forte dependência de jovens com tecnologias móveis [Rideout et al. 2010]. Por serem de fácil acesso e simples uso, os aplicativos (*apps*) para dispositivos móveis atingiram grande popularidade, intermediando tarefas úteis em áreas variadas como segurança, educação e lazer. Dados publicados em 2015 apontam que, no mundo todo, o número de “viciados em usar aplicativos”, isto é, pessoas que os utilizam pelo menos 60 vezes ao dia, é maior do que a população do Brasil [Khalaf 2015].

De fato, a atual geração de jovens espera que existam aplicativos (*apps*) para tudo e muitas vezes recusa formas menos diretas e eficientes de obterem o que desejam, sendo

considerados por estudiosos como a “Geração App” [Gardner e Davis 2013]. Por outro lado, os benefícios dos *apps* são relevantes, pois eles podem promover um senso de identidade, manter amizades e estimular a criatividade [Gardner e Davis 2013].

Na área de educação, há algum tempo vislumbra-se que aplicativos móveis podem ser aliados para estimular o aprendizado [Hulme e Traxler 2005]. Mais do que isso, na educação em Computação, acredita-se que a **criação** de aplicativos móveis possa ser uma forma de incentivar as novas gerações a desenvolver o pensamento computacional [Wolber 2011, Fronza et al. 2015].

Assim, no contexto de um projeto de extensão universitária, desenvolveu-se uma experiência de criação de aplicativos com alunos de Ensino Fundamental e Médio, despertando-os para a ideia de que eles estão aptos a não somente utilizar, mas também a produzir essas tecnologias. Para isso, utilizou-se a ferramenta MIT App Inventor, que propõe uma abordagem lúdica e auto-explicativa, usando programação visual com blocos que se encaixam. Apesar da plataforma apresentar apenas recursos para criação de aplicativos simples, estima-se que ela também sirva de incentivo para futuros programadores, pois pode despertar a busca por funcionalidades mais avançadas fazendo uso de outras linguagens de programação.

Neste artigo, relata-se a experiência de um ano de exploração do MIT App Inventor no cenário delineado acima. A Seção 2 descreve a ferramenta trabalhada e suas potencialidades. A Seção 3 apresenta trabalhos que também utilizam esta ferramenta, mas com algumas diferenças de abordagem. A Seção 4 apresenta o método de trabalho e as atividades realizadas nas experiências. A seção 5 apresenta e discute resultados de avaliação das experiências, ressaltando alguns obstáculos identificados. Por fim, a Seção 6 traz considerações finais, destacando lições aprendidas.

2. MIT App Inventor

O MIT App Inventor é um ambiente web lançado em 2010 pela empresa Google em parceria com o Instituto de Tecnologia de Massachusetts (MIT). Essa ferramenta é utilizada para criar aplicativos funcionais em dispositivos móveis de sistema operacional Android, possibilitando que pessoas sem conhecimento sobre linguagens de programação possam desenvolver aplicativos através de uma abordagem visual, com encaixe de blocos.

O ambiente de criação dos aplicativos é dividido em tela de “Design” e tela de “Blocos”, para adicionar elementos visuais e lógicos, respectivamente. Dentre os elementos visuais (Figura 1) estão desde funcionalidades simples como inserir botões e imagens até mais elaboradas como localização, sensores e chamadas para celular.

Na tela de elementos lógicos (Figura 2), estão os blocos responsáveis pelo funcionamento do aplicativo, identificados por cor e formato e separados em blocos lógicos, de controle, textuais, aritméticos, para a criação listas, dentre outros.

O App Inventor disponibiliza tutoriais para iniciantes aprenderem o básico da programação com blocos e um emulador, visto que nem todos possuem um aparelho de celular. Além disso, o usuário pode publicar aplicativos na chamada “Galeria”, facilitando o compartilhamento de suas criações e a descoberta de outras possibilidades com a ferramenta.

3. Trabalhos Relacionados

A disseminação da lógica computacional a crianças e adolescentes é alvo de iniciativas em escala internacional, como por exemplo Hour of Code (Hora do Código)¹ ou Europe Code Week². Essas iniciativas multiplicam experiências bem sucedidas, como é o caso da programação visual por meio de blocos que se encaixam, presente não só no MIT App Inventor mas também na ferramenta precursora Scratch ou na mais recente Blockly. Dado o caráter inovador e o potencial dessas ferramentas, elas também têm sido alvo de trabalhos científicos e acadêmicos.

¹hourofcode.com

²codeweek.eu

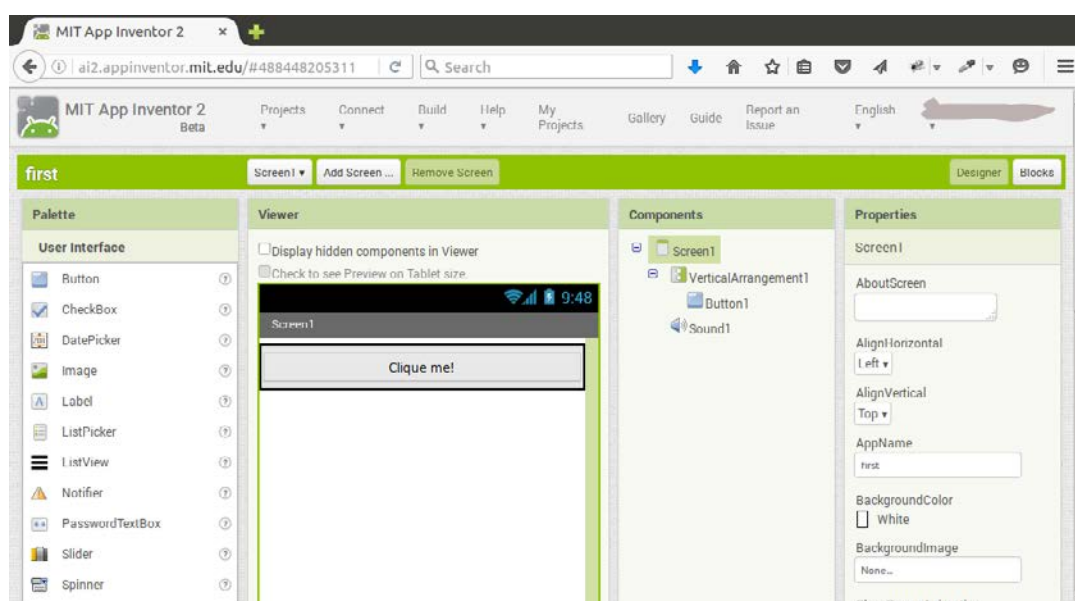


Figura 1. Tela “Design” do MIT App Inventor

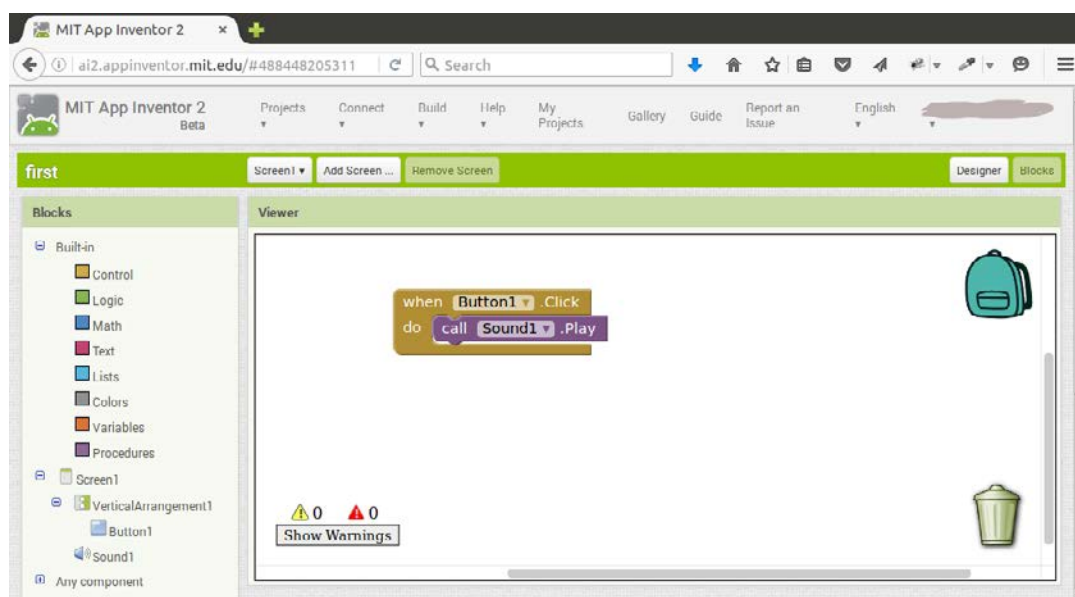


Figura 2. Tela “Blocs” do MIT App Inventor

No Brasil, encontram-se relatos de experiência de uso do App Inventor em projetos de instituições de ensino. Por exemplo, há trabalhos que relatam a utilização da ferramenta para ensino de lógica e programação no ensino médio e também a calouros no ensino superior [Gomes e de Melo 2013a, Gomes e de Melo 2013b, Finizola et al. 2014, de Cerqueira Cardoso et al. 2015]. Já em [Duda et al. 2014], a ferramenta é usada por bolsistas de cursos técnicos, para a criação de jogos digitais envolvendo tópicos em matemática. Em [Ramos et al. 2015], o App Inventor é utilizado em um projeto visando estimular alunas de ensino médio a considerarem carreiras em ciências exatas, engenharias e, especialmente, computação.

Nota-se, nos trabalhos citados, que os autores se preocupam em ressaltar aspectos positivos das experiências. Apenas no trabalho de [Ramos et al. 2015] são comentadas dificuldades, algumas relacionadas à ferramenta e outras ao contexto no qual ela se inseriu. Ao nosso entendimento, para facilitar a reprodução das experiências em outros tempos e espaços, é importante que os relatos discutam os obstáculos que, naturalmente, podem ocorrer nesse tipo de experiência. Outro aspecto observado, nos trabalhos citados, é que o público-alvo mais atendido é o de ensino médio ou superior. Em nosso trabalho, buscamos estender as experiências também ao ensino fundamental, como será discutido a seguir.

4. Método e Atividades

As experiências aqui relatadas se desenvolveram no contexto de um projeto de extensão universitária, do qual participam bolsistas e voluntários, graduandos de cursos da área de Computação da Universidade Federal de Santa Maria, orientados por professores desta instituição. A equipe do projeto possui mais de 2 anos de experiência em contatos com escolas de ensino fundamental e médio, desenvolvendo atividades variadas com diferentes ferramentas (Scratch, Python, etc.). Esse público, em geral, nunca teve contato com fundamentos de computação antes de participarem das atividades do projeto. Em 2015, buscou-se ampliar a experiência, com o planejamento e execução de atividades com o App Inventor.

Pelas experiências anteriores, notou-se que os jogos para iniciantes oferecidos pelo projeto Hora do Código são eficazes para comunicar noções de programação visual, com encaixe de blocos. Por outro lado, essa abordagem oferece limitações ao potencial criativo dos alunos, pois baseia-se em tutoriais com objetivos em níveis crescentes de dificuldade. Com o MIT App Inventor, esperava-se contornar essa limitação, com a possibilidade de os alunos criarem seus próprios aplicativos, apoiados pela equipe do projeto (tutores).

Também pelas experiências anteriores, observou-se que a duração e frequência das atividades com os alunos é um aspecto crítico nesse tipo de projeto de extensão. Por um lado, atividades intensivas/frequentes são praticamente inviáveis, pois tanto nas escolas como na universidade os envolvidos possuem outras atividades. Por outro lado, atividades muito curtas e/ou espaçadas no tempo podem se tornar improdutivas e desestimulantes.

Considerando-se todas essas observações, decidiu-se planejar uma oficina com encontros semanais, somando de 4 a 6 horas no total. Essa oficina inicia com uma atividade da Hora do Código a ser realizada no período de uma hora e depois introduz o App Inventor, com o qual cada aluno cria um aplicativo na forma de um “Quiz” (jogo de

perguntas e respostas), personalizado de acordo com seus interesses. Um tal aplicativo é apresentado num dos tutoriais do App Inventor³, servindo como referência para tutores do projeto, mas não diretamente para o público-alvo, por ser em inglês e usar uma abordagem que considerou-se “pesada” para o público das escolas.

Visto que o projeto visa executar as atividades de maneira informal e colaborativa, não apresenta-se uma introdução teórica à ferramenta, mas sim um primeiro contato através de uma visualização conjunta, com demonstrações e apresentação de slides, nas quais observam-se os principais recursos do App Inventor. Após, os participantes ficam livres para explorar a ferramenta, que por si só é auto-explicativa, pois cada recurso acompanha um botão que apresenta uma breve instrução sobre sua funcionalidade.

A atividade foi nomeada “Oficina de Criação de Aplicativos Móveis” e aplicada em duas ocasiões em 2015: com alunos de ensino médio na metade do ano e com alunos de ensino fundamental no final do ano.

4.1. Primeira experiência

Durante quatro semanas, em aproximadamente uma hora semanal, desenvolveu-se a experiência com uma turma de primeiro ano do Ensino Médio do Colégio Politécnico da UFSM, num laboratório de informática da própria escola. Para a escolha das datas e da turma mais apropriada, discutiu-se previamente com a coordenação da escola, que se mostrou interessada e disposta a ajudar. Esse apoio foi fundamental para que a oficina atingisse seus objetivos. De fato, em experiências anteriores, com menor envolvimento de professores ou coordenação das escolas, houve atividades com pouquíssimos ou nenhum participante.

De 31 alunos inscritos, 19 compareceram em pelo menos um encontro. Havia participantes de ambos os sexos. No primeiro encontro, apresentou-se o projeto, o cronograma de atividades e trabalhou-se a programação visual através da Hora do Código. A introdução ao MIT App Inventor ocorreu a partir do segundo encontro, quando lançou-se a proposta de criação de jogos personalizados de perguntas e respostas (*Quizzes*), cuja criação engloba mecanismos de raciocínio lógico similares aos mais utilizados para a programação em geral. Nesse ponto, forneceu-se um *template* (Figura 3) do jogo, com alguns blocos básicos mostrando como implementar uma pergunta/resposta no aplicativo, restando vários outros a serem inseridos ou configurados. Os encontros posteriores foram dedicados à customização e otimização dos aplicativos, e a execução desses no dispositivo móvel ou no emulador.

Na figura 3, é possível visualizar os blocos iniciais que foram disponibilizados aos alunos. Nesse *template*, estão contidos os blocos de inicialização de variáveis (número e listas) para armazenar as perguntas e respostas (*initialize*), blocos de controle para executar ações ao iniciar a tela e ao clicar o botão de responder (*when*) e bloco de procedimento (*Prox.Pergunta*, para percorrer as listas). Nota-se também um bloco condicional (*if*) aninhado, para verificar se a resposta inserida está correta.

Durante os encontros, os alunos eram acompanhados quase individualmente pelos tutores, após uma breve explanação sobre os objetivos do encontro. Ou seja: a apresentação dos recursos da ferramenta foi a mais curta possível, abrindo espaço para

³<http://appinventor.mit.edu/explore/ai2/quizme.html>

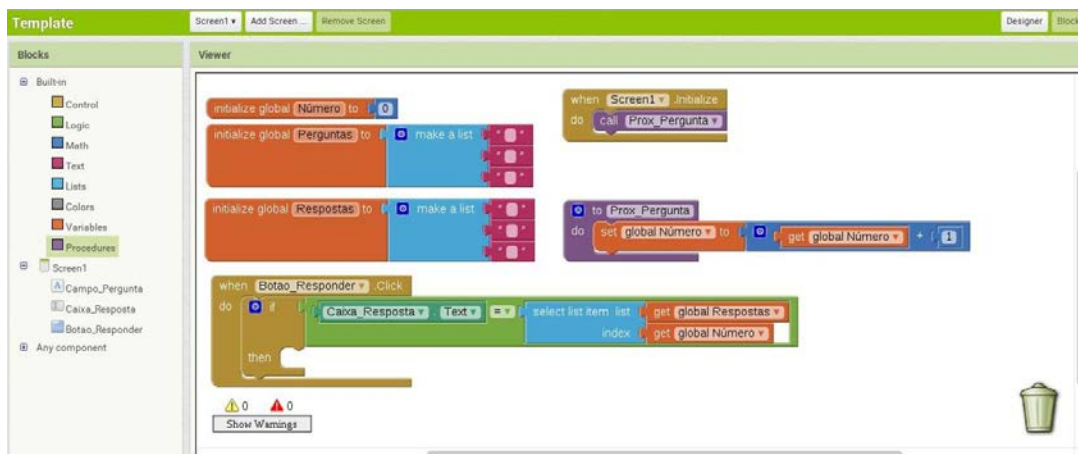


Figura 3. Tela “Blocs” com o *template* para a criação do aplicativo

que os alunos explorassem o MIT App Inventor na consecução dos objetivos propostos. Esses objetivos, por sua vez, eram moldáveis pelo interesse dos alunos. Por exemplo: o tema dos *Quizzes* poderia ser escolhido individualmente, assim como a quantidade de perguntas e o visual do aplicativo, entre outras características.

Entre os encontros, propôs-se um prêmio (uma camiseta do projeto) para os alunos que enviassem suas criações para o grupo de tutores. Dois alunos participaram, demonstrando autonomia em continuar a atividade em casa. No último encontro, com 9 alunos presentes, 6 deles enviaram seus aplicativos para a galeria do App Inventor (os demais não reservaram tempo para isso). Ao final desse encontro, todos os alunos preencheram um formulário avaliativo. A presença dos alunos no decorrer da oficina foi muito variada, portanto também enviou-se formulários àqueles que se inscreveram mas nunca compareceram e àqueles que compareceram em outros encontros, exceto no último. As questões do formulário visavam coletar impressões dos alunos sobre a oficina, identificando possíveis dificuldades e perspectivas de continuidade.

Na Figura 4, apresenta-se um exemplo de aplicativo criado por um aluno, com perguntas sobre matemática e geografia. Neste exemplo, é possível identificar os blocos utilizados e a interface do jogo no emulador.

4.2. Segunda experiência

Após novo planejamento, organizou-se outro cronograma para a segunda experiência. Dessa vez, a atividade se desenvolveu nas dependências da Universidade e foi aberta a alunos de diferentes escolas, com divulgação voltada principalmente ao público de Ensino Fundamental. A oficina ocorreu em dois sábados pela manhã, por ser um período livre para a maioria dos estudantes, com cerca de 3 horas por encontro, incluindo intervalo para lanche. De 16 alunos inscritos, 5 compareceram em ambos os encontros, sendo 2 meninas e 3 meninos. Esses alunos eram provenientes de 3 escolas e cursavam o 5º ou 6º ano do Ensino Fundamental.

A sequência do trabalho com o App Inventor foi semelhante à da experiência anterior, porém com uma melhoria: para a segunda oficina, desenvolveu-se um guia sobre o App Inventor, que foi entregue impresso a cada participante. Essa iniciativa apoiou-se num estudo [Xie et al. 2015] que sugere que usar tutoriais no primeiro contato com o App

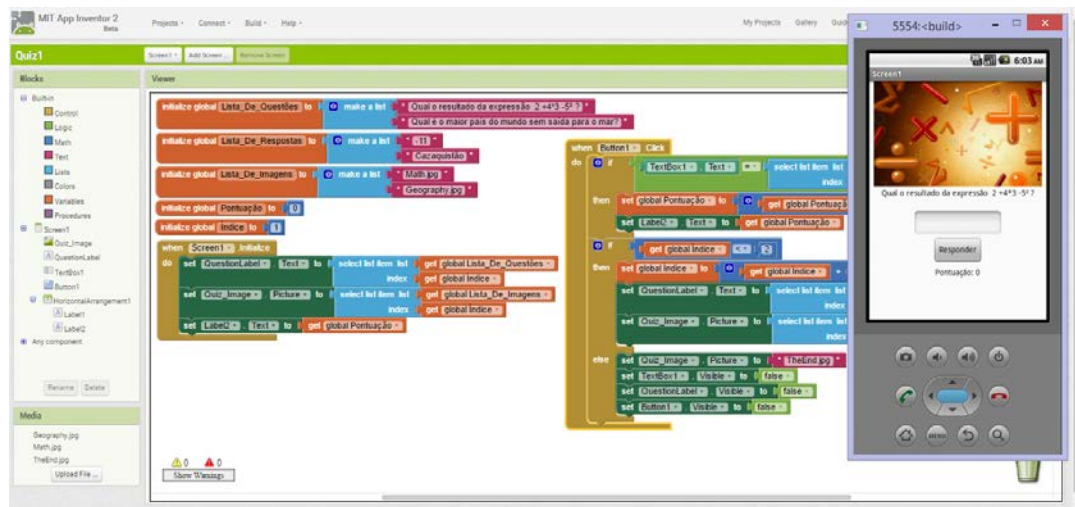


Figura 4. Exemplo de aplicativo produzido por um aluno

Inventor influencia na usabilidade percebida pelo usuário, no momento de implementar funcionalidades particulares. Esse guia foi organizado nas seguintes seções:

1. Primeira vez no MIT App Inventor?
2. Meu primeiro aplicativo
3. Aplicativo Quiz
4. Conhecendo melhor os blocos
5. Testando o aplicativo
6. Aprendendo a programar
7. Postando seu aplicativo na galeria
8. Mais informações

Entre os encontros, e mesmo após, manteve-se contato com os alunos via rede social, resolvendo dúvidas que surgiram quando esses tentaram utilizar o App Inventor em suas residências. A avaliação usou o mesmo formulário usado na primeira experiência.

5. Resultados e Discussão

De forma geral, as respostas dos alunos no formulário avaliativo revelaram boa aceitação da oficina, com níveis de dificuldade classificados como 3 ou 4, numa escala de 1 (mais baixo) a 5 (mais alto). Alguns alunos registraram comentários no formulário, que transcrevemos a seguir:

“Só acho que seria interessante uma continuação, onde pudéssemos nos aprofundar mais nos assuntos.”

“Seria interessante novos encontros mais dinâmicos.”

“A condução das atividades foi excelente. A quantidade de pessoas auxiliando colaborou bastante.”

Embora os públicos nas duas experiências tivessem idades e níveis de escolaridade distintos, o desempenho dos alunos nas oficinas não revelou muita diferença. Em ambas as oficinas, alguns alunos progrediram mais rapidamente que outros, mas todos conseguiram atingir os objetivos propostos. Na segunda oficina, notou-se que a disponibilidade do guia impresso foi um grande facilitador. Uma observação que merece atenção

	Inscritos	Maior número de presenças	Menor número de presenças
Oficina 1: Ensino Médio	31	19	9
Oficina 2: Ensino Fundamental	16	5	5

Tabela 1. Números de inscritos e presenças nos encontros

diz respeito aos números de inscritos e presentes nos encontros, reunidos na tabela 5. Nessa tabela, colocou-se lado a lado o maior e o menor número de presenças em cada oficina, considerando-se todos os encontros.

O número de inscritos, nas 2 oficinas, foi maior do que o número de presenças. Esse comportamento já foi observado em todas as atividades anteriores do projeto de extensão e pode ser atribuído ao fato de que as atividades são gratuitas e o formulário de inscrição é disponível *online*, facilitando a inscrição de alunos que, depois, mudam de ideia.

Nota-se que os alunos do ensino fundamental, embora menos numerosos desde o início, estiveram presentes na totalidade dos encontros. O mesmo não se verificou com alunos do ensino médio, cujas presenças variaram mais. Mesmo considerando que a primeira oficina teve mais encontros, vale lembrar que a segunda ocorreu fora das escolas e aos sábados. Embora não tenhamos a pretensão de generalizar, essas observações são alinhadas com uma hipótese de que o público de ensino médio seja mais difícil de se manter estimulado com a proposta. Um dos motivos para isso pode ser a proximidade do ingresso no ensino superior, que se torna uma prioridade para muitos. De outro lado, os alunos de ensino fundamental, além de parecerem bastante atraídos pela proposta, são também influenciados pelos seus responsáveis, que os conduziram e, possivelmente, os estimularam a participar dos encontros.

Alguns obstáculos envolvendo a utilização do App Inventor foram observados ao longo dos encontros, pelos tutores, mesmo não sendo explicitamente apontados como dificuldades pelos alunos. Os mais frequentes foram:

- **Interface em inglês:** à época da realização das oficinas, o App Inventor não possuía interface em português, por isso foi usado o idioma inglês. Isso causou dispersão em alguns alunos, ao buscarem traduções paralelamente ao desenvolvimento dos aplicativos. Essa dificuldade também foi observada em outro trabalho [Ramos et al. 2015]. Atualmente, o App Inventor já possui interface em português, mas notou-se que a primeira caixa de diálogo da ferramenta continua em inglês.
- **Teste dos aplicativos:** conforme já mencionado, o App Inventor inclui um emulador de dispositivo móvel, no qual os aplicativos podem ser testados. Outra opção é baixar o aplicativo no próprio celular. Em ambos os casos, notou-se dificuldades: o emulador muitas vezes interrompe a comunicação com a ferramenta ou leva um tempo considerável para atualizar a visualização, o que causa muitas dúvidas aos alunos; já a instalação no celular não é feita como os alunos estão habituados (via Google Play), por isso também causa dúvidas.

- **Blocos condicionais:** a abordagem utilizada não partiu de uma exposição detalhada dos blocos, mas mesmo assim a lógica do *Quiz* e muitos blocos foram rapidamente compreendidos pelos alunos. No entanto, blocos condicionais geraram muitas dúvidas não somente pela lógica, mas pela necessidade de muitos encaixes e configurações menos intuitivas.

É importante ressaltar que a presença dos tutores foi fundamental nas oficinas. Pela abordagem utilizada, as crianças e jovens foram estimulados a explorar o App Inventor e solicitar ajuda quando tivessem dificuldades. Na prática, cada tutor assistiu um ou dois alunos, acompanhando as criações durante os encontros. Essa abordagem inviabiliza oficinas com grande número de participantes, que necessitaria de muitos tutores. Felizmente, nossa experiência com esse tipo de ação extensionista tem mostrado que os graduandos são atraídos pela possibilidade de interagir com o público externo e, mais do que isso, contribuir para disseminar fundamentos da área. Assim, pudemos contar com 10 graduandos voluntários atuando como tutores em diferentes ocasiões.

6. Considerações Finais

As experiências realizadas aproximaram alunos de ensino médio e fundamental a uma possibilidade que até então lhes parecia distante: eles/elas foram capazes de criar aplicativos para seus celulares. Assim como outras intervenções extensionistas relatadas na literatura, estas experiências não só aproximaram a Universidade da comunidade, mas também estimularam o interesse de crianças e jovens em aprofundar seus conhecimentos na área de Computação. Para graduandos, a experiência também revelou-se estimulante, atraindo voluntários e não somente bolsistas.

Visando facilitar a replicação deste tipo de experiência, o material desenvolvido e utilizado nas oficinas encontra-se disponível em: facebook.com/CompClubUFSM. Sugere-se, em ações futuras, levar em consideração as seguintes lições aprendidas:

- Alunos de ensino fundamental ficam tão à vontade quanto os do ensino médio frente ao App Inventor. Assim, caso se tenha que priorizar uma ou outra faixa etária beneficiada com a experiência, sugere-se iniciar com os de ensino fundamental, por ser um público com menos preocupações (ingresso no ensino superior) e que, potencialmente, seguirá mais facilmente participando de novas ações extensionistas que visem aprofundar o conhecimento de fundamentos de Computação.
- Usar material impresso durante as atividades, embora tenha um custo adicional, é uma estratégia vantajosa para estimular a autonomia dos alunos. Nota-se que, disponibilizando esse material fisicamente (e não somente *online*), os alunos passam a consultá-lo antes de solicitar ajuda de um tutor.
- Embora o App Inventor seja uma ferramenta com recursos intuitivos, ele impõe também alguns obstáculos a usuários iniciantes. Nesses pontos, mesmo que exista documentação explicando os blocos e outros componentes, é fundamental o apoio de um tutor, pois do contrário os alunos podem facilmente se desestimular.

Referências

de Cerqueira Cardoso, J. P. P., Morais, L. C., Rodrigues, C. A., e Bittencourt, R. A. (2015). Introduzindo programação de computadores e PBL com python e App Inventor. In *Anais do XLIII Congresso Brasileiro de Educação em Engenharia – (COBENGE)*, pages 1 – 10, Feira de Santana, BA.

- Duda, R., Zontini, D. D., de Carvalho Rutz da Silva, S., e Grossi, L. (2014). Elaboração de aplicativos para Android com o uso do App Inventor: uma experiência no instituto federal do paran  - c mpus irati. In *Anais do IV SINECT – Simp sio Nacional de Ensino de Ci ncia e Tecnologia*, pages 1 – 12, Ponta Grossa, PR.
- Finizola, A. B., Raposo, E. H. S., Pereira, M. B. P. N., Gomes, W. S., de Ara jo, A. L. S. O., e Souza, F. V. C. (2014). O ensino de programac o para dispositivos m veis utilizando o MIT-App Inventor com alunos do ensino m dio. In *Anais do Workshop sobre Inform tica na Escola – (WIE)*, pages 337 – 341.
- Fronza, I., El Ioini, N., e Corral, L. (2015). Students want to create apps: Leveraging computational thinking to teach mobile software development. In *Proceedings of the 16th Annual Conference on Information Technology Education, SIGITE '15*, pages 21–26, New York, NY, USA. ACM.
- Gardner, H. e Davis, K. (2013). *The App Generation: How Today's Youth Navigate Identity, Intimacy, and Imagination in a Digital World*. Yale University Press.
- Gomes, T. C. S. e de Melo, J. C. B. (2013a). L gica de programac o: Inicia o l dica com app inventor for android. *Revista da Escola Regional de Inform tica de Pernambuco*, pages 155 – 161.
- Gomes, T. C. S. e de Melo, J. C. B. (2013b). O pensamento computacional no ensino m dio: Uma abordagem blended learning. In *Anais do Workshop sobre Educa o em Computa o – (WEI)*, pages 1 – 10.
- Hulme, A. K. e Traxler, J. (2005). *Mobile Learning: A Handbook for Educators and Trainers*. Routledge.
- Khalaf, S. (2015). Mobile addicts multiply across the globe. Flurry Analytics.
- Ramos, N., Freitas, C., Avila, S., Costa, P. D. P., Testoni, V., e Borin, J. F. (2015). Ensino de programac o para alunas de ensino m dio: Relato de uma experi ncia. In *Anais do Workshop sobre Educa o em Computa o – (WEI)*, pages 1 – 10.
- Rideout, V. J., Foehr, U. G., e Roberts, D. F. (2010). Generation M2: Media in the Lives of 8- to 18-Year-Olds. Technical report, Henry J. Kaiser Family Foundation.
- Wolber, D. (2011). App inventor and real-world motivation. In *Proceedings of the 42Nd ACM Technical Symposium on Computer Science Education, SIGCSE '11*, pages 601–606, New York, NY, USA. ACM.
- Xie, B., Shabir, I., e Abelson, H. (2015). Measuring the usability and capability of app inventor to create mobile applications. In *Proceedings of the 3rd International Workshop on Programming for Mobile and Touch, PROMOTO 2015*, pages 1–8, New York, NY, USA. ACM.

Experiências de Ensino da Computação Desplugada em Diferentes Séries da Educação Fundamental Maior

Alberto Alan Raiol¹, Ana Carina Lima¹, Breno Sousa¹, João Sarges¹,
Rafael Santos¹, Fábio Bezerra¹

¹ Instituto Ciberespacial – Universidade Federal Rural da Amazônia
Avenida Presidente Tancredo Neves, 2501 – Belém-Pará

fabio.bezerra@ufra.edu.br

Resumo. *O estudo da computação é visto como algo complexo e que possui conceitos de difícil abstração. Esta leitura da área da computação poderia ser melhorada se os alunos da educação básica, do ensino fundamental ou médio, tiverem o contato com a computação como uma ciência. No entanto, em que momento do ciclo de formação a computação poderia ser apresentada a esse alunado da educação básica? Com a intenção de mostrar a importância da computação e facilitar o ensino e o aprendizado desta área, o presente trabalho relata a aplicação da Computação Desplugada como método de ensino bem aceito pelos estudantes do ensino fundamental maior. Os resultados dessa experiência apontam que o sétimo ano foi a série que melhor respondeu bem ao conteúdo trabalhado em sala de aula.*

Abstract. *The study of computing is seen as complex and of hard abstraction. This understanding of computing area could be improved if the students of basic education, elementary or high school, had contact with computing as a science. However, when in the training cycle computing science could be presented to these students of basic education? In order to show the importance of computing science and to facilitate the teaching and the learning of this area, this paper describes the application of computing science unplugged with students of upper elementary education. The results of this experiment indicate that the seventh year was the moment that best responded to the content worked in the classroom.*

1. Introdução

O ano de 1998 teve como destaque o fim da primeira etapa do processo da informatização das escolas públicas brasileiras. A partir de então, há uma preocupação sobre a forma adequada de se aplicar a computação nas escolas, para que o ensino de conceitos computacionais nas mesmas possa caminhar de forma legítima ao que a disciplina oferece [BORGES NETO 1999]. Passada esta fase de primeiro contato com a informatização das escolas, observa-se que atualmente no Brasil as práticas de ensino de conceitos da Computação nas escolas não proporcionam o conhecimento necessário afim de que os alunos avancem para um futuro promissor nessa área, especialmente porque a computação ainda não é um elemento curricular obrigatório. No ambiente escolar pouco se observa o ensino da computação enquanto ciência, uma vez que os alunos realizam atividades mais práticas envolvendo hardwares e softwares específicos, talvez porque ainda se acredita que a computação resume-se à habilidade de manusear e configurar equipamentos e

programas de computador, podendo também difundir o pensamento incorreto que os computadores são autossuficientes e solucionadores de qualquer problema, não necessitando da intervenção humana para funcionar [da Conceição Araújo et al. 2015].

Acreditamos que um grande ponto que dificulta a inserção de fundamentos computacionais na educação básica é o fato do tema não fazer parte da diretriz curricular brasileira, fazendo com que as crianças convivam quando possível, com a mera utilização dos recursos básicos da máquina, sem exploração da interdisciplinaridade existente entre as outras disciplinas e amadurecimento de habilidades estimulados pela ciência da computação, como raciocínio lógico e o pensamento computacional [Costa et al. 2012]. Muitos trabalhos discutem a necessidade de a Computação ser trabalhada como ciência desde o ensino básico [França et al. 2012, Bell et al. 2011, da Siva et al. 2015, de França and do Amaral 2013]. Neste contexto, a Sociedade Brasileira de Computação (SBC) apresenta-se como um importante instrumento, pois promove anualmente eventos para buscar debates em diferentes temas relacionados ao ensino de Computação no país.

Neste cenário de inserção da computação como ciência na educação, a computação desplugada apresenta-se como um importante método de ensino. O material que foi desenvolvido por três professores de Ciência da Computação – Tim Bell, Lan H. Witten e Mike Fellows – e disponível gratuitamente no livro “*Computer Science Unplugged*”, com tradução para o português, destaca-se como uma alternativa eficaz na finalidade de atenuar os desafios encontrados para o ensino dos conceitos de computação nas escolas. Conforme [Bell et al. 2011], sem o uso do computador e de forma descontraída, a Computação Desplugada pode ser empregada em uma sala de aula normal e fundamentos importantes da computação podem ser ensinados de forma fácil aos alunos. Por isso, as atividades da computação desplugada estão sendo utilizadas por educadores e não especialistas em computação em diversos países do mundo, tanto dentro quanto fora dos ambientes escolares. No caso das escolas públicas brasileiras, o método da computação desplugada torna-se especialmente importante, pois democratiza o acesso a um importante conteúdo da ciência sem precisar investir em um laboratório de ensino especializado.

Este artigo tem como objetivo relatar a experiência que os autores vivenciaram com quatro turmas do ensino fundamental maior (sexto, sétimo, oitavo e nono ano) de uma instituição pública de ensino em Belém, utilizando o método da computação desplugada. O objetivo dessa experiência na escola era aproximar os estudantes de conteúdos da ciência da computação, fomentar a afinidade pela ciência da computação, mas principalmente, avaliar qual ano/série melhor responderia, quanto a aprendizagem e motivação, aos conteúdos apresentados.

Este trabalho está organizado da seguinte forma: na Seção 2, descrevemos os materiais e métodos utilizados e o público que participou de nossas experiências; na Seção 3, apresentamos as lições aprendidas com as intervenções; na Seção 4, apresentamos as experiências de trabalhos com ações semelhantes as relatadas aqui; por fim, na Seção 5, apresentamos as nossas considerações finais e propostas para trabalhos futuros.

2. Materiais e Métodos

Nesta seção abordaremos o método de ensino aplicado em nossas intervenções (Seção 2.1) e apresentaremos o planejamento das oficinas, o público participante (Seção 2.2) e processo de ensino nas aulas (Seção 2.3).

2.1. Método de Ensino: Computação Desplugada

A computação desplugada é uma técnica que tem como objetivo ensinar os fundamentos da computação para estudantes e pessoas de diferentes idades, apresentando com uma linguagem simples como os computadores funcionam, porém sem utilizar o computador como instrumento de ensino [Bell et al. 2009]. Assim, as atividades podem ser aplicadas em qualquer escola, justamente por não necessitar de equipamentos eletrônicos, ocasionando a desvinculação da ideia do aprendizado de computação que é popularmente atrelada à utilização de um computador.

O livro “Computação Desplugada” foi criado por Tim Bell, Ian H. Witten e Mike Fellows. Possui 12 ricas atividades que cobrem diferentes áreas ou temas da computação, e pode ser baixado gratuitamente em www.csunplugged.org. O livro está organizado em três partes: dados (cinco atividades), algoritmos (cinco atividades) e linguagens (duas atividades).

Este trabalho limitou-se a primeira parte do livro, que é fundamental para entender como os computadores representam, armazenam e trocam dados. Cada uma das 12 atividades geralmente possui cinco estruturas: atividades (sumário, matérias correlacionadas, habilidades, idades e material), conteúdo (introdução e discussão), folha de atividade (instruções, discussão e atividades), uma explicação final e mais aprofundada do conteúdo, e resoluções e dicas de aprimoramento da atividade.

2.2. Ambiente e Escopo da Experiência: Escola, Turmas e Atividades

As experiências expostas neste trabalho foram realizadas com alunos do ensino fundamental maior, entre os meses de novembro de 2014 e fevereiro de 2015 na Escola Estadual de Ensino Fundamental e Médio Barão de Igarapé Miri, localizada na periferia de Belém.

Tiveram acesso às atividades com a computação desplugada apenas uma turma de cada uma das séries, do sexto ao nono ano. Essa limitação – apenas uma turma de cada série e apenas três atividades – está relacionada à disponibilidade de horário, no mesmo turno das aulas, que a escola poderia ceder a execução das oficinas. No entanto, mesmo considerando que apenas quatro turmas foram atendidas, ao final da oficina, 110 alunos tiveram contato com os conteúdos planejados para a ação. No caso, apresentamos as três primeiras atividades, das 12 disponíveis no livro: “Números Binários”, “Representação de Imagens” e “Compressão de Texto”.

Cada atividade do livro foi planejada em formato de oficina, compreendendo aproximadamente a duração de 90 minutos. As oficinas eram realizadas duas vezes na semana na escola, nos dias de quarta e sexta-feira, sendo que no primeiro momento da semana, a oficina era aplicada para as turmas de sexto e nono ano, e no segundo momento, a oficina era realizada com os alunos das turmas de sétimo e oitavo ano (ver Tabela 1).

Com um público dividido em quatro séries escolares, as oficinas contemplaram alunos com idades que variaram entre onze a quatorze anos. Todavia, o planejamento

Tabela 1. Quantidade de Participantes por Turma

Turmas	Numero de Participantes
Sexto ano	31
Sétimo ano	30
Oitavo ano	25
Nono ano	24
Total	110

das aulas, bem como sua execução, se fez de forma semelhante para todas as quatro turmas. Entretanto, vale ressaltar que apesar de termos planejado e agendado antecipadamente as aulas, junto à direção da escola e aos professores, que cederam alguns de seus horários, houveram alterações constantes na agenda planejada, motivadas tanto por parte dos mesmos docentes que em período de avaliações acabaram utilizando os horários que em outrora haviam cedido, quanto por programações da Secretaria Estadual de Educação e eventos da própria instituição. Na Tabela 2 apresentamos os conteúdos trabalhados, os materiais utilizados e os objetivos de cada aula.

Tabela 2. Planejamento Resumido da Oficina

Aula	Conteúdo	Materiais	Objetivo
1º Dia	Atividade I - Números Binários - Conversão de decimal para binário, binário para decimal; desafio do código, decodificação e envio de mensagem.	6 cartões de tamanho A4, crachás para identificação dos grupos, lápis, borracha, marcadores, palito de churrasco, rolo de fita crepe, folha de isopor e percevejos.	Compreender o sistema binário, realizando a conversão de valores decimais em binários e binários em decimais.
2º Dia	Atividade II - Representação de Imagens - Codificando e Decodificando imagens (formas geométricas e letras do alfabeto).	40 cartões 5*5m, isopor, fita crepe, percevejos, borracha e lápis.	Conhecer e entender como os computadores representam as imagens, identificando as diferentes quantidades de pixels das figuras.
3º Dia	Atividade III - Compressão de Texto - Identificando modelos de vocábulos e frases.	Textos (poemas e cantigas), lápis, caneta, borracha.	Compreender a forma eficiente do computador salvar as informações, com a reutilização de dados, fazendo a supressão de duas ou mais palavras na compressão de textos.

Adaptamos as atividades para serem aplicadas em formato de competição entre equipes, havendo a formação de pequenos grupos de alunos, que interagiam entre si para solucionarem os desafios propostos nas aulas. Dessa forma, as competições foram organizadas visando criar uma atmosfera de aprendizagem divertida[Scaico et al. 2012].

2.3. Dinâmica das Aulas

O envolvimento dos estudantes com as atividades aconteceram de seguinte maneira: inicialmente uma explicação do conteúdo de forma expositiva e com demonstrações; em seguida um exercício individual para aproximar o aluno do conteúdo; finalmente, foram realizadas dinâmicas competitivas entre os grupos para fixação dos conteúdos abordados. A dinâmica em grupo previa uma premiação simbólica, então o grupo que executasse a tarefa em menor tempo ganhava um brinde a mais.

Semelhante ao trabalho apresentado em [Bezerra 2014], para aproximar os alunos do vocabulário ligado às tecnologias e ajudar na coordenação das dinâmicas em sala, criamos crachás para identificar os grupos, cujos nomes estavam associados: às linguagens de programação *Python, Ruby, Java e Php*; aos provedores de e-mail *Yahoo, Gmail e Hotmail*; às redes sociais *Twitter, WhatsApp, Facebook e Instagram*. Para nossa surpresa, mesmo considerando o contexto social, econômico e geográfico dos alunos, com exceção das linguagens de programação, tanto os provedores quanto as redes sociais já eram expressões familiares aos mesmos.

Durante as duas tarefas iniciais, as avaliações das apresentações pelos estudantes ocorreram com base em dois conceitos, S (Sim) para indicar que o aluno entendeu ou assimilou as atividades lúdicas trabalhadas e N (Não) para indicar que não entendeu. Na terceira tarefa foi acrescentada a resposta R (Regular) à avaliação, a fim de aumentar a precisão da opinião do aluno quanto a aprendizagem e motivação do conteúdo.

3. Resultados e Discussões

Os resultados produzidos pela experiência estão relacionados às seguintes observações: resposta a aprendizagem do conteúdo trabalhado e motivação dos alunos na participação das atividades. A aprendizagem foi um item coletado por dois métodos: (i) auto-avaliação dos alunos, que respondiam em formulário próprio se conseguiram ou não absorver o conteúdo apresentado; (ii) observação dos instrutores durante as gincanas. Enquanto a motivação foi um item medido através da frequência dos alunos às atividades.

3.1. Aprendizagem do Conteúdo

Ao final de cada oficina os alunos respondiam em formulário próprio se haviam entendido o conteúdo da atividade apresentada. Com respostas simples, representadas por “Sim” ou “Não” os alunos responderam esse questionamento. Para a última atividade a alternativa “Regular” foi adicionada ao formulário, para indicar que o aluno entendeu o conteúdo de forma incompleta ou não sente domínio do mesmo.

Apresentamos na Tabela 3, as respostas dos alunos presentes em cada dia de atividade, para cada turma, o percentual de respostas positivas e negativas. Nessa tabela podemos observar que todas as turmas envolvidas responderam bem à aprendizagem de todos os conteúdos apresentados. Com exceção da atividade de “Representação de Imagens”, para a turma do sexto ano, que respondeu positivamente para 78,3% dos alunos presentes. No entanto, não chega a ser um índice ruim.

Na observação dos instrutores, os alunos da turma do sétimo se desenvolveram melhor na resolução das atividades, pois apresentaram-se mais articulados, desenvolvidos e participativos nas aulas. Em outro contexto, notamos que os alunos da turma do sexto

Tabela 3. Respostas da auto-avaliação por cada atividade

Turmas	Atividade I				Atividade II				Atividade III					
	Sim	%	Não	%	Sim	%	Não	%	Sim	%	Não	%	Regular	%
Sexto ano	18	81,8	4	18,2	18	78,3	5	21,7	11	100	0	0	0	0
Sétimo ano	18	90	2	10	22	100	0	0	16	100	0	0	0	0
Oitavo ano	18	90	2	10	19	100	0	0	6	100	0	0	0	0
Nono ano	17	100	0	0	11	100	0	0	10	83,3	0	0	2	16,7

ano tiveram mais dificuldades em compreender algumas atividades propostas, recorrendo muitas vezes a ajuda dos monitores, para sanar as dúvidas e realizar as tarefas.

As turmas do oitavo e nono ano também apresentaram um bom desempenho na realização das tarefas, contudo, em ambas as turmas percebemos uma certa desmotivação de alguns alunos na evolução das aulas, sobretudo na primeira e segunda atividade – ex.: relatos de falta de interesse para realizar a tarefa. Assim no decorrer das aulas, apenas os alunos motivados foram os mais presentes e participativos.

Tabela 4. Respostas da auto-avaliação de todo o conteúdo

Turmas	Sim		Não		Regular	
	Respostas	%	Respostas	%	Respostas	%
Sexto ano	47	83,93	9	16,07	0	0
Sétimo ano	56	96,55	2	3,45	0	0
Oitavo ano	43	95,55	2	4,45	0	0
Nono ano	38	95	0	0	2	5

Na Tabela 4, encontram-se as respostas dos alunos por turma sobre o que foi questionado somando-se os três dias de atividades. Nesse contexto, nota-se que em todas as turmas o percentual de respostas positivas (Sim) dos alunos é bem significativa, visto que a turma do sexto ano apresentou 83,93% de confirmação de entendimento do conteúdo pelos alunos, a turma do sétimo ano 96,55%, a do oitavo ano 95,55% e nono ano 95%.

3.2. Motivação das turmas

Entendemos que a motivação é uma aspecto humano e pessoal, portanto, é algo individual de cada aluno das turmas que essa experiência relata. No entanto, para o contexto da avaliação dessa experiência, aplicaremos o significado de motivação para a turma, ou conjunto de alunos, de forma claramente simplificada. No caso, o aspecto da motivação da turma relatado neste trabalho refere-se à frequência da turma em cada uma das atividades, ou seja, 70% de frequência significa que 70% dos alunos estiveram presentes na aula.

Apesar do conteúdo das oficinas ser extra-curricular, observamos uma participação razoável dos alunos às oficinas. Acreditamos que esse interesse ao longo das três atividades esteve ligado as novas práticas de ensino proporcionada pelo método das aulas de computação desplugada, uma vez que as mesmas são desenhadas para promover a interação e curiosidade dos aprendizes. A Tabela 5 ilustra a frequência das turmas, das quais a do sétimo ano ganha destaque por apresentar maior frequência às aulas.

3.3. Discussões

Os resultados relatados nessa experiência corroboram com os relatos de outros trabalhos da literatura, que apontam a riqueza e vantagens da aplicação do método da computação

Tabela 5. Frequências das turmas nas oficinas

Turmas	Frequência da turma (%)
Sexto ano	60,22
Sétimo ano	64,44
Oitavo ano	60
Nono ano	55,55

desplugada na educação básica.

Especificamente nessa experiência, acreditamos que a computação desplugada apresentou-se mais apropriada aos alunos do sétimo ano, enquanto que a turma do sexto ano teve maior percentual de rejeição. Contudo, acreditamos que o método apresenta-se como uma alternativa viável para qualquer série do ensino fundamental maior, especialmente quando consideramos que a rejeição foi, de uma forma geral, pequena.

4. Trabalhos Relacionados

Alguns autores ultimamente vêm discutindo a importância da inclusão da ciência da computação no currículo da educação básica [de França and do Amaral 2013, de Souza Barbosa et al. 2015, da Siva et al. 2015]. Diversos países desenvolvidos já apresentam um currículo mínimo para o ensino de computação na educação básica. No Brasil, diversas práticas vem sendo aplicadas para alunos do ensino fundamental e médio, trazendo contribuições através de estudos de métodos que podem futuramente ser incluídos no currículo de computação para a educação básica, tais práticas estão relatadas nos trabalhos de [de Paiva et al. 2015, Scaico et al. 2011, de França et al. 2013, Andrade et al. 2013, Scaico et al. 2012]. Em [de França and do Amaral 2013] os autores realizaram um mapeamento sistemático da produção científica brasileira nos anos de 2009 à 2012, com o objetivo de verificar quais práticas podem ser adotadas no país. Nesse mapeamento os autores notaram que a computação desplugada ganhou destaque nos anos de 2011 e 2012.

Em [de Souza Barbosa et al. 2015] os autores apresentam diversos métodos utilizados no ensino de computação na educação básica. No caso, a computação desplugada vem ganhando destaque nas pesquisas, porém ainda há poucos resultados sobre a avaliação e estratégias de resolução das atividades desplugadas que estão sendo desenvolvidas. Em [de Souza Barbosa et al. 2015] foi realizado o planejamento, execução e identificação das estratégias desenvolvidas pelos alunos do primeiro ano do ensino médio, em uma gincana desplugada. Mesmo com algumas dificuldades, os estudantes conseguiram desenvolver as técnicas propostas pelos monitores e também utilizaram técnicas próprias para a resolução das atividades.

As conexões interdisciplinares da ciência da computação são colocadas em evidência em [de Paiva et al. 2015]. No caso, os autores relatam a aplicação interdisciplinar da computação com sete disciplinas regulares da educação básica. Em [Scaico et al. 2011], através do projeto "Sem matemática não existe computação", os autores também reforçam o papel interdisciplinar da computação.

Em [da Siva et al. 2015] relatam uma experiência com alunos do nono ano da ensino fundamental maior, baseada no *Model Curriculum for k12 Computer Sci-*

ence¹ e nas atividades do livro *Computer Science Unplugged*. Os trabalhos de [de França et al. 2013, Andrade et al. 2013, Scaico et al. 2012] relatam experiências semelhantes com a computação desplugada, a fim de disseminar a computação na educação básica e desenvolver o pensamento computacional.

No contexto de programação de computadores, em [de França et al. 2013] os autores realizaram um projeto de programação com Scratch. No caso, as atividades foram desenvolvidas abordando alguns trabalhos realizados na escola nos anos de 2010 a 2012, envolvendo a programação de computadores de forma lúdica, desenvolvimento do raciocínio lógico e o ensino de conceitos computacionais.

Em [Andrade et al. 2013] os autores relatam uma proposta metodológica para desenvolver o *Computational Thinking Toolkit*. Para isso, foram realizadas atividades lúdicas para estudantes do ensino fundamental, envolvendo conceitos básicos da computação. Em [Scaico et al. 2012] os autores combinam o uso da computação desplugada e do ensino de programação com scratch, para que os alunos do ensino médio possam desenvolver habilidades e competências com a área da computação, bem como aumentar o interesse pela área.

5. Considerações Finais e Trabalhos Futuros

Neste trabalho levamos a uma escola pública de Belém o ensino da ciência da computação, o que causou bastante entusiasmo nos alunos, pois responderam positivamente as expectativas. Além disso, pudemos observar que em todas as turmas investigadas, as respostas dos alunos quanto à compreensão dos conteúdos abordados, foi bastante expressiva, dando indícios que a aprendizagem dos conceitos trabalhados foram alcançados. Nesse cenário, em ênfase, a turma do sétimo ano (antiga sexta série) apresentou os melhores resultados, ficando também acima das demais em assiduidade nas aulas e na avaliação dos monitores, sendo a turma que teve maior aceitação e entendimento dos conteúdos ensinados nas aulas. Portanto, uma hipótese que pode ser melhor investigada é se o sétimo ano do ensino fundamental é o momento mais favorável para aplicação do método da computação desplugada.

Mais do que isso, este trabalho relata a experiência docente dos autores com estudantes de diferentes séries do ensino fundamental maior. Aos autores ainda em formação, houve claramente um amadurecimento acadêmico e pedagógico, pois permitiu o exercício prático das teorias apresentadas na universidade, a experiência docente através do contato professor-aluno, e a vivência do contexto escolar com todas as suas peculiaridades. Para os alunos, notamos o entusiasmo com “o novo”, pois a ciência da computação foi apresentada como um conteúdo interessante e útil, que os alunos não têm naturalmente acesso, uma vez que não faz parte do currículo da educação básica. Um terceiro ente beneficiado com as ações relatadas neste artigo foi a escola, pois foi uma oportunidade para dialogar e repensar as ações que melhoram o processo de ensino e aprendizagem entre a escola, os alunos e os bolsistas.

Acreditamos que o contato com a computação, desde cedo, pode ser um estimulante para que esses alunos optem pela computação como carreira profissional. Assim, re-

¹Modelo proposto pela Association for Computing Machinery (ACM), que serve como base e fornece orientações sobre as práticas de ensino de computação em diferentes etapas da educação [Tucker et al. 2003]

alizer a mesma experiência com alunos do último ano do ensino médio pode ajudar a identificar o quanto com contato com esses conteúdos influenciam na escolha da computação como atividade profissional. Outro objetivo futuro é aprimorar o método de ensino e os materiais e atividades utilizadas nas oficinas, a fim de melhorar a compreensão dos fundamentos da ciência da computação pelos educandos por meio da Computação Desplugada.

Ainda como trabalhos futuros, pretendemos explorar todos os conteúdos presentes no livro de [Bell et al. 2011], bem como aplicar nossas intervenções nas demais séries do ensino fundamental menor, não contempladas neste trabalho. Outro fator a ser explorado seria verificar se as atividades trabalhadas no livro, poderiam ajudar os alunos em conteúdos relacionados a matemática, já que algumas atividades colocam em prática alguns conteúdos como a soma, a exponenciação, os símbolos matemáticos, dentre outros.

Agradecimentos

Os autores agradecem o programa PIBID-UFRA, que financiou a realização deste trabalho através das bolsas de iniciação à docência.

Referências

- Andrade, D., Carvalho, T., Silveira, J., Cavalheiro, S., Foss, L., Fleischmann, A. M., Aguiar, M., and Reiser, R. (2013). Proposta de atividades para o desenvolvimento do pensamento computacional no ensino fundamental. In *Anais do Workshop de Informática na Escola*, volume 1, page 169.
- Bell, T., Witten, I. H., and Fellows, M. (2011). Computer science unplugged. ensinando ciência da computação sem o uso do computador. *Tradução por: Luciano Porto Barreto. Disponível em: <http://csunplugged.org/books>. Acesso: Março de.*
- Bezerra, F. (2014). Bem mais que os bits da computação desplugada. In *Anais do Workshop de Informática na Educação 2014*, Dourados, MS.
- BORGES NETO, H. (1999). Uma classificação sobre a utilização do computador pela escola. *Revista Educação em Debate*, ano, 21:135–138.
- Costa, T., Batista, A., Maia, M., Almeida, L., and Farias, A. (2012). Trabalhando fundamentos de computação no nível fundamental: experiência de licenciandos em computação da universidade federal da paraíba. In *XX Workshop de Educação em Computação-WEI. Curitiba, PR, Brasil.*
- da Conceição Araújo, D., Rodrigues, A. N., de Araújo Silva, C. V., and Soares, L. S. (2015). O ensino de computação na educação básica apoiado por problemas: Práticas de licenciandos em computação. *XXIII do WEI - Workshop sobre Educação em Computação.*
- da Siva, S. F., Barbosa¹², A. F., de Souza¹², A. A., Lauyse, E. G. d. S. M., and Rogério, S. d. O. S. (2015). Relato de experiência de ensino de computação no ensino fundamental em estágio supervisionado da universidade de pernambuco no campus garanhuns. In *XXI Workshop sobre Educaç ao em Computaç ao.*
- de França, R. S., da Silva, W. C., and do Amaral, H. J. C. (2013). Despertando o interesse pela ciência da computação: Práticas na educação básica. In *Proceedings of International Conference on Engineering and Computer Education*, volume 8, pages 282–286.

- de França, R. S. and do Amaral, H. J. C. (2013). Ensino de computação na educação básica no brasil: Um mapeamento sistemático. In *XXI Workshop sobre Educação em Computação*.
- de Paiva, L. F., Ferreira, A. C., Rocha, C., Barreto, J., Melhor, A., Lopes, R., and Matos, E. (2015). Uma experiência piloto de integração curricular do raciocínio computacional na educação básica. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 4, page 1300.
- de Souza Barbosa, A. V., Neto, A. F. P., de Oliveira, R. N. R., da Costa, T. L. S., Souto, A. L., De Araújo, O., and Costa, F. V. S. (2015). O ensino de conceitos computacionais para alunos do ensino médio: relato de experiência de uma gincana e das estratégias utilizadas pelos alunos na resolução das atividades desplugadas. In *XXI Workshop sobre Educação em Computação*.
- França, R. d., Silva, W. d., and Amaral, H. d. (2012). Ensino de ciência da computação na educação básica: Experiências, desafios e possibilidades. In *XX Workshop sobre Educação em Computação*.
- Scaico, P. D., Corlett, E. F., Paiva, L. F., Raposo, E. H. S., and Alencar, Y. (2012). Relato da utilização de uma metodologia de trabalho para o ensino de ciência da computação no ensino médio. In *Anais do Workshop de Informática na Escola*, volume 1.
- Scaico, P. D., Maia, M., Duarte, A. C., de Azevedo Silva, M. A., and da Silva, J. C. (2011). Sem matemática não existe computação. In *Anais do Workshop de Informática na Escola*, volume 1, pages 1424–1427.
- Tucker, A., Deek, F., Jones, J., McCowan, D., Stephenson, C., and Verno, A. (2003). A model curriculum for k-12 computer science. *Final Report of the ACM K-12 Task Force Curriculum Committee, CSTA*.

Explorando a Evolução do Problema de Pesquisa de Doutorado: Um Estudo Qualitativo no LER/UFPE

Enyo Gonçalves^{1,2}, Ana Carla Holanda^{1,3}, Dorgival Netto¹, Helena Cristina Bastos¹, André Calisto¹, Mayara Santos¹, Fabio Q. B. Silva¹, Carla Silva¹, Jaelson Castro¹

¹ Centro de Informática – Universidade Federal de Pernambuco, Recife-PE, Brasil

² Universidade Federal do Ceará, Campus Quixadá

³ Instituto Federal de Educação Ciência e Tecnologia do Acre

{ejtg, acah, dpsn2, hcab2, acsmg, mwas, fabio, ctlls, jbc}@cin.ufpe.br

***Resumo.** Neste artigo, vamos explorar como evoluem os problemas de pesquisa de doutorado, mais especificamente, no grupo de pesquisa LER (Laboratório de Engenharia de Requisitos) da Universidade Federal de Pernambuco (UFPE). Para tanto, foi realizado um estudo qualitativo para identificar fatores que podem influenciar na evolução do problema de pesquisa dos alunos de doutorado. Entrevistas semiestruturadas foram realizadas e dados foram coletados com o objetivo de termos subsídios para responder às questões de pesquisa. Os resultados indicam que existem fatores que se relacionam e que juntos podem facilitar a evolução do problema de pesquisa do doutorado. Por exemplo, o conhecimento prévio do aluno que, mesmo não sendo determinante, auxilia nessa evolução. Além disso, a definição do tema pode ser influenciada por um acompanhamento sistemático das atividades do aluno, pode influenciar positivamente no fator psicológico do orientando. A participação em workshops de teses e dissertações em computação e a interação com a indústria também podem influenciar de forma positiva.*

1. Introdução

Segundo dados do Ministério da Educação (MEC), o Brasil conta atualmente com 10.161 discentes matriculados em cursos de doutorado na área de ciências exata e da terra, sendo que cerca de 1.600 são da área de computação [GEOCAPES, 2015]. Um fato comum, no contexto de doutorado em computação no Brasil, se refere a pedidos de prorrogação de prazo da defesa da tese. Este pode ser um indicador de que há dificuldade na evolução dos problemas de pesquisa de doutorado nesta área.

É importante entender como evoluem os problemas de pesquisa de doutorado na área de computação e os fatores que os influenciam. Esse entendimento pode contribuir para uma melhor condução da pesquisa do doutorado em computação por parte dos discentes, além de possibilitar que aspectos que influenciem negativamente na evolução do seu problema de pesquisa sejam identificados.

Assim sendo, um estudo qualitativo pode contribuir na identificação de fatores que estão relacionados e influenciam diretamente ou indiretamente na evolução do problema. Este artigo apresenta os resultados de um estudo qualitativo realizado junto

ao Laboratório de Engenharia de Requisitos (LER) do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco (UFPE).

A contribuição deste trabalho se concentra na educação em informática em nível de pós-graduação. Os resultados deste trabalho podem ajudar alunos e orientadores de programas de doutorado a entender melhor como evoluem os problemas de pesquisa por meio de fatores que podem influenciar na evolução.

O restante do artigo está organizado como segue. A Seção 2 apresenta o referencial teórico e a Seção 3 os trabalhos relacionados. A Seção 4 caracteriza a metodologia seguida no desenvolvimento deste trabalho. Na Seção 5 são apresentados os resultados e discussões. E por fim, na Seção 6 são apresentadas as conclusões.

2. Referencial Teórico

Uma tese de doutorado pode trazer muitos benefícios para a vida acadêmica e profissional para o discente de uma Pós-Graduação *strictu sensu*. Neste sentido, muitas são as motivações para um candidato almejar essa conquista.

Porém, sabe-se que se trata de uma missão demorada, complexa e árdua, onde há muita “transpiração” como também inspiração [Duque, Brondani e Luna, 2005]. A Inteligência Emocional [Goleman, 1996] é um fator que merece ser destacado. O conceito de Inteligência Emocional surgiu no âmbito acadêmico, em 1990, formalizado pelos pesquisadores Peter Salovey (Yale University) e John Mayer (University of New Hampshire). Neta, García e Gargallo [2008] definem inteligência emocional como sendo a habilidade para controlar os sentimentos e emoções em si mesmo e nos demais, discriminar entre elas e usar essa informação para guiar as ações e os pensamentos.

Assim, Neta, García e Gargallo [2008] consideram a evolução de um mestrado ou doutorado como a realização de um projeto, este sendo definido como um esforço temporário empreendido para criar um produto, serviço ou resultado exclusivo. Os autores apresentam uma lista de verificação e riscos que podem surgir no decorrer do doutorado e propõem boas práticas no desenvolvimento de teses e dissertações.

Estes autores também procuraram identificar e avaliar possíveis fatores desencadeantes de estresse em pós-graduandos de diferentes áreas. Observam durante a pesquisa que a pressão por publicar leva a um grau exagerado de competitividade, induzindo uma distorção cultural, na qual a “corrida por produção” prevalece sobre o conhecimento.

3. Trabalhos Relacionados

O trabalho de Godoy, Silva e Figueiredo [2011] procura analisar o impacto de uma proposta de incorporação de prática docente e formação pedagógica aos cursos de pós-graduação na Unicamp – o Programa de Estágio Docente (PED). Os autores visam avaliar diferentes fatores relacionados à confiança na prática de pesquisa e ensino do mestrando e do doutorando em Engenharia Elétrica e em Ciência da Computação. O método utilizado envolveu a aplicação de um questionário autoaplicado para melhor entendimento da confiança. Nosso trabalho, assim como este, também está relacionado ao entendimento de fenômenos da pós-graduação em computação, porém nos concentramos no entendimento de como evoluem os problemas de pesquisa de doutorado.

Em Stump et al. [2011], os autores abordam a necessidade de se proporcionar formação em Cognição e Aplicações Computacionais aos alunos de pós-graduação stricto sensu em Computação. As estruturas conceituais e práticas relativas aos processos cognitivos foram abordadas. Também é apresentada uma forma de aperfeiçoamento da capacidade de construir um conhecimento organizado, por meio de uma proposta curricular para cursos stricto sensu na área da Computação e correlatas. O trabalho de Stump et al. [2011] assim como o nosso foca na melhoria da pós-graduação por meio do suporte a cognição dos alunos, no entanto o trabalho deste autores se concentra na defesa da criação de uma disciplina para tal finalidade.

Segundo Silveti et al. [2013], participar de um Doutorado Sanduíche ou de uma colaboração internacional é uma experiência enriquecedora, profissional e pessoalmente. Esta experiência permite que alunos de pós-graduação conheçam outros grupos de pesquisa, experimentem diferentes processos de trabalho acadêmico, aprendam formas alternativas de financiamento de pesquisa, além de estreitar contatos e criar parcerias com pesquisadores em diferentes países. O trabalho de Silveti et al. [2013] dá ênfase a somente um dos fatores que identificamos neste artigo.

4. Metodologia

[Problema de Pesquisa] O propósito deste trabalho é contribuir para um melhor entendimento de como evoluem os problemas de pesquisa de doutorado em computação. *[Design do estudo]* Nós utilizamos uma abordagem de pesquisa qualitativa básica realizada no grupo de pesquisa LER, o qual foi escolhido de forma intencional, por se tratar de um grupo de pesquisa consolidado no centro de informática da UFPE. *[Amostra dos participantes]* A unidade de análise do estudo é composta pelos alunos do grupo LER que se encaixem no seguinte intervalo: dois anos ou mais de ingresso no doutorado e menos de seis meses após a defesa. Os respectivos orientadores dos alunos entrevistados também foram considerados como participantes da pesquisa, como forma de realizar uma triangulação das informações. O propósito de utilizar uma amostra não probabilística encontra-se na seleção da informação, pois é possível escolher casos ricos para o estudo realizado [Merriam, 2009].

[Preparação da Coleta] Entrevistas semiestruturadas foram realizadas utilizando roteiros de entrevista especificamente concebidos e compostos de perguntas abertas. Os roteiros de entrevista foram pré-testados, através de uma entrevista piloto com um dos alunos de doutorado que é membro do LER e que estava no intervalo da amostra. Esta experiência ajudou a refinar os roteiros de entrevista (ver Quadro 1 e Quadro 2). Após a realização das entrevistas, ocorreu a transcrição e codificação e categorização dos dados. Também realizamos entrevistas de retrospectiva para clarificar e complementar as informações identificadas na análise dos dados. *[Coleta de Dados]* As entrevistas foram realizadas presencialmente nas instalações do centro de informática da UFPE (Com exceção de uma das entrevistas que foi realizada remotamente via skype), nos meses de outubro e novembro 2015. Oito participantes foram entrevistados: cinco alunos (dos quais dois já haviam concluído o curso) e três orientadores. Todas as entrevistas foram gravadas (com autorização prévia do entrevistado) e juntas totalizam 5 horas e 20 minutos de tempo de áudio. *[Análise de Dados e Síntese]* Seguimos diretrizes fornecidas por Strauss e Corbin [2007] para categorizar e sintetizar dados, com o objetivo de construir uma teoria baseada em

evidências como evolui o problema de doutorado em computação. O áudio das entrevistas foi transcrito e QSR NVIVO 11 foi usado para apoiar o processo de análise. Nós consideramos porções rotuladas de texto usando códigos (Iniciamos com codificação aberta, depois fizemos codificação fechada e por fim foi utilizada a codificação axial) e em seguida estes códigos foram relacionados dando origem às categorias que foram nomeadas seguindo um método de comparação constante [Strauss e Corbin, 20070]. Em seguida, as relações entre categorias foram mapeadas, levando a proposições que sustentam a estória central.

[*Credibilidade, Consistência e Transferência dos Resultados*] Nós utilizamos triangulação para aumentar a credibilidade dos dados. Além disto, os dados foram discutidos em grupo durante reuniões e inconsistências identificadas foram tratadas tanto nas discussões quanto em explicações complementares dos participantes.

Quadro 1. Roteiro de Entrevista dos Alunos

- | |
|--|
| <ol style="list-style-type: none"> 1. Descreva em detalhes sua trajetória acadêmica e profissional desde a graduação até hoje. 2. Descreva em detalhes sua trajetória no doutorado, desde a decisão de fazer o doutorado até agora. 3. Em sua opinião, o que é um problema de pesquisa de doutorado? 4. Qual o seu problema de pesquisa? 5. Descreva em detalhes como evoluiu a definição do seu problema de pesquisa? 6. Quais os fatores que influenciaram positivamente na evolução do problema de pesquisa? 7. Quais os fatores que influenciaram negativamente na evolução do problema de pesquisa? 8. Tem algo que eu não mencionei que você gostaria de falar? 9. Você tem alguma pergunta a nos fazer a respeito da pesquisa? |
|--|

Quadro 2. Roteiro de Entrevista dos Orientadores

- | |
|--|
| <ol style="list-style-type: none"> 1. Em sua opinião, o que é um problema de pesquisa para o doutorado? 2. Como um problema de pesquisa é definido? 3. Quais os fatores que influenciam positivamente a definição de um problema de pesquisa? 4. Quais os fatores que influenciam negativamente a definição de um problema de pesquisa? 5. Cite um exemplo de um orientado que você considera o ideal e que não teve dificuldade para definir o problema de pesquisa e no terminou no tempo ideal. 6. Cite um exemplo de um orientado que você considera o mais difícil, que teve dificuldade para definir o problema de pesquisa e não terminou no tempo ideal. 7. Você aplica algum método específico para ajudar seus orientandos a definir o problema de pesquisa deles? 8. Tem algo que eu não mencionei que você gostaria de falar? 9. Você tem alguma pergunta a nos fazer a respeito da pesquisa? |
|--|

5. RESULTADOS E DISCUSSÃO

Nós realizamos uma investigação qualitativa para descobrir quais possíveis fatores podem facilitar positivamente e negativamente a evolução do problema e os dados estão apresentados nas subseções seguintes.

5.1. Descrição do Contexto

Desde 1997, o grupo LER visa estabelecer bases para aumentar a qualidade do processo de produção de software por meio da cooperação com universidades e empresas e focando em atividades Engenharia de Requisitos. O foco de pesquisa do grupo trata-se de Engenharia de Requisitos, uma das fases mais críticas do ciclo de desenvolvimento de software. Exige que os clientes pensem sobre suas necessidades com cuidado e o objetivo da fase de Engenharia de Requisitos é produzir uma especificação completa sobre o software para resolver o problema do cliente. O grupo possui 15 teses de doutorado defendidas e 134 dissertações concluídas.

5.2. Como os indivíduos entendem um problema de doutorado e a evolução do problema

Antes de cada entrevista começar foi dada uma explicação sobre o propósito da pesquisa e solicitada concordância dos entrevistados para realização da gravação. Inicialmente abordamos o entendimento dos entrevistados sobre o que é um problema de doutorado e como este problema evolui. No roteiro de entrevista dos alunos temos três perguntas relacionadas a este tema (Q3, Q4 e Q5 - Quadro 1) e no roteiro dos orientadores temos duas perguntas (Q1 e Q2 - Quadro 2).

Não encontramos uma definição clara na literatura para problema de pesquisa de doutorado, nem uma diferenciação do que seria um problema adequado. Assim sendo, é compreensível que doutorandos ou mesmo recém-doutores tenham dificuldade de responder a este questionamento. Nos discursos de dois alunos participantes podemos ter esta constatação:

— “O que é um problema de pesquisa... [Longo momento de pausa] Descobrir... É ... às vezes tem um problema já existente na indústria ou na academia e outra coisa é encontrar um problema.” (Entrevistado 3)

— “Um problema de pesquisa é uma lacuna que ficou de pesquisas anteriores ou um problema novo que você decide investigar.” (Entrevistado 4)

Vale a pena destacar no discurso do Entrevistado 3 o fato de ter mencionado esta questão de problemas na indústria. Este olhar para a indústria é uma característica cada vez presente nas pesquisas de computação, seja de modo a embasar/entender um fenômeno ou validar uma técnica proposta.

Um dos alunos entrevistados conseguiu conceituar problema de pesquisa de doutorado de forma consistente aos conceitos dados pelos orientadores entrevistados.

— “Um problema de doutorado tem que ser relevante, um problema real e que tenha uma solução original. Não tem que ser totalmente construído do zero, pois pode partir de uma pesquisa já realizada.” (Entrevistado 6)

Vejamos discursos de alguns orientadores entrevistados.

— “É um problema que a gente consegue ter evidências de que existe e consegue mensurar sua relevância com base em referências. E ainda tem outra questão, que é a originalidade, que é outra dificuldade. Tem que mostrar que é algo que ninguém propôs ou mostrar que a sua solução é melhor.” (Entrevistado 7)

Um dos professores foi além da definição dada pelos demais (do problema de definição geral, único e relevante) e destacou que muitos programas de alto nível pelo mundo aceitam um problema de doutorado se o aluno consegue publicar em conferências consolidadas:

— “Seu doutorado é doutorado se você publicar um paper em certos lugares, inclusive vários programas são assim... Se você tiver três papers, você não precisa escrever tese, são três papers e pronto. Vários programas de alto nível são assim.” (Entrevistado 8)

A pergunta seguinte diz respeito à descrição do problema de pesquisa por parte dos alunos. Cada um conseguiu descrever de forma clara e precisa seu problema de pesquisa, sendo que dois dos entrevistados ainda encontram-se no processo de evolução do problema, uma vez que estão com um pouco mais de dois anos de doutorado.

5.3. Categorias Centrais da Evolução do Problema de Doutorado

Depois de analisar o entendimento dos participantes em relação ao conceito de problema de doutorado, o próximo passo foi identificar os fatores que afetam positivamente e negativamente a evolução do problema de pesquisa. Nesta subseção os códigos foram agrupados em fatores que influenciam a evolução do problema de pesquisa. Foram usados trechos de entrevistas com evidências relacionadas a cada categoria. A seguir apresentaremos as três categorias centrais.

1) O Conhecimento Prévio do aluno representa uma categoria central dentro do problema de pesquisa se relacionando com algumas categorias tais como: estágio, publicações no mestrado e continuidade na área e com o mesmo orientador, iniciação científica, intercâmbio, monitoria, conhecer a área de pesquisa na graduação. A colaboração com grupo no exterior faz parte do background, pois a realização de intercâmbio durante a graduação ou mestrado é uma das formas consideradas de realizar esta colaboração. Alguns trechos apresentados a seguir estão relacionados ao conhecimento prévio.

— “Além de assistir aula eu fiz um monte de atividades. Particpei da empresa júnior daqui.... Ai depois, quando tava já perto de sair da empresa jr eu entrei no PET aqui do CIN, né... Nos últimos três semestres (graduação) aí, foi que comecei a entrar na área de requisitos... Estagiei no CESAR, trabalhando com TV digital.” (Entrevistado 3)

— “No doutorado eu continuei com o tema. Isso permitiu o que? Que eu tivesse resultados no início do doutorado. Então não tinha que passar um tempão procurando...” (Entrevistado 4).

Publicar artigos científicos é muito importante para os programas de pós-graduação, para os orientadores e principalmente para os estudantes. A publicação é uma oportunidade de receber um feedback da comunidade acadêmica sobre seu trabalho. Esta categoria pode ser constatada nos trechos transcritos abaixo.

— “Peguei um probleminha e uma publicação e depois outro probleminha e outra publicação e assim foi. Até encontrar um problema maior que desse conta de ser o meu trabalho de doutorado.” (Entrevistado 3)

— “...outra coisa é você ter publicações que demonstrem a evolução do seu tema. Quando conseguir definir seu problema de doutorado, manda logo para um fórum de doutorado, só para avaliar. À medida que você vai evoluindo é interessante que tenham esses marcos representados por artigos que você submeteu, ajuda a você a se alto regular, você tá cumprindo prazo e tendo feedback, você vai construindo...” (Entrevistado 7).

Uma característica bem interessante da área de computação é a existência de workshops de teses e dissertações, como mencionado pelo Entrevistado 7. Estes são importantes veículos para se obter um feedback de pesquisadores experientes da área quanto a relevância do problema a ser atacado.

2) O Acompanhamento Sistemático dos Resultados é alcançado por meio de pontos de checagem como os seminários realizados semestralmente pelo grupo LER, a qualificação, reuniões periódicas com o orientador ou co-orientador e por publicações (publicações no doutorado), o que permite uma orientação mais objetiva e focada em resultados concretos e incrementais. A experiência do orientador pode fazer com que ele

consiga direcionar melhor o aluno durante os acompanhamentos, tendo impacto na evolução por meio do acompanhamento sistemático dos resultados. O doutorado sanduíche (intercâmbio) pode contribuir com o acompanhamento dos resultados, pois envolve um coorientador, o que é um fato capaz de aumentar a cobrança e o feedback.

Vale ressaltar ainda que participar de grupo de pesquisa é sinônimo de ter contato mais próximo com diversos professores e alunos, permitindo um maior acompanhamento da pesquisa. Quanto mais o aluno tem debates sobre seu tema, mais ele estimula a visão crítica e a capacidade de argumentar e defender ideias, que são fundamentais para o desenvolvimento da pesquisa. Recortes das entrevistas evidenciam o impacto do acompanhamento sistemático como fator na evolução do problema:

— “Participar de grupos de pesquisa é muito importante. Um doutorado sanduíche é interessante desde que seja em um lugar que se trabalhe no contexto de sua pesquisa. Realizar publicações e fazer a qualificação o mais cedo possível é muito importante para receber feedback e fazer os ajustes necessários.” (Entrevistado 6).

— “O mais interessante foi que eu pude fazer um intercâmbio, e foi aí que eu pude explorar vários pontos de computação autônoma. Aí quando eu voltei para o CIN tinha mais teoria, conhecia mais trabalhos... Fiquei mais focado. No final do doutorado eu fui para Trento que o pessoal era especialista nisso, tem projetos nesta área... E lá eles me direcionaram bastante no meu problema”. (Entrevistado 3)

Este acompanhamento sistemático aumenta a pressão por resultados que o aluno deve apresentar. O estresse gerado por essa cobrança exige uma maturidade emocional do aluno, percebe-se uma relação entre as categorias Acompanhamento Sistemático de Resultados e Maturidade Emocional, uma influenciando diretamente a outra.

3) Maturidade Emocional foi um fator identificado nas transcrições como sendo relacionado a inteligência emocional e também bastante relacionado a capacidade de reagir com equilíbrio diante de fatores externos como nascimento de filho, ir trabalhar em outra cidade, perder emprego, perder um parente, separação. A Dedicção, Autonomia e Iniciativa foram identificadas nos discursos e por se tratarem de aspectos subjetivos individuais também podem ser categorizados como Maturidade Emocional e tem uma influência positiva. A Flexibilidade em relação ao problema de pesquisa foi mencionada como um fator positivo relacionado a maturidade do aluno.

Não há um consenso entre os entrevistados em relação à influência da dedicação, como pode ser verificado nos trechos abaixo.

— “Ao mesmo tempo estava como professora em outra cidade e não conseguia me dedicar bem ao Doutorado.” (Entrevistado 2).

— “Tudo depende do perfil do aluno, tem aluno que funciona melhor trabalhando e estudando. Depende do aluno é muito da personalidade da pessoa, isso daí de ser dedicação exclusiva é do aluno. O aluno tem que ver com o orientador”(Entrevistado 7).

— “Aí eu acho que fator número um é dedicação, se o cara não se dedicar o mínimo que são 40 por semana no doutorado por 4 anos, é impossível fazer.... eu acho que o número um é dedicação” (Entrevistado 8).

A autonomia e iniciativa do aluno também são fatores determinantes para a evolução do problema de pesquisa. Os trechos a seguir corroboram com essa afirmação.

— “eu decidi que eu queria trabalhar com elicitação de requisitos então, foi quando eu propus para o meu orientador que queria trabalhar com elicitação” (Entrevistado 1).

— “o aluno tem que ter maturidade para não se perder em meio as muitas opiniões.” (Entrevistado 7).

A Maturidade Emocional é um fator que influencia positivamente na evolução do problema de pesquisa durante o Doutorado. Quando o aluno é instável emocionalmente, o resultado para a evolução do problema de pesquisa é muito negativo.

5.4. Construindo uma História Central

As ideias apresentadas nas subseções anteriores foram relacionadas de forma a construir a história central que apresenta como evoluem os problemas de doutorado no LER e os fatores relacionados.

A evolução dos trabalhos de doutorado ocorre de formas bastante distintas no LER. Um dos entrevistados caracteriza a forma como seu problema de doutorado evoluiu como sendo uma abordagem incremental, na qual um pequeno problema foi encontrado e resolvido resultando em uma publicação e, posteriormente, problemas relacionados a este primeiro foram sendo acoplados, resolvidos e publicados até ter o resultado final que foi a tese. Outros dois alunos citam a evolução de forma um pouco diferente deste primeiro, onde o problema foi definido em um determinado momento, sendo um deles no início do doutorado e outro no decorrer dele, e que este problema foi sendo resolvido ao longo do doutorado. Já os dois alunos restantes tiveram dificuldade com a definição do problema de pesquisa, por isso, mudaram de tema e de orientação.

Dependendo do perfil do aluno, haver uma flexibilidade em relação ao problema pode ser algo extremamente positivo, uma vez que alunos flexíveis tem a possibilidade de abandonar problemas inviáveis ou que não são relevantes para uma pesquisa de doutorado. Ter equilíbrio emocional, para saber lidar com os momentos de alta produtividade e com os de desestímulo, pode ser um diferencial que juntamente com a postura pró-ativa podem compensar a ausência de conhecimento prévio do aluno e fatores externos que atrapalhem. Os fatores externos como problemas pessoais, nascimento de filho, ir trabalhar em outra cidade e problemas de saúde na família, podem atrapalhar a evolução do problema de pesquisa, uma vez que tem influência direta sobre o fator psicológico e a inteligência emocional.

A motivação do aluno com relação ao trabalho de pesquisa que está sendo realizado é fator determinante para o sucesso da pesquisa, pois junto a maturidade emocional o aluno saberá superar essas adversidades de forma equilibrada. O fato de computação ser uma ciência relativamente nova, onde novas linhas de pesquisa são fomentadas constantemente, também pode influenciar na evolução.

Outro aspecto importante é o acompanhamento sistemático dos resultados do aluno representado por meio de publicações no doutorado e pontos de checagem. A participação em workshops de teses e dissertações pode ser bastante útil no momento de definição do problema de pesquisa. A influência do orientador é determinante neste aspecto uma vez que sua experiência, pode levá-lo a identificar se o problema que está sendo definido pode ter desdobramentos para um doutorado e fazer sugestões de alterações/mudança no problema. A realização de doutorado sanduíche (Colaboração com grupo no exterior) foi citada tanto de forma positiva quanto de forma negativa.

Foram mencionados vários benefícios para a evolução do problema de pesquisa como o acompanhamento do co-orientador do país de destino e contato com alunos de outros grupos. No entanto, a questão da adaptação foi identificada como tendo uma influência negativo associada a este fator. Apesar disto, o saldo é extremamente positivo. O período de colaboração com um grupo de pesquisa no exterior foi realizado na graduação ou no mestrado, chamado por alguns entrevistados de intercâmbio, contribui para o conhecimento prévio do aluno, que também é um dos fatores que influenciam na definição do problema de doutorado. Quando é realizado na forma de doutorado sanduíche, propicia o acompanhamento de um co-orientador, bem como o foco passa a ser a publicação de artigos. Além disso, ajuda na validação do problema de pesquisa.

Há um entendimento de que atividades realizadas na graduação como iniciação científica, atividades de extensão, estágio, monitoria, colaboração com grupo no exterior e a forma como o mestrado do aluno aconteceu (se houveram publicações no mestrado, por exemplo), podem fazer com que o aluno chegue ao doutorado com um melhor amadurecimento em pesquisa. Do mesmo modo, a continuidade na área de pesquisa e com o mesmo orientador podem influenciar positivamente, uma vez que o aluno consegue iniciar o doutorado com uma fase de adaptação mais branda. As disciplinas de metodologias de pesquisa são fundamentais para a evolução do problema de pesquisa. Na Figura 1, abaixo, procuramos sintetizar essas relações.

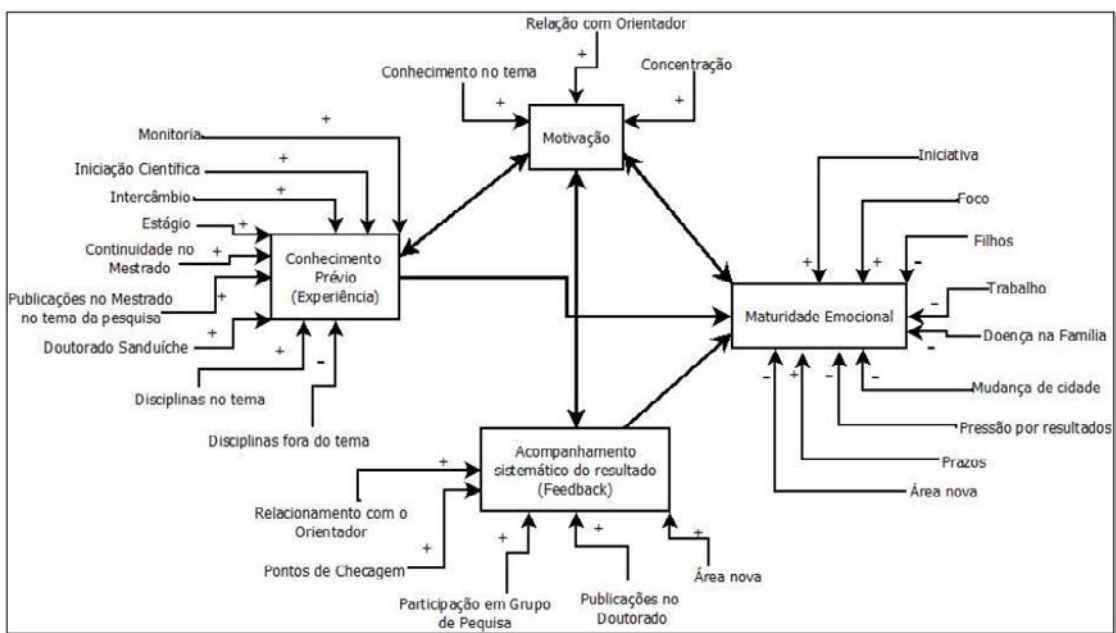


Figura 1. Núcleo de Fatores e seus Relacionamentos

5.5. Usando Resultados para Melhorar a Evolução dos Problemas de Doutorado

Nesta subseção discutimos como o resultado desta análise qualitativa pode ser utilizado para melhorar a forma como os problemas de doutorado em computação evoluem. Contudo, estas recomendações ainda não foram testadas na prática, assim elas somente definem estratégias derivadas dos resultados desta pesquisa.

O período de doutorado envolve, em média, quatro anos e existe uma grande possibilidade que algum tipo de fator externo venha acontecer e impactar na evolução do problema de pesquisa. Infelizmente, parte dos fatores externos não tem como ser

evitados, mas recomenda-se atenção com aqueles possíveis de se mitigar. A motivação com relação a pesquisa e a maturidade emocional do aluno ajudam a superar os obstáculos. Em algumas situações, a ajuda de um e psicanalista pode ser útil.

O conhecimento prévio no tema da pesquisa é um fator de sucesso para o doutorado. A formação prévia é importante, ter feito uma graduação e um mestrado de forma dedicada ajudam a ter uma base sólida para uma evolução do problema de pesquisa. Apoio do orientador, co-orientador ou grupo de pesquisa é de grande valia para validação da pesquisa que está sendo realizada. As publicações realizadas e a participação em grupos de pesquisa facilitam a evolução do problema de pesquisa.

6. CONCLUSÕES

Neste artigo, apresentamos os principais resultados de uma pesquisa qualitativa básica sobre a evolução do problema de pesquisa no doutorado de no grupo LER, vinculado ao Programa de Pós-Graduação em Computação da UFPE.

Com os resultados obtidos neste trabalho, mostramos uma teoria que explica como ocorre a evolução de um problema de pesquisa e a complexa interação entre os fatores positivos e negativos que influenciam neste processo. Embora os resultados da pesquisa qualitativa apresentado neste artigo não podem ser generalizados, os princípios centrais da teoria e do método de pesquisa podem ajudar outros pesquisadores a reinterpretar a teoria em contextos específicos.

Como trabalho futuro, pretendemos replicar este estudo em outros grupos do programa de pós-graduação da UFPE para ampliar o entendimento sobre a evolução dos problemas de doutorado na área de Computação.

Referências

- Duque, J. C., Brondani, J. T., Luna, S. P. L. Estresse e pós-graduação em medicina veterinária. RBPG, v.2, n. 3, p.134-148, 2005.
- GEOCAPES. Ministério da Educação. Disponível em <http://geocapes.capes.gov.br/geocapes2/> Acessado em 23 de Novembro de 2015.
- Godoy, A., da Silva, E. E., Figueiredo, R. C. Efeitos do Estágio Docente na Formação do Pós-Graduando. Workshop de Educação em Informática (WEI), Natal-RN, 2011.
- Goleman. D. Inteligência Emocional. A teoria revolucionária que redefine o que é ser inteligente. Ed. Objetiva. 1996.
- Merriam, S. B. Qualitative research: a guide to design and implementation. 2009.
- Neta, N. F. A.; García, E. G.; Gargallo, I. S. A inteligência emocional no âmbito acadêmico: uma aproximação teórica e empírica. Psicol. Argum. 2008.
- Salveti, M., Bueno, M.; Gastaldo, D., Kimura, A., Pimenta, C. Doutorado Sanduíche: Considerações para uma experiência de sucesso. Rev. Gaúcha Enferm., 2013.
- Strauss, A., Corbin, J. M. Basics of Qualitative Research: Second Edition: Techniques and Procedures for Developing Grounded Theory. Sage Publications, 2007.
- Stump, S., Knihs, E., Omar, N., Mustaro, P., Silveira, I. Discutindo a Construção do Pensamento em Pós-Graduandos. Workshop de Educação em Informática, 2011.

Hall of Fame/Shame: um Padrão Pedagógico para o Ensino de Programação

Andrea S. Charão^{1,2}, Alberto F. Kummer Neto²
Benhur de O. Stein^{1,2}, Patrícia P. de A. Barcelos¹

¹ Departamento de Linguagens e Sistemas de Computação

² Programa de Pós-Graduação em Informática
Universidade Federal de Santa Maria
Santa Maria, RS, Brasil

{andrea, alberto, benhur, pitthan}@inf.ufsm.br

Resumo. Padrões pedagógicos visam aproveitar o conhecimento especializado em práticas de ensino e aprendizagem, em um formato que favoreça sua reutilização. Neste trabalho, propõe-se um padrão pedagógico voltado ao ensino de programação, em cursos de nível superior na área de Computação. Este padrão é centrado na exposição e discussão de bons e maus exemplos de códigos produzidos pelos alunos, formando o que denominou-se respectivamente de “Hall of Fame” e “Hall of Shame”. Ao longo do texto, apresenta-se uma caracterização do padrão, relacionando-o com outros padrões pedagógicos descritos na literatura, e descreve-se sua aplicação no ensino de programação orientada a objetos, em uma instituição de ensino superior. Os resultados indicam uma avaliação positiva do padrão pelos alunos e revelam sua eficácia em incentivar boas práticas de programação.

Abstract. Pedagogical patterns propose to take advantage of the expertise in teaching and learning practices, in an organized manner that can be easily reused. In this paper, we propose a pedagogical pattern targeted to teaching programming in higher level courses in Computer Science. This pattern is centered on the presentation and discussion of good and bad examples of codes produced by the students, forming what was named respectively “Hall of Fame” and “Hall of Shame”. Throughout the text, we presents a characterization of HoFS pattern, relating it to other pedagogical patterns described in the literature. We also report its application in teaching object oriented programming in a higher education institution. The results indicate a positive evaluation by students and reveal the pattern helps to encourage good programming practices.

1. Introdução

Os processos de ensino-aprendizagem de programação de computadores são alvo de discussões recorrentes em comunidades de pesquisadores e professores da área de Computação [Pears et al. 2007, Aureliano e Tedesco 2012]. É consenso que a programação impõe várias dificuldades a alunos iniciantes e, aos que já possuem alguma experiência prévia, apresenta-se também o desafio de aprimorar suas habilidades e aprender novos paradigmas [Jenkins 2002, Robins et al. 2003, Lahtinen et al. 2005, Tan et al. 2009]. Aos professores e instituições de ensino, cabe a responsabilidade de permanentemente refletir

sobre as metodologias e as práticas de ensino-aprendizagem de programação, que envolve também a tomada de decisões com impacto em todo o processo [Denning 1989]. Questões frequentes neste contexto incluem, por exemplo: Que linguagens de programação adotar para iniciantes? Que estratégias de ensino-aprendizagem são mais eficazes?

Em buscas de respostas para estas e outras questões, há exemplos de pesquisadores que têm se dedicado a compreender melhor os processos cognitivos envolvidos na aprendizagem de programação [Soloway e Spohrer 1988, Winslow 1996], enquanto outros investigam e propõem formas de lidar com as dificuldades recorrentes [Wilson e Shrock 2001, Caspersen e Kolling 2009]. Nessa segunda direção, há diversos autores que sustentam uma abordagem centrada nos chamados “padrões pedagógicos” (*pedagogical patterns*) [Sharp et al. 1996], originários de uma comunidade interessada no ensino-aprendizagem de programação orientada a objetos, e gradativamente estendidos para um escopo mais amplo [Larson et al. 2008, Köppe 2015]. Essa abordagem propõe-se a capturar boas práticas de ensino-aprendizagem em domínios específicos, sob uma forma compacta e organizada, para que possa ser facilmente comunicada e reutilizada. Seu foco principal é, portanto, facilitar a disseminação de boas práticas.

Desde a proposta inicial, em 1996, vários foram os padrões pedagógicos descritos, catalogados e publicados, com diferentes intenções. No domínio específico de aprendizagem de programação, há padrões variados, como por exemplo: *Mistake* [Bergin 2000], que propõe que os alunos criem programas (ou outros artefatos) com erros; e *Show Programming*, que sugere mostrar programação em tempo real aos alunos, e não somente slides sobre programação [Schmolitzky 2007]. Conforme autores da área [Bergin et al. 2012], alguns padrões podem parecer triviais para educadores experientes, mas se tornam referências úteis para quem tem menos experiência ou busca diversificar suas práticas. A forma sucinta de apresentação dos padrões facilita este compartilhamento; por outro lado, a ausência de dados complementares sobre avaliação de certos padrões pode suscitar dúvidas quanto à sua efetividade.

Neste artigo, apresenta-se um padrão pedagógico voltado ao ensino de programação, em cursos de nível superior na área de Computação. Este padrão, que denominou-se “Hall of Fame/Shame” (HoFS), é centrado na exposição e discussão de bons e maus exemplos de códigos produzidos pelos alunos. Sua motivação é principalmente estimular os alunos a aprimorarem suas habilidades em programação. Além de descrever o padrão de forma sucinta, num formato que facilite o reuso, apresenta-se também resultados experimentais sobre sua aplicação prática numa instituição de ensino superior.

Este artigo encontra-se assim organizado: na seção 2 apresenta-se sucintamente uma visão histórica de padrões pedagógicos no ensino de programação, indicando as principais fontes de referência sobre o assunto, até os dias atuais; na seção 3 apresenta-se o padrão HoFS num formato tipicamente utilizado pela comunidade da área; na seção 4 descreve-se a aplicação e a avaliação do padrão em uma instituição de ensino superior; na seção 5, por fim, apresentam-se considerações finais sobre o trabalho.

2. Padrões Pedagógicos no Ensino de Programação

Padrões pedagógicos (*pedagogical patterns*) [Sharp et al. 1996] surgiram no contexto de discussões sobre ensino de programação orientada a objetos e suas tecnologias. A inspiração para isso veio dos chamados padrões de projeto (*design patterns*) [Gamma et al. 1993],

que constituem soluções bem sucedidas e reusáveis para problemas no contexto da orientação a objetos. A ideia inicial dos padrões pedagógicos era, portanto, usar essa mesma abordagem para catalogar soluções reusáveis a problemas comumente encontrados no ensino desse paradigma.

Gradativamente, a ideia dos padrões pedagógicos estendeu-se a um escopo mais amplo, capturando soluções para problemas encontrados em diferentes situações e domínios além da orientação a objetos. Pesquisadores e educadores passaram a propor padrões em conferências tais como PLoP (Pattern Languages of Programs) e EuroPLoP (European Conference on Pattern Languages of Programs), chegando a conferências tradicionais sobre ensino de computação, tais como SIGCSE (ACM Technical Symposium on Computing Science Education) e ITiCSE (ACM Conference on Innovation and Technology in Computer Science Education). Os padrões foram sendo catalogados em sites¹² e, mais recentemente, vários deles foram reunidos em livros [Bergin et al. 2012, Mor et al. 2014].

No cenário brasileiro, um dos primeiros trabalhos que encontramos sobre padrões pedagógicos foi uma dissertação de mestrado [de Oliveira Neto 2000], seguida mais adiante por publicações no Simpósio Brasileiro de Informática na Educação [de Barros et al. 2004, Medeiros et al. 2007]. No Workshop de Educação em Computação, os padrões pedagógicos também são abordados em alguns trabalhos [de Barros e Delgado 2006, dos Santos Júnior et al. 2009]. De forma geral, nota-se que esses trabalhos em âmbito nacional não propõem ou avaliam padrões, mas sim ferramentas ou estratégias de apoio ao aprendizado com base em padrões pedagógicos.

Um aspecto importante sobre os padrões pedagógicos é seu formato de descrição, que é sempre sucinto e organizado para facilitar o reuso. Esse formato não é rígido, mas comumente inclui as seções: nome do padrão, objetivo/motivação, aplicabilidade, estrutura, consequências, implementação, recursos necessários, exemplos e padrões relacionados [Sharp et al. 1996]. Alguns autores organizam a descrição em menos seções, porém mantendo essas informações básicas [Schmolitzky 2007]. Para facilitar a comunicação, a redação do padrão geralmente usa o pronome “você” para dirigir-se diretamente ao professor.

3. O Padrão “Hall of Fame/Shame”

O padrão proposto é descrito a seguir, usando um formato semelhante ao utilizado por [Bergin 2000].

- **Nome:** HALL OF FAME/SHAME
- **Problema:** Alunos que vencem barreiras iniciais no aprendizado de programação ficam satisfeitos quando conseguem resolver novos problemas e seus programas funcionam. No entanto, esses alunos não costumam ser críticos quanto a seus códigos, ou julgam-se incapazes de produzir algo melhor no tempo disponível. Muitas vezes, seus programas são resultado de tentativas e erros, revelando más práticas de programação que são repetidas sucessivamente. Os alunos recebem orientações sobre boas práticas, porém não as relacionam com os problemas que devem resolver. O *feedback* fornecido para esses alunos ocorre muitas vezes sob

¹<http://www.pedagogicalpatterns.org>

²<http://educationalpatterns.org>

- forma de uma nota, um comentário ou um gabarito, que não estimulam o aluno a vislumbrar diferentes alternativas de soluções, algumas melhores que outras.
- **Contexto:** Você está ensinando um novo tópico de programação a alunos que já adquiriram alguma experiência com uma linguagem (tipicamente, alunos de segundo ano de cursos de graduação em Computação). Por exemplo, podem ser novas metodologias e técnicas de programação, um novo paradigma ou uma nova linguagem. Você requer que os alunos produzam código, de diferentes tamanhos, para resolver problemas propostos. Seu grupo de alunos tem de 20 a 40 indivíduos (não é uma turma muito pequena, nem muito grande). Você deseja estimulá-los a conhecer e adotar boas práticas, aproveitando também suas experiências anteriores em programação.
 - **Solução:** Depois que o grupo de alunos entregar suas soluções para um dado problema, analise os códigos buscando identificar boas e más práticas de programação. Preferencialmente, permita que os alunos entreguem soluções parciais antes da entrega final. A cada entrega, escolha trechos de códigos para formar uma exposição, separada em duas “galerias”: *Hall of Fame* (boas práticas) e *Hall of Shame* (más práticas). Em aula, para cada aspecto da solução, apresente e comente os trechos em ambas as galerias, sem revelar os autores, discutindo com os alunos os motivos para classificar cada trecho em uma ou outra galeria. Apresente também trechos de código sem classificação e discuta com os alunos o enquadramento como *Fame* ou *Shame*. Use recursos visuais contrastantes para identificar ambas as galerias. Ressalte que, num mesmo programa, é possível encontrar exemplos de boas e más práticas, reforçando que a galeria não é de bons ou maus programadores, mas sim de trechos de código com bons e maus exemplos. Faça com que a galeria fique disponível para consultas futuras. Repita este padrão sempre que possível.
 - **Discussão:** Este padrão provê *feedback* aos alunos com base em problemas que eles se dedicaram a resolver. Ao contrário de recomendações gerais e abstratas sobre boas práticas, o padrão garante exemplos reais, provenientes do próprio grupo de alunos. Com a separação clara e repetitiva entre *Fame/Shame*, a intenção é estimular processos cognitivos que, nas próximas atividades de programação, resultem na adoção de uma boa prática ou, no mínimo, alimentem o senso crítico dos alunos frente a seus programas. Este padrão não visa substituir uma avaliação individual dos programas (que, por outro lado, nem sempre é possível), mas constitui uma alternativa para prover *feedback* de forma rápida e proveitosa para o grupo. O padrão tem também limitações, a saber: (i) é necessário um tempo considerável para preparar e expor/comentar as galerias, imediatamente após a entrega das soluções pelos alunos; (ii) em grupos de alunos com experiências muito homogêneas, a diversidade de exemplos produzidos pode ser pequena.
 - **Recursos necessários:** Para exibição das galerias, é necessário projetor multimídia para uso em aula, além de infraestrutura para disponibilizar o material para consulta futura. Os problemas propostos aos alunos devem ser especificados de tal forma a dar margem suficiente para que se produzam soluções diversificadas. O apoio de um monitor ou assistente é recomendável na preparação das galerias, para que a exibição ocorra o mais rápido possível, idealmente em um encontro subsequente à entrega das soluções.
 - **Exemplos:** Um exemplo de aplicação do HoFS é no ensino de programação procedimental em linguagem C, depois que os alunos já souberem estruturar o

código em funções e procedimentos. Neste ponto, é comum que os alunos tenham que programar um jogo simples, ou um editor de textos. A galeria *Shame*, por exemplo, possivelmente irá revelar trechos de código com os seguintes problemas: divisão do código em poucos subprogramas, procedimentos que realizam mais ou menos instruções do que o nome sugere, procedimentos que acessam variáveis globais quando deveriam se ater a seus argumentos, código mal endentado, etc. Na galeria *Fame*, ficarão os trechos de código com características positivas, opostas a essas.

Outro exemplo de aplicação é no ensino de programação funcional, em linguagens Haskell ou Lisp. Habitualmente, propõem-se que os alunos resolvam vários tipos de problemas com listas, envolvendo percursos recursivos ou via funções de ordem superior (*higher order functions*). Quando os alunos já conhecem ambas as alternativas, a galeria *Shame* poderá apontar, por exemplo, trechos de código que usaram recursão explícita para implementar o que seria mais facilmente escrito e reusável usando funções de ordem superior, como `map` ou `filter`.

- **Padrões relacionados:** A aplicação do padrão HoFS supõe que os alunos devam produzir código para implementar algum software, mesmo que parcialmente. Assim, padrões relacionados ao ensino de desenvolvimento de software podem ser aliados no processo de ensino-aprendizagem [Schmolitzky 2007]. Além disso, o padrão HoFS é centrado na ideia de que os alunos se beneficiam de *feedback* frequente para progredirem e desenvolverem suas habilidades em programação. Desta forma, outros padrões pedagógicos que também focam em *feedback* contínuo [Larson et al. 2008, Köppe et al. 2015] podem complementar a aplicação do HoFS.

4. Aplicação e Avaliação do HoFS

O padrão aqui apresentado foi aplicado em 4 semestres de uma disciplina que trata de Programação Orientada a Objetos (POO), oferecida na Universidade Federal de Santa Maria, para alunos cursando a partir do terceiro semestre dos cursos de graduação em Paradigmas de Programação. Essa disciplina proporciona um primeiro contato com POO aos alunos e, para isso, adota a linguagem Java. Observou-se, nas primeiras instâncias de aplicação, que o padrão pareceu ter aceitação positiva por parte dos alunos, evidenciada principalmente por: (i) uma maior atenção e participação dos alunos em sala de aula durante a exposição dos códigos, em contraste com exposições de conceitos ou trechos de código do professor e (ii) interesse de alguns alunos em saber se seus códigos entrariam ou não para as galerias.

Diante dessas observações, na última oferta da disciplina buscou-se realizar uma avaliação do padrão HoFS com critérios e procedimentos previamente estabelecidos, visando coletar mais dados sobre seu impacto no aprendizado de POO. O restante desta seção se refere à avaliação do HoFS aplicado à última turma.

Na literatura sobre padrões pedagógicos, não se encontram critérios e procedimentos sistematicamente utilizados para avaliar padrões propostos. Assim, decidiu-se avaliar 2 aspectos complementares que, em nossa experiência, poderiam fornecer indicadores sobre a eficácia (ou não) do padrão. Um desses aspectos foi a **aceitação** da abordagem por parte dos alunos, avaliada por meio de um questionário respondido ao final da disciplina. Outro aspecto foi a **produção** de código pelos alunos, contendo bons e/ou maus exemplos de práticas de programação, avaliada pela análise continuada do código produzido para

resolver problemas propostos. Entende-se que esses aspectos são complementares pois, por exemplo, os alunos poderiam avaliar positivamente a experiência, mas suas produções de código não revelarem melhoria de suas práticas.

4.1. Avaliação da aceitação: método e resultados

Para avaliar a aceitação, elaborou-se um questionário dividido em 2 partes, cada parte apresentada sob forma de itens da escala de Likert de 5 pontos, variando de -2 (discordo totalmente) a +2 (concordo totalmente). A primeira parte continha 5 questões relacionadas à percepção geral do padrão pelos alunos, relacionando-o com outras experiências de aprendizado. A segunda parte apresentava 4 questões sobre o formato utilizado para exibição e discussão das galerias, que compreendeu aulas expositivas e participativas usando *slides* com trechos de códigos, seguida de publicação do material *online* para consultas futuras (conforme apresentado na seção 3). Nessa segunda parte, buscou-se levantar a necessidade de possíveis ajustes na aplicação do padrão.

O questionário foi respondido anonimamente por 15 alunos, num total de 22 que cursaram a disciplina até o final. Esses alunos tiveram aulas teóricas e práticas, realizaram exercícios e desenvolveram 2 trabalhos/projetos individuais de POO. O padrão HoFS foi aplicado em 3 oportunidades: (1) após a entrega do primeiro trabalho, (2) após uma entrega parcial do segundo trabalho (50% dos requisitos implementados) e (3) após a entrega final do segundo trabalho. A avaliação dos alunos na disciplina considerou seu desempenho nos trabalhos e em uma avaliação escrita. O questionário foi apresentado após a divulgação de notas, ao fim do semestre.

A figura 1 apresenta os resultados obtidos na primeira parte do questionário. Pode-se notar que uma das questões continha uma proposição com viés negativo (“Achei ruim ver meus erros expostos, mesmo anonimamente”). De forma geral, as respostas dos alunos reforçaram a hipótese inicial sobre a boa aceitação da abordagem. Além disso, os alunos perceberam um caráter original na abordagem, ao não concordarem com a afirmação “Já tive aulas assim em outras disciplinas”.

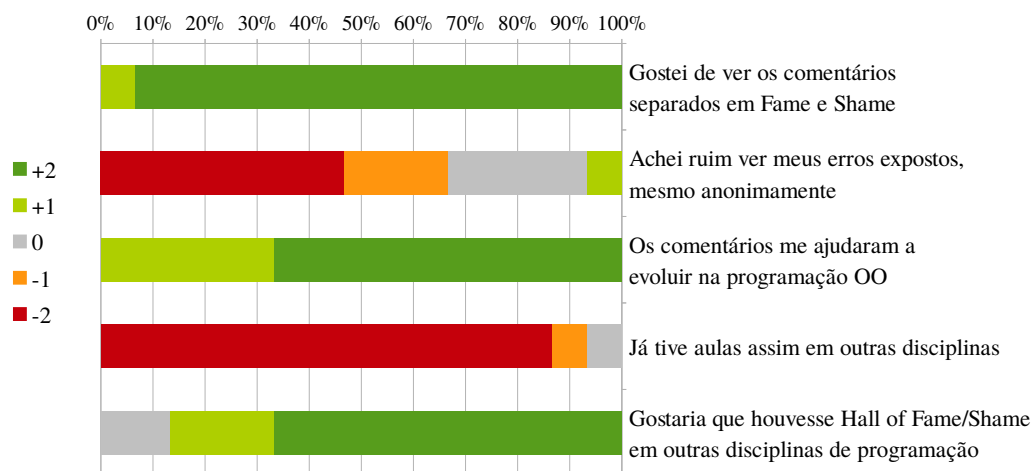


Figura 1. Avaliação da percepção geral do padrão pelos alunos

Na figura 2, apresenta-se as respostas para as questões sobre o formato de aplicação do padrão. Neste quesito, de forma geral, nota-se também que houve boa aceitação do

formato, porém alguns itens merecem mais atenção: (1) os comentários *Fame/Shame* antes da entrega final de um trabalho foram entendidos como mais úteis do que aqueles apresentados após a entrega; (2) muitos alunos responderam “Concordo parcialmente” ao item “O material disponibilizado para consultas posteriores foi suficiente”. O item (1) sugere que, quando possível, o padrão seja aplicado enquanto os alunos ainda estão programando suas soluções. O item (2), por sua vez, revela um possível descompasso entre o aproveitamento das galerias durante e após as aulas. Os motivos disso podem ser investigados em novas experiências mas, como alternativa imediata, estima-se que a disponibilização das galerias em formato de vídeo poderia reduzir esse descompasso.

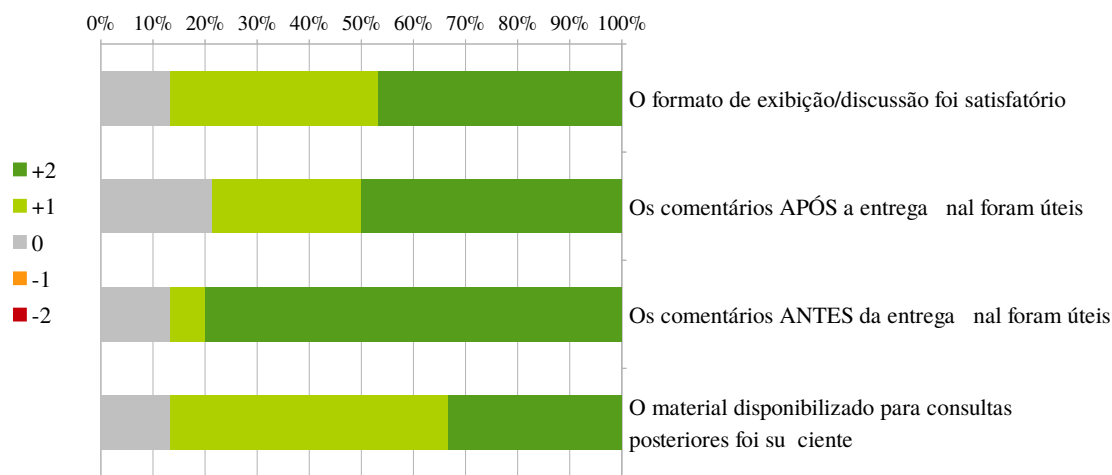


Figura 2. Avaliação do formato adotado para as aulas que seguiram o padrão HoFS

4.2. Avaliação da produção: método e resultados

Para avaliar o impacto do HoFS na produção dos alunos, considerou-se 2 trabalhos/projetos de POO desenvolvidos pelos alunos. Os programas referentes a esses projetos foram criados pelos alunos no prazo de 1 e 3 semanas, respectivamente, e entregues ao professor em 3 oportunidades, conforme explicado na seção anterior.

O primeiro projeto tinha como tema a simulação de entradas e saídas de veículos num estacionamento, compreendendo o cálculo do valor a ser cobrado para cada veículo. Esse projeto não explicitava requisitos quanto à interface com o usuário, enfatizando principalmente o projeto e implementação orientados a objetos para o domínio em questão. O padrão HoFS foi aplicado após a entrega final deste trabalho.

O segundo projeto, por sua vez, consistiu em desenvolver uma aplicação com interface gráfica para *desktop*, em Java, visando gerenciar um cadastro de postos de combustíveis em um município, com seus respectivos históricos de preços. O projeto deveria utilizar o padrão MVC (*Model-View-Controller*) e persistir os dados em arquivo e/ou banco de dados. Para este trabalho, aplicou-se o padrão HoFS antes e após a entrega final.

Para análise da produção dos alunos, considerou-se todos os códigos entregues pelos alunos nos 3 momentos descritos anteriormente, num total de 17 e 20 projetos entregues, respectivamente, para o primeiro e segundo trabalhos. Para cada exposição

Fame e Shame em aula, escolheu-se em média 15 trechos de código para discussão, entre bons e maus exemplos, versando sobre diversos aspectos da POO.

Pelos temas dos projetos, há boas práticas de POO que não eram necessárias nos 2 trabalhos (por exemplo, classes de acesso aos dados persistidos em arquivo ou banco de dados). Por outro lado, algumas práticas relacionadas aos fundamentos da POO deveriam estar presentes em todos os trabalhos. Nessa situação, a hipótese era de que o padrão HoFS poderia estimular os alunos a corrigir suas más práticas, entre um trabalho e outro. Assim, dentre todas as boas e más práticas identificadas nos trabalhos, escolheu-se 3 más práticas cujas ocorrências foram contabilizadas nos códigos **antes** do primeiro HoFS (isto é, no primeiro trabalho entregue) e **depois** do último HoFS (isto é, no último trabalho entregue). Essas práticas foram: (1) duplicação de código ao invés de polimorfismo, (2) violação do princípio da responsabilidade única na implementação de uma classe e (3) visibilidade pública para atributos de instância.

Na tabela 1, apresenta-se as ocorrências identificadas, sendo que um aluno pode ser responsável por mais de uma ocorrência. Pode-se notar que as ocorrências destas más práticas diminuíram, como esperado. Sabe-se que os processos de ensino-aprendizagem são por natureza complexos e, por isso, não se pode afirmar com certeza que a diminuição de más práticas seja consequência unicamente da aplicação do HoFS. Mesmo assim, interpreta-se estes resultados como indícios de que o padrão pode ser eficaz. Mais importante do que isso é o evidente progresso demonstrado pelos alunos e estimulado pelo HoFS.

Má prática	Ocorrências antes	Ocorrências depois
Duplicação de código	5	1
Violação do princípio da responsabilidade única	10	2
Visibilidade pública para atributos de instância	3	0

Tabela 1. Ocorrências de más práticas nos códigos antes e depois da aplicação do HoFS

5. Considerações Finais

Neste trabalho, buscou-se caracterizar um padrão pedagógico focado no ensino de programação. Esse padrão relaciona-se com outros voltados ao *feedback* contínuo, porém difere desses à medida em que se fundamenta essencialmente na análise e exibição de trechos de código que representam boas e más práticas de programação.

O padrão foi aplicado numa instituição de ensino superior e, na última instância de aplicação, realizou-se uma avaliação baseada na aceitação do padrão pelos alunos e na análise de códigos por eles produzidos. Os resultados apontam para a eficácia da abordagem e destacam seu caráter original, embora seja necessário repetir o experimento para aferir tais afirmações.

Tem-se consciência de que, isoladamente, um padrão pedagógico não pode garantir o sucesso do processo de ensino-aprendizagem. No entanto, padrões documentados, aplicados e discutidos podem constituir recursos úteis para professores que procuram diversificar e aprimorar suas práticas. Assim, como perspectivas futuras relacionadas a este trabalho, pode-se vislumbrar que mais professores também compartilhem padrões originais e/ou relatem experiências de aplicações de padrões pedagógicos em Computação.

Visando facilitar a reprodução do padrão em outros tempos e espaços, as galerias produzidas e outros materiais utilizados encontram-se disponíveis em: <http://www-usr.inf.ufsm.br/~andrea/elc117-2015b/>.

Referências

- Aureliano, V. e Tedesco, P. (2012). Ensino-aprendizagem de programação para iniciantes: uma revisão sistemática da literatura focada no SBIE e WIE. In *Anais do Simpósio Brasileiro de Informática na Educação (SBIE)*.
- Bergin, J. (2000). Fourteen pedagogical patterns. In *Proceedings of the Fifth European Conference on Pattern Languages of Programs*, Irsee, Germany.
- Bergin, J., Eckstein, J., Volter, M., Sipos, M., Wallingford, E., Marquardt, K., Chandler, J., Sharp, H., e Manns, M. L. (2012). *Pedagogical Patterns: Advice For Educators*. Joseph Bergin Software Tools.
- Caspersen, M. E. e Kolling, M. (2009). Stream: A first programming process. *Trans. Comput. Educ.*, 9(1):4:1–4:29.
- de Barros, L. N. e Delgado, K. V. (2006). Aprendizado de Programação. In *Anais do Workshop sobre Educação em Computação (WEI)*, pages 31 – 40.
- de Barros, L. N., Delgado, K. V., e Machion, A. C. G. (2004). An ITS for programming to explore practical reasoning. In *Anais do Simpósio Brasileiro de Informática na Educação (SBIE)*.
- de Oliveira Neto, J. A. (2000). Suporte de ferramenta de software para o padrão pedagógico aula em mapa de conceitos. Master's thesis, Universidade Federal da Paraíba, Campina Grande.
- Denning, P. J. (1989). A debate on teaching computing science. *Commun. ACM*, 32(12):1397–1414.
- dos Santos Júnior, G. P., Fachine, J. M., e de Barros Costa, E. (2009). Analogus: Um ambiente para auxílio ao ensino de programação orientado pelo raciocínio por analogia. In *Anais do Workshop sobre Educação em Computação (WEI)*, pages 499 – 508.
- Gamma, E., Helm, R., Johnson, R., e Vlissides, J. (1993). Design patterns: Abstraction and reuse in object-oriented designs. In Nierstrasz, O., editor, *Proceedings of ECOOP'93*, Berlin. Springer-Verlag.
- Jenkins, T. (2002). On the difficulty of learning to program. In *Proceedings of the 3rd Annual Conference of the LTSN Centre for Information and Computer Sciences*, pages 53–58.
- Köppe, C. (2015). Towards a pattern language for lecture design: An inventory and categorization of existing lecture-relevant patterns. In *Proceedings of the 18th European Conference on Pattern Languages of Program*, EuroPLoP '13, pages 3:1–3:17, New York, NY, USA. ACM.
- Köppe, C., Portier, M., Bakker, R., e Hoppenbrouwers, S. (2015). Lecture design patterns: More interactivity improvement patterns. In *Preprints of the 22nd Pattern Languages of Programs conference, PLoP*, volume 15.

- Lahtinen, E., Ala-Mutka, K., e Järvinen, H.-M. (2005). A study of the difficulties of novice programmers. In *Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education*, ITiCSE '05, pages 14–18, New York, NY, USA. ACM.
- Larson, K. A., Trees, F. P., e Weaver, D. S. (2008). Continuous feedback pedagogical patterns. In *Proceedings of the 15th Conference on Pattern Languages of Programs*, PLoP '08, pages 12:1–12:14, New York, NY, USA. ACM.
- Medeiros, F. M., Hernández-Domínguez, A., de Medeiros, F. N., e da Silva, A. G. (2007). Um sistema de ensino na web baseado no padrão pedagógico exposição teórica-exemplos-atividade-apresentação-avaliação. In *Anais do Simpósio Brasileiro de Informática na Educação (SBIE)*.
- Mor, Y., Mellar, H., Warburton, S., e Winters, N. (2014). *Practical Design Patterns for Teaching and Learning with Technology*. Sense Publishers.
- Pears, A., Seidman, S., Malmi, L., Mannila, L., Adams, E., Bennedsen, J., Devlin, M., e Paterson, J. (2007). A survey of literature on the teaching of introductory programming. In *Working Group Reports on ITiCSE on Innovation and Technology in Computer Science Education*, ITiCSE-WGR '07, pages 204–223, New York, NY, USA. ACM.
- Robins, A., Rountree, J., e Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, 13:137–172.
- Schmolitzky, A. (2007). Patterns for teaching software in classroom. In *Proceedings of the 12th European Conference on Pattern Languages of Programs (EuroPLoP 2007)*, EuroPLoP '07, pages B5:1–B5:10, Irsee, Germany. Hillside Europe.
- Sharp, H., Manns, M. L., McLaughlin, P., Prieto, M., e Dodani, M. (1996). Pedagogical patterns – successes in teaching object technology: A workshop from OOPSLA '96. *SIGPLAN Not.*, 31(12):18–21.
- Soloway, E. e Spohrer, J. C. (1988). *Studying the Novice Programmer*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.
- Tan, P. H., Ting, C. Y., e Ling, S. W. (2009). Learning difficulties in programming courses: Undergraduates' perspective and perception. In *Computer Technology and Development, 2009. ICCTD '09. International Conference on*, volume 1, pages 42–46.
- Wilson, B. C. e Shrock, S. (2001). Contributing to success in an introductory computer science course: A study of twelve factors. In *Proceedings of the Thirty-second SIGCSE Technical Symposium on Computer Science Education*, SIGCSE '01, pages 184–188, New York, NY, USA. ACM.
- Winslow, L. E. (1996). Programming pedagogy – a psychological overview. *SIGCSE Bulletin*, 28(3):17–22.

Information organization via computational thinking: case study in a primary school classroom*

Simone A. C. Cavalheiro¹, Ana M. Pernas¹, Marilton S. de Aguiar¹, Luciana Foss¹,
André Du Bois¹, Renata H. S. Reiser¹, Clause de Brum Piana¹, Yuri Weissahn¹

¹Technology Development Center - Federal University of Pelotas (UFPEL),
1, Gomes Carneiro St - 96.010-610 - Pelotas, RS, Brazil

Abstract. *Technological innovations have changed the profile of professionals required by economy sectors demanding skills related to problem solving and logical reasoning. Concurrently with this reality, Computational Thinking emerges as a methodology for solving problems, which is able to develop many of these skills. This article reports the experience in applying the activities referred as Binary Numbers, Color by Numbers, Sorting Algorithms and Battleships, which aims at developing some of necessary skills for the practice of Computational Thinking by fourth-graders. The article resumes the activities, and evaluates the obtained results.*

1. Introduction

Computational Thinking (CT) is the idea of using basic concepts of Computer Science (CS) to solve problems, develop systems and to understand human behavior. In a general way, abilities once developed and used to design programs are applied in a more general context, solving general problems from different domains of knowledge [Bundy 2007, Denning 2009]. Wing argues that CT is in fact human thinking instead of machine/mechanized thinking. It consists of conceptualized thinking instead of programmed thinking, i.e., it is a way of thinking instead of a product. More than a specialty of computer scientists [Yinnan and Chaosheng 2012], CT can be perceived as a human capability. In this context, CT is thought as a basic human ability such as reading, writing and solving basic arithmetic problems. CT requires a high level of abstract thinking. As it is a new way to organize thought and solve problems, and as humans are more intelligent and imaginative than computational devices, CT helps us to solve problems once unsolvable before the computing era.

Primary education is an important stage to begin developing CT as children are starting to develop logical and deductive reasoning. Companies like Google and Microsoft are promoting CT in different levels of the educational system. For example, Google is proposing a set of activities to develop CT in primary and secondary education in the USA [Google 2013]. In Brazil, there are also many projects in this direction [França and Amaral 2013, de Francca and Tedesco 2014, Ramos and da Silva Teixeira 2015, Rodrigues et al. 2015, Rodriguez et al. 2015]. Although there are some advances in the area, the research is still in its early stages and far from making CT a regular subject in basic education.

This paper describes our experiences with a pilot project for introducing some basic concepts of CT in Brazilian public schools. More precisely, we discuss the application

*The authors gratefully acknowledge financial support received from PROEXT - MEC/SESu.

of activities based on fundamental concepts of CT as proposed on [CSTA et al. 2010], mainly restricted to two classes and four activities: (i) data and image representation, referred as Binary Numbers (BN) and Colour by Numbers (CN); and (ii) sorting and searching algorithms, Battleships (BS) and Sorting Algorithms (SA). From the results obtained we can conclude that these activities promoted different abilities such as to correlate, count, apply, memorize and sort. The students' achievement assessment was obtained by analysis of pre- and post-tests. Case studies and results of our research are included to illustrate how schools and academia can be integrated to plan for instructional improvement in basic education, especially in the early school years.

The paper is organized as follows. Section 2 presents a brief theoretical framework of CT. In Section 3 the methodological aspects for developing the selected activities is described. Section 4 reports the case study: the application of the activities in two classes of elementary school. Finally, Section 5 the main conclusions and further work are presented.

2. Computational Thinking

The International Society for Technology in Education (ISTE) and the Computer Science Teachers Association (CSTA) have collaborated to develop an operational definition of CT for K-12 education. CT includes (but is not limited to) the following specific skills [ISTE 2013] and [Deng et al. 2009]: (i) reformulating an apparently difficult problem into another one we know the solution by reduction, simulation, or transformation; (ii) thinking recursively and using parallel processing as approaches for data and code interpretation; (iii) using abstraction and decomposition for solving a large complex task or designing a large system via piecewise analysis; (iv) choosing an appropriate representation for a problem or modeling the relevant aspects of a problem to make it tractable; (v) thinking in terms of prevention, protection, and recovery from worst-case scenarios using redundancy, damage containment, and error correction concepts; (vi) using heuristic reasoning to discover a solution – planning, learning, and scheduling in the presence of uncertainty; (vii) using massive amounts of data to speed up solution and computation processes.

ISTE and CSTA supported by National Science Foundation (NSF), in collaboration with leaders from higher education, industry, and K-12 education created an important reference in CT called *Computational Thinking in K-12 Education Leadership Toolkit* [CSTA et al. 2010]. This toolkit provides a theoretical foundation for the process of developing CT and resources for creating a systemic change. In this toolkit, nine fundamental concepts of CS are presented in a progression chart for CT development: data collection, data analysis, data representation, problem decomposition, abstraction, algorithms and procedures, automation, simulation, and parallelization.

Programs and projects focusing on designing activities for the development of CT in distinct levels of education may be cited [Caetano and Oliveira 2015, Google 2013, Carnegie Mellon, Barr and Stephenson 2011, Bell et al. 2010, França and Amaral 2011, Lee et al. 2011, Kafura and Tatar 2011, Repenning et al. 2010, Hambruch et al. 2009]. In Brazil, it is possible to cite different projects aimed at introducing CT in basic education. Just to give some examples, *The Game Design Brasil*, hosted at PUC-RJ, aims to develop a Brazilian technology to stimulate teaching and learning of comput-

3.1. Activity 1 - Binary Numbers

Summary: The objective of this activity is to communicate using binary codes, working with the conversion of binary codes to decimal numbers, further their conversion to letters and the reverse conversion of letters into numbers and binary codes. For that, students are invited to participate in a secret club, where the binary codes are used for communication. The activity is composed of three tasks, each one with concepts, exercises and evaluation.

Educational Material: the activity is presented using boards and each board contains five colored cards (with 1, 2, 4, 8 and 16 black dots on one side). An enlarged version of the material was made for demonstration in class and a smaller version was made for each student (Figure 2). Also, a table was designed associating decimal numbers (1 to 26) to letters (A to Z).

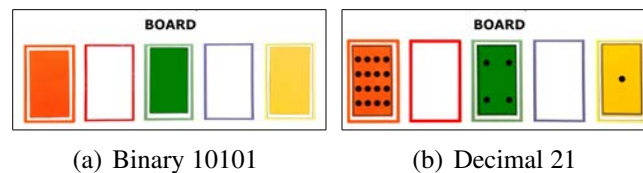


Figure 2. Student' Material

Methodology: In order to convert from binary codes to decimal, we initially present the binary codes, with some examples. Then, each student gets five coloured cards and a board, knowing that each card can only be placed in the board space of the same color. Next, the teacher explains how we can represent a binary code on the board: when the code digit was 0 they should not place the card on the board; when the digit was 1 they should place the corresponding card. The next step is to find out what is the secret number represented by the code. For this, the cards must be flipped and the black dots counted. They use the table described in the Educational Material to convert the secret number into letters. In the reverse conversion - letters to binary code - the students receive two pages where, in the first, they write his/her name and favorite fruit. For each letter used, they fill the respective number (using the table) and convert to binary code. After that, the second page containing only the binary code is delivery randomly to another student, who must decipher the message. Concluding, a method of converting decimal numbers to binary code by successive divisions was also practiced with the students.

3.2. Activity 2 - Colour by Numbers

Summary: The objective of this activity is to show students how computer recognizes images, working with the process of coding and decoding pixels that form images on the screen. For this, black and white images are presented to the students, simulating a computer screen. First, from a drawing, the students learn to determine the code that represents the arrangement of pixels on the screen. So, they do the reverse process, determining the drawing from the arrangements of pixels, given by a code. This activity is developed in two tasks, considering learning concepts, practice exercises and evaluation.

Educational Material: the activity is presented using polystyrene boards measuring 48cm wide by 48cm high. In each board, a grid is drawn with a specific number of squares, depending on the number of pixels to be simulated on the screen. In the presen-

tation of the images on the screen, we used small black squares built on cardboard, which were fixed to the board, forming the drawings used to explain the activity.

Methodology: using the polystyrene boards we first present a simple drawing, such as a letter of the alphabet, to show how each line of the image can be represented by a sequence of numbers: the first representing the number of white pixels (or squares without black cardboard) on the line from the first small square on the grid; the second representing the number of black pixels; and so on until the end of the first line. This first example is explained from the first to the last line. Then, the students receive some exercises to do in class and go through an evaluation. In the second task, the reverse is shown: from a code, with the number of white and black squares on each line, we derive an image. After the explanation, they do some exercises and an evaluation.

3.3. Activity III - Battleships

Summary: the aim of this activity is to introduce through a battleships game three different search methods - linear search, binary search and search by hashing - often employed to find information in (large) collections of data. The activity is split into three tasks one for each search method.

Educational Material: the necessary materials for the development of the activity are: (i) game boards, as illustrated in Figure 3, in large format for the teacher, posted on Styrofoam, and in A4 sheets for students; (ii) squares made of black cardboard with dimensions 6 cm × 6 cm; (iii) pins to secure the squares in the Styrofoam; (iv) beans.

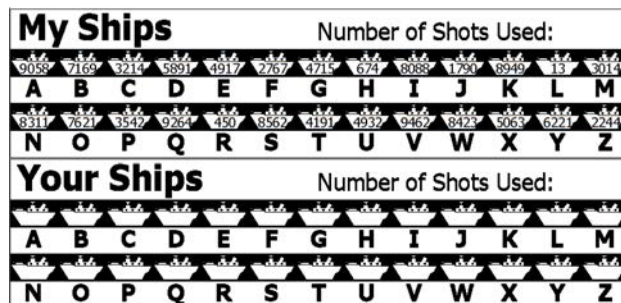


Figure 3. Battleships board [Scaico et al. 2012]

Methodology: the battleships game is a guessing game for two players. The goal of the game is to sink a ship chosen by the partner. Each player must tell his/her partner the boat number (of their ships) that the opponent must overthrow. The players take turns to guess where the partner's ship is. One player says the letter name of a ship and the partner tells him the number of the ship at that letter. The numbers of shots taking to locate the partner's ship is the score of the game. The player who hits fewer shots wins. In the first task the boats are randomly numbered, and the players execute a linear search. They go through along the positions, one by one. In the second task ship numbers are sorted, hence they learn that the best strategy to win is to use binary search. That is, the player must shoot in the middle of the ships list and see whether the number is greater than or less than the value he is looking for. This determines if the ship is in the first or second half of the board. This strategy should be repeated until the end of the game. In the third task, the ships are organized in ten columns (0 to 9) and it is possible to find out which

column the ship is in simply by adding all the digits of the ship's number and computing the sum modulo 10 (the modulo operation finds the remainder after division of the sum by 10). Each task is presented in a different meeting. At the beginning of each lesson the teacher presents the game playing with a volunteer, using the larger board posted in front of the classroom. The black cards are used to mark where the shots have been taken, allowing the reuse of the board. With the same purpose the students can use beans. Then, some game rounds must be performed between the students and the teacher must collect and discuss the scores. The idea is to follow up a discussion, questioning what would be the minimum and maximum scores possible, if there is a strategy which leads to a lower score and what the advantages of each search method are. The binary search is usually faster than the linear search, but the linear search doesn't require the ships to be sorted into order. The search by hashing is usually faster than the other two, but it is possible, by chance, for it to be very slow. In the worst case, if all the ships end up in the same column, it is just as slow as the first strategy.

3.4. Activity IV - Sorting Algorithms

Summary: this activity introduces two sorting algorithms, named as selection sort and quicksort. The purpose of this activity is to show that there are different strategies to sort a set of data and that the choice of the method to be applied directly impacts the speed at which you get the sorted data. According to the method chosen, it can take a long time to sort a large list into order, even on a fast computer.

Educational Material: the material used for the activity was: scales; kits of 8 bottles, all with different weights, in a box; kits of 30 squares made of black cardboard with dimensions 6 cm x 6 cm; board containing 8 horizontally aligned positions, one for each bottle; and rubber bands. We considered one material set for each trio of students.

Methodology: at the beginning of each meeting, before distributing the material to the students, the teacher presents the sorting algorithm for the whole class. Particularly, in the first task, before presenting the method, is reviewed the concept of comparison. The teacher chooses two bottles randomly and asks which is the lighter (or heavier). He weighs the two bottles and then points out that a comparison operation was done to find the answer. Following, the selection sort is introduced. Two bottles are removed from the box, a black square is picked up and the bottles are weighed. The lighter remains with the teacher and the heaviest is placed on the floor. Another bottle is removed from the box, a black square is picked up, and the bottle is compared with the one that the teacher has at hand. The teacher gets the lighter one and the other is placed on the floor. This process is repeated for each remaining bottle in the box. When the box is empty, the teacher has the lightest bottle in his hand. Then, he puts the lighter in the leftmost free position of the board, places the other bottles back into the box and repeats the process until it has all the bottles placed on the board. The black squares are used to count the total number of comparisons required to sort the whole kit of bottles. For each comparison made, a black square is picked up. After the teacher's presentation, students apply the method to sort subsets of 3 to 8 bottles, always recording the cost of ordering each subset. In the selection sort they note that the comparison cost for each bottle subset is the same for all groups of students. Before the final evaluation of the first class, the number of comparisons obtained by subset are discussed. The teacher concludes the presentation, describing that by using this ordering method, the number of comparisons required to sort

x bottles is given by $\sum_{i=1}^{x-1} i$. In the second task quicksort is introduced. A bottle of the kit is selected and chosen as pivot. The teacher puts a rubber band on the bottle to identify that it is the pivot. Then the bottles are removed from the box one by one and compared to the pivot. For each comparison a black card is picked up. If the bottle is lighter than the pivot, it is placed to the left of the pivot. If it is heavier than the pivot, it is placed to the right of the pivot. When the box is empty, the pivot has on its left hand side all the bottles that are lighter than it, and on its right all the heavier. With this information it is possible to put the pivot on the board in its correct position. If we have x lighter bottles than the pivot, its position on the board from left to right is $x + 1$. Then, the remaining bottles are replaced into the box and the process is repeated until all bottles are correctly positioned on the board. In such case, the students note that the comparison cost varies among groups, even considering the same number of bottles in the subset. So, a final discussion shall be conducted. The teacher guides a comparison between the two approaches and emphasizes that according to the method chosen, the ordering can be achieved in fewer steps (time).

4. Results about the pilot project

First, the profile of classroom practices is described. Then, the main results concerned with pre- and post-tests are presented.

4.1. Profile of classroom practices

Decisions taken for the implementation of this pilot project took into consideration meetings with our City Department of Education and Sports. The activities were applied in a primary public school, which is located near our university, making it easier not only the access of undergraduate students and professors' support to the practices in school classes, but also increasing the integration between school and academia communities.

Regular meetings were held with the management and coordination of the school, identifying the profiles of the groups and preparing preliminary work schedule. The study was conducted on two fourth year classes since in order to develop the chosen CT concepts, the children should know how to read, write and basic arithmetic. The survey addressed a sample of 46 students divided into a morning class (22 students) and an afternoon class (24 students). The age of the students ranged from 9 to 13 years old, but the vast majority (82%) was 9 or 10 years old. Regarding gender, 63% of students were male while 37% were female.

Regarding the schedule, the activities were held in the 2nd semester of 2013, organized as follows: (i) two meetings, one for the pre- and the other for the post-test; (ii) three meetings (one per week) for each activity and evaluations (a day for each task). Moreover, the practice of all activities began in the second meeting, after applying the pre-test. The meetings were held with the presence of two or three Computer Science undergraduate students sharing the following responsibilities: (i) the oral presentation of the activity; (ii) the support with concrete material related to the development of activities (only when necessary); and (iii) the individual assistance (when required by students). In addition, two professors/researchers, together with the class teacher were always present. The former supervised the undergraduate students and helped collecting data for the project; and the latter, helped with previous knowledge about behaviour and personal characteristics of each student.

4.2. Results about pre- and post-tests

The test included four questions (Q1 and Q2, worth 2 points each; Q3 and Q4, worth 3 points each) devised to evaluate the following abilities: association, correlation, count, application, identification, memorization, comparison and sorting. The conception of the assessment instrument was based on these skills and the weights of the questions have been distributed according to the abilities involved. In particular, Q1 required from students the skills to correlate and count; Q2 to associate, correlate and apply; Q3 to identify, memorize and compare; and Q4 to correlate, count, apply, memorize, compare and sort. Based on these skills, Q1 and Q2 evaluated activities BN and CN, Q3 evaluated BN, and Q4, activity SA. Students' total scores and by question, in pre- and post-test, including the gains and the results of the test of significance are provided in Table 1.

Table 1. Description of students scores. Total and by question scores in the pre- and post-test including gain and results of significance test.

Group	Question (weight)	Before		After		Gain		t-test	
		Mean	SD	Mean	SD	Mean	SD	T	p
Morning n=16	1 (2,0)	1,1	0,3	1,1	0,6	0,0	0,5	0,24	0,8172
	2 (2,0)	1,6	0,8	1,5	0,9	-0,1	0,5	-0,72	0,4847
	3 (3,0)	2,3	0,9	2,6	0,7	0,4	1,3	1,14	0,2708
	4 (3,0)	1,2	1,2	2,4	0,9	1,2	1,6	2,92	0,0105
	Total	6,1	1,8	7,6	1,6	1,5	1,6	3,67	0,0023
Afternoon n=19	1 (2,0)	1,1	0,5	1,3	0,5	0,2	0,5	1,37	0,1868
	2 (2,0)	1,5	0,8	1,7	0,8	0,2	1,0	0,84	0,4141
	3 (3,0)	2,5	0,7	2,7	0,5	0,2	0,7	1,25	0,2284
	4 (3,0)	1,6	1,1	2,1	1,0	0,5	1,4	1,64	0,1185
	Total	6,7	2,1	7,8	1,1	1,1	1,8	2,59	0,0186

n = number of students; SD = standard deviation

The pre-test score describes the initial condition of the student in relation to the skills worked in the activities. The gain with the development of these activities is measured by the difference between the scores of students in the post-test and pre-test. The significance of these gains is verified by the t-test for paired samples, with $\alpha = 0.05$. The hypothesis under verification in this test assumes that the average gain of the students (μ) does not differ from zero, i.e. $H_0 : \mu = 0$.

It is possible to verify for both groups that the average performance of students in the pre-test was lower than 7.0, but the afternoon class (average 6.7) was slightly better than the morning class (average 6.1). In the post-test, the gain from the morning class (average 1.5) was higher than the one from the afternoon class (average 1.1). Although they are both low, these average gains were significant. However, when one analyses the gain by question, it is possible to note that only in Q4, and only for the morning class, the average gain was significantly higher than zero ($p = 0.0105$). This result suggests that the morning class responded better to the skills worked in the SA activity. This analysis also reveals that questions that the students could not solve in the pre-test are still incorrect in the post-test. For example, only 11% of students got the first question fully right in the pre-test. Although the percentage increased to 23% in the post-test, most students could not answer the question that asked to complete a numerical sequence. This means that activities NB and CN, did not help the ability to correlate and count that were needed to solve the question. Questions 2 and 3 required six of the eight abilities mentioned, and students got at least 75% of it right already in the pre-test, with no significant increase in

the post-test. Only in question 4 there was a major increase in the result from 40% in the pre-test to 80% percent in the post-test. This question required different abilities such as to correlate, count, apply, memorize and sort.

These results may suggest that the pre and post-tests used were not a good instrument to measure the gain achieved by the students. This gain was clearly observable in the class assignments. This observation is important so we can seek a better mechanism to measure the knowledge and abilities of students in the next years.

5. Final remarks

This paper contributes with a detailed description, from modeling to development of appropriate methodology to teach some concepts of CT. Specially, we focus on four activities related to data and image representation, sorting and searching algorithms, developing deductive logical reasoning for fourth grade students in a public elementary school. The students achievement assessment is obtained by analysis of pre- and post-tests. Case studies and results of our research illustrate how schools and academia can be integrated to plan for instructional improvement in basic education, especially in early school years.

Further work considers improvements in the evaluation tests and development of new activities promoting other potential abilities of logical reasoning in elementary school students. We hope that this project can be one step forward in consolidating CT in primary schools in Brazil in the future.

References

- Andrade, D. et al. (2013). Proposta de atividades para o desenvolvimento do pensamento computacional no ensino fundamental. In *WIE 2013*, pages 169–178.
- Barr, V. and Stephenson, C. (2011). Bringing computational thinking to k-12: what is involved and what is the role of the computer science education community? *ACM Inroads*, 2(1):48–54.
- Bell, T., Witten, I. H., and Fellows, M. F. (2010). *Computer Science Unplugged*.
- Bundy, A. (2007). Computational thinking is pervasive. *Journal of Scientific and Practical Computing*, 1:67–69.
- Caetano, G. and Oliveira, O. (2015). Computational thinking and child performance in a preschool. In Dagiene, V., Schulte, C., and Jevsikova, T., editors, *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education, ITiCS 2015, Vilnius, Lithuania, July 4-8, 2015*, pages 57–62. ACM.
- Carnegie Mellon. Center for Computational Thinking. <http://www.cs.cmu.edu/CompThink/>. Accessed March, 2016.
- CSTA, ISTE, and NSF (2010). Computational Thinking Leadership Toolkit. <http://www.csta.acm.org/Curriculum/sub/CompThinking.html>. Accessed March, 2016.
- de Francca, R. S. and Tedesco, P. (2014). Um modelo colaborativo para à aprendizagem do pensamento computacional aliado a autorregulação. In *Anais do SBIE 2014*, pages 1133–1142.
- Deng, Z., Huang, W., and Dong, R. (2009). Discussion of ability cultivation of computational thinking in course teaching. In *Proc. ICETC'09*, pages 197–200.

- Denning, P. J. (2009). The profession of it: Beyond computational thinking. *Commun. ACM*, 52(6):28–30.
- França, R. S. and Amaral, C. J. H. (2011). Proposta Metodológica de Ensino e Avaliação para o Desenvolvimento do Pensamento Computacional com o Uso do Scratch. In *XIX Workshop de Informática na Escola*.
- França, R. S. and Amaral, C. J. H. (2013). Ensino de computação na educação básica no brasil: Um mapeamento sistemático. In *XXI Workshop de Informática na Escola*.
- Google (2013). Exploring Computational Thinking. <http://www.google.com/edu/computational-thinking/>. Accessed March, 2016.
- Hambruch, S. et al. (2009). A multidisciplinary approach towards computational thinking for science majors. In *SIGCSE 2009*, pages 183–187. ACM.
- ISTE (2013). Operational Definition of Computational Thinking. <https://csta.acm.org/Curriculum/sub/CurrFiles/CompThinkingFlyer.pdf>. Accessed March, 2016.
- Kafura, D. and Tatar, D. (2011). Initial experience with a computational thinking course for computer science students. In *SIGCSE 2011*, pages 251–256. ACM.
- Lee, I. et al. (2011). Computational thinking for youth in practice. *ACM Inroads*, 2(1):32–37.
- PUC-Rio (2014). Scalable Game Design Brasil. <http://www.serg.inf.puc-rio.br/wiki>. Accessed March, 2016.
- Ramos, F. and da Silva Teixeira, L. (2015). Significação da aprendizagem através do pensamento computacional no ensino médio: uma experiência com scratch. In *Anais do WIE 2015*, pages 207–217.
- Repenning, A., Webb, D., and Ioannidou, A. (2010). Scalable game design and the development of a checklist for getting computational thinking into public schools. In *41st Tech. Symp. on Comp. Sci. Education*, pages 265–269, New York, USA. ACM.
- Rodrigues, R., Andrade, W., Guerrero, D., and Sampaio, L. (2015). Análise dos efeitos do pensamento computacional nas habilidades de estudantes no ensino básico: um estudo sob a perspectiva da programação de computadores. In *Anais do SBIE 2015*, pages 111–121.
- Rodriguez, C., Zem-Lopes, A. M., Marques, L., and Isotani, S. (2015). Pensamento computacional: transformando ideias em jogos digitais usando o scratch. In *Anais do WIE 2015*, pages 62–71.
- Scaico, P. D. et al. (2012). Um Relato de Experiências de Estagiários da Licenciatura em Computação com o Ensino de Computação para Crianças. *RENOTE*, 10(3).
- UFPel (2014). EXP-PC - Explorando o Pensamento Computacional para a Qualificação do Ensino Fundamental. <http://wp.ufpel.edu.br/pensamentocomputacional>. Accessed May, 2016.
- Yinnan, Z. and Chaosheng, L. (2012). Training for computational thinking capability on programming language teaching. In *Proc ICCSE'12*, pages 1804–1809.

L²PM: relato de uma experiência sobre o ensino integrado de Lógica, Programação e Matemática para Computação

Mateus Raeder¹, Mônica Py¹, Sandro Rigo¹, Josaine Pinheiro¹

¹ – Universidade do Vale do Rio dos Sinos (UNISINOS)
Av. Unisinos, 950 – CEP 93.022-750– São Leopoldo – RS – Brazil

{mraeder, mpy, rigo, josaine}@unisinos.br

Abstract. *This paper aims at describing and analyzing a brand new experience in teaching Logics, Programming and Mathematics to IT students at university. The experience started with an extra class study group mediated by three teachers (one from each of the mentioned subjects). The distinction of it was the dynamics of the meetings: interdisciplinary classes. These integrated classes promoted rounds of fifteen minutes dealing with topics related to each theme. The results show that the applied dynamics fostered other ways to build knowledge. Students revealed interest in taking part in such meetings and also enthusiasm for improving such specific knowledge.*

Resumo. *Este artigo descreve e analisa uma experiência diferenciada de ensino de Lógica, Programação e Matemática para Computação, cujos sujeitos envolvidos são alunos da graduação das áreas de TI na UNISINOS. Constitui-se um grupo de estudos chamado L²PM, que teve como mediadores três professores e ocorreu em período extraclasse. Os encontros contemplaram em um mesmo ambiente aulas integradas das disciplinas em questão, realizadas em rodadas de 15 minutos, alternadas entre os professores e temas, finalizando com uma rodada que serve de integração, a qual reúne os conteúdos tratados no grupo. O resultado desta experiência aponta que a dinâmica adotada no L²PM possibilitou outras condições para a construção de conhecimento.*

1. Introdução

Em cursos de Tecnologia da Informação (TI), um dos principais focos é que o aluno deve sugerir e implementar soluções para problemas cotidianos, empresariais e de diversas outras naturezas [Júnior and Rapkiewicz 2004]. Entretanto, alunos que ingressam em cursos superiores da área da TI muitas vezes não estão preparados para o novo paradigma de ensino que as universidades propõem, uma vez que a maioria está acostumado ao ritmo e formato do Ensino Médio tradicional.

O caráter disruptivo desta mudança, atrelado à dificuldade inerente dos cursos da área, são fatores críticos e que interferem diretamente no aprendizado dos estudantes. Em especial, na área em questão, o ensino de programação e disciplinas correlatas não é uma tarefa trivial. A resolução de problemas, segundo [Martins et al. 2010], “é antes de tudo uma competência que envolve processos cognitivos como criatividade e racionalidade”. Para isto, as habilidades de leitura, interpretação e raciocínio dos problemas propostos devem ser desenvolvidos e incentivados, respeitando o ritmo de cada indivíduo [Martins et al. 2010].

Outro fator estreitamente ligado à dificuldade de aprendizado de conteúdos relacionadas à programação (como lógica e matemática), trata-se do aspecto motivacional. O fato de os professores do ensino superior simplesmente repassarem conhecimento na sala de aula de forma convencional é um problema a ser resolvido, pois faz com que os estudantes sejam atores passivos no processo de aprendizado [da Silva 2003].

Neste contexto, um grupo de professores da UNISINOS realizou um levantamento sobre as disciplinas de Programação, Lógica e Matemática para identificar essas questões. Estas atividades encontram-se, em geral, no primeiro semestre dos cursos de TI da universidade em questão, e são disciplinas extremamente importantes para o futuro profissional e acadêmico dos alunos.

Além de índices altos de reprovação, alguns discursos são recorrentes entre alunos e professores, e esses sugerem que outras metodologias deveriam ser utilizadas em determinados momentos. Essas recorrências são testemunhadas por:

- *“Não vejo relação entre o conteúdo de Matemática para Computação e Programação”*
- *“Em alguns momentos não acho que a aula do professor seja motivadora”*

Além disto, em certa ocasião um dos professores de programação proferiu o seguinte depoimento:

- *“17 aprovados, de um total de 42 que fizeram a prova [...] Acho que a prova não foi difícil [...] Espero forte evasão nos próximos dias/semanas”*

Todos estes fatos e argumentos foram grandes impulsionadores para a criação do grupo de estudos L²PM. A sigla refere-se às disciplinas que são estudadas no grupo: Laboratório I, Lógica, Programação I e Matemática para Computação. A ideia do grupo é realizar aulas integradas em períodos extraclasse, para motivar alunos a aumentarem seu engajamento com as disciplinas, bem como para que visualizem a grande relação de uma disciplina com as outras. A experiência foi realizada pela primeira vez no primeiro semestre de 2015, e será descrita em maiores detalhes no decorrer deste artigo.

O restante do artigo está organizado da seguinte maneira: a Seção 2 apresenta a fundamentação teórica para a realização da experiência; a Seção 3 elenca alguns trabalhos relacionados ao tema proposto neste artigo; a Seção 4 descreve em detalhes a dinâmica e a organização das aulas do L²PM; a Seção 5 relata a percepção dos alunos e professores sobre a experiência realizada, e; a Seção 6 conclui o trabalho, apontando algumas considerações finais e trabalhos futuros.

2. Fundamentação teórica

Na atualidade, há um descompasso entre o que é ensinado na escola e o que o aluno aspira aprender, tornando o ensino e a aprendizagem dois processos que estão cada vez mais distantes. Essa distância entre o que é ensinado e o que se quer (enquanto aluno) aprender constitui um paradigma que encontra-se naturalizado no processo de ensino-aprendizagem, ou ainda, duas ações que são gerenciadas, em sua maioria, por sujeitos diferentes – professor e aluno – são percebidas, compreendidas como uma única.

A naturalização desse paradigma, de forma insipiente, pode provocar certo desconforto quando o aluno ingressa na universidade, visto que o sujeito – aluno carrega marcas de um ensino no qual muitos dos temas abordados não foram aprendidos.

Nessa direção pontuamos que uma das dificuldades detectadas nos alunos, no que se refere a aprendizagem de Lógica, Matemática para Computação e Programação é a linguagem. Uma das características comuns dessas disciplinas é que cada uma possui uma linguagem específica da área. Por exemplo, em Matemática para a Computação, os alunos devem ter habilidade em ler as questões e realizar a tradução para a linguagem matemática. Isso também verifica-se em Lógica e Programação. Detectando esse fator e como as linguagens possuem muitas aproximações foi avaliada a possibilidade de construir um ambiente que tornasse possível estudos integrados para auxiliar os alunos a realizarem as conexões existentes entre diferentes disciplinas.

A inspiração para a construção desse ambiente foi a sala de aula invertida de Jonathan Bergmann e Aaron Sams [Bergmann and Sams 2014]. A inversão ocorre quando os alunos estudam os conteúdos em casa, através de livros, vídeos, filmes, músicas etc., e vão para a escola resolver exercícios, problemas, ou seja, testar seus conhecimentos.

A experiência vivenciada pelos alunos no L²PM, que teve uma concepção diferente da sala de aula tradicional, reforçou a importância do aluno possuir autonomia e de ocupar espaços nos quais possam testar seus conhecimentos, com a mediação de profissionais competentes. Ressaltamos que quando nos referimos à experiência estamos considerando “o que nos passa, o que nos acontece, o que nos toca. Não o que se passa, não o que acontece, ou o que toca”[Larrosa 2002].

3. Trabalhos relacionados

Na literatura, podem ser encontrados diversos trabalhos que analisam metodologias diferenciadas para o aprendizado de programação e disciplinas correlatas. Na sequência, alguns destes trabalhos serão destacados, em virtude do seu relacionamento com o presente estudo e do teor inovador de suas ideias.

Em [Raabe and da Silva 2005] é abordada a dificuldade de aprendizagem dos alunos. Os autores relatam que o aprendizado de algoritmos é desafiador tanto para alunos quanto para professores. Para os alunos, o entendimento da parte lógica e matemática é distante do seu cotidiano. Para os professores, a maior dificuldade está no acompanhamento e atendimento aos alunos, em virtude da quantidade e das condições individuais de aprendizado. Para tentar minimizar as dificuldades supracitadas, os autores utilizaram um Assistente Inteligente de Ensino (ITA - Intelligent Teaching Assistant) chamado ALICE [Raabe and Giraffa 2006], que permite o acompanhamento individualizado dos estudantes. O ambiente ALICE foi aplicado em algumas turmas da instituição de ensino dos autores e ajudou a identificar melhor os problemas dos alunos em sala de aula, com a contrapartida de que a demanda sobre o professor aumentou, em virtude do acompanhamento de cada aluno.

Os autores de [Helmer 2014] também mostram preocupação com novas formas de construção do conhecimento em sala de aula. No trabalho, são exploradas novas estratégias no ensino de Estrutura de Dados, como Philips 66, recordatório, simulado, dinâmica de grupo, dentre outras. O maior motivador para os experimentos utilizando novas metodologias de ensino é a necessidade de ensinar aos alunos os conceitos de abstração. As estratégias propostas foram realizadas em uma turma e os resultados mostraram que a experiência foi extremamente positiva, pois cada estudante possui formas diferentes para obter conhecimento. A diversificação da forma de ensino abre um leque

maior de formas de aprendizado para os alunos.

Seguindo a mesma linha, o trabalho proposto em [Lopes 2006] apresenta a proposta de diversificação das estratégias no ensino de programação. Os autores falam sobre a “complexidade inerente aos processos de ensino-aprendizagem”, e que é função dos professores a busca por novos métodos menos “burocráticos” de ensino. A ideia dos autores é de que a ênfase do processo de avaliação deveria ser alterado, pois, segundo os autores, a aplicação de provas pode não medir a capacidade de programação dos alunos. Assim, os autores propuseram estratégias de aulas expositivas dialogadas, estudo de textos e soluções de problemas. O principal benefício percebido pelos autores é de que a motivação dos alunos é uma forte aliada do conhecimento, uma vez que eles mesmos constroem seu aprendizado, além de obterem uma participação mais ativa nas aulas.

Outro trabalho desenvolvido na mesma direção da experiência realizada pode ser visto em [Gomes et al. 2008]. Os autores do trabalho apresentam uma proposta para auxílio aos alunos com dificuldades na aprendizagem inicial de programação. Apontam as grandes taxas de insucesso nas disciplinas de programação básica, novamente em virtude dos conceitos de abstração. A proposta dos autores para minimizar o problema descrito concentra-se na utilização de um modelo que determina o estilo de aprendizagem de um indivíduo, chamado “*The Felder-Silverman Learning Style Model*”. Com este modelo, os autores procuram motivar os alunos através da resolução de problemas, utilizando-se de aspectos lúdicos e do respeito ao seu próprio ritmo de aprendizagem.

Os trabalhos abordados possuem uma convergência em suas análises que é a motivação. Para isto, diversos autores propõem metodologias diferenciadas para o ensino destes conteúdos. Assim, diferentemente dos trabalhos descritos acima, a presente experiência difere-se dos demais na concepção do modelo de ensino, pois tenta inverter, ou ainda, subverter a lógica da sala de aula tradicional, ao propor uma dinâmica diferenciada e desafiadora. Entendemos que na contemporaneidade “(. . .) a figura cartesiana do sujeito humano dissolve-se nos fluxos de informação e nos sistemas de relações”[Vincent 2013], o que torna as barreiras tradicionais da sala de aula tradicional, objetivos para serem transpostos, de maneira a construir outras formas de ensinar e aprender em um mundo, no qual a tecnociência está na ordem do discurso.. Este artigo apresenta uma outra maneira de mediar os processos de ensino e aprendizagem, inspirada na sala de aula invertida, mas com a peculiaridade de ter três disciplinas ocorrendo concomitantemente. Não estamos tentando dizer que essa é a melhor forma de ensinar, ou a mais criativa. Gostaríamos apenas de compartilhar uma experiência que, segundo testemunhas realizadas pelos próprios alunos, foi diferenciada e que alcançou resultados além do esperado.

4. Grupo de estudos L²PM

Nesta seção serão detalhados os aspectos inerentes ao grupo de estudos L²PM. Serão descritos os aspectos motivacionais para a concepção do grupo (Seção 4.1) e, também, o ambiente diferenciado criado para as aulas integradas das disciplinas em questão (Seção 4.2).

4.1. Motivação para criação

Conforme indicado anteriormente, o aprendizado de programação e disciplinas correlatas (como Lógica e Matemática para Computação) é uma tarefa complexa, que é composta por várias relações. Além disso, muitos dos alunos que ingressam nos cursos de

Computação nas universidades chega sem qualquer conhecimento sobre o assunto e não estão preparados para a mudança de paradigma que ocorre entre o Ensino Médio e a Universidade.

Observando a realidade encontrada, os cursos de Computação da UNISINOS oferecem aulas de reforço aos alunos em algumas disciplinas, por elas serem classificadas pelos alunos como complexas, principalmente para aqueles alunos do primeiro semestre. Essas aulas ocorrem em períodos em que nenhum aluno está em aula, para que todos os alunos interessados possam comparecer. São aulas tradicionais, em outras palavras, aulas ministradas por um professor à frente da turma e abordando uma única disciplina, reforçando alguns aspectos mais importantes da matéria.

Entretanto, nos dois últimos anos notou-se uma redução no número de alunos participantes nestes encontros para reforço do conteúdo. O comparecimento nestas aulas era de cerca de 20 alunos, contabilizadas as aulas que ocorreram desde 2013. Entretanto, em 2015, o número de alunos que compareciam nestas aulas diminuiu bastante, chegando até mesmo a 8 alunos em um encontro.

Para os professores, não se tratava de um problema o baixo número de alunos, uma vez que aqueles que sentiam dificuldades no conteúdo compareciam nas aulas e recebiam todo o apoio e ensinamento desejado. Entretanto, o número de alunos nas aulas não refletia o que acontecia em sala de aula, principalmente ao se observarem as notas das avaliações.

Assim, estimou-se que a diminuição na motivação dos alunos poderia estar diretamente ligada ao modelo de aula, que não era mais atrativo para eles, os quais procuravam resolver dúvidas pontuais sobre a matéria. Aliando a isso, através do questionamento com alguns alunos que não compareciam e que não obtiveram notas satisfatórias, percebeu-se que o nome “aula de reforço” não estava recebendo uma boa aceitação por parte dos estudantes. Podemos inferir que esse fato ocorreu, pois eles sentiam-se menos capazes (ou inteligentes) do que aqueles que não compareciam nas aulas. Desta forma, muitos alunos não frequentavam os encontros para evitar uma resposta positiva para “Você foi na aula de reforço?”, possivelmente feita por colegas.

Com todas essas mudanças na sociedade, é importante rever o formato dos encontros, bem como o objetivo e a metodologia das aulas extraclasse teve de ser revista. A ideia não é que apenas os alunos que precisam de reforço nas notas compareçam em estudos extraclasse, mas sim realizar encontros de estudos em que alunos de diferentes disciplinas possam comparecer para revisar o conteúdo, ter contato mais direto com os professores para sanar dúvidas e poder trocar experiências e ideias com outros alunos. Com estas premissas, criou-se o grupo de estudos L²PM.

4.2. Ambiente e dinâmica das aulas

O ambiente do L²PM foi pensado para que os alunos não se sentissem em uma sala de aula convencional, o que remeteria a eles a ideia de que se tratava puramente de uma aula comum para reforço de conteúdos. A Figura 1 mostra a disposição do local do grupo de estudos.

Diferentemente de uma aula tradicional, o modelo do grupo de estudos L²PM ocorre em 3 ilhas de conteúdo. Uma ilha de Lógica, uma ilha de Programação I e

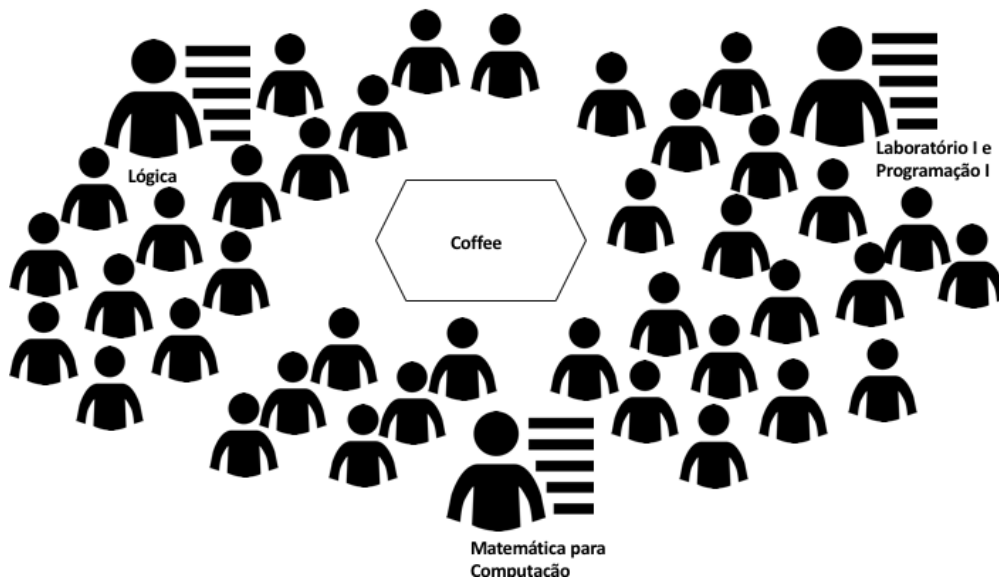


Figura 1. Disposição do espaço físico do L²PM

Laboratório I (pois são conteúdos relacionados, com assuntos teóricos e práticos sobre programação) e uma ilha de Matemática para Computação. Em cada uma das ilhas existe um professor responsável por aquele conteúdo e um quadro branco. Cada professor, inspirado na sala de aula invertida, expõe exercícios, de vários graus de complexidade, para que os alunos o resolvam. Caso haja dúvidas, o professor procura responder aos questionamentos, como outros questionamentos e com explicações dos exercícios. No centro, uma mesa com café e alguns biscoitos torna o momento mais descontraído e menos monótono.

O aspecto que mais chama a atenção dos alunos para comparecerem ao grupo de estudos é o fato de 4 disciplinas serem abordadas no mesmo ambiente. As disciplinas de Lógica, Laboratório I, Programação I e Matemática para Computação são geralmente cursadas no mesmo semestre pelos alunos. Assim, alunos que cursam 1, 2, 3 ou as 4 disciplinas podem comparecer aos encontros do grupo.

A dinâmica dos encontros do L²PM é a seguinte: o encontro dura de 3 horas e meia a 4 horas. No ambiente integrado, os professores de cada disciplina abordam conteúdos específicos e previamente selecionados de acordo com as dúvidas e dificuldades identificadas em sala de aula. Cada professor apresenta a temática do momento em sua ilha, em uma microaula expositiva de 15 minutos, que possui na centralidade a resolução de exercícios, e ao final deste tempo, disponibiliza outros exercícios. Assim sendo, os alunos interessados na matéria em questão sentam-se nos lugares próximos ao quadro daquele professor.

Logo em seguida, quando os 15 minutos da microaula do primeiro professor chegam ao fim, este dá a vez para o próximo professor. O segundo professor, então, introduz sua temática em sua ilha, no mesmo formato relatado acima, realizando o fechamento da sua microaula com a disponibilização de exercícios aos alunos interessados.

Ao final da microaula do terceiro professor, os próximos 15 minutos ficam livres para que os alunos transitem entre as ilhas da forma que preferirem. Neste tempo, os alunos podem fazer um lanche, resolver exercícios, tirar dúvidas com algum(ns) dos pro-

fessores, ir ao banheiro etc. Quando estes 15 minutos acabam, dá-se o fim de uma rodada, que possui a duração de 1 hora (3 microaulas de 15 minutos mais os 15 minutos livres ao final). Quando uma rodada termina, o primeiro professor dá início a mais uma rodada de microaulas. No tempo que separa as microaulas de um professor, este fica em sua ilha para sanar dúvidas dos alunos.

Em um encontro do L²PM, acontecem 3 rodadas de 1 hora. Ao término de todas as rodadas, os professores permanecem disponíveis por cerca de 30 minutos até 1 hora para esclarecimento de dúvidas dos alunos.

A disposição das ilhas permite que os alunos transitem entre os conteúdos livremente. Deste modo, os alunos têm a opção de escolher dar continuidade no desenvolvimento do exercício e esclarecimento de dúvidas em uma ilha específica ou acompanhar o professor da próxima microaula. Para um aluno que está cursando apenas uma das disciplinas, por exemplo, ele assiste a microaula desejada e tem 45 minutos para realizar exercícios e conversar com o professor na ilha específica, até que a próxima microaula do seu conteúdo seja iniciada. As fotos visualizadas na Figura 2 são imagens reais de um encontro do L²PM no primeiro semestre de 2015.¹ do L²PM.

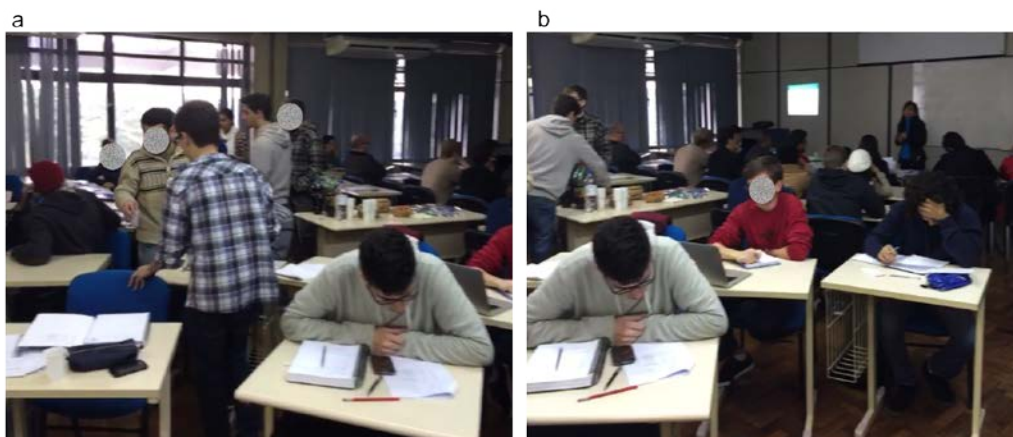


Figura 2. Fotos reais de um encontro do grupo de estudos L²PM

Na Figura 2a, é possível observar que alguns alunos estão voltados para o lado esquerdo da foto (onde encontrava-se a ilha de Programação I e Laboratório I) e alguns para a frente da foto (onde encontrava-se a ilha de Lógica). Na Figura 2b, por sua vez, é possível notar que ao fundo existem outros alunos voltados para outra ilha (onde estava ocorrendo a microaula de Matemática para Computação). Em ambas as fotos é possível verificar o local onde o café e biscoitos estavam dispostos.

A ideia do L²PM, além de promover uma dinâmica mais atrativa de aula aos alunos, é que os professores integrem os conteúdos, seja em seus exercícios ou seja ao final das 3 rodadas de microaulas, no tempo restante do encontro. Este tipo de abordagem proporciona ao aluno uma melhor percepção sobre o relacionamento muito próximo que a Programação, a Lógica e a Matemática para Computação têm entre si.

Por enquanto, foram realizadas duas edições do L²PM, ambas no primeiro semestre de 2015. A partir de 2016, serão realizadas edições semestrais, ocorrendo sem-

¹Os rostos dos alunos foram borrados na foto, para fins de preservação de suas identidades.

pre próximas ao término do semestre, quando o conteúdo é maior, mais complexo e a integração dos temas nas diferentes disciplinas pode ser melhor percebida.

5. Percepção dos envolvidos

A partir dos relatos dos professores e alunos, percebe-se que a dinâmica adotada nas aulas qualifica o conhecimento do aluno para conteúdos discutidos em cada disciplina, além de enriquecer a interdisciplinaridade. Nesse contexto, o trabalho descreve a percepção dos professores e a percepção dos alunos.

A percepção dos professores foi positiva, pois atingiram o objetivo que era discutir diferentes temáticas em uma única sala, proporcionando aos alunos um aprendizado “múltiplo”. Também ficou evidente a proximidade da relação professor-aluno para esclarecimento de questões pontuais do conteúdo, sendo importante destacar que as microaulas eram iniciadas com exercícios e esses desencadeavam os questionamentos dos alunos. Para os professores, os alunos que participaram dessa dinâmica mostraram-se mais comprometidos e empenhados nas atividades de sala de aula.

Nessa direção, um fato extremamente importante foi a motivação e o interesse dos alunos na atividade oferecida em período extraclasse. Conforme dito anteriormente, nas aulas de reforço tradicionalmente realizadas pelos cursos de TI da UNISINOS, a participação dos alunos estava diminuindo consideravelmente, passando de 20 alunos no começo da proposta deste tipo de aula para 8 em certas ocasiões. No grupo de estudos L²PM, entretanto, o número de estudantes que participaram da segunda edição do grupo de estudos foi de mais de 40. Se somadas, as duas edições realizadas ultrapassam 60 alunos.

Infelizmente, os nomes dos alunos que participaram das primeiras edições do L²PM não foram registrados, uma vez que os professores envolvidos não imaginavam a grande repercussão que o grupo teria perante os alunos, em especial a segunda edição. Entretanto, alguns alunos reportaram-se aos professores e coordenações envolvidas agradecendo a oportunidade e solicitando que a atividade fosse realizada com mais frequência e também com outras disciplinas do currículo.

A ilha que os alunos mais procuraram foi a ilha de Matemática para Computação. Entretanto, vários alunos seguiam o roteiro da aula e migravam entre as ilhas conforme as microaulas começavam, o que mostrou que a disposição e o formato escolhido para o grupo de estudos surtiu o efeito desejado.

É interessante pontuar que, em diversas ocasiões, alunos de diferentes turmas encontravam-se discutindo conteúdos no centro da sala, local onde encontrava-se o café. Este tipo de interação também foi considerado um ponto positivo, pois o encontro promoveu novos contatos e uma importante troca de experiência entre os estudantes envolvidos.

Finalmente, professores que ministram as disciplinas em questão em sala de aula mas não estavam envolvidos diretamente no L²PM gostaram da ideia do encontro, promovendo-o em suas turmas e solicitando aos professores envolvidos que novas edições sejam realizadas.

6. Considerações finais

O fator motivacional dos alunos, proporcionado pelo tipo de dinâmica, ambiente e estratégias de ensino, foi um dos fatores que ajudou para um efetivo comparecimento e dedicação nas atividades. Os métodos de ensino devem ser repensados e adequados ao tipo de aluno que existe na universidade.

Alunos de ensino superior, especialmente de programação e disciplinas relacionadas, demonstraram necessitar um pouco mais de tempo para aprenderem, além do que possuem em sala de aula. O tipo de aprendizagem de cada aluno deve ser considerado, pois cada um experimenta os fatos de maneira única. A criação de um ambiente integrado como o L²PM mostrou que novas metodologias podem ser aplicadas dentro da sala de aula, para que o ensino de Programação, Lógica e Matemática para Computação seja efetivo e que possibilite um ambiente propício para despertar o interesse do aluno.

Alunos de ensino superior, especialmente de programação e disciplinas relacionadas, não possuem o mesmo ritmo de aprendizado. A maneira singular de aprendizado de cada estudante deve ser levada em conta e a instituição, na figura dos seus professores, deve preocupar-se com tal fato. A criação de um ambiente integrado como o L²PM mostrou que novas metodologias podem ser aplicadas dentro da sala de aula, para que o ensino de Programação, Lógica e Matemática para Computação não seja monótono e tedioso, tanto para alunos quanto para professores.

A próxima edição do L²PM será realizada no primeiro semestre de 2016 e novas abordagens estão sendo estudadas para motivar ainda mais os alunos. Quanto aos professores, espera-se que a cada encontro novos professores participem do grupo de estudos, pois acredita-se que este tipo de ação engrandece o trabalho dos professores também dentro da sala de aula.

Como trabalhos futuros relacionados ao experimento realizado, acredita-se ser possível integrar as atividades do grupo de estudos L²PM com sistemas de acompanhamento dos alunos no decorrer das aulas durante o período letivo. Desta forma, os professores chegarão ao encontro com um panorama dos alunos envolvidos, o que pode aumentar significativamente o resultado do L²PM, pois os assuntos poderão ser voltados às dificuldades já percebidas pelo sistema.

Além disto, outras disciplinas estão sendo estudadas para serem integradas neste formato, como as disciplinas de Complexidade de Algoritmos, Teoria da Computação e Estruturas Avançadas de Dados. Tratam-se de atividades mais avançadas, porém com conteúdos correlacionados e não triviais, que se encaixam perfeitamente nos moldes de um grupo de estudos como o L²PM.

Referências

- Bergmann, J. and Sams, A. (2014). *Flipped learning: Gateway to student engagement*. International Society for Technology in Education.
- da Silva, F. M. (2003). Aspectos relevantes das novas tecnologias aplicadas à educação e os desafios impostos para a atuação dos docentes. *Akrópolis - Revista de Ciências Humanas da UNIPAR*, 11(2). ISSN: 1982-1093.

- Gomes, A., Henriques, J., and Mendes, A. J. (2008). Uma proposta para ajudar alunos com dificuldades na aprendizagem inicial de programação de computadores. *Educação, Formação & Tecnologias*, 1(1):93 – 103.
- Helmer, E. A. (2014). Utilização de estratégias de ensino diversificadas na sala de aula : Estudo de caso em um curso superior de tecnologia. *Seminário Nacional de Educação Profissional e Tecnológica*.
- Júnior, J. C. R. P. and Rapkiewicz, C. E. (2004). O processo de ensino-aprendizagem de fundamentos de programação: Uma visão crítica da pesquisa no brasil.
- Larrosa, J. (2002). Notas sobre a experiência e o saber da experiência. *Revista Brasileira de Educação*, 1(19).
- Lopes, M. C. (2006). Uma experiência de ensino-aprendizagem em uma disciplina de programação. *Anais do XV SEMINCO - Seminário de Computação*, pages 109 – 118.
- Martins, S. W., Mendes, A. J., and Figueiredo, A. D. (2010). Comunidades de investigação em programação: Uma estratégia de apoio ao aprendizado inicial de programação. *Revista Iberoamericana de Tecnologias da Aprendizagem*, 5(1):39 – 46.
- Raabe, A. L. A. and da Silva, J. M. C. (2005). Um ambiente para atendimento as dificuldades de aprendizagem de algoritmos. *Congresso da Sociedade Brasileira de Computação (WEI 2005)*.
- Raabe, A. L. A. and Giraffa, L. M. M. (2006). Uma arquitetura de tutor para promover experiências de aprendizagem mediadas. *Simpósio Brasileiro de Informática na Educação*, 16(2002):268–277.
- Vincent, B. B. (2013). *As vertigens da tecnociência*. São Paulo: Ideias Letras.

LFApp: Um Aplicativo Móvel para o Ensino de Linguagens Formais e Autômatos

Juventino Neto, Ricardo Terra

Departamento de Ciência da Computação,
Universidade Federal de Lavras (UFLA), Brasil

jneto@computacao.ufla.br, terra@dcc.ufla.br

Resumo. *Linguagens Formais e Autômatos (LFA) é uma importante área que aborda modelos matemáticos que possibilitam a especificação e reconhecimento de linguagens, suas propriedades e características. Embora o conhecimento sólido em LFA seja de extrema importância para a formação de um bacharel em Ciência da Computação e áreas afins, os algoritmos e técnicas abordadas na disciplina são complexos e de difícil assimilação. Diante disso, este trabalho apresenta LFApp, um aplicativo móvel para ensino de LFA. O aplicativo – desenvolvido para celulares e tablets com sistema operacional Android – provê aos alunos não somente a resolução de problemas envolvendo Linguagens Regulares e Linguagens Livres de Contexto, mas também, uma interface com caráter acadêmico que descreve e ilustra cada etapa da execução dos algoritmos de modo a apoiar os alunos no processo de aprendizagem.*

Abstract. *Formal Languages and Automata (FLA) deals with mathematical models that enable the specification and recognition of languages, their properties and characteristics. Although solid knowledge in FLA is extremely important for a B.Sc. degree in Computer Science and similar fields, the algorithms and techniques covered in the course are complex and difficult to assimilate. Therefore, this paper presents LFApp, a mobile application for teaching FLA. The application—developed for mobile phones and tablets running Android—provides students not only the solving of problems involving Regular and Context-free Languages, but also an Educational environment that describes and illustrates each step of the algorithms in order to support students in the learning process.*

1. Introdução

Linguagens Formais e Autômatos (LFA) é uma importante área da computação que aborda modelos matemáticos que possibilitam a especificação e reconhecimento de linguagens, suas propriedades e características [10]. Considerando a importância que um conhecimento sólido de LFA tem sobre o perfil de um bacharel da área de Ciência da Computação e observando um nível de reprovação de quase 50% na disciplina [11], foi proposta uma plataforma que possa auxiliar os alunos no processo de aprendizagem.

Diante disso, este artigo apresenta LFApp, um aplicativo móvel para ensino de LFA que provê aos alunos não somente a resolução de problemas envolvendo Linguagens Regulares e Linguagens Livres de Contexto, mas também uma interface com caráter acadêmico – sendo essa a principal contribuição deste trabalho – que descreve e ilustra cada etapa da execução dos algoritmos de modo a apoiar os alunos no processo de

aprendizagem. É importante mencionar que o aplicativo foi desenvolvido para celulares e *tablets* com sistema operacional (SO) Android e possui o código aberto visando a ampla divulgação, uma vez que o Android é o SO mais utilizado por dispositivos móveis [7].

Em estudo prévio, notou-se que estudantes apresentavam baixo rendimento em algoritmos relacionados a gramáticas [11]. No intuito de cobrir tal deficiência, LFApp foca em Linguagens Livres de Contexto, implementando as seguintes funcionalidades: identificação do tipo de gramática, derivação, remoção de recursão no símbolo inicial, remoção de produções vazias, remoção de regras de cadeia, remoção de símbolos não-terminais, remoção de símbolos não alcançáveis, Forma Normal de Chomsky, remoção de recursão direta à esquerda, remoção de recursão indireta à esquerda, Forma Normal de Greibach e algoritmo de reconhecimento CYK (Cocke-Younger-Kasami) [10].

Este artigo está organizado como a seguir. A Seção 2 apresenta a análise do contexto em que o aplicativo melhor se adaptaria e as motivações para a sua criação, enquanto que a Seção 3 apresenta o aplicativo desenvolvido e o que espera-se com seu uso. A Seção 4 apresenta seu projeto, implementação e ilustra os testes de unidade realizados. A Seção 5 lista os trabalhos relacionados. Por fim, a Seção 6 conclui relatando as principais contribuições e enumerando trabalhos futuros.

2. Contextualização

O aplicativo foi desenvolvido com a motivação de apoiar os alunos no processo de aprendizagem de algoritmos relacionados a gramáticas, os quais os estudantes mais se queixavam [11]. Dessa forma, o aplicativo – como principal elemento pedagógico – reflete exatamente os passos seguidos pelo professor em sala de aula. A seguir, destacam-se conceitos importantes para a compreensão correta deste trabalho.

Hierarquia de Chomsky: Proposta por Noam Chomsky em 1956, a Hierarquia de Chomsky constitui uma classificação para linguagens, gramáticas e autômatos [2] que é composta por quatro categorias, sendo cada uma delas um subconjunto próprio da categoria superior. As quatro categorias são: Linguagens Recursivamente Enumeráveis (0), Linguagens Sensíveis ao Contexto (1), Linguagens Livres de Contexto (2) e Linguagens Regulares (3). Os níveis 2 e 3 são amplamente empregados na descrição de linguagens de programação e na implementação de interpretadores e compiladores [1]. Este trabalho mantém como foco a categoria de nível 2, Linguagens Livres de Contexto, empregando algoritmos e técnicas que atuam sobre esse nível da hierarquia.

Gramática: É um conjunto de regras de produções no seguinte formato $\mu \rightarrow \nu$ que permitem gerar uma linguagem. A definição formal de uma gramática é uma quádrupla (V, Σ, P, S) , onde V representa o conjunto de símbolos não-terminais (variáveis), Σ o conjunto de símbolos terminais, P o conjunto de regras e S o símbolo não-terminal inicial [10]. Segundo a Hierarquia de Chomsky, a Tabela 1 descreve os quatro tipos de gramáticas, em que cada tipo representa uma categoria.

Tabela 1. Hierarquia de Chomsky

Tipo	Linguagem	Gramática	Regras $\mu \rightarrow \nu$
0	Rec. Enumerável	Irrestrita (GI)	$\mu \in (V \cup \Sigma)^+, \nu \in (V \cup \Sigma)^*$
1	Recursiva	Sensível ao Contexto (GSC)	$\mu \in (V \cup \Sigma)^+, \nu \in (V \cup \Sigma)^+, \mu \leq \nu $
2	Livre de Contexto	Livre de Contexto (GLC)	$\mu \in V, \nu \in (V \cup \Sigma)^*$
3	Regular	Regular (GR)	$\mu \in V, \nu \in \lambda \mid \Sigma \mid \Sigma V$

Forma Normal: GLCs são muito flexíveis na formação de suas regras, o que pode ser vantajoso no processo de criação de uma gramática. No entanto, essa flexibilidade não

define uma estrutura padrão para esse tipo de gramática, o que não permite estabelecer relações gerais sobre a gramática, derivações e linguagens [10]. Com o objetivo de tornar as GLCs mais concisas, formas normais impõem regras e simplificam a representação, sem alterar o poder de expressão da linguagem. A Tabela 2 descreve a Forma Normal de Chomsky (FNC) e Greibach (FNG), e o formato de suas regras.

Tabela 2. Formais Normais

Forma Normal de Chomsky (FNC)		Forma Normal de Greibach (FNG)	
$A \rightarrow BC$	onde $B, C \in V$	$A \rightarrow aA_1A_2...A_n$	onde $A_i \in V - \{S\}$
$A \rightarrow a$	onde $a \in \Sigma$	$A \rightarrow a$	onde $a \in \Sigma$
$S \rightarrow \lambda$		$S \rightarrow \lambda$	

Derivação: Consiste no processo de substituição dos símbolos lidos em uma dada palavra pelas regras de uma gramática. Dada uma gramática G e uma palavra p , o objetivo é verificar se a partir de seu símbolo inicial S deriva-se p , i.e., $S \Rightarrow^* p$. As derivações podem ser mais à esquerda ou mais à direita. A derivação mais à esquerda consiste em substituir os símbolos mais à esquerda da forma sentencial, enquanto que, de forma análoga, a mais à direita consiste em substituir os símbolos mais à direita. O processo de derivação ocorre em todas as categorias das linguagens presentes na Hierarquia de Chomsky [10].

3. Solução Proposta

Esta seção apresenta LFApp, um aplicativo que provê a implementação de diversas funcionalidades em um ambiente de caráter acadêmico. O objetivo é prover o aluno com uma descrição sobre o processo de resolução de forma a estimular seu aprendizado. Nas subseções seguintes são descritas e ilustradas as funcionalidades implementadas. Para ilustrar as funcionalidades, adota-se a seguinte GLC, a qual foi projetada de forma a ser possível observar as transformações propostas pela maioria dos algoritmos implementados:

$$\begin{array}{lll}
 S \rightarrow aABC \mid a \mid S & B \rightarrow bcB \mid bc & D \rightarrow abc \mid E \\
 A \rightarrow aA \mid \lambda & C \rightarrow cC \mid c & E \rightarrow Ea \mid a
 \end{array}$$

Basicamente, a interação entre o usuário (aluno) e a ferramenta ocorre conforme ilustrado pela Figura 1. O aluno informa a gramática e todos os algoritmos são computados.

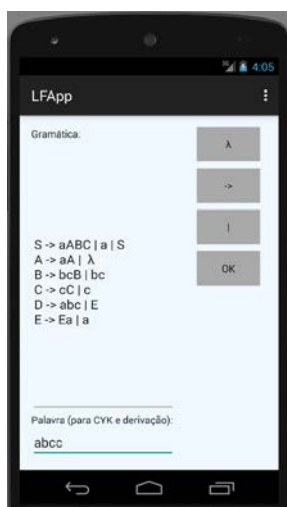


Figura 1. Tela inicial do LFApp

Identificação da gramática

A classificação de uma gramática é feita pelo tipo de suas regras ($u \rightarrow v$). A tabela abaixo mostra o formato de regras características de cada nível:

(3) GR	$u \in V$	$v \in \lambda \mid \Sigma \mid \Sigma V$
(2) GLC	$u \in V$	$v \in (V \cup \Sigma)^*$
(1) GSC	$u \in (V \cup \Sigma)^+$	$v \in (V \cup \Sigma)^+$
(0) GI	$u \in (V \cup \Sigma)^+$	$v \in (V \cup \Sigma)^*$

Resultado:

- Na gramática informada, a regra $S \rightarrow aABC$ não pertence ao conjunto das gramáticas regulares. Logo, a gramática inserida é uma Gramática Livre de Contexto (GLC).

Figura 2. Identificação do tipo da gramática

Identificação do tipo de gramática: Essa funcionalidade consiste em identificar em qual nível da hierarquia de Chomsky a gramática se encontra. A Figura 2 ilustra o resultado dessa funcionalidade na gramática de referência. Basicamente, o algoritmo analisa o formato das regras da gramática e a classifica no nível mais restrito possível. Nesse exemplo, além de classificar a gramática como GLC, proveu-se uma tabela de classificação de regras e um exemplo de uma regra que não se enquadra na definição do nível superior (GR).

Derivação: Essa funcionalidade realiza o processo de derivação mais à esquerda, i.e., tenta gerar uma dada palavra a partir de uma dada gramática. Essa funcionalidade também avalia se a gramática é ambígua. Verificar se uma gramática é ambígua é considerado um problema indecidível, logo, são realizadas até 1.000 tentativas de encontrar formas diferentes de gerar a mesma sentença. Em caso positivo, a gramática é classificada como ambígua, caso contrário, nada se conclui. A Figura 3 ilustra o processo de derivação sobre a gramática de referência e a palavra *abcc*.

Derivação

Resultado:

```
S => aABC
=> aλBC
=> abcC
=> abcc
```

Palavra: abcc

Ambiguidade: Uma Gramática Livre de Contexto é ambígua caso possua mais de uma derivação mais à esquerda. Para a gramática inserida foram realizadas até 1000 tentativas de encontrar uma ambiguidade, porém, não foi encontrada outra forma de gerar a palavra abcc.

Figura 3. Derivação

Remoção de Recursão no Símbolo Inicial

Resultado:

```
S' -> S
S -> a | aABC | S
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> abc | E
E -> Ea
```

Algoritmo:
O símbolo inicial deve se limitar a iniciar derivações, não podendo ser uma variável recursiva. Logo, não deve ser possível ter derivações do tipo $S \Rightarrow^* \alpha S \beta$.

A gramática $G = (V, \Sigma, P, S)$ possui o símbolo inicial S recursivo. Logo, existe uma GLC $G' = (V \cup \{S'\}, \Sigma, P \cup \{S' \rightarrow S\}, S')$, tal que $L(G) = L(G')$ e o novo símbolo inicial S não é recursivo.

Figura 4. Remoção de recursão no símbolo inicial

Remoção de recursão no símbolo inicial: Essa funcionalidade consiste em transformar a gramática de forma que o símbolo inicial se limite a iniciar derivações. A Figura 4 ilustra o resultado dessa funcionalidade aplicada à gramática de referência. O algoritmo para remoção de recursão no símbolo inicial consiste em verificar se a gramática gera uma derivação do tipo $S \Rightarrow^* \alpha S \beta$. Caso possua, é criado um novo símbolo não-terminal S' e inserida uma nova regra do tipo $S' \rightarrow S$, removendo então a recursão no símbolo inicial. Caso contrário, a gramática não possui recursão no símbolo inicial e não é alterada. Nesse exemplo, além da transformar a gramática, proveu-se uma explicação do problema e do algoritmo descrito acima [10].

Remoção de produções vazias: Durante o processo de derivação de uma palavra, a forma sentencial pode conter variáveis que não contribuem para a formação de uma sentença, as quais são classificadas como produções vazias. Essa funcionalidade remove as produções vazias em uma GLC, garantindo a não contratibilidade da gramática. Para a realização dessa transformação, implementou-se o algoritmo λ -rules [10] que torna a gramática não-contrátil ou essencialmente não-contrátil. A Figura 5 ilustra o resultado da execução dessa funcionalidade na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo proveu uma explicação do processo de determinação das variáveis anuláveis (1), da substituição das variáveis (2) e da remoção dos símbolos λ (3).

Remoção de regras de cadeia: Ocasionalmente ocorrem em GLCs regras com o formato $A \rightarrow B$, as quais são denominadas regras de cadeia. Essas regras não contribuem de forma direta para o processo de formação de uma palavra e devem ser removidas por meio de um algoritmo que localize as regras de cadeia e as substitua pelas produções re-

Remoção de Produções Vazias

Resultado:

```
S -> aBC | a | aABC | S
A -> aA | a
B -> bcB | bc
C -> cC | c
D -> abc | E
E -> a | Ea
```

Algoritmo:

O algoritmo para remoção de regras λ consiste em 3 passos:

(1) Determinar o conjunto das variáveis anuláveis.

NULL = {A | $\{A \rightarrow \lambda\} \in P$ }

repita

PREV = NULL

para cada A $\in V$ faça

se A $\rightarrow w$ e $w \in PREV^*$ faça

NULL = NULL $\cup \{A\}$

até NULL == PREV

	NULL	PREV
(0)	{A}	\emptyset
(1)	{A}	{A}

(2) Adicionar regras em que as ocorrências de variáveis nulas são omitidas. Por exemplo, assumamos a regra $A \rightarrow BABA$ e B é uma variável anulável. Logo, são inseridas as seguintes regras: $A \rightarrow ABA$, $A \rightarrow BAa$ e $A \rightarrow Aa$.

S -> aBC | a | aABC | S

A -> λ | aA | a

B -> bcB | bc

C -> cC | c

D -> abc | E

E -> a | Ea

(3) Remover as regras λ .

S -> aBC | a | aABC | S

A -> λ | aA | a

B -> bcB | bc

C -> cC | c

D -> abc | E

E -> a | Ea

Figura 5. Remoção de produções vazias

ferentes a variável destino da cadeia [10]. Por exemplo, em uma gramática que possua a seguinte regra $A \rightarrow B$, será realizada a substituição de B em A por todas as produções existentes em B. A Figura 6 ilustra o processo de remoção de regras de cadeia na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo proveu o conjunto de cadeias de cada variável (1), o destaque das regras de cadeia (2) e uma explicação do processo de substituição das mesmas (3).

Remoção de Regras de Cadeia

Resultado:

```
S -> a | aABC
A ->  $\lambda$  | aA
B -> bc | bcB
C -> c | cC
D -> a | Ea | abc
E -> a | Ea
```

Algoritmo:

A remoção de regras de cadeia substitui as ocorrências de uma cadeia diretamente pelas regras da variável renomeada.

(1) O primeiro passo do algoritmo é montar as cadeias de cada variável.

CHAIN(A) = {A}

PREV = \emptyset

repita

NEW = CHAIN(A) - PREV

PREV = CHAIN(A)

para cada B \in NEW faça

para cada C \in B faça

CHAIN(A) = CHAIN(A) \cup {C}

até CHAIN(A) == PREV

Variável Cadeia

Variável	Cadeia
(S)	{S}
(A)	{A}
(B)	{B}
(C)	{C}
(D)	{D, E}
(E)	{E}

(2) Destacar as cadeias encontradas.

S -> a | aABC | S

A -> λ | aA

B -> bc | bcB

C -> c | cC

D -> abc | E

E -> a | Ea

(3) Substituir as cadeias encontradas:

S -> a | aABC

A -> λ | aA

B -> bc | bcB

C -> c | cC

D -> a | Ea | abc

E -> a | Ea

Figura 6. Remoção de regras de cadeia

Remoção de variáveis que não geram terminais: Essa funcionalidade consiste em remover as variáveis que não geram terminais da gramática. Para ilustrar o funcionamento dessa funcionalidade sem grandes alterações na gramática de referência, a regra $E \rightarrow a$ foi removida, fazendo com que a variável E não gere terminais. A Figura 7 ilustra o processo de remoção de variáveis que não geram terminais na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo apresenta a construção do conjunto TERM que possui as variáveis que produzem terminais diretamente ou indiretamente (1), e o processo de remoção das variáveis que não estão em TERM (2).

Remoção de símbolos não alcançáveis: Essa funcionalidade remove símbolos não alcançáveis a partir do símbolo inicial S, i.e., aqueles não-terminais A que não aparecem em uma derivação da forma $S \Rightarrow^* \alpha A \beta$. A Figura 8 ilustra o processo de remoção de

Remoção de variáveis que não geram terminais

Resultado:

```
S -> a | aABC | S
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> abc
```

Algoritmo:
 Remove as regras que não geram terminais. Consiste de dois passos:
 (1) Determinar quais variáveis geram terminais direta e indiretamente.
 TERM = {A | existe uma regra A → w ∈ P, com w ∈ Σ*}
 repita
 PREV = TERM
 para cada A ∈ V faça
 se A → w ∈ P e w ∈ (PREV ∪ Σ)* então
 TERM = TERM ∪ {A}
 até PREV == TERM

	TERM	PREV
(0)	{D, A, S, B, C}	∅
(1)	{D, A, S, B, C}	{D, A, S, B, C}

(2) Remover as variáveis que não estão em {S, A, B, C, D}, i.e., {E}.
 S -> a | aABC | S
 A -> λ | aA
 B -> bc | bcB
 C -> c | cC
 D -> abc | E
 E -> Ea

Figura 7. Remoção de símbolos não-terminais

Remoção de símbolos não alcançáveis

Resultado:

```
S -> a | aABC | S
A -> λ | aA
B -> bc | bcB
C -> c | cC
```

Algoritmo:
 Remover as variáveis não alcançáveis no processo de derivação de uma palavra.
 (1) Determinar quais variáveis são alcançáveis a partir do símbolo inicial S.
 REACH = {S}
 PREV = ∅
 repita
 NEW = REACH - PREV
 PREV = REACH
 para cada A ∈ NEW faça
 para cada A → w faça
 adicione as variáveis de w em REACH
 até REACH == PREV

	REACH	PREV	NEW
(0)	{S}	∅	∅
(1)	{A, S, B, C}	{S}	{S}
(2)	{A, S, B, C}	{A, S, B, C}	{A, B, C}

(2) Remover as variáveis que não estão em {S, A, B, C}, i.e., {D, E}.
 S -> a | aABC | S
 A -> λ | aA
 B -> bc | bcB
 C -> c | cC
 D -> abc | E
 E -> a | Ea

Figura 8. Remoção de símbolos não alcançáveis

variáveis inalcançáveis na gramática de referência [10]. Além da transformação realizada, o aplicativo apresenta o construção do conjunto *REACH* que possui as variáveis alcançáveis (1) e o processo de remoção das variáveis que não estão em *REACH* (2).

Forma Normal de Chomsky (FNC): Essa funcionalidade padroniza as regras da gramática em um dos formatos da FNC, como descrito na Seção 2. O algoritmo requer que a GLC de entrada não possua símbolo inicial recursivo, produções vazias, regras de cadeia e símbolos inúteis (variáveis que não geram terminais e símbolos não alcançáveis), logo, a ordem da aplicação dos algoritmos é importante. Portanto, o aplicativo pode automaticamente transformar a gramática para satisfazer as pré-condições do algoritmo utilizado [10]. A Figura 9 ilustra o processo de transformação da gramática de referência para a FNC. Além da transformação realizada sobre a gramática, o aplicativo destaca as regras não válidas na FNC (1) e as novas regras criadas (2).

Forma Normal de Chomsky

Remoção de recursão no símbolo inicial:
Remoção de produções vazias:
Remoção de regras de cadeia:
Remoção de símbolos não terminais:
Remoção de símbolos não alcançáveis:
Forma Normal de Chomsky:

Resultado:

```
S' -> I3 I6 | I3 I4 | a
A -> a | T3 A
B -> T1 T5 | T1 T2
C -> T2 C | c
I1 -> b
T2 -> c
T3 -> a
T4 -> BC
T5 -> T2 B
T6 -> AT7
T7 -> BC
```

Algoritmo:
 Uma GLC G = (V, Σ, P, S) está na Forma Normal de Chomsky se suas regras tem uma das seguintes formas
 - A -> BC onde B, C ∈ V - {S}
 - A -> a onde a ∈ Σ
 - S -> λ

(1) Identificar as regras que não estão na Forma Normal de Chomsky.
 S' -> aBC | a | aABC
 A -> aA | a
 B -> bc | bcB
 C -> c | cC

(2) Transformar tais regras em um dos formatos válidos
 S' -> I3 I6 | I3 I4 | a
 A -> a | T3 A
 B -> T1 T5 | T1 T2
 C -> T2 C | c
 T1 -> b
 T2 -> c
 T3 -> a
 T4 -> BC
 T5 -> T2 B
 T6 -> AT7
 T7 -> BC

Figura 9. Forma Normal de Chomsky

Remoção de recursão direta à esquerda: Essa funcionalidade remove recursões diretas à esquerda, as quais podem produzir *loops* infinitos em analisadores sintáticos descendentes [1]. A Figura 10 ilustra o processo de remoção de recursão direta à esquerda na gramática de referência. Além da transformação realizada sobre a gramática, o aplicativo identifica as regras com recursão (1) e o processo de remoção de tais regras (2).

Remoção de Recursão Direta à Esquerda

Resultado:

```
S -> a | aABC
A -> λ | aA
B -> bcB | bc
C -> cC | c
D -> abc | E
E -> a | aZ1
Z1 -> aZ1 | a
```

Algoritmo:

Recursividade direta à esquerda pode produzir "loops infinitos" em analisadores sintáticos descendentes (top-down).

(1) O primeiro passo é identificar a recursão.

Suponha a regra genérica diretamente recursiva à esq.:
 $A \rightarrow A\mu_1 | A\mu_2 | \dots | A\mu_m | v_1 | v_2 | \dots | v_n$

Regra equivalente não-recursiva à esquerda:
 $A \rightarrow v_1 | v_2 | \dots | v_n | v_1Z | v_2Z | \dots | v_nZ$
 $Z \rightarrow \mu_1Z | \mu_2Z | \dots | \mu_mZ | \mu_1 | \mu_2 | \dots | \mu_m$

Remoção de Recursão Direta e Indireta à Esquerda

Resultado:

```
S -> a | aABC | S
A -> λ | aA
B -> bc | bcB
C -> c | cC
D -> abc | E
E -> a | Ea
```

Algoritmo:

A gramática inserida possui ciclos.
 A regra $S \rightarrow S$ é um ciclo.

Figura 10. Remoção de recursão direta à esquerda

Figura 11. Remoção de recursão indireta à esquerda

Remoção de recursão indireta à esquerda: Essa funcionalidade remove recursões indiretas à esquerda. A Figura 11 ilustra o processo de remoção de recursão indireta à esquerda na gramática de referência. No entanto, o algoritmo utilizado [10] requer que a gramática inserida não possua ciclos, o que impede a execução da funcionalidade na gramática de referência. Caso as pré-condições sejam satisfeitas, além da transformação realizada sobre a gramática, o aplicativo ordena as variáveis (1), localiza as recursões (2) e descreve o processo de alteração das regras (3).

Forma Normal de Greibach (FNG): Essa funcionalidade padroniza as regras da gramática em um dos formatos da FNG, como descrito na Seção 2. O objetivo é garantir que não ocorra recursões diretas e indiretas à esquerda. A Figura 12 ilustra o processo de transformação da gramática de referência para a FNG. Além da transformação realizada sobre a gramática – que usualmente envolve a remoção de recursão à esquerda direta e indireta – o aplicativo destaca as regras não válidas na FNG (1) e as novas regras criadas (2).

Algoritmo de reconhecimento CYK: Essa funcionalidade verifica se uma palavra pode ser gerada por uma GLC. O algoritmo CYK utiliza de uma abordagem *bottom-up* com o intuito de analisar a sentença associando-a com as regras da gramática. A Figura 13 ilustra o processo de reconhecimento da palavra *abcc* na gramática de referência. Além da construção da matriz triangular inferior, o aplicativo ilustra o passo-a-passo de sua construção. Nesse exemplo, o símbolo inicial da gramática (S') encontra-se no topo da matriz, logo pode-se concluir que a palavra *abcc* pertence a gramática.

4. Projeto, Implementação e Testes

Com o intuito de se tornar um recurso educacional aberto, LFApp apoia-se em soluções livres—IDE Android Studio e Android SDK—e utiliza apenas bibliotecas nativas do Android. Mais importante, o aplicativo é classificado como software livre, sendo disponibilizado desde o início de seu desenvolvimento em um repositório público no GitHub.¹ A

¹<https://github.com/rterrahb/LFApp>

Forma Normal de Greibach

Remoção de recursão à esquerda:

Resultado:

$S' \rightarrow T_3T_6 | T_3T_4 | a$
 $A \rightarrow a | T_3A$
 $B \rightarrow T_1T_5 | T_1T_2$
 $C \rightarrow T_2C | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow bT_5C | bT_2C$
 $T_5 \rightarrow cB$
 $T_6 \rightarrow aT_7 | aAT_7$
 $T_7 \rightarrow bT_2C | bT_5C$

Algoritmo:

A remoção de recursão à esquerda consiste em ordenar as variáveis da gramática e organizar as regras da forma que a variável do lado esquerdo sempre possua valor menor do que a variável do lado direito.

(1) Ordenar as variáveis da gramática.

Variável	Valor
S'	1
A	2
B	3
C	4
T1	5
T3	6
T2	7
T6	8
T7	9
T4	10
T5	11

(2) Localizar as recursões.

$S' \rightarrow T_3T_6 | T_3T_4 | a$
 $A \rightarrow a | T_3A$
 $B \rightarrow T_1T_5 | T_1T_2$
 $C \rightarrow T_2C | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow BC$
 $T_5 \rightarrow T_2B$
 $T_6 \rightarrow AT_7$
 $T_7 \rightarrow BC$

(3) Alterar as regras

$S' \rightarrow T_3T_6 | T_3T_4 | a$
 $A \rightarrow a | T_3A$
 $B \rightarrow T_1T_5 | T_1T_2$
 $C \rightarrow T_2C | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow bT_5C | bT_2C$
 $T_5 \rightarrow cB$
 $T_6 \rightarrow aT_7 | aAT_7$
 $T_7 \rightarrow bT_2C | bT_5C$

Forma normal de Greibach:

Resultado:

$S' \rightarrow aT_4 | a | aT_6$
 $A \rightarrow aA | a$
 $B \rightarrow bT_2 | bT_5$
 $C \rightarrow cC | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow bT_5C | bT_2C$
 $T_5 \rightarrow cB$
 $T_6 \rightarrow aT_7 | aAT_7$
 $T_7 \rightarrow bT_2C | bT_5C$

Algoritmo:

Uma GLC G = (V, Σ, P, S) está na FN de Greibach se suas regras têm uma das seguintes formas:

$A \rightarrow aA_1A_2A_3...A_n$ onde $a \in \Sigma$ e $A_1...A_n \in V - \{S\}$
 $A \rightarrow a$ onde $a \in \Sigma$
 $A \rightarrow \lambda$

(1) Localizar as regras que não estão na Forma Normal de Greibach.

$S' \rightarrow T_3T_6 | T_3T_4 | a$
 $A \rightarrow a | T_3A$
 $B \rightarrow T_1T_5 | T_1T_2$
 $C \rightarrow T_2C | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow bT_5C | bT_2C$
 $T_5 \rightarrow cB$
 $T_6 \rightarrow aT_7 | aAT_7$
 $T_7 \rightarrow bT_2C | bT_5C$

(2) Transformar tais regras em um dos formatos válidos

$S' \rightarrow aT_4 | a | aT_6$
 $A \rightarrow aA | a$
 $B \rightarrow bT_2 | bT_5$
 $C \rightarrow cC | c$
 $T_1 \rightarrow b$
 $T_2 \rightarrow c$
 $T_3 \rightarrow a$
 $T_4 \rightarrow bT_5C | bT_2C$
 $T_5 \rightarrow cB$

Figura 12. Forma Normal de Greibach

CYK

Resultado:

	{S, T ₆ }			
		{T ₄ , T ₇ }		
			{B}, {C}	
	{A, S, T ₃ }	{T ₁ }	{C, T ₂ }	{C, T ₂ }
a	b	c	c	

Algoritmo:

(1) O primeiro passo do algoritmo é adicionar à tabela as variáveis que produzem os respectivos terminais diretamente.

Há alguma regra que gere a diretamente? Sim.

$T_3 \rightarrow a, A \rightarrow a, S' \rightarrow a$

	{A, S, T ₃ }	{T ₁ }	{C, T ₂ }	{C, T ₂ }
a	b	c	c	

(2) O segundo passo do algoritmo é adicionar à tabela as variáveis que produzem as sentenças de tamanho dois. Há alguma regra que gere AT₁, ST₁, T₃T₁? Não.

		{B}, {C}	
{A, S, T ₃ }	{T ₁ }	{C, T ₂ }	{C, T ₂ }
a	b	c	c

(3) O terceiro passo do algoritmo é adicionar à tabela as variáveis que produzem as sentenças de tamanho três. Há alguma regra que gere AB, S'B, T₃B, -C, -T₂? Não.

			{T ₄ , T ₇ }
		{B}, {C}	
{A, S, T ₃ }	{T ₁ }	{C, T ₂ }	{C, T ₂ }
a	b	c	c

E assim por diante.

Figura 13. Reconhecimento CYK

Figura 14 apresenta as duas principais classes do aplicativo: Grammar que representa uma gramática de qualquer nível da Hierarquia de Chomsky que é composta por um conjunto de Rule que representa uma regra do tipo $\mu \rightarrow \nu$. Objetos Grammar são imutáveis, logo todas as funcionalidades sempre retornam um novo objeto Grammar como resposta.

Com o objetivo de evitar erros de codificação, todos os algoritmos desenvolvidos foram testados por meio de testes de unidade, que verificam cada unidade da aplicação [8], garantindo o correto funcionamento do LFA. Portanto, utilizando o *framework* JUnit², foram desenvolvidos 39 casos de testes, os quais são executados antes da publicação de uma nova versão do aplicativo. A Figura 15 ilustra um *screenshot* da execução de um teste de regressão no aplicativo. O teste de regressão consiste em submeter as versões mais recentes do software aos mesmos testes que as versões anteriores já foram submetidas [8]. O objetivo é garantir que as funcionalidades previamente existentes permanecem corretas, i.e., o software não regrediu.

5. Trabalhos Relacionados

Existem diversas ferramentas similares à proposta neste artigo. Popular no âmbito de Teoria da Computação e LFA, a ferramenta JFLAP disponibiliza funcionalidades com

²http://www.junit.org

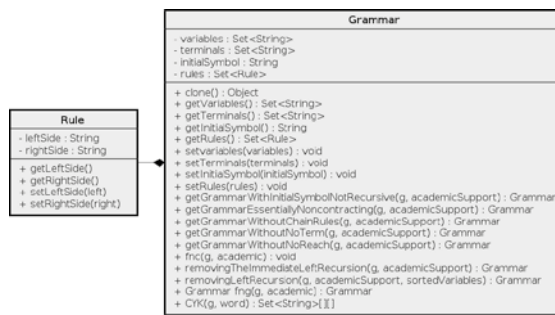


Figura 14. Diagrama de classes

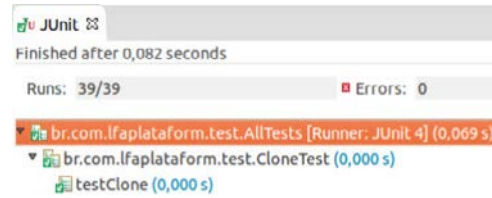


Figura 15. Testes de unidade

foco principal em autômatos. Para a classe das LLCs, o JFLAP disponibiliza quatro tipos de transformações, sendo elas: conversão de APs para GLCs e mais três algoritmos para conversão de GLCs em APs [5]. Mesmo possuindo foco na classe das LLCs, ao contrário de LFApp, o JFLAP não apresenta como foco principal as transformações de GLCs, o que é amplamente abordado neste trabalho.

Yandre et al. [4] propuseram a ferramenta SCTMF (Software para a Criação e Teste de Modelos Formais) com o objetivo de auxiliar o ensino de LFA. A ferramenta consiste em um software para a criação e testes de modelos formais em máquinas aceitadoras da hierarquia de Chomsky, tais como Autômatos Finitos (AFs), Autômatos com Pilha (APs), Autômatos Linearmente Limitados (ALLs) e Máquinas de Turing (MTs). Em contraste com LFApp, seu principal foco é em máquinas aceitadoras, sem possuir algoritmos sobre gramáticas.

Menezes et al. [12] propuseram *Language Emulator*, um ambiente com foco em Linguagens Regulares (LRs) que apoia o ensino de LFA. Dentre suas funcionalidades estão algoritmos sobre Expressão Regular (ER), Gramática Regular (GR), AFs, e máquinas de Mealy e Moore. A ferramenta é completa no que se refere à classe das LRs, implementando as técnicas e algoritmos abordados na área de conhecimento. No entanto, em contraste com LFApp, não possui implementação de modelos referentes à classe das LLCs.

Jukemura et al. [6] descrevem GAM, uma ferramenta que simula diferentes tipos de máquinas abstratas para apoiar o ensino de Teoria da Computação. Dentre suas funcionalidades estão simulação de AFs, transformação de AFND para AFD, transformação de AFD para ER, transformação de ER para AFND- λ , simulação de APs e simulação de MTs. Mesmo possuindo funcionalidades no nível de LLCs (e.g., simulação de AP), o qual é o foco de LFApp, GAM não possui funcionalidades que lidam com transformações de gramáticas e formas normais.

Ainda é importante mencionar alguns jogos educacionais. Morais et al. [3] descrevem o projeto de um jogo educativo baseado em *mobile learning*, visando melhorar o ensino de alguns tópicos básicos em disciplinas introdutórias de Teoria da Computação. Silva et al. [9] descrevem o jogo educacional *Automata Defense 2.0* e a experiência de sua utilização como apoio pedagógico em LFA. Os autores ainda apresentam os resultados de testes de usabilidade e uma avaliação preliminar da sua eficácia pedagógica.

6. Considerações Finais

Linguagens Formais e Autômatos (LFA) é uma importante disciplina da computação que aborda modelos matemáticos que possibilitam a especificação e reconhecimento de lin-

guagens, suas propriedades e características. No entanto, os algoritmos e técnicas abordadas na disciplina são complexos e de difícil assimilação, o que implica em um alto índice de reprovações. Diante disso, este trabalho apresentou LFApp, um aplicativo – desenvolvido para celulares e *tablets* com SO Android – que provê aos alunos não somente a resolução de problemas envolvendo Linguagens Regulares e Linguagens Livres de Contexto, mas também, uma interface com alto caráter acadêmico que descreve e ilustra cada etapa da execução dos algoritmos de modo a apoiar os alunos no processo de aprendizagem.

Como principais contribuições do LFApp, espera-se (i) prover um ambiente de estudos interativo para os alunos de forma a auxiliar no processo de ensino e aprendizagem de LFA e (ii) promover a ideia do uso de aplicativos móveis educacionais. Como trabalhos futuros, pretende-se (i) avaliar o aplicativo no apoio ao ensino da disciplina de LFA, (ii) estender LFApp para máquinas aceitadoras desenvolvendo, por exemplo, algoritmos para simulação de AF/AP/ALL/MT, conversão de AFND para AFD, minimização de AFD, etc., (iii) expandir a ferramenta para abordar todos os níveis da Hierarquia de Chomsky e (iv) desenvolver versões para outras plataformas (iOS e Windows Mobile).

A ferramenta LFApp e seu código fonte estão publicamente disponíveis em:

http://www.dcc.ufla.br/~terra/papers/2016_wei_lfapp

Agradecimentos: Este trabalho foi apoiado pela FAPEMIG, CAPES e CNPq.

Referências

- [1] Alfred V Aho, Ravi Sethi, and Jeffrey D Ullman. *Compiladores: Princípios, técnicas e ferramentas*. LTC, 2 edition, 2008.
- [2] Noam Chomsky. Three models for the description of language. *IRE Transactions on Information Theory*, 2(3):113–124, 1956.
- [3] Dyego Carlos Sales de Moraes, Antônio DPC Alencar, and Rodrigo de Souza. Jogo baseado em m-learning e aprendizado tangencial para auxílio ao ensino de teoria da computação. In *XXII Simpósio Brasileiro de Informática na Educação*, pages 554–557, 2011.
- [4] Yandre M. e G. da Costa, Rafael C. de Meneses, and Flávio R. Uber. Uma ferramenta para auxílio didático no ensino de teoria da computação. In *XVI Workshop sobre Educação em Computação (WEI)*, pages 208–217, 2008.
- [5] Eric Gramond and Susan H Rodger. Using JFLAP to interact with theorems in automata theory. *ACM SIGCSE Bulletin*, 31(1):336–340, 1999.
- [6] Anibal S. Jukemura, Hugo A. D. do Nascimento, and Joaquim Q. Uchôa. GAM: Um simulador para auxiliar o ensino de linguagens formais e de autômatos. *XIII Workshop sobre Educação em Computação (WEI)*, pages 2432–2443, 2005.
- [7] Farhad Manjoo. A murky road ahead for Android, despite market dominance. *The New York Times*, 2015.
- [8] Roger S Pressman. *Engenharia de software*. Bookman, 7 edition, 2011.
- [9] Rômulo César Silva, Ricardo Luis Binsfeld, Izaura Maria Carelli, and Rodrigo Watanabe. Automata defense 2.0: reedição de um jogo educacional para apoio em linguagens formais e autômatos. In *XXI Simpósio Brasileiro de Informática na Educação*, 2010.
- [10] Thomas A. Sudkamp. *Languages and machines: an introduction to the theory of Computer Science*. Addison-Wesley, 3 edition, 2005.
- [11] Ricardo Terra. Dados da disciplina de Linguagens Formais e Autômatos. Technical report, Universidade Federal de Lavras (UFLA), 2016.
- [12] Luiz Filipe M. Vieira, Marcos Augusto M. Vieira, and Newton J. Vieira. Language emulator, a helpful toolkit in the learning process of computer theory. In *35th SIGCSE Technical Symposium on Computer Science Education (SIGCSE)*, pages 135–139, 2004.

Maratona para a Popularização da Ciência da Computação na Educação Básica

¹Michel Girotto Brum ²Marcia E. J. Kniphoff da Cruz,

^{1,2}Departamento de Computação Curso de Licenciatura em Computação
Universidade de Santa Cruz do Sul – UNISC
Santa Cruz do sul - RS – Brasil

^{1,2} michelgb@unisc.br e marciakniphoff@gmail.com ou mcruz@unisc.br

Resumo. Diferentes ações têm sido empreendidas para difundir o ensino de Ciência da Computação na Educação Básica. Com o objetivo de fomentar essa proposta, este relato apresenta a Maratona UNISC de Computação – MUC. A MUC é uma olimpíada científico/tecnológica que objetiva proporcionar novos desafios na área da Computação a estudantes de escolas de Ensino Fundamental e incentivar o interesse pela área. A primeira edição realizada no ano de 2015, oportunizou aos estudantes a resolução de desafios teóricos e problemas em linguagem de programação. Doze escolas da região do Vale do Rio Pardo e Taquari participaram. Uma nova edição está prevista para o ano de 2016 com a participação do Ensino Médio. Pretende-se ampliar a abrangência da MUC para outras regiões do país e incentivar a competição saudável na área da Computação, junto à Educação Básica.

Palavras chave: Scratch, Maratona, Computação.

Abstract. Different actions have been undertaken to spread the Computer Science education in basic education. In order to promote this proposal, this report presents the first Maratona UNISC de Computação - MUC. The MUC is a scientific/technological Olympiad which aims to provide new challenges in the field of Computing students from elementary schools and encourage interest in the area. The first edition held in 2015, provided an opportunity to students to solve theoretical problems and challenges in programming language. Twelve schools in the Rio Pardo and Taquari Valley region participated. A new edition is scheduled for the year 2016 with the participation of high school. It is intended to expand the scope of MUC to other regions of the country and encourage healthy competition in computer area, next to the Basic Education.

Keywords: Scratch, Marathon, Computer.

1. A CIÊNCIA DA COMPUTAÇÃO E A EDUCAÇÃO BÁSICA

O desenvolvimento social e econômico de um país está diretamente relacionado ao seu potencial científico e tecnológico. No Brasil, mesmo com todos os esforços e ações para difundir o ensino da Ciência da Computação, observa-se que um há longo caminho a ser trilhado para efetivar estudos da área, na Educação Básica. Em análise dos artigos publicados nos anais do XXIII Workshop sobre Educação em Computação - WEI observa-se as grandes dificuldades do ensino da programação nos cursos de ensino superior, conforme citado por Barbosa (2011) e Correia (2015). Esses estudos, ainda, apontam dados alarmantes sobre a evasão dos cursos de Computação; o assunto tem se tornado pauta de encontros nacionais, segundo Massa (2015). Os problemas referentes podem ser equalizados com a introdução de estudos de Ciência da

Computação na Educação Básica, através do desenvolvimento do Pensamento Computacional e de lógica para programação. Segundo Andrade et al (2013), o Pensamento Computacional se utiliza de atividades que introduzem os conceitos de coleta, análise e representação de dados, decomposição de problemas, abstração, automação, simulação, algoritmos e paralelismo, sem o uso de computadores. Em Ciência da Computação essas habilidades são básicas para o desenvolvimento posterior de algoritmos e programação.

Algumas escolas já estão incorporando o ensino de Computação nas suas rotinas diárias. O Scratch tem sido utilizado para o ensino da programação, por ser uma ferramenta visual e lúdica, mas poderosa. Permite a elaboração de algoritmos complexos. Outros trabalhos como os de Rabêlo (2015) e Batista (2015), relatam formas criativas e inovadoras de levar o ensino da computação de forma atrativa aos estudantes, utilizando Scratch, para elaboração de animações e jogos. A necessidade de oportunizar o desenvolvimento do Pensamento Computacional aos estudantes da Educação Básica está relacionada ao grau de complexidade dos problemas enfrentados no cotidiano. Ribeiro (2013), afirma que, noções matemáticas trabalhadas nas séries iniciais já foram conhecimentos apenas de intelectuais em séculos passados, contudo, hoje a programação de computadores está se tornando essencial.

Códigos de programação são partes integrantes de muitos dos recursos cotidianos, mas poucos tem o conhecimento par elaborá-los, o que justifica a necessidade emergente de repensar o currículo da Educação Básica do Brasil e atualizá-lo. É importante lembrar, segundo Ribeiro (2013), que o pensamento lógico-matemático difere do Pensamento Computacional; o raciocínio computacional utiliza a lógica como uma ferramenta no processo de resolução de problemas. Significa decompor os problemas em partes menores, para que possam ser resolvidos e combinados em função do resultado final, tornando o processo todo muito mais complexo que a simples tomada de decisão. Muitas escolas, ainda veem o ensino de informática apenas como a necessidade de dominar o uso de softwares e não como uma forma de desenvolver habilidades que estão se tornando necessárias, cada vez mais cedo, na sociedade, o que justifica a necessidade de incansáveis esforços para popularizar a Ciência da Computação na Educação Básica. Portanto, as iniciativas que promovem eventos com a finalidade de popularizar da Ciência da Computação são, cada vez mais, necessárias. Essas iniciativas incentivam as instituições de Educação Básica a incorporá-la em seus currículos regulares. Com a intenção de incentivar tais iniciativas, este artigo apresenta o desenvolvimento e os resultados de uma maratona de computação para estudantes das séries finais do Ensino Fundamental, que ocorre. A maratona ocorreu através de uma competição saudável¹.

2. OLIMPÍADAS CIENTÍFICAS

Olimpíadas científicas objetivam difundir o ensino da ciência e da tecnologia e identificar talentos em potencial, para incentiva-los a seguir carreiras nas áreas de ciência e tecnologia. No Brasil, as olimpíadas tiveram seu início em 1978 e desde 2002 vem recebendo apoio do poder público. Hoje, diversas olimpíadas científicas como a Olimpíada Brasileira de Robótica – OBR, Olimpíada Brasileira de Informática – OBI são fomentadas por órgãos como o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Ministério da Educação em parceria com a Fundação Nacional de Desenvolvimento da Educação (FNDE/MEC). Alguns

¹ A competição saudável incentiva o desafio como forma de superação pessoal e a disputa com demais colegas torna-se um benefício comum, pois os resultados da competição incentivam a busca por mais conhecimento.

eventos como a OBI utilizam os resultados das provas na modalidade teórica para selecionar e formar equipes que irão representar o Brasil em outras olimpíadas de nível internacional, como a International Olympiad in Informatics – IOI. A primeira IOI foi proposta na vigésima quarta conferência da United Nations Educational, Scientific and Cultural Organization (UNESCO), em Paris no ano de 1987, que também patrocinou a primeira IOI, realizada na Bulgária em 1989. Desde então, outras edições vêm acontecendo anualmente, sendo que, a cada ano a IOI é sediada em um país diferente. As tarefas nessas olimpíadas versam sobre algoritmos e desafios em tecnologias de informação. No caso da IOI, versam também sobre análise de problemas, estruturas de dados e design de algoritmos. Além dos eventos nacionais, diversas instituições de nível superior promovem olimpíadas científicas regionais voltadas à Educação Básica.

3. MARATONA UNISC DE COMPUTAÇÃO

A Maratona UNISC de Computação – MUC foi promovida pelo curso de Licenciatura em Computação e Projeto UNISC Inclusão Digital². Aconteceu na Universidade de Santa Cruz do Sul – UNISC e envolveu escolas de Educação Básica do Vale do Rio Pardo e Taquari. A MUC teve por objetivos:

- Proporcionar novos desafios na área da Computação a estudantes de escolas de Educação Básica e de organizações educativas.
- Incentivar o interesse pela área da Computação, introduzindo o pensamento lógico, através da resolução de problemas e de técnicas de programação de computadores.
- Identificar e premiar talentos na área da Computação e incentivá-los a seguir carreiras nas áreas de ciência e tecnologia.

No primeiro ano, a MUC aconteceu em duas modalidades: Nível 1 e Nível 2, contando ambas as modalidades com dois desafios. O primeiro desafio foi a resolução da atividade “circuito de questões” realizado no turno da manhã. O segundo, a resolução de atividade em laboratório de informática, no desenvolvimento de “Desafios de Programação”, com tarefas apropriadas para cada modalidade, no turno da tarde. Para a próxima edição, estuda-se a ampliação para três níveis, com o objetivo de envolver também o público do Ensino Médio. No Nível 1, as questões do circuito versaram sobre problemas de Lógica e Computação. Na atividade desafio de programação, as duplas receberam programas já criados, devendo montar os algoritmos para que o programa funcionasse, conforme objetivo indicado nas tarefas, utilizando a linguagem visual Scratch 2. No Nível 2, as questões do circuito versaram sobre problemas de Lógica e Computação. Na atividade desafio de programação, as duplas recebem enunciados de programas, tendo como desafio desenvolver algoritmo e a interface para que o programa funcionasse.

Foi desenvolvido um *hotsite* para a divulgação da MUC, inscrições e divulgação dos resultados. Para a elaboração dos formulários de inscrição, foi utilizado a plataforma web JotForm, que permite a integração de formulários web em qualquer site. O site da MUC

² Foram apoiadores da MUC, ainda: Núcleo de Socialização Ciência e Tecnologia – NTSC, Programa UNISC Escola, Projeto de Pesquisa O uso de estilos cognitivos e agentes pedagógicos no processo de ensino-aprendizagem em sistemas tutores inteligentes e Subprojeto de Informática PIBID/UNISC.

encontra-se disponível na Internet e foi desenvolvido utilizando a plataforma web gratuita Weebly. A maratona aceitou equipes formadas por duplas de alunos, acompanhadas por professor responsável. Para popularizar o evento não foi estimulado limite de inscritos por escola nem por professor orientador, que pode inscrever alunos de mais de uma modalidade ou escola.

3.1 Elaboração das atividades

Para a elaboração das atividades, realizou-se um estudo das últimas cinco edições da OBI e OBR, baseando-se nos assuntos abordados nessas provas de modalidade teórica. As questões não envolvem conteúdos que não estejam em livros ou que não sejam abordados na escola, mas explora a capacidade dos competidores em decompor um problema maior e resolvê-lo, abstrair e apresentar uma solução.

A MUC foi dividida em três partes distintas: O circuito de questões, a tarefa cultural e os Desafios de Programação. A primeira, aconteceu na parte da manhã, em auditório, com espaço adequado. A segunda em laboratórios de informática no período da tarde. Ambas as partes tiveram duração de três horas. Além das provas, o evento alocou um tempo de duas horas para o intervalo entre o Circuito de Questões e os Desafios de Programação. O amplo tempo livre foi dedicado para a integração entre os participantes. Nesse momento, realizaram uma tarefa extra denominada “Tarefa Cultural”, com um propósito lúdico, que foi liberada ao final da manhã.

Para a elaboração do edital e das questões e desafios, reuniu-se um comitê organizador, sendo este composto por estudantes do curso de Licenciatura em Computação, integrantes do diretório acadêmico e professores do Departamento de Computação.


3.1.1 Circuito de questões

O circuito de questões foi composto por três provas. Cada uma das provas iniciou com uma situação problema com cinco alternativas de respostas, sendo apenas uma correta. O primeiro desafio para o estudante foi interpretar e entender o problema. Para a resolução precisaram escolher uma abstração, empregar lógica e algoritmos. Cada prova foi entregue em horário pré-determinado, sendo que, para o recebimento da próxima prova, a dupla deveria entregar no ato a prova anterior, com sua resolução. Para cada nível, foram desenvolvidas provas distintas, com diferentes graus de dificuldade. A Figura 1 apresenta uma questão teórica do circuito de questões:

DESAFIO 2 (20 PONTOS)

Baralho Digital

Ana e Bernardo criaram um novo jogo usando um baralho. No novo jogo, o baralho deve conter 48 cartas, igualmente divididas entre pretas e vermelhas. Nesse jogo, as cartas são dispostas na mesa viradas para baixo, e cada jogador deve virar uma carta por vez. Ao virar a carta, ele retira da mesa e conta a pontuação de acordo com o algoritmo ao lado. Uma rodada completa acontece quando os dois jogadores retiraram uma carta. Para definir o número de rodadas um dos participantes deve jogar 4 dados comuns na mesa e somar os valores que ficaram para cima dos 4 dados.



```

quando receber Nova Carta
se Nova Carta = Vermelha então
se Nova Carta = Par então
  adicione a Jogador A 2
senão
  adicione a Jogador B 1
senão
se Nova Carta = Impar então
  adicione a Jogador A 1
senão
  adicione a Jogador B 2
                    
```

Figura 1. Exemplo de questão teórica presente na prova do circuito de questões

Na Figura 1 pode-se observar uma situação problema, na qual é necessário aplicar um algoritmo para responder corretamente a perguntas. Para a correta resolução, os estudantes precisaram empregar habilidades relacionadas ao Pensamento Computacional. A pontuação foi dada em função de condições estabelecidas previamente, nas regras.

3.1.2 Tarefa Cultural

A Tarefa Cultural solicitou que os participantes fotografassem frases pintadas em prédios da universidade e publicassem no modo público, em redes sociais online, marcadas com *hashtags* #MUCÉShow. As *hashtags* são utilizadas em meios de publicação com a finalidade de organizar e agrupar conteúdo durante a busca pelos usuários, servindo como filtro. Essa tarefa teve importante papel na divulgação do evento. A Figura 2 apresenta um exemplo de fotografia:



A Figura 2 apresenta exemplo de fotografia esperada dos estudantes

3.1.3 Desafios de Programação

Para a resolução dos Desafios de Programação, cada dupla teve à disposição um computador, em laboratório de informática, com o software Scratch 2.0, ferramenta de programação em linguagem visual desenvolvida pelo Massachusetts Institute of Technology – MIT. A Figura 2 apresenta uma questão do Desafios de Programação:

Desafio 2:

Adalberto é um menino muito preguiçoso, mas muito inteligente. Em seu quarto, há três móveis de mesmo tamanho, mas em função de sua desorganização, sempre os deixa em lugares diferentes. Para realizar a limpeza do ambiente ele criou um robô com um aspirador de pó, capaz de se locomover sozinho. Na sequência, pode-se ver uma simulação do robô. Você deverá:

- Criar o programa de computador que fará com que o robô ande pelo quarto limpando o ambiente, sem passar por cima dos móveis.
- Nessa simulação, você deverá riscar com a caneta o local, pelo qual, o robô está passando.
- O quarto é considerado limpo quando o robô passar, ao menos 1 vez, por todos os azulejos. Não há tempo limite para limpeza.
- Toda vez que o robô iniciar uma limpeza, os móveis devem reaparecer em lugares aleatórios no quarto.

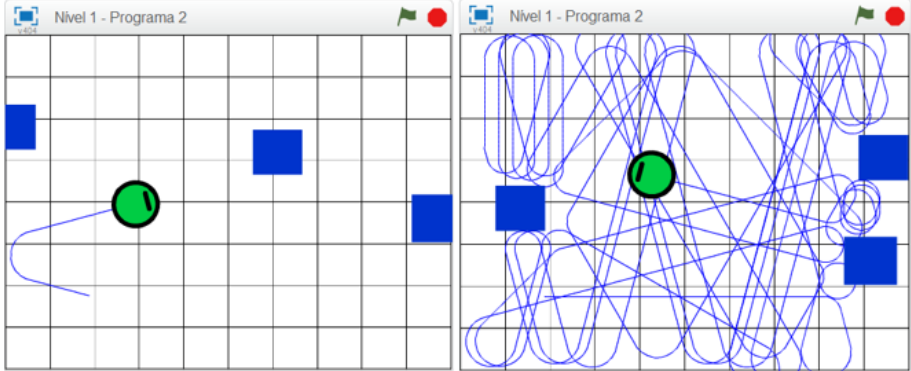


Figura 2. Desafio da sessão “Desafios de Programação”, apresentado nos dois níveis da MUC

Esse desafio exigiu dos participantes conhecimento sobre *loop*, condicionais, operadores lógicos, aritméticos e sensores. Para cada nível o desafio 1 apresentou grau de dificuldade menor, envolvendo programação, apenas de movimento e comandos de aparência para criar animações. No último desafio o grau de complexidade e dificuldade foi elevado. Para a resolução foram necessárias habilidades para trabalhar com dados de entrada do usuário. No caso do Nível 1 o desafio envolveu operações matemáticas com variáveis e tomadas de decisões. No caso do Nível 2 operações lógicas e aritméticas envolvendo listas de dados, tomadas de decisões e noções de estrutura de dados.

4. RESULTADOS, AVALIAÇÃO E PROPOSIÇÕES FUTURAS

Participaram da MUC doze escolas públicas e privadas, e instituições educacionais, quinze professores da Educação Básica, dezesseis bolsistas e alunos do curso de Licenciatura em Computação e cinco professores do Departamento de Computação, além dos cento e seis estudantes envolvidos nas oito horas de maratona.

Todas as provas foram corrigidas por três integrantes do comitê avaliador. O Quadro 1 apresenta o cronograma da MUC e avaliação:

Atividade	Início	Fim	Pontuação
Abertura e Validação da Inscrição	8 horas	9 horas	-
Problema 1	9 horas	10 horas	20
Problema 2	10 horas	10h45min	15
Problema 3	10h45min	11h30min	15
Tarefa cultural	11h30min	13h30min	5
Intervalo para Almoço	11h45min	13h30min	-
Desafio de Programação em <i>Scratch</i>	13h30min	16h30min	45
Envio da Atividade de Programação	16h30min	17 horas	-

Quadro 1: Horários e pontuação das tarefas da MUC

Ao final da maratona, a soma das tarefas foi de 100 pontos, distribuídos em 50 pontos para Circuito de Questões, 5 pontos para Tarefa Cultural e 45 pontos para os Desafios de Programação. Após a avaliação das 61 duplas participantes, constatou-se a grande heterogeneidade dos estudantes, relativa aos conhecimentos de lógica de programação e Pensamento Computacional. A Figura 3 apresenta dados: 49% dos participantes não obtiveram 25 pontos; apenas 20% dos participantes tiveram mais de 50 pontos, conforme apresenta a Figura 3:

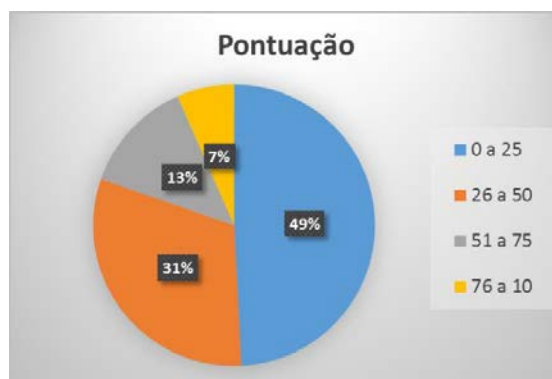


Figura 3: Porcentagem de estudantes por faixa de pontuação

O resultado da primeira edição demonstra a necessidade de uma atualização no currículo escolar; há urgência em inserir Pensamento Computacional e técnicas de programação junto ao ensino de informática nas escolas de Educação Básica.

Na avaliação dos professores orientadores das escolas, a maratona foi muito positiva. Consideraram a duração do evento apropriada. Quanto ao tempo disponível para a resolução das atividades, 67% indicaram ter sido apropriado para as tarefas, 33% indicaram que para o tempo disponível as tarefas poderiam ter sido maiores. Quanto ao grau de dificuldade das tarefas, 67% indicaram ter sido apropriado para o nível dos estudantes, e 33% indicaram que a dificuldade foi alta. Foi questionado também o grau de satisfação geral dos participantes quanto à organização do evento. Neste quesito, solicitou-se a atribuição de uma nota de 1 a 5, sendo 1 para muito insatisfeito, e 5 para muito satisfeito. Conforme a Figura 5, a maior parte atribuiu nota 4.

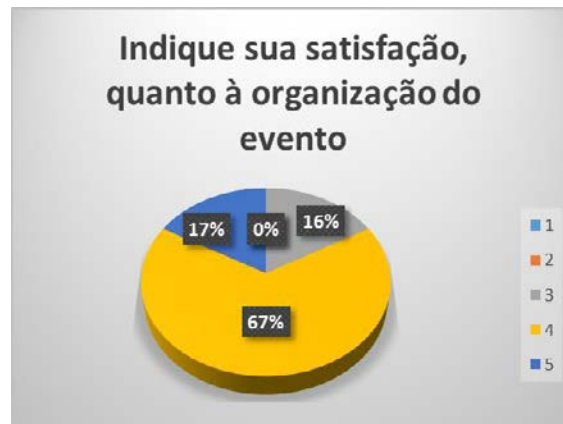


Figura 5: Pesquisa com professores e responsáveis

Durante o intervalo os organizadores conversaram com os alunos participantes, professores e responsáveis, a fim de ouvi-los sobre o andamento da maratona. Os participantes descreveram o evento de forma muito positiva. Entre os aspectos mais importantes foi citada a integração com alunos de outras escolas que também aprendem computação, os desafios de trabalho em conjunto, a proximidade dos desafios com o mundo real. Também relataram terem visto o evento como uma forma de incentivo ao aprendizado de programação, pouco enfatizado na maioria das escolas. “Foi uma experiência muito divertida que me ensinou a trabalhar em equipe e me motivou a continuar aprendendo sobre programação”, comentou um estudante. Para um dos ganhadores, “[...] foi uma maratona que me ensinou duas coisas importantes, que não importa se ganhamos ou perdemos, e a trabalhar em equipe com minha dupla”. Durante a premiação uma campeã acrescentou que “a MUC foi um grande evento educacional, onde muitos alunos tiveram a oportunidade de participar e acrescentar conhecimento sobre algo tão interessante como a computação”.

Com a realização da MUC, conclui-se que a oferta de competição saudável é uma excelente estratégia educativa, no que se refere ao incentivo do Pensamento Computacional para a Educação Básica. Os estudantes mostraram muito interesse em ampliar conhecimentos de programação; os professores sensibilizaram-se sobre a importância do ensino de computação para a formação dos alunos.

Como proposição futura pretende-se estender a participação para o Ensino Médio e gradativamente, ampliar o grau de dificuldade dos desafios e exigências na programação. Pretende-se, ainda, ampliar a divulgação científica nos meios de comunicação de massa para incentivar o ensino da computação na Educação Básica.

Referências

Batista, W. P. et. al. (2015). Oficinas de Aprendizagem de Programação em uma Escola Pública através do Ambiente Scratch In: Anais do XXIII WEI. SBC.

Barbosa, L. S., Fernandes, T. C., and Campos, A. M. (2011). Takkou: Uma ferramenta proposta ao ensino de algoritmos. In Anais do CSBC/XIX WEI.

Correia, A. L. et. al. (2015). Uso de avaliação por pares em disciplinas introdutórias de programação. In Anais do CSBC/XXIII WEI.

International Olympiad in Informatics – IOI, acessado em 04 de março de 2016, disponível em <http://www.ioinformatics.org/history.shtml>

JotForm. Disponível em <http://www.jotform.com/>. Acesso em abril de 2016.

Massa, S. Monica. (2015). A formação didático-pedagógica do docente da área de computação: um estudo de caso em uma Universidade Brasileira. In: Anais do XXIII WEI. SBC.

MUC. Edital Maratona UNISC de Computação – MUC, 2015. Disponível em: <http://i-muc.weebly.com/>. Acesso em abril de 2016.

Rabêlo, H. M. et al. (2015). Scratch na produção de recursos interdisciplinares com disciplinas indígenas. In: Anais do XXIII WEI. SBC.

Ribeiro, Leila. (2013) Painel sobre "O Pensamento Computacional no Ensino Fundamental e Médio" realizado em outubro de 2013 no II Workshop-Escola de Informática Teórica (WEIT). Instituto de Informática / UFRGS

SCRATCH 2. Disponível em scratch.mit.edu. Acesso em dezembro de 2015.

Mão na Massa: Dinâmica Vivencial para Apoio ao Ensino de Gerenciamento de Projetos de Software

Pablo Schoeffel¹, Raul Sidnei Wazlawick²

¹Departamento de Engenharia de Software
Universidade do Estado de Santa Catarina (UDESC) – Ibirama, SC – Brasil

²Departamento de Informática e Estatística
Universidade Federal de Santa Catarina (UFSC) – Florianópolis, SC – Brasil

pabloschoeffel@yahoo.com, wazlawick@gmail.com

Abstract. *As Project Management teaching is usually theoretical, approaches to motivate students and help them to practice those concepts are considered necessary. This paper describes experimental activities to support project management teaching, which consist of a team developing a complete software project, from its beginning to its end including management practices. The paper describes the activities in detail, as well as a proof of concept including a posttest evaluation made with 35 graduation students from two Software Engineering classes. The study produced positive results in terms of motivation and learning, implying that this approach is really effective.*

Resumo. *Como o ensino de Gerenciamento de Projetos é essencialmente teórico, requer de abordagens que permitam aos alunos praticarem os conceitos e se motivarem para a disciplina. Este artigo descreve uma dinâmica vivencial para apoiar o ensino de gerenciamento de projetos, que consiste no desenvolvimento de um projeto de software completo, em equipe, do início ao fim, vivenciando as práticas de gerenciamento. O trabalho descreve o detalhamento da dinâmica, assim como o relato de aplicação e uma avaliação pós-teste realizada com 35 alunos de duas turmas de um curso de Engenharia de Software. O estudo identificou resultados positivos tanto em aspectos motivacionais quanto de aprendizagem, sugerindo a efetividade da atividade.*

1. Introdução

As estatísticas mostram que os projetos de Tecnologia da Informação continuam tendo um índice significativo de falhas e cancelamentos. Somente 39% dos projetos foram concluídos com sucesso e 18% cancelados (STANDISH GROUP, 2013).

Embora, segundo Tatnall e Reyes (2005), a maioria dos guias curriculares modelos dar um peso significativo para o ensino de Gerenciamento de Projetos (GP), pode-se supor que o ensino do tema não está ou está afetando num grau pequeno a utilização de melhores práticas nos projetos reais, aumentando o sucesso destes.

Segundo Ibrahim (2011), a disciplina de GP é teórica, em sua natureza, sendo um desafio ensinar seu conteúdo, uma vez que pode ser desinteressante quando ensinado teoricamente. Ojiako et al. (2011) diz que iniciativas de ensino e aprendizagem em GP

requerem novas e não tradicionais formas de pensar. “Um dos principais desafios no ensino de Engenharia de Software é proporcionar aos alunos experiências significativas, os quais irão achar úteis quando entram no mercado de trabalho” (BOVOTA, 2012).

Atividades práticas permitem que o aluno seja mais proativo e se envolva ao mesmo tempo em que aprende. A combinação de várias abordagens pode ser considerada uma chave para facilitar o processo de ensino e aprendizagem (Ibrahim, 2011). Segundo NTL *Institute for Applied Behavioral Science* (?), a atividade prática tem um desempenho melhor (75%) que a maioria das demais, ficando abaixo somente de ensinar outras pessoas.

Uma dessas abordagens práticas, criada por Kolb (1984), é definida como a Teoria da Aprendizagem Experiencial ou Vivencial e é o processo pelo qual o conhecimento é criado através da transformação da experiência. O conhecimento resulta da combinação de compreender e transformar a experiência” (Kolb, 1984).

Percebe-se os grandes desafios no ensino de GP e na necessidade de utilizar diferentes abordagens, com ênfase para abordagens práticas. Foram encontrados diversos trabalhos que utilizam diferentes abordagens para o ensino de GP, como: jogos (Junior e Petri, 2014), simuladores (Dantas, Barros e Werner, 2004) e atividades vivenciais (Cidral, 2003; Schoeffel, 2014). As principais contribuições desse artigo frente aos trabalhos similares é a vivência de um projeto completo de software, passando por todas as etapas, e aplicado em um curto espaço de tempo numa disciplina de pós-graduação.

Esse trabalho visa demonstrar a dinâmica vivencial *Mão na Massa*, criada para explorar o ensino do gerenciamento de projetos por meio da prática, assim como um relato de sua aplicação em duas turmas de um curso de Pós-Graduação *lato sensu* em Engenharia de Software. O artigo está organizado da seguinte forma: na Seção 2 são apresentados os principais trabalhos correlatos, na Seção 3 é apresentada a metodologia utilizada, na Seção 4 é descrita a dinâmica vivencial, na Seção 5 são mostrados os resultados da aplicação da atividade e na Seção 6 são descritas as considerações finais.

2. Trabalhos Correlatos

O trabalho descrito em Corrêa et al. (2011) relata uma abordagem vivencial como ferramenta aplicada ao ensino de Gerenciamento de Projetos, cujo escopo do projeto é uma festa de formatura. A abordagem proposta abrange somente o planejamento e é desenvolvida ao longo de um semestre, com alunos de graduação.

Cano et al. (2006) descreve a aplicação de uma abordagem para treinamento em Gerenciamento de Projetos baseada na resolução de problemas de clientes reais. O trabalho não deixa claro o escopo dos projetos, mas sim a metodologia utilizada. Também é focada para alunos de graduação e desenvolvida durante todo o semestre.

Cook e Olson (2006) relata uma dinâmica intitulada “*The Sky’s the Limit*”, cujo objetivo é ensinar GP por meio do planejamento e construção de uma estrutura com materiais fornecidos. É uma atividade rápida (cerca de 60 minutos) e os objetivos de aprendizagem são conceitos básicos de acompanhamento e planejamento de projetos.

Outro trabalho é o de Cidral (2003), que descreve uma “metodologia de aprendizagem vivencial para o desenvolvimento de competências para o gerenciamento da implementação de sistemas de informação”. Trata da forma de aplicação de uma

abordagem vivencial para TI, aplicada em 124 alunos de graduação. A atividade é baseada em simulações de problemas reais de projetos de implantação de software e os resultados das avaliações com os participantes foram positivos.

Schoeffel (2014) descreve a dinâmica vivencial PizzaMia, que tem por objetivo simular a execução de um projeto, baseado no PMBOK, na realização de uma refeição pelos alunos. O trabalho mostra os resultados na aplicação com 35 alunos de duas turmas de um curso de especialização, mostrando resultados positivos tanto na motivação quanto na aprendizagem.

Esse trabalho descreve uma dinâmica vivencial para apoio ao ensino de gerenciamento de projetos baseado na execução de um projeto de software, do início ao fim. Como diferenciais dos trabalhos correlatos encontrados, a dinâmica proposta não é uma simulação, mas uma execução completa de um projeto; permite sua execução em curto espaço de tempo; aplicada e avaliada com alunos de pós-graduação *lato sensu*).

3. Metodologia

Esse trabalho relata o desenvolvimento de uma atividade pedagógica, classificada como dinâmica vivencial, que visa ensinar os conceitos de GP executando um projeto de software completo. A atividade é utilizada como ferramenta complementar a outras abordagens, a fim de aplicar os conceitos e, dessa forma, ajudar os alunos a reforçarem o conhecimento e entendimento do assunto.

O desenvolvimento da unidade instrucional seguiu o modelo ISD (*Instructional Systems Design*) de Dick e Carey (2000), composto por 10 etapas: i) identificar objetivos instrucionais; ii) conduzir análise instrucional; iii) identificar perfil dos participantes; iv) escrever objetivos de desempenho; v) desenvolver instrumentos de avaliação; vi) desenvolver estratégia instrucional; vii) desenvolver e selecionar materiais; viii) projetar e conduzir avaliações formativas; ix) avaliar e conduzir avaliações somativas; e x) revisar instrução.

A atividade foi aplicada em duas turmas da disciplina de Gerenciamento de Projetos, com carga horária de 45 horas/aula, do curso de pós-graduação *lato sensu* em Engenharia de Software da Universidade do Estado de Santa Catarina (UDESC), nos anos de 2014 e 2015. Para avaliação da atividade, foi aplicado um questionário pós-teste para os alunos participantes, adaptado de Cidral (2003), por ser a abordagem mais próxima encontrada em termos de avaliação e Savi (2011), utilizando a avaliação da aprendizagem com base na taxonomia de Bloom.

Foram respondidos 35 questionários, sendo 19 da turma 2014 e 16 da turma 2015. A avaliação foi composta de 40 perguntas, seguindo metodologia proposta por Cidral (2003), utilizando a escala *likert* com quatro possibilidades de respostas cada uma: i) discordo inteiramente; ii) discordo parcialmente; iii) concordo parcialmente; iv) concordo inteiramente. As perguntas foram divididas nas seguintes categorias: i) propósito geral; ii) natureza dos problemas; iii) relação dos papéis com problemas; iv) controle dos eventos; v) formato; vi) sequência dos eventos; vii) materiais; viii) *feedback*; e ix) coordenador.

Para avaliação da aprendizagem foi utilizada uma auto avaliação de diversos conceitos de GP, na qual os alunos indicaram um valor de 1 a 5 para o conhecimento, entendimento e aplicação dos conceitos antes e depois da realização da atividade.

Na próxima seção serão descritos os detalhes da unidade instrucional e da dinâmica vivencial *Mão na Massa* proposta.

4. A Dinâmica Vivencial Mão na Massa

A dinâmica *Mão na Massa* é uma atividade vivencial que visa aplicar os conceitos de gerenciamento de projetos, baseado no PMBOK (PMI, 2014), em um projeto de software que os próprios alunos desenvolvem.

A atividade tem como público alvo estudantes de cursos de graduação e pós-graduação. Como objetivos de aprendizagem principais, tem-se: i) executar as atividades e utilizar os artefatos das atividades de gerenciamento de projetos; ii) vivenciar problemas reais e aplicar conceitos de GP para solucioná-los; iii) utilizar ferramentas de gestão de projetos; iv) entender a dinâmica de trabalho em equipe e a importância da gestão; v) entender o processo de interação com cliente e importância da comunicação.

Para o planejamento da unidade instrucional, foram realizadas as seguintes etapas do modelo de Dick e Carey (2000): i) identificar metas instrucionais: as metas instrucionais foram identificadas, seguindo o objetivo pedagógico do curso e andamento da turma. A meta almejada para a unidade foi que os alunos consigam desenvolver um projeto de software seguindo práticas gerenciamento de projetos baseado no PMBOK; ii) conduzir análise instrucional: identificou-se que, para atingir as metas, os alunos precisariam entender os conceitos básicos de gerenciamento de projetos e saber aplicá-los utilizando práticas e artefatos baseados no guia PMBOK; iii) escrever objetivos de desempenho: utilizando a técnica GQM (*Goal – Question – Metric*) (Basili, Caldiera e Rombach, 1994), foram identificados as seguintes objetivos a serem alcançados pelos alunos: aplicar técnicas de gerenciamento de projetos para gerenciar um projeto de software, como gerente de projetos; desenvolver artefatos para apoiar o gerenciamento de projetos de software e entender a participação com membro da equipe em um projeto de software; iv) analisar aprendizes e contextos: foi realizada por meio de uma enquete com a turma, para identificar o contexto atual da turma, identificando a experiência e conhecimento prévio dos alunos, tanto em gerenciamento de projetos quanto em desenvolvimento de software. Foi identificado que todos os alunos tinham experiência em desenvolvimento de software, a maioria no papel de programador, mas a minoria tinha conhecimento ou experiência no uso de práticas e técnicas de GP.

No projeto e desenvolvimento da dinâmica foram desenvolvidos os planos de aula, instrumentos de avaliação e os materiais didáticos utilizados. Seguindo o modelo de Dick e Carey (2000), foram realizadas as seguintes etapas: i) desenvolvimento da estratégia instrucional; ii) desenvolvimento dos instrumentos de avaliação; e iii) desenvolvimento e seleção de materiais instrucionais.

4.1.1 Desenvolvimento da Estratégia Instrucional

A unidade instrucional foi planejada contemplando a explicação introdutória aos conceitos de GP, e a aplicação da dinâmica em paralelo com aulas expositivas,

demonstrando a teoria a ser aplicada em cada fase da atividade. O Quadro 1 apresenta o cronograma previsto para a unidade.

Quadro 1: cronograma da unidade instrucional

Atividade	Carga horária
Aulas expositivas	
Apresentação, motivação e introdução a gerenciamento de projetos	3 horas
Iniciação	1 hora
Planejamento	4 horas
Execução, Monitoramento e Controle	2 horas
Encerramento e Apresentação	1 hora
Dinâmica Mão na Massa	
Desenvolvimento do projeto	29 horas
Fechamento da unidade	
Avaliação e feedback	5 horas

Conforme mostrado no Quadro 1, percebe-se que a disciplina teve um total de 16 horas de teoria e 29 de prática, enfatizando a importância que foi dada para a atividade aplicada. Durante as atividades práticas, o professor deve estar disponível e interagir com os alunos, a fim de esclarecer dúvidas, apoiar e gerenciar o andamento da atividade.

4.1.2 Desenvolvimento e Seleção de Materiais Instrucionais

Para a aula expositiva, foram desenvolvidos *slides* com o conteúdo previsto no plano de aula. Esses *slides* foram baseados nos grupos de processos, áreas, atividades e práticas do guia PMBOK.

Para a dinâmica vivencial Mão na Massa foram desenvolvidos artefatos para apoiar o gerenciamento do projeto. Esses artefatos foram desenvolvidos em documentos textos e planilhas, a fim de não restringir ao uso de alguma ferramenta específica. Os artefatos desenvolvidos foram: i) termo de abertura; ii) cronograma; iii) plano de gerenciamento de projetos; iv) solicitação de mudança; v) *status report*; vi) lições aprendidas; vii) termo de encerramento/entregas. O documento do plano de gerenciamento de projetos contempla todos os planos das demais áreas do PMBOK: escopo, tempo, custo, qualidade, recursos humanos, riscos, integração, partes interessadas, aquisição e comunicação.

4.1.3 Desenvolvimento dos Instrumentos de Avaliação

Para cada objetivo de aprendizagem, foram definidas metas e questões a serem avaliadas, utilizando a técnica GQM:

- a) executar as atividades e utilizar os artefatos das atividades de gerenciamento de projetos: - Os produtos de gerenciamento foram desenvolvidos e de forma correta? – O resultado final do produto foi satisfatório?
- b) vivenciar problemas reais e aplicar conceitos de GP para solucioná-los: - Os estudantes souberam contornar situações adversas? – Como foi a vivência como gestor de projetos?
- c) utilizar ferramentas de gestão de projetos; - Os estudantes souberam elaborar os artefatos de GP de maneira correta?
- d) entender a dinâmica de trabalho em equipe e a importância da gestão; - As equipes gerenciaram suas atividades de forma eficaz?

- e) entender o processo de interação com cliente e importância da comunicação: -
O aluno soube negociar e interagir com sua equipe e cliente?

Como resultado, chegou-se a um instrumento com critérios e pesos para avaliação da dinâmica (métricas), conforme Tabela 1.

Tabela 1: Instrumento de Avaliação da Dinâmica

Critério	Peso
Produto (aspectos relacionados ao software gerado e artefatos de GP)	20
Escopo (entrega do escopo negociado e atividades de GP desenvolvidas)	10
Qualidade (qualidade do software e uso corretos das práticas de GP)	10
Satisfação (referente à comunicação e interação com cliente e equipe)	20
Cliente (comunicação, negociação)	10
Equipe (motivação, sinergia, gestão, desempenho)	10
Execução do Projeto	60
Organização da equipe (comunicação, distribuição de papéis)	15
Metodologia e artefatos (informações do projeto usando técnicas corretamente, completude e qualidade dos artefatos entregues)	15
Participação (envolvimento individual, nível de entendimento dos conceitos de GP)	15
Atuação de gerência (comunicação, gestão das pessoas, negociação com cliente)	15

A avaliação deve ocorrer no final da dinâmica. Para medir a satisfação da equipe foi desenvolvido um instrumento de auto-avaliação, contemplando os itens: i) motivação; ii) desempenho; iii) sinergia; iv) organização/gerência. Todos os participantes devem responder cada item, seguindo uma escala *Likert* de 0 a 5, possibilitando gerar a média de cada equipe.

4.2 A Estrutura da Dinâmica Vivencial Mão na Massa

Os alunos devem ser divididos em grupos de 5 a 6 pessoas. As equipes podem definir os papéis a serem utilizados, porém o papel de Gerente de Projetos é obrigatório e deverá ser desempenhado por todos os membros da equipe.

A atividade consiste no desenvolvimento de um projeto de software, seguindo um escopo pré-determinado pelo professor, que fará o papel de cliente do projeto. O escopo dos projetos foi desenvolvido para ser desafiador e diferente para cada equipe, a fim mostrar diversidades de situações em diferentes projetos. O escopo sugerido foi de uma solução que permita aos pais de bebês e crianças registrarem o acompanhamento de seus filhos via Internet. Nesse contexto, o escopo foi dividido em 4 subprojetos, sendo que cada equipe desenvolvesse um deles:

- Calendário de vacinas: sistema que permite pré-configurar as vacinas necessárias por idade e sexo e manter um registro da vacinação do bebê, enviando alertas antes das datas de vacinação;
- Acompanhamento de peso e altura: sistema que permite o registro de peso e altura em datas específicas e comparar com as curvas de crescimento padrão, de acordo com o sexo, mostrando gráficos comparativos;
- Histórico de saúde: sistema que permita o registro de doenças, medicações, dentições, consultas médicas, alergias e outras ocorrências de saúde da criança;
- Momentos marcantes: sistema que permita registrar os momentos marcantes e fotos do desenvolvimento da criança, como: primeiros passos, primeiras palavras, primeiro banho de mar, além de datas comemorativas.

A fim de enfatizar as práticas de GP, não foram definidos aspectos técnicos, como: linguagem de programação, banco de dados, etc. A execução do projeto fica sob inteira responsabilidade da equipe. A equipe poderá optar pelo ciclo de vida mais apropriado, pela tecnologia e pela metodologia a ser seguida. Porém, sugere-se que sejam utilizadas as técnicas e artefatos de GP, com obrigatoriedade para: i) termo de abertura; ii) documento de aprovação do escopo; iii) planejamento de custo, tempo e recursos humanos, riscos; iv) monitoramento através de métricas (SPI – Índice de Desempenho de Prazo, CPI – Índice de Desempenho de Custo); v) documento de aceitação de entrega e vi) termo de encerramento.

O professor desempenha o papel de cliente, estando disponível para tirar dúvidas do escopo, fazer negociações, etc. A cada período de 4 horas troca-se o Gerente de Projetos, havendo um rodízio que deve contemplar todos os participantes das equipes. Em cada período deverá ser entregue um relatório do andamento, a ser apresentado pelo gerente de projetos.

Ao final, as equipes devem apresentar os projetos, contemplando aspectos de gerenciamento de projetos (indicadores, evolução, lições aprendidas, etc.) e também o produto gerado (funcionalidades, interface, funcionamento, etc.).

A última etapa da dinâmica é a avaliação, que é realizada pelo professor, baseada nos critérios definidos e também a discussão/*feedback* da atividade, onde deve-se promover uma discussão com os alunos, a fim de identificar situações de aprendizagem, problemas ocorridos, importância da comunicação e das atividades de gerenciamento, resultados obtidos, entre outros aspectos que se julgar necessário.

5. Resultados

Nas duas turmas aplicadas, a atividade foi realizada dividindo as turmas em quatro equipes de cinco alunos. Como resultado, além do ambiente de integração e comprometimento, todas as equipes entregaram os projetos ao final, com softwares prontos para utilização e perfazendo a totalidade ou grande parte dos requisitos iniciais.

Considerando as respostas dos 35 questionários, percebeu-se que quase 75% dos participantes concordaram fortemente com as afirmações da avaliação, sendo que apenas cerca de 5% das respostas foram negativas, conforme Tabela 2, que exhibe o resultado por cada uma das nove categorias avaliadas.

A categoria *Desenvolvimento Pessoal*, refere-se à percepção do aluno com relação ao desenvolvimento de habilidades pessoais para análise e solução de problemas. A categoria *Natureza dos Problemas* avalia a percepção dos alunos sobre os desafios e problemas abordados na atividade. A categoria *Relação dos Papeis com Problemas* avalia a percepção dos alunos sobre os papéis que desempenharam no desenvolvimento do projeto de software. Já a categoria *Controle dos Eventos* avalia a percepção dos alunos sobre a autonomia de tomada de decisão durante a atividade e a consequência de suas escolhas. A categoria *Formato* avalia a percepção dos alunos sobre a estrutura da atividade e motivação. A categoria *Sequência dos Eventos* avalia se o fluxo da atividade foi similar a uma situação real. A categoria *Materiais* avalia a percepção dos alunos sobre as informações e materiais da atividade. A categoria *Feedback* avalia a percepção dos alunos sobre a reflexão e discussão dos resultados da atividade e, por fim, a

categoria *Coordenador* avalia a percepção dos alunos sobre o papel desempenhado pelo professor na atividade, no papel de coordenador da dinâmica.

Tabela 2: Resultados da avaliação por categoria

<i>Categoria</i>	<i>Discordo</i>	<i>Discordo</i>	<i>Concordo</i>	<i>Concordo</i>
	<i>Inteiramente</i>	<i>Parcialmente</i>	<i>Parcialmente</i>	<i>Inteiramente</i>
Desenvolvimento Pessoal	0,0%	0,0%	12,0%	88,0%
Natureza dos Problemas	1,3%	1,4%	24,4%	64,6%
Relação dos papéis com problemas	0,0%	0,0%	20,0%	80,0%
Controle dos Eventos	0,0%	0,2%	17,9%	81,4%
Formato	0,0%	0,3%	20,0%	78,9%
Sequência dos Eventos	0,0%	0,0%	20,0%	80,0%
Materiais	0,0%	0,0%	28,6%	71,4%
Feedback	0,0%	1,1%	30,0%	65,7%
Coordenador	1,7%	1,9%	26,4%	58,9%
TOTAL	2,0%	2,7%	21,1%	74,2%

Com relação à aprendizagem, foi solicitado aos alunos uma auto avaliação com relação ao seu avanço no conhecimento pré e pós atividade (de 1 a 5), considerando três dos níveis de aprendizagem da Taxonomia de Bloom (conhecer, entender e aplicar). Os resultados são mostrados na Tabela 3, mostrando um significativo nível de concordância com a contribuição da atividade para a evolução da aprendizagem.

Tabela 3: Evolução da Aprendizagem

<i>Conceitos</i>	<i>Lembrar o que é</i>			<i>Compreender como funciona</i>			<i>Aplicar na prática</i>		
	<i>Antes</i>	<i>Depois</i>	<i>Dif.</i>	<i>Antes</i>	<i>Depois</i>	<i>Dif.</i>	<i>Antes</i>	<i>Depoi</i>	<i>Dif.</i>
	Iniciação de projeto	2,06	4,26	2,20	2,00	4,20	2,20	1,60	3,89
Planejamento de Escopo	2,37	4,40	2,03	2,14	4,17	2,03	1,86	4,03	2,17
Planejamento de Tempo	2,51	4,26	1,74	2,20	4,26	2,06	1,97	4,11	2,14
Planejamento de Custo	2,06	4,03	1,97	1,91	3,77	1,86	1,49	3,63	2,14
Planejamento de Qualidade	1,94	3,94	2,00	1,91	3,86	1,94	1,86	3,69	1,83
Planejamento de Riscos	2,03	4,11	2,09	1,89	4,23	2,34	1,51	3,89	2,37
Planejamento de RH	1,86	3,69	1,83	1,71	3,60	1,89	1,54	3,40	1,86
Planejamento de Partes	1,97	3,83	1,86	1,77	3,77	2,00	1,60	3,66	2,06
Planejamento de Aquisição	1,77	3,71	1,94	1,71	3,69	1,97	1,54	3,54	2,00
Planejamento de Integração	1,66	3,51	1,86	1,60	3,49	1,89	1,37	3,29	1,91
Planejamento de Comunicação	1,97	3,83	1,86	1,83	3,83	2,00	1,60	3,71	2,11
Gerenciamento da execução do	2,14	4,26	2,11	2,00	4,20	2,20	1,74	3,94	2,20
Acompanhamento e	2,14	4,29	2,14	2,00	4,29	2,29	1,83	4,06	2,23
Encerramento do projeto	1,91	4,11	2,20	1,77	4,17	2,40	1,63	4,03	2,40

Calculando-se a média total de cada nível de aprendizagem, percebe-se uma evolução média significativa tanto no conhecimento (1,99), entendimento (2,08) e aplicação (2,12). Os resultados das avaliações sugerem que a atividade teve uma boa aceitação pelos alunos, sendo que quase a totalidade das respostas foi positiva (95,3% de concordância total ou parcial com os quesitos avaliados). Além disso, segundo a auto avaliação dos alunos, a dinâmica vivencial *Mão na Massa* contribuiu significativamente para a aprendizagem, nos três níveis de aprendizagem (conhecimento, entendimento e aplicação) e em todos os conceitos avaliados, destacando-se: iniciação do projeto (evolução = 2,29), planejamento de riscos (evolução = 2,37), gerenciamento da

execução do projeto (evolução = 2,20), acompanhamento e monitoramento (evolução = 2,29) e encerramento do projeto (evolução = 2,40).

Além desses resultados, foi solicitado aos alunos que descrevessem pontos fortes, pontos fracos e sugestões de melhoria da atividade. Dos pontos positivos, destacaram-se: i) a vinculação da prática com teoria; ii) utilização de um projeto completo similar a um contexto real e iii) trabalho em equipe. Os pontos negativos mais citados foram: i) pouco tempo para realizar as atividades; ii) defasagem do conteúdo teórico, devido a priorização da atividade na disciplina; iii) cada membro não vivencia todas as situações como gerente de projetos, devido ao rodízio. As sugestões de melhorias mais frequentes foram: i) definição melhor das entregas e artefatos; ii) inserir mais restrições; iii) mais ênfase no controle e acompanhamento do projeto.

6. Considerações Finais

O artigo descreveu uma dinâmica vivencial para apoiar no ensino da disciplina de GP. A atividade consistiu no desenvolvimento de quatro projetos de software, sendo uma para cada equipe, no decorrer de uma disciplina em duas turmas de especialização. Todos os 8 projetos foram entregues ao final da disciplina, de maneira satisfatória, sendo que a mesma teve 45 horas/aula concentradas em apenas uma semana de aula.

Numa avaliação, após a aplicação da atividade, os alunos responderam um questionário que demonstrou resultados positivos com relação à percepção dos participantes em relação à atividade (95% de respostas positivas). Da mesma forma, a evolução da aprendizagem, na percepção dos alunos, foi significativa. Numa escala de 1 a 5, a evolução média descrita foi de 2,06 (1,86 para 3,92).

Além dos resultados expressos em números e comentários dos alunos, o fato das equipes de alunos conseguirem entregar um projeto completo funcionando ao final da disciplina foi uma motivação extra. Ficou claro a motivação e engajamento das equipes, que trabalharam diversas horas extra sala de aula nos seus projetos. Após a entrega e apresentação, durante a reflexão e *feedback* (última etapa da atividade), muitos alunos relataram estarem impressionados com o desempenho da equipe, sendo que na indústria, um projeto de complexidade similar, talvez seria finalizado em semanas ou até meses.

Apesar dos resultados positivos e promissores, mostrando que atividades vivenciais podem ser uma alternativa viável para melhorar o processo de ensino-aprendizagem de GP, existem algumas ameaças aos resultados desse trabalho. O fato das avaliações serem auto avaliações podem ter direcionado os alunos para responder de forma displicente ou imparcial. Outra questão é que o baixo número de alunos atingidos e num único contexto pode ter direcionado os resultados, devendo ser realizadas mais aplicações em diferentes contextos para poder generalizar os resultados.

Como sugestão de trabalhos futuros, seria interessante a aplicação da dinâmica Mão na Massa em turmas de graduação, outros cursos de pós-graduação ou mesmo treinamentos da indústria. Além disso, uma sugestão de evolução da dinâmica seria a criação de novos contextos e escopo, englobando as sugestões de melhoria dos alunos.

7. Referências

Basili, V., Caldiera G. e Rombach, D. (1994). The Goal Question Metric Paradigm.

- Cano, J. L. Lidon, I. Rebollar, R. Roman, P. Saenz, M. J. (2006) "Student Groups Solving Real-life Projects: a case study of experiential learning". In *Int. J. Engng Ed.* v. 22, n. 6, p. 1252-1260. Tempus Publications. 2006.
- Cidral, A. (2003) "Metodologia de Aprendizagem Vivencial para o Desenvolvimento de Competências para o Gerenciamento de Projetos na Implementação de Sistemas de Informação". Tese de Doutorado do Programa de Pós-Graduação em Engenharia da Produção da Universidade Federal de Santa Catarina, 2003.
- Cook, L. Olson, J. (2006) "The Sky's the Limit: An Activity for Teaching Project Management". In *Journal of Management Education*, v. 30 n. 3, Jun 2006, p. 404-420.
- Corrêa, F.R. Barçante, L.C. Alves, C.S.C. Maciel, M.S.D. Souza, F. S. (2011) "A Abordagem Vivencial como Ferramenta Aplicada ao Ensino de Gerenciamento de Projetos". In: XXXIX Congresso Brasileiro de Educação em Engenharia (COBENGE2011). Blumenau/SC.
- Ibrahim, I. (2011) "Teaching Project Management for IT Students: Methods and Approach". In *IPEDR vol.13*, IACSIT Press, Singapore.
- Kolb, D. (1984) "Experiential learning: experience as the source of learning and development". Prentice-Hall, New Jersey.
- Dantas, A. Barros, M. e Werner, C. (2004) "A Simulation-Based Game for Project Management Experiential Learning". In *Proceedings of the Sixteenth International Conference on Software Engineering and Knowledge Engineering (SEKE'04)*.
- NTL Institute for Applied Behavioral Science, Learning Pyramid. 300 N. Lee Street, Suite 300, Alexandria, VA 22314. 1-800-777-5227. Disponível em: <<http://homepages.gold.ac.uk/polovina/learnpyramid/about.htm>>.
- Ojiako, U. Ashleigh, M. Chipulu, M. Maguire, S. (2011) "Learning and teaching challenges in project management". In *International Journal of Project Management*, v. 29, n. 3, Abril 2011, p. 268–278.
- Petri, G., Marcon Júnior, R.P. (2014). "Um Jogo Educacional Para o Ensino de Metodologias Ágeis". FEES - Fórum de Educação em Engenharia de Software, evento integrante do XXIII Simpósio Brasileiro de Engenharia de Software (SBES),
- PMI. Project Management Institute (2014). *Um Guia do Conhecimento em Gerenciamento de Projetos (PMBOK)*. 5ª Ed. Project Management Institute, 2014.
- Savi, R. (2011) "Avaliação de Jogos Voltados para a Disseminação do Conhecimento". Tese (doutorado) - Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. Florianópolis, SC, 2011.
- Schoeffel, P. (2014). *PizzaMia: Dinâmica Vivencial para Apoio ao Ensino de Gerenciamento de Projetos Baseado no PMBOK*. In *XXII Workshop de Educação em Computação (WEI 2014)*.
- The Standish Group (2013) "Chaos Manifesto 2013". The Standish Group International. Disponível em: <<http://versionone.com/assets/img/files/CHAOSManifesto2013.pdf>>. Acessado em: 20 jan 2016.

MEGA GP: Aplicando a Gamificação no Ensino de Gerência de Projetos

Sara Midori Mendes Tomisaki, Adler Diniz de Souza, Rodrigo Duarte Seabra

Instituto de Matemática e Computação – Universidade Federal de Itajubá (UNIFEI)
Caixa Postal 50 – 37.500-903 – Itajubá – MG – Brasil
saramidori.hime@gmail.com, adlerunifei@gmail.com, rodrigo@unifei.edu.br

***Abstract.** Teaching the concepts of software engineering applied to project management to computer science course students and related fields is a challenging task due to motivational factors and resources. From this reality, different methodologies that stimulate the study of the theme are investigated. This research presents the using experience of gamification in project management lecturing on the impact of the strategy explored on the content teaching. The main results observed show that the gamification resources explored in the classroom achieved good acceptance in accordance with the opinion of the students participating in the experimental research carried out.*

***Resumo.** Ensinar os conceitos de engenharia de software aplicados à gerência de projetos para estudantes de cursos de computação e áreas afins se apresenta como uma tarefa desafiadora devido a fatores motivacionais e de recursos. A partir desta realidade, buscam-se diferentes metodologias que incentivem o estudo do tema. Esta pesquisa apresenta a experiência do uso da gamificação no ensino de gerência de projetos, dissertando sobre o impacto da estratégia explorada no ensino do conteúdo. Os principais resultados observados mostraram que os recursos de gamificação explorados em sala de aula alcançaram boa aceitação, de acordo com a opinião dos estudantes participantes da pesquisa experimental realizada.*

1. Introdução

Atualmente, as organizações têm investido expressivamente na busca de profissionais que gerenciem de modo eficiente projetos empresariais, visando à vantagem competitiva e a excelência no mercado de trabalho. Nesse cenário, espera-se que o profissional apresente certas qualificações que vão desde sua capacitação até a experiência julgada como satisfatória na área. Por conta disso, esses fatores se apresentam como obstáculos que dificultam o ingresso dos recém-graduados no mercado profissional.

Com base nas dificuldades impostas pela área profissional e no que tange ao objetivo primordial esperado do universo acadêmico – a formação do profissional do futuro por meio de um ensino de qualidade – aponta-se, então, a necessidade de oferecer aos estudantes aulas mais dinâmicas e tangíveis sobre gerenciamento de projetos, promovendo melhor entendimento do assunto e dos conceitos presentes no PMBOK – *Project Management Body of Knowledge* (PMI 2013). Este guia foi elaborado com o objetivo de organizar e catalogar o conjunto de conhecimentos relacionados à área de gerenciamento de projetos.

Estudos relatados na literatura apontam variadas estratégias exploradas no ensino de gerenciamento de projetos e engenharia de software com base, por exemplo, em jogos educacionais. Pesquisas recentes mostram que por meio de mecanismos de jogos que

estimulem o aprendizado, é possível motivar o aluno de maneira lúdica, estratégica e efetiva frente ao conteúdo exposto (Bonetti e Wangenheim 2013; Dubois e Tamburrelli 2013), contribuindo para a redução da monotonia das aulas convencionais.

Ao se falar em jogos, remete-se ao conceito contemporâneo denominado gamificação, que, se bem explorado, caminha paralelamente à motivação. A estratégia apresenta potenciais benefícios não apenas no âmbito educacional, mas também no ambiente corporativo real, objetivando instigar a participação do indivíduo submetido ao processo. Com isso, na seara profissional, aumenta-se a probabilidade de resultados positivos para a equipe e, conseqüentemente, para a empresa. No contexto acadêmico, o uso da gamificação pode proporcionar vantagens significativas no processo de formação dos discentes. Segundo Mattar (2010), o envolvimento proporcionado pelos jogos supera os benefícios oferecidos por qualquer outra mídia. O autor relata que os principais fatores envolvidos durante um jogo são: diversão, regras e estrutura, objetivos, interatividade e resultados.

O uso de recursos diferenciados na sala de aula potencializa o processo de ensino e aprendizagem, promovendo o desenvolvimento do relacionamento entre o estudante e o professor. As estratégias utilizadas visam atender as necessidades de ambas as partes: dos estudantes, pela motivação e o desafio que os incentivem à prática do estudo; do docente, como um recurso para auxiliar o ensino de maneira mais eficaz. Considerando somente a área da educação, o mapeamento sistemático recente realizado por Borges et al. (2013) identificou 26 trabalhos relacionados ao uso da gamificação nesta área. No cenário nacional, algumas pesquisas recentes podem ser destacadas, tais como as de Figueiredo et al. (2015), Campos, Gardiman e Madeira (2015) e Jucá et al. (2014).

Devido a sua importância, torna-se fundamental o ensino de gerenciamento de projetos nos cursos de computação e afins. A partir do cenário exposto, esta pesquisa tem como objetivo geral apresentar o desenvolvimento e a aplicação de um novo método de ensino dos conceitos de gerenciamento de projetos no ensino superior, por meio da gamificação, com o intuito de aumentar a motivação e o desempenho dos estudantes. Adicionalmente, o trabalho ainda se propõe a investigar, sob a ótica discente, os potenciais benefícios apontados em decorrência do uso do método proposto no apoio ao ensino do tema. Como principais contribuições do estudo destacam-se: (i) contribuir com a disseminação da ideia do uso de jogos educacionais, para que futuramente haja aumento de sua utilização nos variados contextos de ensino; (ii) avaliar qualitativamente a opinião dos estudantes envolvidos no estudo em questão, verificando se o jogo desenvolvido nesta oportunidade atendeu as expectativas esperadas, bem como seu grau de aceitação e contribuição no processo de ensino.

Nesse sentido, a seção 2 discorre sobre o conceito de gamificação, correlacionando este recurso à inerente variação de abordagens sobre gerenciamento de projetos. A seção 3 apresenta detalhes ligados à concepção e dinâmica do jogo MEGA GP. A seção 4 expõe a análise dos dados coletados e discute os principais resultados da pesquisa experimental realizada. Finalmente, a seção 5 apresenta as conclusões do trabalho.

2. Gamificação

O termo gamificação, do original em inglês *gamification*, diz respeito ao uso de jogos para diferentes fins de entretenimento (Vianna et al. 2013). Em síntese, se refere ao uso

de mecânicas e dinâmicas de jogos com o objetivo de resolver problemas, despertando o engajamento do público envolvido e, potencialmente, melhorando o aprendizado. Os jogos servem de plataforma para o modelo de gestão que apoia o aspecto social da equipe, diferentemente do modelo hierárquico burocrático com funções estritamente definidas e limitadas. Em ambientes empresariais, os jogos têm a função de organizar pessoas para alcançar um objetivo, estimulando a superação de limites, a cooperação e promovendo a criatividade no processo de inovação.

Cunhado, em 2002, por Nick Pelling, o termo gamificação só ganhou fama após uma apresentação realizada oito anos depois por Jane McGonigal, autora da obra considerada a maior referência da área – “A realidade em jogo: por que os games nos tornam melhores e como eles podem mudar o mundo”. Na ocasião, a autora apresentou dados significativos sobre o número de horas gastas pelos jogadores de um famoso jogo *online*, tempo este que poderia ser investido na resolução de problemas reais. Com esses dados, a autora demonstrou como as pessoas se sentem atraídas por jogos e a imersão que eles propiciam. Além de engajar as pessoas, a gamificação possibilita que tarefas tediosas se tornem mais agradáveis. A intenção envolvida não consiste em apenas criar ou participar de jogos, mas sim, utilizar de seus mecanismos para provocar os objetivos que necessitam ser alcançados. A estratégia, portanto, busca desenvolver novas posturas dentro de uma equipe, analisando o comportamento humano, e não apenas aplicando processos mecânicos pré-estabelecidos.

2.1 Gamificação em Gerência de Projetos

Até meados do século XX, a função de gerente de projetos não requeria profissionais especializados e certificados, todavia, o exercício desta prática transformou essa responsabilidade em uma profissão distinta e de grande importância no mercado profissional. Atualmente, problemas de dimensões variadas são resolvidos por meio de projetos. A importância do gerenciamento de projetos pode ser comprovada pelo fato de as empresas investirem cada vez mais no treinamento e na capacitação de seus funcionários para melhorar a eficiência de seus projetos, mensurada, em sua maioria, por algumas variáveis, tais como prazo, orçamento e escopo.

O Guia PMBOK (PMI 2013) oferece uma série de diretrizes para o gerenciamento adequado de projetos, e o *Capability Maturity Model Integration for Development*, ou CMMI-DEV (SEI 2010), fornece um conjunto de práticas específicas e genéricas que aumentam a probabilidade de gerenciar projetos de software com sucesso. Normalmente, esses guias podem ser usados em conjunto para alcançar o sucesso em um projeto de software. Contudo, de modo geral, os estudantes sentem dificuldades em assimilar os conceitos envolvidos da engenharia de software por serem termos distantes de suas realidades.

Uma alternativa que tem se mostrado eficiente na superação deste problema é o uso de jogos no ensino. Variadas pesquisas recentes encontram-se sistematicamente relatadas na literatura com o objetivo de apoiar essa estratégia e, por conseguinte, contribuir com a motivação dos estudantes. De modo geral, essas pesquisas abrangem projetos de jogos educacionais que simulam processos de engenharia de software e gerência de projetos.

O estudo de Prikladnicki *et al.* (2007) apresenta uma ferramenta denominada *Planager*, que oferece apoio ao treinamento de gerência de projetos de software. O principal público alvo da pesquisa são gerentes de projeto iniciantes e experientes a

compreenderem os conceitos envolvidos em suas atividades por meio de cenários e de práticas do PMBOK. Uma proposta semelhante pode ser observada em Mira *et al.* (2012), por meio do jogo *ProMEG*, cujo foco está no gerenciamento de recursos humanos. A pesquisa desenvolvida por Kohwalter *et al.* (2011) propõe o uso de um jogo de simulação em que o jogador é proprietário de uma empresa de desenvolvimento de software que conta com o auxílio de uma equipe, administrada pelo jogador, para desenvolver produtos requisitados pelos clientes. Em Bonetti e Wangenheim (2013), verifica-se a apresentação de um repositório digital – *Instructional Games Repository* – com o objetivo de compartilhar informações sobre jogos educacionais para o ensino de gerenciamento de projetos.

A estratégia de apenas memorizar os conceitos presentes no PMBOK e CMMI-DEV não se apresenta como uma alternativa eficiente. Pelo contrário, é necessário que o estudante, ainda na universidade, pratique ao máximo os conceitos inerentes à gerência de projetos, visando lapidar sua sensibilidade acerca das práticas gerenciais fundamentais para alcançar o objetivo final de todo e qualquer projeto. Nessa seara, a aplicação da gamificação contribui para que os discentes compreendam, de fato, os principais termos e conceitos relacionados ao tema, mediante as possíveis competições estabelecidas entre os colegas na sala de aula. Acrescenta-se ainda ao exposto que a gamificação vem sendo aplicada não apenas em sala de aula. Empresas de diversos setores também têm usufruído dos benefícios da gamificação em seus processos a fim de engajar as equipes na busca de soluções para os problemas. Como exemplo, a equipe da MVJ Tecnologia & Inovação (Vianna *et al.* 2013) traz em seu site artigos sobre a implementação da gamificação em empresas com estudos ligados a alguns temas, tais como gestão de mudanças, planejamento estratégico, metodologia ágil, entre outros.

A partir das pesquisas arroladas, é fato que o uso de jogos na educação tem despertado o interesse de pesquisadores nos últimos anos. Isso se deve, sobretudo, à ascensão da geração Y ao mercado (Mattar 2010), pelo fato de os indivíduos pertencentes a ela estarem habituados ao uso da tecnologia. A gamificação motiva o estudante pelo simples fato de que o ser humano sente-se atraído por jogos (Jayasinghe e Dharmaratne 2013). De modo geral, como as pessoas gostam de competir e vencer, essa expectativa influencia no preparo para a realização perfeita de uma tarefa (Kapp 2012).

O presente trabalho representa mais uma alternativa frente a esse desafio, abrangendo uma nova oportunidade para o estudo de gerenciamento de projetos.

3. O MEGA GP

O MEGA GP representa uma proposta de jogo baseada nos moldes do *game show* MEGA Senha, um programa de televisão que consiste em um jogo de palavras que devem ser adivinhadas mediante o uso de dicas. Diferentemente dos jogos citados na seção 2, o MEGA GP não explora recursos computacionais em sua dinâmica, pelo contrário, utiliza poucos artefatos para viabilizar sua aplicação em sala de aula, sem qualquer investimento financeiro em termos de tecnologia. O jogo é conduzido pelo professor, que controla seu andamento e exerce a figura de moderador das regras. De modo similar ao *game show*, os jogadores, neste caso, os estudantes, devem adivinhar palavras-chave relacionadas a uma área de conhecimento em gerenciamento de projetos com base no PMBOK ou às técnicas e práticas específicas do CMMI-DEV.

O MEGA GP consiste de duas etapas, sendo a primeira eliminatória e, a segunda,

caracterizada como a etapa final. Inicialmente, os jogadores são previamente informados sobre a área de conhecimento em gerenciamento de projetos em questão para que se preparem para o jogo. A cada rodada, os participantes têm 90 segundos para adivinhar três palavras-chave relacionadas à área de conhecimento envolvida naquela rodada e, por conseguinte, serem classificados para a etapa final. A Tabela 1 apresenta exemplos de informações disponíveis nas cartas do jogo na forma de palavras-chave e dicas.

Em sua dinâmica, o jogo é composto por 35 cartas divididas em quatro áreas de conhecimento, além de cartas relativas a uma introdução à gerência de projetos. As cartas são distribuídas nas seguintes áreas: (i) Introdução à Gerência de Projetos, oito cartas; (ii) Gerenciamento de Escopo, oito cartas; (iii) Gerenciamento de Tempo, oito cartas; (iv) Gerenciamento de Custo, cinco cartas; e (v) Gerenciamento de Comunicação e Qualidade, seis cartas.

Tabela 1 – Exemplos de informações presentes nas cartas do MEGA GP¹.

Carta	Palavras-chave	Dicas	
1	Caminho Crítico	Sem folga	Cronograma
	Estimativa de Três Pontos	Estimativa otimista	Pessimista e mais provável
	<i>Poker Planning</i>	Estimativa	<i>Extreme Programming</i>
2	Prototipação	Modelo funcional	Elaboração progressiva
	<i>Brainstorming</i>	Ferramentas e técnicas	Definição do escopo
	Matriz de Rastreabilidade Bidirecional	Relacionamento + requisitos	Matriz
3	Pacote de Trabalho	EAP	Menor nível
	Estimativa análoga	Estimativa	Menos precisa
	Valor agregado	Medida básica	Trabalho realizado
...

É importante destacar que a variação na quantidade de cartas em cada área se deve ao total de informações técnicas presentes em cada área de processo. As dicas apresentadas nas cartas variam de uma a três, dependendo da dificuldade das palavras-chave disponíveis na carta. Durante a experimentação do MEGA GP, os estudantes participantes do jogo não se limitavam às dicas apresentadas nas cartas e, frequentemente, elaboravam novas dicas. A Figura 1 apresenta algumas cartas disponíveis no jogo, bem como exemplos de palavras-chave e dicas.

As duplas que avançam à etapa final são submetidas a quatro *rounds* de 90 segundos. No primeiro *round*, o jogador conta com uma ficha composta por seis palavras-chave, e deve adivinhar três delas. A cada *round*, o número de palavras disponíveis diminui até alcançar o último *round*, no qual o participante deve acertar as palavras-chave na íntegra. Ao tentar descobrir a palavra-chave, o estudante exercita seu raciocínio enquanto assimila em qual contexto se aplicam as dicas. Esse processo de associação permite melhor absorção e entendimento do conteúdo em questão. Por exemplo, ao oferecer a dica “ferramentas e técnicas”, o jogador recorrerá à própria

¹ A relação completa das cartas disponíveis no jogo pode ser obtida em: <http://megagp.xpg.uol.com.br/index.htm>

memória quais ferramentas e técnicas são usadas no gerenciamento de projetos de software. A partir desta estratégia, o jogo propicia uma exposição abrangente do conteúdo, bem como o refinamento da resposta a cada nova dica disponibilizada.

Um dos mecanismos que caracteriza a gamificação é o fator motivacional oferecido ao participante por meio de um sistema de recompensa. Cada *round* da etapa final do MEGA GP possui uma premiação revertida em pontos extras na nota dos estudantes regularmente matriculados na disciplina de Gerência de Projetos de Software.



Figura 1 – Exemplos de cartas do MEGA GP.

Finalmente, após o uso do MEGA GP, foi aplicado um simulado visando mensurar o desempenho dos estudantes, sendo composto por: (i) 20 questões de múltipla escolha, contendo quatro alternativas cada; (ii) as questões foram relacionadas à área de conhecimento estudada; e (iii) exemplos de questões das provas de certificação PMP do PMI. Os jogadores contaram com 20 minutos disponíveis para o simulado e as questões exigiram conhecimento teórico e prático sobre gerência de projetos.

4. Discussão dos Resultados

A disciplina de Gerência de Projetos de Software, do curso de Sistemas de Informação da Universidade Federal de Itajubá – UNIFEI, foi utilizada como estudo de caso para a validação do jogo MEGA GP. Essa disciplina tem como objetivo estudar as principais áreas de conhecimento do PMBOK, bem como as principais ferramentas e técnicas aplicadas à Gerência de Projetos de Software. A disciplina de Gerência de Projetos envolveu dois componentes de avaliação, a saber: (i) um conjunto de mini simulados relacionados às áreas de conhecimento do PMBOK; e (ii) duas provas dissertativas.

Para avaliar a eficiência do jogo MEGA GP no aprendizado dos estudantes, comparou-se as notas obtidas pelos alunos nos mini simulados a partir de duas turmas, referentes aos anos de 2013 e 2014. A turma de 2013 não utilizou o jogo MEGA GP como objeto de aprendizado. A turma de 2014 utilizou o jogo e, em seguida, realizou o mini simulado relativo a uma área de conhecimento. Com o objetivo de minimizar a influência de qualquer tipo de viés na condução do estudo, tanto os mini simulados

quanto o material didático utilizado para conduzir a disciplina foram iguais em ambas as turmas. Ademais, esses materiais não sofreram modificações no decorrer dos dois anos do estudo. Em ambas as turmas, não havia alunos repetentes na disciplina e nenhum deles possuía experiência prévia em gerência de projetos de software.

Para possibilitar a análise da contribuição do jogo no aprendizado, os estudantes foram agrupados em cinco categorias, com base nas notas obtidas: (i) **Péssimo**: alunos com notas que variaram entre 0-20; (ii) **Ruim**: alunos com notas que variaram entre 21-40; (iii) **Regular**: alunos com notas que variaram entre 41-60; (iv) **Bom**: alunos com notas que variaram entre 61-80; (v) **Excelente**: alunos com notas que variaram entre 81-100.

Em razão da quantidade variável de alunos presentes nos mini simulados, optou-se por calcular o percentual de estudantes classificados nas categorias supramencionadas em cada área de conhecimento em que eles foram avaliados. A Tabela 2 mostra a classificação dos discentes em termos percentuais, em função de suas notas, que não utilizaram o jogo MEGA GP como objeto de aprendizagem nos mini simulados realizados (turma 2013), e a classificação dos alunos que utilizaram o jogo MEGA GP e, posteriormente, fizeram os mini simulados (turma 2014).

Tabela 2 – Classificação dos estudantes nos mini simulados.

	Introdução à GP		Gerência de Escopo		Gerência de Tempo		Gerência de Custo	
	2013	2014	2013	2014	2013	2014	2013	2014
Péssimo	0%	0%	0%	0%	0%	0%	0%	0%
Ruim	3,70%	3,33%	3,70%	5%	31,25%	12,50%	10,71%	23,81%
Regular	51,85%	23,33%	77,78%	19,05%	40,63%	41,67%	57,14%	14,29%
Bom	44,44%	56,67%	18,52%	66,67%	28,13%	41,67%	25%	47,62%
Excelente	0%	16,67%	0%	14,29%	0%	4,17%	7,14%	14,29%

No caso do mini simulado de Introdução à Gerência de Projetos, é perceptível que as notas dos alunos que utilizaram o MEGA GP antes do simulado foram maiores que os desempenhos dos alunos que não o utilizaram. Os estudantes da turma de 2013 tiveram grande parte das notas classificadas como desempenho “Regular” e “Bom”; nenhum aluno da turma foi classificado com desempenho “Excelente”. Contrariamente a esse cenário, a turma de 2014, que utilizou o MEGA GP, obteve maior distribuição de alunos classificados com desempenhos “Bom” e “Excelente”. Esse mesmo comportamento foi observado nos mini simulados relativos aos conteúdos de Gerência de Escopo e Gerência de Tempo.

Nessas três áreas de conhecimento, isto é, Introdução à Gerência de Projetos, Gerenciamento de Escopo e Gerenciamento de Tempo, existe uma ampla quantidade de conceitos que precisam ser memorizados para a realização de um mini simulado. Nesse contexto, a turma que utilizou o jogo MEGA GP teve claramente notas maiores em todos os mini simulados aplicados, com desempenhos concentrados nas categorias “Bom” e “Excelente”, e menos alunos nas classificações “Ruim” e “Regular”. A única exceção detectada ocorreu no mini simulado de Gerência de Custos, no qual houve um grande percentual de estudantes que apesar de terem utilizado o MEGA GP, ainda assim obtiveram notas classificadas como “Ruim”. Uma possível explicação para esse resultado pode ser atribuída ao fato de que a área de conhecimento de Gerência de Custos é a única, dentre as estudadas, que contém um conjunto menor de conceitos que pode ser

memorizado, sendo que grande parte das questões presentes em seu mini simulado envolve a aplicação de equações de uma técnica conhecida como *Earned Value Management* – EVM (PMI 2011).

Durante a aplicação do MEGA GP nessa área de conhecimento, optou-se por abordar o significado e a interpretação dos conceitos envolvidos nas equações. Posteriormente, foram realizados exercícios práticos com a aplicação das equações estudadas durante as aulas. A utilização de exercícios práticos de EVM é uma prática adotada pelo docente da disciplina há vários anos. Essa prática visa aumentar a média das turmas no mini simulado de Gerenciamento de Custo que, historicamente, sempre foi a mais baixa. O resultado observado sugere que a aplicação do MEGA GP foi adequada para um conjunto de alunos que obteve notas consideradas como desempenhos “*Bom*” e “*Excelente*”. O percentual geral de alunos que alcançou as notas máximas foi maior que o da turma de 2013, que não utilizou o MEGA GP. Entretanto, o percentual de alunos com notas consideradas como desempenho “*Ruim*” também foi maior na turma que utilizou o MEGA GP na área de Gerência de Custos. A explicação para esse comportamento pode estar no fato de um grupo de alunos não ter realizado os exercícios práticos dessa área de conhecimento, o que impediu que eles absorvessem o conhecimento prático de EVM necessário para essa área de conhecimento.

Após quatro rodadas do jogo em sala de aula, ainda foi aplicado um questionário avaliativo para fins de coleta da opinião dos estudantes sobre o uso experimentado do jogo MEGA GP. O questionário foi desenvolvido com base na norma ISO/IEC 25010 e preocupou-se em avaliar o impacto da gamificação no ensino de Gerência de Projetos de software com relação aos seguintes atributos: *eficácia*, *satisfação* (utilidade e prazer) e *cobertura do contexto* (completude do contexto e flexibilidade). O questionário, composto por 13 questões, foi aplicado aos estudantes participantes do estudo, perfazendo um total de 23 entrevistados. As possíveis respostas em cada questão variaram no intervalo de 1 (*discordo totalmente*) a 5 (*concordo totalmente*), sendo caracterizadas em função de seus valores mínimos, máximos, médias e desvios padrão (Tabela 3). De modo geral, a análise dos questionários permitiu inferir uma avaliação satisfatória sobre o uso experimentado do MEGA GP em sala de aula. A Figura 2 permite a visualização das respostas com base em cada atributo avaliado.

Tabela 3 – Distribuição das respostas para cada questão.

Questão	Mínimo	Máximo	Média	Desvio Padrão	Questão	Mínimo	Máximo	Média	Desvio Padrão
1	2	5	4,1	0,85	8	2	5	4,2	1
2	3	5	4,5	0,63	9	1	5	4,2	0,97
3	2	5	3,8	0,95	10	2	5	3,8	0,86
4	2	5	3,7	1	11	2	5	3,9	0,88
5	1	5	3,1	1,08	12	2	5	4,6	0,68
6	3	5	4,1	0,77	13	1	5	2,2	1,16
7	2	5	3,4	0,84					

Em relação ao atributo *eficácia*, cinco questões investigaram se o jogo permitiu que seus jogadores atingissem seus objetivos de acordo com suas necessidades: (Q1) *As regras do MEGA GP são claras e objetivas?* (Q2) *A rodada do MEGA GP é rápida e simples de jogar?* (Q3) *A relação entre as senhas e as palavras-chave garantem a fixação do conteúdo?* (Q4) *O MEGA GP permite melhor entendimento dos conceitos presentes no PMBOK?* (Q5) *O MEGA GP facilitaria o gerenciamento de um projeto de software?* Aproximadamente 67% dos alunos responderam satisfatoriamente as questões

sobre *eficácia* (Figura 2).

Quatro questões investigaram se os jogadores se sentiram satisfeitos com o MEGA GP, e se ele foi útil para auxiliar os estudos e alcançar os objetivos dos jogadores: (Q6) *Os simulados se tornam mais fáceis após uma rodada do MEGA GP?* (Q7) *O MEGA GP direciona os estudos para a prova?* (Q8) *É agradável participar do MEGA GP durante as aulas?* (Q9) *O uso do MEGA GP em sala de aula tem efeito motivacional?* Nesse contexto, a subcaracterística “*utilidade*” apresentou desempenho mediano. Ao analisar as respostas, foi verificado que a maioria (58,9%) dos estudantes reconheceu a utilidade do jogo na disciplina, destacando sua relevância para o estudo de Gerenciamento de Projetos de software. Ademais, nenhum aluno discordou totalmente sobre esse aspecto (Figura 2).

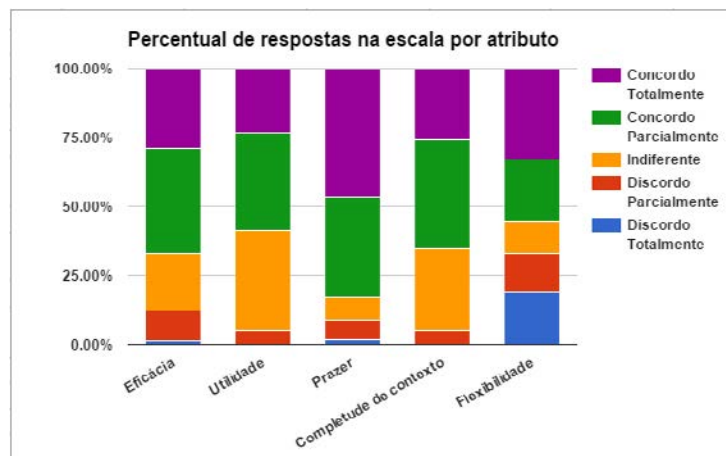


Figura 2 – Distribuição das respostas por atributo pesquisado.

Sobre a *cobertura do contexto*, as questões investigadas foram: (Q10) *É mais fácil estudar depois de jogar o MEGA GP?* (Q11) *As aulas são mais proveitosas com o MEGA GP?* (Q12) *A dinâmica das aulas com o MEGA GP prende a atenção dos estudantes?* (Q13) *É necessário o uso de muitos recursos para a rodada do MEGA GP?* Nesse cenário, a questão 13 revelou um desvio considerável (1,16) nas respostas (Tabela 3). Essa variação pode ser explicada devido a uma possível incompreensão dos estudantes na interpretação do enunciado da referida questão. Todavia, ressalta-se que a dinâmica do jogo exige apenas o uso de um cronômetro e as cartas com as senhas para a realização de uma partida.

5. Considerações Finais

Jogos, de maneira geral, são atrativos, estimulam a competitividade e motivam seus participantes a vencer. Desta forma, o uso da gamificação no processo de ensino-aprendizagem tem resultado positivo, pois motiva o estudante e melhora seu desempenho acadêmico de maneira lúdica e eficaz.

Diferentemente das propostas de jogos educacionais que propõem simular as atividades de engenharia de software, o MEGA GP visa à fixação de conceitos e suas importâncias. Por meio da experiência vivenciada, o jogo MEGA GP propiciou aos estudantes participantes do estudo a possibilidade de exercitarem seus conhecimentos acerca dos conceitos do PMBOK e do CMMI-DEV de maneira lúdica e motivadora. Em relação à receptividade dos alunos, a estratégia foi bem aceita e utilizada de maneira satisfatória.

Finalmente, a aplicação da gamificação no ensino de gerência de projetos e em outras áreas deve ser mais explorada, visto que sua utilização traz benefícios aos indivíduos envolvidos. A partir das pesquisas e experimentos relacionados ao tema e os resultados obtidos no presente trabalho, comprova-se que a estratégia é benéfica, segundo a visão dos discentes, e deve ser utilizada com proveito no meio acadêmico.

Referências

- Bonetti, T. M.; Wangenheim, C. G. (2013). Desenvolvimento de um repositório de jogos educacionais para o ensino de gerenciamento de projetos. *Anais do XXIV Simpósio Brasileiro de Informática na Educação – SBIE*, Campinas.
- Borges, S. S. et al. (2013). Gamificação aplicada à educação: um mapeamento sistemático. *Anais do XXIV Simpósio Brasileiro de Informática na Educação – SBIE*, Campinas.
- Campos, A.; Gardiman, R.; Madeira, C. (2015). Uma ferramenta gamificada de apoio à disciplina introdutória de programação. *Anais do XXIII Workshop sobre Educação em Computação – WEI*, Recife.
- Dubois, D. J.; Tamburrelli, G. (2013). Understanding gamification mechanisms for software development. *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*, p. 659-662.
- Figueiredo, K. S. et al. (2015). Uma abordagem gamificada para o ensino de programação orientada a objetos. *Anais do XXIII Workshop sobre Educação em Computação – WEI*, Recife.
- Jayasinghe, U.; Dharmaratne, A. (2013). Game based learning vs. gamification from the higher education students' perspective. *Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on*, p. 683-688.
- Jucá, P. M. et al. (2014). Aplicação da gamificação na disciplina de empreendedorismo. *Anais do XXII Workshop sobre Educação em Computação – WEI*, Brasília.
- Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. Pfeiffer. Hoboken, NJ.
- Kohwalter, T. C. et al. (2011). SDM – An educational game for software engineering. *Games and Digital Entertainment (SBGAMES), 2011 Brazilian Symposium on*, p. 222-231.
- Mattar, J. (2010). *Games em educação: como os nativos digitais aprendem*. São Paulo: Pearson Prentice Hall.
- Mira, S. H. B. et al. (2012). ProMEG: um jogo para ensino de gerência de projetos com foco na gerência de recursos humanos. *V Fórum de Educação em Engenharia de Software (FEES 2012)*, Natal.
- PMI. (2011). *Practice Standard Earned Value Management*, Second Edition, Pennsylvania, Project Management Institute.
- PMI. (2013). *Project Management Body of Knowledge – PMBOK Newton Square*. Pennsylvania: Project Management Institute. 5 ed.
- Prikladnicki, R. et al. (2007). Ensino de gerência de projetos de software com o Planager. *Anais do XVIII Simpósio Brasileiro de Informática na Educação – SBIE*, São Paulo.
- SEI. (2010). *CMMI for Development*, version 1.3. Carnegie Mellon University, Software Engineering Institute (SEI).
- Vianna, Y. et al. (2013). *Gamification, Inc: como reinventar empresas a partir de jogos*. Rio de Janeiro: MJV Press.

Melhora dos índices de aprovação pela formação de turmas com perfis de conhecimento semelhantes

Robinson Vida Noronha , Hilton José Silva de Azevedo

Universidade Tecnológica Federal do Paraná (UTFPR)

Departamento Acadêmico de Eletrônica

Curitiba - Paraná - Brazil

{vida, hilton}@utfpr.edu.br

Resumo. *Investigou-se se a formação de turmas de acordo com o coeficiente de rendimento do aluno poderia ter um desempenho acadêmico superior às turmas que não possuem esse tipo de critério de composição. Todas as turmas envolvidas nesse estudo tiveram os mesmos instrumentos de avaliação. As turmas formadas por alunos de baixo coeficiente de rendimento tiveram uma abordagem didática diferenciada. Os resultados obtidos apontam para um aumento nos índices de aprovação e reprovação das turmas e uma redução nos índices de desistência dos alunos.*

Abstract. *This paper report research results of the impact of clustering students by student's score criterion to student's success, failure and abandon of a particular class. The students with low student's score received a particular didactic strategy. The results of this research indicate that when classes are formed by student's score criterion, the index of number of student's success and student's failure increase. The results also indicate a reduction of the index of number of student's abandon of the class. All students that participate of this research (322) was assessed with the same tests and project's defense.*

Introdução

Ao ingressar em um curso superior em uma universidade, o aluno traz consigo a expectativa, da formatura. Porém, nem sempre isso ocorre. De acordo com [Gaioso 2006], [Manhães et al. 2012], [Costa et al. 2014], [Souza et al. 2015], entre outros, um dos problemas enfrentados pelas universidades é o crescente número da Evasão Escolar. Investigar as causas e possíveis estratégias que minimizem esse problema tem sido alvo de diversas investigações sem que o problema ainda tenha sido resolvido. Conforme apresentado na Seção 2 deste artigo, alguns relatos de pesquisa associam a Evasão Escolar ao elevado número de disciplinas reprovadas pelos alunos em um curso e à motivação de seus alunos.

Nesse contexto, Jesus (2008) destaca que os professores deveriam utilizar um ritmo de ensino adequado às capacidades e conhecimentos anteriores dos alunos como forma de motivá-los e em decorrência disso, reduzir os índices de reprovação. Porém, como fazer isso com as turmas heterogêneas de alunos? Nos dias de hoje, é comum o professor identificar nas disciplinas de programação de computadores, por exemplo, alguns alunos que já possuem conhecimento e habilidades de programação, enquanto que outros não as apresentam.

A presença de pelo menos esses dois tipos de alunos na mesma turma faz com que essa turma seja um desafio ao professor e em alguns casos, uma possível fonte de desmotivo para alguns alunos. O conjunto de atividades, a forma de abordagem dos conteúdos deveriam ser adaptados por esse professor aos diversos tipos de alunos. Essa adaptação poderia ter o objetivo de melhorar o desempenho dos alunos e que eles tivessem a sensação de serem capazes de executar as tarefas solicitadas.

Este trabalho de pesquisa apresenta um estudo de caso que busca identificar alguns efeitos no desempenho acadêmico de alunos quando o critério de seleção dos alunos na turma é Coeficiente de Rendimento Acadêmico desses alunos. Define-se assim, duas turmas, uma composta por alunos com coeficiente de rendimento elevado e outra com alunos com coeficiente de rendimento acadêmico baixo. Investiga-se também se uma pequena mudança nas estratégias pedagógicas do professor é capaz de modificar o desempenho acadêmico dos alunos com coeficiente de rendimento baixo.

A disciplina escolhida para o experimento é a Fundamentos de Programação 2. Essa disciplina é ofertada no segundo período do curso de Engenharia Eletrônica da Universidade Tecnológica Federal do Paraná. Os alunos que cursam essa disciplina são divididos em duas turmas. O início do estudo foi no segundo semestre acadêmico de 2013 e continua em andamento. Participaram desse estudo quatro professores e 322 alunos no total. Os dados coletados são contrastados nessa pesquisa com uma terceira turma (turma referência ou controle) que não teve o mesmo critério de formação, o coeficiente de rendimento acadêmico dos alunos.

Para descrever o estudo de caso e seus resultados, este artigo está estruturado da seguinte maneira: a Seção 2 analisa alguns trabalhos a respeito de formação de grupo e desempenho e possíveis fatores associados à Evasão Escolar. Na Seção 3, o experimento é descrito. A análise dos dados coletados é feita na Seção 4 e uma discussão é apresentada na Seção 5. Conclui-se este artigo com trabalhos futuros na Seção 6.

2. Estado da Arte

Um dos motivos que poderia ser considerado como uma das causas da Evasão Escolar, é o baixo desempenho acadêmico do aluno [Manhães et al 2012], [Souza et al. 2015] e [Santos et al 2015]. Porém, outro ponto de vista desse fenômeno poderia indicar que tanto a Evasão Escolar como o baixo desempenho acadêmico do aluno estão relacionados com a falta de engajamento do aluno com o curso ou com a falta de motivação desse aluno. Independentemente do ponto de vista considerado, há de se constatar que existem indícios da existência de relacionamento entre a Evasão Escolar e o baixo desempenho acadêmico do aluno.

Ao estudar a evasão em um curso de Licenciatura em Informática do IFRN, o trabalho de [Souza et al 2015] identificou, por meio de entrevistas com esses alunos dois motivos destacados neste trabalho que indicam que o desempenho acadêmico individual do aluno pode estar associado à evasão escolar. Os motivos identificados pelos alunos como causas do abandono foram, entre outros: i) a dificuldade de acompanhar a disciplina de algoritmos e outras disciplinas técnicas e ii) os métodos de ensino, métodos de avaliação e dificuldade de relacionamento com os professores.

Outro trabalho que também relaciona o abandono com a dificuldade encontrada pelos alunos em disciplinas é o de Manhães et al. (2012). Por meio de técnicas de mineração de dados, Manhães et al. identificaram em bases de dados computacionais uma redução gradual da quantidade de disciplinas aprovadas até o momento da evasão do curso dos alunos. Essa redução apontada por Manhães et al, bem como a pouca quantidade de empréstimos de livros na biblioteca feitas pelos alunos que desistem da universidade [Oliveira Junior et al. 2014], poderiam ser interpretadas como um indício da pouca motivação desses alunos em relação às disciplinas e ao curso. O efeito da motivação nos índices de aprovação em disciplinas de programação é constatado por Santos et al. (2015).

Além desses aspectos, [Barker et al, 2009] apontam que o relacionamento entre os alunos é um fator importante a ser considerado quando o assunto é Evasão Escolar, ou retenção de alunos. O trabalho de Barker et al, relata que as chances do aluno permanecer no curso aumentam quando ele se sente parte de uma comunidade de mesmo nível social e intelectual. O que é esperado que aconteça neste experimento quando as turmas são divididas pelo coeficiente de rendimento. Ou seja, que os alunos desistam menos da disciplina e, conseqüentemente, haja um aumento no número de aprovados ou reprovados.

Diversos estudos a respeito de formação de grupos tem apresentado resultados que, de acordo com [Citadin et al. 2014], não podem ser generalizados. Depende da tarefa que o grupo realizará. Por exemplo, de acordo com [Manukian et al, 2013] grupos homogêneos (grupos com as mesmas características) são melhores para disseminar conhecimento.

3. O Experimento

Investigou-se o impacto que duas abordagens: uma de gestão e outra didática teriam sobre o desempenho de estudantes de uma disciplina de programação de segundo semestre do primeiro ano de um curso de Engenharia Eletrônica. A abordagem de gestão consistiu em dividir parte dos estudantes em dois grupos, um com os maiores coeficientes de rendimento escolar e o outro com os menores e observar se essa divisão tinha impacto no desempenho acadêmico.

A abordagem didática consistiu em oferecer um tratamento diferenciado para um grupo de estudantes com menores coeficientes de rendimento. O estudo nasceu de uma iniciativa espontânea de um dos professores da disciplina que buscavam alternativas para reduzir a quantidade de estudantes repetentes nessa disciplina.

O método empregado foi o experimental investigativo. A pergunta de pesquisa procurou responder se o fato de se separar os alunos fortes dos fracos em uma turma poderia ter um impacto positivo no desempenho dos estudantes de uma disciplina de programação. Como o fato motivador era a busca de ações que pudessem contribuir para a redução da quantidade de alunos repetentes em turmas iniciais do curso de engenharia, decidiu-se que a divisão não seria o foco principal mas o que se poderia fazer para auxiliar os prováveis futuros repetentes. Isso provocou o desenho de uma abordagem diferenciada para o grupo dos estudantes mais fracos. Para assegurar o rigor do processo de ensino aprendizagem, todos os participantes foram submetidos aos mesmos instrumentos de avaliação.

3.1 Caracterização da disciplina.

A disciplina Fundamentos de Programação 2, objeto deste estudo, é organizada em dois encontros de duas aulas semanais cada. Nessa disciplina, ofertada no segundo semestre do curso de engenharia eletrônica, os estudantes são apresentados ao paradigma Orientado a Objeto, tendo como plataforma a linguagem “C++”. Os aprovados devem demonstrar compreender e empregar os conceitos de classe, herança, polimorfismo, abstração e habilidades no desenvolvimento e documentação de projetos de software, entre outros. A avaliação é feita mediante provas escritas e um projeto de desenvolvimento de "software" que deve ser defendido perante banca formada pelos professores da disciplina. O estudo cobriu cinco semestres consecutivos em que a disciplina foi ofertada em três turmas por vez.

3.2 Organização das turmas.

A distribuição dos estudantes nas turmas foi condicionada ao regulamento de matrícula vigente na universidade. Foram ofertados: um horário noturno com um terço das vagas disponíveis e um horário vespertino com dois terços das vagas disponíveis. O horário noturno serviria como grupo de referência (o preenchimento da turma teve seus coeficientes variados em função das demandas espontâneas dos estudantes que se matricularam e o professor conduziu a disciplina sem mudanças de abordagem em relação aos semestres anteriores ao experimento). O horário vespertino foi organizado em duas turmas, intencionalmente divididas em função dos maiores ou menores coeficientes de rendimento dos inscritos.

Devido a peculiaridades administrativas, o corpo docente participa de vários cursos (por exemplo: Engenharia Eletrônica, Engenharia de Computação, Engenharia Mecatrônica e Tecnologia de Telecomunicações) e nem sempre o mesmo professor pode assumir as mesmas disciplinas ao longo dos semestres. Esse fato fez com que quatro professores atuassem na disciplina de Fundamentos de Programação 2 ao longo dos cinco semestres do experimento.

3.3. Caracterização da mudança de abordagem pedagógica.

A alteração da abordagem consistiu na distribuição, para as turmas com menores coeficientes de rendimento, de tarefas semanais de resolução de exercícios em casa e entrega para o professor, correção dos exercícios pelo professor (extra classe) e posterior devolução dos mesmos e apresentação de exemplos de resolução dos exercícios em aula.

3.4. Dados considerados no experimento.

Os dados do experimento foram obtidos diretamente dos diários de classe: Média de Desempenho, situação do estudante ao final da disciplina (aprovado, reprovado, desistente). Todos os estudantes realizaram as mesmas avaliações. A nota relativa ao projeto no final do semestre, das três turmas de cada semestre, foi construída mediante defesa perante banca composta pelos quatro professores.

3.5. Critérios para análise dos dados.

Como somente um dos quatro professores mudou sua abordagem pedagógica quando trabalhava com as turmas com baixo coeficiente de rendimento, adotou-se o

agrupamento apresentado na Tabela 1, para as 15 turmas que participaram do experimento (3 turmas a cada semestre entre 2013/2 e 2015/2).

As variáveis principais deste experimento são: número de aprovados, reprovados e desistentes nos três grupos. A variável secundária considerada é a média de rendimento nos grupos dos aprovados e reprovados.

Tabela 1 - Agrupamento das Turmas

GrupoBN	Grupo Baixo coeficiente abordagem Não diferenciada (46 estudantes)
GrupoBD	Grupo Baixo coeficiente abordagem Diferenciada (64 estudantes)
GrupoCA	Grupo Coeficiente ALTO e sem abordagem diferenciada (100 estudantes)
GrupoRE	Grupo de referência. Coeficientes misturados e sem abordagem diferenciada(112 estudantes). Os estudantes que participam dessa turma são estudantes que já cursaram a disciplina pelo menos uma vez e repetiram por nota ou nota e frequência.

4. Análise dos dados obtidos

Nessa seção apresenta-se os dados coletados a partir de consultas em um Sistema de Gerenciamento de Informações Acadêmicas. Esses dados estão agrupados e são apresentados nas seguintes seções:

- * Seção 4.1: Dados de desempenho acadêmico entre os grupos com baixo coeficiente de rendimento acadêmico
- * Seção 4.2: Dados de desempenho acadêmico do grupo com coeficiente de rendimento acadêmico elevado.
- * Seção 4.3: Dados de desempenho acadêmico dos grupos com coeficiente de rendimento acadêmico alto e baixo com abordagem diferenciada e da turma com os coeficientes misturados (grupo controle).
- * Seção 4.4: Dados de desempenho acadêmico total das turmas divididas por coeficiente acadêmico alto e baixo com e sem abordagem diferenciada e a turma com os coeficientes misturados (grupo controle).

4.1. Dados relativos de desempenho entre os grupos com baixo coeficiente acadêmico.

A Tabela 2 compara os resultados gerais de desempenho, entre os grupos com baixo coeficiente acadêmico, tendo como variável de controle a adoção ou não de abordagem diferenciada. Lembrando que as provas são as mesmas nos dois grupos, observou-se um desempenho levemente melhor no grupo de estudantes com baixo coeficiente que fez exercícios extras de fixação:

- Uma média de rendimento ligeiramente superior (6,9 contra 6,6),

- 5,5% a mais de estudantes aprovados,
- 16% a menos de estudantes reprovados,
- um aumento de 10,5% no número de desistentes e
- apesar de bastante baixa, uma melhora na média de rendimento dos reprovados que passou de 1,9 para 2,8 (aumento de 80%).

Tabela 2 Comparação de desempenho e frequências médios entre turmas com atividade específica para os mais fracos.

	BAIXO COEF. ABORDAGEM PADRÃO - GrupoBN			BAIXO COEF. ABORDAGEM DIFERENCIADA - GrupoBD		
	%	MÉDIA DA TURMA	DESV. PADRAO	%	MÉDIA DA TURMA	DESV. PADRAO
APROVADOS	19,57%	6,6	0,5	25,00%	6,9	0,9
REPROVADOS	63,04%	1,9	1,2	46,88%	2,8	1,7
DESISTENTES	17,39%	----	----	28,13%	---	---

4.2. Dados Referentes ao grupo de coeficiente alto

A Tabela 3 compara os resultados gerais de desempenho acadêmico médio, entre os grupos Referência (GrupoRE) e Coeficiente Alto - GrupoCA. Observou-se algumas melhoras nos índices do GrupoCA, quando comparado ao Grupo RE, a saber:

- A quantidade de aprovados foi maior - 74% para o GrupoCA e 43% para o GrupoRE,
- A quantidade de reprovados foi menor - 16% para o GrupoCA e 27% para o GrupoRE,
- A quantidade de desistentes foi menor - 10% para o GrupoCA e 30% para o GrupoRE,
- A média das turmas permaneceu praticamente a mesma (7,8) nos grupos GrupoCA e GrupoRE.

Tabela 3 - Comparação de desempenho entre turmas de Coeficiente de Rendimento Elevado e o Grupo de Referência

	GrupoRE			GrupoCA		
	%	MÉDIA DA TURMA	DESV. PADRAO	%	MÉDIA DA TURMA	DESV. PADRAO
APROVADOS	43%	7,8	1,1	74%	7,8	1,2
REPROVADOS	27%	2,2	1,2	16%	2,3	1,4
DESISTENTES	30%			10%		

4.3. Dados Relativos à Reunião das Turmas com Coeficiente alto e baixo com abordagem diferenciada e a turma com os coeficientes misturados.

A Tabela 4 compara o impacto geral que a separação de estudantes em fortes e fracos (os fracos com abordagem diferenciada) com um grupo de estudantes que tem

os fortes e fracos misturados (sem mudança de abordagem pedagógica). Como anteriormente, as provas foram as mesmas nos dois grupos. Observou-se um desempenho levemente melhor no grupo que teve seus estudantes fortes e fracos separado e teve abordagem diferenciada para os mais fracos:

- Uma média de rendimento levemente inferior para o grupo com coeficiente separados (7,7 contra 7,8 para o grupo),
- 10% a mais de aprovados,
- 1% a mais de reprovados e
- 13% a menos de desistentes

Tabela 4: Comparação de desempenho e frequências médios do grupo de referência com a grupo reunião dos grupos alto coeficiente (sem mudança de abordagem) e baixos coeficientes (com mudança de abordagem).

	GrupoCA + GrupoBD			GrupoRE		
	%	MÉDIA DA TURMA	DESV. PADRAO	%	MÉDIA DA TURMA	DESV. PADRAO
APROVADOS	55%	7,7	1,2	43%	7,8	1,1
REPROVADOS	28%	2,6	1,7	27%	2,2	1,2
DESISTENTES	17%			30%		

4.4. Dados Relativos à Reunião das Turmas com Coeficiente alto e baixo com e sem abordagem diferenciada e a turma com os coeficientes misturados.

A Tabela 5 totaliza e compara os resultados gerais de desempenho, entre todos os grupos envolvidos no experimento.

Observou-se algumas melhoras em alguns índices, a saber:

- A quantidade de aprovados foi maior (47%) quando a turma foi dividida por coeficiente de rendimento alto e baixo quando comparada com a turma que não teve essa divisão (43%).
- A quantidade de reprovados das turmas divididas pelo coeficiente de rendimento foi maior (37%) quando comparada com a turma que não teve essa divisão (28%).
- A quantidade de desistentes foi menor para as turmas que foram divididas pelo coeficiente de rendimento (16%) quando comparada com a turma controle (29%).

5. Discussão

Apesar do número de estudantes envolvidos ser baixo (322 no total) o experimento trouxe luz a alguns pontos que merecem ser aprofundados em estudos futuros. No que se refere aos estudantes que chegam na disciplina de Fundamentos de Programação 2, com baixo coeficiente de rendimento, os dados obtidos sugerem que a adoção de abordagem didática diferenciada é um estratégia que auxilia os estudantes a melhorarem seu desempenho.

5.1. Impacto da separação de fortes e fracos juntamente com abordagem didática diferenciada no fluxo dos estudantes no curso

Separando-se os estudantes com baixo e alto coeficientes de desempenho acadêmico, observou-se que o número de aprovados aumentou de 10% ao mesmo tempo em que se diminuiu o número de desistentes e que a média do desempenho acadêmico dos reprovados aumentou de 2,2 para 2,6. Isso pode ser um tipo de ação a ser investigado com o intuito de se minimizar a quantidade de estudantes repetentes em certas disciplinas.

5.2 Considerações sobre o conceito consolidado de que “misturar estudantes com perfis de desempenho acadêmico variados auxilia os mais fracos a superarem as suas dificuldades”

Foi uma surpresa observar que tanto os estudantes fortes quanto o fracos, saíram-se melhor quando separados (os fracos com abordagem didática diferenciada). A hipótese que se pode fazer, e necessita de estudos para sua verificação é a de que a manutenção de fortes e fracos só tem sentido quando o conhecimento prévio do grupo

Tabela 5 - Resultado total do experimento - Quantidade de Alunos Aprovados, Reprovados e Desistentes

	(GrupoCA + GrupoBD + GrupoBN) Valor Absoluto	(GrupoCA + GrupoBD + GrupoBN) %	(GrupoRE) Valor Absoluto	GrupoRE%
APROVADOS	99	47%	48	43%
REPROVADOS	77	37%	31	28%
DESISTENTES	34	16%	33	29%
TOTAL DE ALUNOS	210	100%	112	100%

de estudantes caracteriza uma curva normal de desempenho. No caso estudado, devido ao processo de entrada na universidade dividir os ingressantes em 5 grupos de cotas (regulamentado por lei federal), a distribuição normal de desempenho dos estudantes que compõem uma turma de entrada deixa de existir. Justificando que a abordagem didática diferenciada apresente melhor resultado do que um tratamento didático igualitário para todos os estudantes.

5.3. Considerações sobre a natureza da abordagem didática diferenciada.

Neste experimento, a única mudança realizada foi a solicitação de resolução de exercícios em casa, sua correção e devolução e a posterior socialização de exemplos de como poderiam ter sido resolvidos. Os pesquisadores tem consciência de que a abordagem didática pode ir muito além de designação de tarefas, passando por materiais didáticos específicos, adoção de gêneros de discurso mais adequados, atitudes motivadoras, desenvolvimento de autonomia e autoestima junto aos estudantes.

5.4 Considerações sobre o domínio dos pré-requisitos para a disciplina

Neste experimento não se procedeu a um teste de nivelamento inicial que pudesse identificar a distribuição do nível de competência dos estudantes no que se referia a programação estruturada (tema da disciplina precedente). Este ponto é importante porque a distribuição das notas mostram duas concentrações de notas, uma na parte baixa da escala de desempenho e outra na parte altas, ficando o centro da escala com poucos indivíduos.

6. Conclusão

Aumentar os índices de alunos aprovados em qualquer disciplina de um curso superior é um desejo de muitos gestores educacionais (professores, coordenadores de curso, entre outros). Alguns argumentam que esse aumento de índice estaria associado a uma possível piora no nível de exigência que os professores empregam ou que estaria associado com um conceito de “facilitar” para o aluno passar.

Nessa pesquisa, relata um pequeno aumento no desempenho dos alunos quando turmas são formadas pelo critério de seleção por coeficiente de rendimento acadêmico quando comparada com a turma onde isso não ocorreu. No cômputo geral, houve uma melhora no índice de aprovados (47% para o grupo que recebeu a intervenção e 43% para o grupo que não recebeu a intervenção), uma redução nos índices de desistentes (16% para o grupo que recebeu a intervenção e 29% para o grupo que não recebeu a intervenção) e um aumento na quantidade de reprovados para o grupo que recebeu a intervenção.

Um aspecto precisa ser destacado ao analisarmos esses números: o grupo que não recebeu a intervenção era formado basicamente por alunos que já tinham cursado a disciplina sem conseguir a aprovação pelo menos uma vez. Se o grupo controle fosse formado por alunos que ainda não cursaram a disciplina, como foi o caso do grupo que recebeu a intervenção, esses números poderiam ser mais promissores.

A redução do índice de desistência vem confirmar os resultados da pesquisa de [Barker et al 2009] que relata que as chances do aluno permanecer no curso aumentam quando ele se sente parte de uma comunidade de mesmo nível social e intelectual. Nesse caso, a turma de alunos onde a intervenção não ocorreu, o GrupoRE, apresentou índice de 30% de desistentes. Um valor bem superior aos índices apresentados pelas turmas onde ocorreu a intervenção (índice de 10% para a turma GrupoCA, 17% para turma GrupoBN e 28% para a turma GrupoBD).

O trabalho futuro a ser desenvolvido é investigar quais técnicas ou abordagens pedagógicas poderiam ser empregadas para a turma dos alunos com coeficiente de rendimento baixo de tal forma a melhorar ainda mais esses índices obtidos até o presente momento. Uma dessas técnicas candidatas é a Aprendizagem Baseada em Problemas. Um grupo de estudo foi criado com o objetivo de se capacitar ao uso dessa técnica em sala de aula.

Referências

- Barker, L., J., McDowell, C., Kalahar, K. (2009). Exploring Factors that Influence Computer Science Introductory Course Students to Persist in the Major. *ACM SIGCSE Bulletin*, 41(2), pp. 282-286.
- Citadin, J., R., Kemczinski, A., Matos, A., V. (2014). Formação de Grupos para Aprendizagem Colaborativa: Um mapeamento sistemático da literatura. *Nuevas Ideas en Informática Educativa TISE 2014*.
- Costa, S., S., Cazella, S., Rigo, S.J. (2014). Mineração de dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre a evasão escolar na UNA-SUS. *CINTED - Novas Tecnologias na Educação*, V. 12, No.2, dezembro de 2014.
- Gaios, Natalicia Pacheco de Lacerda (2006) O Fenômeno da Evasão Escolar na Educação Superior no Brasil. Unesco, 2006. Disponível em: www.iesalc.unesco.org.ve/programas/Deserción/Informe. Acesso em 19/03/2016
- Jesus, S. (2008). Estratégias para motivar os alunos. *Revista Educação*, 31 (1), 21-29.
- Santos, A., Gorgônio, A., Lucena, A., Gorgônio F. (2015). A Importância do Fator Motivacional no Processo Ensino-Aprendizagem de Algoritmos e Lógica de Programação para Alunos Repetentes. *Anais do Workshop de Educação em Informática - WEI 2015. XXXV Congresso da Sociedade Brasileira de Computação. 20 a 23 de julho de 2015 - Recife - PE*.
- Souza, O., S. Morais, P. S., Silva Júnior, F., C. (2015). Um Estudo sobre a Evasão no Curso de Licenciatura em Informática do IFRN – Campus Natal – Zona Norte. *Anais do Workshop de Educação em Informática - WEI 2015. XXXV Congresso da Sociedade Brasileira de Computação. 20 a 23 de julho de 2015 - Recife - PE*.
- Manukyan, N., Eppstein, M. J., e Horbar, J. D. (2013). Team structure and quality improvement in collaborative environments. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, IEEE, pp. 523-529.
- Manhães, L., M., B., Cruz, S., M., S., Costa, R., J., M., Zavaleta, J., Zimbrão, G. (2012). Identificação dos Fatores que Influenciam na Evasão em Cursos de Graduação Através de Sistemas Baseados em Mineração de Dados; Uma Abordagem Quantitativa. *VIII Simpósio Brasileiro de Sistemas de Informação*.
- Oliveira Junior, J., G., Noronha, R., V., Kaestner, C., A., A., (2014) Análise da Correlação da Evasão de Cursos de Graduação com o Empréstimo de Livros em Bibliotecas. *Anais do Congresso Brasileiro de Informática na Educação. Workshop de Mineração de Dados em Ambientes Virtuais do Ensino e Aprendizagem*, pag. 601-610. Dourados - Mato Grosso do Sul, 03 a 06 de novembro de 2014.

Modos contemporâneos de aprendizado e construção do conhecimento: reflexões sobre o ensino de Teoria da Computação para Sistemas de Informação

Isabel Cafezeiro^{1,2}, Leonardo Cruz da Costa^{1,3}, Ricardo Kubrusly²

¹ Instituto de Computação – Universidade Federal Fluminense (UFF)
Niterói – RJ – Brasil

² Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia
(UFRJ) Rio de Janeiro – RJ – Brasil

³ Programa de Pós-Graduação em Ciências da Informação (UFF)
Niterói – RJ – Brasil

{isabel,leo}@ic.uff.br, risk@hcte.ufrj.br

Abstract. *This article describes the experience of designing and implementing a discipline in the under graduation course in Information Systems that seeks to address the contents of the Theory of Computation, a topic that has been, over the years, deprecated in Computer Science courses and that is not part the usual curricula in Information Systems. We argue that a proposal based on interdisciplinarity through sociotechnical approach can reframe these contents bringing out a great contribution to the understanding of the current configuration of computer systems through its historical process of construction.*

Resumo. *Este artigo descreve a experiência de concepção e implantação de uma disciplina em curso de Sistemas de Informação que procura abordar o conteúdo de Teoria da Computação, um tema que vem sendo ao longo dos anos preterido nos cursos de Ciências da Computação e que não faz parte dos currículos usuais em Sistemas de Informação. Argumentamos que uma proposta ambientada na interdisciplinaridade através da abordagem sociotécnica pode reposicionar estes conteúdos colocando em evidência uma grande contribuição na compreensão da configuração atual dos sistemas computacionais através do seu processo histórico de construção.*

1. As duas culturas

Nossas análises têm por base a observação dos cinco primeiros anos de operação do curso de Sistemas de Informação da Universidade Federal Fluminense (UFF), na cidade de Niterói, RJ. Arriscamos um diagnóstico ao problema do alto grau de retenção de estudantes nos primeiros três períodos do curso. Trata-se justamente do momento em que se concentram as disciplinas de base matemática, como cálculo, lógica e matemática discreta, que apresentam um conteúdo não familiar tanto para o estudante recém-saído do ensino médio quanto para os que já possuem alguma experiência no mercado de trabalho em Sistemas de Informação e retornam à universidade em busca da

titulação. Completam o quadro de retenção as disciplinas iniciais de programação¹, também ministradas nos primeiros períodos.

A retenção não é uma característica exclusiva dos cursos de Sistemas de Informação. Ela também se faz presente nos cursos de Ciências da Computação, referente a este mesmo conjunto de disciplinas. De modo geral em computação, retenção, dificuldade de aprendizagem e evasão vem demandando atenção nos fóruns de discussão de currículos e metodologias de ensino², e surpreende que este problema não tenha ainda encontrado solução satisfatória em cursos de longa tradição como os de Ciência da Computação (o curso de graduação em Ciências da Computação da UFF, em Niterói, foi implantado em 1986, mas os primeiros do Brasil datam de 1969). Porém é possível que os cursos de Sistemas de Informação, por fundamentarem-se em uma abordagem diferente e recente quando comparada ao tradicional enfoque em Ciências da Computação, sejam capazes de atrair atenção para uma nova dinâmica de compreensão do conhecimento, algo que pode ser elaborado de maneira a sugerir transformações e metodologias inovadoras no ensino tanto em Sistemas de Informação quanto nas Ciências da Computação. Trata-se da abordagem sociotécnica.

Desde os primeiros contatos com as apresentações dos cursos de Sistemas de Informação, o aluno percebe que se trata de um campo de conhecimento não completamente incluso na área de Ciências da Computação. As páginas de apresentação dos cursos costumam indicar uma forte interdisciplinaridade, em que a computação é situada como atividade meio, capaz de capacitar o profissional para a utilização instrumental da tecnologia na solução criativa de problemas diversos. É uma abordagem que encontrou suporte nos documentos referenciais da área, como o currículo de referência da Sociedade Brasileira de Computação (SBC, 1999). Essa multiplicidade de campos de conhecimentos manifestou-se na grade curricular do Curso de Sistemas de Informação da UFF pela colaboração com diversos departamentos (direito, sociologia, psicologia, administração, ciências da informação) através do oferecimento de disciplinas em turmas mistas (não exclusivas a estudantes de Sistemas de Informação), o que veio a fortalecer o diálogo dos estudantes com as demais áreas. Fora das salas de aula, o curso contou com o apoio de um Núcleo Docente Estruturante, e com o Colegiado de Curso, que, atentos ao caráter híbrido requerido na formação do profissional de Sistemas de Informação, contribuíram para oferecer com sucesso a diversificação do corpo de conhecimentos presente na grade curricular.

O que não se verificou, entretanto, foi a efetivação da abordagem sociotécnica no interior de cada disciplina, e principalmente daquelas disciplinas consideradas “exatas”, como as matemáticas. O estudante se deparou com uma perceptível fronteira:

¹ No caso da programação de computadores, precisamos notar que o mecanismo de raciocínio que esta disciplina exige difere da práxis do ensino médio, onde os conteúdos considerados exatos são apresentados em uma abordagem predominantemente teorematizada (isto é, que não deixa aparentes processo e motivação de construção das formulas, focando-se o modo de aplicação, e, quando muito, justificativas dedutivas). Diferentemente, a atividade de programação exige um pensamento algorítmico, centrado no “como fazer”, raciocínio aderente ao problema. Sobre a diferença entre as abordagens teorematizada e problemática, ver Deleuze & Guatarri (1997, p. 26).

² Para um exemplo recente, remetemos ao Curso de Qualidade da Sociedade Brasileira da Computação (CQ, 2015), cujo tema foi “Evasão e estratégias para permanência estudantil” (http://csbc2015.cin.ufpe.br/eventos_descricao/8). Percebemos também que o tema é recorrente nas diversas edições do WEI. No ano de 2015, por exemplo, podemos ver Souza, Morais, & Junior (2015).

as disciplinas ministradas pelos departamentos de computação e matemáticas são apresentadas como “técnicas” (ou “exatas”). O aprendizado é centrado em mecanismos abstratos ou metodologias validadas no cenário acadêmico mundial e as avaliações focalizam a aplicação destes mecanismos. As outras disciplinas, principalmente as ministradas nos departamentos das áreas humanas são apresentadas como disciplinas “humanísticas”, “não-técnicas”. Neste caso, o debate e a expressão oral e escrita são incentivados, e as avaliações não seguem necessariamente o formato das provas tradicionais e a abordagem demanda um posicionamento crítico pessoal que se afasta de um caráter supostamente neutro, usual no ensino das técnicas. Sem que se efetive o diálogo entre estas abordagens, modo de apresentação, conteúdos, avaliações, tudo se altera quando se passa de um universo para o outro. É uma diferença que acompanha a configuração dos saberes modernos, e que na UFF se faz visível também espacialmente, na separação dos campi universitários (as disciplinas humanísticas são oferecidas no Campus do Gragoatá, onde se concentram os departamentos e cursos correspondentes. Por outro lado, o Instituto de Computação se localiza no Campus da Praia Vermelha, e lá são ministradas as disciplinas da matemática e da computação). Latour (2014, tradução nossa) observou:

[P]arece que estamos sofrendo de uma divisão do trabalho intelectual: se você aprendeu história, ciências sociais, literatura, direito, arte ou qualquer ramo do que é chamado de "humanidades", tenho certeza que você aprendeu um monte de coisas, mas não necessariamente sobre como a tecnologia e a ciência têm interferido em todos esses vários campos; por outro lado, se você assistiu aulas ou teve formação em engenharia, informática, ciências naturais ou negócios aplicados, não tenho a certeza que você vai ter aprendido algo sobre quando, onde e por quem esses elementos de conhecimento foram produzidos e que eles relações entretêm com a história, literatura, arte, política ou ciências sociais. (...) Qualquer que seja o campo de onde você vem, há chances de que você tenha percebido de que existe uma divisão entre aqueles que vêm "das letras" e aqueles que vêm "das ciências". E não apenas uma divisão, mas muitas vezes, infelizmente, alguma forma de condescendência, mesmo de rancor contra os do outro lado: você poderia ter sido tratado como um "nerd analfabeto" por algum estudioso literário brilhante que, por sua vez, poderia ter sido tratado como um "idiota romântico" por alguns colegas do sério departamento de matemática.

Diante desse quadro de cisão entre os conhecimentos, recusamos qualquer diagnóstico referente à retenção, evasão e dificuldades de aprendizagem que se situe na falta de preparo dos estudantes de Sistemas de Informação, e arriscamos dizer que o baixo desempenho nas disciplinas das ditas “ciências exatas” nasce no conflito entre a expectativa por um conhecimento híbrido e problematizador e o defrontamento com um conhecimento que se pretende puramente técnico. Este conflito é desmotivador, uma vez que não conduz à compreensão da utilidade destes conhecimentos na profissão. Traz para a reflexão o contraste com o alto desempenho nas ditas “humanísticas”, um fato que, ignorando-se a familiaridade destes conteúdos a estes estudantes desde o ensino médio, ainda desencadeia o argumento de que as chamadas “humanísticas” são “mais fáceis” do que as ditas “exatas”. A este impasse, referindo-se especificamente à Engenharia de Software, Cukierman et al., em 2007 já haviam proposto um caminho:

A saída para os impasses criados pela separação entre o técnico e o social consiste em mudar o ângulo de aproximação e, assim, percebê-los por um novo enquadramento. Resumindo brevemente, um enquadramento em que o técnico e o social constituem um movimento de "comodificação", somente percebido por uma abordagem concomitantemente social e técnica, por um olhar sociotécnico. Um olhar que busca apreender a ES sem fragmentá-la em "fatores ou aspectos técnicos" de um lado, e "fatores ou aspectos não-técnicos" de outro, sem fatorá-la em quaisquer outras dualidades ("fatores técnicos" versus "fatores humanos, organizacionais, éticos,

políticos, sociais, etc.") que terminem por desfigurar o "pano sem costura" que imbrica na ES o técnico e o social em um mesmo e indivisível tecido.

2. Teoria da Computação e os currículos de Sistemas de Informação

Se por um lado a saída apontada por Cukierman et al. encontra acolhimento em áreas como a Engenharia de Software, Interação Humano-Computador, Jogos, e outras especialidades que, de uma forma ou de outra, admitem algum diálogo com a sociedade, o mesmo não se verifica com a área da Teoria da Computação. Aqui, nos referimos precisamente aos conhecimentos relacionadas ao estudo da Computabilidade.

Muitas questões abordadas no campo da Computabilidade se estabeleceram antes da construção dos primeiros computadores, em um ambiente de busca pela compreensão da capacidade de expressão formal (sistemas formais). Naquele momento, início do século XX, percebiam-se certas idiossincrasias, como os paradoxos e enunciados não provados, que não eram bem-vindas numa ciência que buscava controle, exatidão e objetividade. Mas, em consonância com a concepção da ciência moderna de uma matemática completa e totalizante, buscavam-se no seio da própria matemática, explicações para essas esquisitices, dando margem aos programas de fundamentação da matemática do início do século XX (Cafezeiro et al, 2010). Resultados concebidos no contexto da afirmação da matemática enquanto disciplina autônoma migraram para a computação como fundamentação para o campo incipiente. Dali para a década de 1960, já com cerca de 15 anos de operação dos computadores, outros personagens, os fundadores da computação, reconheceram no estudo de sistemas formais e suas propriedades o coração da computação e a justificativa para a institucionalização do campo. Daí se compreende uma certa resistência para a abordagem híbrida: herança da tradição matemática como um campo purificado, que não se contamina com os demais, fortalecimento do caráter científico do novo campo a partir da aderência a um campo já estabelecido. Embora alguns cientistas, como Alan Perlis, insistissem na interação com saberes relacionados à gestão e à informação, outros pioneiros do campo da computação, como Edsger Dijkstra, Tony Hoare, Nicklaus Wirth, apostaram no caráter "exato" e se opuseram à institucionalização do campo em co-construção com outros saberes "não exatos".

Argumentos em favor das abordagens formais podem ser lidos nos relatórios das famosas conferências da OTAN (Organização Militar do Atlântico Norte) de 1968 e 1969 (NATO 1969,1970), onde, sob o pretexto de discutir a "crise do *software*", terminou-se por delimitar os contornos do que se passou a chamar de "Ciência da Computação". O trecho a seguir mostra o quão frágil era, naquele momento a compreensão do campo que vinha então se configurando (NATO,1970, tradução nossa,p.49):

Strachey: (...). Acho que devemos nos perguntar seriamente a seguinte questão: é a computação uma ciência? (...) Até termos um corpo suficiente de tópicos que sejam importantes, relevantes e bem desenvolvidos, não podemos chamar o assunto de "ciência".

Woodger: Eu me pergunto com o se parece que um princípio básico em informática, se eu me deparasse com um. Sério! Em que termos ele poderia ser expresso? Não é bom usar termos inexplicáveis como "língua" ou "compilador"; temos muito poucos termos claros em que para expressar qualquer princípio fundamental da ciência da computação. Será que algum professor aqui pode me dizer um único princípio básico?

Grande parte das discussões girava em torno de um impasse entre “teoria” e “prática”. Os defensores da “teoria” encontravam nos métodos matemáticos a garantia de um campo bem fundamentado e uma atividade de programação livre de erros. Aqui Hoare e Dijkstra, em divergência com Perlis, ironizam os testes informais em programas e argumentam em favor das provas formais (NATO, 1970,p.16):

Hoare: Pode-se, muito facilmente, construir provas convincentes da futilidade de testes exaustivos em um programa e até mesmo de testes por amostragem. (...) Teste dos casos base poderiam, por vezes, serem automatizados. Atualmente, isto é basicamente teoria (...) Esta área de trabalho teórico parece mostrar uma possibilidade de resultados práticos, embora provar correção seja um processo trabalhoso e caro. Talvez não seja um luxo para certas áreas cruciais de um programa.

Perlis: Grande parte da complexidade do programa é espúria e uma série de casos de teste devidamente estudados esgota o problema dos testes. O problema consiste em isolar casos de teste adequados, e não provar o algoritmo, pois isto [a correção] se segue após a escolha dos casos de teste adequados.

Dijkstra: Teste mostra a presença, não a ausência de bugs.

Possivelmente em função do forte apelo dos pioneiros que argumentavam em favor de métodos formais, questões relacionadas à decidibilidade ganharam importância nos currículos da computação, no sentido da compreensão das capacidades e limitações das abordagens formais. No Brasil, a maioria dos cursos em computação que começaram a surgir na década de 1970 (as exceções são Ciências da Computação da UNICAMP e Processamento de Dados da UFBA que iniciaram em 1969) contava com disciplinas referentes tanto à parte de linguagens formais e autômatos, onde aparecem questões de decidibilidade, quanto aos conteúdos relacionados à enumerabilidade, Tese de Church, Máquina Universal. Mas, da década de 1970 aos dias de hoje, estes conhecimentos vêm perdendo espaço nos currículos da computação, e são praticamente inexistentes nos cursos de Sistemas de Informação, um campo que toma como periférica a importância destes saberes na formação profissional.

Ao mesmo tempo que verificamos que a abordagem usual da computabilidade é extremamente matematizada, centrada na apresentação de resultados e provas formais, e portanto, se afasta do enfoque sociotécnico buscado nos cursos de Sistemas de Informação, vemos também que a ausência destes assuntos causa um lapso nesta formação no que tange a compreensão das possibilidades de sistemas computacionais, a configuração atual de sistemas computacionais e suas perspectivas futuras e também a compreensão desenvolvimento da computação e do próprio computador. Acrescentamos ainda que a presença ubíqua da computação em nossas vidas vem demandando uma compreensão em termos teóricos (no suporte matemático), filosóficos e sociológicos que capacite os profissionais da computação a lidar com a nova dinâmica social fortemente amparada nas tecnologias. Por isso, convém refletir sobre a concepção de uma maneira sociotécnica de trabalhar estes conhecimentos. Longe de pretender desenvolver uma abordagem “suavizada” da abordagem formal, o que está aqui em questão é o resgate destes conteúdos devido à sua importância nas formações em computação. A estratégia é desfazer a cisão entre os resultados matemáticos e as questões do mundo que lhes motivaram, buscar maneiras de apresentar conteúdos formais das disciplinas de Computabilidade inseridos na conjuntura que suscitou a sua formulação para que sejam compreendidos mais amplamente. Por exemplo: os números transfinitos de Cantor, suas propostas de compreensão do infinito, e suas percepções

sobre cardinalidades nos parecem ainda hoje esquisitos porque fugiam do que era considerado “intuitivo”. No entanto, as formulações de Cantor não buscavam refletir diretamente o que a grande maioria considerava “intuitivo”, mas eram inspiradas nos dilemas religiosos de caráter muito intimista que Cantor vivia (Junior, 2006). Portanto, sob a luz destas inquietações as construções matemáticas de Cantor se deixam explicar mais claramente. Da mesma forma, se percebermos os movimentos artísticos que se desenrolavam na Europa dos anos 1930, que tem uma das primeiras expressões na revolta do dramaturgo Antonin Artaud com relação à compreensão da potência da representação (Cafezeiro, Gadelha & Chaitin, 2016), podemos verificar que as indagações formuladas pelos matemáticos com relação à representação matemática acompanhavam uma tendência daquela sociedade naquele momento. A Europa, em consequência da guerra recente, do autoritarismo dos governos e na iminência de uma nova guerra vivia um momento de crise nas representações, incluindo-se aí as representações políticas. Isto, embora seja claro em estudos sociológicos, artísticos e filosóficos, para os matemáticos é um ponto que ainda necessita ser debatido. Alguns cientistas matemáticos preferem considerar os resultados matemáticos como questões livres de qualquer influência do mundo em que viviam os matemáticos. Seriam questões limitadas ao escopo da própria disciplina: “Muitas referências ao teorema da incompletude fora do campo da lógica formal são obviamente absurdas e parecem basear-se em mal-entendidos graves ou algum processo de livre associação” (Franzen, 2005). Por fim, cabe mencionar a própria construção do computador, cuja compreensão não se dissocia da conjuntura da segunda guerra (Dyson, 2012). Quando acompanhados das circunstâncias complexas em que foram concebidos, aquilo que vem sendo usualmente apresentado como sendo “essencialmente matemático” ou “essencialmente técnico” passa a mostrar um caráter inevitavelmente interdisciplinar. Por esse caminho, abre-se espaço a novas traduções, interpretações e conexões com questões da contemporaneidade. Segue-se um exemplo.

Em meados do ano de 2013 começaram a aparecer na mídia carioca diversas notícias referentes a um novo campo da computação denominado “Computação Humana” (Setti, 2013). As notícias enfatizavam o trabalho do cientista guatemalteco Luis von Ahn, que concebeu as *captchas* (Ahn et al., 2004) e as *recaptchas* (Ahn et al., 2008), e consta como um pioneiro da Computação Humana. O termo se refere a sistemas computacionais cada vez mais populares entre nós, que se utilizam do processamento humano para efetuar pequenas tarefas componentes de um objetivo maior (Law & Ahn, 2011). O humano é convocado a participar em situações que a máquina não seria capaz de resolver sozinha por serem incomputáveis ou computacionalmente inviáveis. Esta situação nos remete ao pioneirismo de Alan Turing, ao conceber em 1936, o modelo formal do computador. Poucos anos mais tarde, em 1938, em sua tese de doutorado (Turing, 1938), Turing vinha trabalhando numa hierarquia de sistemas formais. Partindo de um sistema (incompleto) ele descreveu um processo de “completação” de sistemas inicialmente sugerido por Gödel em seu artigo de 1931 (Gödel, 1965). O processo consiste em gerar um novo sistema um pouco mais completo³ que seu precedente pela introdução de um enunciado que o fazia incompleto.

³ Aqui, os termos “completo” e “incompleto” referem-se à sistemas formais: um sistema formal é completo se a qualquer enunciado formulado no sistema corresponda uma prova de sua veracidade ou falsidade. Remetendo-se aos Teoremas da Incompletude de Gödel, Turing argumentou que embora cada

Isso demandaria alguma forma de decidir a veracidade ou falsidade do enunciado em questão. Turing propôs um oráculo: alguma maneira não-maquínica que seria capaz de fornecer o passo não computável de modo a permitir o prosseguimento da tarefa de completação. Em sua tese de doutorado, Turing propôs: “algum meio não especificado de resolver problemas de teoria dos números; como se fosse uma espécie de oráculo. Não vamos mais longe na natureza deste oráculo além de dizer que ele não pode ser uma máquina” (Turing, 1938, p.18). Hoje, passados quase 80 anos, vemos a Computação Humana materializar a sugestão de Turing: o pequeno passo que ele delegou ao oráculo é o que hoje se delega ao humano. Como no caso da Computação Humana, é possível que muito do que vemos hoje na área de sistemas de informação, coisas que nos parecem essencialmente inovadoras, encontrem bases de fundamentação nos trabalhos matemáticos de um momento em que se dedicou grandes esforços criativos nos estudos relativos aos sistemas formais. Em outras palavras, fundamentos conceituais para o paradigma do século XXI denominado “computação humana” podem ser encontrados nas propostas de Alan Turing, em 1938. Assim, pode ser bastante interessante voltarmos os olhos estes estudos de modo a buscar uma melhor compreensão do nosso tempo presente.

Propomos aqui um enfoque sociotécnico que consiste em deixar aparente o percurso de construção dos conteúdos. Com isso, abrimos mão de uma concepção em que os resultados buscam explicar-se por si só, e, como bem colocou Cukierman et al., possibilitamos a verificação de que o que se diz “técnico” se conforma em co-construção com o que se diz “social”.

3. Modos contemporâneos de compreensão e construção do conhecimento

A questão da retenção comentada na seção 1 junta-se com a necessidade de buscar uma dinâmica em sala de aula que melhor se adapte aos modos contemporâneo de comunicação e entendimento. É possível que o recorte disciplinar em que os conteúdos estão tradicionalmente organizados já não mais dê conta de sensibilizar o estudante. As respostas aos anseios do nosso tempo não são encontradas nas disciplinas isoladas, mas no encontro entre elas, na conjunção de saberes, e por isso, o estudante está constantemente buscando, pesquisando em seus dispositivos móveis, questões relacionadas (ou não) ao que se fala em aula. Muitos dos conteúdos tradicionais já estão disponíveis em video-aulas na Internet, a qualquer tempo, em qualquer lugar e isto põe em cheque o papel do professor como enunciador de conteúdo. Os estudantes reivindicam diálogo, construções colaborativas, descentramento, e é neste espaço que o professor precisa se reposicionar para fazer circular a sua mensagem, não mais como centro catalizador, mas como parte da rede. Além disso, é preciso acompanhar o diálogo, e no fluxo dos interesses de cada turma, introduzir as questões que se quer comunicar. Por tudo isso sequência fixada de uma ementa ou programa passa a adquirir um papel orientador, apresentando questões, e não determinando o andamento das aulas.

Presenciamos hoje na sala de aula uma multiplicidade de recursos que disputam a atenção do aluno. Entretanto, o estudante parece mobilizar-se quando tem consciência

sistema de sua hierarquia fosse um pouco mais completo que o precedente, a hierarquia jamais alcançaria a completude.

da importância de um determinado assunto na sua formação profissional. Por isso, os conteúdos programáticos precisam ser justificados. O professor que ensina precisa saber para quem e para o quê ensina, e isto invalida o modelo de “aulões”, “provões” e disciplinas pré-formatadas, que colocam em primeiro plano o conteúdo e não o aluno. Invalida também a maneira como grande parte das disciplinas de conteúdo “exato” são ministradas. Uma ciência é considerada “exata” se seus conteúdos já estão de tal forma estabilizados que não admitem questionamentos. São fórmulas, cálculos, que podem ser apresentados de modo a serem diretamente aplicados. Este conceito de “ciências exatas” amplifica o conflito entre um modo estático de transmissão, e a expectativa nos dias de hoje de uma construção em fluxo.

Problematizar a exatidão e rigor, ou seja, deixar aparente o processo de construção (a historicidade) dos conhecimentos quebra este conflito na medida em que coloca à vista os problemas e os embates que motivaram aquelas formulações. Destitui a imponência de um conhecimento que dispensa justificativas, abrindo possibilidades de traduções às situações inesperadas do tempo presente. Desta maneira torna-se possível ao professor situar a importância do assunto em questão à formação pretendida. No caso dos cursos de formação em computação, já foi dito, mas convém repetir que conhecer o processo de construção do computador e da computação ajuda a compreender o mundo em que vivemos hoje, onde a computação vem ocupando um espaço cada vez mais amplo, e assim verificar possibilidades futuras e conceber inovações.

Teoria da Computação, uma abordagem sociotécnica.

Durante este ano de 2016, o Curso de Sistemas de Informação da Universidade Federal Fluminense está passando por uma reforma curricular que visa, entre outros objetivos, minimizar a retenção nos primeiros períodos e adaptar os conteúdos às demandas locais. Dentre diversas mudanças elaboradas pela coordenação, em conjunto com o Núcleo Docente Estruturante, o Colegiado de Curso, e considerando as direções apontadas pelos estudantes, citamos a reorganização dos conteúdos de Matemática Discreta e Lógica Formal, em uma disciplina denominada Fundamentos Matemáticos para Computação, a ser ministrada no primeiro período do ingressante, concomitante com Teoria da Computação para Sistemas de Informação, outra disciplina recém-criada. Com base nas justificativas já comentadas nas seções anteriores, apresentamos o conteúdo programático de Teoria da Computação para Sistemas de Informação. A disciplina organiza-se em torno da linha do tempo, porém seu objetivo não é fazer um relato da sequência temporal purificada dos acontecimentos, mas trabalhar as construções matemáticas ao longo dessa diretriz. Ou seja, pretende-se uma abordagem sociotécnica, em que a matemática explica-se e explica a conjuntura em que foi concebida, e vice-versa. A elaboração desta proposta não teria sido possível sem as discussões realizadas no âmbito do Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia, onde as questões matemáticas são abordadas sob pontos de vistas diversos.

Objetivo: Percorrer a história do desenvolvimento matemático no decorrer do século XX, no que diz respeito às discussões sobre a fundamentação da matemática. Apresentar os conceitos fundamentais da Teoria da Computação imbricados nas situações que motivaram a sua formulação. Abordar não somente as maneiras como as construções teóricas deram suporte ao surgimento da tecnologia (computadores), mas também como as tecnologias deram suporte à

novas construções teóricas. A partir da compreensão da construção histórica dos fundamentos da computação buscar a compreensão de situações contemporâneas envolvendo sistemas computacionais.

Embora o foco da disciplina esteja no decorrer do século XX, o curso inicia ambientando a concepção da ciência ao final do século XIX, para evidenciar como as questões emergentes no campo da matemática colocaram em cheque os anseios da ciência moderna que se queria completa e totalizante. O curso avança pelo século XXI para abranger questões da contemporaneidade em uma proposta de conteúdo que não pode ser completamente fixada. Embora os títulos se refiram a eventos históricos, pretende-se trabalhar em detalhe o conteúdo matemático relacionado a cada evento. Ressaltamos entre parênteses o conteúdo matemático a ser abordado, com a ressalva de que há interseções com a disciplina Fundamentos Matemáticos para Computação.

Conteúdo Programático:

Do século XIX ao século XX: As questões que intrigavam os matemáticos ao final do século XIX e os embates com a Ciência Moderna (aritmética da análise, axiomatização da aritmética, formalização da geometria. Conjuntos indutivos e recursão primitiva a partir dos axiomas de Peano). Georg Cantor e os infinitos (coleções infinitas: enumerabilidade, cardinalidade, provas por diagonalização).

O século XX: Russell e o paradoxo (conjuntos, elementos, pertinência, conjuntos das partes). O nascimento dos programas de fundamentação da matemática: O programa Logicista, O programa Intuicionista (comparações com a lógica clássica), O programa Formalista (revisitar funções primitivas recursivas. Abordar também as parciais recursivas. Completude, Consistência, Decidibilidade de sistemas formais). O Teoremas de Gödel (numeração de Gödel, a prova da incompletude, contrapontos ao programa de Hilbert). Compreensões da matemática sob a percepção da incompletude. Propostas de formalização do conceito de computável. O cálculo λ (Church/Kleene). As funções recursivas gerais (Gödel). A máquina de Turing como um programa (construção da máquina de turing). A máquina de Turing Universal (contraponto ao programa de Hilbert). Tese de Church. Evidências para a Tese de Church (equivalências entre os formalismos abordados). Compreensões filosóficas sob a percepção da incompletude. O Oráculo de Turing e a necessidade de operar com indecidíveis. A dicotomia homem-máquina e o nascimento da IA. O embate entre IA forte e IA fraca e a concepção de híbridos. Computadores, tecnologias e Segunda Guerra. Dilemas de Einstein, Russell, e Wiener. John von Neumann e a bomba atômica. John von Neumann e a implementação da máquina universal. A computação no pós-Guerra. O nascimento das CC, as Conferências da OTAN. O não-computável e os computadores. Número Ω (a construção do número Ω). A situação do Brasil no cenário da computação na década de 1970. Representação computacional dos números reais. As controvérsias sobre as abordagens matematizadas e os processos sociais.

O século XXI: A teoria da computação e o suporte às novas tendências da computação: Processos colaborativos, Abordagens sociotécnicas, Computação humana.

A construção desta proposta vem ocorrendo ao longo de três períodos consecutivos em que foi oferecida como disciplina optativa. Utilizou-se como bibliografia básica, os livros textos usuais da área para cobrir o conteúdo formal com uma notação atual. Além disso, artigos da época disponíveis na internet (cartas, discursos, etc) auxiliaram não propriamente no conteúdo formal, mas para apresentar pequenos trechos onde esses autores expõem motivações e intuições das suas propostas. Trechos de filmes e documentários também contribuíram para trazer um pouco do ambiente em que se vivia a cada época. Fotografias, cartas e documentos do arquivo digital de Turing, e por fim, alguns romances (Doxiadis, 2001, 2010) também contribuíram para ilustrar os dilemas a época.

Referências

- Ahn, L.; Blum, M.; Langfort, J. (2004) “Telling Human and Computers apart” Em: Comm. of the ACM. Vol. 47, No. 2, 5860. 2004.
- Ahn, L.; Maurer, B.; Mcmillan, C.; Abrahm, D.; Blum, M. (2008) “reCAPTCHA: Human Based Character Recognition via Web Security Measure”. Science v321.
- Cafezeiro, I.; Haeusler, H.; Marques, I.; Cukierman, H. (2010) “Recontando a Computabilidade”. Em: Revista Brasileira de História das Ciências, Rio de Janeiro, v. 3, n. 2, p. 231-251.
- Cukierman, H.L.; Teixeira, C.; Prikladnicki, R. (2007) “Um olhar sociotécnico sobre a engenharia de software”. Em: Rev. de Inf. Teórica e Aplicada, RS, UFRGS.
- Curso de Qualidade (2015) “Evasão e estratégias para permanência estudantil” In: XXXV CSBC. Bento Gonçalves – RS.
- Deleuze, G. & Guatarri, F. (1997) “Mil Platôs”, vol 5. São Paulo: Editora34.
- Dyson, G. Turing’s Cathedral: The Origins of the Digital Universe. Vintage Books.
- Doxiadis, A. (2001) “Tio Petrus e a conjectura de Goldbach”. São Paulo: Editora 34.
- Doxiadis, A., Papadimitirus, C. (2010) Logicomix, uma jornada épica em busca da verdade. Martins Fontes.
- Franzen, T., (2005) “Gödel’s Theorem: An incomplete guide to its use and abuse”. Wellesley, Mass.: A K Peters.
- Gödel, K. (1965) On formally undecidable propositions of Principia Mathematica and related systems. Em M. Davis (Ed), The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions. NY: Dover Pubs.
- Junior, W. G. N. (2006) “O Infinito contado por Deus. Uma interpretação dedekindiana do conceito de número ordinal transfinito de Cantor”. Tese Dout. Fil., PUC, RJ.
- Latour, B. (2014) “Scientific Humanities, Introduction”. <http://www.bruno-latour.fr/courses>.
- Law, E. Ahn, L. (2011) Human Computation. Morgan & Claypool Publishers.
- NATO (1969,1970) “Software Engineering: Report of a conference sponsored by the NATO Science Committee”, Scientific Affairs Division.
- Setti, R. (2013) “Após digitalizar livros combatendo spams criador do duolingo quer traduzir web ensinando línguas”. O Globo, 04/06/2013
- SBC (1999) “Currículo de Referência da SBC para Cursos de Graduação em Computação”. Sociedade Brasileira de Computação. Bento Gonçalves – RS.
- Souza, O. S., Morais, P. S., Junior, F. C. S., (2015) “Reprovações e Trancamentos nas Disciplinas de Introdução à Programação da Universidade de São Paulo: Um Estudo Preliminar”, In: XXIII WEI – XXXV CSBC. Bento Gonçalves – RS.
- Turing, A. (1936) On computable numbers, with an application to the Entscheidungsproblem, Proceedings of the London Mathematical Society, Series 2, n.42, p 230-265.
- Turing, A. (1938) Systems of Logic Based on Ordinals (Doctoral thesis, Princeton University).

O que Explica o Sucesso de Jogos no Ensino de Engenharia de Software? Uma Teoria de Motivação

Márcia Souza, César França

Mestrado Profissional em Engenharia de Software
CESAR.EDU – Recife – PE – Brazil

Universidade Federal do Rural de Pernambuco (UFRPE)
Recife – PE – Brazil

marciavr.souza@gmail.com, cesar@franssa.com

***Abstract.** The development of games for teaching software engineering disciplines has proven to be a quite successful instructional alternative. It is known, however, that the attractiveness of the games is highly associated with sophisticated social and psychological mechanisms. This article presents a systematic review conducted in 2008 WEI Annals to 2015, which shows that there is a significant convergence between the features of educational games in this literature and the Theory of Motivation and Satisfaction of Software Engineers. Thus, this theory can serve as a basic mechanism for the conceptual design of even more attractive educational games in this area.*

***Resumo.** O desenvolvimento de Jogos para ensino das disciplinas de engenharia de software tem se mostrado como uma alternativa instrucional de bastante sucesso. É sabido, no entanto, que a atratividade dos jogos está associada a mecanismos sociais e psicológicos altamente sofisticados. Neste artigo, apresentamos uma revisão sistemática conduzida nos anais do WEI de 2008 a 2015, que demonstra que há uma convergência significativa entre as características dos jogos educacionais propostos nesta literatura e os pressupostos da Teoria de Motivação e Satisfação de Engenheiros de Software. Isto posto, esta teoria poderá servir como mecanismo de base para o design conceitual de jogos educacionais ainda mais atrativos nesta área.*

1. Introdução

Para garantir a formação de um bom profissional na área tecnológica, bem como um engenheiro de software, é preciso que durante a graduação ele tenha uma boa base de conhecimentos básicos, desde programação até processos de desenvolvimento de software. A finalidade é que futuramente no mercado ele possa ser capaz de atuar profissionalmente desenvolvendo com qualidade produtos de softwares.

Os professores, das disciplinas dos cursos na área de tecnologia, são desafiados constantemente a instigar os estudantes para que eles possam prosseguir de maneira positiva no curso, dando-lhes uma boa base de conhecimento aliada a motivação, onde segundo Silva et al. (2011) pode estar relacionado à carência de práticas educacionais alternativas às tradicionais. Os alunos esperam dos professores não apenas a teoria,

porém o maior desafio é na utilização de uma metodologia na sala de aula que possa favorecer um maior número de alunos e conseqüentemente o professor.

Uma das propostas de melhoria do aprendizado em sala de aula são os jogos educacionais. Na literatura encontram-se vários trabalhos que demonstram profissionais de educação utilizando-os como ferramenta de auxílio ao aprendizado. Dentre as vantagens de utilizar jogos educacionais, podemos destacar a motivação e o aprendizado vivencial. Os jogos permitem que os alunos possam vivenciar experiências teóricas e práticas. No entanto, trabalhos recentes têm apontado que a atratividade dos jogos advém fundamentalmente da sua capacidade de simulação de fatores psicossociais relevantes para a motivação humana [Hamari, 2013][Chou, 2014].

Nesta pesquisa, nos propusemos então a revisar a literatura sobre jogos educacionais das disciplinas dos cursos de computação, e avaliar se a presença/ausência dos fatores previstos pela Teoria de Motivação e Satisfação de Engenheiros de Software (TMS-ES) por França (2014) estão associados ao sucesso reportado do jogo. A TMS-SE é um trabalho que envolve as principais teorias de motivação do campo da psicologia desenvolvidas no último século adaptando-os para a realidade específica do trabalho na engenharia de software. A TMS-SE foi desenvolvida no campo do comportamento organizacional, mas há razões para acreditar que ela possa ter um impacto significativo na área da educação [Richter, 2014]. Neste artigo, apresentamos uma análise preliminar resultantes da literatura específica do Workshop de Educação em Computação (WEI) sob este conjunto de características da TMS-SE.

Este artigo está organizado da seguinte forma: será explicado sobre a importância da engenharia de software e dos jogos educacionais e serão abordados os elementos motivacionais que servirão de base para a condução desta pesquisa. Na terceira seção será explicada a metodologia que foi conduzida para este trabalho. Na quarta seção serão exibidos os resultados desta pesquisa e as discussões dos resultados. Por fim, na quinta seção, serão explicadas a conclusão desta pesquisa.

2. Referencial Teórico

2.1. Desafios no ensino da Engenharia de Software

Engenharia de software é uma das áreas mais importantes da Tecnologia da Informação, sendo especialmente central para a formação adequada de Engenheiros de Software [Sommerville 2011]. Tem como objetivo apoiar o desenvolvimento profissional de software, cobrindo todos os aspectos da produção de um software. Monsalve et al. (2010), comentam que a engenharia de software é uma área com aspectos teóricos e práticos, tornando fundamentais as experiências e decisões da prática no desenvolvimento de softwares de qualidade, com economia, utilidade e no prazo esperado. Porém, Benitti e Molléri (2008) concordam que a engenharia de software é uma área muito jovem e sofre contínuas mudanças nos seus fundamentos tecnológicos concretizadas nos métodos e ferramentas de suporte, portanto necessita de métodos de ensino lúdicos e dinâmicos que possam contribuir na aprendizagem do estudante.

Contudo, a forma como algumas disciplinas de engenharia de software comumente são ensinadas para os alunos não permite que a prática seja trabalhada podendo gerar insatisfação e conseqüentemente profissionais despreparados com as

técnicas utilizadas no mercado de trabalho. Andrade et al. (2008) em seu trabalho reforça a necessidade da relação pedagógica entre professor e aluno em uma estrutura metodológica e planejada com teorias modernas de aprendizagem. Como a finalidade na sala de aula é estimular ideias dos alunos, ensinar apenas com aulas expositivas tradicionais pode dificultar o aprendizado.

Uma das propostas metodológicas para ensino de engenharia de software e suas disciplinas, são os jogos educacionais. Sabe-se que os jogos educacionais, segundo Nunes e Parreira (2015), têm sido intensamente utilizados por profissionais da área de educação como auxílio para a construção do conhecimento. Em sua pesquisa, Fukusawa et al. (2015) apontam alguns dos benefícios que os jogos educacionais podem trazer ao processo de ensino e aprendizagem como, por exemplo, a motivação e o aprendizado por descoberta. Portanto, os jogos podem proporcionar a vivência em experiências de aprendizagem concretas [Monsalve et al. 2010].

2.2. Jogos para auxílio nas disciplinas de Engenharia de Software

Na literatura é possível encontrar inúmeros trabalhos utilizando jogos como auxílio ao aprendizado nas disciplinas dos cursos de engenharia de software. Os autores Rangel et al. (2015) observaram que utilizar jogos na educação proporcionam aos estudantes e profissionais um ambiente interativo e dinâmico provocando a motivação para avançar em um enredo, por meio de simulações e situações relacionadas ao tema abordado. Silva et al. (2015) comentam que uma abordagem alternativa às aulas tradicionais, devido elas serem mais teóricas e expositivas, é a utilização dos jogos, pois esta abordagem preza por uma teoria de motivação humana como ponto de partida.

Pode-se observar que para alcançar algum sucesso no que é proposto, além de atender aos objetivos educacionais, os jogos devem principalmente transmitir motivação e satisfação aos estudantes. Segundo Vahldick et al. (2015), se estes objetivos não forem alcançados, o jogador ficará aborrecido, desmotivado e desconcentrado, não atingindo os objetivos da aprendizagem. No entanto, a motivação é um elemento quase que acidental resultante do processo de design de jogos *function-focused*, uma vez que fatores motivacionais não são necessariamente conhecidos ou levados em conta no início deste processo [Chou, 2014]. Chou (2014) aponta que a abordagem *human-focused* é capaz de garantir o sucesso do jogo no aspecto motivacional, pois esta abordagem adota uma teoria de motivação humana como ponto de partida.

2.3. Teoria da Motivação e Satisfação dos Engenheiros de Software

De acordo com França (2014), o sucesso dos projetos de software depende do gerenciamento apropriado da motivação e satisfação no trabalho. O estudo realizado pelo autor, investigou um conjunto de características presentes no ambiente de trabalho dos engenheiros de software que podem influenciar positivamente ou negativamente seu desempenho. Relevantes teorias advindas do campo da psicologia, tais como a Job Satisfaction Theory por Locke (1969) e Job Characteristics Theory por Hackman (1980) foram então combinadas e adaptadas a uma nova teoria explanatória da motivação e satisfação no trabalho de engenheiros de software (TMS-SE), a qual tem como objetivo esclarecer as características do trabalho que influenciam a motivação dos engenheiros de software.

Segundo esta teoria, a motivação atua como um elemento moderador na transformação do desempenho potencial de um engenheiro, em seu desempenho real. De forma concreta, a motivação do engenheiro é representada por uma combinação de engajamento e concentração na execução de uma tarefa, e as características específicas da tarefa, como percebidas pelos engenheiros de software, condicionam este comportamento. As características influentes seriam então: a *utilidade do conhecimento* gerado, a demanda por *criatividade*, o *impacto social*, a *variedade da tarefa*, uma *definição clara* do que é esperado, a *autoconfiança* técnica e o *engajamento de seus pares*. Por se tratar de uma atividade fundamentalmente intelectual, a *sobrecarga cognitiva* atuaria de forma negativa sobre a motivação destes profissionais.

A satisfação, por sua vez, é representada concretamente pela felicidade resultante do julgamento do engenheiro sobre os resultados atingidos pelo seu desempenho dadas as condições conjunturais (físicas, organizacionais e sociais) nas quais a tarefa foi conduzida. O principal resultado da satisfação é o reforço (positivo ou negativo) sobre o desempenho potencial do engenheiro de software para execução de outras tarefas no futuro.

O uso desta teoria como fundamento para o design de jogos educacionais poderia ser útil, uma vez que a teoria permite orientações, como por exemplo:

- (i) A TMS-ES afirma que a forma mais simples de melhorar a confiança técnica de um engenheiro de software é proporcionando oportunidades de experimentar e aprender novas tecnologias, ferramentas e métodos antes de adotá-las. A partir desta teoria, os jogos que constroem a autoconfiança técnica dos alunos tem mais chance de serem úteis em sala de aula.
- (ii) Com a tecnologia em constante evolução, França (2014) observou que os engenheiros de software necessitam de aprendizado constante e a utilidade deste aprendizado podem refletir na qualidade da sua produção, desempenho, ou até mesmo aumentar a sua oportunidade de encontrar outros empregos. O aluno, na sala de aula, ao perceber a utilidade daquilo que está aprendendo, pode refletir também no seu desempenho.
- (iii) A criatividade, segundo a TMS-ES, é uma característica intrínseca do processo produtivo dos engenheiros de software e é composto por uma série de atividades criativas ou não. Há uma maior concentração na parte criativa quando se refere a suas tarefas favoritas e tendem a falar quando não gostam das tarefas. O desafio seria manter a motivação na execução de parte que não são criativas. Na sala de aula, estimular a criatividade do aluno é um grande desafio e dependendo do jogo, é possível estimular a criatividade do aluno para certas tarefas.

3. Metodologia

Foram pesquisados nos Anais da base WEI e selecionados os artigos que tivessem trabalhos com jogos que auxiliassem no aprendizado de ensino superior das disciplinas de Engenharia de Software, preferencialmente os que validassem com alunos. Ficaram de fora os artigos que utilizaram o Scratch, ainda que o produto final fosse um jogo, é considerado um ambiente de programação fácil que não requer conhecimentos prévios

de algoritmos ou de programação para o seu manuseio [Cristovão, 2008]. O período das publicações foi entre 2008 a 2015 nos anais disponíveis de forma eletrônica na internet¹.

O propósito dos jogos estudados neste trabalho estão compreendidos entre as disciplinas básicas para todos os cursos e as específicas do curso de Engenharia de Software. Portanto os objetivos educacionais dos jogos selecionados deste trabalho fazem parte das Diretrizes Curriculares (2003) dos cursos Engenharia de Software.

Foram incluídos na pesquisa os artigos que apresentavam no conteúdo um relato da experiência do uso de algum jogo educacional com os alunos dos cursos de engenharia de software. Inclusive os que não tomavam como base uma validação necessariamente experimental e formal. Consideramos igualmente os jogos independentemente da plataforma digital ou analógico. Descartamos os trabalhos com alunos do ensino médio e fundamental e incluímos dois trabalhos do curso técnico, sendo um deles para ensino à distância. A seleção inicial considerou os artigos que explicitavam no título ou no resumo que se tratavam de ferramentas em geral que apoiassem o ensino de disciplinas de engenharia de software. Em um segundo momento, foi efetuada uma prévia leitura da metodologia dos trabalhos para averiguar se eram jogos. Artigos sobre ferramentas de apoio à execução de tarefas e/ou processos que não se tratavam de jogos foram então removidos. A lista final dos 20 trabalhos selecionados pode ser encontrada no Apêndice deste.

Estes resultados ainda são passíveis de falhas, uma vez que as avaliações dos artigos foram conduzidas por apenas um pesquisador, mesmo que com o apoio do seu orientador. Representando, portanto, uma ameaça à validade deste trabalho.

Apoiado na TMS-SE, verificamos a presença ou ausência dos fatores motivacionais em cada um dos jogos relatados nos trabalhos, com uma leitura mais aprofundada na metodologia, resultados e discussões. Foi construído então uma tabela com os artigos e em cada uma das características apresentadas acima, com exceção do objeto de aprendizagem, colocamos os valores SIM para os trabalhos que confirmassem a característica, NÃO para os que não confirmassem a características e ‘?’ para os que não apresentaram o resultado para a característica. Conservamos nesta pesquisa os trabalhos [11] e [20], apesar de utilizarem a mesma ferramenta, CyberCIEGE, as abordagens na sala de aula foram diferentes. Nesta tabela, também anotamos os objetivos da aprendizagem, referindo-se ao conteúdo da disciplina que o jogo se propõe a auxiliar. Sendo assim, os elementos analisados foram:

- a. *Objeto da aprendizagem.* Apenas os trabalhos que tratassem de jogos voltados para o ensino de disciplinas dos cursos de engenharia de software;

Elementos Sociais e Individuais

- b. *Engajamento dos colegas;*

No que diz respeito ao engajamento, segundo a TMS-SE um engenheiro motivado é um engenheiro engajado, conforme observado no jogo SE-RPG [1], em seu trabalho, o jogo foi executado em equipe contribuindo para engajamento entre os colegas. Já no trabalho [8], o iTestLearning é um jogo single-player

¹ <http://www.lbd.dcc.ufmg.br/colecoes/wei/>

(jogo para somente um jogador) onde é realizado um planejamento de teste de software a partir da especificação de projeto.

c. Construção da autoconfiança técnica do aluno;

Sobre a autoconfiança técnica do aluno, [2] em seu trabalho constatou que foi através do jogo que os alunos puderam lembrar e ressignificar conteúdos importantes para as disciplinas seguintes contribuindo para a construção da autoconfiança dos alunos. Já no trabalho de [5], constataram nas avaliações que o jogo não fez diferença em relação a aprendizagem quando comparado com uma aula prática.

Características da Tarefa

d. O aluno compreende o impacto social daquilo que está aprendendo;

Em relação a compreender o impacto social do que está aprendendo, [5] relataram que os alunos acreditaram que o jogo aliado com as aulas práticas os ajudaram a aprender o conteúdo. Nos demais trabalhos esta característica não estava declarada.

e. O aluno percebe a utilidade daquilo que está aprendendo;

No que diz respeito à percepção da utilidade daquilo que está aprendendo, os alunos deste trabalho [10] perceberam a utilidade do ambiente de investimentos entre equipes que o jogo proporcionou. Já [13] apontaram alguns problemas na adoção do jogo, como por exemplo a desistência de alguns alunos no desafio proposto.

f. Trata-se de uma quebra de rotina;

Considerando que a ferramenta de auxílio ao aprendizado de algumas disciplinas de engenharia de software trata-se de um jogo, pode-se afirmar que é uma quebra de rotina às aulas tradicionais. Todos os artigos concordaram que o jogo tratava-se de uma quebra de rotina. Em seu trabalho, [18] comentaram que os discentes se mostraram receptivos às inovações e explicaram que esse tipo de iniciativa é louvável frente a disciplinas teóricas, que demandam o estudo de conteúdos e que, em aulas expositivas, é feito de modo cansativo.

g. Explora a criatividade do aluno;

Dentre os jogos que exploram a criatividade do aluno, [15] afirmaram que uma das vantagens experimentais está a possibilidade de simular os cenários criados pelo professor, contribuindo para aguçar a criatividade. A maioria dos trabalhos apresentaram jogos com estratégias que exploram a criatividade do aluno. Poucos trabalhos não deixaram explícito no texto esta característica.

h. Possui tarefas bem definidas;

Com relação à definição das tarefas, [12] dividiu seu jogo em duas fases em que o aluno deverá jogar 16 missões e verificou a imersão dos alunos e com o conteúdo dentro dos objetivos da aprendizagem.

Características Ambientais

i. Possui a carga cognitiva desbalanceada negativamente.

No trabalho [09], quando perguntado aos estudantes se utilizar o jogo simultaneamente ao estudo da disciplina facilitaria o processo de ensino-aprendizagem da disciplina, 100% responderam positivamente e alegaram que o jogo teria ajudado bastante principalmente nos seus estudos para realização das avaliações. Todos os trabalhos apresentaram carga cognitiva balanceada, com desafios relacionados ao conteúdo de sala de aula.

Com essas informações podemos avaliar se estes elementos estão presentes nos trabalhos selecionados e principalmente se o sucesso do uso desses jogos estão relacionados com a presença ou não desse conjunto de elementos.

4. Resultados e Discussão

Obsevou-se que de 2008 a 2011, poucos trabalhos envolveram jogos e que tivessem validação. Nenhum trabalho do ano de 2009 foi selecionado. A partir do ano de 2013 houve muito mais trabalhos com os critérios de pesquisa que especificamos no capítulo anterior. Na Tabela 1, com relação aos objetos de aprendizagem, entre os artigos selecionados observamos um número maior de jogos de aprendizagem de algoritmos. Em seguida estão os jogos voltados para auxílio de programação. Na tabela 2, encontram-se as respostas da avaliação de cada artigo (de 1 à 20), para cada fator motivacional (de *b* à *i*). No Apêndice estão numerados os trabalhos selecionados para esta pesquisa e para as análises utilizaremos os índices das características de A a I descritos no capítulo anterior.

Tabela 1. Análise dos artigos sobre o Objeto da Aprendizagem

Objeto da Aprendizagem	Ref.*	Objeto da Aprendizagem	Ref.*
Processo de desenvolvimento de software	[1]	Segurança em Redes de Computadores	[11 e 20]
Gestão de Projetos	[1, 15]	Teoria da Computação	[9]
Algoritmos	[2, 3 e 4]	Empreendedorismo	[10]
Estrutura de Dados	[3, 14]	Lógica de Programação	[12]
Engenharia de Requisitos	[5]	Teste de Software	[8]
Programação	[7 e 19]	Circuitos Digitais	[13]
Programação Orientada a Objetos	[6, 17]	Métodos de Ordenação	[16]
Tipos de Variáveis	[7]	Sistema da Informação	[18]

**artigos referenciados no Apendice A*

Nesta pesquisa, encontramos 9 dos 20 trabalhos que afirmaram ter observado o aumento da autoconfiança técnica dos alunos. Foram observados 19 trabalhos em que a prática dos jogos não provocou a sobrecarga cognitiva dos alunos. Apenas um trabalho não continha essa informação explícita.

Observamos que dos 20 artigos, 8 confirmaram a presença da percepção do impacto social do que estavam aprendendo. Na pesquisa, 10 trabalhos afirmam que os alunos percebem a utilidade do que estão aprendendo e 18 trabalhos permitem explorar a criatividade dos alunos na sala de aula. 19 artigos apresentaram jogos com tarefas bem definidas.

Tabela 2. Análise dos Fatores Motivacionais

Artigo / Fator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total
----------------	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	-------

b.	✓	✓	✓	?	✓	✓	✓	×	×	✓	×	×	×	×	✓	✓	✓	✓	✓	12/20	
c.	✓	✓	×	?	×	?	?	✓	✓	✓	?	?	?	✓	✓	?	?	?	✓	✓	9/20
d.	✓	✓	?	?	✓	?	?	✓	?	✓	?	?	?	✓	?	✓	?	?	?	✓	8/20
e.	✓	✓	?	?	?	?	?	✓	?	✓	✓	?	×	✓	✓	?	?	✓	✓	✓	10/20
f.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20/20
g.	✓	✓	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	18/20
h.	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	19/20
i.	×	×	?	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	0/20

✓ - SIM, × - NÃO, ? – Não foi possível inferir

Todas as características descritas são elementos importantes que podem ser utilizados para avaliação de jogos que apoiam o ensino de disciplinas de engenharia de software. Todos os jogos tiveram sucesso em algum aspecto e a maioria dos trabalhos apontaram as falhas e se propuseram a corrigir e/ou repetir o experimento com os alunos. A partir dessas informações podemos afirmar que estes elementos propostos por França (2014) em sua teoria TMS-SE estão largamente presentes nos trabalhos selecionados.

5. Conclusão

Este trabalho efetuou uma avaliação dos jogos para auxílio de aprendizagem de disciplinas de engenharia de software, publicado nos anais do WEI entre os anos de 2008 a 2015, sob o conjunto de características propostas na teoria de motivação e satisfação dos engenheiros de software. Estes conjuntos de características que estão presentes pelo menos parcialmente em todos os jogos pesquisados, podem atestar que uma das principais preocupações ao desenvolver um jogo é se provocará a motivação e a satisfação dos alunos.

É importante notar que este trabalho é um esforço preliminar no sentido de revelar os aspectos motivacionais relevantes para o design de jogos educacionais nas disciplinas de engenharia de software. Como forma de continuar este trabalho, planejamos conduzir uma revisão sistemática mais ampla, expandindo a base de artigos para incluir outras conferência e jornais relevantes nacionais e internacionais, e observar as evidências reportadas sobre a experiência de aprendizado, buscando identificar a sua relação com a TMS-ES. Em seguida, pretendemos avaliar o uso da teoria em um processo real de concepção de um jogo educacional na área de engenharia de software, bem como os efeitos desse uso no potencial motivador do jogo nos alunos.

Referências

Andrade, R. M. de C., Marinho, F. G., Leitão, V. L. e Rocha, L. S. (2008). Uma Proposta de Metodologia para o Ensino de Engenharia de Software. *Fórum de Educação em Engenharia de Software*, n. Mdcc.

Benitti, F. e Molléri, J. (2008). Utilização de um RPG no ensino de gerenciamento e processo de desenvolvimento de software. WEI - Workshop sobre Educação em Computação. (pp. 358-267).

Chou, Yu-kai (2014) Actionable Gamification - Beyond Points, Badges, and Leaderboards. Leanpub.

Cristovão, H. M. (2008). Aprendizagem de Algoritmos num Contexto Significativo e Motivador: Um Relato de Experiência. WEI - Workshop sobre Educação em Computação. (pp.30).

Diretrizes Curriculares Conselho Nacional de Educação e Câmara de Educação Superior (2003). Ministério da Educação. <http://www.sbc.org.br/documentos-da-sbc/send/131-curriculos-de-referencia/761-diretrizes-curriculares-consulta-publica>. Maio, 2016.

França, C. (2014). A Theory of Motivation and Satisfaction of Software Engineers. p. 200.

Fukusawa, J., Carniello, A. e Carniello, A. (2015). JOGOS DIGITAIS NO ENSINO E APRENDIZAGEM DE ENGENHARIA. *Gestão Universitária*, v. 1, p. 1–8.

HACKMAN, J. R. Work redesign and motivation. *Professional Psychology*, v. 11, n. 3, p. 445-455, 1980.

Hamari, J., & Koivisto, J. (2013). Social motivations to use gamification: an empirical study of gamifying exercise. Proceedings of the 21st European Conference on Information Systems, Utrecht, Netherlands, June 5–8, 2013.

LOCKE, E. A. What is job satisfaction? *Organizational Behavior and Human Performance*, p. 309-336, 1969.

Monsalve, E. S., Werneck, V. M. B. e Cesar, J. (2010). SimulES-W : Um Jogo para o Ensino de Engenharia de Software. *Anais do III Fórum de Educação em Engenharia de Software*, p. 17–26.

Nunes, I. F. e Parreira, P. A. (2015). RPG4Sorting - Um Jogo Educacional para Auxílio ao Ensino de Métodos de Ordenação. *XXIII Workshop sobre Educação em Informática*.

Rangel, D., Leite, A. e Delfino, S. S. (2015). Gsprojects - ambiente para simulação da gestão de projetos de software. *23º WEI - Workshop sobre Educação em Computação*.

Richter, G., Raban, D., & Rafaeli, S. (2014). Studying Gamification: The Effect of Rewards and Incentives on Motivation. *Gamification In Education And Business*, 21-46. doi:10.1007/978-3-319-10208-5_2.

Silva, R. J. M., Azevedo, W. L. V, Nascimento, R. P. e Paxiúba, C. M. C. (2015). MercadoSI : Um jogo de tabuleiro para o ensino de Sistemas de Informação. *23º WEI - Workshop sobre Educação em Computação*.

Silva, T. G., Müller, F. M., e Bernardi, G. (2011). Panorama do Ensino de Engenharia de Software em Cursos de Graduação Focado em Teste de Software: Uma Proposta de Aprendizagem Baseada em Jogos. *RENOTE*, 9(2).

Sommerville, I. (2011). *Engenharia de Software*. Pearson/ Addison Wesley. Teixeira, M. R. W., Aguiar, G. D. F. e Ferlin, E. P. (2010). O Jogo Imagem & Ação como Estratégia de Aprendizagem para o Conteúdo de Algoritmos Genéticos. *WEI – XVIII Workshop sobre Educação em Computação*, n. 843, p. 843–846.

Vahldick, A., Mendes, A. J. e Marcelino, M. J. (2015). Testando a Diversão em um Jogo Sério para o Aprendizado Introdutório de Programação. *23º WEI - Workshop sobre Educação em Computação*.

Apêndice

- [1] Benitti, F. B. V., & Molléri, J. S. (2008) Utilização de um RPG no ensino de gerenciamento e processo de desenvolvimento de software.
- [2] Teixeira, M. R. W., de França Aguiar, G., & Pedro, E. (2010). O Jogo Imagem & Ação como Estratégia de Aprendizagem para o Conteúdo de Algoritmos Genéticos.
- [3] Digiampietri, L. A., Kropiwiec, D. D., & Silva, R. A. (2010). O uso de jogos como fator motivacional em cursos de computação.
- [4] Barbosa, L. S., Fernandes, T. C., & Campos, A. M. (2011). Takkou: uma ferramenta proposta ao ensino de algoritmos.
- [5] Silva, J. C., Sousa, S. P., Kulesza, R., & Brito, A. V. (2012). Uma avaliação do emprego do jogo Modelando como apoio ao ensino de Engenharia de Requisitos.
- [6] Oliveira, E. D., & de Barros, R. M. (2013). ABstração Game Um Jogo Para Facilitar o Processo de Ensino e Aprendizagem da Orientação a Objetos.
- [7] Kahwage, C., de França, E. L., Nunes, R. C., Carvalho, R., & Souza, D. T. (2013). Jogo Baralho das Variáveis.
- [8] Bezerra, C. I. M., & Coutinho, E. (2013). Avaliação do Jogo iTestLearning: Um Jogo para o Ensino de Planejamento de Testes de Software.
- [9] Leite, L., Sibaldo, M., Carvalho, T., Souza, R. (2014). Montanha de Chomsky: jogo tutor para auxílio no ensino de Teoria da Computação.
- [10] Jucá, P. M., Alexandrino, A. O., Alexandrino, A. O., Rolim, G. F., & de Almeida, C. D. A. (2014). Aplicação da Gamificação na Disciplina de Empreendedorismo.
- [11] Herpich, F., Jardim, R. R., Silva, R. F., Voss, G. B., Nunes, F. B., Medina, R. D. (2014). Jogo SériO na Educação: Uma Abordagem para Ensino-Aprendizagem de Redes de Computadores (Fase II).
- [12] Vahldick, A., Mendes, A. J., Marcelino, M. J., Hogenn, M., & Schoeffel, P. (2015). Testando a Diversão em um Jogo SériO para o Aprendizado Introdutório de Programação.
- [13] Cagnini, H. E., Charao, A. S., Patricia, P. D. A., & de Azevedo, B. R. (2015). Mundo virtual Minecraft: uma Experiência no Ensino de Circuitos Digitais.
- [14] Barbosa, W. A., Nunes, I. F., Inocencio, A. C., Oliveira, T. B., Parreira Júnior, P. A. (2015). DEG4Trees: Um Jogo Educacional Digital de Apoio ao Ensino de Estruturas de Dados.
- [15] Leite, D. R. A., Delfino, S. S., Mélo, C. B., Medeiros, Á. F., & de Andrade, H. G. R. (2015). GSPROJECTS-Ambiente para simulação da gestão de projetos de software.
- [16] Freitas Nune, I., & Júnior, P. A. P. (2015). RPG4Sorting-Um Jogo Educacional para Auxílio ao Ensino de Métodos de Ordenação.
- [17] Silva Figueiredo, K., Ribeiro, J. M., Souza, R., & Angelo, V. R. (2015). Uma Abordagem Gamificada para o Ensino de Programação Orientada a Objetos. XXXV Congresso da SBC. XXIII Workshop sobre Educação em Computação.
- [18] Silva, R. J., Azevedo, W. L., Nascimento, R. P., & Paxiúba, C. M. (2015). MercadoSI: Um jogo de tabuleiro para o ensino de Sistemas de Informação.
- [19] Campos, A., Gardiman, R., & Madeira, C. (2015). Uma ferramenta gamificada de apoio à disciplina introdutória de programação.
- [20] Krassmann, A. L., Falcade, A., da Silva, L. E., & Medina, R. D. (2015). Serious Games to Computer Networks Learning With CyberCIEGE: A Case Study in Brazilian Higher Education.

O Uso de Recursos Lúdicos Para o Ensino de Processos em Engenharia de Software

Ronney Moreira de Castro^{1,2}, Gleison dos Santos Souza²

¹Faculdade Metodista Granbery – Juiz de Fora (FMG)
Rua Batista de Oliveira, 1145 – Granbery - Juiz de Fora – MG – Brasil

²Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Av. Pasteur, 456 – Urca – Rio de Janeiro – RJ – Brasil

ronney.castro@granbery.edu.br, gleison.santos@uniriotec.br

Abstract. *Teachers and students faces considerable difficulty teaching the use of processes in Software Engineering courses because most of adopted methodologies are based on expository classes with low efficiency. In this article a different approach is presented, which uses ludic resources to aid the teaching of processes, in order to improve students knowledge absorption and consequently bring to practice knowledge only seen in theory.*

Resumo. *Docentes e alunos têm enfrentado grande dificuldade no que se refere à abordagem de ensino-aprendizagem do uso de processos na disciplina de Engenharia de Software, pois a maioria das metodologias adotadas atualmente se baseiam em aulas expositivas com pouca eficiência. Neste artigo é apresentada uma abordagem diferente, que utiliza recursos lúdicos para o auxílio ao ensino de processos, com o objetivo de melhorar a absorção de conhecimento pelos alunos e, conseqüentemente, trazer para a prática conhecimentos somente vistos na teoria*

1. Introdução

A Engenharia de Software (ES) é parte do currículo de muitos cursos superiores de computação. Ela é ensinada geralmente em uma ou mais disciplinas nos cursos de Sistemas de Informação e Ciência da Computação e, muitas vezes, o aprofundamento em seus conhecimentos é feito em nível de Pós-Graduação. A ES está mais diretamente relacionada ao desenvolvimento de software propriamente dito, englobando conteúdos de qualidade, processos de desenvolvimento, técnicas para gerenciamento de projetos, metodologias, entre outros. Cada um desses conteúdos possui diferentes tipos de problemas com características específicas de seu contexto.

Os métodos tradicionais de ensino de ES ainda estão excessivamente centrados no professor, focados na aplicação de teorias para os alunos e, muitas vezes, a parte prática é feita de forma simples utilizando algumas ferramentas em laboratórios ou exercícios em da sala de aula. Isso leva os alunos a uma falta de oportunidade de aplicação prática do conteúdo, podendo gerar desmotivação pela disciplina. Uma forma de melhorar a situação descrita é o uso de métodos alternativos para o ensino da disciplina, tais como estudos de caso, atividades envolvendo projetos de empresas, uso de jogos (tabuleiro, cartas, eletrônicos), aplicação de dinâmicas de grupo, uso de simuladores. Já existem pesquisas como as citadas em Wangenheim [2012], cujo

objetivo é alcançar uma melhoria no ensino de ES. Em sua maioria estão relacionadas ao uso específico de jogos, como, por exemplo, jogos computacionais (projetados para serem jogados através de um software) que envolvem conteúdos da área [FERNANDES e WERNER, 2009].

Este artigo aborda uma forma lúdica de ensinar o que são processos de software fazendo analogia ao uso destes dentro da ES. A Seção 2 aborda metodologias de ensino e aprendizagem, a Seção 3 discorre sobre processos de software, a Seção 4 mostra a dinâmica desenvolvida e sua aplicação, a Seção 5 apresenta os resultados obtidos por este trabalho e, por fim, as conclusões são apresentadas na Seção 6.

2. Evolução das Metodologias de Ensino

A compreensão do funcionamento da mente humana foi um dos principais objetivos e temas de pesquisa de Jean Piaget. Ele buscava compreender, de forma analítica, todas as fases do desenvolvimento humano e suas características psicológicas e mentais [PIAGET, 2003]. A Psicologia da Aprendizagem possui uma série de teorias, dentre as quais, este artigo apresenta três das mais conhecidas: teoria comportamental, teoria cognitivista e teoria construtivista.

2.1 - Psicologia Comportamental

Segundo Baum [1999], a teoria teve início em 1913 por John B. Watson, na qual defendia que a psicologia não deveria estudar os processos internos da mente, mas do comportamento, que era visível. Nesta época a investigação se baseou na observação do comportamento mediante a situação e limites impostos. Sua crença o levou a acreditar que seria possível prever e controlar a conduta humana.

A base foi no estudo e nas teorias do russo Ivan Pavlov sobre condicionamento, como observar aos estímulos gerados ao ver, por exemplo: um cachorro salivar não somente ao ver comida, mas também com o som, gesto ou sinal de chegada da refeição. Qualquer alteração que resulta no estímulo do meio-ambiente pode provocar modificação no comportamento. Nem toda conduta pode ser detectada seguindo esta teoria. Eduard C. Tolman propõe o Neobehaviorismo Mediacional. Nessa teoria o organismo trabalha como mediador entre o estímulo e a resposta, ou seja, ele atravessa etapas que Tolman denominava de variáveis intervenientes - elos conectivos entre estímulos e respostas, estas sim consideradas ações internas [BAUM, 1999] [CATÂNIA, 1999].

2.2 - Psicologia Cognitiva

Segundo Matlin [2004], a psicologia cognitiva é uma área da psicologia que pesquisa as variadas formas de como o ser humano aprende, pensa, percebe e a forma de representação da realidade. Esse ramo abrange também como principais focos de pesquisa, a percepção, o pensamento e a memória, de forma a explicar como o indivíduo compreende o mundo e faz uso do conhecimento adquirido para várias funções cognitivas, tais como falar, raciocinar, memorizar e solucionar problemas.

O processo de transformação das estruturas cognitivas ocorre através da assimilação e acomodação. A assimilação se baseia na interpretação de eventos das estruturas cognitivas existentes. Pelo contrário, a acomodação se baseia na modificação desta estrutura com o objetivo de compreensão do meio.

2.3 - Psicologia Construtivista

A ideia de investigar o surgimento do conhecimento e, como nasce à necessidade da pessoa buscar o mesmo, fez com que Jean Piaget se propusesse a estudar esse fenômeno e entender, nas diversas fases do desenvolvimento do ser humano, como surge o conhecimento e como ele se desenvolve. Algo extremamente interessante observado por Piaget [2003] foi que qualquer pessoa, para poder tentar realizar uma tarefa ou buscar o conhecimento a envolvendo, tem a necessidade de ter pelo menos visto alguém desempenhá-la, ou seja, precisa de um contato inicial, mesmo que de forma bem sucinta para que, a partir daí, seja despertado nessa pessoa a necessidade de aprender.

2.4 Aprendizagem

O ato de adquirir conhecimento, informação e habilidades através de ensino, estudo ou até vivência no dia a dia é denominado aprendizagem. De acordo com Tapajós [2008], no âmbito pedagógico, a aprendizagem é abordada de diversas formas onde cada uma possui sua peculiaridade.

A aprendizagem receptiva, onde o indivíduo é capaz de receber uma instrução e compreendê-la, é possível reproduzir o que lhe foi passado, mas é incapaz de descobrir algo; a aprendizagem por descoberta, onde a pessoa tem que trabalhar para poder conseguir aprender algo, nada é fornecido facilmente. Outra aprendizagem que merece destaque é a repetitiva, na qual existe a memorização do conteúdo, embora a compreensão inicial seja um pouco confusa. Por último, a aprendizagem significativa onde à pessoa consegue relacionar os seus conhecimentos prévios com novas informações passadas [TAPAJOS, 2008].

3. O Ensino da Definição de Processo

Para Filho [2009], processo é uma sequência de etapas que deverão ser seguidas a fim de se atingir um objetivo e agrupa uma série de ferramentas e técnicas. Scheer *et al.* [2006] relata que a ideia de analisar e melhorar os processos veio da indústria da manufatura e foi difundida para outros setores. Pressman [2011] define processos como um conjunto de atividades, ações e tarefas realizadas na criação de algum produto de trabalho. Gerenciá-los implica em uma verificação detalhada de atividades importantes e dos recursos da organização, tais como mercado onde está inserida, suas atividades, pessoas, entre outros.

Pela afirmação de vários autores compreende-se que processos são atividades organizadas e utilizadas de forma adequada para obtenção de um determinado resultado a um alto nível qualitativo. O seu uso está sendo difundido nas empresas que estão em busca de uma melhor qualidade em seus produtos [DAVIS, 2008]. Jones [2010] relata que estudos atuais mostram que as empresas estão procurando compreender os processos que possuem, o que pode afetar sua capacidade de planejar, justificar e implementar as alterações que impactarão nos mesmos, deixando-os mais coerentes com as necessidades e fornecendo mais do que o necessário para uma empresa conquistar um produto bem feito. Segundo o autor, se eles forem seguidos de forma correta irão prevenir as ocorrências de falhas que podem afetar o produto final.

3.1 Aplicações de Processos

Um dos fatores que impulsionaram o constante crescimento do uso de sistemas nas empresas é a competitividade que eles oferecem. Com isso, houve também um grande crescimento na construção dos softwares, produzidos de forma corriqueira e sem nenhuma qualidade aparente, caso conhecido na literatura como “Crise do Software” [WAZLAWICK, 2013].

Fernandes e Teixeira [2011] descrevem o processo de software fazendo um paralelo entre as operações de manufatura. Da mesma forma que se decompõe as etapas de operações de manufatura, o processo de software pode ser dividido em várias partes. Júnior [2002] descreve procedimentos, subprocessos, como a identificação das funções, a complexidade e classificação de um ponto, representando a funcionalidade provida ao usuário, para atender seus requisitos internos e externos em relação aos dados.

3.2 Abordando as Definições em Sala de Aula

De acordo com Behrens [1999], existem dois métodos utilizados para o ensino: um é focado no professor e o outro no aluno. De fato, observa-se uma maior utilização nos dias atuais da metodologia voltada para aulas expositivas, onde o professor fala e os alunos somente ouvem. Nesse tipo de aula não se tem grande aproveitamento pelos alunos, pois estes se dispersam nos primeiros minutos de aula.

A disciplina Qualidade de Software (QS) é de grande relevância para formação dos conhecimentos de ES e, a maior parte de seu conteúdo, é lecionada pelos docentes de forma expositiva, teórica e bastante abstrata e, com isso, acaba sendo menos eficiente, pois ativa somente a audição do aluno [ALBUQUERQUE, 2009]. Para entendimento do conteúdo de QS, é necessário o conhecimento de processos, que é a base para um software com qualidade. No entanto, lecionar processos, assim como outros tópicos da disciplina, envolve um conteúdo muito teórico sendo considerado um desafio para os docentes dos cursos de graduação nos dias atuais.

Para Albuquerque [2009], o ensino do que é processo inicia-se pela conceituação, para que os alunos entendam sua função dentro do contexto em que se encontra. O professor utiliza geralmente slides para a transmissão do conhecimento. Em paralelo, para complementar o entendimento, são buscadas informações em outras fontes, como em artigos. Durante o aprendizado são demonstrados alguns exemplos de empresas que utilizam processos para melhoria da qualidade do produto final e, de forma paralela, é justificado o uso dos mesmos no cenário da ES. Em suma, ensinar processos no ambiente de graduação tem como principal objetivo fazer com que o aluno entenda a sua importância para a qualidade de um produto.

4. O uso de uma Dinâmica de Grupo como Alternativa para Ensino de Processos

A metodologia que mais se utiliza no processo de ensino e aprendizado da ES é a expositiva, onde o professor fala e os alunos escutam [WANGENHEIM, 2012]. A intenção deste trabalho é apresentar uma técnica alternativa de ensino, na qual os alunos tenham uma maior eficácia no aprendizado promovendo interação social, desafio, competição e, conseqüentemente, a recompensa, o que os mantém mais ativos, auxiliando na aprendizagem.

A dinâmica construída consiste na elaboração de um boneco montado por partes. O objetivo é que os alunos produzam o mesmo utilizando a abordagem de processo. No

decorrer da dinâmica são inseridas algumas variações cujo objetivo é demonstrar como o processo não é estático em uma organização, devendo sofrer melhorias constantes, de acordo com o que é observado durante seu uso. Os detalhes da dinâmica, assim como sua preparação, podem ser observados na Tabela 1.

Tabela 1. Detalhes e preparação para a dinâmica

Título	Criação do boneco para ensinar a abordagem de processos e sua utilização.
Objetivo	Mostrar que um processo desorganizado pode afetar a criação de um produto e, que processos alinhados, podem ser bem melhores para as organizações. Mostrar aos alunos, por analogia, que processo em software pode contribuir positivamente para a geração de um produto final de melhor qualidade.
Participantes	Grupos de cinco a seis alunos.
Tempo	50 a 100 minutos.
Material utilizado	13 cartões que conterão as partes do boneco, 13 envelopes de cartão, tesouras, jornal, lápis ou caneta, fita adesiva e um quadro para fixação das peças que serão montadas.
Preparação	Cada envelope deverá conter um cartão escrito com uma peça do boneco a ser confeccionada: 1ª peça: cabeça, 2ª peça: orelha direita, 3ª peça: orelha esquerda, 4ª peça: pescoço, 5ª peça: corpo (tronco), 6ª peça: braço direito, 7ª peça: braço esquerdo, 8ª peça: mão direita, 9ª peça: mão esquerda, 10ª peça: perna direita, 11ª peça: perna esquerda, 12ª peça: pé direito, 13ª peça: pé esquerdo.

Descrição: A dinâmica começa com o professor dividindo a turma em grupos de 5 a 6 pessoas, dispostas em lugares distintos na sala, de forma que um grupo não possa ver o que outro estará fazendo. Ele deve explicar a todos que será aplicada uma atividade que envolverá definições da disciplina ES, mais especificamente: O Uso de Processos para a Melhoria da Qualidade. É necessário também, embaralhar os cartões antes de distribuí-los aos participantes a cada fase. Logo após, deverá distribuir dois a três envelopes para cada grupo e solicitará aos participantes para não abri-los até que seja dada ordem. Será entregue, nesse momento, folhas de jornal e tesouras. O professor, então, informa aos alunos que eles devem construir com o material disponibilizado, o que está escrito nos cartões.

Aplicação: A dinâmica foi aplicada em turmas do 7º e 8º períodos do Curso de Bacharelado em Sistemas de Informação da Faculdade Metodista Granbery nos anos de 2014 e 2015. Em todas as aplicações foram feitas quatro etapas.

Na primeira etapa todas as equipes foram proibidas de se comunicar umas com as outras. O objetivo era de simular a dificuldade que se tem no processo de desenvolvimento de software, onde não existe comunicação com os envolvidos no processo, criando um ambiente nada colaborativo e completamente desorganizado.

O professor estipulou um prazo de 15 minutos para término da tarefa, simulando o tempo corrente para a criação de um produto em um ambiente real. Após o término desse prazo, foi solicitado que todas as equipes aguardassem para poder iniciar a montagem do produto final no quadro. As peças, uma a uma, foram fixadas com fita adesiva no quadro e, no fim, o produto obtido foi de qualidade muito baixa, sendo facilmente visualizado por todos. O objetivo da primeira etapa, que era simular uma produção independente, sem comunicação e padronização foi obtido com total ênfase, já que o produto criado foi um produto ruim, sem aceitação de todos e sem nenhuma

padronização. A Figura 1 mostra os participantes durante a primeira fase da dinâmica e o boneco gerado nessa fase.



Figura 1 – Participantes e boneco gerado na primeira fase da dinâmica

Na etapa posterior foram eleitos líderes de equipes e somente estes estavam liberados para efetuar a comunicação com os demais líderes das outras equipes, simulando o papel de gerente de projetos. Esse, por sua vez, tinha como objetivo compreender as demandas, tentar alinhar sua equipe com as demais e, no final, produzir um produto de qualidade. Foi observado que as pessoas encontraram dificuldades por não haver documentação ou algo que as orientassem.

Isso demonstrou a importância de documentação para a realização de qualquer desenvolvimento de software, pois a padronização dos processos envolvidos é extremamente vital para o sucesso de um projeto que visa qualidade. A Figura 2 ilustra o boneco gerado na segunda fase.



Figura 2 - Boneco gerado na segunda fase da dinâmica

Na terceira etapa, uma pessoa foi eleita como modelo. Essa por sua vez serviu como inspiração para as demais equipes e também simulou uma documentação, na qual os envolvidos poderiam observar e copiar as medidas. Com isso, observou-se uma melhoria da qualidade do produto final, já que as equipes de trabalho tinham algo em que se basear, análogo a uma documentação. A Figura 3 ilustra a utilização do modelo e o boneco que estava sendo gerado nessa fase.



Figura 3 – Uso de um modelo e boneco sendo gerado na terceira fase da dinâmica

Nessa mesma fase foram eleitos também gerentes de projetos e de qualidade, com as tarefas de gerenciar e auditar o processo de todas as equipes, simulando uma forma bem similar à produção atual de um software no dia a dia. Ambos tinham o poder de aprovar ou até mesmo reprovar as demandas realizadas pelas equipes.

Na última etapa os participantes foram inseridos em um ambiente que era bem próximo da realidade, com seus gerentes, líderes e analistas de qualidade, todos desempenhando tarefas inerentes ao seu cargo e as equipes sempre envolvidas no processo de produção. No final, o produto construído, com o auxílio de padronização, sob uma supervisão de gerentes e guiados por analistas focando a qualidade, ficou muito melhor com relação as demais etapas anteriores, conforme ilustra a Figura 4.



Figura 4 - Boneco gerado na última fase da dinâmica

4.1 Abordagens utilizadas

Segundo Wangenheim [2012], o uso de recursos lúdicos, como a dinâmica de grupo, por exemplo, favorece positivamente o processo de aprendizagem, pois mantém os alunos mais ativos.

Objetivando-se uma maior eficiência no processo de ensino e aprendizado, a dinâmica elaborada baseou-se na ideia central da psicologia cognitiva, que estuda a forma como o ser humano assimila as coisas que a realidade oferece. Os alunos foram convidados a aprender de maneira prática o que é um processo e fazer a analogia ao processo de desenvolvimento de software. Eles foram motivados a cada etapa e, à

medida que o projeto se desenvolveu, foram assimilando a ideia de trabalho em equipe, gerência de requisitos, elaboração de documentação, além da importância do uso de processos na produção de software.

5. Resultados

Ao fim da dinâmica foi feita uma análise a partir do primeiro boneco até o quarto. Foi perceptível, através da visualização dos bonecos no quadro, a melhora no desenvolvimento do processo e, conseqüentemente, do produto. Foram acrescentadas várias definições entre as discussões para esta evolução. Os alunos puderam perceber que era necessário continuar evoluindo para obter um melhor produto. Através desta dinâmica foram observados diversos benefícios como: bem-estar, motivação, melhor aprendizado, aprender com o erro e a dificuldade em lidar com pessoas. Os alunos tiveram a chance de cometer erros e analisar o fracasso, dando oportunidade de vivenciar uma experiência próxima do mercado de trabalho e que, em geral, o ambiente acadêmico não possibilita.

Após a dinâmica, os participantes foram convidados a preencher um questionário com 25 afirmativas, cada uma delas com opções de notas: 0 (muito ruim), 1 (ruim), 2 (razoável), 3 (bom), 4 (muito bom) e 5 (excelente), para serem respondidas como feedback do trabalho realizado. A Tabela 2 mostra o questionário aplicado.

Tabela 2. Questionário aplicado

1	A dinâmica atendeu as expectativas em relação ao tema proposto: processos.
2	Estou satisfeito com a dinâmica apresentada porque foi uma atividade prática diferente das aulas convencionais.
3	Ao passar pelas etapas da dinâmica senti confiança de que estava aprendendo.
4	Foi fácil entender a dinâmica e ela pode ser utilizada como material de estudo.
5	O conteúdo da dinâmica agregou outros conhecimentos que já possuía sobre o tema.
6	A utilização desta dinâmica está mais adequada ao meu jeito de aprender.
7	O conteúdo da dinâmica é relevante para entendimento do que é processo.
8	A utilização de uma dinâmica me fez ficar atento à aula ministrada.
9	Algo interessante no início da aplicação da dinâmica capturou minha atenção.
10	A forma como a dinâmica foi feita é atraente (Apresentação, desenrolar, tarefas, finalidade).
11	Tive sentimentos positivos no desenrolar da dinâmica e percebi que é mais eficiente para o aprendizado.
12	Consegui atingir os objetivos da dinâmica sem maiores dificuldades.
13	Gostaria novamente de utilizar mais dinâmicas para aprendizado das disciplinas.
14	Eu recomendaria a aplicação desta dinâmica aos meus colegas.
15	Me diverti com a dinâmica.
16	A dinâmica evolui em um ritmo adequado e não fica monótona.
17	As tarefas envolvidas na dinâmica não são muito fáceis nem muito difíceis.
18	A dinâmica promove momentos de cooperação entre as pessoas que participam.
19	Me diverti junto com as outras pessoas do meu grupo.
20	Foi possível interagir com outras pessoas durante a dinâmica.
21	Eu não percebi o tempo passar durante a dinâmica e quando vi ela acabou.
22	Temporariamente esqueci minhas preocupações do dia a dia. Fiquei totalmente concentrado na dinâmica.
23	As definições trabalhadas na dinâmica irão contribuir para minha vida profissional.
24	A dinâmica foi eficiente para minha aprendizagem em comparação com outras atividades da disciplina.
25	A dinâmica contribuiu para a minha aprendizagem na disciplina.

O questionário teve o objetivo de verificar se, com a aplicação desta dinâmica, o conteúdo foi ensinado de um modo diferente; se foi possível assimilar melhor o conteúdo transmitido em relação ao método tradicional; se os alunos se sentiram motivados a estudar, utilizando artifícios diferentes do método tradicional de ensino. O universo de alunos da pesquisa realizada consistia em 32 participantes na turma A, 35 participantes na turma B, 25 participantes na turma C. O método utilizado foi de contagem do número de respostas em cada item. Após tratamento e análise dos dados, obteve-se 95% de sucesso a partir das questões respondidas, que estão entre ter marcado a nota 4 (muito bom) ou 5 (excelente). Alguns itens merecem destaque como é o caso da questão 2. Cerca de 90% dos respondentes marcou nota 5, o que leva a concluir que as aulas tradicionais, centradas em conteúdo expositivo, tendem a ser mais enfadonhas. O mesmo ocorreu na questão 13, com cerca de 91% dos participantes marcando nota 4 ou 5. Isso também significa que os alunos preferem as atividades lúdicas e, que esse tipo de atividade, possibilita uma melhor absorção do conteúdo trabalhado em sala de aula. A questão 21, com cerca de 93% dos participantes marcando nota 5, mostra que os alunos não percebem o passar da aula quando são utilizados recursos lúdicos como recurso para o aprendizado.

Como ponto positivo mais importante da aplicação da dinâmica nas turmas, pode-se destacar a boa aceitação por parte dos participantes de uma outra forma de aprendizado (recursos lúdicos) em relação ao método tradicional (aula expositiva) e, conseqüentemente, uma maior participação durante a aula. Um ponto negativo observado foi uma certa ociosidade dos membros das equipes gerada na segunda fase da dinâmica, na qual os gerentes poderiam conversar entre si e depois passar as tarefas para suas equipes. Os gerentes ficaram um tempo considerável conversando e não se preocuparam em passar atividades para suas equipes. Em decorrência disso, boa parte da turma ficou ociosa. Os alunos sugeriram uma melhoria na qual os gerentes teriam um tempo estipulado para conversar e, logo após, deveriam passar as tarefas para os membros das equipes.

6. Conclusão

O objetivo do artigo foi apresentar uma dinâmica que consiste, através da construção de um boneco por partes, mostrar o ensino de processos e fazer uma analogia ao processo de produção de um software mostrando sua evolução. Além disso, demonstrar que o processo deve sofrer melhorias para que fique cada vez melhor.

Em geral, o método de aprendizado mais aplicado na área educacional é o tradicional com a utilização da forma verbal, que faz uso da escrita ou da fala como meio de comunicação, com avaliações escritas [BEHRENS, 1999]. O uso de recursos lúdicos permite transpor este paradigma e levar o conhecimento de uma forma mais natural e motivadora. Quando é possível realmente atingir um número maior de alunos, sem distorcer o assunto a ser transmitido, percebe-se que o uso do recurso lúdico tem maior eficácia, já que o público fica mais envolvido e participante.

A dinâmica criada teve a intenção que os alunos tivessem uma maior eficácia no processo de ensino e aprendizado. Eles foram convidados a aprender de maneira prática sobre o que é um processo. Além disso, puderam fazer uma analogia com o processo de desenvolvimento de software, sendo motivados a cada etapa. À medida que o projeto se desenvolveu, foram assimiladas noções de trabalho em equipe, a importância da

documentação, gerência de projetos e requisitos e também a importância do uso de processos na produção de software.

Espera-se também que essa metodologia seja aplicada por outros docentes na área de Engenharia de Software e afins, de forma a motivar os alunos ao aprendizado desta disciplina tão importante.

Referências

- Albuquerque, A. B. (2009) Atividades Lúdicas no Ensino de Processos de Software. VIII Simpósio Brasileiro de Qualidade de Software, Fortaleza, Brasil, 318-325.
- Baum, W. M. (1999) Compreender o Behaviorismo Ciência, Comportamento e Cultura (M.T.A. Silva, M.A. Matos, G.Y. Tomanari, E.Z. Tourinho) Porto Alegre.
- Behrens, M. A. (1999) O Paradigma Emergente e a Prática Pedagógica. Curitiba: Champagnat.
- Davis, R. (2008) ARIS Design Platform: Advanced Process Modelling and Administration. Springer-Verlag, London.
- Fernandes, L. & Werner, C. (2009) Sobre o uso de Jogos Digitais para o Ensino de Engenharia de Software. In: II FEES, SBES, Fortaleza, Brasil, 17-24.
- Fernandes, A. A.; Teixeira, D. S. (2011) Fábrica de Software: Implantação e gestão de operações. 1. ed. São Paulo: Atlas.
- Filho, P. W.P. (2009) Engenharia de Software: Fundamentos, Métodos e Padrões. 3. ed. Rio de Janeiro: LTC, 1248p.
- Júnior, O. S. L.; Farias P. P. M.; Belchior, A. D. (2002). Análise de Pontos por Função Fuzzy. I Simpósio de Qualidade de Software, Gramado, Brasil, 225-236.
- Jones C. (2010) Software Engineering Best Practices. McGraw-Hill, Inc., New York, NY, USA, 1 edition.
- Matlin, M. W. (2004) Psicologia Cognitiva. LTC.
- Piaget, J. (2003) Seis Estudos de Psicologia. 24. ed. Rio de Janeiro: Forense Universitária.
- Pressman, R. S. (2011) Engenharia de Software: Uma Abordagem Profissional. 7. ed. São Paulo: McGraw-Hill.
- Scheer W.A., Kruppke H., Jost W., Kindermann H. (2006) Agility by ARIS Business Process Management. Berlin: Springer-Verlag.
- Wangenheim, C. G. V. (2012) Ensinando Computação com Jogos. 1. ed. São Paulo: Câmara Brasileira do Livro.
- Wazlawick, R. S. (2013) Engenharia de Software: Conceitos e Práticas. 368p. Campus. 2013.

O uso de Robótica para aprendizado de Programação integrando alunos de Educação Básica e Ensino Superior

Luemy Avila^{1,2}, Flavia Cristina Bernardini³, Patrick Moratori³

¹Faculdade de Educação
Universidade Estadual do Rio de Janeiro – UERJ
Rio de Janeiro, RJ, Brasil

²Secretaria de Educação de Macaé
Prefeitura Municipal de Macaé
Macaé, RJ, Brasil

³Laboratório de Inovação no Desenvolvimento de Sistemas – LabIDeS
Instituto de Ciência e Tecnologia – ICT
Universidade Federal Fluminense – UFF, Campus Rio das Ostras
Rio das Ostras, RJ, Brasil

luemyavila@gmail.com, {fcbernardini, pmoratori}@id.uff.br

Abstract. *Nowadays, it is been hard to keep both engagement and attention of students during their learning process due to the broad information access provided by the internet. Moreover, students in general have been having difficulties in logical reasoning, concepts abstraction and problem solving. On the other hand, Computer Science undergraduate students have been reporting a lack of motivation during their initial academic terms, because it is difficult to them to understand applications of their learned knowledge. Consequently, the use of robotic programming allow interactions between students from basic education and undergraduate courses, which allow a learning process through discussions and collaboration between them in performing tasks using technology tools. In this context, a case study is presented in this article. A project was carried out to join local students from undergraduate courses and basic education aiming to learn robotic programming. The results indicate positive effects on participants, improving their self-confidence, abstraction skills and problem solving capability.*

Resumo. *Atualmente, há uma dificuldade em manter o interesse e a atenção dos estudantes no seu aprendizado devido ao amplo acesso à informação trazido pela internet. Ainda, os alunos em geral apresentam dificuldade em raciocínio lógico, abstração de conceitos e resolução de problemas. Por outro lado, alunos de graduação em Ciência da Computação apresentam falta de motivação nos períodos iniciais, já que a aplicação direta dos conhecimentos obtidos na universidade muitas vezes não é percebido. Daí, a interação entre alunos com diferentes níveis de formação para programação de robôs pode permitir um processo de aprendizado por meio de discussões e colaborações na realização de tarefas, utilizando ferramentas tecnológicas. Nesse contexto, um estudo de caso é apresentado neste trabalho. Foi executado um projeto que visava*

uma ampla interação entre alunos de graduação e de escolas do município local para aprendizado de programação em robótica. Os resultados indicaram efeitos positivos nos participantes, onde se observou melhoria nos quesitos autoconfiança, abstração de conceitos e resolução de problemas.

1. Introdução

Nas últimas décadas a educação tem estado carente de inovações para lidar com os desafios da contemporaneidade, existindo um descompasso entre os avanços tecnológicos e a realidade escolar. Observa-se que as gerações atuais convivem cotidianamente com diversos aparatos tecnológicos e um amplo acesso à informação para realizar suas atividades diárias, seja para ler um livro digital ou para comprar um produto pela internet. Porém, tanto o ambiente quanto a estrutura e as metodologias aplicadas nas escolas nem sempre estão alinhados à esta modernidade [MORIN 1996, DEMO 1998, FRAWLEY 2000], principalmente quando se considera a realidade das instituições públicas brasileiras. Consequentemente, existem demandas por estratégias inovadoras que permitam inovar, motivar e manter o interesse deste novo perfil de estudantes.

Para que um aluno possa ter uma posição ativa no processo ensino-aprendizado é preciso que ele seja capaz de abstrair problemas do seu dia-a-dia, a partir do desenvolvimento de seu raciocínio lógico. Neste contexto, a utilização de tecnologias na área educacional tem se tornado uma questão de estado em diversos países da Europa e na América do Norte, onde a utilização de ensino de programação e robótica aplicados à problemas do mundo real tem permitido melhorias em abstrações, atraindo e motivando estudantes em diversos contextos [PEREIRA 2010, BENITTI et al. 2009, AGUIAR et al. 2015, MARINS 1999, RIBEIRO et al. 2011].

As parcerias entre universidades e escolas permitem ações nesse sentido, onde alunos com diferentes níveis de formação podem ser integrados para um aprendizado por discussões e colaborações na realização de tarefas, tornando possível: (i) atrair e motivar estudantes para resolver problemas em diversos contextos; (ii) aproximar alunos da Educação Básica em relação à universidade, conhecendo melhor a realidade e a potencialidade de cursos de graduação; (iii) motivar alunos do Ensino Superior à contribuir com a sociedade, vendo aplicabilidade de seus conhecimentos; e (iv) diminuição de evasões escolares e universitárias. Tais iniciativas são relevantes no âmbito acadêmico-profissional, pois existe uma relação direta entre os setores de tecnologia e o desenvolvimento de ações inovadoras. Porém, um número decrescente de profissionais tem estado disponível no mercado, principalmente pela elevada desistência de alunos em cursos de graduação em áreas relacionadas, como Ciência da Computação e Engenharias [da SILVA et al. 2013].

Este artigo apresenta um estudo de caso de uma atividade de extensão. Trata-se da execução de um projeto durante o ano de 2015 que visava a integração entre alunos da graduação em Ciência da Computação da Universidade Federal Fluminense (UFF) – Campus Rio das Ostras e estudantes da Educação Básica (ensino fundamental e médio) de escolas locais da cidade. A ideia principal é utilizar a robótica no ensino de programação como uma estratégia inovadora na relação educação-tecnologia, observando os fatores motivacionais. A iniciativa do projeto de robótica no Campus Universitário envolve alunos do ensino fundamental, médio e superior, além de seus respectivos docentes. Para a execução deste projeto, é apresentada uma metodologia, identificando as atividades

realizadas bem como a sua avaliação à posteriori pelos participantes.

As seções subsequentes possuem a seguinte organização: a Seção 2 apresenta a motivação e os trabalhos relacionados ao aprendizado de programação pela robótica; a Seção 3 descreve a metodologia aplicada na integração Universidade-Escola no uso da tecnologia associada ao ensino; a Seção 4 discute os resultados obtidos; e por fim, a Seção 5 apresenta as conclusões e trabalhos futuros.

2. O aprendizado de programação na Educação Básica e no Ensino Superior

Atualmente existem demandas para o uso de tecnologias no processo educacional. Iniciativas como o CODE no ensino de programação [CODE 2015] conta com o apoio de diversas empresas e importantes personalidades da atualidade, como Google, Microsoft, Salesforce, Omidyar Network, Reid Hoffman, Mark Zuckerberg and Priscilla Chan, Bill Gates, Drew Houston, Ballmer Family Giving, Ali Partovi and Hadi Partovi, Sean N. Parker Foundation, Diane Tang and Ben Smith, John and Ann Doerr, Quadrivium Foundation, dentre outros. Projetos como “Scratch” [MIT Media Lab 2015], do MIT, e “Eu posso programar” [MICROSOFT 2015], da Microsoft, reforçam essa perspectiva considerando a importância da introdução de conceitos de programação a jovens estudantes, no intuito de permitir uma melhor compreensão de abstrações presentes no mundo real. Tal fato permite compreender o grau de importância que a programação e conceitos associados à área de Ciência da Computação podem impactar no cenário de uma nova concepção do uso e geração de tecnologias e inovações, principalmente quando ensinados em séries iniciais da Educação Básica. Entretanto, um estudo divulgado pela SOFTEX aponta que o Brasil pode chegar em 2020 com um déficit de mão de obra qualificada em TI de 408 mil profissionais. Tal estimativa, que destaca uma posição pessimista, mostra que essa escassez se tornará um problema educacional caso o país não reforce programas para reverter tal situação [COMPUTERWORLD. 2013]. Tal situação é recorrente, pois um estudo da IDC, encomendado pela Cisco em abril de 2013, já havia destacado que até 2015 a demanda na América Latina por profissionais de TI especializados seria 35% maior que a força de trabalho disponível no mercado, chegando a um déficit de aproximadamente 296 mil trabalhadores [NOTÍCIAS DA CISCO. 2013].

Neste sentido, a robótica pode contribuir para a motivação de alunos tanto da Educação Básica quanto do Ensino Superior, pois a aplicação de tecnologias a problemas do mundo real viabiliza as seguintes características [CASTILHO 2002, DEMO 1998, MAISONNETTE 1999, FRANCO 1998]: (i) promover a qualificação profissional do cidadão; (ii) democratizar o acesso aos conhecimentos produzidos para a melhoria da qualidade de vida; (iii) promover a cidadania e valores democráticos aos diferentes atores sociais que se envolvem de forma direta e indireta nas ações; (iv) enriquecer a proposta educacional escolar, implantado ou fortalecendo a aprendizagem pela resolução de problemas, de forma contextualizada, além de desenvolver os valores estéticos, políticos e éticos, organizados sob as premissas de sensibilidade, igualdade e identidade; (v) promover a construção flexível dos saberes com foco na mediação da aprendizagem; e (vi) incentivar que jovens em diferentes faixas etárias se enveredem por áreas relacionadas ao desenvolvimento da Ciência e Tecnologia.

Em diversos trabalhos na literatura são apresentados estudos nos quais se utiliza a programação associada à robótica com intuítos motivacionais. Em [PEREIRA 2010]

é apresentado um trabalho sobre o uso do Lego Mindstorms como uma ferramenta de robótica para o aprendizado de conceitos básicos de computação em uma escola da Zona Rural de Catalão, GO. Em [BENITTI et al. 2009] são apresentados os resultados de um experimento do uso de robôs no ensino médio em diferentes disciplinas do currículo escolar. Em [MARINS 1999], destaca-se a importância da robótica na escola, descrevendo que entre as dez escolas curitibanas que obtiveram as melhores notas no ENEM em 2012, sete oferecem a robótica como disciplina extracurricular. Um representante destas escolas afirma que tal disciplina vem sendo incluída na grade curricular por ser uma atividade lúdica e que, quando o aluno se sente livre, ele aprende mais facilmente. Em [AGUIAR et al. 2015] é apresentada uma proposta de uso de programação em robótica para ensinar a lógica de programação de uma forma indutiva, ou seja, construção de inferências à partir de casos observados. O diferencial da metodologia descrita neste artigo, em contraponto aos trabalhos já disponibilizados na literatura, se dá pela integração entre universidade e escola, identificando os benefícios proporcionados por esta parceria. Em [RIBEIRO et al. 2011], foi realizado um estudo do uso de robótica em disciplinas de programação básica no primeiro período de cursos de graduação, que permitiu observar que o uso da robótica motiva os alunos iniciantes na graduação na programação, bem como facilita o desenvolvimento do raciocínio lógico.

3. Uma metodologia para programação de robôs na integração entre Universidade e Escola

A estratégia principal aplicada no aprendizado de programação utilizando robótica consiste na criação de oficinas que permitem aos alunos debater conceitos relacionados ao seu dia-a-dia. Desta maneira, a sistematização de informações se torna necessária, viabilizando a aquisição de novos conhecimentos durante o processo de construção de um robô. Para se atingir esse objetivo, os envolvidos devem percorrer diversas etapas para viabilizar o desenvolvimento do pensamento criativo, como: (1) a elaboração visual de protótipos a partir de desenhos e gráficos; (2) a utilização do pensamento sistemático que define sequências de ações para resolução de problemas; (3) o uso da escrita para documentar, de maneira clara e objetiva, os passos envolvidos nos projetos realizados; (4) a prática do raciocínio lógico para transpor sistematizações em códigos de programação nos dispositivos que estão sendo construídos; e (5) o processo de pesquisa para a identificação de demandas, onde soluções e facilitadores possam ser propostos. Essas etapas propiciam aos estudantes uma maior maturidade para lidar com desafios da vida, pois os mesmos são instigados à ter uma postura crítica e analítica frente a problemas.

O objetivo principal deste projeto é preparar estudantes da Educação Básica para o desenvolvimento de suas habilidade cognitivas, além de viabilizar a capacitação dos acadêmicos extensionistas. Além da geração de um campo de estágio lúdico para os universitários, existe o estímulo pelo estudo de ciências entre os beneficiados pelo projeto. Desta maneira, jovens passam a considerar a universidade como uma opção viável e favorável à sua formação como indivíduo, possibilitando impactos positivos para a elevação de níveis socioeconômicos e potencializando a qualidade de vida no município. Vale ressaltar que Rio da Ostras está em uma área estratégica na produção de petróleo no país, fazendo parte de um nicho de grande utilização profissional de robôs na indústria.

Para que o projeto atinja um de seus objetivos, que é despertar o interesse nas áreas de Ciências Exatas e Tecnológicas, mais especificamente em Computação e Engenharia,

para alunos da rede pública, é proposta a criação de oficinas direcionadas aos alunos para construção de conceitos e aplicabilidade do conhecimento para a construção de um robô. Tais oficinas podem evidenciar o objetivo de solucionar problemas com ideias inovadoras a partir do levantamento da demanda social, tornando a engenharia e a computação algo lúdico, mais palpável que os cálculos matemáticos, que podem ser assustadores para iniciantes. Para se atingir esse objetivo, os alunos percorrem diversas etapas que possibilitam o desenvolvimento do pensamento criativo, que envolvem: (1) a elaboração de desenhos para protótipos; (2) a sistematização do pensamento para a resolução de problemas; (3) a escrita para documentar, de modo claro e sólido, os passos dos projetos de construção de robôs; (4) o uso do raciocínio lógico para programar os dispositivos construídos; e (v) o processo de pesquisa para levantamento das demandas que necessitam solucionar e ou facilitar o desenvolvimento do robô. Esse desenvolvimento propicia maturidade e uma postura crítica e analítica para os desafios da vida. Para uma execução inicial do projeto, é proposto que o grupo seja composto por até 20 alunos dos Ensinos Fundamental e Médio da cidade de Rio das Ostras. O grupo pode utilizar kits de robótica para trabalhos educacionais. Kits comumente utilizados são os da linha Lego® Mindstorms, utilizados para a execução deste projeto. Esses kits são compostos por (1) blocos de montagem padrão Lego; (2) peças móveis como engrenagens, roldanas, eixos e rodas; (3) sensores de toque e ultrassônico, e (4) módulo principal, o qual é um mini computador programável. Uma IDE (Integrated Development Environment) para programação em blocos acompanha o kit educacional. Essas peças podem ser combinadas de diversas formas, de acordo com a criatividade dos alunos, para a formação de diferentes dispositivos robóticos. Deve ser observado que os kits Lego® Mindstorms não é um simples brinquedo. Tais kits foram concebidos para que estudantes pensem sobre padrões e estruturas não apenas de robótica, mas também de programação e resolução de problemas de forma mais geral. Como o projeto visa levar o aprendizado de programação e o uso do raciocínio lógico para soluções de problemas a alunos de Educação Básica, esses alunos devem (1) pesquisar e desenvolver programas utilizando programação em robótica utilizando os kits; e (2) desenvolver robôs e estruturas que podem ser trabalhadas para a resolução de algumas tarefas que possam ser aplicadas no cotidiano.

Durante os encontros são elaborados alguns esquemas e táticas para realização de tarefas, onde cada aluno participa integralmente do processo, ou seja, desde a montagem de robôs até sua programação e cálculos necessários para construir soluções. Para realizar tais missões, os alunos recebem orientações dos monitores de graduação para desenvolverem tanto soluções lógicas quanto táticas de programação. Este processo é facilitado pelo uso da IDE que acompanha o Lego Mindstorm, na qual as soluções são desenvolvidas a partir de sequências de blocos programáveis, permitindo, de maneira intuitiva, a implementação dos mais diversos comportamentos desejados. Por exemplo, para que o robô ande para alguma direção, é utilizado o bloco de ação que envia um comando para os motores responsáveis pela locomoção. Similarmente, pode-se implementar que o robô apenas realize uma determinada tarefa após a detecção de um determinado obstáculo, ou seja, terá que aguardar uma resposta de algum dos sensores utilizados, seja ele de toque ou ultrassônico, para que a ação seja desempenhada. Dessa maneira, os alunos podem acompanhar o desenvolvimento e execução de tarefas, partindo de algo conceitual e teórico para soluções práticas. Com isso, várias alternativas são exploradas para a resolução de problemas específicos.

Além do acompanhamento durante a execução do projeto, ao final é indicada uma reunião com todos os participantes do projeto, ou a aplicação de questionários, para identificar os pontos fortes e fracos durante a execução. Também, atividades pouco trabalhadas por alguns alunos podem ser variadas no ano subsequente para outros conhecimentos serem trabalhados, retirando os alunos de suas zonas de conforto.

4. Resultados

Para compor a turma para participação no projeto, foram visitadas duas escolas, uma de ensino fundamental do município, e outra de ensino médio. A execução do projeto teve duração de março a dezembro de 2015. Participaram ao longo do projeto onze alunos das duas escolas e cinco alunos de graduação. É importante observar que alguns alunos das escolas desistiram do projeto no decorrer da execução, quanto outros entraram no meio do ano. Como não havia notas associadas no certificado de participação, não havia a obrigatoriedade da participação. Um dos critérios para participação era a oportunidade de trabalhar com desenvolvimento de programas e resolução de problemas com robótica para participar tanto de torneios quanto olimpíadas de programação de robôs. Para o desenvolvimento dos robôs, foram utilizados três kits Lego Mindstorms – um kit NXT e dois kits EV3. Observou-se que os conhecimentos adquiridos durante as reuniões faziam com que os alunos ficassem mais confiantes em suas habilidades interpessoais e intelectuais, e conseqüentemente, mais entusiasmados com o projeto.

Em pouco tempo de projeto, notou-se grande empenho nos participantes para participar de eventos relacionados, vista a formação de duas equipes para participar das Etapa Regional e Estadual das Olimpíadas Brasileiras de Robótica – OBR e uma equipe para participar da competição First Lego League – FLL. As equipes que participaram da OBR regional, em junho de 2015, ficaram em segundo e décimo primeiro lugar na competição de desenvolvimento de tarefas pelo robô. Já na etapa estadual, em outubro de 2015, uma das equipes conquistou o prêmio de robô com o melhor design. Para a etapa regional da FLL, os alunos tinham que fazer uma pesquisa sobre a temática da competição. Escolheram o tema lixo eletrônico. Foi construído um painel com exemplos de lixo eletrônico, um folder de explicação dos problemas desse tipo de lixo, e um jogo programado em Scratch também sobre essa temática. Essa competição ocorreu em Novembro de 2015. O time não foi premiado, mas foi fundamental a participação dos alunos no evento para perceberem a importância do trabalho em equipe, fundamental para as carreiras profissionais atualmente. Observa-se que a experiência dos alunos nesses eventos trouxeram uma maior maturidade para os alunos, deixando mais cientes de possíveis desafios que estão por vir na competição.

O entusiasmo dos alunos ficou evidente quando se observou seu comportamento nesses eventos, desde o momento da inscrição, a escolha do nome da equipe, até o começo dos treinos para a competição. Percebe-se melhorias nas formas de trabalho e na interação com os outros parceiros de equipe, onde são construídos vínculos de lealdade e confiança. Em outra perspectiva, observa-se o impacto positivo observado por uma professora de Física do colégio estadual que acompanhou as atividades. A professora afirmou: “Pude perceber uma melhoria quanto ao [meu] aluno em sala de aula, uma vez que, antes dele frequentar o projeto não tinha tanta motivação no aprendizado e era muito dependente de mim para realizar os exercícios propostos em sala de aula. Depois que ele começou a ir aos encontros pude perceber que estava bem menos dependente para a execução de

atividades e mais comunicativo com a sua turma.”

Já para os alunos da graduação, pôde-se perceber que a vontade de ensinar e aprender com esses alunos é muito gratificante, pois mesmo em um projeto piloto, os alunos possuem grande liderança, dedicação e proatividade. Um dos alunos monitores do projeto, do curso de Ciência da Computação, afirmou: “O legal do projeto é que em nossas reuniões não existe a definição de aluno e professor, todos estamos ali para ensinar e aprender. Eu mesmo aprendi muito com um dos alunos do ciclo básico de apenas 15 anos e que tem uma enorme facilidade com a programação dos blocos em Lego”. Outro aluno da graduação também considerou: “[...] tem sido muito gratificante e proveitoso participar deste processo, pois desenvolvemos nossas habilidades de comunicação e colaboramos com a difusão do conhecimento no decorrer do estudo. Na questão de qualificação profissional, sentimos que não estamos presos a um único trabalho específico e sim a vários desafios em paralelo, visto que durante o desenvolvimento das tarefas vão surgindo novas possibilidades de pesquisas. Neste momento, compreendemos de forma mais objetiva a importância de um projeto [como este] para a formação mais integral de um profissional em relação a um mercado de trabalho cada vez mais competitivo. Conseguimos melhorar o nosso desempenho acadêmico, já que aprendemos a ser mais comunicativos, trabalhamos em equipe e resolvemos problemas proativamente.”

Ao final da execução do projeto, foi feita uma chamada para uma reunião com todos os membros participantes do projeto. Compareceram à reunião seis alunos das escolas, denominados e_1, \dots, e_6 , e três alunos da graduação, denominados g_1, \dots, g_3 . Foi ouvido um depoimento de cada aluno, cujo resumo está descrito a seguir:

- e_1 (participou de todas as competições do time), que observou-se executar um papel de líder no time, disse que aprendeu os valores de trabalho em equipe, e que passou a aplicar tais valores na escola, tendo bons resultados nas suas relações interpessoais;
- e_2 (participou da FLL) ficou interessado no projeto por ter gosto pelas disciplinas de matemática e física, mas também gostou de programação e da participação na FLL;
- e_3 (participou da FLL) gostou do projeto por fazer novos amigos, mas também sentiu que tinha que levar as atividades mais a sério que a própria escola, pois tinha mais interesse e motivação. Aprender lógica de programação com os alunos da graduação ajudou na programação em Scratch e no Lego;
- e_4 (participou da FLL) afirmou que gostou muito da participação no projeto e no torneio FLL. Atuou bastante na montagem do robô, o que gostou bastante, mas achou complicada a programação;
- e_5 (participou da OBR regional e estadual) disse que compôs uma equipe sozinho para participar dos eventos, o que tornou as atividades mais difícil. Ainda assim, gosta de programar, e ressaltou suas atividades realizadas para o time participar da FLL, apesar de não ter participado;
- e_6 disse que ajudou bastante nas pesquisas de materiais na temática de lixo eletrônico, pois tinha dificuldade em programar e montar o robô;
- g_1 disse que percebia os diferentes perfis entre os alunos, e tentava ser rigoroso para garantir o bom andamento do trabalho;
- g_2 observou que a equipe era grande e que ninguém se conhecia, e observou que, com a interação, subgrupos foram sendo construídos. Um dos alunos de escola

tinha tido problemas com a equipe, e g_2 acredita que um dos motivos foi essa questão do subgrupo;

- g_3 disse que sempre chamava a atenção da equipe para construírem o robô, fazendo com que os alunos voltassem sua atenção para as atividades da competição.

Vale salientar que os alunos colaboraram entre si, passando conhecimentos para colegas, ou até mesmo absorvendo, de forma mais intuitiva, o que tinha ficado previamente com dúvidas. Assim, fica facilitado o acúmulo de conhecimentos nas tentativas de resolução dos desafios lançados em grupo. Adicionalmente, nota-se que os alunos ficam mais proativos e estimulados na busca de novas informações e, assim, vão deixando a timidez e ficando cada vez menos dependentes dos monitores da graduação.

5. Conclusões e trabalhos futuros

Neste trabalho, uma metodologia foi apresentada para execução de um projeto para aprendizado de robótica que visa a integração entre alunos de Educação Básica e Ensino Superior. Tal integração pretende atrair e motivar estudantes para resolver problemas em diversos contextos através de atividades em grupo, aproximando alunos da Educação Básica à Universidade e motivando alunos do Ensino Superior à contribuir com a sociedade, vendo aplicabilidade de seus conhecimentos. Na execução deste projeto, foi possível observar para alguns participantes da Educação Básica, uma melhoria na autoconfiança, no raciocínio lógico e na resolução de problemas. Para os alunos de graduação, foi possível observar também que viam a aplicabilidade de seus conhecimentos também na área da educação. Como trabalhos futuros, pretende-se ampliar a investigação do impacto deste projeto no percurso escolar dos alunos participantes do projeto.

Agradecimentos

Agradecemos à Universidade Federal Fluminense pelas bolsas concedidas para a execução deste projeto e à Odebrecht Óleo e Gás pelo apoio financeiro para participação dos alunos nas competições.

Referências

- [AGUIAR et al. 2015] AGUIAR, Y., MACIEL, B., MATTOS, S., SOARES, L., and OLIVEIRA, V. (2015). Introdução à robótica e estímulo à lógica de programação no ensino básico utilizando o kit educativo lego® mindstorms. In *Anais do IV Congresso Brasileiro de Informática na Educação (CBIE 2015)*. Disponível em <http://www.br-ie.org/pub/index.php/wie/article/view/2166>.
- [BENITTI et al. 2009] BENITTI, F., VAHLICK, A., URBAN, D., KRUEGER, M., and HALMA, A. (2009). Experimentação com robótica educativa no ensino médio: ambiente, atividades e resultados. In *WIE 2009*. Disponível em <http://www.br-ie.org/pub/index.php/wie/article/view/2166>.
- [CASTILHO 2002] CASTILHO, M. (2002). Robótica na educação: Com que objetivos? Dissertação de Mestrado — Universidade Federal do Rio Grande do Sul. Disponível em <http://www.proinfo.gov.br/upload/biblioteca.cgd/192.pdf>. Acessado em 28 de fevereiro de 2016.
- [CODE 2015] CODE (2015). Code: Anybody can learn. [online]. Disponível em <https://code.org/>. Acessado em 28 de fevereiro de 2016.

- [COMPUTERWORLD. 2013] COMPUTERWORLD. (2013). Déficit de talentos de ti no brasil pode chegar a 408 mil em 2020. [online]. Disponível em <http://computerworld.com.br>. Acessado em 06 de março de 2016.
- [da SILVA et al. 2013] da SILVA, H. M., de ALEMIDA, L. M., BATISTA, I. D. S., and GORGÔNIO, F. L. (2013). Uma reflexão sobre o crescente desinteresse e a constante evasão em cursos de Computação e Informática. In *Proceedings of International Conference on Engineering and Technology Education*, volume 12, pages 166–170, Dili, East Timor.
- [DEMO 1998] DEMO, P. (1998). *Educar pela Pesquisa*. Autores Associados, Campinas, SP, 3ª edition.
- [FRANCO 1998] FRANCO, S. (1998). *O Construtivismo e a Educação*. Mediação, Porto Alegre, RS.
- [FRAWLEY 2000] FRAWLEY, W. (2000). *Vygotsky e a Ciência Cognitiva: linguagem e integração das mentes social e computacional*. Artmed.
- [MAISONNETTE 1999] MAISONNETTE, R. (1999). A utilização dos recursos informatizados a partir de uma relação inventiva com a máquina: A robótica educativa. [online]. Disponível em <http://www.proinfo.gov.br/upload/biblioteca.cgd/192.pdf>. Acessado em 28 de fevereiro de 2016.
- [MARINS 1999] MARINS, L. (1999). Robôs enriquecem o currículo escolar e estimulam o cérebro. [online]. Disponível em <http://www.gazetadopovo.com.br>. Acessado em 28 de fevereiro de 2016.
- [MICROSOFT 2015] MICROSOFT (2015). Eu posso programar. [online]. Disponível em <https://www.eupossoprogramar.com/>. Acessado em 28 de fevereiro de 2016.
- [MIT Media Lab 2015] MIT Media Lab (2015). Scratch: Imagine, program, share. [online]. Disponível em <https://scratch.mit.edu/>. Acessado em 28 de fevereiro de 2016.
- [MORIN 1996] MORIN, E. (1996). *O problema epistemológico da complexidade*. Lisboa: Publicações Europa-América.
- [NOTÍCIAS DA CISCO. 2013] NOTÍCIAS DA CISCO. (2013). Estudo revela demanda crescente por profissionais de tecnologia no brasil, mas há falta de mão de obra qualificada. [online]. Disponível em <http://globalnewsroom.cisco.com>. Acessado em 06 de março de 2016.
- [PEREIRA 2010] PEREIRA, G. (2010). O uso da robótica educacional no ensino fundamental: relatos de um experimento. Trabalho de Conclusão de Curso do Bacharelado em Ciência da Computação — Universidade Federal de Goiás, Campus de Catalão. Disponível em <https://dcc.catalao.ufg.br/up/498/o/Gabriela2010.pdf>. Acessado em 05 de abril de 2015.
- [RIBEIRO et al. 2011] RIBEIRO, P. C., MARTINS, C., and BERNARDINI, F. C. (2011). A robótica como ferramenta de apoio ao ensino de disciplinas de programação. In *22º SBIE / 17º WIE, 2011*, Aracaju.

Oficina do Código: Um projeto para o ensino e integração de alunos do ensino fundamental e médio na área de tecnologia

Stéphanie Alencar Braga¹, Caio Raveli F. Barbosa¹, Priscylla da Silva Tavares¹,
Mikhail José Pires Pedrosa de Oliveira¹

¹Instituto Federal de Educação Ciência e Tecnologia do Ceará (IFCE)
Av. 13 de Maio, 2081 – Fátima – CEP 60040-531 – Fortaleza – CE – Brasil

{stephanie.abraga, caio.raveli, priscyllatavares, mikhailj.pedrosa}@gmail.com

Abstract. *This paper is a report of the experience acquired in the "Oficina do Código" project: an action of education, integration and formation of Middle and High school students that are included in low-income group of Fortaleza city, located in Ceará state. The project's motivation was the inclusion of these children and teenagers in order to make them capable to contribute to the society growth, improve their creativity, group work and logical and critical thinking. The project is a partnership between the Fortaleza City Government, that sponsored the robotic kits used in class, as well as the others costs and the IFCE (Federal Institute of Ceará), that provided the computer labs. The activities included both theoretical and practical classes aiming to increase the interest of the students in basic projects of the robotics field developed by themselves. In order to help the teaching of programming, electronic circuits and robotics, it was used the open source platform Arduino.*

Resumo. *O presente trabalho é um relato de experiência do projeto Oficina do Código: um projeto que tem como categoria de ação o Ensino, Integração e formação de Crianças do Ensino Fundamental e médio de baixa renda da cidade de Fortaleza-CE. A grande motivação do projeto foi a inclusão dessas crianças para que essas sejam capazes de contribuir para o crescimento da sociedade, estimular a criatividade, trabalho em grupo, pensamento lógico e crítico. O projeto é uma parceria da Prefeitura Municipal de Fortaleza, que custeou os kits de robótica utilizados nas aulas, bem como outros custos com os alunos beneficiados e o IFCE (Instituto Federal do Ceará), que cedeu os laboratórios para as aulas acontecerem. As atividades do Projeto foram feitas através de Aulas Teóricas e Práticas, visando a ambientação e despertar de interesse a projetos básicos de robótica desenvolvidos pelos próprios alunos. Para auxiliar no ensino de programação e prática de eletrônica e robótica foi utilizada a plataforma open source Arduino.*

1. Introdução

Segundo França e Amaral (2013), no Brasil, o debate sobre a implantação de um currículo de Computação na educação básica ainda é incipiente. Em 2001, foi criado em Roraima um projeto de lei que viabilizou a implantação da disciplina de Computação no ensino fundamental do 6º ao 9º ano e em todos os anos do ensino médio nas escolas da rede pública do Sistema Estadual de Educação. Entretanto, ainda não se observa a publicação

de nenhum panorama sobre o assunto nos veículos científicos nacionais em que Educação e Computação são tratadas conjuntamente, apesar de outros estados também terem promovido ações voltadas ao ensino de Computação com estudantes do ensino fundamental e médio [França e Amaral 2013].

O contato com a Computação permite o desenvolvimento de diversas capacidades, úteis para qualquer atividade profissional, como é o caso do aprendizado e a aplicação de técnicas para resolver problemas, o contato com a abstração de conceitos e outras diversas competências decorrentes do uso das tecnologias [Scaico et al. 2012]. Para a SBC a Computação deve ser tratada como qualquer outra ciência que auxilie diretamente no desenvolvimento do educando e como consequência, na evolução do país.

Para França et al. (2012), a grande maioria das profissões do século 21 exige uma compreensão de instrumentos ou de conceitos tecnológicos. Profissionais de arte e entretenimento, comunicação, saúde, entre outros, precisam ter conhecimentos tecnológicos. Logo, mesmo para os alunos que não tiveram completa afinidade com estudo e desenvolvimento inicial de tecnologia, tiveram chance de conhecer a forma em que a mesma tem se integrado em outras profissões como ferramentas para agilizar, melhorar e automatizar seus processos específicos.

O incentivo do aprendizado de tecnologia pode beneficiar muitas classes ainda no Brasil. Em seu trabalho Rabelo et al (2015) utilizaram a ferramenta *Scratch* para produzir recursos interdisciplinares com disciplinas indígenas estimulando o pensamento computacional através destas atividades, integrando programação, arte e cultura e língua Tupi. Já em Maliuk (2009), o autor mostrou a robótica no ensino de matemática para alunos do nono ano da escola EMEF José Mariano Beck em Porto Alegre. O objetivo era melhorar o rendimento dos alunos na disciplina de matemática e, posteriormente, despertar o interesse deles no mundo tecnológico.

No presente trabalho é apresentado o projeto Oficina do Código, que foi planejado para introduzir alunos de escolas públicas da cidade de Fortaleza no meio da Computação e Tecnologia, apresentando aos alunos desde aspectos teóricos de *Hardware* e *Software* até o desenvolvimento prático de um carrinho microcontrolado utilizando a plataforma *open source Arduino*. A ideia fundamental do projeto é despertar o interesse dos jovens no mundo tecnológico de maneira profissional, visto que, no caso do Brasil, o aumento na demanda por profissionais de tecnologia reforça a necessidade de atraí-los desde cedo [Meneses et al. 2015].

O trabalho está apresentado da seguinte maneira: a seção 2 descreve o projeto e o quê o motivou, a seção 3 apresenta a plataforma *Arduino* no auxílio da educação, a seção 4 descreve as etapas de planejamento e metodologia, a seção 5 as fases de execução do projeto, a seção 6 apresenta os resultados e análise e, por fim, a seção 7 apresenta as considerações finais.

2. Motivação e Descrição do Projeto

A Sociedade Brasileira de Computação - SBC atua e defende o ensino da Computação enquanto ciência desde a educação básica, pois segundo a mesma, o domínio da Computação e das tecnologias da informação é fundamental e estratégico para o desenvolvimento social e econômico mundial [Silva et al. 2015]. Nesse contexto, a grande

motivação do projeto foi a inclusão de crianças do Ensino Fundamental de baixa renda para que essas sejam capazes de contribuir para crescimento da sociedade, estimular a criatividade, trabalho em grupo, pensamento lógico e crítico.

O projeto Oficina do Código aprovado no edital 04/2014 da prefeitura municipal de Fortaleza como um dos projetos do Programa “Nossa ação muda a cidade”, buscava ensinar às crianças das escolas públicas de Ensino Fundamental (5º a 9º ano) de Fortaleza, a lógica e linguagem básica de programação e fundamentos de eletrônica atrelando, posteriormente, à Robótica. A equipe do projeto era composta por 4 integrantes, dois coordenadores e 2 professores, que visavam proporcionar conhecimento básico nessas áreas citadas, para despertar a curiosidade e capacitar alunos de baixa renda na área da Tecnologia.

3. A Plataforma *Arduíno* e a educação

O fato do projeto Oficina do Código ter como objetivo principal a introdução de conceitos voltados à computação e tecnologia para iniciantes, foi escolhida a plataforma microcontrolada *Arduíno*. Essa plataforma de prototipagem serviu de auxílio no ensino de crianças, adolescentes, jovens ou adultos, por se tratar de uma ferramenta de código aberto (*open source*), de baixo custo e de fácil utilização [Perez et al. 2013]. Com essa ferramenta, foi possível mostrar na prática conceitos de programação e eletrônica.

O *Arduíno* teve seu início comercialmente em 2005 no Interaction Design Institute na cidade de Ivrea, na Itália, desenvolvido pelo professor Massimo Banzi e pelo pesquisador da Universidade de Malmö, na Suécia, chamado David Cuartielles. Anteriormente, o *Arduíno* foi desenvolvido como recurso auxiliar no ensino dos estudantes. Porém, como os produtos no mercado eram caros e difíceis de usar, Banzi e Cuartielles desenvolveram um microcontrolador com as principais exigências a um preço baixo e com uma facilidade de uso de tal forma que qualquer pessoa pudesse utilizá-lo [Evans et al. 2013].

A plataforma *Arduíno* tem outras facilidades que a torna bastante flexível para inseri-la no meio educacional quando o ensino é relacionado à tecnologia. Esta possui em seu *hardware* uma interface de comunicação USB (*Universal Serial Bus*) para acessar de forma mais simples o seu microcontrolador, tem um próprio ambiente de desenvolvimento integrado (IDE), do inglês *Integrated Development Environment*, que é apresentado de forma bem intuitiva, utilizando uma linguagem de programação baseada nas linguagens C e C++. Assim, torna-se descomplicado fazer recepção e envio de sinais analógicos/digitais para componentes como sensores, atuadores, leds, motores, entre outros [Kalil et al. 2013].

4. Etapas de Planejamento e Metodologia

Após a aprovação do projeto Oficina do Código, iniciaram-se as etapas de planejamento, antes do mesmo começar a ser executado. Nas seções subsequentes estão descritas essas etapas.

4.1. Escolha e Aquisição do Material Didático

Após o recebimento dos recursos financeiros para o projeto, foi iniciada a compra do material necessário para as aulas, para que estes chegassem a tempo da realização do curso. Foi realizada uma pesquisa de preços, pois o fomentador do projeto exigia 3 orçamentos,

dos quais seria escolhido o com preço mais acessível. Devido as limitações no orçamento, foram comprados os materiais da Tabela 1.

Tabela 1. Materiais utilizados nas aulas

Material	Quantidade
Kit de Desenvolvimento <i>Arduíno Uno</i>	8
Kit de componentes eletrônicos	8
Chassi com rodas e motores	8
Multímetros	4
Baterias para alimentação dos kit de robótica	8

4.2. Inscrições e Seleção dos alunos

Depois de um consenso entre a equipe executora, foi tomada a decisão de que os alunos que seriam selecionados para participar do projeto deviam obedecer aos seguintes critérios:

- Os alunos seriam de escolas diferentes, ou seja, não escolheu-se uma escola específica para selecionar os alunos;
- Os alunos teriam que comprovar baixa renda, por meio de documentos;
- Tinham que estar cursando entre o 6º ano do ensino fundamental e 3º ano do ensino médio;

Os alunos inscritos foram selecionados por meio da análise de notas. Primeiro analisou-se os que tinham melhores notas em matemática. Em caso de empate, os que tinham as melhores notas em português. Permanecendo, a média geral de todas as disciplinas.

O projeto foi divulgado por meio de cartazes, redes sociais, rádios locais e pessoalmente em algumas escolas, por um período de aproximadamente dois meses. Ao todo, se inscreveram 54 alunos (32 do sexo masculino e 22 do sexo feminino) e foram selecionados 17. Dos selecionados apenas 2 estavam no ensino médio, o restante ainda cursava o ensino fundamental II. Os alunos selecionados não tinham conhecimento prévio ou alguma experiência sobre programação/robótica.

O perfil da turma, após serem selecionados os alunos, está ilustrado na Figura 1.

4.3. Ementa do Curso

Como o objetivo do curso era aliar conhecimentos de computação/programação e eletrônica, a escolha da ementa buscou inserir conteúdos básicos de computação e eletricidade nas primeiras aulas, o que daria suporte às aulas práticas de programação com a plataforma *Arduíno*. A Tabela 2 mostra como ficou a ementa inicial do curso.

5. Execução do Projeto

Após a turma selecionada, ementa definida e materiais adquiridos, o projeto começou a ser executado. Houve uma parceria com o Instituto Federal do Ceará (IFCE) - Campus Fortaleza, que cedeu o espaço de seus laboratórios para que as aulas fossem ministradas.

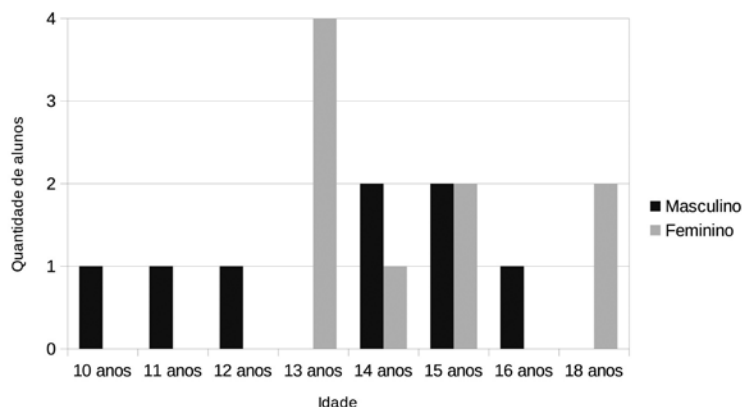


Figura 1. Perfil dos Alunos

Tabela 2. Ementa do Curso

	Conteúdo
Aula 1	<ul style="list-style-type: none"> ● Apresentação Inicial: Explicar o projeto, o que irá ser estudado. ● Mostrar exemplos de Tecnologias existentes. ● Apresentar as ferramentas que serão utilizadas (Placas, exemplos de programação, etc). ● Apresentação de vídeos e exemplos práticos, visando motivar as crianças.
Aula 2	<ul style="list-style-type: none"> ● Explicar o conceito de eletricidade: Corrente, Tensão contínua e Alternada. ● Apresentar alguns componentes eletrônicos e mostrar o funcionamento de alguns: resistor, leds, fontes de bancada. ● Explicar conceitos básicos de eletrônica digital.
Aula 3	<ul style="list-style-type: none"> ● Mostrar aplicações com todos os conceitos mencionados e introduzir o conceito de microcontroladores. ● Prática com o multímetro: mostrar o funcionamento e realizar algumas medições.
Aula 4	<ul style="list-style-type: none"> ● Apresentação do que é programação. ● Práticas com o site “Hora do Código”.
Aula 5	<ul style="list-style-type: none"> ● Conceitos de programação. ● Práticas.
Aula 6	<ul style="list-style-type: none"> ● Conceitos de programação. ● Práticas.
Aula 7	<ul style="list-style-type: none"> ● Conceitos de programação. ● Práticas.
Aula 8	<ul style="list-style-type: none"> ● Apresentação da plataforma <i>Arduíno</i>. ● Identificação dos conceitos estudados utilizando <i>Arduíno</i>. ● Práticas.
Aula 9	<ul style="list-style-type: none"> ● Introdução à robótica. ● Definição do projeto prático final. ● Práticas.
Aula 10	<ul style="list-style-type: none"> ● Execução da Projeto final com os kits de Robótica.

5.1. Aulas

O projeto foi desenvolvido por meio de aulas apenas teóricas e aulas em que houve uma parte teórica e prática. As aulas aconteciam aos sábados e tinham a duração de 4 horas. Iniciavam às 8h e terminavam às 12h, com pausa de 15 minutos para lanche.

No laboratório havia computadores suficientes para que os alunos ficassem no máximo aos pares. Havia também um *data show* para a utilização de *slides* e ilustrações na hora da aula. O conteúdo era ministrado pelos dois professores do projeto, contando também com o auxílio dos dois coordenadores. A Figura 2 mostra uma foto registrada da primeira aula do projeto.



Figura 2. Foto da primeira aula

5.2. Mudanças na ementa

Apesar de ter sido planejada uma ementa, na prática a mesma foi modificada e adaptada, de acordo com as necessidades observadas em cada aula.

Até a aula 4 não houve mudanças, mas a partir da aula 5 a ementa mostrada na Tabela 2 foi praticamente toda modificada. Na aula 5 já foi apresentada a plataforma *Arduíno*, pois na aula 4 os alunos conseguiram assimilar bem alguns conceitos básicos de programação, como o que são variáveis, tipos de variáveis, declaração, etc. Como os alunos iriam praticar a programação utilizando a plataforma *Arduíno*, decidiu-se começar as práticas diretamente na plataforma. As práticas foram definidas e inseridas nas aulas, pois ao final iriam ser integradas ao projeto final com o carrinho. A Tabela 3 mostra as mudanças que ocorreram na ementa a partir da aula 5.

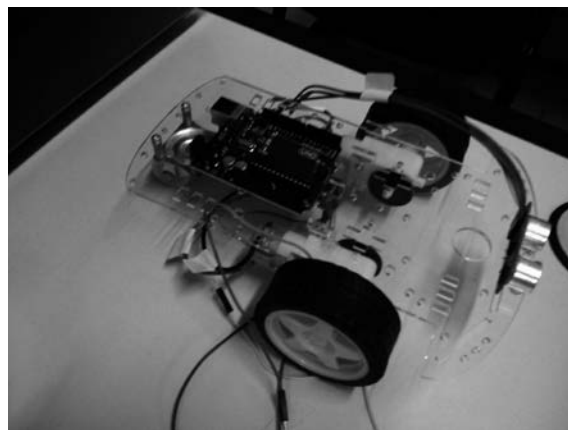
5.3. Projeto prático

No decorrer das aulas, foi estabelecido o objetivo de, ao fim do curso, executar um projeto prático utilizando o carrinho e os sensores adquiridos. A partir da aula 7, as práticas realizadas serviram para compor o projeto final: “Botar o carrinho pra andar”.

Ao fim, os próprios alunos realizaram a montagem de 8 unidades do projeto final, com o auxílio dos professores caso tivessem alguma dúvida. Os materiais utilizados em cada um foram os seguintes: carrinho, 2 motores CC, *Arduíno Uno*, sensor ultrassônico, 4 pilhas. O código final do projeto foi uma junção dos códigos que iam sendo construídos nas aulas práticas. A Figura 3 ilustra a versão final de um dos carrinhos montados.

Tabela 3. Ementa modificada

	Conteúdo
Aula 5	<ul style="list-style-type: none"> • Apresentação da plataforma <i>Arduíno</i> e programações iniciais.
Aula 6	<ul style="list-style-type: none"> • Práticas iniciais utilizando a plataforma <i>Arduíno</i>: Led + botão.
Aula 7	<ul style="list-style-type: none"> • <i>Arduíno</i> e sensores: estudo do sensor ultrassônico. • Prática com o sensor ultrassônico.
Aula 8	<ul style="list-style-type: none"> • Montagem do robô.
Aula 9	<ul style="list-style-type: none"> • Prática: <i>Arduíno</i> e motor CC junto com o carrinho.
Aula 10	<ul style="list-style-type: none"> • Prática final: “Botando o carrinho pra andar”. Junção do sensor ultrassônico + motor.

**Figura 3. Versão final de um dos carrinhos montados**

6. Análise dos Resultados

Durante o período de realização do projeto, aproximadamente 3 meses, foi possível observar, comprovar e aprender muitas coisas. Com relação à equipe executora do projeto (professores e coordenadores), a experiência foi de grande valia e a oportunidade de lidar com um público que ainda não se tinha trabalhado foi de grande importância e aprendizado. Isso pode ser observado no fato da mudança da ementa do curso. As ações planejadas nem sempre se realizavam como se esperava, levando a equipe a mudar um pouco a metodologia.

Com relação aos alunos, pode-se comprovar que o número de alunos que concluíram o curso foi inferior ao número de alunos que iniciaram. Apesar do esforço da equipe em motivá-los, ainda houve 2 desistências, totalizando 15 alunos concludentes. Uma das desistentes informou que desistiu por questões de dificuldade de locomoção até o IFCE e o outro não informou a respeito. Quanto aos outros houve boa frequência, apenas 1 aluno faltou 3 vezes, 4 alunos faltaram 2 vezes e 2 alunos faltaram 1 vez. A Figura 4 mostra os alunos recebendo o certificado de conclusão no fim do curso.

Mesmo o curso apresentando boa frequência, o impacto que a falta de um aluno causa na aula é significativo, já que a turma é pequena. Principalmente nas aulas iniciais, onde abordava-se mais a teoria, era necessário no começo da aula, fazer uma espécie de revisão do conteúdo da aula anterior. Porém, apesar de atrasar um pouco a aula servia para os alunos que não faltaram lembrarem o conteúdo e assim assimilarem melhor.

Quanto ao interesse e desempenho dos alunos durante as aulas, foi possível observar que havia um interesse maior pelas aulas práticas, por exemplo utilizando o multímetro, os componentes eletrônicos e a placa arduíno e até mesmo com programação. Na hora de montar a parte mecânica do carrinho os alunos tiveram um desempenho surpreendente, a maioria sendo capaz de apenas ler o manual e montar corretamente. Também se mostravam bastante pró-ativos com relação aos equipamentos, demonstrando total interesse pelas coisas novas que estavam vivenciando.

Em relação à interação dos alunos com a linguagem de programação, os resultados foram bastante satisfatórios. A partir da aula 4 já foram introduzidos os primeiros conceitos com o desafio “Programação com Anna e Elsa”, disponível em [A Hora do Código]. Os alunos tiveram ótimo desempenho no desafio e conseguiram resolver rapidamente, mostrando-se bastante interessados.

Já quando foi introduzida a plataforma *Arduíno*, os alunos tiveram um pouco de dificuldade, sendo necessário lembrá-los quase sempre as funções. A maior dificuldade encontrada foi devido ao idioma (inglês) das funções utilizadas no *Arduíno*, pois os alunos em sua grande maioria não sabiam inglês ou sabiam muito pouco. Esse fato dificultou a interação dos alunos quando iam implementar os códigos, pois eles tiveram dificuldade de assimilar o nome das funções e a relacionar o nome à sua real função.



Figura 4. Entrega dos certificados aos alunos no fim do curso

7. Considerações Finais

Este artigo apresentou um relato de experiência descrevendo as etapas de desenvolvimento e resultados obtidos com o projeto Oficina do Código. Pelos resultados apresentados, apesar de algumas dificuldades encontradas, o objetivo do projeto foi atingido, pois os alunos mostraram-se bastante interessados pela área de computação. Alguns mostraram-se bastante interessados pela plataforma *Arduíno*, e outros disseram que queriam seguir carreira nessa área de computação.

Pelo perfil da turma foi possível comprovar que os alunos não tinham conhecimento, ou tinham muito pouco sobre programação/robótica. Os assuntos ensinados eram novidade para a grande maioria, apenas 1 dos 17 alunos já tinha participado de uma iniciativa de ensino a robótica na escola onde estudava. Os demais ainda não tinham vivenciado esse tipo de experiência, nem projetos nem na escola onde estudavam. Isso mostra que as escolas públicas de Fortaleza ainda têm muito a avançar nessa área.

Agradecimentos

A Prefeitura Municipal de Fortaleza e ao Banco Interamericano de Desenvolvimento (BID) pelo financiamento do projeto, ao IFCE por ter cedido o espaço e equipamentos dos laboratórios e ao professor Edson Almeida, chefe do departamento de Telemática do IFCE, por se mostrar bastante prestativo às necessidades logísticas do projeto.

Referências

- A Hora do Código. Programação com Anna e Elsa. <https://studio.code.org/s/frozen/stage/1/puzzle/1>. Acessado em março de 2016.
- Evans, M., Noble, J., and Hochenbaum, J. (2013). *Arduíno in Action*. Manning Publications. New York.
- França, R. S. and Amaral, H. J. C. (2013). Ensino de Computação na Educação Básica no Brasil: Um Mapeamento Sistemático. In *XXI Workshop Sobre Educação em Computação (WEI)*, Maceió - AL.
- França, R. S., da Silva, W. C., and do Amaral, H. J. C. (2012). Ensino de Ciência da Computação na Educação Básica: Experiências, Desafios e Possibilidades. In *XX Workshop Sobre Educação em Computação (WEI)*, Curitiba - PR.
- Kalil, F., Hernandez, H., Antunez, M. F., Oliveira, K., Ferronato, N., and dos Santos, M. R. (2013). Promovendo a robótica educacional para estudantes do ensino médio público do Brasil. In *Nuevas Ideas en Informática Educativa TISE*. Porto Alegre - RS.
- Maliuk, K. D. (2009). Robótica educacional como cenário investigativo nas aulas de matemática. In *Dissertação (Mestrado em ensino de matemática) - Instituto de Matemática*. Universidade Federal do Rio Grande do Sul. Porto Alegre.
- Meneses, L. F., Mai, L. F. F., Rosario, J., Oliveira, E., and Gomes, R. L. (2015). IntroComp: Atraindo Alunos do Ensino Médio para uma Instigante Experiência com a Programação. In *XXIII Workshop Sobre Educação em Computação (WEI)*, Recife - PE.
- Perez, A. L. F., Darós, R., Puntel, F. E., and Vargas, S. R. (2013). Uso da Plataforma Arduino para o Ensino e o Aprendizado de Robótica. In *International Conference on Interactive Computer aided Blended Learning*.
- Rabelo, H. M., de Oliveira, W. K. F., Santos, L. L., de Araújo, A. L. S. O., and Souza, F. V. C. (2015). Scratch na produção de recursos interdisciplinares com disciplinas indígenas. In *XXIII Workshop Sobre Educação em Computação (WEI)*, Recife - PE.
- Scaico, P. D., Corlett, E. F., Paiva, L. F., Raposo, E. H. S., and Alencar, Y. (2012). Relato da Utilização de uma Metodologia de Trabalho para o Ensino de Ciência da Computação no Ensino Médio. In *XVIII Workshop de Informática na Escola*, Rio de Janeiro- RJ.
- Silva, S. F., Barbosa, A. F., Souza, A. A., Silva, E. G., Oliveira, M. L. S., Silva, S. R., and Santos, W. O. (2015). Relato de Experiência de Ensino de Computação no Ensino Fundamental em Estágio Supervisionado da Universidade de Pernambuco no Campus Garanhuns. In *XXIII Workshop Sobre Educação em Computação (WEI)*, Recife - PE.

PBL e robótica no ensino de conceitos de Lógica de Programação

Flávia Zenaro Nogueira e Silva, Marcos Augusto Francisco Borges

Faculdade de Tecnologia - Unicamp

R. Paschoal Marmo, 1888 – 13484-332 – Limeira – SP – Brasil

flavia.zenaro@pos.ft.unicamp.br, marcosborges@ft.unicamp.br

Abstract. *Studies show that programming logic school subjects have high failure levels, increasing demotivation and dropout of student risks. The use of robotics for educational purposes may help to promote Teaching-Learning Processes and knowledge creation improvements by using activities that arouse the interest of students. This paper's goal is to show programming logic concepts teaching proposal using robotics kits as pedagogical tool based on Problem-Based Learning (PBL) method and Collaborative Learning.*

Resumo. *Estudos apontam que a disciplina de lógica de programação possui altos índices de reprovação, contribuindo para a desmotivação e possível evasão de alunos. O uso de robótica com finalidades pedagógicas pode ajudar a promover situações de ensino-aprendizagem e aumento da construção do conhecimento, pela introdução de atividades que despertam o interesse e aguçam a curiosidade e motivação dos alunos. Este trabalho tem por objetivo apresentar uma proposta de ensino de conceitos de lógica de programação utilizando kit robótico como ferramenta pedagógica baseada no método Problem-Based Learning (PBL) e na Aprendizagem Colaborativa.*

1. Introdução

Os jovens estudantes, fascinados e mergulhados em um mundo de tecnologia e inovação, onde tudo acontece “on-line”, requerem um novo tipo de abordagem pedagógica.

O avanço da tecnologia nos permitiu empregar novas ideias na área educacional, como, por exemplo, a utilização da robótica, que é utilizada como ferramenta pedagógica no ensino de disciplinas como lógica e linguagem de programação. O emprego de computadores e kits de robótica em ambientes educacionais pode ser entendido como um caminho natural, uma ferramenta adequada para o desenvolvimento de atividades que envolvam criar, projetar e planejar, favorecendo assim o processo de ensino-aprendizagem (Chella, 2002).

A disciplina de Lógica de Programação (LP), também conhecida como Algoritmos, está presente em quase todas as grades de cursos de Graduação e Técnico da área de Computação. Essa disciplina também pode ser encontrada em grande parte dos cursos de engenharia, por ajudar a desenvolver o raciocínio lógico e matemático, fator importante para ambas as áreas. Vista por muitos alunos como uma matéria complexa e de difícil entendimento, a disciplina em questão tem papel fundamental para prepará-los nos conceitos básicos de programação para que, posteriormente, possam

aprender uma linguagem de programação e escrever códigos para computador e dispositivos móveis.

Rapkiewicz *et.al*, (2006) comentam que a falta de motivação e a dificuldade de desenvolver o raciocínio lógico, necessários para a construção de algoritmos, são os fatores principais da desistência dos ingressantes em programação. Outro fator importante são os altos índices de reprovações nas disciplinas iniciais (Gomes *et.al*, 2008)(Peixoto *et.al*, 2013).

O presente artigo propõe formas de trabalhar alguns conceitos de LP utilizando a metodologia PBL (*problem-based learning*) e *kit* robótico como ferramenta pedagógica. O restante do artigo está organizado da seguinte forma: a seção 2 apresenta os trabalhos relacionados que contribuíram para o desenvolvimento desta pesquisa, a seção 3 relata o embasamento teórico, a seção 4 descreve a análise dos *kits* e a proposta de ensino, por fim, a seção 5 discute as considerações finais e os trabalhos futuros.

2. Trabalhos Relacionados

Existem várias pesquisas sobre a utilização do *kit* robótico *Lego Mindstorms* no ensino, o uso da robótica na educação e ferramentas de apoio para o ensino de LP.

Os autores De Jesus & Cristaldo (2014) relatam pontos positivos em suas experiências na utilização do *Lego Mindstorms* em uma olimpíada de robótica para aprendizagem nas áreas de Ciência e Algoritmos com turmas de ensino médio integrado de informática. Constatou-se um aumento no índice das notas acima de 9.0 em aproximadamente 8% após a participação dos alunos no projeto. Outro fator positivo, observado por esses autores, foi a aceitação do projeto por parte dos alunos, com 80% das opções de avaliação marcadas como excelente, o que mostra a grande aceitação na participação de atividades com o uso de robótica.

Almeida & Netto (2015), fizeram uma revisão sistemática na literatura sobre a Robótica no contexto Educacional, Aprendizagem de Programação e Programação de Robôs a Distância. Analisando 82 artigos, segundo esses autores, a robótica no contexto educacional tem sido utilizada de forma interdisciplinar, associada a aplicações desenvolvidas nos *kits Lego Mindstorms* ou Arduino. Sobre a aprendizagem de programação com robôs, perceberam a utilização de diferentes tecnologias, porém, apenas um trabalho auxiliava no ensino de LP e outros dois faziam uso das linguagens Python e C para a programação. Chegaram à conclusão que um grande número de métodos e ferramentas vem sendo utilizado como apoio na aprendizagem de programação para robôs.

Neto *et. al* (2015) também apresentaram uma revisão sistemática do uso da robótica na educação durante os anos de 2004 até 2014. Os autores notaram que a maior parte da pesquisa estava localizada nas regiões Sudeste e Nordeste do país. Identificaram também a Robótica Educacional (RE) como uma matéria interdisciplinar, sendo utilizadas por outras doze disciplinas, como Física e Química, sem contar a própria RE.

A comparação entre as publicações estudadas mostrou que a RE está sendo mais utilizada no ensino fundamental, do que nos ensinos médio, técnico e superior. Outro item importante observado é o grande volume de propostas, principalmente de novos

produtos de *Software*, metodologias e robôs para os diversos níveis de ensino, em contrapartida, existem poucas experiências práticas relatadas na área.

De maneira geral, os autores tem relatado grande aceitação dos alunos na utilização da robótica como meio de ensinar conceitos de diversas disciplinas. Para a disciplina de LP, no entanto, observou-se a menção de apenas uma aplicação prática. O presente artigo apresenta sugestões de utilização do ambiente de programação do *kit Lego Mindstorms* que podem ser usadas em sala de aula para ensinar alguns conceitos de LP.

3. Embasamento teórico

3.1 – Problem-Based Learning (PBL)

A PBL é uma abordagem para o ensino que tem crescido em todo o mundo desde os anos 70, a qual permite que os estudantes sejam capazes de perceber as situações e estruturas particulares de como eles aprendem e como eles se veem como futuros profissionais (Savin-Baden, 2000).

Segundo Norman e Schmidt (2000), a PBL fornece uma abordagem à educação mais desafiadora, motivadora e interessante. A resolução de problemas reais que podem ser vivenciados no dia-a-dia de uma profissão, faz com que o aluno tenha interação com a vida real, praticando a interação social através da vivência na resolução destes problemas. A PBL também estimula a motivação intrínseca (epistêmica) dos alunos, através da colocação e discussão em sala de aula de problemas relevantes a sua futura área de atuação profissional. Schmidt (2001) afirma em seu trabalho que a motivação epistêmica funciona como uma força interna que leva as pessoas a conhecerem melhor o mundo.

3.2 – Pensamento Computacional

O Pensamento Computacional (PC) é o processo de pensamento envolvido na formulação de problemas e das suas soluções, tendo como base os fundamentos da Ciência da Computação. A combinação do pensamento crítico com os fundamentos da computação define uma metodologia para resolver problemas, denominada Pensamento Computacional ou *Computational Thinking* (Wing, 2006). Define-se como PC “saber usar o computador como um instrumento de aumento do poder cognitivo e operacional humano” (Blikstein, 2008). Esta visão defende que a cultura do computador ajuda a sociedade não somente a aprender, mas especialmente oferece uma nova maneira de aprender a aprender (Lu & Fletcher, 2009).

PC pode ser considerado, conforme Wing (2006), como uma das habilidades intelectuais básicas de um ser humano, comparada a ler, escrever, falar e fazer operações aritméticas, habilidades estas que servem para descrever e explicar situações complexas. O PC pode ser visto como uma linguagem (como as linguagens escrita e matemática) que pode ser utilizada para as pessoas falarem (descreverem) sobre o Universo e seus processos complexos.

3.3 - Aprendizagem Colaborativa

Segundo Torres, Alcantara e Irala (2004, p. 3) a Aprendizagem Colaborativa “é uma estratégia de ensino que encoraja a participação do estudante no processo de aprendizagem e que faz a aprendizagem um processo efetivo e ativo.” Ainda segundo os

autores, os alunos de forma coletiva, constroem seu próprio conhecimento por meio da troca de informações, de pontos de vista, de opiniões divergentes, de resoluções de questões e de avaliações. Outra definição para Aprendizagem Colaborativa é “[...] uma situação em que duas ou mais pessoas aprendem ou tentam aprender alguma coisa juntas” (Dillenbourg, 1999, p. 1, tradução nossa)¹.

4. Ferramentas e Métodos

O presente projeto de pesquisa visa, através do *kit* robótico escolhido, demonstrar propostas de maneiras para trabalhar alguns conceitos da LP com objetivo de identificar o melhor uso dos recursos oferecidos pelo *kit*.

O projeto foi dividido em três etapas, sendo elas: análise de *kits* robóticos disponíveis no mercado, análise dos recursos disponíveis para o *kit* escolhido e uma formulação de Proposta de Ensino.

4.1. Análise de *kits* robóticos disponíveis no mercado

A primeira etapa consistiu da análise dos *kits*: Arduino², *Lego Mindstorms*³ e *Modelix Robotics*⁴.

1. Arduino

Foi originalmente desenvolvido como recurso pedagógico para o ensino e depois no ano de 2005 foi lançado comercialmente por Massimo Banzi e David Cuartielles (Monk, 2013).

A placa Arduino pode variar dependendo do modelo escolhido, de maneira geral, é uma plataforma de prototipagem eletrônica de hardware livre e de placa única, projetada com um microcontrolador Atmel AVR com suporte de entrada/saída embutido, o Arduino *Software IDE* (ambiente de desenvolvimento integrado) pode ser instalado nos sistemas operacionais Windows, Macintosh OSX e Linux, porém, a maioria dos sistemas microcontroladores são limitados para uso no sistema operacional Windows.

Sua linguagem de programação oficial é baseada na linguagem C/C++, com algumas particularidades para geração automática de protótipos de funções. Os programas escritos usando o Arduino *Software IDE* são chamados de *sketches* e salvos com a extensão de arquivo .ino (Arduino, 2016). Um exemplar da placa UNO é ilustrado na Figura 1(a) e do Arduino *Software IDE* na Figura 1(b).

¹ Texto original: “[...] it is a *situation* in which *two or more* people *learn* or attempt to learn something *together*.”

² <https://www.arduino.cc/>

³ <http://www.lego.com/en-us/mindstorms>

⁴ <http://modelix.cc/>



(a) Arduino Uno



(b) Arduino Software IDE

Figura 1 - Ilustração da placa e do ambiente de desenvolvimento do Arduino. Fonte: <https://www.arduino.cc>

2 – Lego Mindstorms

O kit *Lego Mindstorms* é comercializado pela empresa dinamarquesa LEGO e foi criado em 1980, em parceria com o Instituto de Tecnologia de Massachusetts (MIT) (Lego, 2016).

Sua programação é feita de forma visual através de um ambiente de programação próprio, que pode ser obtido no site oficial da empresa Lego. Sua programação é feita de forma visual, com a utilização de *blocos* funcionais que estão agrupados em: blocos de ação, blocos de fluxo, blocos de sensores, blocos de operações de dados e blocos avançados (Lego, 2016).

O *Brick* é a unidade programável do *kit* onde os programas são carregados, armazenados e executados. Esta unidade contém microprocessador, quatro portas de entrada nomeadas como 1, 2 3 e 4, quatro portas de saída A, B, C e D, uma porta mini USB, para conectar do *Brick* a um computador, uma porta para cartão de memória e uma porta USB (Lego, 2016).

O *kit* possui diversas peças com vários formatos de forma a possibilitar a montagem de estruturas como tanques, carros e robôs ou em formatos como de escorpião e de cobra. Alguns tipos de montagem encontram-se ilustrados na Figura 2.



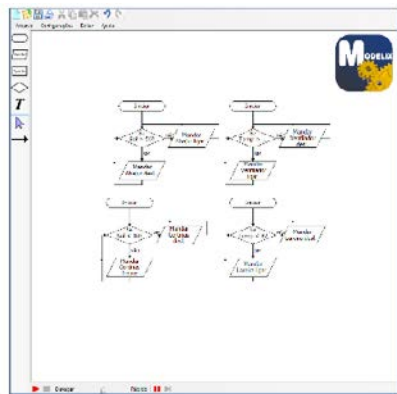
Figura 2 - Opções de montagens com o kit Lego Mindstorms EV3. Fonte: <http://www.lego.com/>

Atualmente, o *kit Lego Mindstorms* está em sua terceira geração, chamada de EV3, e traz como inovação a possibilidade dos robôs serem controlados por dispositivos como *smartphones* e *tablets*.

3 – Modelix Robotics

O *kit* robótico Modelix Laboratório de Robótica 411 PLUS é voltado para ensino de robótica com ambiente de programação próprio, onde a programação é feita através

da construção de fluxogramas, conforme ilustra a Figura 3(a), sem a necessidade de conhecimento de nenhuma linguagem de programação. As peças que compõem o *kit* permitem alguns tipos específicos de montagem, ilustrados na Figura 3(b). (Modelix, 2016).



(a) Ambiente de programação

(b) Opções de montagens

Figura 3 – Ambiente de programação e opções de montagens com *kit* Laboratório de Robótica 411 PLUS - Fonte: <http://modelix.cc/>

4 - Análise das principais características entre os *kits*

Na Tabela 1 são listadas algumas características observadas em cada *kit* as quais são de fator determinante para a escolha do mais apropriado para o ensino de LP.

Tabela 1 - Comparação entre os *kits* robóticos analisados

<i>Kits</i>	Montagem	Programação	Material de instruções para montagem	Preço
Arduino	É necessário noções de eletrônica. Sem limitação de opções de montagem.	Linguagem C/C++	Ausente	R\$ 199,00
<i>Lego Mindstorms</i>	Montagem simples com peças de encaixe. Sem limitação de opções de montagem.	Ambiente próprio de programação. Programação de modo Visual.	Grande quantidade de conteúdo disponível no site do fabricante.	R\$ 2.856,00
Modelix Robotics	É necessário noções de eletrônica. Limitação de opções de montagem.	Ambiente próprio de programação. Uso de Fluxograma.	Conteúdo fornecido junto ao <i>kit</i> .	R\$ 2.950,00

Para análise de custo, serão considerados os valores para aquisição de *kits* de robótica. Para o Arduino, uma opção de *kit* pode ser encontrada pelo valor de R\$

199,00⁵ que vem com o Arduino Uno R3 e com componentes de eletrônica. O *kit Lego Mindstorms Education EV3* pode ser adquirido no valor de R\$ 2.856,00⁶. Uma opção de *kit* oferecida pela empresa do Modelix é o *kit* de robótica 411 PLUS pelo preço de R\$ 2.950,00⁷. Em questão de custo, a opção do *kit* Arduino se revela a mais vantajosa em comparação com as demais, no entanto, levando em consideração a usabilidade (facilidade de uso) nos itens de montagem e programação para alunos ingressantes, o *kit* da empresa LEGO se destaca o mais apropriado como já mencionado anteriormente.

4.2 - Análise dos recursos disponíveis para o *kit* escolhido

Para este projeto, foi escolhido o *kit Lego Mindstorms*, por oferecer vários tipos de sensores prontos para serem usados como, por exemplo, sensores de toque, luz e presença, como ilustrado na Figura 4, e pelo fato de não precisar de conhecimentos prévios de programação ou do conhecimento de alguma linguagem de programação, como é o caso do Arduino.

Outro ponto positivo para a escolha do *kit* LEGO, ao invés do *kit* Arduino e do Modelix, é sua facilidade de montagem, não exigindo conhecimentos de eletrônica, como é caso do Arduino e sendo mais fáceis de encaixar, não necessitando de uso de ferramenta como chave de fenda e o uso de parafusos, como acontece com o Modelix. A não necessidade de conhecimentos de eletrônica é um fator importante, pois a disciplina de LP é oferecida nos semestres iniciais e os alunos, em sua maioria, não possuem conhecimentos avançados de eletrônica. Também existe a facilidade de encontrar exemplos de montagem, tutoriais passo-a-passo e demais materiais de apoio para iniciantes, fornecidos pela própria empresa, através do site da Lego.



Sensor de cor

Sensor de rotação

Sensor de toque

Sensor Infravermelho

Figura 4 - Sensores do *kit* robótico EV3. Fonte: <http://www.lego.com/>

4.3 - Proposta de Ensino

A proposta está focada em três assuntos da LP: programação sequencial, estrutura de decisão e de repetição.

1. Programação sequencial

⁵ Arduino Uno R3 Fonte: https://www.robocore.net/modules.php?name=GR_LojaVirtual&prod=105. Acessado em: 23 de maio de 2016.

⁶ Fonte: <http://zoom.education/produtos/conjunto-lego-mindstorms-education-ev3>. Acesso em: 23 de maio de 2016.

⁷ Fonte: <http://www.kitsderobotica.com.br/kit-escolar-411-laboratorio-de-robotica>. Acesso em: 23 de maio de 2016.

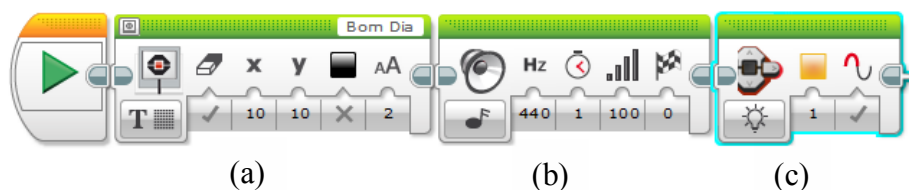


Figura 5 - Programação Sequencial. Fonte: do Autor

Para exemplificar uma programação sequencial é proposto o uso dos blocos: *Display*, *Sound* e *Brick Status Light* encontrados na aba *Action*. A programação proposta está ilustrada na Figura 5.

Desta forma, o aluno pode perceber que o robô executa as ações representadas pelos blocos (a), (b) e (c), ilustrados na Figura 5, de forma sequencial. Ao final da execução do programa, o robô terá feito sequencialmente, a exibição do texto “Bom Dia” na tela do *Brick*, utilizando o bloco *Display*, a emissão de um som, através do bloco *Sound*, e ter acendido a luz amarela do *Brick*, com o uso do bloco *Brick Status Light*.

2. Programação com estrutura de decisão

Como proposta para utilizar a estrutura de decisão em uma programação, foi criada a programação utilizando o sensor de toque ou *Touch Sensor*, encontrado na aba *Sensor*, o qual através de seus estados: 0 – Liberado, 1- Pressionado e 2 Pressionado e Solto, é possível programar uma ação para cada uma das opções de verdadeiro ou falso do estado testado.

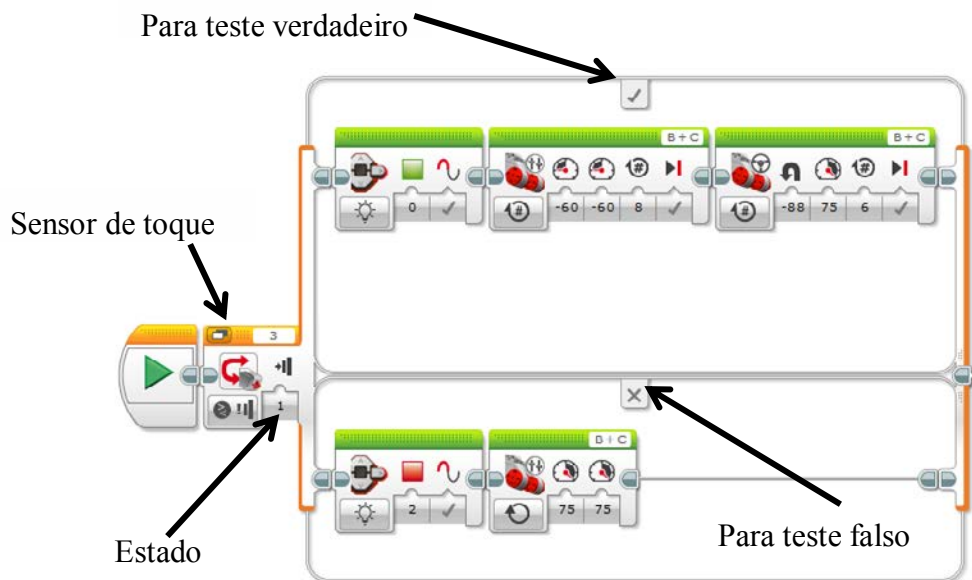


Figura 6 - Programação com Estrutura de Decisão. Fonte: Do autor.

Na Figura 6, a programação pode ser entendida da seguinte maneira: Se o sensor de toque que está com o estado 1, for verdadeiro, que significa que o sensor foi pressionado, então a execução do programa será a da parte superior, fazendo com que acenda uma luz verde do *Brick* e em seguida o motor gire em sentido anti-horário, fazendo o carro andar de marcha ré e depois faça uma curva.

Em caso falso, ou seja, se a situação não for de sensor pressionado, a execução executada será a da parte de baixo da imagem, a qual possui um bloco para acender a luz vermelha do *Brick* e fazer o motor funcionar com rotação para frente.

3 – Programação com estrutura de repetição

Para ilustrar como uma estrutura de repetição pode ser utilizada no ambiente de programação *Legó Mindstorms Home Edition*, foi proposta a criação da programação ilustrada da Figura 7. Nesta proposta foi utilizado o bloco *Loop*, encontrado na aba fluxo, o qual tem a função de fazer com que todos os blocos contidos dentro dele sejam repetidos conforme o tipo de repetição escolhida, podendo ser: ilimitada, por um período de tempo, por uma quantidade definida de vezes, entre outras opções.

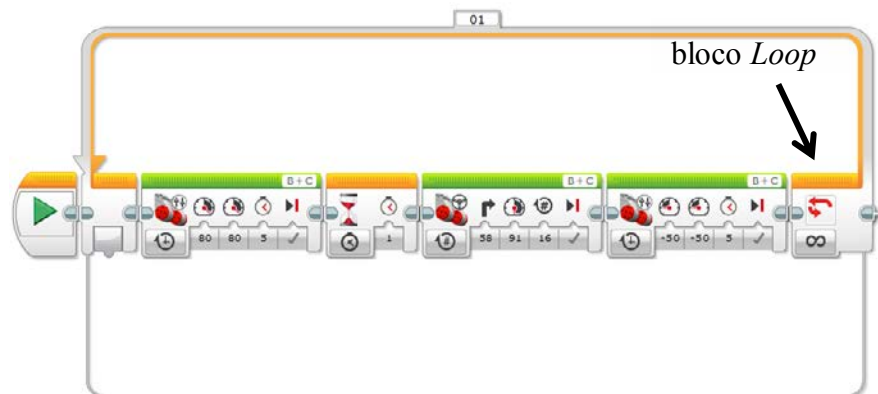


Figura 7 - Proposta de programação com Estrutura de Repetição. Fonte: Do autor.

Com este exemplo pretende-se deixar claro para o aluno o que uma estrutura de repetição faz com a fácil identificação de todos os itens que serão repetidos.

Para aplicação das propostas relatadas neste trabalho é sugerida a divisão dos alunos em grupos, podendo ser de 3 a 7 alunos, de forma a propiciar a Aprendizagem Colaborativa, onde os alunos trabalham de forma conjunta e sem hierarquia definida. Outra sugestão, agora para a aplicação da PBL, seria propor um problema no qual um robô faria o monitoramento e vigilância em uma área determinada. O robô deve percorrer um trajeto, chegando até um ponto definido e retornando ao ponto de partida, contornando possíveis barreiras.

5. Considerações Finais e Trabalhos Futuros

Para o ensino de LP é importante atentar para a escolha de um *kit* de robótica que se encaixe com o perfil dos alunos que farão uso. Para que o objetivo de ensinar LP a alunos iniciantes seja alcançado, a identificação e o uso de estruturas simples e claras são fundamentais para que o assunto trabalhado seja de fácil entendimento.

O presente artigo demonstrou que é possível trabalhar conceitos de LP de uma maneira prática, de forma que possa despertar o interesse dos alunos propiciando a melhor compreensão dos conceitos trabalhados, sem que, para isso, seja necessária a utilização de grande quantidade de recursos do ambiente de programação e tão pouco de recursos complexos.

Como sugestão de trabalhos futuros, podem-se criar atividades de ensino para os conceitos não abordados na proposta apresentada. Além disso, sugere-se a elaboração

de mais exemplos com a utilização de outros sensores e recursos para os conceitos de lógica descritos deste trabalho.

Referências

- Arduino (2016). Projeto Arduino. <http://arduino.cc>. Acessado em 21 de fevereiro de 2016.
- Almeida, T., Netto, J. F. (2015). Robótica Pedagógica Aplicada ao Ensino de Programação: Uma Revisão Sistemática da Literatura. In Anais do Simpósio Brasileiro de Informática na Educação (SBIE) (Vol. 26, No. 1, p. 597).
- Chella, M. T.(2002). Ambiente de Robótica Educacional com Logo. In: XXII Congresso da Sociedade Brasileira de Computação - SBC2002. Florianópolis - SC
- De Jesus, L., Cristaldo, M. F. (2014). Uma abordagem utilizando LEGO Mindstorms Education EV3 para verificar o desempenho acadêmico dos estudantes do Instituto, pp. In Anais do Simpósio Brasileiro de Informática na Educação (SBIE) (Vol. 25, No. 1, pp. 1198-1202).
- Gomes, A., Areias, C. M., Henrique, J., Mendes, A. (2008). Aprendizagem de programação de computadores: dificuldades e ferramentas de suporte. Revista Portuguesa de Pedagogia, 42, 2, 161-179.
- Lego. (2016). *LEGO*. Acesso em 03 de Outubro de 2015, disponível em The Lego group: <http://www.lego.com/en-us/mindstorms/products/ev3/31313-mindstorms-ev3/>
- Monk, S. (2013). Programação com Arduino: Começando com Sketches-Série Tekne. AMGH Editora.
- Neto, R. P., Santana, A. M., Rocha, D. P., Souza, A. (2015). Robótica na Educação: Uma Revisão Sistemática dos Últimos 10 Anos. In Anais do Simpósio Brasileiro de Informática na Educação (SBIE) (Vol. 26, No. 1, p. 386).
- Norman, G. R.; Schmidt, H. C. (2000). Effectiveness of problem based learning curricula theory, practice and paper darts. Medical Education, 34, 721–728.
- Peixoto, M., Scaico, P., Souza, F., Peixoto, H. (2013) Uso de Estratégias de Aprendizagem e Motivacionais pelos Alunos em Disciplinas de Programação: Um Estudo de Caso na Licenciatura em Computação. In: Anais do XXI WEI, p. 460-465. SBC.
- Rapkiewicz, C. E. (2006). Estratégias pedagógicas no ensino de algoritmos e programação associadas ao uso de jogos educacionais. Novas Tecnologias na Educação, 1-2.
- Savin-Baden, M. (2000). Problem-Based Learning In Higher Education: Untold Stories: Untold Stories. McGraw-Hill Education (UK).
- Torres, P. L.; Alcantara, P. R.; Irala, E. A. F. (2004) Grupos de consenso: uma proposta de aprendizagem colaborativa para o processo de ensino-aprendizagem. Revista Diálogo Educacional. Curitiba, v. 4, n. 13
- Wing, J. (2006). Computational thinking. Communications os the ACM Vol 49, 33-35.

Pensamento Computacional Praticado com um Jogo Casual SériO no Ensino Superior

Adilson Vahldick^{1,2}, António José Mendes², Maria José Marcelino², Paulo Roberto Farah¹

¹Centro de Educação do Alto Vale do Itajaí (CEAVI)
Universidade do Estado de Santa Catarina (UDESC) – Ibirama, SC – Brasil

²Centro de Informática e Sistemas da Universidade de Coimbra (CISUC)
Universidade de Coimbra – Portugal
{adilson.vahldick, paulo.farah}@udesc.br, {toze, zemar}@dei.uc.pt

Resumo. Nesse artigo é discutida a importância para os alunos das disciplinas introdutórias de programação de cursos de graduação iniciarem o seu aprendizado com o desenvolvimento em habilidades de Pensamento Computacional (PC). Devido a restrições de tempo no semestre letivo, é difícil para os professores oferecerem essa oportunidade através de ambientes de autoria específicos, como Alice e Scratch. Por essa razão, foi criado um jogo casual sério para apoiar no desenvolvimento dessas habilidades. Jogos casuais são caracterizados pelas missões curtas e rápidas, e isso pode ser favorável como atividades extraclasse. Um ciclo de avaliação já foi concluído em um curso de graduação e outro está em andamento noutro curso. Os detalhes do jogo, da sua utilização, a opinião dos alunos e o parecer do professor são apresentados nesse artigo. A principal contribuição desse artigo é elencar alguns princípios que possam auxiliar no desenvolvimento de jogos para o aprendizado de programação através de tarefas com PC.

Abstract. This paper discusses the importance to develop Computational Thinking (CT) skills in undergraduate introductory programming courses. Due to time constraints, it is hard for teachers provide this opportunity through specific authoring environments, like Alice and Scratch. For this reason, it was created a serious casual game to support the development of these skills. Casual games are characterized by fast and short missions, and this can be favorable as extra classes activities. An evaluation cycle was finished in an undergraduate course and another is underway. The details of the game, the opinion of students and the teacher are presented in this article. The main contribution of this work is to list some principles that supports the development of games for programming learning through tasks with CT.

1. Introdução

Os jogos são experiências interativas fáceis e divertidas. Normalmente não exigem um treino ou ajuda externa, a não ser pelos comentários das pessoas que os apreciaram jogando (Prensky, 2001). A aprendizagem baseada em jogos (*Game Based Learning – GBL*) é uma alternativa em que os alunos podem aprender de forma personalizada (o jogo acompanha a evolução do jogador) e auto assistida (o jogador percebe os seus erros ao falhar nas missões, ou o jogo instrui o jogador em como realizar alguma ação) (Whitton, 2010). Os jogos sérios (ou educacionais) oferecem um contexto relevante que

ajuda a transformar conceitos abstratos em mais concretos (Eagle & Barnes, 2009). Usar um jogo envolve naturalmente atividades de resolução de problemas (Schell, 2008). É justamente a falha na apropriação da habilidade de resolução de problemas que resulta no maior fracasso na aprendizagem de programação (Robins et al., 2003).

O aprendizado em programação é um processo em que se exigem transformações cognitivas profundas e desenvolvimento no domínio de habilidades complexas (Gomes & Mendes, 2007). Escrever programas envolve compreender o problema e abstrair um modelo, dividir o problema em partes menores, decidir pela melhor estratégia para resolver cada uma das partes, e aplicar ou adaptar soluções já conhecidas (Winslow, 1996). As habilidades e competências (abstração, decomposição, análise, etc) para desempenhar esse processo é descrito como Pensamento Computacional (PC) (Wing, 2008). A investigação para o aprendizado do PC tem sido focada na educação básica (França & Tedesco, 2015) devido a possibilidade de agregá-lo com outras disciplinas como matemática e biologia.

Apesar de haverem vários jogos disponíveis para a prática e aprendizagem de programação e PC (Vahldick et al., 2014), existem lacunas para a sustentabilidade do processo ensino-aprendizagem. Esses jogos não abrangem com profundidade os assuntos na disciplina, e nem oferecem um sistema de acompanhamento do progresso dos alunos. Este artigo expõe as vantagens no aprendizado do PC no ensino superior nas disciplinas introdutórias de programação e apresenta a experiência no uso de um jogo casual sério para promover esse aprendizado. Jogos casuais exigem pouquíssimas instruções para aprender a jogá-los (Kadle, 2009). Para fins educacionais, essa característica favorece para que o aluno se foque mais nas tarefas instrucionais do que gastar seu tempo em aprender as mecânicas e demais elementos do jogo (Landers & Callan, 2011). O artigo descreve o jogo, quanto aos seus elementos e design instrucional. O trabalho está estruturado da seguinte forma: na segunda seção é justificada a importância no uso de blocos para o aprendizado de programação no ensino superior; a terceira seção fundamenta o uso de jogos no aprendizado de programação e PC; na quarta seção é apresentado o jogo desenvolvido na investigação; na quinta são descritos os experimentos com o jogo; na sexta os resultados dessas experiências; e por fim são realizadas as conclusões sobre os resultados.

2. Blocos para o aprendizado de programação no ensino superior

Tem-se observado a omissão na capacidade de resolver problemas em disciplinas como matemática e ciências nos jovens do ensino médio (Grover & Pea, 2013). Ensinar habilidades computacionais na educação básica brasileira tem sido o desafio e a investigação para tentar amenizar essa situação (França & Tedesco, 2015). Em relação a abordagem do ensino de PC nos cursos de graduação, existem propostas para as disciplinas de matemática visando atenuar os altos índices de evasão e reprovação em tais disciplinas (Barcelos et al., 2015).

Essas propostas ainda são esforços recentes e isolados. Várias gerações de ingressantes nunca tiveram contato com esse conhecimento, o que pode ser apreciado na Tabela 1 onde 66,7% corroboram com esse argumento. Essa tabela apresenta os dados demográficos de dois experimentos, descritos mais adiante, que envolveram 96 alunos de primeiro semestre dos cursos de graduação. As reprovações e evasões são altas nas disciplinas introdutórias de programação. Um dos problemas graves dos alunos é abstraírem a noção de como a máquina interpreta um programa (Du Boulay, 1986). Para apoiar no desenvolvimento dessa noção de máquina abstrata, a estratégia adotada por

ambientes educacionais de programação como Alice e Scratch é que os alunos iniciem com elementos concretos e gradativamente sejam conduzidos para a compreensão dos aspectos abstratos da computação (Dann & Cooper, 2009). Nesses tipos de ambientes o aluno pratica e desenvolve a construção de soluções computacionais por meio de elementos concretos. Para alcançar essa solução é preciso desempenhar tarefas de resolução de problemas, e assim praticar e aprender abstração, decomposição e análise, que são competências essenciais para escrever programas de computador. Esses ambientes utilizam uma linguagem visual baseada na construção dos programas usando blocos (*Block-Based Programming – BBP*). Esses blocos representam fisicamente os comandos das linguagens de programação o que torna mais fácil a iniciação dos alunos principiantes e mantê-los atraídos à tarefa (Weintrop & Wilensky, 2015). Além disso, as animações funcionam como uma ferramenta natural de depuração, mostrando passo-a-passo a execução do programa e o estado das variáveis.

3. Jogos como meios para praticar o pensamento computacional

Alice e Scratch são ambientes de autoria, ou seja, permitem que o aluno livremente construa sua solução. Esses ambientes são ricos de recursos e oportunidades, porém, demandam tempo para o aluno desenvolver os programas. A sua adoção na disciplina de introdução de programação exige que se consumam horas/aula da ementa que foram planejadas no curso para o ensino de resolução de problemas com uma linguagem de programação real.

Jogos vem sendo usados na educação para introduzir e reforçar conceitos, e para praticar como uma alternativa aos exercícios tradicionais (Whitton, 2010). O maior desafio é a manutenção da autoestima e motivação para que os alunos insistam em tentar e vencer o jogo. Entretanto, a diversão permite que o aluno realize as suas tarefas de forma mais fácil, mesmo quando exige mais esforço (Prensky, 2001).

O principal fator para promover a aprendizagem da programação é que o aluno disciplinadamente pratique intensivamente (Robins et al., 2003). Entretanto, muitas vezes falta-lhes motivação para se envolverem nessas tarefas. A prática em resolver os problemas nos jogos, que são mais motivadores do que os exercícios tradicionais, promove a confiança e a experiência, construindo um conjunto estruturado de padrões de soluções, o que lhe será muito útil quando se deparar com problemas usando linguagens de programação reais (Barnes et al., 2007).

Os jogos casuais compartilham essa ideia de terem poucos objetivos nas tarefas. O jogador é estimulado a vencer rapidamente as missões e permanecer envolvido com o jogo. As recompensas com as vitórias podem ser convertidas em novos poderes, personalizações do seu personagem, ou em pontos para classificá-lo num ranking entre outros jogadores (Trefry, 2010). Para a educação, os jogos casuais podem ser úteis, porque, além da facilidade de uso, e poucas tarefas para manter o jogador sempre motivado, eles fornecem *feedback* imediato, a rejogabilidade de missões permite que se aprimorem e testem novas possibilidades, e são fáceis de serem instalados ou são executados em navegador (Landers & Callan, 2011).

4. Jogo NoBug's Snack Bar

Considerando os benefícios e oportunidades expostas nas seções 2 e 3, foi desenvolvido um jogo para o aprendizado e prática das técnicas do PC. A mecânica do jogo é inspirada nos jogos casuais de gestão de tempo. Neste tipo de jogo, as tarefas são divididas em passos sequenciais em que o jogador precisa cumpri-las dentro de um limite de tempo (Trefry, 2010). O contexto do jogo é baseado em uma lanchonete. O jogador controla o

funcionário e o jogo controla os clientes que efetuam pedidos. As ações do funcionário são: perguntar ao cliente o que ele deseja comer e/ou beber, ir às máquinas preparar ou pegar o alimento e/ou bebida, cobrar a conta do cliente, devolver-lhe o troco, e falar (imprimir mensagens na tela). Esse pequeno conjunto de ações permite desenvolver uma gama imensa de missões. Algumas missões são inspiradas em exercícios tradicionais como cálculo de totais, médias, menor e maior valores.

Todas as ações do funcionário são programadas com blocos. Para exemplificar uma missão, a Figura 1a ilustra a solução e a Figura 1b apresenta a área da lanchonete para essa missão. Essa é uma missão em que ainda não foram aprendidas as estruturas de repetição, por isso aparecem blocos duplicados. Seguindo passo-a-passo os blocos da Figura 1a: o atendente vai até o cliente da posição 1 e se ele estiver com sede então executa um conjunto de blocos que pergunta o que deseja beber, vai até a geladeira, pega a bebida, volta ao cliente e entrega a bebida. Depois o atendente vai até a posição 3 e realiza as mesmas ações para esse cliente.

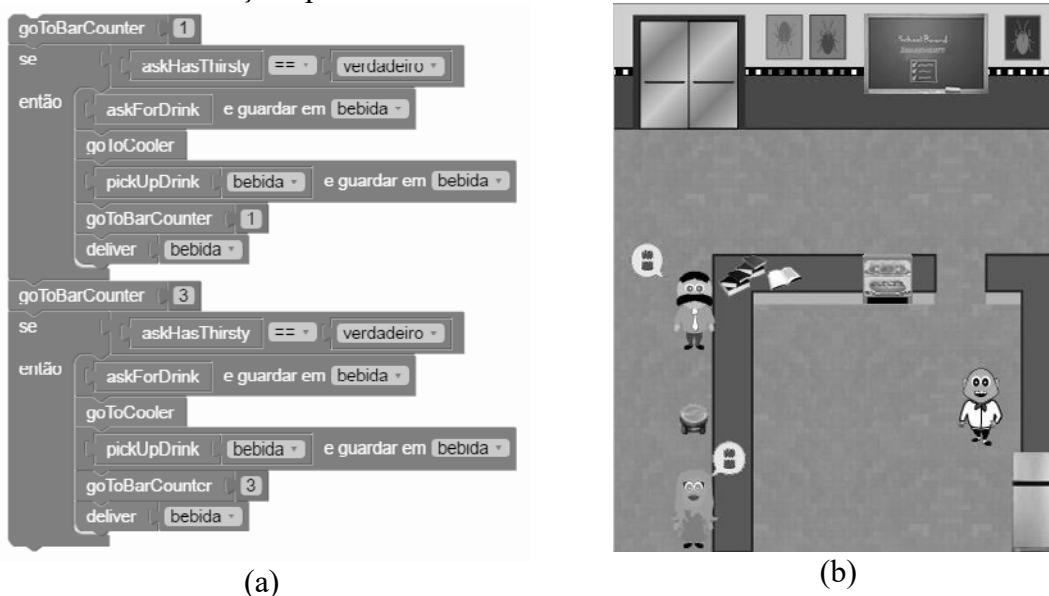


Figura 1. (a) blocos para resolver a missão; (b) área da lanchonete

O ambiente do jogo é apresentado na Figura 2. Ao centro está a área de trabalho, onde o aluno constrói a sua solução; à esquerda está a área de animação, com os botões para executar, depurar, parar e controlar a velocidade de animação, além da área da animação da lanchonete que é refletida conforme a solução vai sendo executada; à direita são apresentadas as variáveis quando o aluno depura a sua solução. Na parte superior existe um quadro com três estrelas, que se refere ao sistema de pontuação na missão. O jogador inicia com três estrelas quando entra a primeira vez em uma missão. Cada estrela corresponde a uma quantidade de pontos. A figura ilustra que cada estrela vale 30 pontos e o jogador já perdeu uma estrela. O consumo das estrelas pode variar entre as missões, mas existem duas formas: (1) por tempo ou (2) por tentativas. Mesmo que todas as estrelas tenham sido consumidas, ainda existe uma pontuação de motivação, que é bem menor, mas é garantida ao jogador quando finalizar a missão. Nas missões por tempo (1) é determinado um prazo para concluí-la. Cada terço desse prazo equivale a uma estrela. Quando encerra o tempo, todas as estrelas estão consumidas, mas o jogador continua a missão até finalizar, para então receber a pontuação de motivação. Nas missões por tentativas (2), como ilustrado na Figura 2, existe um contador de tentativas ao lado das estrelas. Cada vez que o aluno inicia uma execução ou depuração, é consumida uma

tentativa. Pode-se observar que o jogador já consumiu quatro tentativas: três para a primeira estrela e mais uma tentativa na segunda estrela.

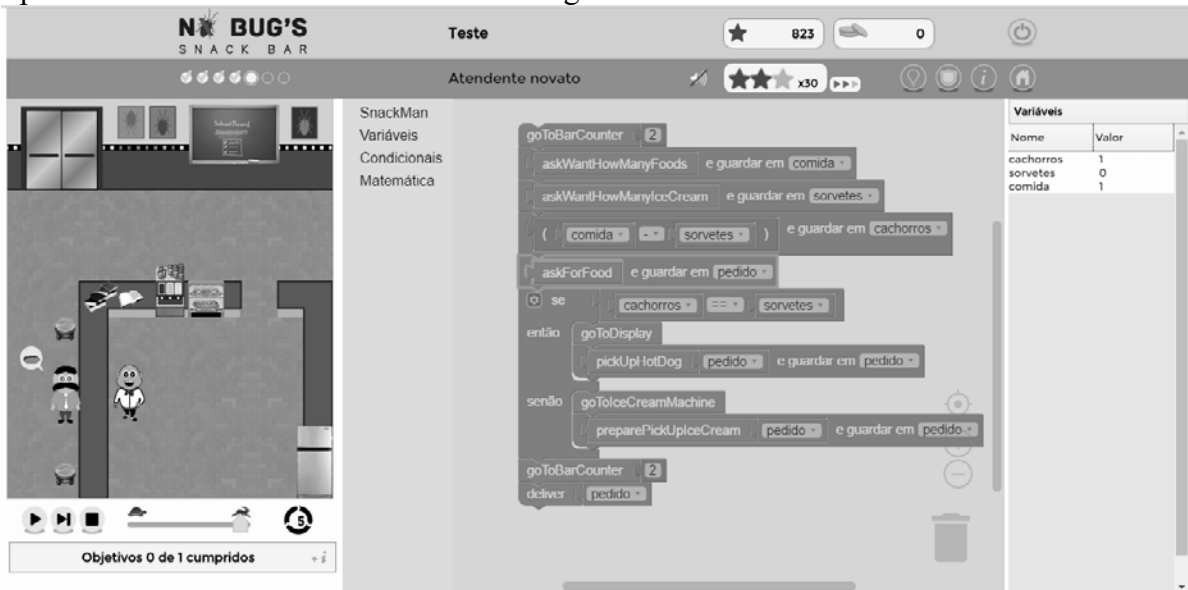


Figura 2. Ambiente do jogo

O jogo apresenta três rankings ordenando os alunos pela pontuação (1), pelo tempo total gasto para cumprir as missões (2) e pela quantidade total de tentativas (3). Além disso, conforme o aluno vai acumulando pontos são liberadas opções para personalizar seu *avatar*: cor dos olhos, estilo e cor dos cabelos/bigode, cor da pele, estilo e cor do chapéu e cor das roupas.

O professor pode acompanhar o desempenho dos alunos de duas formas: consultando as suas conquistas, ou a situação dos alunos nas missões. O jogo tem um sistema de conquistas, ou recompensas, em que fornece medalhas ao aluno em algumas situações definidas pelo professor como, por exemplo, finalizar todas as missões de uma fase dentro de um prazo. A Figura 3 apresenta a tela usada pelo professor para consultar a situação dos alunos. Cada quadrado representa uma missão e indica quantos alunos estão nela. Ao clicar na missão são listados os alunos com suas quantidades de tentativas e tempo consumido na missão. O professor pode consultar a resposta de cada tentativa e contribuir com o aluno enviando-lhe um e-mail ou mesmo discutindo o problema em sala de aula.



Figura 3. Área de administração do professor

As fases foram classificadas em três níveis de dificuldade: (1) introdutório, que abrange de forma superficial todo o conteúdo básico de um assunto; (2) aperfeiçoamento em que se aprofunda no assunto; e (3) domínio que é formada por missões que desafiam com situações complexas e restritivas (por exemplo, limite de comandos e/ou variáveis). As fases introdutórias são pré-requisitos umas das outras, e das fases de aperfeiçoamento e domínio. Por exemplo, se o aluno concluir a fase 2, ele pode decidir jogar livremente entre as fases 3, 4 ou 5.

O processo de aprendizagem dentro das fases também foi revisto. A quantidade de missões foi reduzida em cada fase. Os tipos de tarefas a serem realizadas em cada missão foram inspiradas no desenvolvimento cognitivo proposto na Taxonomia de Bloom (Ferraz & Belhot, 2010). A Tabela 2 apresenta a relação entre cada categoria do domínio cognitivo com o tipo de tarefa nas missões. As fases introdutórias e de aperfeiçoamento possuem aproximadamente oito missões, obedecendo a sequência das categorias, mas variando na sua distribuição. Por exemplo, uma fase pode iniciar com duas missões de conhecimento, e uma de compreensão. Em seguida alterna entre duas de aplicação e análise. Essas fases sempre finalizam com uma missão da categoria síntese. As fases de domínio possuem menos missões (em torno de quatro) e não seguem essa sequência.

Tabela 2. Relação entre as categorias dos domínios cognitivos da Taxonomia de Bloom e os tipos de tarefas das missões

Categorias	Tipos de Tarefas
1. Conhecimento	Corrigir erros: é fornecida a solução com erros. O aluno precisa alterar a variável que está sendo referenciada, trocar operadores de comparações ou lógicos, trocar blocos de posição ou inserir novos blocos.
2. Compreensão	Organizar blocos: todos os blocos já estão disponíveis espalhados na área de trabalho e o aluno precisa organizá-los na sequência correta.
3. Aplicação	Construir iniciando com sugestões: são fornecidos alguns blocos que podem ser o início, o fim ou uma parte central da solução. O aluno cria a solução, inserindo novos blocos, para completar a parte que falta.
4. Análise	Construir: o aluno cria sua solução a partir do zero.
5. Síntese	Construir com restrições: as tarefas têm restrições quanto a quantidade de blocos usados, quantidade de vezes que pode ser usado um determinado bloco, e/ou a quantidade de variáveis que podem ser usadas.

Para melhorar a integração curricular, foi desenvolvido um sistema de conquistas: o aluno recebe um distintivo quando finaliza uma fase até uma data definida pelo professor. Em compensação, o aluno recebe bonificações na nota de uma avaliação (teste/prova). Além disso, o professor utiliza exemplos do jogo (desenho dos blocos ou soluções) nas aulas introdutórias dos assuntos da disciplina.

Ainda existem outras três diferenças significativas entre os experimentos: foram melhoradas a interface com o usuário e a qualidade das imagens (as Figuras 1 e 2 são do segundo experimento), a forma de oferecer *feedback* e de consumir estrelas. No primeiro experimento, durante a missão, o jogo fornecia muitas dicas para o aluno ultrapassar as suas dificuldades e resolver os erros. O consumo das estrelas nas missões era na totalidade pelo tempo. No segundo experimento, o consumo das estrelas é frequentemente por tentativas. O jogo praticamente não fornece dicas. Em contrapartida, foi desenvolvido um novo recurso que mostra a resposta em português estruturado nas missões do tipo “construir” das fases introdutórias e de aperfeiçoamento. O custo para usar esse recurso é consumir todas as estrelas da missão.

6. Resultados

O primeiro aspecto avaliado é quanto à dedicação e interesse dos alunos no jogo. Para isso, foram comparadas as quantidades de alunos que concluíram as missões. A Figura 5 apresenta os resultados entre as duas experiências respectivamente. A Figura 5a apresenta somente as primeiras 19 missões porque a partir dessa somente dois alunos seguiram

adiante e concluíram o jogo. A Figura 5b apresenta a frequência de alunos em todas as missões em que algum aluno já tenha concluído. As duas últimas missões do gráfico correspondem a fase 8 disponibilizada na véspera da redação desse artigo. Como existe uma certa liberdade nas fases em que jogam, é possível notar que a missão 23 tem menos alunos concluintes que as posteriores. A missão 23 é a última da fase 4, que é a fase de domínio no assunto de manipulação de variáveis.

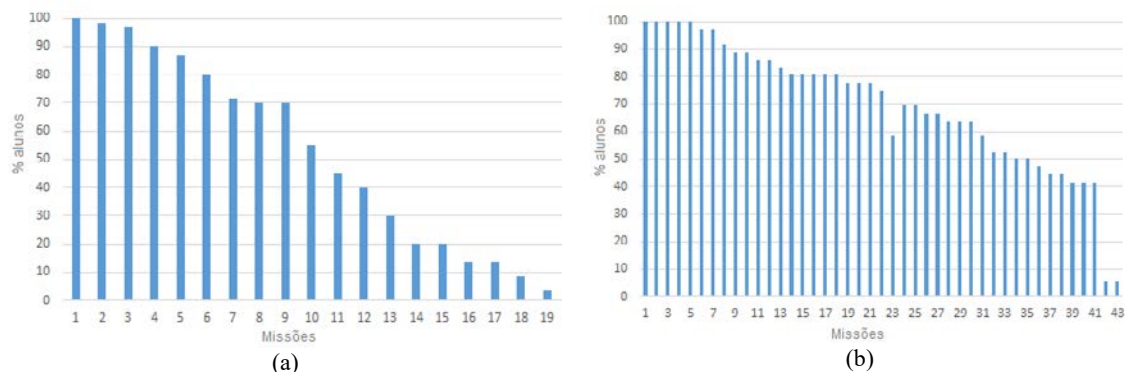


Figura 5. Relação entre alunos e missões concluídas na experiência (a) portuguesa e (b) brasileira

O segundo aspecto avaliado na experiência brasileira foi quanto a opinião dos alunos na aprendizagem percebida por eles. Após os alunos concluírem três fases, o jogo fornece um inquérito com seis questões na escala *Likert* (1-discordo completamente a 5-concordo completamente). A Figura 6 apresenta a distribuição das respostas dos 26 alunos (72,2%) que responderam ao inquérito.

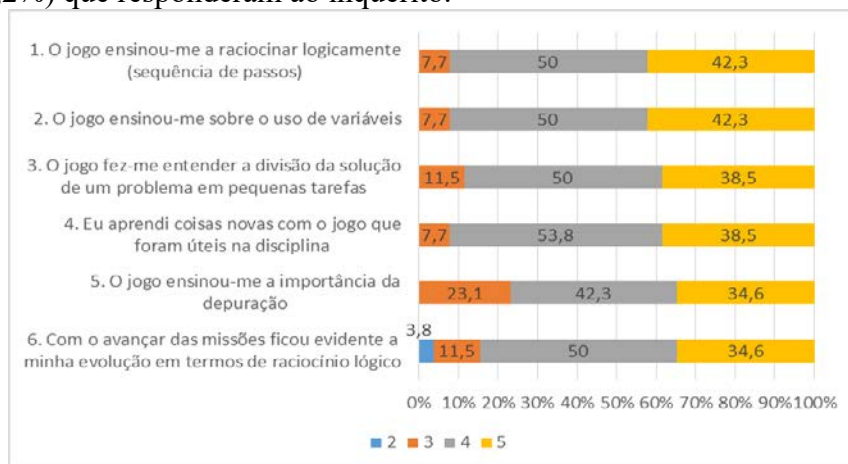


Figura 6. Aprendizagem percebida pelos alunos

O terceiro aspecto avaliado verificou quanto o recurso para mostrar o resultado em português estruturado foi intensamente usado, podendo representar alguma deficiência no processo instrucional. 15 missões (31,3%) têm esse recurso. Esse conjunto de missões foi concluído 380 vezes e o recurso foi usado 40 vezes (10,5%).

O último aspecto avaliado foi a percepção do professor em relação ao comportamento dos alunos. Segundo ele, “os alunos conectam rapidamente o conteúdo que estamos vendo em sala com os conceitos vistos anteriormente no jogo. Como professor tive a impressão que a maioria dos alunos já tinham noção do que eu estava explicando e até se arriscavam com perguntas mais avançadas”.

7. Discussão dos resultados

Existem várias investigações sobre o desenvolvimento de habilidades de PC na educação básica. Contudo, há uma lacuna em pesquisas consoantes ao público-alvo do ensino superior, e como apresentado nesse trabalho, a maior parte desse público não é familiarizado com o assunto. Nesse artigo foi apresentado um jogo casual para a prática dos conceitos de PC. O jogo foi desenvolvido e evoluído através de uma experimentação durante dois meses com 60 alunos de graduação numa universidade portuguesa. Os resultados dessa experiência propuseram mudanças no jogo e na sua forma de usar. Um novo experimento em andamento está sendo conduzido em um curso de graduação numa universidade brasileira. As três semanas de experimento mostraram avanços na motivação em jogar e nas atitudes e comportamentos dos alunos em sala de aula.

A principal contribuição desse artigo para o desenvolvimento de jogos sérios no ensino de programação é quanto ao design instrucional e integração curricular. A forma como estava organizada a sequência de fases proposta na segunda experiência disponibilizava ao aluno um maior conjunto de missões para jogar. Isso influenciou para que o aluno não parasse de jogar caso estivesse com dificuldade em vencer uma missão. No intuito de direcionar de forma coerente, clara e concisa o processo de ensino, a sequência instrucional dentro das fases seguiu a Taxonomia de Bloom que definiu os tipos de tarefas de cada missão. Segundo a opinião dos alunos, eles acreditavam que estavam aprendendo com o jogo (Figura 6). Esse sentimento vem a ser confirmado com o julgamento do professor que observou maior confiança e autoestima nos alunos, a ponto de relacionar essas atitudes com alunos do segundo semestre. Acreditamos que um componente que ajudou na motivação em continuar jogando é a integração do jogo com a disciplina, seja na questão didática ao utilizar os blocos e problemas do jogo nas aulas introdutórias, seja na relação de serem bonificados com pontos extras na próxima avaliação. Outro indício de que os alunos estejam aprendendo é a utilização de um recurso para mostrar o resultado em português estruturado. O recurso consome toda a pontuação da missão. Em 31,3% das missões é oferecido esse recurso. Entretanto, ele foi utilizado somente em 10,5% das vezes que os alunos concluíram essas missões.

O uso do jogo para praticar o PC na disciplina introdutória de programação permitiu de forma mais concreta que os alunos aprendessem conceitos abstratos de programação. Desenvolver a solução com blocos evitou problemas com sintaxe e facilitou a composição de estruturas mais complexas, como os condicionais e repetições. A animação integrada à solução permitiu que os alunos observassem o resultado de cada comando e percebessem a noção de máquina sequencial. A maior parte dos alunos ainda chega no curso superior sem nenhum contato com PC. O jogo facilitou aos alunos aprenderem conceitos de PC antes de alcançarem o mesmo assunto em aula. A vantagem no Ensino Superior em usar o jogo ao invés de um ambiente como Scratch ou Alice é que permitiu mais oportunidades de aprendizado aos alunos sem dispender aulas específicas para isso, e assim o professor continuar cumprindo com a ementa original da disciplina.

Agradecimentos

AV agradece a bolsa de doutorado apoiada pelo CNPq/CAPES – Programa Ciência sem Fronteiras – CsF (6392-13-0) e autorização de afastamento da UDESC (688/13). Os autores também agradecem a disponibilidade de professores e alunos envolvidos no experimento.

Referências

Barcelos, T., Muñoz, R., Villarroel, R., & Silveira, I. (2015). Relações entre o Pensamento Computacional e a Matemática através da construção de Jogos Digitais. In *IV Congresso*

- Brasileiro de Informática na Educação* (pp. 1369–1378).
- Barnes, T., Powell, E., Chaffin, A., Godwin, A., & Richter, H. (2007). Game2Learn: Building CS1 learning games for retention. In *12th SIGCSE Conference on Innovation and Technology in Computer Science Education* (pp. 121–125). Dundee, Scotland.
- Dann, B. W., & Cooper, S. (2009). Education Alice 3: Concrete to Abstract. *Communications of the ACM*, 52(8), 27–29.
- Du Boulay, B. (1986). Some difficulties of learning to program. *Journal of Educational Computing Research*, 2(1), 57–73.
- Eagle, M., & Barnes, T. (2009). Experimental evaluation of an educational game for improved learning in introductory computing. In *40th ACM Technical Symposium on Computer Science Education* (pp. 321–325). Chattanooga, TN.
- Ferraz, A. P. D. C. M., & Belhot, R. V. (2010). Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. *Gestão & Produção*, 17(2), 421–431.
- França, R., & Tedesco, P. (2015). Desafios e oportunidades ao ensino do pensamento computacional na educação básica no Brasil. In *IV Congresso Brasileiro de Informática na Educação* (pp. 1464–1473).
- Gomes, A., & Mendes, A. J. N. (2007). Learning to program-difficulties and solutions. In *International Conference on Engineering Education* (pp. 1–5). Coimbra, Portugal.
- Grover, S., & Pea, R. (2013). Computational Thinking in K-12: A Review of the State of the Field. *Educational Researcher*, 42(1), 38–43.
- Kadle, A. (2009). Casual and Serious Digital Games for Learning – Some Considerations. Retrieved December 27, 2015, from <https://www.upsidelearning.com/blog/index.php/2009/04/18/casual-and-serious-digital-games-for-learning-some-considerations/>
- Landers, R. N., & Callan, R. C. (2011). Casual social games as serious games: The psychology of gamification in undergraduate education and employee training. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious Games and Edutainment Applications* (pp. 399–423). Springer London.
- Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
- Robins, A., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education*, 13(2), 137–172.
- Schell, J. (2008). *The art of game design: A book of lenses*. Burlington, MA: Elsevier Inc.
- Trefry, G. (2010). *Casual game design: Designing play for the gamer in all of Us*. Burlington, MA: Elsevier.
- Vahldick, A., Mendes, A. J., & Marcelino, M. J. (2014). A review of games designed to improve introductory computer programming competencies. In *44th Annual Frontiers in Education Conference* (pp. 781–787). Madrid, Spain.
- Weintrop, D., & Wilensky, U. (2015). To block or not to block , that is the question : Students’ perceptions of blocks-based programming. In *ACM SIGCHI Interaction Design and Children* (pp. 199–208). Medford, MA.
- Whitton, N. (2010). *Learning with digital games: A practical guide to engaging students in higher education*. New York: Taylor & Francis.
- Wing, J. M. (2008). Computational thinking and thinking about computing. *Philosophical Transactions of The Royal Society A*, 366(July), 3717–3725.
- Winslow, L. E. (1996). Programming pedagogy: A psychological overview. *ACM SIGCSE Bulletin*, 28(3), 17–22.

PERTool - Uma Ferramenta Educacional para Manipulação de Redes PERT/CPM

Paulo Henrique Lima Oliveira¹, João Darcy T. Sant'Anna¹,
 Flaviane Vicente T. Sant'Anna¹, Diego Bevilaqua¹,
 Ana Carolina G. Inocêncio¹, Heitor Costa², Paulo A. Parreira Júnior²

¹ Instituto de Ciências Exatas - Universidade Federal de Goiás – Regional Jataí –
 Caixa Postal 03 – 75801-615 – Jataí-GO – Brasil

² Departamento de Ciência da Computação - Universidade Federal de Lavras
 Caixa Postal 3.037 – 37.200-000 – Lavras – MG – Brasil

{ph.lima.tecnico, diego.bevilaqua86, jdtsncomp, flaviane.vicente,
 anacarolina.inocencio}@gmail.com, {heitor, pauloa.junior}@dcc.ufla.br

***Abstract.** Teachers, students and project managers related to computer science courses, engineering and related fields have reported difficulties in teaching and learning some content using some interactive features such as lectures based on books and presentations with slides, as well as costly replication of examples presented in class without the aid of computational tools. This article presents the PERTool, a digital educational tool that allows users to create and manipulate networks PERT / CPM, which is one of the issues addressed in disciplines such as Operations Research and Software Design. This technique is used to optimize cost and time schedules for the development of various types of projects.*

***Resumo.** Professores, estudantes e gerentes de projetos relacionados aos cursos de Ciência da Computação, Engenharias e áreas afins têm relatado dificuldades no ensino-aprendizagem de alguns conteúdos utilizando recursos pouco interativos como aulas expositivas baseada em livros e apresentações com slides, além da onerosa replicação dos exemplos apresentados em sala sem o auxílio de ferramentas computacionais. Este artigo apresenta a PERTool, uma ferramenta educacional digital que permite ao usuário criar e manipular redes PERT/CPM, a qual é um dos assuntos abordados em disciplinas como Pesquisa Operacional e Projeto de Software. Esta técnica é utilizada para otimizar cronogramas de tempo e custo durante o desenvolvimento de variados tipos de projetos.*

1. Introdução

Redes PERT/CPM (*Program Evaluation and Review Technique/Critical Path Method*) são técnicas utilizadas para auxiliar no planejamento, no agendamento, no monitoramento e no controle das atividades de um projeto que estejam relacionados entre si [1]; esse conjunto de atividades é comumente conhecido como **cronogramação** [3]. No contexto de ensino na área de Computação, essas atividades são apresentadas em disciplinas relacionadas à Engenharia de Software; mais especificamente, quando o assunto é a construção e o acompanhamento de cronogramas de projetos de software. A

gerência de projetos é uma atividade essencial e complexa [4], pois o bom gerenciamento não garante o sucesso de um projeto, mas o mau gerenciamento aumenta significativamente as chances de um projeto falhar. Um projeto de sucesso é finalizado no prazo e no orçamento planejados, gerando um produto de qualidade com uma equipe motivada ao final do projeto [4].

No âmbito do gerenciamento de projetos de software, muitas vezes a precisão nos cronogramas é mais importante do que a precisão nos custos, uma vez que o não cumprimento do cronograma pode [4]: (i) reduzir o impacto de um produto/serviço no mercado; (ii) provocar insatisfação dos clientes; e (iii) gerar problemas de integração entre sistemas. Por exemplo, uma equipe de meteorologistas brasileiros pode deixar de realizar experimentos em um período propício para observação da Amazônia, caso o software para análise das imagens de satélite não seja entregue até o mês de setembro, quando as chuvas e a nebulosidade na região diminuem [13]. Sendo assim, ressalta-se a importância do correto entendimento dos conceitos de cronogramação, em especial, das redes PERT/CPM, por parte dos alunos e futuros profissionais. Tal importância é corroborada pela incidência desse assunto em questões recentes das provas de avaliação de desempenho acadêmico no Brasil, como o ENADE (Exame Nacional de Desempenho de Estudantes) - provas de 2011 e de 2014 [11, 12].

Uma ferramenta educacional digital é um conjunto de recursos computacionais projetados com a intenção de serem utilizados no contexto do processo de ensino-aprendizagem [2]. A simulação é um recurso utilizado por este tipo de software e consiste em atividades que colocam o aluno diante do computador como manipulador de situações que “imitam” ou aproximam-se do cenário real. O uso de ferramentas educacionais digitais pode trazer alguns benefícios para o processo de ensino-aprendizagem [5], como por exemplo:

- Os alunos podem relacionar de forma mais adequada o conteúdo teórico, aprendido nas disciplinas, com a prática;
- A interatividade proporcionada por esse tipo de ferramenta pode facilitar o aprendizado de conceitos relacionados a características dinâmicas e/ou abstratas de um sistema;
- A flexibilidade proporcionada por esse tipo de recurso educacional permite que os alunos testem novas entradas para determinado problema e vejam os resultados em tempo real, melhorando a interatividade entre aluno e conceito ensinado; e
- O aluno pode continuar estudando e experimentando novas configurações de entrada/saída para determinado problema fora do ambiente de sala de aula, o que não é tão simples de ser realizado com a utilização de outras ferramentas pouco interativas, tais como, livros-textos, apresentações de slides ou vídeo-aulas.

Assim, neste trabalho, o objetivo é trazer os benefícios comentados anteriormente para a área de ensino de cronogramação. Para isso, foi desenvolvida uma ferramenta educacional para construção e manutenção de redes PERT/CPM, denominada PERTool. O diferencial oferecido por essa ferramenta é o seu enfoque educacional, haja vista que as demais ferramentas existentes para esse tipo de conteúdo [7, 8, 12], comentadas na Seção 3 deste trabalho, não possuem esse enfoque. Além

disso, outro aspecto que ressalta a importância da `PERTool` é o fato de ela ser *open-source*, uma vez que a maioria das ferramentas para se trabalhar com redes PERT/CPM são *plug-ins* para ferramentas proprietárias, tais como, *MS Project*. Por ser *open-source*, o acesso à `PERTool` por professores e por alunos é facilitado, bem como a contribuição de pesquisadores e desenvolvedores para seu aprimoramento.

O restante deste trabalho está organizado da seguinte forma. Na Seção 2, são apresentados conceitos básicos sobre redes PERT/CPM. Na Seção 3, são descritos alguns trabalhos relacionados. Na Seção 4, é apresentada a ferramenta `PERTool`, descrevendo suas funções, Na Seção 5, são discutidos o planejamento e os resultados obtidos da avaliação realizada com alunos de graduação em computação sobre a ferramenta `PERTool`. Na Seção 6, estão presentes as considerações finais e sugestões de trabalhos futuros.

2. Redes PERT/CPM

Redes PERT/CPM são representações gráficas de um cronograma, nas quais são apresentadas a sequência lógica e as interdependências entre as tarefas desse cronograma. Essas redes são compostas por **atividades** e **eventos** [1] em um grafo orientado. Uma atividade compõe uma aresta do grafo, sendo a execução efetiva de uma operação, a qual representa o consumo de tempo e/ou recursos no projeto e possuem duração, podendo estar relacionadas com outras atividades. Um evento corresponde ao vértice do grafo, sendo um marco dentro da rede, representando o início e/ou o fim de uma atividade, e não consome tempo nem recursos em um projeto.

Exemplos de atividades e eventos são apresentados na Figura 1 (todas as figuras apresentadas neste trabalho foram geradas com o auxílio da ferramenta `PERTool`). Por exemplo, a atividade “Pintar porta” é representada com uma seta (aresta direcionada) ligando dois eventos (dois nós), sendo o evento 1 o marco de início da atividade e o evento 2 o marco de fim da atividade. Pode haver dependências entre as atividades de uma rede PERT/CPM, como apresentado na Figura 2. Nesse caso, é possível perceber que a atividade “B” só poderá ser iniciada quando a atividade “A” for terminada, pois o evento 2, que corresponde ao início da atividade “B”, também é o evento final da atividade “A”. Nesse caso, diz-se que as atividades “A” e “B” são executadas sequencialmente.

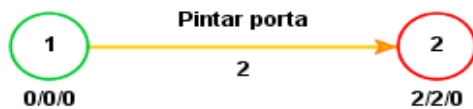


Figura 1 - Exemplo de uma Atividade e seu Evento Inicial e Final

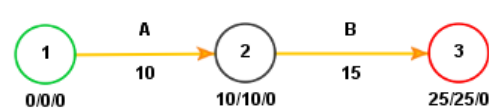


Figura 2 - Exemplo de Dependência entre Atividades

Há atividades que podem ser executadas em paralelo, conforme apresentado na Figura 3. Nessa figura, as atividades “A” e “B” podem ser executadas ao mesmo tempo, pois existem duas setas saindo de um mesmo evento de início (evento 1), caracterizando o paralelismo entre elas. Além disso, há atividades chamadas “atividades fantasmas”, um tipo especial de atividade que não consome tempo nem recursos. Atividades fantasmas são representadas por uma linha pontilhada (Figura 4). Um dos casos em que

uma atividade fantasma deve ser utilizada é para mostrar a interdependência entre múltiplas atividades. Por exemplo, na Figura 4, a atividade “Café com leite” depende das atividades “Moer café” e “Comprar leite”, porém a atividade “Coalhada” depende apenas da atividade “Comprar leite”.

Um caminho de uma rede PERT/CPM é uma rota que segue as setas do evento inicial ao evento final. O comprimento de um caminho é a soma das durações das atividades desta rota; nos exemplos das Figuras 1 à 4, a duração é apresentada abaixo do nome das atividades. O caminho com maior comprimento é denominado **caminho crítico** da rede, pois as atividades sobre esse caminho são críticas, ou seja, são as atividades que consomem a maior quantidade de recursos do projeto. O caminho crítico é representado nas figuras anteriores pelas linhas de cor laranja.

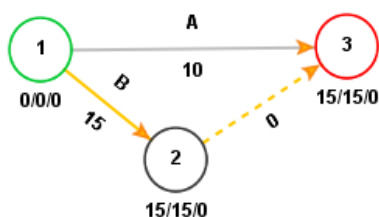


Figura 3 - Exemplo de Atividades Executadas em Paralelo

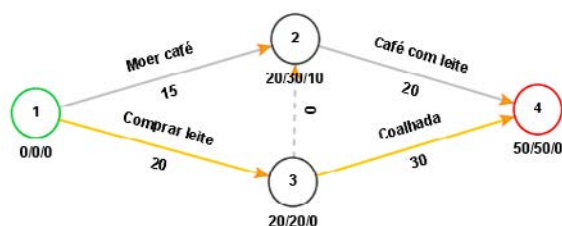


Figura 4 - Exemplo de interdependência entre Atividades

Uma das características interessantes das redes PERT/CPM é o cálculo do tempo de execução. O tempo de execução de uma rede PERT/CPM está relacionado com os conceitos de **tempo cedo**, **tempo tarde** e **folga**. O **tempo cedo** é o tempo necessário para um evento ser atingido, sem atrasos. Para cada evento da rede, o tempo cedo é calculado por $\text{tempo cedo} = \text{máximo}(\text{tempo cedo do evento anterior} + \text{duração})$. O **tempo tarde** corresponde ao tempo limite para atingir um evento (*deadline*), de modo que não haja atrasos no projeto. Para cada evento da rede, o tempo tarde é calculado por $\text{tempo tarde} = \text{mínimo}(\text{tempo tarde do evento posterior} - \text{duração})$. A **folga** corresponde à diferença entre o tempo tarde e o tempo cedo e é calculada por $\text{folga} = (\text{tempo tarde} - \text{tempo cedo})$. As informações colocadas abaixo dos eventos nas redes das figuras anteriores descrevem o **tempo cedo/tempo tarde/folga** de cada evento, respectivamente.

3. Trabalhos Relacionados

Nesta seção, são descritas as principais ferramentas computacionais existentes na literatura, similares à PERTool, destacando-se seus pontos fortes e fracos.

WBS Schedule Pro [7] é um sistema focado no planejamento e no gerenciamento de projetos. Com ele, podem ser tratados diversos conceitos para a concepção e o planejamento de projetos de software; dentre eles, estão as redes PERT/CPM. Inicialmente, esse sistema se chamava PERT Chart Expert, desenvolvido pela empresa Critical Tools [9]. O funcionamento do WBS Schedule Pro acontece por meio da integração do sistema com o MS Project. Algumas de suas limitações é que ele é um sistema proprietário (e pago) e não possui enfoque educacional.

ProjectLibre [12] é um projeto *open-source* disponibilizado de forma gratuita a seus usuários. Como uma de suas funções, podem ser manipuladas redes PERT/CPM. Por um lado, o seu enfoque é o gerenciamento de projetos; por outro lado, não há elementos educacionais que o tornam uma alternativa apropriada para ser utilizado em um ambiente de ensino.

Visio PERT foi o outro software analisado [8], sendo este um projeto implementado pela empresa Microsoft em seu pacote Office. Com o Visio PERT, o usuário consegue organizar tarefas, estabelecer metas, construir e manipular graficamente redes PERT/CPM. Entretanto, esse software não realiza cálculos sobre o caminho crítico, tempo cedo, tempo tarde e folga das redes criadas e não possui foco educacional. Além disso, é dependente de um software proprietário (Microsoft Visio).

Na Tabela 1, é apresentado um resumo das diferenças e das similaridades existentes entre a ferramenta `PERTool` e as demais analisadas nesta seção. Os critérios de comparação utilizados estão na segunda coluna. As demais colunas apresentam os resultados da comparação para cada ferramenta descrita anteriormente na ordem (A) WBS Schedule Pro, (B) ProjectLibre, (C) Visio PERT e (D) `PERTool`. Os símbolos “X” e “-” representam que uma ferramenta “atende” ou “não atende” determinado requisito, respectivamente. Além dos critérios, são apresentadas a portabilidade e a data da última atualização dessas ferramentas. Quanto à portabilidade, foram atribuídos os símbolos “L” para o sistema operacional Linux, “W” para o sistema operacional Windows e “M” para o sistema operacional Mac.

Tabela 1 - Comparativo de Ferramentas para Manipulação de Redes PERT/CPM

#	Critério	Ferramentas			
		A	B	C	D
1	Aborda redes PERT/CPM	X	X	X	X
2	Apresenta os cálculos para caminho crítico, tempo cedo, tempo tarde e folga.	X	X	-	X
3	É software livre	-	X	-	X
4	Apresenta enfoque educacional	-	-	-	X
5	Data da última atualização da ferramenta	2013	2016	2013	2016
6	Plataformas suportadas pela ferramenta	W; M	L; W; M	W; M	L; W; M

De acordo com os dados apresentados na Tabela 1, nota-se que as ferramentas WBS Schedule Pro e VisioPERT estão obsoletas, visto que a data da última atualização foi há 3 anos (2013). Além disso, quanto à portabilidade das ferramentas analisadas, apenas a ProjectLibre e a `PERTool` oferecem suporte para o sistema operacional Linux, um sistema amplamente utilizado, principalmente no meio acadêmico. A ferramenta `PERTool` diferencia-se das demais analisadas por ser a única com enfoque educacional, cujos elementos são descritos na Seção 4. Além disso, é uma ferramenta especializada para construção e manutenção de redes PERT/CPM, está atualizada, é *open-source* e portátil para as plataformas Linux, Windows e Mac, visto que é executada na máquina virtual Java (JVM - *Java Virtual Machine*).

4. A Ferramenta `PERTool`

Nesta seção, são descritas as principais funções existentes na ferramenta `PERTool`, destacando as diferenças entre os demais softwares analisados. Além disso, à medida que as funções são descritas, discussões a respeito dos benefícios educacionais que podem ser obtidos com a utilização dessa ferramenta são apresentados. É importante

ressaltar que PERTool pode ser obtida gratuitamente por meio de *download* e seu código está disponível em um repositório de acesso público: <https://gitlab.com/joaoarcy/perttool>.

A internacionalização (*internationalization - i18n*) do conteúdo é uma característica positiva e importante da PERTool, pois a falta de material para estudo no idioma nativo do estudante pode tornar-se uma barreira para o entendimento do conteúdo abordado. Dessa forma, o aluno deve superar as barreiras naturais impostas pelo aprendizado e as barreiras linguísticas, a qual nem sempre é superada de forma fácil por todos. Apesar das demais ferramentas contemplarem essa característica, elas não possuem material didático sobre redes PERT/CPM internacionalizado. O recurso de *i18n* foi implementado na PERTool, disponibilizando rótulos na interface gráfica e material didático sobre redes PERT/CPM em três idiomas (Português, Inglês e Espanhol) (Figura 5 - A). No material didático disponível na PERTool, foram disponibilizados um tutorial sobre a técnica PERT/CPM, um manual de uso da ferramenta e um conjunto de três redes de exemplo (internacionalizadas). Essa característica é positiva, pois torna PERTool acessível a maior quantidade de pessoas, minimizando a barreira linguística como fator limitante ao aprendizado sobre os conceitos abordados.

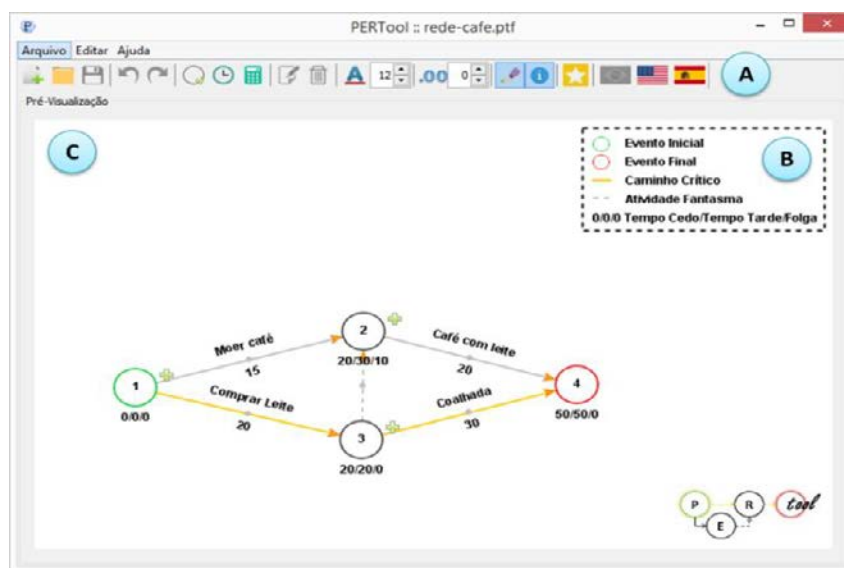


Figura 5 - Interface Gráfica da Ferramenta PERTool

Em PERTool, são oferecidos recursos para auxiliar alunos e professores a criarem visualizações de redes PERT/CPM para utilização em trabalhos acadêmicos, slides, vídeo-aulas, entre outros. Para isso, em PERTool, é disponibilizada a função de exportação das redes criadas para imagens com formato PNG (*Portable Network Graphics*). Além disso, com a PERTool, os usuários podem salvar (armazenar persistentemente) seu trabalho em um arquivo com extensão *.ptf*, sendo interessante, pois permite ao usuário (i) continuar seu trabalho posteriormente; e (ii) compartilhar o seu projeto com outras pessoas, tornando o trabalho de construção de redes PERT/CPM colaborativo. As imagens utilizadas na Seção 2 deste artigo foram exportadas com PERTool.

Outro recurso importante na `PERTool`, não disponível nas demais ferramentas analisadas, é as legendas explicativas sobre os elementos gráficos de uma rede PERT/CPM (Figura 5 - B), as quais podem ser desabilitadas pelo usuário. Essas legendas contribuem de forma positiva do ponto de vista educacional, pois permitem ao usuário relembrar os principais conceitos durante a construção de suas redes.

Realizar os cálculos de tempo da rede (tempo cedo, tempo tarde e folga) manualmente é um dos desafios ao se confeccionar uma rede PERT/CPM por ser onerosos e suscetíveis a falhas. Dessa forma, pode ser vantajoso utilizar `PERTool`, visto que esses cálculos são realizados automaticamente, à medida que usuário constrói sua rede (Figura 5 - C). Esse é um ponto positivo, pois o usuário consegue identificar quais são os gargalos e os atrasos, ao montar a sua rede e em tempo real. Na ferramenta `PERTool`, o caminho crítico da rede é calculado à medida em que a mesma é construída pelo usuário (setas de cor laranja da Figura 5 - C). Isto é interessante, pois o usuário consegue saber, visualmente, quais são os caminhos que consomem mais recursos, sendo uma informação valiosa no decorrer do projeto de software. Ao criar um cronograma de um projeto de software, deve-se começar com um conjunto de tarefas [4]. Se forem utilizadas ferramentas automáticas, a subdivisão do trabalho entra como uma rede ou resumo de tarefas. Dados de esforço, de duração e de data de início são definidos para cada tarefa. Além disso, as tarefas podem ser atribuídas a indivíduos específicos. Com base nessas informações, é gerado um gráfico de Gantt, utilizado para ilustrar as diferentes etapas de um projeto de software, permitindo uma análise do progresso conforme o tempo de execução do projeto. Esse gráfico pode ser exportado como uma imagem no formato PNG em `PERTool`.

5. Avaliação e Resultados

Nesta seção, é discutida a avaliação realizada com a ferramenta `PERTool`, descrevendo o método utilizado, os resultados obtidos e as análises feitas.

5.1 Metodologia de Avaliação

A avaliação de `PERTool` foi conduzida de acordo com um modelo de avaliação de jogos educacionais proposto por Savi *et al.* [15]. Uma das vantagens desse modelo é não consumir muito tempo em sua aplicação, uma vez que, em geral, não há tempo hábil para longas sessões de avaliações nas disciplinas. Com esse modelo, a ferramenta é avaliada para saber se: (i) consegue motivar os alunos a utilizarem o recurso como material de aprendizagem; (ii) proporciona boa experiência nos usuários; e (iii) gera percepção de utilidade educacional (se os alunos acham que a ferramenta é útil e sabem utilizá-la).

Algumas adaptações foram realizadas neste modelo para adequá-lo ao contexto de uma ferramenta educacional. Algumas questões foram adaptadas/excluídas para a avaliação de `PERTool`, por tratarem de elementos específicos de jogos educacionais digitais, os quais não eram adequados ao contexto de ferramenta educacional. Para avaliação, foi proposto um questionário com o intuito de coletar dados dos entrevistados sobre os parâmetros que compõem o modelo de avaliação proposto por Savi [15]. O questionário é composto por 34 afirmações (Tabela 2), para as quais os participantes da avaliação deveriam indicar o quanto concordam com as mesmas, de acordo com uma escala *Likert* de 7 pontos variando de “discordo fortemente” a “concordo fortemente”.

Tabela 2 - Perguntas do Questionário de Avaliação

#	Perguntas	Categoria
1	Houve algo interessante na ferramenta que capturou minha atenção.	M
2	O <i>design</i> da ferramenta é atraente.	M
3	Ficou claro para mim como o conteúdo da ferramenta está relacionado com coisas que eu sabia.	M
4	Eu gostei tanto da ferramenta que gostaria de aprender mais sobre o assunto abordado por ela.	M
5	O conteúdo abordado na ferramenta é relevante para meus interesses.	M
6	Eu poderia relacionar o conteúdo do jogo com coisas que já vi, fiz ou pensei.	M
7	O conteúdo abordado na ferramenta será útil para mim.	M
8	A ferramenta foi mais fácil de entender do que eu imaginava.	M
9	Foi fácil identificar e lembrar dos pontos importantes do conteúdo na ferramenta.	M
10	O conteúdo da ferramenta foi exposto de forma tão interessante que foi difícil tirar a atenção dele.	M
11	Eu consegui entender boa parcela do material da ferramenta.	M
12	Realizar minhas tarefas na ferramenta me deu um sentimento de realização.	M
13	Eu aprendi algumas coisas com a ferramenta positivamente surpreendentes ou inesperadas.	M
14	Os textos de <i>feedback</i> ou outros comentários da ferramenta me ajudaram.	M
15	Eu me senti bem ao completar minhas tarefas na ferramenta.	M
16	Eu não percebi o tempo passar enquanto usava a ferramenta.	E
17	Eu perdi a consciência do que estava ao meu redor enquanto usava a ferramenta.	E
18	Eu me senti mais no ambiente da ferramenta do que no mundo real.	E
19	Em nenhum momento, eu quis abandonar o uso da ferramenta.	E
20	Eu me senti estimulado a aprender com a ferramenta.	E
21	Eu gostei da ferramenta e não me senti ansioso ou entediado.	E
22	A ferramenta me manteve motivado a continuar utilizando-a.	E
23	Minhas habilidades melhoraram gradualmente à medida que ia usando a ferramenta.	E
24	Eu me senti bem-sucedido ao utilizar a ferramenta.	E
25	Eu alcancei rapidamente meus objetivos com a ferramenta.	E
26	Eu me senti competente.	E
27	Eu senti que estava tendo progresso à medida que ia usando a ferramenta.	E
28	Eu gostaria de utilizar esta ferramenta por mais tempo.	E
29	Quando interrompido, fiquei desapontado, pois tive que parar de usar a ferramenta.	E
30	Eu usaria esta ferramenta novamente.	E
31	Há poucas coisas na ferramenta que me irritaram.	E
32	Depois do uso da ferramenta, consigo lembrar de mais informações relacionadas ao tema apresentado nela.	A
33	Depois do uso da ferramenta, consigo compreender melhor os temas apresentados nela.	A
34	Depois do uso da ferramenta, sinto que consigo aplicar melhor os temas relacionados com nela.	A

5.2 Análise dos Resultados

A avaliação da ferramenta `PERTool` foi realizada por 18 alunos do curso Bacharelado em Ciências da Computação da Universidade Federal de Goiás - Regional Jataí (graduação incompleta). Os participantes cursavam a disciplina de Projeto de Software, os quais estudavam sobre cronogramação, incluindo redes PERT/CPM. Para essa avaliação, foi ministrado um treinamento em sala de aula para relembrar os conceitos sobre redes PERT/CPM e apresentar aos alunos as funções de `PERTool`. Durante a avaliação de `PERTool`, cada aluno a utilizou por cerca de 30 minutos, nos quais os alunos deveriam montar algumas redes PERT/CPM propostas pela equipe que conduziu o experimento. Em seguida, os alunos responderam ao questionário de avaliação apresentado na Tabela 2. Por questão de espaço, as respostas dadas pelos participantes foram omitidas, apresentando-se apenas os resultados por categoria (**M** - Motivação, **E** - Experiência do Usuário e **A** - Aprendizagem). Contudo, a planilha completa com os resultados pode ser visualizada em <https://goo.gl/vaydP5>.

As dimensões da categoria “Motivação” são Atenção (perguntas 1 e 2 da Tabela 2), Relevância (perguntas 3 à 7), Confiança (perguntas 8 à 14) e Satisfação (perguntas 12 à 15). De acordo com os resultados da Tabela 3, a avaliação mostrou que houve forte concordância por parte dos avaliadores com relação à Atenção (83,3%) e Confiança (88,9%). A Atenção refere-se às respostas cognitivas dos alunos a estímulos racionais. A

Confiança refere-se à criação de expectativas positivas nos alunos, ao proporcionar-lhes experiências de sucesso decorrentes de suas próprias habilidades e esforço [15]. Além disso, a avaliação mostrou também que Satisfação (69,4% das avaliações positivas) não obteve bons resultados quanto aos demais dessa categoria. Os alunos precisam ter sentimentos positivos sobre a experiência da aprendizagem e isso pode vir por meio de recompensas e reconhecimentos [15]. As propostas para melhorar os resultados desta dimensão podem vir com a gamificação, uma técnica que compreende a aplicação de elementos de jogos em atividades que não sejam jogos [14]. Por fim, a categoria “Motivação” obteve 80,9% de avaliações positivas, o que indica que o aluno se sente motivado ao utilizar a ferramenta educacional como auxílio para aprendizagem dos conteúdos sobre redes PERT/CPM.

Tabela 3 - Média de Avaliações por Categoria

CATEGORIAS / DIMENSÕES		AVALIAÇÕES		
		Positivas (5 até 7)	Neutras (4 apenas)	Negativas (1 até 3)
MOTIVAÇÃO	Satisfação	69,4%	16,7%	13,9%
	Relevância	81,1%	10,0%	8,9%
	Confiança	88,9%	8,3%	2,8%
	Atenção	83,3%	13,9%	2,8%
MÉDIA		80,7%	12,2%	7,1%
EXPERIÊNCIA DO USUÁRIO	Desafio	69,4%	13,9%	16,7%
	Habilidade/Competência	77,8%	12,9%	9,3%
	MÉDIA	73,6%	13,4%	13%
APRENDIZAGEM	Aprendizagem	74,1%	20,4%	5,7%

A categoria “Experiência do Usuário” obteve, em média, 73,6% das avaliações positivas, tendo como pior resultado a dimensão Desafio (afirmações 21 à 31 da Tabela 2), cujo valor foi de 69,4% de avaliações positivas. Esse resultado se deve ao fato de a mecânica dos jogos não estar aplicada à ferramenta PERTool. Para isso, como proposta de solução, também pode ser aplicada a gamificação para obter melhores resultados para essa categoria. Por fim, a categoria “Aprendizagem” obteve 74,1% das avaliações positivas, o que pode indicar a sedimentação dos conhecimentos do usuário utilizando PERTool.

Além dessas questões, uma questão aberta foi disponibilizada para os avaliadores deixarem sua opinião sobre PERTool. Alguns desses comentários são “Ao aumentar o tamanho da fonte, deverá expandir também a distância dos eventos.”, “Quando tento ligar dois eventos sucessivos em duas atividades ele não permite sabendo que o ideal seria criar uma fantasma, por conta de já saber disso falta orientação do software do porque isso acontece, e porque não criar a atividade fantasma automaticamente?” e “A ferramenta poderia auto ajustar a rede pert conforme fosse aumentando o tamanho das letras para que não ficasse desproporcional.”.

6. Considerações Finais

Neste trabalho, foi apresentada a ferramenta educacional PERTool para manipulação de redes PERT/CPM, uma técnica utilizada quando se trata de gerenciamento de projetos de software. Com PERTool, o objetivo é auxiliar os alunos e os professores a conduzirem melhor seus estudos sobre esse assunto. Para constatar o alcance desse objetivo, uma avaliação com alunos de uma instituição federal de ensino superior foi realizada. Nessa avaliação, foi constatada que a PERTool obteve razoável avaliação

positiva, em que as categorias (Motivação, Experiência do Usuário e Aprendizagem) sugeridas por um modelo de avaliação alcançaram a pontuação 80,7%, 73,6% e 74,1% (Positivas - 5 a 7, na escala Likert). Por outro lado, propostas de melhorias foram detectadas com a opinião fornecida pelos avaliadores mediante resposta da questão aberta disponibilizada. Algumas dessas melhorias foram remetidas para sugestões de trabalhos futuros.

Como trabalhos futuros, planeja-se duas novas principais funções: (i) repositório para redes de exemplo, em que usuários poderão compartilhar suas redes PERT/CPM criadas pela ferramenta `PERTool`, armazenando-as em um repositório *online*, o qual será mantido pela equipe de desenvolvimento. Dessa forma, outros usuários poderão encontrar essas redes compartilhadas e utilizá-las com o intuito de se obter novas ideias para criar suas próprias redes, tornando assim, o uso da `PERTool` colaborativo; e ii) criar mecanismos de ajuda para `PERTool` ferramenta para auxiliar o usuário enquanto ele a utiliza. Esse assistente pode ajudar o usuário com dicas ou com correção automática de erros, informando ao usuário qual erro ocorreu e como ele foi resolvido. Além disso, detectar possíveis erros e problemas de usabilidade da `PERTool` e corrigi-los em uma próxima versão.

7. Referências

- [1] H. S. Swanson, R. E. D. Woolsey (1969), “A PERT-CPM Tutorial”
- [2] Sancho, J. M (1998) “Para Uma Tecnologia Educacional”. Porto Alegre, Artmed.
- [3] Lima, A et al. (2010) “Uma Ferramenta Baseada em Simulação Gráfica para Suporte ao Ensino de Redes de Computadores”. In Workshop de Educação em Informática Bahia-Alagoas-Sergipe (WEIBASE).
- [4] Sommerville, “Engenharia de Software”, 8ª edição, Editora Pearson.
- [5] Judith S. Liebman, (1998) Teaching Operations Research: Lessons from Cognitive Psychology. *Interfaces* 28(2):104-110.
- [6] Paas, F. G. W (1992) Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *Journal of Educational Psychology*, Vol. 84, Nº. 4, 429-434.
- [7] WBS Schedule Pro, Disponível em: <http://www.criticaltools.com/download.html>. Acessado em: Abril/15.
- [8] Visio PERT, Disponível em: <http://goo.gl/PVEhGf>. Acessado em: Abril/15.
- [9] Critical Tools, Disponível em: <http://www.criticaltools.com/>. Acessado em: Abril/15.
- [10] INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. ENADE 2014. Disponível em: <http://portal.inep.gov.br/web/guest/enade/provas-e-gabaritos-2014>. Acessado em: Abril/15.
- [11] INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. ENADE 2011. Disponível em: <http://portal.inep.gov.br/web/guest/enade/provas-e-gabaritos-2011>. Acessado em: Abril/15.
- [12] ProjectLibre, Disponível em: <http://sourceforge.net/projects/projectlibre/>. Acessado em: Abril/15.
- [13] Jornal O Globo. “Falta de software empaca dados de satélite brasileiro”. Disponível em: <http://goo.gl/3koFCr>. Acessado em: Abril/15.
- [14] Fadel, Ulbricht, Batista, Vanzin, “Gamificação na educação” (2014), Editora Pimenta Cultural.
- [15] Savi, R., Von Wangenheim, C. G., Ulbricht, V., and Vanzin, T. (2010) “Proposta de um Modelo de Avaliação de Jogos Educacionais”. In: RENOTE - Revista Novas Tecnologias na Educação, vol. 8, no. 3.

Plataforma Arduino como apoio ao ensino de programação no curso de Técnico em Informática integrado

Otávio Soares Papparidis, Matheus Eloy Franco

Área de Computação – Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Campus Machado – 37750-000 – Machado – MG – Brasil

otaviopa@gmail.com, matheus.franco@ifsuldeminas.edu.br

Resumo. Programação de computadores é uma das disciplinas essenciais e introdutórias em cursos de tecnologia da informação. Entretanto, muitos alunos enfrentam dificuldades nessa disciplina, causando altos índices de reprovação e evasão. Diante disso, a plataforma Arduino aparece como um meio de colocar a programação em prática. Algumas pesquisas já validaram essa proposta para estudantes de nível superior. Nesse cenário, este artigo propõe a utilização do Arduino para demonstrar conceitos de programação a alunos de nível técnico do IFSULDEMINAS. Os resultados obtidos se mostram satisfatórios e constata a capacidade dessa abordagem prática para uma possível minimização das taxas de reprovação e evasão.

Abstract. Computer programming is one of the essential and introductory courses in information technology degrees. However, many students struggle in this discipline, resulting in high rates of failures and dropouts. Given this scenario, Arduino platform appears as a way to put into action basic programming concepts. Some researches already validated this proposal for students at undergraduate level. In such context, this paper proposes to use Arduino to demonstrate computer programming topics to secondary technical level students of IFSULDEMINAS. The results were satisfactory and attest the capability of this hands-on approach for a possible reduction in failure and dropout rates.

1. Introdução

Devido à vasta utilização de computadores nas tarefas do mundo atual, disciplinas de computação estão presentes nas matrizes curriculares dos mais variados cursos de todos os níveis de educação, do ensino fundamental à pós-graduação (VALENTE, 1999). Uma dessas disciplinas é a programação de computadores, marcada por dificuldades enfrentadas por educadores e educandos (Castro et al, 2003; Robins, Rountree, J. e Rountree, N., 2003; Delgado et al, 2004; Miliszewska e Tan, 2007; Macedo e Prietch, 2013).

Dentre as dificuldades notadas está a assimilação de conceitos básicos de algoritmos, que formam a base para a programação de computadores. Com a finalidade de melhorar a assimilação de conteúdos, os educadores recorrentemente utilizam objetos de aprendizagem, que podem ser definidos como uma coletânea de itens de conteúdo, prática e avaliação que são combinados com base em um único objetivo de aprendizagem (Gazzoni et al, 2006).

2. Ensino de Programação

O ensino de computação é considerado como desafiador tanto para alunos de nível superior quanto para os do médio/técnico. Para Delgado et al (2004), as dificuldades “aflorem principalmente no transcorrer das disciplinas ligadas ao estudo de algoritmos e programação e consideradas essenciais em todos os currículos de cursos de graduação em computação”.

A complexidade dos conceitos estudados na computação é a razão pela qual Winslow (1996, apud Robins, Rountree, J. e Rountree, N. 2003), reitera que “uma boa pedagogia requer ao professor manter os fatos, modelos e regras simples no início, e somente expandi-los e melhorá-los à medida que os alunos ganham experiência”. Como consequência das dificuldades enfrentadas, as disciplinas de programação enfrentam altas taxas de reprovação, que por sua vez afetam diretamente as taxas de evasão do curso (Castro et al, 2003; Robins, Rountree, J. e Rountree, N., 2003; Miliszewska e Tan, 2007; Macedo e Prietch, 2013).

Segundo Castro et al (2003), o modo como a programação é ensinada tem impactos negativos em cursos que possuem esta disciplina na matriz curricular dos primeiros anos. Ainda segundo os autores, esse impacto também atinge alunos que foram introduzidos à programação ainda no ensino médio.

Para Miliszewska e Tan (2007), outra dificuldade enfrentada pelos alunos “[...] é a necessidade de imaginar e compreender vários termos abstratos que não têm equivalentes na vida real: como uma variável, um tipo de dado, ou um endereço de memória são relacionados a um objeto da vida real?”

Diferentes estudos como o de Crouch et al (2004), miram formas de facilitar o ensino de disciplinas que exigem abstração. Uma das abordagens mais utilizadas pelos educadores é a utilização de demonstrações e exemplos práticos para facilitar a assimilação de conteúdos abstratos. Porém, segundo Crouch et al (2004), o simples uso de demonstrações e exemplos não é suficiente: alunos que observam passivamente demonstrações compreendem tanto quanto alunos que sequer as observam. Em contrapartida, a assimilação do conteúdo prospera quando o aluno se envolve com a demonstração.

Com o intuito de aprimorar esse processo de ensino, Robins, Rountree, J. e Rountree, N. (2003) reiteram a necessidade do uso de demonstrações em sala. Enfatizam também a importância de explicar cada passo do processo de criação do algoritmo a ser estudado. Desta maneira, observa-se que o ensino de programação se mostra como uma disciplina importante que apresenta dificuldades em seu processo de ensino aprendizagem; assim, novas abordagens podem ser utilizadas a fim de melhorar a exemplificação do conteúdo e estimular a motivação dos educandos.

3. A Plataforma Arduino

O cofundador da companhia Arduino, Massimo Banzi, descreve a missão do Arduino como plataforma de prototipagem para “encontrar uma forma simples e rápida de se criar protótipos da maneira mais barata possível.” Arduino é uma plataforma de prototipagem de código aberto baseada em hardware e software de fácil utilização. Placas Arduino são capazes de ler entradas – a luz em um sensor, uma mensagem no Twitter – e transformá-las em uma saída – a ativação de um motor, o acendimento de

um LED. Você pode dizer à placa o que fazer por meio do envio de um conjunto de instruções para o microcontrolador da placa (Banzi, 2009).

Outro fator que faz com que a plataforma Arduino seja versátil é a possibilidade de se acoplar *shields* à placa do Arduino, expandindo assim a sua funcionalidade. As *shields* podem trazer funções de rede (com ou sem fio), visores de LCD, câmeras, leitores de cartão, entre outros (Banzi, 2009).

Dessa forma, a plataforma em questão possui um vasto rol de aplicações, que vai tanto de simples protótipos eletrônicos a grandes automações industriais, ou de pequenos carrinhos de controle remoto à central de voo de veículos aéreos não tripulados (Arduino LLC, 2015).

4. Arduino no Apoio ao Ensino

Como foi observado, o ensino de programação é constantemente alvo de estudos que visam identificar a melhor forma de fazê-lo. A fim de auxiliar o processo de exemplificação do conteúdo, em consonância com os estudos de Crouch et al (2004) e Robins, Rountree, J. e Rountree, N. (2003), os professores lançam mão de equipamentos que não o computador. Dentre as possibilidades, figura a plataforma Arduino.

A simplicidade e a versatilidade, características do Arduino, fizeram com que este fosse objeto de estudos que visam utilizá-lo como ferramenta pedagógica. Segundo Macedo e Prietch (2013); Santos (2013) e Santos et al (2015), a plataforma Arduino faz com que os alunos se sintam mais motivados a aprender programação em comparação com aulas tradicionais que utilizam apenas o computador. Ainda de acordo com dois destes estudos, Prietch (2013) e Santos (2013), a assimilação do conteúdo das aulas é maior quando a plataforma Arduino está presente. Isso evidencia o envolvimento dos educandos com a programação entrando em consonância com a afirmação de Robins, Rountree, J. e Rountree, N. (2003) de que “precisamos motivar os estudantes, envolvê-los no processo, e fazê-los querer aprender a ser programadores eficazes”.

Esse envolvimento dos educandos com a programação na prática vai ao encontro da teoria freiriana e de seus preceitos. Em sua obra, Freire (1987) salienta a importância da educação problematizadora, em contraponto ao que chama de “educação bancária”. Segundo Freire, na educação bancária “não há criatividade, não há transformação, não há saber. Só existe saber na invenção, na reinvenção, na busca inquieta, impaciente, permanente, que os homens fazem no mundo, com o mundo e com os outros”. Ainda segundo o autor, a criatividade é um elemento de papel preponderante na educação problematizadora e deve ser estimulada pelo educador-educando, de forma a potencializar o processo de ensino.

Segundo Narayan (2013), até os dias atuais o sistema de ensino sufoca a criatividade e o espírito inovador dos educandos, sugerindo a necessidade de uma mudança nos currículos escolares. Nesse contexto, o autor diz que a plataforma Arduino “pode desempenhar um papel fundamental no desenvolvimento cognitivo e criativo dos estudantes”. Em seu estudo, Narayan combinou elementos sociais e tecnológicos a fim de potencializar o aprendizado dos estudantes. Essa combinação trouxe resultados positivos, visto que, além de estimular o ensino, os próprios alunos passaram a atuar como educadores e formadores de conteúdo, em consonância com a teoria freiriana.

De acordo com a teoria de construção do conhecimento de Piaget (1967), o desenvolvimento cognitivo dos alunos é fruto da maturação biológica, em adição à interação com o meio, evidenciando a importância da experimentação no processo de aprendizagem. Tendo como base a teoria de Piaget, Papert (1980) cria uma extensão do construtivismo, denominado construcionismo. Ao definir o construcionismo, Papert (1980) afirma que a aquisição do conhecimento tem uma ligação muito forte com a interação com o objeto de estudo, seja refletindo sobre ele, seja agindo sobre o mesmo.

Correa (2012) afirma que Papert usou o termo construcionismo “para mostrar outro nível de construção do conhecimento, aquele que acontece quando o aluno constrói um objeto de seu interesse, como uma obra de arte, um relato de experiência ou um programa de computador”. Ainda de acordo com Papert (1980, apud Correia, 2012) “o aprendiz constrói alguma coisa e o aprendizado só acontece quando o aprendiz constrói algo do seu interesse, porque o envolvimento afetivo torna a aprendizagem significativa”.

Desse modo, dada uma padronização na utilização da plataforma Arduino para fins pedagógicos, um Objeto de Aprendizagem (OA) aflora. Segundo o Instituto de Engenheiros Elétricos e Eletrônicos (2000, apud Gazzoni et al, 2006) os OAs “[...] são definidos como uma entidade, digital ou não, que pode ser utilizada, reutilizada ou referenciada durante o processo de aprendizagem que utilize tecnologia”. Para a elaboração deste OA, Gazzoni et al (2006) mostram a importância de levar em conta alguns aspectos, como a estrita ligação do OA com o conteúdo abordado em sala.

Seguindo esta mesma linha, trabalhos recentes como os de Cardoso e Antonello (2015) e Santos et al (2015) tiveram resultados positivos quanto à utilização do Arduino como OA. Entretanto, ambos analisaram o uso do Arduino por alunos de cursos superiores – no caso, bacharelados em Sistemas de Informação. Já o presente trabalho examina impactos deste OA sobre alunos de nível médio técnico.

5. Materiais e Métodos

Considerando o teor prático e os objetivos pedagógicos deste projeto, a metodologia do presente trabalho seguiu os pilares de pesquisa-ação: Investigação e Ação. Segundo Tripp (2005), a pesquisa-ação tem como ciclo básico as ações de: “Planejar uma melhora na prática, agir para implantar a melhora planejada, monitorar e descrever os efeitos da ação e avaliar os resultados da ação” (Tripp, 2005).

Assim, este trabalho visa alinhar os objetivos apresentados com as ações do ciclo supracitado. Inicialmente foi realizado um levantamento de demanda com três professores de disciplinas de programação de nossa instituição. Este levantamento foi realizado via entrevistas. O conteúdo da entrevista pode ser visto no formulário eletrônico via <<http://goo.gl/forms/8UJul5Smu8>>.

A partir do levantamento realizado com os professores, que culminou na confirmação unânime da demanda, deu-se início aos planos de elaboração dos projetos com Arduino a serem implementados nas aulas de programação dos alunos do IFSULDEMINAS – *Campus* Machado. Duas classes foram escolhidas para a implementação: turmas A e B do 1º ano do curso integrado de Técnico em Informática. O critério para escolha foi o fato de ambas estarem passando pelo processo de introdução à programação, exatamente o público-alvo deste trabalho. A partir daí, foram

estabelecidos três temas do ensino de programação a serem abordados: estrutura sequencial, estruturas condicionais e estruturas de repetição. Não por acaso, os temas escolhidos são a base da programação de computadores.

Para a criação dos modelos foi utilizada a ferramenta 123D Circuits, disponível no endereço <<http://123d.circuits.io>>. Trata-se de uma ferramenta web criada e mantida pela empresa Autodesk. Com ela, o usuário pode construir circuitos eletrônicos e simular o seu funcionamento. Além de ser gratuito, o simulador proporciona economia devido ao custo zero de investimento e manutenção. Como alguns alunos não têm familiaridade com eletrônica, é interessante que sempre simulem os projetos antes de montá-los efetivamente, para evitar a “queima” de placas e componentes. Isso deixa os alunos livres para trabalhar sua criatividade sem terem medo de provocar danos, e não abram mão da prática com os componentes reais, considerando que os projetos seriam replicados com os componentes reais nos casos das simulações bem-sucedidas.

Apesar de estar disponível apenas em inglês, o idioma não é um grande obstáculo para a utilização do simulador, pois sua interface é rica em recursos gráficos em oposição ao uso de textos, além de utilizar o formato “arrasta-e-solta” para a inclusão e manipulação dos componentes. A ferramenta ainda conta com mais duas opções de organização de tela: a utilizada para edição do código da placa Arduino (*code editor*) e a utilizada para a simulação propriamente dita, que pode ser iniciada pelo botão com um sinal de *play* e a inscrição “*Start Simulation*”. A figura 1 representa a interface no modo de edição de código.

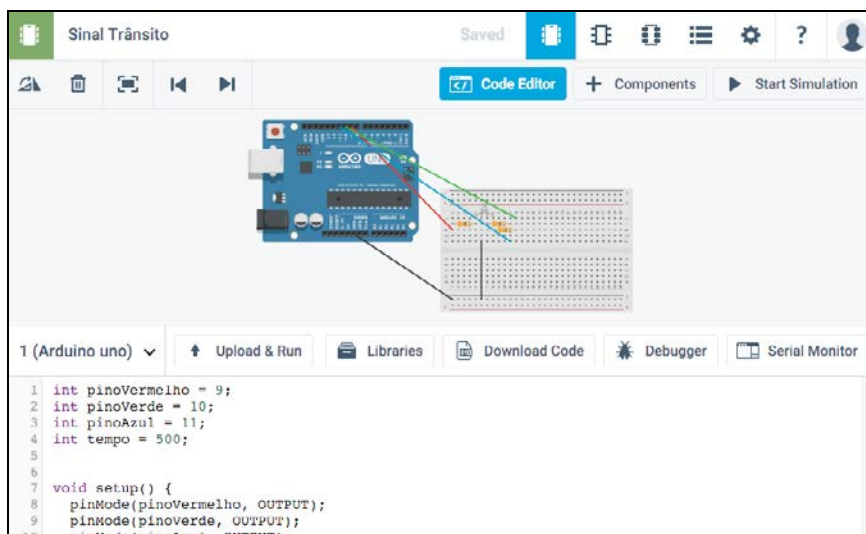


Figura 1: Interface do simulador 123D Circuits – Código.

Fonte: Autor (captura de tela)

Após a definição dos projetos e do simulador a ser utilizado, foram criados três projetos. Para exemplificar a estrutura sequencial, foi criado um simulador de semáforo. Para as estruturas condicionais, foi elaborado um simulador de poste de iluminação pública, que acende durante a noite e se apaga durante o dia. Por fim, para aulas sobre as estruturas de repetição foi utilizado um LED multicolor que troca de cor repetidamente, mostrando um amplo espectro de cores.

6. Descrição dos Projetos Arduino

Todos os projetos apresentados nesta seção encontram-se disponíveis publicamente nos seus respectivos endereços na plataforma 123D Circuits, por meio dos endereços: Projeto 1 – Estrutura Sequencial: <<https://goo.gl/JdIcqS>>, Projeto 2 – Estrutura Condicional: <<https://goo.gl/QJVRaJ>>, e Projeto 3 – Estrutura de Repetição: <<https://goo.gl/X3n6QT>>. Nas páginas é possível visualizar a estrutura física, a lista de componentes utilizados, o código-fonte e a simulação do funcionamento do projeto.

Os modelos elaborados foram apresentados aos alunos. A demonstração visava deixar claro que a estrutura utilizada no algoritmo daquele projeto específico – por exemplo, a estrutura “Se-Então-Senão” – era exatamente a mesma que eles haviam aprendido ao estudá-la utilizando o método tradicional na tela do computador.

Como a estrutura física necessária para o funcionamento dos projetos 1 e 3 era a mesma, foi possível a utilização de apenas uma *protoboard* (ou placa de ensaio) para a demonstração de todos os três projetos, ficando uma montagem à esquerda e outra à direita. Cada projeto resultou em uma demonstração, identificada pelo mesmo número daquele. Para trocar de um projeto para outro bastava reconectar os fios na placa conforme etiquetas e fazer o carregamento do código correspondente ao projeto no Arduino. Todo o processo de troca leva cerca de um minuto, não atrapalhando, portanto, o andamento da aula. A figura 2 mostra a *protoboard* contendo as montagens supracitadas.

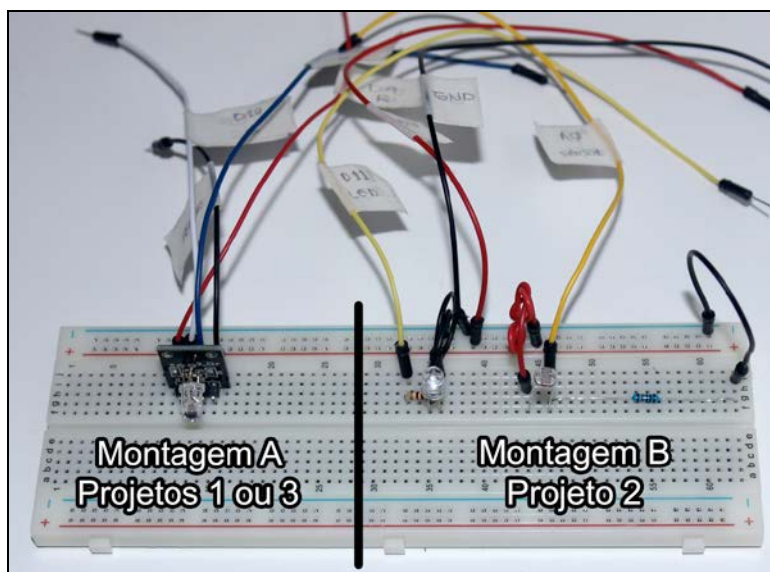


Figura 2: *Protoboard* com os projetos montados.

Fonte: Elaborada pelo autor.

6.1. Demonstração 1 – Estrutura Sequencial

A demonstração 1 tinha como objetivo explicar a estrutura sequencial de programação utilizando-se da sequência lógica de um semáforo. O código continha basicamente três pontos de execução: luz verde, luz amarela e finalmente a luz vermelha. Ao final, a própria estrutura de execução da plataforma Arduino se incumbia de repetir a sequência.

Ao apresentar o funcionamento do algoritmo aos alunos, foram feitas algumas mudanças sutis no código a fim de atestar o papel desempenhado por cada trecho em específico: mudança no tempo de cada cor, inverter cores, pular uma das cores, etc. Dessa forma, os alunos puderam compreender que cada linha da sequência em questão produzia um efeito durante a execução do algoritmo.

6.2. Demonstração 2 – Estrutura Condicional

A demonstração 2, assim como na anterior, fazia uma analogia com um elemento que está presente no dia de todos que vivem em um ambiente urbano: um poste de iluminação. Assim como os postes da cidade, que acendem automaticamente quando está escuro e se apagam quando o sol nasce, o LED presente neste projeto acendia e apagava conforme a luminosidade ambiente.

A fim de reforçar o conceito da estrutura condicional, foram feitas várias mudanças no código, que tinha como base uma única estrutura “Se-Então-Senão”. Salientou-se como uma simples mudança de um símbolo menor (<) para maior (>) invertia toda a lógica de funcionamento: o LED acendia quando a sala tinha as luzes acesas e se apagava quando estava com as luzes apagadas.

6.3. Demonstração 3 – Estrutura de Repetição

A terceira e última demonstração trazia um LED que mudava de cor rapidamente, exibindo um grande espectro de cores formadas pela manipulação das cores-fonte vermelho, verde e azul. Tal demonstração teve o propósito não só de exemplificar uma estrutura de repetição, mas também de mostrar como essas estruturas se comportam quando estão aninhadas dentro umas das outras. Para tal, o código trazia três estruturas “Para” (*for*) aninhadas de tal forma que cada uma delas controlava uma das cores-fonte.

Sendo definidos valores máximos para a intensidade de cada cor-fonte, o aluno pôde observar cada mudança destes valores e seu efeito sobre a formação da cor final, exibida pelo LED. Para facilitar essa percepção, durante a apresentação diminuiu-se o tempo de mudança entre as cores, além de exibir a intensidade de cada cor-fonte em tempo real na tela do computador, projetada para toda a sala.

7. Avaliação e Resultados

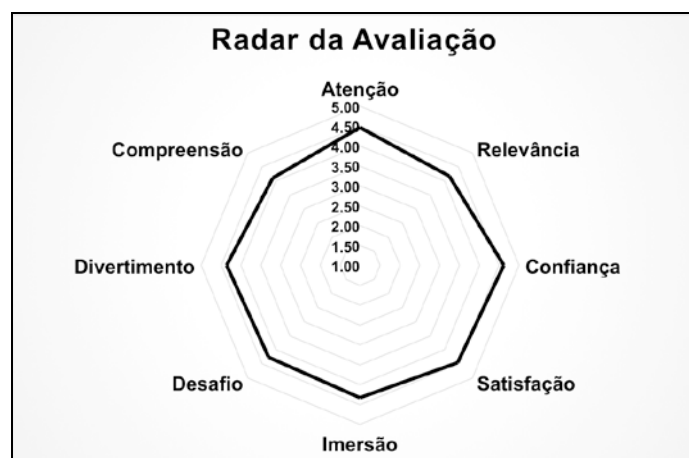
A proposta pedagógica foi avaliada por um total de 71 alunos. Para essa avaliação, foi aplicado um questionário semiestruturado, disponível em <<http://goo.gl/forms/8UJul5Smu8>>, que teve como referência o modelo de avaliação de jogos educacionais proposto por Savi et al (2010). Tal modelo, conforme relatado pelos autores, segue os “princípios da taxonomia de Bloom para avaliação do impacto na aprendizagem do aluno”, sendo a escala Likert de 5 pontos utilizada para análise dos dados. O quadro a seguir mostra as questões utilizadas no formulário e seus respectivos aspectos avaliados, divididos em três grandes grupos: motivação, experiência e conhecimento.

Quadro 1. Aspectos avaliados pelo questionário.

Motivação	Atenção	Houve algo interessante no início das demonstrações com o Arduino que capturou minha atenção.
	Relevância	Ficou claro para mim como o conteúdo das demonstrações com Arduino está relacionado com coisas que eu já sabia.
	Confiança	A quantidade de informações foi adequada aos exemplos.
	Satisfação	Eu aprendi algumas coisas com os exemplos que foram surpreendentes ou inesperadas.
Experiência	Imersão	Senti-me estimulado a aprender com o Arduino.
	Desafio	Eu gostei de aprender com o Arduino e não me senti ansioso ou entediado.
Conhecimento	Divertimento	Gostaria de ver mais demonstrações com Arduino para complementar meu aprendizado em programação.
	Compreensão	Depois de ver exemplos com o Arduino consigo compreender e aplicar melhor os temas apresentados nos exemplos.

Como proposto, ao final das apresentações nas salas de aula, os alunos foram convidados a responder o questionário; cerca de 80% dos alunos o fizeram. O interesse demonstrado pelos alunos em utilizar o Arduino para criar seus próprios experimentos foi nítido. Uma das preocupações iniciais do presente trabalho foi que, embora o Arduino seja uma plataforma de simples programação, talvez fosse visto como algo complicado para os alunos do nível técnico. Entretanto, não foram apresentadas dificuldades quanto a isso. Um dos alunos ainda vinculou a simplicidade do Arduino à compreensibilidade do mesmo: “Como demonstrado, o Arduino é bem fácil de aprender e aplicar, ou seja, bem fácil de se compreender”. Outro destacou o estímulo proporcionado pelo Arduino como algo positivo: “Arduino é capaz de estimular a conhecer ainda mais a programação”. As opiniões vão ao encontro das observações de Prietch (2013) e Santos (2013) no tocante ao papel do Arduino como facilitador da compreensão do conteúdo.

Ademais, os resultados obtidos nos segundo e terceiro eixos (experiência e conhecimento) evidenciaram o envolvimento dos alunos com o OA, além do estímulo à criatividade dos mesmos, fatores preponderantes para um ensino eficaz (Narayan, 2013; Freire, 1987). A figura 3 mostra um gráfico do tipo radar apresentando os resultados de todos os aspectos avaliados conforme o quadro 1.

**Figura 3: Gráfico Radar das Avaliações.**

Fonte: Elaborada pelo autor.

8. Considerações Finais

Diante dos dados obtidos, constatou-se que a abordagem proposta tem, de fato, capacidade de potencializar o ensino de programação não somente no ensino superior, mas também no ensino técnico integrado. O fator motivacional manifestado pelos alunos foi preponderante dentre os aspectos que confirmam essa capacidade. Além de se sentirem motivados, os educandos mostraram ter uma oportunidade de externar sua criatividade com a plataforma Arduino, seja em trabalhos propostos pelo professor, seja em projetos independentes.

O trabalho deixa um legado que poderá ser analisado e, eventualmente, incorporado a planos de ensino de qualquer curso que tenha em sua matriz curricular disciplinas de programação de computadores e tenha uma realidade similar às encontradas nos cursos de nossa instituição.

Como possíveis trabalhos futuros, sugere-se a utilização contínua da abordagem experimentada neste trabalho, fazendo com que os alunos tenham contato com a prática da programação com Arduino do início ao fim da disciplina, e não apenas pontualmente como foi neste estudo.

Referências

- Arduino LLC. Arduino Forum. Arduino Website, 2015. Disponível em: <https://forum.arduino.cc>. Acesso: Dezembro 2015.
- Banzi, M. Getting Started with Arduino. [S.l.]: O'Reilly Media, 2009.
- Cardoso, R. e Antonello, S. L. Interdisciplinaridade, programação visual e robótica educacional: relato de experiência sobre o ensino inicial de programação. Workshops do IV Congresso Brasileiro de Informática na Educação (CBIE). Maceió, 2015.
- Castro, T., Castro, A., Menezes, C. Boeres, M. e Rauber, M. Utilizando Programação Funcional em Disciplinas Introdutórias de Computação. XI Workshop de Educação em Computação – WEI. Campinas, 2003.
- Correa, M. D. C. Tecnologia e Práticas Educativas: O Projeto Mundo do Saber. Americana, 2012.
- Crouch, C., Fagen, A. P., Callan, J. P. e Mazur, E. Classroom demonstrations: Learning tools or entertainment? American Journal of Physics, 72, 2004. 835-838.
- Delgado, C., Xexeo, J. A. M., Souza, I. F., Campos, M. e Rapkiewicz, C. E. Uma Abordagem Pedagógica para a Iniciação ao Estudo de Algoritmos. XII Workshop de Educação em Computação – WEI. Salvador, 2004.
- Freire, P. Pedagogia do oprimido. 17ª. ed. Rio de Janeiro, 1987.
- Gazzoni, A., Canal, A. P, Falkembach, G. A. M., Fioreze, L. A., Pincolini, L. B. e Antoniazzi, R. Proporcionalidade e semelhança: Aprendizagem via objetos de aprendizagem. Revista Novas Tecnologias na Educação. Porto Alegre, 2006.
- IEEE Standards Association. IEEE Learning Technology Standards Committee. IEEE.org, 2000. Disponível em: <https://ieee-sa.imeetcentral.com/ltsc>.

- Macedo, R. S. D. e Prietch, S. S. Utilizando a Placa Arduino como Ferramenta de Apoio ao Ensino de Programação. Anais da IV Escola Regional de Informática da SBC – Regional de Mato Grosso. 37-42. Alto Araguaia, 2013.
- Miliszewska, I. e Tan, G. Befriending Computer Programming: A Proposed Approach to Teaching Introductory Programming. Issues in Informing Science and Information Technology. Melbourne, 2007.
- Narayan, V. Technology as a creative partner: Unlocking learner potential and learning. 30th ascilite Conference. 612-621. Sydney, 2013.
- Papert, S. Mindstorms. Children, Computers and Powerful Ideas. New York, 1980.
- Piaget, J. e Inhelder, B. The Child's Conception of Space. New York: W. W. Norton & Company, 1967.
- Robins, A., Rountree, J. e Rountree, N. Learning and teaching programming: a review and discussion. Computer Science Education, 13, 2003. 137-172.
- Santos, A., Gorgônio, A., Lucena, A. e Gorgônio, F. A Importância do Fator Motivacional no Processo Ensino-Aprendizagem de Algoritmos e Lógica de Programação para Alunos Repetentes. XXIII Workshop sobre Educação em Informática – WEI. Recife, 2015.
- Santos, R. R. D. Utilização de Plataformas de Hardware Aberto no Ensino: Um Estudo de Caso na Disciplina de Introdução à Engenharia de Computação. International Journal of Computer Architecture Education (IJCAE), v. 2, p. 37-40, 2013.
- Savi, R., Wangenheim, C. G., Ulbricht, V., Vanzin, T. Proposta de um Modelo de Avaliação de Jogos Educacionais. RENOTE - Revista Novas Tecnologias na Educação, v. 8, n. 3. Porto Alegre, 2010.
- Tripp, D. Pesquisa-ação: uma introdução metodológica. Educação e Pesquisa, v. 31, 443-466. São Paulo, 2005.
- Valente, J. A. O computador na sociedade do conhecimento. UNICAMP/NIED. Campinas, 1999.
- Winslow, L. E. Programming pedagogy - a psychological overview. SIGCSE Bulletin, 28, n. 3. 17-22. New York, 1996.

POOGame: Um Jogo Sério para o Ensino de Programação Orientada a Objetos

Leuson M. P. da Silva , Brendo C. Bonfim , Rogério C. Silva , Jefferson B. da Silva ,
Wellington L. Moura , Carla I. M. Bezerra , Paulyne M. Jucá

¹Universidade Federal do Ceará (UFC) – Campus Quixadá
CEP: 63900-000 – Quixadá – CE – Brazil

{leuson, brendocbb, rogerio_carvalho15, jefferson_barbosa,
wellington.tylon}@alu.ufc.br, {carlailane, paulyne}@ufc.br

Abstract. *This paper describes the serious game developed to support the teaching of basic object oriented development (POO in portuguese) concepts, known as POOGame. It is a RPG based game where the player controls a character that summons and controls a creature using JAVA command during battles to combat enemies. The game was developed during an introduction to game development course using a game development process. The paper also presents the results of the POOGame evaluation that demonstrates that the game has the potential do support the POO basic concept learning.*

Resumo. *Este trabalho tem como objetivo apresentar a elaboração de um jogo sério para o ensino de programação orientada a objetos (POO), denominado POOGame. O desenvolvimento do jogo foi baseado em jogos de RPG, de forma que o jogador controla um personagem e a partir de batalhas é possível invocar criaturas e controlar suas ações utilizando comandos baseado na linguagem JAVA. O jogo foi elaborado e avaliado no contexto da disciplina de Introdução ao Desenvolvimento de Jogos de um curso de Engenharia de Software, seguindo uma metodologia para o desenvolvimento de jogos. Os resultados da avaliação do POOGame indicaram evidências de que o jogo contribui para o aprendizado de POO.*

1. Introdução

A Programação Orientada a Objetos (POO) mostra-se um paradigma de programação influente, pelo qual a maior parte dos cursos da área de computação incluem a POO como parte de seus currículos [Figueiredo et al. 2015]. No entanto, o ensino de POO não é uma tarefa trivial.

Há alertas sobre as as dificuldades encontradas por alunos e professores no ensino de POO [Holland et al. 1997]; os autores apontam que é difícil para os alunos entenderem conceitos abstratos como classes, instâncias, encapsulamento, herança e construtores. Apesar de muitos esforços terem sido feitos desde então, as disciplinas de programação ainda são responsáveis por um alto índice de reprovação nos cursos de computação.

Algumas das tentativas mais recentes [Lopes 2012] e [Al-Linjawi and Al-Nuaim 2010] envolvem aliar o interesse dos alunos por jogos

como motivação para o ensino e prática de exercícios de programação. Jogo Sériu ou *Serious Game*, segundo Rodrigues et al. (2012) é aquele “cujo objetivo principal é o aprendizado, mas que também utiliza o entretenimento como fator importante de integração e imersão do jogador no ambiente de aprendizado”.

Nesse contexto, esse artigo apresenta um jogo sério desenvolvido para auxiliar na fixação de conceitos relacionados à OO, denominado POOGame. O artigo também apresenta os resultados da avaliação realizada com alunos que apontam pontos positivos e pontos de melhoria ao jogo proposto. Este trabalho está organizado da seguinte forma: os trabalhos relacionados são discutidos na seção 2, enquanto o jogo POOGame, bem como as tecnologias utilizadas em seu desenvolvimento, são apresentadas na seção 3. Na seção 4 é apresentada a avaliação realizada do jogo. As conclusões e trabalhos futuros são apresentados na seção 5.

2. Trabalhos Relacionados

Vários jogos têm sido propostos na literatura para apoiar o processo de aprendizagem de alunos na área de computação [Voss et al. 2013, Lopes 2012, Al-Linjawi and Al-Nuaim 2010]. Voss et al. (2013) apresentam a proposta de um jogo sério também desenvolvido em ambiente virtual 3D, mas aplicado ao ensino de redes de computadores. Trata-se de um jogo de fases sequenciais com diferentes níveis de dificuldade. A parte inicial do jogo funciona como uma parte mais teórica, enquanto as últimas exploram mais as habilidades/capacidades práticas do jogador.

Contudo, com relação ao ensino da disciplina de Programação Orientado a Objetos (POO), foram identificados na literatura poucos jogos educacionais [Lopes 2012, Al-Linjawi and Al-Nuaim 2010]. Lopes (2012) também se atenta em desenvolver um jogo para o ensino dos conceitos de OO. Para tanto, o autor utiliza as técnicas de jogos de cartas convencionais para desenvolver seu trabalho; entretanto, não há o uso de quaisquer suporte de meios eletrônicos. Al-Linjawi e Al-Nuaim (2010) apresentam um jogo também em ambiente 3D para o ensino dos conceitos de OO. O jogo desenvolvido busca ensinar estes conceitos por meio da criação de mundos virtuais e, posteriormente, da colonização destes mundos por meio de código informado pelo jogador. O processo para a escrita de código ocorre por meio de um editor onde há instruções que encontram-se disponíveis ao jogador. Contudo, a presença permanente dessa forma de ajuda pode não beneficiar a autonomia do jogador de pensar sobre os conceitos aprendidos, considerando que parte dos termos a serem informados no código (palavras-chaves, por exemplo), encontram-se sempre disponíveis ao jogador durante o jogo.

Além dos jogos identificados na literatura para o ensino de POO, também foi encontrada uma abordagem gamificada do ensino de POO, proposta por Figueiredo et al. (2015). A abordagem utilizou o *Gamification Design Framework* [Werbach and Hunter 2012] como base metodológica para a produção do design da gamificação da disciplina de POO. Foi utilizada a plataforma Moodle como plataforma para gamificação, aplicando essa abordagem em duas (2) turmas. No entanto, essa abordagem não propõe um jogo sério para melhorar a motivação da aprendizagem dos conceitos de OO.

3. O Jogo POOGame

Nesta seção, será apresentado o planejamento e execução do jogo denominado POOGame. O jogo tem o objetivo de ensinar conceitos básicos de Programação Orientada a Objetos.

3.1. Desenvolvimento do Jogo

O jogo foi desenvolvido utilizando a *engine* de desenvolvimento de jogos *Unity* 3D¹. Uma *engine* de jogos pode ser entendida como uma ferramenta ou biblioteca que encapsula um conjunto de funcionalidades que facilitam a criação de jogos. Como a *Unity* oferece suporte para mais de 20 plataformas, e conseqüentemente, permite que um mesmo projeto possa ser compilado e executado nessa variedade de plataformas. Devido a essas características a *Unity* foi a engine escolhida para apoiar o desenvolvimento deste projeto. Para o desenvolvimento dos elementos da interface gráfica e do terreno do mundo virtual, a ferramenta Blender² foi utilizada. Para realizar o processo de controle de mudanças e gerenciamento de configuração foi utilizado a versão gratuita do GitHub³.

Como o projeto foi desenvolvido ao longo da disciplina de Introdução ao Desenvolvimento de Jogos, foi adotada uma abordagem de desenvolvimento iterativa e incremental para que as mudanças identificadas no escopo do projeto pudessem ser realizadas. Inicialmente foi definido o *game design* do jogo, e em seguida, foi realizada uma avaliação inicial da proposta por meio da técnica de prototipação em papel. Nessa avaliação foi possível ter um *feedback* direto dos participantes envolvidos, como pontos de melhorias e novas ideias. Depois iniciou-se o desenvolvimento do jogo, que encontra disponível online⁴. Após o término do desenvolvimento, ocorreu a apresentação e avaliação final do jogo, apresentada na seção 4.

3.2. Descrição do Jogo

O jogo foi modelado com base em outros jogos de RPG (*Role Playing Game*), em que o jogador controla um personagem, e a partir de batalhas é possível invocar criaturas e controlar suas ações utilizando comandos baseados na linguagem Java.

Na estória, um mago fica preso em uma caverna e precisa da ajuda do jogador para sair de maneira segura. O mago está fraco e sem poderes, mas pode treinar o jogador para combater as criaturas presentes no mundo exterior. Na primeira fase, o mago vai ensinar o jogador a identificar habilidades corretas (classes Java sintaticamente corretas) que serão utilizadas posteriormente pelo jogador no controle de um *pet* (instanciação e controle de um objeto de uma Classe Aranha, por exemplo) para combater os inimigos na floresta. O jogo acaba quando o jogador consegue derrotar todos os inimigos presentes na floresta.

Fase 01 - Caverna: O objetivo principal desta fase é permitir que o jogador identifique classes sintaticamente corretas, de acordo com os padrões definidos para a linguagem Java [Gosling et al. 2013]. Esse objetivo será satisfeito por meio da identificação de uma classe, a partir de um conjunto de classes possivelmente corretas, distribuídas pelo

¹Unity: <http://unity3d.com/>

²Blender: <http://www.blender.org/>

³GitHub: <https://github.com/>

⁴<https://github.com/Brendo94/GamePOO>

cenário. Com a identificação da classe correta, o jogador terá adquirido a habilidade para invocar seu primeiro *pet* e como isso poderá passar para a próxima fase do jogo.

Esta primeira fase é composta por dois (2) componentes: mago e pergaminhos. Na Figura 1 é apresentada a representação gráfica desta fase. A responsabilidade de apresentar a atividade ao jogador nesta fase é do mago, (i) na Figura 1. O mago também é responsável por verificar se o jogador conseguiu identificar a classe correta, e posteriormente, apresentar ao jogador seu *feedback*. Já nos pergaminhos são apresentadas ao jogador as descrições do conjunto de classes envolvidas nesta fase; estes pergaminhos encontram-se espalhados por todo o mapa. Para coletar um determinado pergaminho, o jogador precisa entrar na área de colisão do pergaminho. Dessa forma, será apresentada a descrição da classe, e a opção de coletar o pergaminho, (ii) na Figura 1. Com o pergaminho coletado, o jogador deve retornar ao mago para verificar se ele encontrou o pergaminho correto. Apenas um dos pergaminhos da fase está sintaticamente correto e pode ser utilizado para aprender a invocar *pets* na Batalha. Dentre os erros presentes nas classes representadas nos pergaminhos, têm-se erros na declaração de métodos, uso de tipos de variáveis e uso de palavras reservadas.

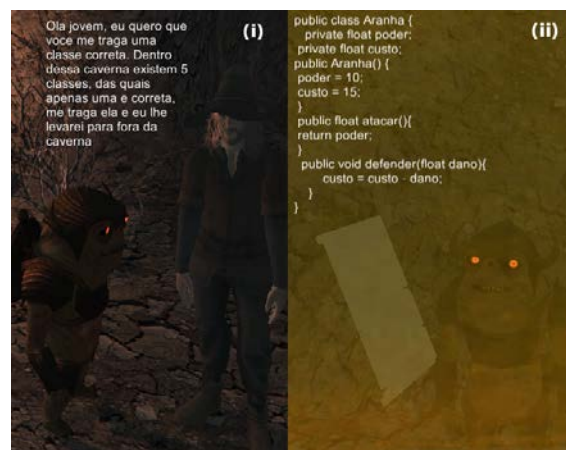


Figura 1. Fase da Caverna - Apresentação e Cumprimento de tarefa

Fase 02 - Floresta Tropical: Esta cena corresponde ao trecho inicial da segunda fase do jogo. Sua principal responsabilidade é apresentar ao usuário o conjunto de criaturas que o jogador poderá duelar. Assim, quando o jogador estabelece um contato com a área de colisão de uma determinada criatura, os dois serão redirecionados para a cena onde acontece a batalha (campo de batalha). Estas criaturas encontram-se espalhadas por toda a extensão do mapa. Na Figura 2 é mostrado como esse processo ocorre. No caso, o jogador encontra-se próximo à área de colisão que é representado pela esfera verde em volta da criatura disponível para combate.

Cada criatura no mapa representa um novo duelo (batalha) para o jogador. Para ganhar o jogo, ele precisa combater todas as criaturas presentes no cenário até chegar ao chefe (*boss*). Para garantir um bom balanceamento do jogo, as primeiras criaturas enfrentadas tem um poder de ataque baixo, o que permite que o jogador aprenda a invocar seu *pet* e ordenar ataques e defesas de forma coordenada. Cada batalha representa uma chance de exercitar os conhecimentos de POO. Uma vez que o jogador ganhe o duelo, a criatura com a qual ocorreu a batalha ficará indisponível para um novo duelo. Caso o

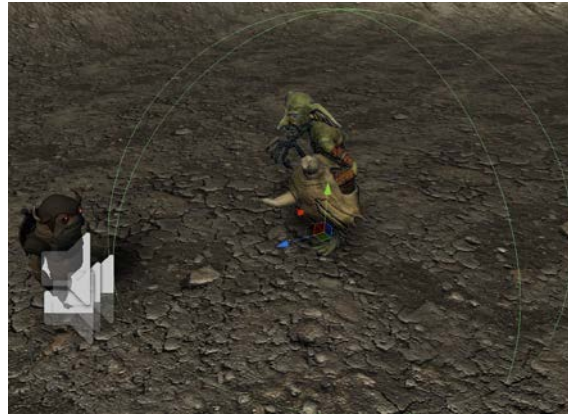


Figura 2. Fase 2: Floresta Tropical - Colisão entre jogador e criatura

jogador seja derrotado pela criatura, ele poderá entrar em combate com a mesma criatura novamente.

Campo de Batalha: Esta cena ainda faz parte da segunda fase do jogo e compreende seu ápice, em que o jogador deverá usar seus conhecimentos de POO para duelar com seus adversários. Na Figura 3 é demonstrado por meio do diagrama de atividades, as atividades que podem acontecer durante uma batalha.

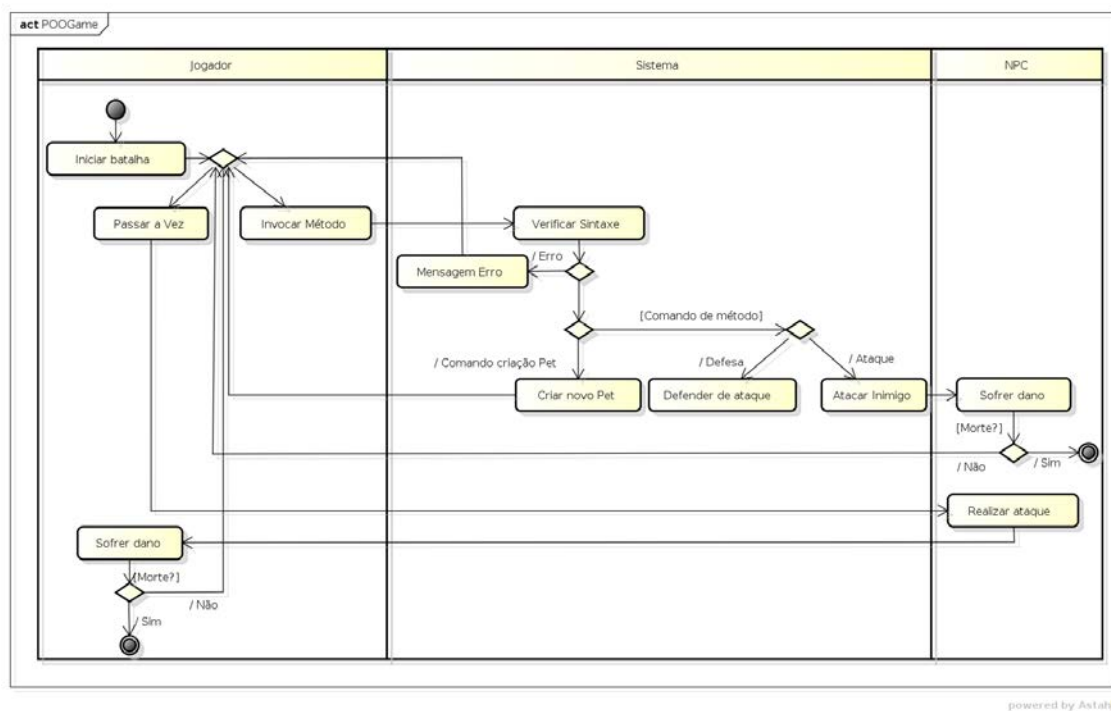


Figura 3. Diagrama de atividades da batalha do jogo

Esta cena é formada pelos seguintes componentes: inventário, terminal, *health bar*, *mana point* e passar vez. Na Figura 4 é apresentado como estes componentes encontram-se dispostos na cena.

No componente inventário, (i) na Figura 4, o jogador pode visualizar as classes



Figura 4. Fase 2: Campo de batalha

disponíveis para uso na batalha naquele instante do jogo. Em futuras versões do jogo, o jogador poderá também criar novas classes a partir da herança múltipla de classes que ele aprenderá em pergaminhos anteriores.

O processo de instanciação de uma classe ocorre através do componente terminal, (ii) na Figura 4. Assim, quando o jogador escreve um comando, instanciação de um objeto a partir de uma classe ou chamada de um método associado ao objeto criado, será realizado internamente uma verificação da sintaxe do código informado, retornando uma mensagem de erro, caso a sintaxe esteja incorreta; caso contrário, ocorre a criação de uma nova criatura no campo de batalha, para indicar que agora o jogador pode manipular este novo objeto, ou a execução de um método chamado pelo jogador. A instanciação de criaturas tem como finalidade fazer com que o jogador pratique a instanciação de objetos a partir de classes. Dessa maneira, o jogo busca separar o conceito de classe e objeto.

Já a chamada de método tem por finalidade fazer com que o jogador pratique e entenda chamada de métodos; neste caso utilizadas para manipular um objeto. Para tanto é necessário fornecer o identificador da criatura criada (objeto) e a ação que ela deve realizar (método). Com a chamada do método de ataque, um combatente atacará seu adversário fazendo com que ele sofra danos, resultando ou não em sua morte, e posteriormente, em sua vitória. Já o método de defesa faz com que o ataque do inimigo do jogador tenha seu impacto reduzido; além de que este método eleva as defesas do jogador por duas (2) rodadas. Caso o jogador perca uma batalha, ele terá de voltar ao início da fase. A batalha acontece em turnos, em que o jogador invoca dois (2) tipos de comportamentos (métodos) do *pet*: atacar ou defender.

No componente *health bar*, (iii) na Figura 4, é possível realizar um acompanhamento sobre o estado do jogo. Esse componente é responsável por apresentar o estado de vida dos combatentes. Adicionalmente para o jogador, é apresentado informações de *mana point*, (iv) na Figura 4, coeficiente de energia disponível para o jogador, que é necessário para que o mesmo possa conseguir realizar uma ação no batalha. Quando o coeficiente de energia for insuficiente para atacar ou defender, *mana point*, o jogador poderá utilizar o componente passar a vez, (v) na Figura 4, de modo que seu adversário possa dar continuidade à batalha. Esta ação também pode ser feita como uma forma de economizar magia que será utilizada para potencializar o efeito de ataques posteriores (críticos). O inimigo por sua vez sempre ataca e também tem uma chance de realizar um ataque crítico.

A batalha exige uma escolha estratégica entre ataques, defesas e “passar a vez”. Perde a batalha o combate que perder seus pontos de vida primeiro, *health bar*.

4. Avaliação do Jogo

O processo para avaliação dos objetivos associados com o jogo POOgame foi realizado por meio de uma avaliação de uso do jogo. Esta avaliação foi dividida em três (3) passos principais: (i) Planejamento, (ii) Execução e (iii) Análise de Dados. No planejamento (i), foi estabelecido que a avaliação seria feita por meio da definição das atividades a serem realizadas, especificação do grupo de participantes envolvidos e elaboração de questionário sobre a experiência de uso do jogo⁵. O questionário utilizado foi idealizado com base no modelo proposto em [Savi 2011]. Por se tratar de um jogo single player, que não atende aos requisitos de interação social do modelo de Savi (2011), não foram incluídas no questionário as perguntas associadas à avaliação da interação dos participantes com outros jogadores dentro do jogo. O questionário final era composto de trinta e duas (32) perguntas, divididas entre questões abertas e de múltipla escolha.

Em relação às respostas das questões de múltipla escolha, foram adotados os valores advindos da escala de Likert de 5 pontos para padronização das respostas, variando de -2 (discordo fortemente) até 2 (concordo fortemente). Assim quando a média de uma determinada característica avaliada aproximava-se de 2 (positivo), significava que a característica avaliada teve uma boa avaliação; caso contrário, média próxima de -2 (negativo), significava uma avaliação ruim.

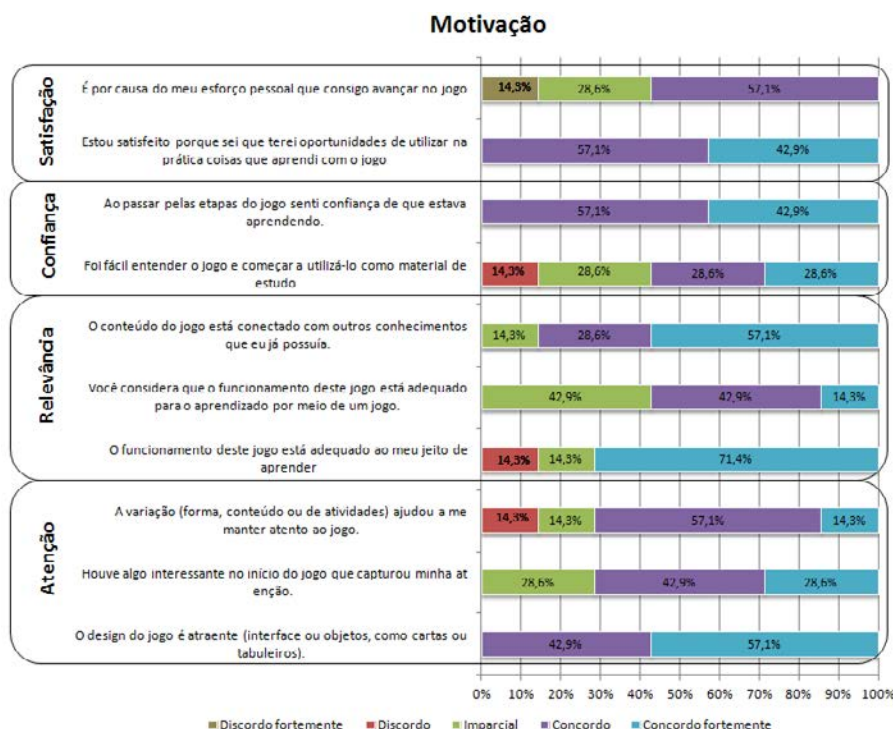


Figura 5. Avaliação da motivação do jogo

Com o planejamento realizado, deu-se início a Execução da avaliação (ii). A avaliação foi feita no dia 17/06/2015 nos laboratórios de informática do Campus da UFC

⁵goo.gl/1YOW9Y

Quixadá. Antes do início da avaliação foi realizada uma explicação sobre a proposta que o jogo apresentava para cada participante. Após a utilização do jogo foi feita a aplicação de um questionário, buscando extrair as experiências adquiridas pelos participantes. Participaram da avaliação sete (7) estudantes, advindos de diferentes cursos de graduação (Engenharia de Software, Sistemas de Informação e Redes de Computadores) e de diferentes semestres (variações entre o 1º e o 6º semestre). Os participantes apresentaram um feedback a respeito das dificuldades encontradas durante as atividades ao longo do jogo, e dos seus fluxos de execução.

Com os dados coletados iniciou-se a Análise dos dados (iii). Para identificar acertos e pontos de melhorias na estrutura do jogo foram gerados alguns gráficos. A partir da leitura dos gráficos foi possível chegar às seguintes conclusões. Em relação à motivação do jogo, como mostrado na Figura 5, é possível perceber que o jogo apresentou um resultado positivo na maioria dos pontos avaliados (satisfação, confiança, relevância, atenção).

O aspecto de satisfação foi o aspecto com a avaliação mais positiva, 92% de pontuação positiva (+1, +2). Isto revela que os alunos se sentiram realmente satisfeitos com as atividades realizadas durante o jogo. O segundo aspecto de maior avaliação positiva foi o aspecto de confiança apresentando 78%. O aspecto que obteve a avaliação mais baixa foi o de relevância, 71%. Isto revela que há a necessidade de incluir mais conteúdos relacionados ao ensino de POO nas atividades realizadas no jogo e melhorar o modo de aprendizagem dos jogadores. Por último, foi avaliado o aspecto atenção, obtendo 80% de avaliação positiva, constatando que o design do jogo foi capaz de atrair e manter o interesse do participante no jogo.

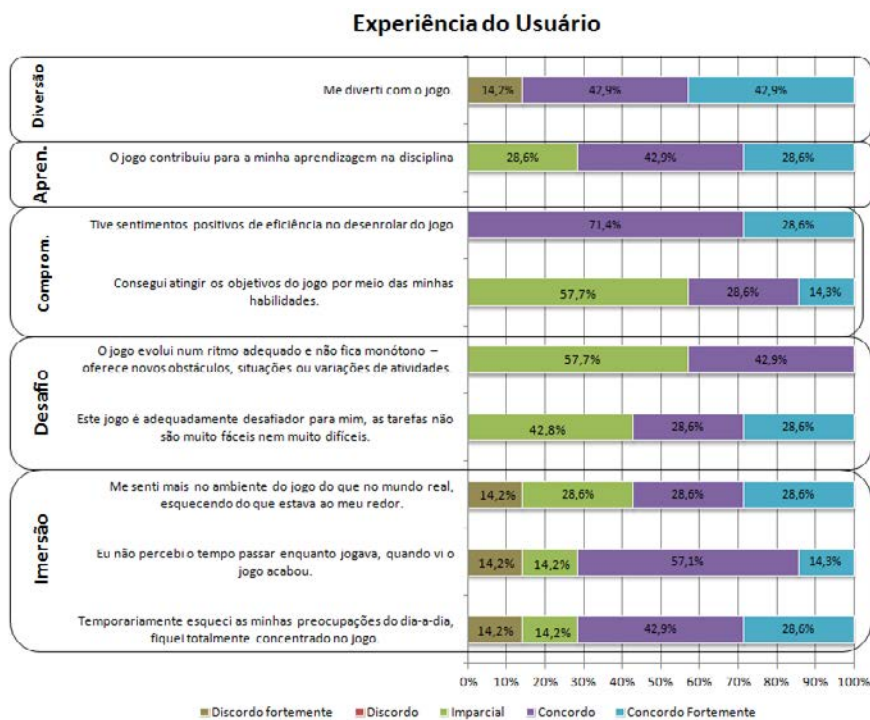


Figura 6. Avaliação da experiência do usuário

Também foram avaliadas as experiências que os participantes tiveram durante o jogo (diversão, aprendizado, competência, desafio e imersão), ilustradas na Figura

6. Nesta avaliação, teve-se uma média resultante geral inferior em relação à escala de motivação, identificando assim pontos de melhorias a serem trabalhados. O aspecto diversão obteve 85% de avaliação positiva, demonstrando que o participante pôde se divertir durante o uso do jogo. Os aspectos de aprendizagem e de competência apresentaram o mesmo percentual de avaliação (71%), evidenciando um bom resultado para ambos os aspectos, mas que ainda há melhoria a serem implementadas.

O aspecto de desafio apresentou o menor índice de avaliação (49%). Isto revela que o conteúdo das atividades precisa sofrer melhorias para futuras atualizações do jogo. O aspecto de imersão também apresentou uma avaliação baixa (66%), o que revela que apesar dos participantes terem gostado da experiência de contato com o jogo, ainda havia momentos do jogo em que os usuários dispersavam-se.

Os participantes também apresentaram sugestões e/ou melhorias que eles identificaram para o progresso do jogo. Estas informações foram consolidadas e são apresentadas na Figura 7. O ponto mais citado pelos participantes foi em relação aos aspectos tecnológicos que o jogo apresentava (movimentação da câmera e do personagem), representando 35% das melhorias citadas. Melhorias em relação à fundamentação do jogo juntamente com ideias de expansão (adição de novos objetivos e personagens) compreenderam cada um 18% das respostas dos participantes. Aspectos relacionados com o design do jogo (melhora dos gráficos em algumas partes específicas do jogo) obtiveram 17%. Por fim, a jogabilidade compreendeu 12% das respostas dos participantes, atentando-se principalmente para a melhoria do tutorial de ensino de uso do jogo e de ajudas recorrentes no momento da batalha.

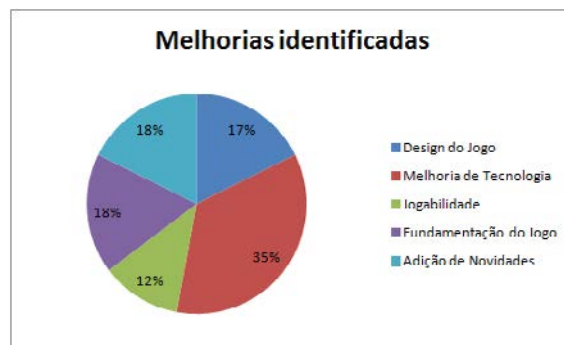


Figura 7. Melhorias identificadas no jogo

Ao final da avaliação concluiu-se que o aspecto de motivação referente ao uso do jogo foi atingido com um alto índice de aprovação (média de 80%), com grande ênfase dada ao aspecto de satisfação, evidenciando que o jogador se sentiu realmente motivado a jogar o jogo. Em relação à experiência do usuário com o jogo, concluiu-se que apesar da maioria dos aspectos avaliados terem tido uma avaliação satisfatória, média de 73%, o aspecto relacionado aos desafios propostos pelo jogo devem sofrer melhorias. Isto torna evidente que as expectativas construídas pelo jogador durante as fases iniciais do jogo não foram correspondidas ao longo do uso do jogo. As dificuldades dos desafios propostos precisam crescer gradativamente ao longo da execução do jogo. Sobre as melhorias apontadas pelos próprios jogadores tornou-se evidente que apesar do jogo ter capturado a atenção do jogador, através dos gráficos e dos mapas utilizados, ainda há a necessidade por um melhor uso da tecnologia utilizada no desenvolvimento do jogo.

5. Conclusão

Esse artigo apresentou um jogo desenvolvido para auxiliar o ensino de programação orientada a objetos (POO). Os resultados apresentados demonstraram que o uso do jogo foi bem recebido pelos alunos e que ajudou a motivar a execução da fixação de conceitos. Os alunos consideraram o jogo atraente e estavam confiantes de que estavam aprendendo durante o jogo. Entretanto, os resultados também apontaram pontos de melhorias, especialmente nos aspectos de diversão e de capacidade de propor desafios adequados ao ensino. Novas versões do jogo podem trazer um maior feedback com sugestões de ações e um tutorial para deixar as regras ainda mais claras.

Como trabalhos futuros, pretende-se evoluir o jogo para englobar um maior número de conceitos de orientação a objetos, e corrigir os problemas encontrados pela avaliação. Também pretende-se realizar outras avaliações do jogo em diferentes turmas de POO de cursos de computação, como uma forma de avaliar constantemente os resultados de uso do jogo no âmbito da sala de aula, e verificando se o uso do jogo tem impacto, por exemplo, nos índices de aprovação das respectivas turmas.

Referências

- Al-Linjawi, A. A. and Al-Nuaim, H. A. (2010). Using alice to teach novice programmers oop concepts. *Journal of King Abdulaziz University: Science*, 22(1):59–68.
- Figueiredo, K. d. S., Ribeiro, J. M., Souza, R., and Angelo, V. R. (2015). Uma abordagem gamificada para o ensino de programação orientada a objetos. *WEI - XXIII Workshop sobre Educação em Computação (WEI 2015)*.
- Gosling, J., Joy, B., Steel, G., Bracha, G., and Buckley, A. (2013). The java language specification.
- Holland, S., Griffiths, R., and Woodman, M. (1997). Avoiding object misconceptions. In *ACM SIGCSE Bulletin*, volume 29, pages 131–134. ACM.
- Lopes, A. C. B. (2012). Desenvolvimento de um jogo didático para ensino de programação orientada a objetos e sua aplicação em cursos técnicos de computação. Master's thesis, Universidade Federal Rural do Semi-Árido, Brasil.
- Savi, R. (2011). *Avaliação de Jogos voltados para a disseminação do conhecimento*. PhD thesis, Universidade Federal de Santa Catarina, Brasil.
- Voss, G., Nunes, F., and Medina, R. (2013). Proposta de um jogo sério para o ensino de redes de computadores no ambiente virtual 3d opensim. In *Proceedings of the XI Brazilian Symposium on Games and Digital Entertainment—SBGames 2013*, pages 37–40.
- Werbach, K. and Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

PortuCol: uma pseudolinguagem inspirada em C ANSI para o Ensino de Lógica de Programação e Algoritmos

Lucas Lemos Barbosa, Christian Marlon Souza Couto, Ricardo Terra

Departamento de Ciência da Computação,
Universidade Federal de Lavras (UFLA), Brasil

lbarbosa@computacao.ufla.br, christianmsc@sistemas.ufla.br, terra@dcc.ufla.br

Resumo. *Linguagens de programação são intrinsecamente complexas, devido à variabilidade, generalidade e completude de suas construções. Isso justifica a adoção de pseudolinguagens de alto nível no ensino de Lógica de Programação e Algoritmos. Embora C ANSI e Java sejam as linguagens de programação mais utilizadas, Portugol – uma pseudolinguagem com instruções em português inspirada na linguagem obsoleta Pascal – é ainda largamente adotada por IES. Diante desse cenário, este artigo propõe PortuCol, uma pseudolinguagem de programação com instruções em português assim como Portugol, porém inspirada em C ANSI. Este artigo também descreve PortuCol2C, uma ferramenta que traduz código escrito em PortuCol para C ANSI e conduz estudos empíricos que demonstram estatisticamente que (i) PortuCol é 27% mais similar a C ANSI do que o Portugol e (ii) PortuCol é para C ANSI o que Portugol é para Pascal.*

Abstract. *Programming languages are intrinsically complex due to the variability, generality, and completeness of their constructions. This justifies the adoption of high-level pseudocode languages when teaching Introduction to Algorithms. Although ANSI C and Java are the most used programming languages, Portugol—a pseudocode language with instructions in Portuguese inspired by the outdated Pascal language—is still widely adopted by Brazilian universities. In view of such circumstances, this paper proposes PortuCol, a pseudocode programming language with instructions in Portuguese as Portugol, although inspired by ANSI C. It also describes PortuCol2C, a tool that translates PortuCol to ANSI C, and conducts empirical studies that statistically demonstrate that (i) PortuCol is 27% more similar to ANSI C than Portugol and (ii) PortuCol is for C ANSI what Portugol is for Pascal.*

1. Introdução

No início do ensino superior, é comum que estudantes da área de Ciência da Computação e afins tenham dificuldade no raciocínio lógico de criação de soluções, uma vez que, no ensino médio, esses estudantes estavam acostumados a apenas aplicar fórmulas previamente existentes na solução de problemas. Um estudo atesta essa dificuldade reportando que, em média, mais de 50% dos alunos das disciplinas de introdução à programação são reprovados a cada semestre [2].

Em virtude de que linguagens de programação são intrinsecamente complexas – devido à variabilidade, generalidade e completude de suas construções – docentes de IES

usualmente adotam pseudolinguagens de alto nível nas disciplinas de introdução à Lógica de Programação e Algoritmos. Pseudolinguagens proporcionam ao aluno um maior espaço para o desenvolvimento de seu raciocínio lógico relacionado à resolução de problemas, sem preocupações relativas ao uso correto de uma linguagem de programação.

Portugol é, hoje no Brasil, a pseudolinguagem mais adotada por universidades com esse propósito [9, 11], a qual conta com diversas ferramentas de apoio ao ensino [3, 5, 9, 8, 4]. No entanto, Portugol é o produto de uma simbiose do português e as linguagens de programação ALGOL e Pascal [5], ambas obsoletas [10]. Pascal, por exemplo, sequer está classificada entre as 50 linguagens de programação mais utilizadas no desenvolvimento de sistemas de software [10]. Assim, devido à origem sintática de Portugol, estudantes *não* terão uma transição natural ao aprender linguagens de programação do estado-da-prática como C ANSI e Java.

Como uma alternativa ao Portugol, este artigo propõe PortuCol, uma pseudolinguagem de programação com instruções em português assim como Portugol, porém inspirada em C ANSI. Estudos empíricos atestam que (i) PortuCol é 27% mais similar a C ANSI do que o Portugol, o que proporciona uma transição mais natural para o estudante ao aprender linguagens do estado-da-prática e (ii) PortuCol é para C ANSI o que Portugol é para Pascal, o que retrata apenas uma adaptação da pseudolinguagem às linguagens do estado-da-prática. Além disso, este artigo descreve PortuCol2C, uma ferramenta que traduz código PortuCol para C ANSI, o que possibilita aos estudantes se familiarizem com a linguagem C ANSI através da observação do seu código PortuCol traduzido.

Este artigo está organizado como a seguir. A Seção 2 apresenta Portugol, a pseudolinguagem mais adotada em disciplinas de programação do Brasil. A Seção 3 descreve o PortuCol, sua sintaxe e a ferramenta PortuCol2C. A Seção 4 reporta os resultados da avaliação de similaridade textual entre algoritmos implementados em PortuCol, C ANSI, Portugol e Pascal. Por fim, a Seção 5 conclui e propõe trabalhos futuros.

2. Portugol

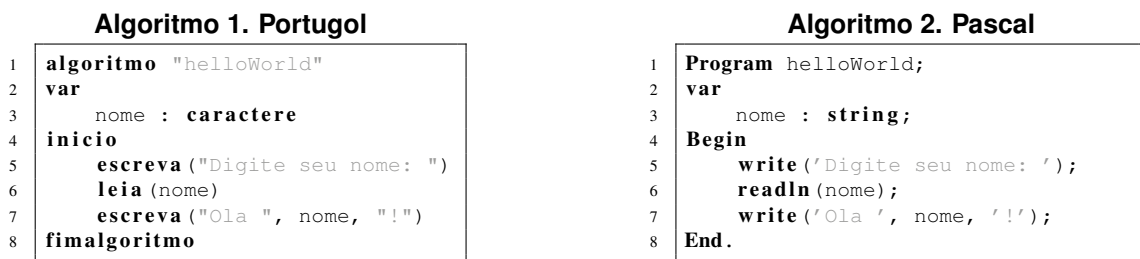
O Portugol é a pseudolinguagem mais adotada em classes introdutórias à programação em IES no Brasil [9, 11]. Mais importante, a maioria dos livros utilizados como referência básica nas disciplinas de introdução à programação em universidades brasileiras possui explicações usando alguma notação de Portugol [9, 11]. Em suma, o objetivo de Portugol é facilitar o aprendizado de Lógica de Programação e Algoritmos para alunos não habituados com programação.

Portugol é o produto de uma simbiose do português e as linguagens de programação ALGOL e Pascal [5], ambas obsoletas [10]. Seu propósito – o qual foi atingido – é ser uma pseudolinguagem simples e com instruções em português para favorecer estudantes não familiares com a língua inglesa. Ao utilizar Portugol, o aluno não precisa se preocupar com a sintaxe e comandos em inglês que uma linguagem de programação mais complexa possui, o que foca sua concentração no *como* resolver um problema proposto.

Apesar de trazer benefícios inquestionáveis no aprendizado de Lógica de Programação e Algoritmos, o Portugol não provê, por definição, uma forma de executar o pseudocódigo escrito para que alunos possam testar seus algoritmos, o que se torna indispensável à medida que os algoritmos se tornam mais complexos e passam a requerer códigos

mais elaborados. Esse problema ocorre pois o Portugol tem diversas variações em sintaxe e em comandos dependendo do contexto de utilização. Isso faz com que as ferramentas não sejam padronizadas, i.e., suportando sintaxes e instruções diferentes [3, 5, 9, 8, 4].

Dentre as ferramentas existentes que apoiam o ensino utilizando Portugol podemos destacar: CIFluxProg [3] disponibiliza um ambiente visual com fluxogramas, além de um ambiente textual com uma variação de Portugol; WEB-UNERJOL [5] provê um ambiente *web* que permite o desenvolvimento e execução de algoritmos em Portugol através de um navegador; Portugol Studio [9] é um ambiente de desenvolvimento integrado (IDE) para uma variante do Portugol que utiliza o compilador Portugol Núcleo; Portugol IDE [8] provê um ambiente que possibilita desenvolver algoritmos em Portugol através do uso de fluxogramas como uma linguagem gráfica, além do tradicional modo textual; e, por fim, VisuAlg[4] que adota uma variante de Portugol bem parecida com Pascal, porém sem pontos-e-vírgulas para separar comandos. Os Algoritmos 1 e 2 ilustram tal similaridade apresentando um mesmo algoritmo escrito em Portugol e em Pascal, respectivamente.



3. PortuCol

Este artigo propõe PortuCol, uma pseudolinguagem de programação com instruções em português assim como Portugol, porém inspirada em C ANSI. O objetivo é adaptar o ensino de Lógica de Programação e Algoritmos através de uma pseudolinguagem cuja sintaxe seja inspirada em uma linguagem de programação do estado-da-prática (i.e., amplamente adotada por fábricas de software).

A Figura 1 ilustra o contexto de utilização de PortuCol. Estudantes desenvolvem o código em PortuCol e, automaticamente, a ferramenta PortuCol2C traduz PortuCol para C ANSI e executa o código correspondente. Isso possibilita que estudantes (i) executem seu pseudocódigo; (ii) se familiarizem com a linguagem C ANSI através da observação do seu código PortuCol traduzido; e (iii) tenham uma transição mais natural ao futuramente aprender linguagens do estado-da-prática como C ANSI e Java.

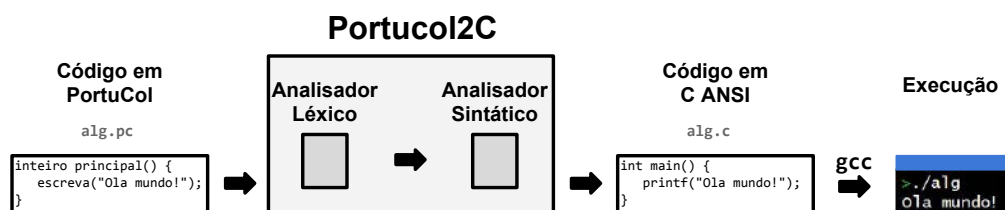


Figura 1. Utilizando PortuCol

3.1. Sintaxe

A definição da sintaxe de PortuCol considerou didática, facilidade de aprendizagem e similaridade com C ANSI de forma que futuramente a transição para essa linguagem de programação seja mais natural para o aluno.

A Tabela 1 apresenta a equivalência de instruções PortuCol, C e Portugol. PortuCol possui instruções e palavras reservadas em português similarmente à Portugol e uma sintaxe similar à C ANSI. Em PortuCol, assim como em C ANSI, é necessário delimitar o escopo do código com os símbolos de abre e fecha chaves. Os operadores aritméticos e lógicos são como em C ANSI; exceções (e.g., operadores lógicos e e ou) ocorreram em virtude do caráter didático que a pseudolinguagem proposta tem como objetivo.

As estruturas básicas de desvio e repetição seguem, em sua maioria, a estrutura da linguagem de programação C ANSI. Contudo, uma das estruturas que foi modificada com o intuito de apresentar um conceito de forma didática foi a estrutura de repetição para (for) que em PortuCol tem condições de parada fixas (menor ou igual, e maior ou igual). O objetivo dessa modificação foi de habituar o aluno a entender arranjos de tamanho n como estruturas que são acessadas a partir da posição 0 até a posição $n - 1$, o que acontece na maioria das linguagens de programação.

Usando o mesmo algoritmo descrito na Seção 2, os Algoritmos 3 e 4 apresentam dois exemplos do mesmo código em PortuCol e C ANSI, respectivamente.

Algoritmo 3. PortuCol

```

1 inteiro principal(){
2     texto nome;
3     escreva("Digite seu nome: ");
4     leia("%texto", nome);
5     escreva("Ola %texto!", nome);
6 }
```

Algoritmo 4. C ANSI

```

1 int main(){
2     char nome[20];
3     printf("Digite seu nome: ");
4     scanf("%s", nome);
5     printf("Ola %s!", nome);
6 }
```

3.2. Ferramenta PortuCol2C

Para que os estudantes possam executar seus códigos escritos em PortuCol, foi criado um protótipo de uma ferramenta denominada PortuCol2C. Essa ferramenta realiza a tradução dos códigos em PortuCol para a linguagem C ANSI e, por conseguinte, já compila e gera o arquivo executável correspondente. Dessa forma, PortuCol2C permite que estudantes (i) “executem” seu código PortuCol e (ii) observem qual seria o código C ANSI correspondente, o que já os familiariza com linguagens de programação reais.

Na atual versão da ferramenta, a execução de PortuCol2C é por linha de comando, conforme pode ser observado na Figura 2. Dado um código em PortuCol (alg.pc), a execução ferramenta (portucol2c alg.pc) gera o código C ANSI correspondente (alg.c) e o arquivo executável (alg).

O tradutor foi implementado usando *Lex* e *Yacc*,¹ que são ferramentas designadas para o desenvolvimento de compiladores e interpretadores [6]. O *Lex* realiza a análise léxica que consiste na leitura dos caracteres de entrada e produção de uma sequência de *tokens*, enquanto o *Yacc* executa a análise sintática que consiste na obtenção de uma cadeia

¹<http://dinosaur.compilertools.net/>

Tabela 1. Equivalência de instruções PortuCol, C e Portugal

Construções	PortuCol	C	Portugal
Estrutura	inteiro principal(){ ... }	int main(){ ... }	Algoritmo "nome" var ... inicio ... fimalgoritmo
Variável tipo inteiro	inteiro i;	int i; short int i; long int i; long long int i;	i: inteiro
Variável tipo real	real d;	float d; double d; long double d;	d: real
Variável tipo lógico	logico b;	bool b; (C99)	b: logico
Variável tipo caractere	texto s;	char s;	s: caractere
Variável tipo <i>string</i>	texto s;	char s[];	s: caractere
Variável tipo vetor	tipo arranjo[linhas];	tipo arranjo[linhas];	v: arranjo[linha1..linhaN] de tipo
Variável tipo matriz	tipo arranjo[linhas][colunas];	tipo arranjo[linhas][colunas];	v: arranjo[linha1..linhaN, colunaa1..colunaaN] de tipo
Ler valor	leia("%tipo", i);	scanf("%tipo", i);	leia(i)
Escrever	escreva("...");	printf("...");	escreva("...")
Escrever com quebra de linha	escreva("...\n");	printf("...\n");	escreval("...")
Escrever com parâmetros	escreva("... %tipo ... %tipo ...", a, b);	printf("... %tipo ... %tipo ...", a, b);	escreval("...", a, "...", b, "...")
Operação de atribuição	a = b;	a = b;	a <- b a := b
Operação de adição	a + b;	a + b;	a + b
Operação de subtração	a - b;	a - b;	a - b
Operação de multiplicação	a * b;	a * b;	a * b
Operação de divisão	a / b;	a / b;	a / b
Operação de divisão inteira	a / b;	a / b;	a div b
Operação de resto da divisão	a % b;	a % b;	a % b a mod b
Operação de potenciação	potencia(a, b);	pow(a, b) <math.h>	a ^ b
Operação de raiz quadrada	raizquadrada(a);	sqrt(a); <math.h>	RaizQ(a)
Operação de concatenação	concatena("texto", "texto"); concatena("texto", a); concatena(a, "texto");	strcat("texto", "texto"); strcat("texto", a); strcat(a, "texto");	"texto" + "texto" "texto" + a (string) a (string) + "texto"
Operação de tamanho do texto	tamanhotexto(a);	strlen(a); <string.h>	compr(a)
Operador menor	a < b;	a < b;	a < b
Operador menor ou igual	a <= b;	a <= b;	a <= b
Operador maior	a > b;	a > b;	a > b
Operador maior ou igual	a >= b;	a >= b;	a >= b
Operador igual	a == b;	a == b;	a = b
Operador diferente	a != b;	a != b;	a <> b
Operador e	a e b	a && b	a e b
Operador ou	a ou b	a b	a ou b
Operador de negação	!a	!a	nao a
Condicional se	se (condição) { ... } senão { ... } fimse	if (condição) { ... } else { ... }	se condição entao ... senao ... fimse
Condicional aninhado	se (condição) { ... } senão se (condição) { ... } senão { ... }	if (condição) { ... } else if (condição) { ... } else { ... }	se condição entao ... senao se condicao entao ... senao ... fimse
Condicional caso	escolha (variável) { caso valor1: ... caso valor2, valor3: ... padrão: ... }	switch (variável) { case valor1: ... break; case valor2: ... case valor3: ... break; default: ... break; }	escolha variável caso valor1 ... caso valor2, valor3 ... outrocaso ... fimescolha
Condicional enquanto	enquanto (condição) { ... }	while (condição) { ... }	enquanto condição faca ... Fimenquanto
Condicional para	para (variável = valor1 ; valorN ; P) { ... }	for (variável = valor1; variável <= valorN; variável += P) { ... }	para variável de valor1 ate valorN passo P faca ... fimpara
Comando de interrupção	interrompa;	break;	interrompa
Criar procedimento	semretorno nome(tipo parâmetro, ...) { ... }	void nome(tipo parâmetro, ...) { ... }	procedimento nome(parâmetro: tipo; ...) var ... inicio ... fimprocedimento
Criar função	tipo nome(tipo parâmetro, ...) { ... retorne valor; }	tipo nome(tipo parâmetro, ...) { ... return valor; }	funcao nome(parâmetro: tipo; ...): tipo var ... inicio ... retorne valor fimfuncao

```

toy_example -- -bash -- 52x17
[toy_example@terra] $ cat alg.pc
inteiro principal(){
    escreva("Ola mundo!");
}

[toy_example@terra] $ ./portuCol2c alg.pc
Codigo C ANSI alg.c gerado e ja compilado com GCC.

[toy_example@terra] $ cat alg.c
#include <stdio.h>

int main () {
    printf("Ola mundo!");
}

[toy_example@terra] $ ./alg
Ola mundo!

```

Figura 2. Execução do PortuCol2C

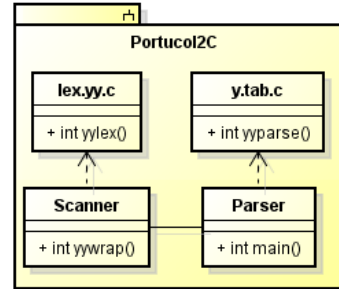


Figura 3. Arquitetura do PortuCol2C

de *tokens* e verificação de quaisquer erros de sintaxe de uma forma inteligível [1]. Conforme ilustrado na Figura 3, a ferramenta segue então uma arquitetura com os seguintes dois módulos principais:

1. *Scanner*: Responsável por realizar a análise léxica do código-fonte escrito em PortuCol. Sua implementação consiste em definições de expressões regulares a serem convertidas em *tokens* tais como números, variáveis e palavras reservadas. Posteriormente, são definidas as regras para cada *token*, sendo que no caso do *Scanner* as regras estabelecidas consistem em retornar cada *token* para o *Parser*. A compilação do *Scanner* pelo *Lex* gera o arquivo `lex.yy.c`, o qual é utilizado para viabilizar a detecção dos *tokens* no código-fonte do PortuCol.
2. *Parser*: Responsável por realizar a análise sintática do código-fonte escrito em PortuCol. O *Parser* é constituído por regras formadas através dos *tokens* recebidos pelo *Scanner*, sendo que cada regra possui subrotinas associadas. No nosso caso, as subrotinas têm como finalidade imprimir, no arquivo de saída, a tradução para código C ANSI da sequência de *tokens* relacionado à cada regra. A compilação do *Parser* pelo *Yacc* gera o arquivo `y.tab.c`, o qual é utilizado para possibilitar a análise sintática dos *tokens*.

4. Avaliação

4.1. Questões de Pesquisa

Um estudo foi projetado para responder às seguintes questões de pesquisa:

QP #1 – PortuCol se aproxima à C ANSI mais do que Portugol?

QP #2 – PortuCol se aproxima à C ANSI mais do que Portugol se aproxima à Pascal?

4.2. Dataset²

Para efetivamente comparar a similaridade entre algoritmos nas linguagens PortuCol, C, Portugol e Pascal, foram especificados 30 algoritmos, os quais foram implementados em cada uma das linguagens. A Tabela 3 descreve sucintamente os algoritmos implementados e aponta as principais estruturas utilizadas.

²Dataset publicamente disponível em: http://www.dcc.ufla.br/~terra/papers/2016_wei_portuCol

Tabela 3. Descrição dos algoritmos

Algoritmo	Descrição	Principais estruturas utilizadas
Alg_1	Cadastro e acesso de produtos	Procedimentos, enquanto, para, caso, se
Alg_2	Equação quadrática	Tipo de dado real, se senao se
Alg_3	Idade por ano de nascimento	Tipo de dado arranjo, para
Alg_4	Bubble sort	Tipo de dado arranjo, para, se
Alg_5	Busca binária	Função, enquanto, se senao se
Alg_6	Busca em arranjo	Faca enquanto, se, senao
Alg_7	Calculadora (+, -, *)	Caso, tipo de dado texto
Alg_8	Média de notas	Leitura de entrada, se, senao
Alg_9	Soma e subtração de reais	Para, se, senao
Alg_10	Combinação	Escrita e leitura de dados, Para
Alg_11	Decomposição de inteiro	Para, faca enquanto, se
Alg_12	Aprovação de aluno por nota	Caso
Alg_13	Fatorial recursivo	Função recursiva, se senao se
Alg_14	Fibonacci iterativo	Função, para
Alg_15	Selection sort	Para, tipo de dado arranjo, se
Alg_16	Mostrar coluna de matriz	Matriz 4x4, para
Alg_17	Média de reais	Função, tipo de dados real
Alg_18	Inverter texto	Tipo de dados texto, para
Alg_19	Números palíndromos	Enquanto, se, senao
Alg_20	Primos até certo número	Para, enquanto, se
Alg_21	Aproximação de Pi	Para, operação de raiz quadrada
Alg_22	Potenciação de inteiros (x^y)	Para
Alg_23	Comb sort	Enquanto, se, para
Alg_24	Soma de matrizes	Matrizes 2x2, para
Alg_25	Soma de algarismos	Enquanto, operação de resto de divisão
Alg_26	Classificação de texto por caso	Caso, comparação de texto
Alg_27	Validação de CPF	Para, se, conversão de texto para inteiro
Alg_28	Divisão de arranjos entre par e ímpar	Repita enquanto, para, se, senao
Alg_29	Aproximação de raiz quadrada	Se, enquanto, senao
Alg_30	Concatenação e comparação de texto	Se, senao, operações de concatenação e comparação de texto

4.3. Similaridade Textual

Para obter a similaridade entre os algoritmos implementados em PortuCol, C, Portugol e Pascal foi utilizada a similaridade de cosseno (*cosine similarity*) – técnica mais adotada em recuperação de informação e mineração de texto para comparação de documentos [7] – a qual mensura a similaridade de dois documentos com base no ângulo entre seus vetores representativos.³

Portanto, a similaridade de cosseno entre dois documentos d_1 e d_2 é computada a partir dos seus vetores representativos $\vec{V}(d_1)$ e $\vec{V}(d_2)$ conforme Equação 1.

$$sim(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{|\vec{V}(d_1)| |\vec{V}(d_2)|} \quad (1)$$

onde $\vec{V}(d_i)$ denota o vetor derivado de um documento d_i , com um componente no vetor para cada termo do dicionário. Assim, cada termo do documento é armazenado uma única vez no vetor e é associado a um peso. O peso dos termos normalmente leva em consideração o $tf-idf_{t,d}$, onde tf é a frequência dos termos (*Term Frequency*), idf é a frequência inversa do documento (*Inverse Document Frequency*), t representa os termos e d o documento.⁴ Enfim, a similaridade entre dois documentos consiste na razão entre (i) o produto escalar e (ii) o produto das normas euclidianas dos dois vetores representativos.

³A similaridade varia de 0 (nada similares) a 1 (totalmente similares).

⁴Particularmente neste artigo, foi considerada apenas a frequência dos termos, uma vez que diminuir o peso de termos frequentes não é desejável.

4.4. Metodologia

Foram conduzidas as seguintes atividades para se responder às questões de pesquisa:

1. Definir uma particular versão de Portugol, devido à existência de variações do Portugol encontradas na literatura, onde diferentes ferramentas [3, 5, 9, 8, 4] possuem suas próprias variações de Portugol. Assim, foi escolhida a variante adotada pelo VisuAlg – um interpretador de Portugol desenvolvido por Souza [4] – a qual é completa e amplamente aceita.
2. Construir um *dataset* contendo uma vasta gama de algoritmos implementados nas linguagens PortuCol, C, Portugol e Pascal (veja Seção 4.2).
3. Comparar a similaridade textual – utilizando *cosine similarity* descrito na Seção 4.3 – dos algoritmos implementados na linguagem PortuCol com C, Portugol com C e Portugol com Pascal. É importante mencionar que textos referentes à inclusão de bibliotecas (e.g. `#include <stdio.h>` e `uses crt`) foram retirados dos códigos C e Pascal.
4. Determinar estatisticamente a relação de similaridade de PortuCol com C e de Portugol com C, a fim de responder a QP #1.
5. Determinar estatisticamente a relação de similaridade de PortuCol com C e de Portugol com Pascal, a fim de responder a QP #2.

4.5. Resultados

A Tabela 4 reporta os resultados da similaridade textual dos algoritmos implementados na linguagem Portugol com C (coluna 1), PortuCol com C (coluna 2) e Portugol com Pascal (coluna 3). A Figura 4 representa visualmente o intervalo de confiança para a média das similaridades textuais de cada par de linguagens.

Tabela 4. Similaridade Textual

Algoritmo	Portugol/C	PortuCol/C	Portugol/Pascal
Alg_1	0.4004	0.6031	0.4365
Alg_2	0.7009	0.6824	0.7607
Alg_3	0.6348	0.7314	0.7291
Alg_4	0.7824	0.8713	0.7850
Alg_5	0.6803	0.7712	0.6591
Alg_6	0.5866	0.7072	0.6256
Alg_7	0.6674	0.6292	0.7227
Alg_8	0.6190	0.5930	0.6347
Alg_9	0.3036	0.7593	0.6543
Alg_10	0.7386	0.6616	0.8338
Alg_11	0.7262	0.8586	0.7523
Alg_12	0.2761	0.3639	0.5355
Alg_13	0.6385	0.6468	0.6156
Alg_14	0.6230	0.6863	0.5823
Alg_15	0.7190	0.8915	0.7118
Alg_16	0.5420	0.6057	0.5809
Alg_17	0.4449	0.4355	0.5536
Alg_18	0.6502	0.7662	0.7963
Alg_19	0.8511	0.8685	0.8348
Alg_20	0.3073	0.7611	0.7110
Alg_21	0.3267	0.7708	0.7780
Alg_22	0.6261	0.6199	0.6581
Alg_23	0.8580	0.9192	0.8247
Alg_24	0.0407	0.8187	0.9045
Alg_25	0.1079	0.8729	0.7621
Alg_26	0.6490	0.7286	0.7338
Alg_27	0.6202	0.8472	0.5964
Alg_28	0.5844	0.6647	0.6306
Alg_29	0.8454	0.8956	0.8604
Alg_30	0.4611	0.5484	0.4665
Média (\bar{x})	0.5671	0.7193	0.6910
Dev. Pad. (s)	0.2089	0.1361	0.1161

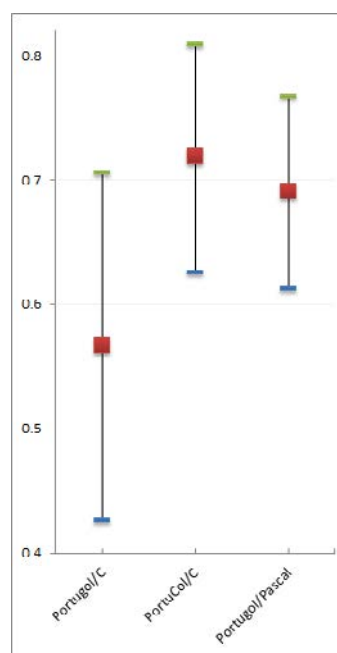


Figura 4. Intervalo de confiança

QP #1 – PortuCol se aproxima à C ANSI mais do que Portugol?

Foram comparadas as similaridades textuais dos algoritmos implementados (i) em Portugol e em C e (ii) em PortuCol e em C, cujos resultados podem ser observados nas primeiras duas colunas da Tabela 4 e nas duas primeiras linhas da Figura 4.

O *dataset* disponível para avaliação apresenta os mesmos algoritmos implementados em diferentes linguagens (veja Seção 4.2). Isso permite que seja conduzida uma observação pareada dos resultados de similaridade textual. Portanto, considerando uma distribuição normal dos resultados, a distribuição probabilística *t-Student* foi utilizada no cálculo do intervalo de confiança para responder à QP #1, conforme Equação 2.⁵

$$IC = \bar{x} \pm t_{[1-\frac{\alpha}{2}; n-1]} \left(\frac{s}{\sqrt{n}} \right) = 0,1523 \pm 3,6590 \left(\frac{0,2158}{\sqrt{30}} \right) = [0,0081, 0,2964] \quad (2)$$

O *dataset* utilizado possui trinta algoritmos ($n = 30$). A diferença da similaridade textual entre cada algoritmo Portugol/C e PortuCol/C (colunas 1 e 2 da Tabela 4) foi calculada. Assim, obteve-se a média aritmética ($\bar{x} = 0,1523$) e o desvio padrão ($s = 0,2158$) das diferenças. A confiança adotada para as amostras da distribuição de *t-Student* foi 99,9% com grau de liberdade de 29, onde $t_{[0,9995; 29]}$ é igual a 3,659. Assim, o teste *t* resultou no intervalo de $[0,0081, 0,2964]$, o qual *não* inclui zero. Isso indica que, com 99,9% de confiança, PortuCol se aproxima 26,85% mais à C ANSI do que Portugol.

QP #2 – PortuCol se aproxima à C ANSI mais do que Portugol se aproxima à Pascal?

Similarmente à QP #1, foram comparadas as similaridades textuais dos algoritmos implementados (i) em PortuCol e em C e (ii) em Portugol e em Pascal, cujos resultados podem ser observados nas últimas duas colunas da Tabela 4 e nas duas últimas linhas da Figura 4. É importante mencionar que os resultados comparativos entre PortuCol e C já tinham sido previamente calculados na resposta à QP #1.

Considerando uma distribuição normal dos resultados, a distribuição probabilística *t-Student* novamente foi utilizada no cálculo do intervalo de confiança para responder à QP #2, conforme Equação 3.

$$IC = \bar{x} \pm t_{[1-\frac{\alpha}{2}; n-1]} \left(\frac{s}{\sqrt{n}} \right) = 0,0283 \pm 3,659 \left(\frac{0,1006}{\sqrt{30}} \right) = [-0,0389, 0,0955] \quad (3)$$

Similarmente à QP #1, a confiança adotada para as amostras da distribuição de *t-Student* foi 99,9% com grau de liberdade de 29. Assim, o teste *t* resultou no intervalo de $[-0,0389, 0,0955]$. Como o intervalo de confiança inclui zero, podemos considerar que o PortuCol se aproxima à C ANSI da mesma maneira que Portugol se aproxima à Pascal.

5. Considerações Finais

O uso de pseudolinguagens no ensino de Lógica de Programação e Algoritmos é uma estratégia muito utilizada em IES. Embora C ANSI e Java sejam as linguagens de programação mais utilizadas no desenvolvimento de sistemas de software [10], a pseudolinguagem mais utilizada no Brasil é o Portugol [9, 11], que é inspirada no obsoleto Pascal.

⁵O teste *t* resulta em um intervalo de confiança. Caso o intervalo inclua zero, as amostras *não* são estatisticamente diferentes. Caso contrário, pode-se afirmar que as amostras *são* estatisticamente diferentes, o que, neste caso, indica que a similaridade em um par de linguagens é maior que na outra.

Diante desse cenário, este artigo propôs PortuCol, uma pseudolinguagem de programação com instruções em português assim como Portugol, porém inspirada em C ANSI. Como contribuição acadêmica, uma análise estatística da similaridade textual demonstrou que: (i) a pseudolinguagem PortuCol é estatisticamente 26,85% mais similar à C ANSI do que Portugol, o que proporciona uma transição mais natural ao estudante ao aprender linguagens do estado-da-prática como C ANSI e Java; e (ii) PortuCol é para C ANSI o que Portugol é para Pascal, o que retrata apenas uma adaptação da pseudolinguagem às linguagens do estado-da-prática. Como contribuição prática, este artigo também apresenta PortuCol2C, uma ferramenta que executa código PortuCol através da prévia tradução para código C ANSI, o que possibilita aos estudantes se familiarizem com a linguagem C ANSI através da observação do seu código PortuCol traduzido.

O *dataset* dos algoritmos utilizados na avaliação e a ferramenta PortuCol2C (binário e código fonte) estão publicamente disponíveis em:

http://www.dcc.ufla.br/~terra/papers/2016_wei_portucol

Como trabalhos futuros, pretende-se (i) avaliar o uso da pseudolinguagem PortuCol no ensino de Lógica de Programação e Algoritmos e (ii) melhorar a ferramenta PortuCol2C com um ambiente de desenvolvimento integrado (IDE) e um depurador.

Agradecimentos

Este trabalho foi apoiado pela FAPEMIG, CAPES e CNPq.

Referências

- [1] Alfred V. Aho, Ravi Sethi, and Jeffrey D. Ullman. *Compiladores: Princípios, técnicas e ferramentas*. LTC, 2 edition, 2008.
- [2] Ricardo Luiz B. L. Campos. Lógica de programação: Há como melhorar o aprendizado fugindo dos padrões estabelecidos nos livros didáticos e adotados pela maioria dos docentes? In *XVII Congresso Iberoamericano de Educación Superior em Computación (CIESC)*, pages 22–25, 2009.
- [3] Rafael de Santiago and Rudimar Luís Scaranto Dazzi. Ferramenta de apoio ao ensino de algoritmos. In *XIII Seminário de Computação (SEMINCO)*, pages 1–8, 2004.
- [4] Cláudio Morgado de Souza. VisuAlg - ferramenta de apoio ao ensino de programação. *TECCEN*, 2(2):1–9, 2009.
- [5] Mauri Ferrandin and Simone Lilian Stephani. Ferramenta para o ensino de programação via internet. In *I Congresso Sul Catarinense de Computação (SULCOMP)*, pages 1–8, 2005.
- [6] John R. Levine, Tony Mason, and Doug Brown. *Lex & Yacc*. O’Reilly, 2 edition, 1992.
- [7] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge University Press, 2008.
- [8] António Manso, Luís Oliveira, and Célio Gonçalo Marques. Ambiente de aprendizagem de algoritmos – Portugol IDE. In *VI Conferência Internacional de TIC na Educação*, pages 969–983, 2009.
- [9] Luiz F. Noschang, Fillipi Pelz, Elieser A. de Jesus, and André L. A. Raabe. Portugol Studio: Uma IDE para iniciantes em programação. In *XXII Workshop sobre Educação em Computação (WEI)*, pages 1287–1296, 2014.
- [10] TIOBE. Programming languages index. http://www.tiobe.com/tiobe_index, March 2016.
- [11] Adriana Salvador Zanini and André Luís Alice Raabe. Análise dos enunciados utilizados nos problemas de programação introdutória em cursos de ciência da computação no Brasil. In *XX Workshop sobre Educação em Computação (WEI)*, 2012.

Práticas Interdisciplinares Apoiadas por Problemas na Formação do Técnico em Informática

Olavo José Luiz Junior¹, Luiz Fernando Delboni Lomba²

¹Instituto Federal do Paraná (IFPR)

²Instituto Federal de Mato Grosso do Sul (IFMS)

olavo.junior@ifpr.edu.br, luiz.lomba@ifms.edu.br

Abstract. *With the implementation of the Federal Institutes, was built in Brazil an educational organization that includes courses ranging from basic education to college, providing opportunities that the concepts of computing are already introduced in basic education. Technical courses of computer area should not hold only teach the "do", but to teach "know-how" and "know seek to do." The objective of the reported activity was to increase the practical scope of application of technical disciplines course curriculum in Integrated Technical Information to high school. The approach was to develop the skills of different disciplines in a joint activity conducted from the perspective of acting in a professional environment, based on the methodology of questioning. The results show the perceptions of teachers about the evaluation process and how to operationalize activities of this nature, concluding that this approach infers more meaningful to learners, overcoming the division of labor between the actions to perform and think.*

Resumo. *Com a implantação dos Institutos Federais, foi construída no Brasil uma organização pedagógica que contempla cursos que vão da educação básica à superior, oportunizando que os conceitos da computação sejam introduzidos já na educação básica. Os cursos técnicos da área da computação não devem se prender apenas a ensinar o "fazer", mas o ensinar para "saber fazer" e "saber buscar o que fazer". O objetivo da atividade relatada foi ampliar o escopo de aplicação prática de disciplinas técnicas do currículo do curso Técnico em Informática Integrado ao Ensino Médio. A abordagem foi desenvolver as competências de diferentes disciplinas em uma atividade conjunta, realizada sob a ótica da atuação em um ambiente profissional, baseada na metodologia da problematização. Os resultados apresentam as percepções dos professores quanto ao processo de avaliação e quanto à operacionalização de atividades desta natureza, concluindo-se que este tipo de abordagem infere mais significado aos aprendizes, superando a divisão do trabalho entre as ações de executar e de pensar.*

1. Introdução

O governo brasileiro propôs um modelo de proposta político-pedagógica na educação profissional, com uma organização verticalizada, que vai da educação básica ao ensino superior. É missão das escolas técnicas federais o compromisso de contribuir no desenvolvimento do país, através da criação de soluções para o desenvolvimento técnico e tec-

nológico de cada região, de forma sustentável e buscando a inclusão social (PACHECO, 2011).

Um dos desafios que este modelo oferece é levar aos diferentes níveis de ensino conteúdos que tradicionalmente são ofertados apenas em cursos de nível superior. Assim, o objetivo deste artigo é relatar as experiências de uma proposta de intervenção realizada para o ensino de disciplinas de computação em um curso técnico de nível médio integrado em informática, utilizando-se as premissas do ensino baseado em problemas.

Na Seção 2 são apresentados conceitos que envolvem o currículo integrado e o ensino dos conteúdos de computação. A Seção 3 apresenta os trabalhos utilizados como base para esta intervenção, baseado no ensino apoiado por problemas. Na Seção 4 relata-se a experiência da abordagem executada. A Seção 5 discute os resultados e a Seção 6 as conclusões do trabalho realizado.

2. Os Conteúdos de Computação e o Ensino Médio Integrado à Educação Profissional

O currículo da educação básica no Brasil não contempla conteúdos da área de Ciência da Computação, exceto nos cursos técnicos de nível médio do eixo tecnológico de Informação e Comunicação. Na abordagem tecnicista, ensinar os conteúdos de computação nos cursos técnicos é ensinar o uso de ferramentas para resolução de problemas do mercado. Há o equívoco de equiparar o ensino de computação ao ensino de informática, sendo este último normalmente direcionado ao ensino do uso do equipamento e das suas ferramentas. Quando se fala em ensino de computação, busca-se um nível em que as técnicas possam ser aplicadas na resolução de problemas, nas diversas áreas do conhecimento.

Nos cursos técnicos de informática há o registro de alunos que desistem por considerarem desnecessária a conclusão do mesmo, uma vez que não desejam seguir ou atuar nessa área técnica de formação. Nestes casos o objetivo do estudante é tão somente obter o título, através do aprender a “fazer”, enquanto deveria contemplar o conhecimento obtido, ou seja, o “saber fazer” e “saber buscar o que fazer”.

Outro aspecto a ser considerado na formação dos profissionais da computação, é a integração entre as disciplinas, valorizando o trabalho desenvolvido em cada unidade componente do currículo e mostrando a importância de cada conteúdo no contexto geral da área. Currículo é definido pelo Ministério da Educação (2007, p. 49) como “um desenho pedagógico e sua correspondente organização institucional à qual articula dinamicamente experiências, trabalho, valores, ensino, prática, teoria, comunidade, concepções e saberes observando as características históricas, econômicas e socioculturais do meio em que o processo se desenvolve”.

Nesse contexto, Frigotto *et al.* (2005) esclarecem que o currículo integrado é aquele que tem como base a compreensão do real, colocando o currículo como a organização do conhecimento, no sentido de compreender a realidade e a própria ciência na sua historicidade. Integrar o currículo é organizar o conhecimento de maneira que os conceitos sejam compreendidos em um contexto que considere o todo. Não é apenas formar para o exercício de um ofício, mas também para a compreensão do mundo do trabalho, para uma atuação crítica e autônoma, de forma que os componentes curriculares,

bem como os conteúdos, devem se complementar. Assim, busca-se desenvolver os fundamentos das diferentes ciências que propiciem aos jovens a capacidade analítica tanto dos processos técnicos quanto das relações sociais.

A ideia de formação integrada sugere superar o ser humano dividido historicamente pela divisão social do trabalho entre a ação de executar e a ação de pensar, dirigir ou planejar. De acordo com Frigotto *et al.* (2005, p. 85): “Trata-se de superar a redução da preparação para o trabalho ao seu aspecto operacional, simplificado, escoimado dos conhecimentos que estão na sua gênese científico-tecnológica e na sua apropriação histórico-social”.

Conforme o Ministério da Educação (2007, p. 41) “o que se pretende é uma integração epistemológica, de conteúdos, de metodologias e de práticas educativas. Refere-se a uma integração entre teoria e prática, entre o saber e o saber-fazer”. Em complemento, não se deve descartar as demandas oriundas do mercado de trabalho, que é dinâmico. A formação integrada deve apresentar elementos que contextualizem a realidade a ser vivenciada no exercício profissional.

Segundo Pacheco (2011, p. 15), a proposta pedagógica da educação profissional deve recusar o conhecimento exclusivamente voltado a formação da mão de obra e deve valorizar a relação do homem com o mundo: “Sua orientação pedagógica deve recusar o conhecimento exclusivamente enciclopédico, assentando-se no pensamento analítico, buscando uma formação profissional mais abrangente e flexível, com menos ênfase na formação para ofícios e mais na compreensão do mundo do trabalho e em uma participação qualitativamente superior nele”.

A escola técnica onde se ambienta esse artigo foi inaugurada no ano de 2010. O início de suas atividades não trouxe novidades apenas à comunidade local, mas também aos profissionais que assumiam os cargos de professores nos primeiros cursos ofertados pela instituição, considerando a estrutura de formação verticalizada que permite ao docente atuar em diferentes níveis de ensino, que podem ir do nível médio ao doutorado. Em geral, os docentes vinham de experiências apenas com docência na graduação em cursos de engenharias e de computação, sem experiências no ensino técnico de nível médio.

Os autores iniciaram suas atividades no curso técnico de nível médio integrado em informática com o desafio de lecionar em um nível de ensino em que nunca haviam atuado. Para superar o paradigma da fragmentação, segundo Ferreira *et al.* (2010, p. 52), é necessário que o docente tenha “uma visão sistêmica, atuando de maneira significativa, onde seu papel de pesquisador e mediador seja evidente e colabore para o desenvolvimento de um ser integral e pleno”. Os autores procuraram pautar sua atuação nessa premissa.

Os autores consideram a avaliação como um evento importante para a aprendizagem, sobretudo na educação profissional. Como diz Moretto (2010, p. 18), a avaliação é um momento privilegiado de estudo, no qual “o ensino proporciona o desenvolvimento de habilidades e a aquisição de conhecimentos, que conduzem às competências almejadas”.

A atividade prática interdisciplinar oportuniza ao estudante um *feedback* imediato sobre seu real aprendizado, bem como, de acordo com a condução da atividade, permite a reflexão sobre a melhor alternativa para resolução dos problemas, ajudando na aprendizagem significativa.

3. O Ensino Apoiado por Problemas

Com o objetivo de quebrar a abordagem tradicional das aulas, centrada no professor como transmissor dos conteúdos, buscou-se uma metodologia que coloque o aluno como o centro do processo de ensino-aprendizagem. Segundo Almeida e Moran (2005, p. 64) “na perspectiva da interatividade, o professor pode deixar de ser um transmissor de saberes para converter-se em formulador de problemas, provocador de interrogações, coordenador de equipes de trabalho, sistematizador de experiências e memória viva de uma educação que, em lugar de prender-se à transmissão, valoriza e possibilita o diálogo e a colaboração”. Sob esta óptica adotou-se uma metodologia baseada em problemas para realizar algumas das atividades de um curso técnico em informática.

A Metodologia da Problematização, proposta por Charlez Maguerez (BORDENAVE e PEREIRA, 1982 *apud* BERBEL, 1995) e baseada no “Método do Arco”, utiliza-se dos problemas como base para o processo de ensino-aprendizagem, em um circuito de etapas: observação da realidade, para identificar o problema; definição de pontos-chave, elencando os tópicos a serem investigados; teorização, para construir respostas concretas para o problema; hipóteses de solução, para alinhar os conceitos teóricos à realidade; e por fim, a aplicação prática, em que efetivamente há a intervenção sobre o problema. Segundo Berbel (1995) ela apoia-se na realidade do contexto, que observada sob diversos ângulos “manifesta-se para alunos e professor, através dos fatos concretos e daí são extraídos os problemas.”

Outra abordagem é o Método de Aprendizagem Baseada em Problemas (*Problem Based Learning* – PBL), que de acordo com Sakai e Lima (1996 *apud* BERBEL, 1998) “é o eixo principal do aprendizado teórico do currículo de algumas escolas de Medicina, cuja filosofia pedagógica é o aprendizado centrado no aluno”. O PBL é uma “metodologia formativa à medida que estimula uma atitude ativa do aluno em busca do conhecimento e não meramente informativa como é o caso da prática pedagógica tradicional.”

Apesar da semelhança entre as duas metodologias há diferenças na concepção e na execução delas. Dentre as diferenças, apontadas por Berbel (1998), destaca-se o fato de que a Metodologia da Problematização não requer grandes alterações materiais ou físicas na escola: “as mudanças são mais na programação da disciplina”, enquanto na PBL há necessidade de adequações nos espaços físicos (que devem ser suficientemente equipados e espaçosos), da equipe de trabalho, da organização dos horários, na distribuição de temas versus tempo, entre outras.

Os trabalhos na literatura sobre a abordagem apoiada por problemas na área de computação, em geral, apontam para o uso do PBL. O trabalho de Araújo *et al.* (2015) relata a experiência de um minicurso realizado por estudantes do curso de Licenciatura em Computação, utilizando a abordagem por problemas para ensinar lógica de programação aos alunos do ensino fundamental, a partir do uso de atividades da computação desplugada¹ e do software *scratch*².

¹A Computação Desplugada (*Computer Science Unplugged*) é um método que defende o ensino de computação sem o uso do computador e permite que o aluno use o pensamento computacional para resolver problemas reais.

²*Scratch* é uma plataforma de programação, com uma sintaxe simples, comumente utilizada para introduzir à programação quem não tem experiência no assunto.

Nos cursos de computação as disciplinas de algoritmos e programação, em geral, são classificadas pelos alunos como difíceis, justamente pela natureza voltada à resolução de problemas. Espelhado na experiência de Araújo *et al.* (2015), entende-se que utilizar esta abordagem em outras áreas do curso é uma oportunidade de retirar destas disciplinas a responsabilidade exclusiva de levar o aluno a raciocinar, já que “a computação, enquanto ciência tem como objeto de estudo a resolução de problemas”. Temos no bojo dos cursos técnicos a oportunidade de introduzir os conceitos da Ciência da Computação já na educação básica, disseminando o pensamento computacional.

Outra experiência é apresentada por Notari *et al.* (2009), que utilizou a Aprendizagem Baseada em Problemas, em uma disciplina do curso de Licenciatura em Computação. O objetivo era ensinar o processo de instalação e configuração de Sistemas Operacionais, com enfoque na aprendizagem colaborativa, despertando a capacidade dos alunos em “trabalhar em conjunto para identificar e analisar problemas e/ou gerar soluções”.

O trabalho de Oliveira *et al.* (2007) apresenta a experiência de um curso de Engenharia da Computação com o PBL, destacando o trabalho desenvolvido na disciplina de Arquitetura de Computadores. Adaptando Venturelli (1997), os autores apontam três adaptações para a aplicação da metodologia em cursos diferente da Medicina e que se encaixam no contexto de um curso de ensino médio integrado: integrar as disciplinas em eixos temáticos; desenvolver estudos baseados em problemas e fundamentados na realidade; e utilizar um mecanismo de avaliação diferenciada, que seja contextualizada e permita uma reflexão do aluno sobre sua prática.

Segundo Berbel (1998), a PBL impacta na reorganização curricular do curso, enquanto a Metodologia da Problematização “pode ser utilizada para o ensino de determinados temas de uma disciplina”. Assim, o presente trabalho se deu baseado Metodologia da Problematização, mas apoiado na experiência de Oliveira *et al.* (2007), a fim de aplicar as adaptações sugeridas do PBL para introduzir uma nova concepção no contexto do curso técnico em informática, sem a necessidade de grandes mudanças na estrutura do curso e da escola.

4. Relato e Metodologia da Intervenção

A primeira etapa foi identificar e relacionar as disciplinas que se encaixam no eixo temático definido como Manutenção de Computadores, considerando a estrutura curricular do curso técnico de nível médio integrado em Informática em que esta experiência foi realizada. Quatro disciplinas foram agrupadas neste eixo e são apresentadas na Figura 1, que também apresenta em qual semestre do curso cada uma delas é oferecida.

Em um segundo passo foi elaborado um plano de ação para o eixo temático, para garantir que os conteúdos de uma disciplina pudessem complementar o de outra e de que não haveria sobreposição entre eles. A Tabela 1 mostra a ementa das disciplinas que compõem o eixo.

A disciplina de Organização de Computadores é a responsável por introduzir os conceitos fundamentais do sistema computacional, com ênfase nos componentes que compõem um computador de mesa (*desktop*) e os seus detalhes técnicos. Diferente da abordagem tradicionalmente abordada nesta disciplina, aqui o foco é apresentar o contexto prático de como o técnico em informática utilizará estes conceitos. Por exemplo,

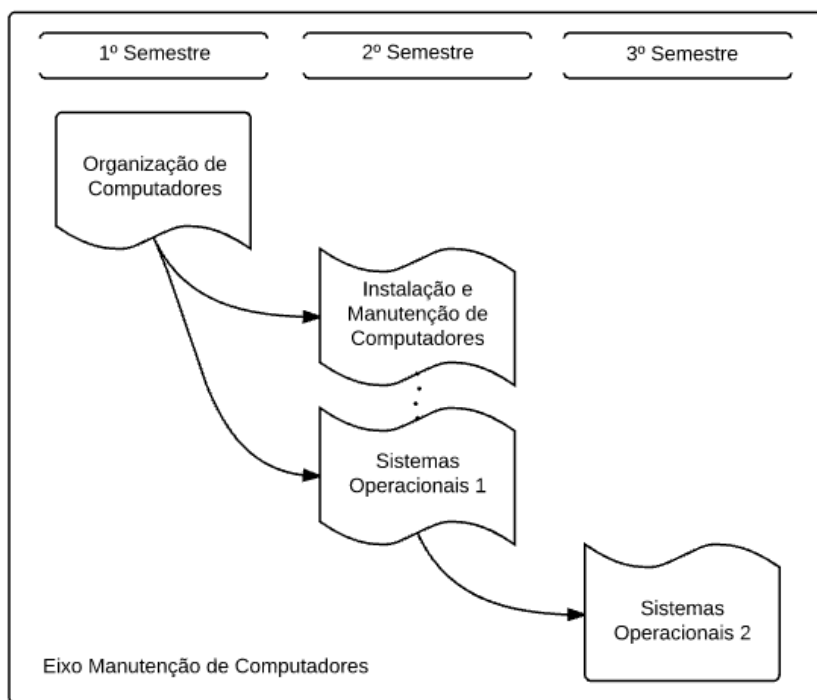


Figura 1. Organização das disciplinas no Eixo Manutenção de Computadores

na apresentação da hierarquia de memórias, o foco é permitir que o aluno identifique as características delas nos componentes do computador: nos conceitos de memória *cache*, por exemplo, o aluno deve identificar que se trata de um detalhe técnico do processador e que através dos manuais dos processadores é possível identificar esta característica.

Nas disciplinas de Instalação e Manutenção de Computadores e Sistemas Operacionais 1, o objetivo é permitir que o estudante faça a instalação e configuração de um computador: configuração da BIOS, instalação do sistema operacional, aplicativos e *drivers*, identificação de falhas físicas e de software (*troubleshooting*).

Já na disciplina de Sistemas Operacionais 2 são abordados os temas mais clássicos da disciplina, porém com o foco em mostrar a aplicação de cada um delas, através do uso de utilitários disponíveis nos sistemas proprietários e livres do mercado.

Tabela 1. Ementa das disciplinas do Eixo Manutenção de Computadores.

Eixo: Manutenção de Computadores	
Disciplina	Ementa
Organização de Computadores	Estrutura básica e conceitos fundamentais da informática. Tipos de componentes e periféricos. Critérios para seleção de hardware.
Instalação e Manutenção de Computadores	Configuração e instalação de microcomputadores e seus periféricos. Técnicas de Manutenção Preventiva e Corretiva. Realização de cópias de segurança, restauração de dados e atividades de prevenção, detecção e remoção de vírus. Seleção de soluções adequadas para corrigir as falhas de funcionamento de computadores, periféricos e softwares.
Sistemas Operacionais 1	Evolução histórica dos Sistemas Operacionais. Arquiteturas computacionais e os Sistemas Operacionais. Tipos de Sistemas Operacionais. Instalação e configuração de Sistemas Operacionais. Utilização de programas utilitários em modo texto e modo gráfico.
Sistemas Operacionais 2	Processamento. Processos, Sincronização, comunicação inter-processos. Gerenciamento de CPU, Memória e Entrada/Saída. Sistemas de Arquivos.

A etapa seguinte foi planejar as atividades a serem realizadas em cada uma das unidades curriculares, baseado em problemas. Apesar das unidades curriculares serem ministradas por docentes diversos e em dias diferentes, o plano de ação permitiu a identificação de problemas que poderiam ser abordados em conjunto, cabendo ao professor relacionar os conteúdos com os das outras disciplinas. Assim, as disciplinas de Instalação e Manutenção de Computadores e Sistemas Operacionais 1, ofertadas simultaneamente no 2º semestre do curso, foram planejadas para que ocorressem de forma integrada.

Posto esta organização, relata-se a seguir a realização de uma avaliação prática conjunta entre os dois componentes curriculares. Além de integrar os conceitos abordados das disciplinas, procurou-se avaliar as competências de ambas através de um problema conjunto, contextualizando o problema em um cenário do mercado, que simulava o ambiente de uma empresa prestadora de serviços.

O problema proposto exigia que o aluno aplicasse técnicas de manutenção corretiva (para solucionar os problemas de funcionamento do computador), manutenção preventiva (para otimizar/personalizar as configurações do computador) e de instalação de softwares (sistema operacional, aplicativos e *drivers*).

Cada aluno recebeu um modelo simulado de contrato e ordem de serviços (Figura 2) com uma série de trabalhos a serem efetuados em um computador, sendo estes complementares e interdependentes. O laboratório foi preparado pelos professores de forma a gerar problemas técnicos no funcionamento dos computadores a serem utilizados na atividade.

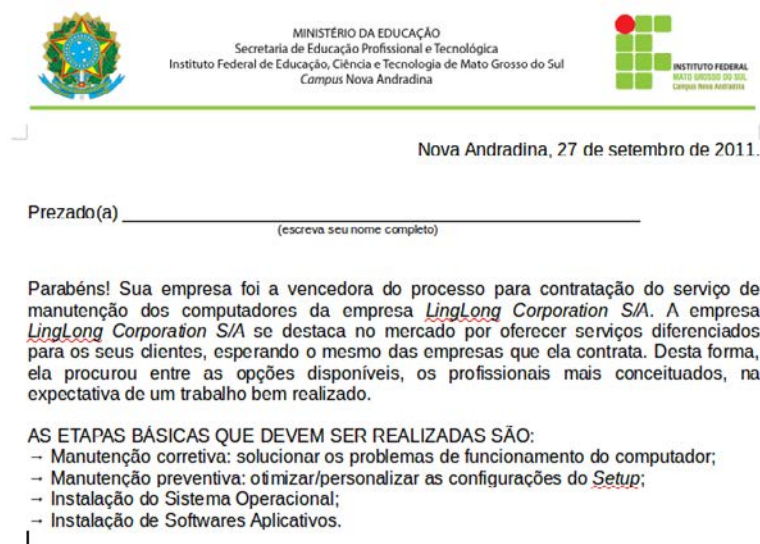


Figura 2. Fragmento do simulacro utilizado na atividade

Além da realização e finalização correta das tarefas técnicas propostas, a atividade deveria ser finalizada com o preenchimento de relatório técnico que procedesse prestação de contas ao cliente do que foi efetivamente realizado e qual a razão de cada intervenção técnica. Com o intuito de considerar outros aspectos da formação integral do indivíduo e do profissional, foram introduzidos parâmetros como compromisso e profissionalismo na atividade.

Outro objetivo focava em que o estudante pudesse mensurar o resultado financeiro do seu trabalho, além de identificar a importância de informar às pessoas que confiaram em seu trabalho, o que e o porquê do serviço que foi realizado, de forma transparente e adequada. O enunciado do exercício sugeria: “Lembrando que conforme combinado na proposta do contrato, o serviço será ‘pago’ somente após a apresentação do relatório dos serviços prestados e a conclusão dos mesmos”, de forma a traçar um paralelo entre o pagamento do serviço na vida real, com a atribuição do conceito escolar referente à atividade.

Após a realização da primeira parte da avaliação, que os alunos consideravam que era única e estaria finalizada, os professores aplicaram um novo desafio, exigindo que novas instalações/configurações fossem realizadas, fazendo um paralelo com solicitações de serviço de pós-venda muitas vezes existentes em ocorrências reais da vida profissional. Essa situação levaria à uma necessária reflexão sobre as consequências de sua intervenção técnica no médio prazo, procurando sempre executar da melhor forma as instalações solicitadas.

5. Resultados e Discussões

Quando a atividade foi revelada aos estudantes, a reação dos mesmos foi apresentar a preocupação que um possível mau desempenho na avaliação resultaria em uma “nota ruim” nas duas disciplinas. Alguns alunos questionaram se a nota seria zero caso não conseguissem finalizar a atividade, Muitos também perguntaram se a prova de recuperação seria feita em um formato similar, demonstrando contrariedade, o que evidencia a cultura ainda existente entre os estudantes de que o que esperam em uma avaliação é a nota à ela atribuída e não o processo e a aprendizagem que a envolve. Mesmo assim, ao final da atividade a percepção dos alunos foi considerá-la uma avaliação mais interessante do que o usual, haja vista que os motiva a utilizar conteúdos de componentes curriculares diversos.

Quanto aos professores, verificou-se que o formato das atividades permitiu uma maior facilidade por parte dos estudantes em demonstrar de fato a aprendizagem e isso se refletiu naturalmente em uma melhoria no resultado, em comparação com avaliações realizadas anteriormente.

Para realizar esta atividade, foram necessárias algumas alterações na organização e logística das aulas. Para este dia, excepcionalmente, as aulas das duas disciplinas foram agrupadas, a fim de permitir tempo suficiente para o estudante identificar os problemas, solucioná-los e para que escrevesse o relatório. Esta situação exigiu que outros professores colaborassem, cedendo o tempo de suas aulas. Este aspecto mostra a necessidade dos professores refletirem sobre sua prática e mudarem as características habituais de preparação das aulas, que tendem a uma rotina de repetição e reprodução (mais do mesmo...).

Em geral, o professor até quer integrar a sua disciplina, mas não quer modificar sua forma de atuar, esperando que a outra área de conhecimento se integre à sua. Uma atividade integrada exige que o professor saia de sua área de atuação e passe a encontrar pontos convergentes em outras áreas do conhecimento. A partir da abordagem por problemas, foi possível criar um ponto C, que agregou as áreas A e B, não apenas somando os conteúdos, mas criando um novo ponto a partir da convergência deles.

Esta experiência envolveu professores que ministram disciplinas diferentes mas atuam em área comum, o que acabou facilitando a execução da experiência. Destarte, a organização do eixo de Manutenção de Computadores possibilitou que todos os professores envolvidos tivessem uma visão macro da área, indo além da sua unidade curricular. É provável que a agregação de conteúdos de outras áreas do conhecimento (através do trabalho dos respectivos docentes) ampliasse fortemente a necessidade de planejamento e preparação mais sofisticados da atividade, bem como demandasse um acompanhamento mais apurado dos resultados.

6. Conclusões

Os professores perceberam que o argumento utilizado, que coloca o aluno no papel de um profissional perante um problema, mostrou ser um formato que representou mais significado aos aprendizes sobre os conteúdos trabalhados. Com a problematização foi possível simular um ambiente mais próximo do encontrado na vida profissional, exigindo que os alunos identificassem os problemas e teorizassem hipóteses de acordo com o conhecimento obtido nas disciplinas, e aplicassem à realidade.

Esta experiência ficou restrita à integração das disciplinas de apenas uma área, sendo que novas experiências deverão ser realizadas para atingir as outras áreas do conhecimento, que envolvem a formação do aluno para uma atuação crítica e autônoma.

Quanto à metodologia utilizada, conclui-se que é necessário quebrar a abordagem tradicional utilizada, que está centrada no conteúdo e avaliada de acordo com a nota obtida na prova. O aluno precisa ser agente ativo do processo e não mero consumidor do conteúdo.

Aplicar este tipo de abordagem em momentos pontuais pode deixar o aluno desorientado. Faz-se necessário que esteja presente em outros momentos e não pontualmente em uma ou outra atividade, para que o aluno se aproprie da mesma e passe a ser o agente principal do processo. Iniciar o trabalho com um grupo de disciplinas é uma oportunidade para buscar a desmistificação do trabalho interdisciplinar, apresentando-se gradativamente como alternativa factível e viável entre todos os docentes.

Esta abordagem mostrou-se como uma oportunidade para superar a divisão do trabalho entre as ações de executar e de pensar, a partir de um problema conjunto, próximo à realidade que será encontrada pelo estudante em sua atuação profissional. Colocar o aluno como agente ativo e crítico do processo, proporcionará a sua aproximação do mundo real, sobrepondo o papel de mero executor de tarefas.

Referências

- Almeida, M. E. B. and Moran, J. M. (2005). *Integração das Tecnologias na Educação*. Ministério da Educação, Seed, Brasília, DF.
- Araújo, D. C., Rodrigues, A. N., Silva, C. V. A., and Soares, L. S. (2015). O ensino de computação na educação básica apoiado por problemas: Práticas de licenciandos em computação. In *XXIII Workshop sobre Educação em Informática - Anais do XXXV Congresso da Sociedade Brasileira de Computação*, Recife. SBC.
- Berbel, N. A. N. (1995). Metodologia da problematização: uma alternativa metodológica apropriada para o ensino superior. *Semina: Ciências Sociais e Humanas*, 16(2):9–19.

- Berbel, N. A. N. (1998). A problematização e a aprendizagem baseada em problemas: diferentes termos ou diferentes caminhos? *Interface — Comunicação, Saúde, Educação*, 2(2):139–154.
- Bordenave, J. D. and Pereira, A. M. (1982). *Estratégias de ensino aprendizagem*. Vozes, Petrópolis.
- Ferreira, J. L., Carpim, L., and Behrens, M. A. (2010). Do paradigma tradicional ao paradigma da complexidade: um novo caminhar na educação profissional. *Boletim Técnico do Senac: a Revista da Educação Profissional*, 36(1):51–59.
- Frigotto, G., Ciavatta, M., and Ramos, M. (2005). *Ensino médio integrado: concepção e contradições*. Cortez, São Paulo.
- Ministério da Educação (2007). *Programa Nacional de Integração da Educação Profissional com a Educação Básica na Modalidade de Educação de Jovens e Adultos*. Ministério da Educação, Brasília.
- Moretto, V. P. (2010). *Prova: Um Momento Privilegiado de Estudo, Não Um Acerto de Contas*. 9 ed. Lamparina, Rio de Janeiro.
- Notari, D. L., Bohn, J., and Boff, E. (2009). Uma abordagem baseada em problemas para aprendizagem colaborativa de sistemas operacionais. In *XVII Workshop sobre Educação em Informática - Anais do XXIX Congresso da Sociedade Brasileira de Computação*, Bento Gonçalves. SBC.
- Oliveira, W. L. A., Arruda, G. H. M., and Bittencourt, R. A. (2007). Uso do método pbl no ensino de arquitetura de computadores. In *Proceedings of the 2007 International Conference on Engineering and Computer Education*, Santos. International Conference On Engineering and Computer Education.
- Pacheco, E. (2011). *Institutos Federais: uma revolução na educação profissional e tecnológica*. Moderna, Brasília.
- Sakai, M. H. and Lima, G. Z. (1996). Pbl: uma visão geral do método. *Olho Mágico*, 2(5/6):51–59.
- Venturelli, J. (1997). *Educación médica*. Serie PALTEX salud y sociedad 2000. Organización Panamericana de la Salud, Washington.

Processo de Desenvolvimento de Software: uma Atividade Prática Supervisionada baseada nos papéis de adquirente e fornecedor

André Roberto Ortoncelli¹, Marisângela Pacheco Brittes¹

¹Universidade Tecnológica Federal do Paraná - (UTFPR) Câmpus Dois Vizinhos
Estrada p/ Boa Esperança, S/n - Zona Rural, Dois Vizinhos - PR, 85660-000

Abstract. *This paper presents an experience report concerning a work applied as Activity Supervised Practice, in a Software Development Process discipline, offered in a Software Engineering undergraduate course. In this activity the students were divided in groups, each group at different times represented the software acquirer and supplier roles (such as is defined in ISO/IEC 12207:2008), and the interaction between different groups it was compulsory. During the development of the activity, the groups had to follow some of the MPS.BR requirements, thus reinforcing concepts of software process quality learned in class. It is noteworthy that according to the opinion of students (obtained through a questionnaire), this work contributed to acquisition and fixation of the knowledge about the discipline, as well as to keep them motivated during its development.*

Resumo. *Esse trabalho apresenta um relato de experiência referente a um trabalho aplicado como Atividade Prática Supervisionada, em uma disciplina de Processo de Desenvolvimento de Software, ofertada em um curso de graduação em Engenharia de Software. Nessa atividade os alunos foram divididos em grupos, sendo que cada grupo, em diferentes momentos representou os papéis de adquirente e fornecedor de um software (tal como é definido na norma ISO/IEC 12207:2008), sendo obrigatória a interação entre diferentes grupos. Durante o desenvolvimento da atividade os grupos tiveram que seguir alguns dos requisitos do MPS.BR, reforçando assim conceitos de qualidade de processo de software aprendidos em aula. Destaca-se que de acordo com a opinião dos alunos (obtida por meio de um questionário), esse trabalho contribuiu para aquisição e fixação dos conhecimentos da disciplina, assim como para mantê-los motivados durante seu desenvolvimento.*

1. Introdução

Cursos de graduação em Engenharia de Software são relativamente novos no Brasil. Tal curso, envolve conhecimentos sobre desenvolvimento, implantação e gerenciamento de projetos de software e também conhecimento para uso de processos organizacionais, passando pela infraestrutura e manutenção de um software, visando um perfil de egresso que deve dominar técnicas de desenvolvimento e aprimoramento de processos e produtos de software.

A disciplina de Processo de Desenvolvimento de Software (PDS) envolve conceitos de produto, processo e projeto de software, apresentando diferentes processos

de desenvolvimento (ágeis e tradicionais), e também modelos que qualidade de processo tal como MPS.BR e CMMi. Trata-se de uma disciplina que dificilmente é encontrada em cursos de graduação na área da computação, apesar de seus conteúdos não serem novos, eles comumente estão divididos em outras disciplinas. De acordo com [Kuhrmann et al. 2013], apesar de processos de desenvolvimento de software serem uma parte essencial da Engenharia de Software, eles são apenas ligeiramente incluídos em currículos de cursos de desenvolvimento de software típicos.

Os conceitos relacionados a PDS são de extrema importância na formação do futuro engenheiro de software, devendo ser considerado alguns requisitos básicos para que seu processo de ensino e aprendizagem seja eficaz, tal como os apresentados em [Kuhrmann et al. 2013], que define cinco requisitos para um curso de PDS : (i) Os alunos precisam compreender a importância de modelos de processo; (ii) Os alunos precisam compreender a integração dos modelos de processos em um ambiente organizacional; (iii) Os alunos precisam experimentar modelos de processos realistas em termos de diferenças de tamanho e complexidade relacionados com as pequenas, médias e grandes organizações e projetos; (iv) Os alunos precisam de conhecimento diferenciado sobre as diferentes abordagens de modelagem; e (v) Os alunos precisam aprender os modelos de processos mais relevantes existentes (tais como padrões nacionais ou internacionais).

Além dos requisitos apresentados em [Kuhrmann et al. 2013], a disciplina de PDS também enfrenta uma série de desafios comuns ao processo de ensino e aprendizagem de Engenharia de Software [Paiva 2011]: (i) distância entre o conteúdo ensinado e a realidade do mercado de trabalho; (ii) falta de interesse do discente em disciplinas maciçamente teóricas; (iii) altas taxas de evasão e reprovação; (iv) índices de aprendizagem efetiva bastante questionáveis; (v) pouca ênfase no trabalho em grupo; (vi) pouco estímulo à interdisciplinaridade; (vii) o planejamento de conteúdos e práticas pedagógicas são ditados pelo interesse de cada docente sem passar, muitas vezes, pelo crivo institucional; (viii) de forma geral, o que se ensina e aprende em um semestre ou disciplina, não é apropriado pela comunidade como um patrimônio intelectual para uso futuro; (ix) pouca ou nenhuma interação entre professores e alunos fora da sala de aula; (x) o conhecimento produzido em um curso/disciplina não é compartilhado com a comunidade externa; e (ix) cada instituição / docente segue sua própria linha de acordo com a sua visão/concepção.

Portanto, ministrar e planejar uma disciplina de PDS é uma tarefa que conta com uma série de desafios, além dos desafios apresentados em [Kuhrmann et al. 2013] e [Paiva 2011], outro desafio importante é não abordar de forma aprofundada conteúdos que são trabalhados com maior ênfase em outras disciplinas, e da mesma forma não se pode abordar de forma superficial conteúdos que serão trabalhados com menor intensidade em outras matérias, tal como a disciplina de Gerência de Projetos, que pode abordar a gestão no contexto de metodologias ágeis, e a disciplina de Qualidade de Software que irá abordar conceitos de qualidade de processo, entre outras matérias de um curso de Engenharia de Software.

A disciplina de PDS contemplada nesse relato de experiência foi ministrada no segundo semestre de um curso de Engenharia Software, sendo assim os alunos ainda não haviam aprendido todos os conceitos relacionados a processos de desenvolvimento de software (por exemplo, os alunos ainda não haviam tido aulas de teste de software, análise e projeto orientado a objetos e gerência de configuração), e não há tempo para se

aprofundar devidamente nesses conteúdos durante a disciplina de PDS. Nesse contexto, destaca-se o desafio de manter os alunos interessados e motivados em relação as aulas da disciplina e também no desenvolvimento da Atividade Prática Supervisionada (APS).

Uma APS consiste em uma ou mais atividades extraclasse, que devem ser acompanhadas pelo professor da disciplina, pois integralizam a carga horária da disciplina, podendo compor também a nota do aluno na disciplina.

Nesse contexto, o objetivo desse trabalho é apresentar um relato de experiência referente a aplicação de uma APS, na disciplina de PDS, baseada na interação entre diferentes grupos de alunos, que em diferentes momentos trabalham como adquirente e fornecedor de um processo de desenvolvimento de software (permitindo assim simular um processo de desenvolvimento de software real), de acordo com a norma NBR ISO/IEC 12207:2008 [ISO 2008], tendo que lidar com requisitos de qualidade do MPS.BR [SOFTEX 2012].

O restante desse trabalho encontra-se organizado da seguinte maneira: A seção 2 apresenta uma breve revisão de literatura referente ao tema. A proposta metodológica da APS que é foco desse relato de experiência é apresentada na seção 3. O questionário de avaliação utilizado e resultados obtidos com sua aplicação serão apresentados na seção 4, sendo uma análise de tais resultados apresentada na seção 5. Por fim, a seção 6, apresenta uma breve conclusão.

2. Revisão de Literatura

Tradicionalmente o ensino de Engenharia de Software se dá por meio de aulas teóricas e aplicação de exercícios práticos. Porém recentemente vêm sendo introduzidas alternativas lúdicas nesse processo, tais como jogos e simuladores. Nesse contexto em [Kohwalter et al. 2014] é apresentada uma análise positiva em relação a como tais técnicas podem levar a respostas mais rápidas e precisas dos estudantes, incluindo os aspectos de aprendizagem que não poderiam ser alcançados da forma tradicional.

Com relação á Engenharia de Software, de forma geral, pode-se destacar vários trabalhos recentes que tem como objetivo tornar o processo de ensino e aprendizagem mais efetivo e motivador. No entanto, quando se foca apenas na área de PDS, o número de trabalhos é ainda um pouco limitado. O objetivo dessa seção é listar alguns desses poucos trabalhos existentes na literatura.

Em [Prikladnicki et al. 2009] são apresentadas diferentes estratégias para o ensino de Engenharia de Software, sendo uma delas focada no uso de LEGO para o ensino de processo de desenvolvimento de software, abrangendo Normas ISO, processos ágeis e de forma geral os modelos de qualidade CMMI e MPS.BR, objetivando consolidar os conceitos dos processos de gerência de projetos, gerência de requisitos, garantia da qualidade, medição e gerência de configuração do MPS.BR. Na atividade apresentada, um produto deve ser construído a partir dos requisitos de forma planejada, monitorada e controlada, de forma similar a um processo de desenvolvimento de software, porém com uso de LEGO.

Técnicas de gamificação são exploradas na abordagem proposta em [Cavalcante et al. 2015], que conseguiu um aumento de 40% na aderência aos processos organizacionais e de 214% na capacitação dos colaboradores. Nesse trabalho se

desenvolveu uma competição lúdica envolvendo todos os projetos de uma organização, com o objetivo de melhorar alguns aspectos relacionados a qualidade do produto. Para pontuação no jogo foram considerados itens em comum a todos os projetos da organização, tais como aderência aos processos organizacionais e capacitação dos colaboradores nos processos.

Em [Werneck et al. 2012] é apresentada uma abordagem interdisciplinar para o ensino de Engenharia de Software para alunos de três disciplinas distintas. Nessa abordagem, a partir da definição de um processo formal centrado nas atividades de gerência de projetos, os alunos da disciplina de Projeto de Sistemas de Informação são responsáveis por desenvolver um sistema com base em uma especificação de requisitos recebida no início do semestre, seu trabalho é gerenciado por um grupo de alunos da disciplina de Gerência de Projetos, e alunos da disciplina de Engenharia de Software II realizam a análise de pontos de função e atuam também em atividades relacionadas a qualidade de software, tal como, inspeção de código.

Outra, atividade interdisciplinar foi aplicada por [Aguilar et al. 2015], na disciplina de Engenharia de Software Aplicada, foi solicitado que os alunos utilizassem um processo de desenvolvimento de software para produção de um software para resolver um problema não fictício, e na disciplina de Gestão da Qualidade de Software os alunos foram guiados na produção e aplicação de um plano de avaliação de qualidade, que avaliou entre outros itens a qualidade do processo de desenvolvimento utilizado e do software codificado.

Em [Bavota et al. 2012] também é apresentada uma atividade interdisciplinar que visa explorar o cenário real de desenvolvimento de software. Alunos da disciplina de Engenharia de Software (de curso de graduação) e alunos da disciplina de Gerenciamento de Projeto de Software (de um programa de mestrado), tiveram que interagir entre si, sendo que os alunos de graduação tiveram funções de desenvolvimento, enquanto os alunos de mestrado foram os responsáveis pelo gerenciamento, visando simular um cenário real, no qual a gestão de qualidade foi um dos conceitos chave.

Além dos trabalhos citados, que tem uma ênfase maior em processos tradicionais, existem na literatura alguns trabalhos focados no ensino de processos ágeis de desenvolvimento de software [Gestal and Barros 2014, Rodriguez et al. 2015].

Em [Rodriguez et al. 2015] é apresentado um ambiente virtual intitulado Virtual Scrum, que utiliza recursos tais como quadros, gráficos e calendários, para simular a sala de uma equipe Scrum. Além disso também permite que avatares 3D interajam uns com os outros. Simulando de forma realista um processo de desenvolvimento ágil com a metodologia Scrum.

Também focado em metodologias ágeis, em [Gestal and Barros 2014] é apresentado um modelo de simulação (e também um criador de simulações) de curta duração (duas a quatro horas) para o ensino da metodologia de desenvolvimento Scrum, com ênfase na tomada de decisões e feedback.

Outro trabalho que simula um processo de desenvolvimento de software por meio de um jogo é apresentado em [Monsalve et al. 2015], nesse jogo o aluno pode assumir diferentes papéis durante o processo de desenvolvimento de um software, lidando com vários problemas recorrentes do desenvolvimento, ao explorar conceitos de complexi-

dade, tamanho e qualidade do produto, de produtividade e maturidade da equipe de desenvolvimento, de gestão de orçamento, de riscos do desenvolvimento, qualidade e aceitação do produto final. Além disso, tal jogo também pode ser jogado de forma individual e coletiva, sendo dividido em rodadas nas quais, os jogadores executam uma série de ações, participando de forma ordenada.

3. Procedimento Metodológico

A APS descrita nesse texto foi aplicada no segundo semestre do ano de 2015, na disciplina de PDS para uma turma de 29 alunos, e está sendo aplicada novamente no primeiro semestre de 2016, para uma turma de 32 alunos.

Nessa APS, os alunos foram divididos em grupos de 3 ou 4 integrantes. Tais grupos interagiram entre si, pois em diferentes momentos representaram os papéis de adquirente e fornecedor em um processo de desenvolvimento de software, de acordo com a norma ISO/IEC 12207:2008 [ISO 2008].

Sendo $S = \{s_1, s_2, \dots, s_x\}$ o conjunto dos x temas de software propostos, e $G = \{g_1, g_2, \dots, g_n\}$ o conjunto dos n grupos formados na disciplina de PDS, por meio de sorteio foi definido quais grupos deveriam interagir entre si.

Inicialmente, para cada software $s_z \in S$ foi sorteado um grupo $g_t \in G$, que devia representar em um primeiro momento o papel de adquirente do software s_z , sendo que em uma fase posterior um $g_y \in G$ foi sorteado para atuar como fornecedor do software s_z , considerando as restrições impostas pelo grupo g_t .

A interação entre os grupos é dividida em quatro fases: (i) solicitar proposta; (ii) desenvolver proposta; (iii) avaliar proposta; e (iv) seminário final. A Figura 1 ilustra a sequência de atividades utilizada na APS, que serão descritas com mais detalhes a seguir, respectivamente nas subseções 3.1 a 3.4.

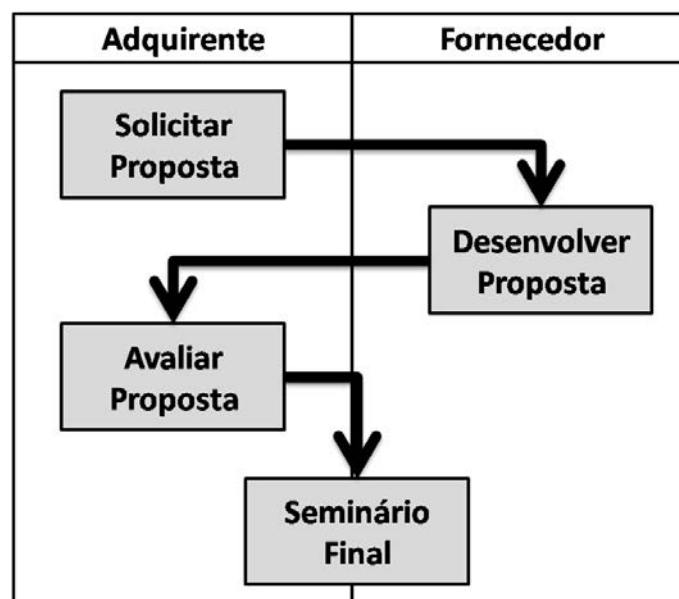


Figura 1. Fases da APS

3.1. Solicitar Proposta

Nessa fase, para cada um dos temas de software contidos no conjunto S , foi sorteado um grupo $g_t \in G$, que foi o responsável por redigir um pedido de proposta para software referente ao tema sorteado, atuando como adquirente.

O pedido de proposta desenvolvido devia contar obrigatoriamente com as quatro seguintes seções: (i) requisitos funcionais do software; (ii) restrições técnicas do software; (iii) responsabilidades dos envolvidos; e (iv) solicitação do protótipo de duas das telas principais do software.

Além de um nome, cada tema contido no conjunto S , contou com uma breve descrição, que devia ser usada como base para os grupos na redação de seus respectivos pedidos de proposta. Por exemplo, o tema software para gerenciamento de eventos, contou com a seguinte descrição: "Uma universidade precisa de um software para gerenciar os eventos que ela organiza. O software deve permitir o cadastro de cursos e palestras, e também gerar a programação com os horários do evento. Tal software também deve permitir a inscrição de alunos no evento, o controle de presença, e a emissão de certificados:

3.2. Desenvolver Proposta

Em uma data pré-definida, todos os grupos deviam entregar os seus respectivos pedidos de proposta, e então, para cada pedido de proposta foi definido por meio de sorteio outro grupo responsável por preparar uma proposta para o desenvolvimento do software solicitado.

Além dos requisitos funcionais e não funcionais do software, nessa fase, cada um dos grupos recebeu um conjunto de requisitos de qualidade de processo, que constam no modelo MPS.BR e também um modelo de processo tradicional (espiral ou prototipagem) que foram utilizados como base para a proposta de um modelo de processo.

Com relação aos requisitos de qualidade de processo, considerou-se apenas os níveis de maturidade G e F do MPS.BR, sendo que cada um dos grupos devia propor um processo de desenvolvimento baseado em um dos processos desses dois níveis (definido por sorteio). O nível G é composto pelos processos de Gerencia de Requisitos e Gerencia de Projetos, e o nível F contempla os processos de Aquisição, Gerencia de Configuração, Gerencia de Portfólio de Projetos, Garantia de Qualidade e Medição.

Destaca-se que tanto o modelo de processo que devia ser seguido, quanto o processo contemplado em dos dois níveis do MPS.BR, foram definidos por meio de sorteio para cada um dos grupos.

A proposta entregue nessa fase deveria contemplar cinco itens: (i) uma proposta de um processo de desenvolvimento de software, modelado com a notação BPMN (Business Process Model and Notation) [White 2008]; (ii) uma descrição textual de cada uma das atividades do processo; (iii) Uma matriz de rastreabilidade que associasse cada resultado esperado de um dos processos do guia MPS.BR com um uma ou mais atividades do processo proposto; (iv) cronograma e orçamento para o desenvolvimento do software; e (v) o protótipo de duas telas (conforme solicitado no pedido de proposta).

3.3. Avaliar Proposta

Cada grupo g_t encaminhou inicialmente um pedido de proposta a um grupo g_y (definido por sorteio), então na fase de avaliação o grupo g_t recebeu a proposta desenvolvida pelo grupo g_y e avaliou a conformidade dessa proposta com o respectivo pedido de proposta e também com os requisitos de processo (referente ao MPS.BR e a um dos modelos de processo tradicionais) que o grupo g_y adicionou a proposta.

Na avaliação cada grupo respondeu as quatro seguintes questões: (i) o modelo de processo sorteado foi seguido corretamente?; (ii) de acordo com o MPS.BR, todos os resultados esperados do processo sorteado para o grupo estão contidos no ciclo de vida proposto?; (iii) o cronograma e orçamento estão corretos?; e (iv) os protótipos propostos estão de acordo com o solicitado no pedido de proposta?

Cada uma das perguntas citadas devia ser respondida com sim ou não, sendo que, se a resposta fosse não, ela deveria ser justificada.

Para avaliar o desempenho dos grupos nessa fase, foi analisada a conformidade da avaliação dos grupos com a avaliação realizada pelo professor da disciplina, que respondeu as mesmas quatro questões para cada uma das atividades.

3.4. Seminário Final

No seminário final, cada um dos grupos apresentou o trabalho que desenvolveu em cada uma das fases, compartilhando assim com os outros grupos, o seu respectivo processo, os erros que foram encontrados durante a avaliação, dificuldades, a metodologia utilizada, entre outras informações.

Destaca-se que no seminário final, o trabalho foi dividido em três partes (de acordo com as fases anteriores da APS): (i) preparação do pedido de proposta; (ii) preparação da proposta; e (iii) avaliação da proposta.

No momento da apresentação, por meio de um sorteio, foi definido quais membros de cada um dos grupos, deveriam apresentar cada uma das partes da APS no seminário.

4. Avaliação dos Procedimentos Metodológicos

Dos 29 alunos que cursaram a disciplina em 2015, 27 deles aceitaram responder um questionário de avaliação da APS. Tal questionário contou com 4 perguntas, sendo as três primeiras objetivas, e a última dissertativa. Para cada uma das três primeiras questões os alunos deviam escolher entre cinco respostas (de acordo com a escala de Likert [Albaum 1997]): (i) discordo totalmente; (ii) discordo; (iii) nem concordo, nem discordo; (iv) concordo; e (v) concordo totalmente.

As questões contempladas no questionário foram as seguintes:

1. O fato de a APS proporcionar interações entre grupos diferentes foi um fator motivador para o desenvolvimento da APS?
2. A APS contribui para aquisição de novos conhecimentos em relação à disciplina?
3. A APS contribui para fixação dos conhecimentos apresentados durante as aulas?
4. Você tem alguma crítica ou sugestão que gostaria de compartilhar em relação a APS?

Tabela 1. Resposta dos discentes ao questionário

	Questão 1	Questão 2	Questão 3
Discordo Totalmente	1	0	0
Discordo	0	0	0
Nem concordo, nem discordo	9	3	5
Concordo	12	10	17
Concordo Totalmente	6	14	5

A Figura 2 apresenta em forma gráfica as respostas dadas as questões objetivas do questionário de avaliação. As mesmas informações são representadas na Tabela 1, na qual cada coluna é referente a uma das três primeiras questões, e as linhas representam o número de respostas em cada uma das alternativas.

Quanto a quarta questão, os alunos relataram que tiveram uma dificuldade maior na atividade de preparação do pedido de proposta em relação as demais atividades da APS (o que já era esperado), e também sugeriram uma etapa formal de interação direta entre os grupos.

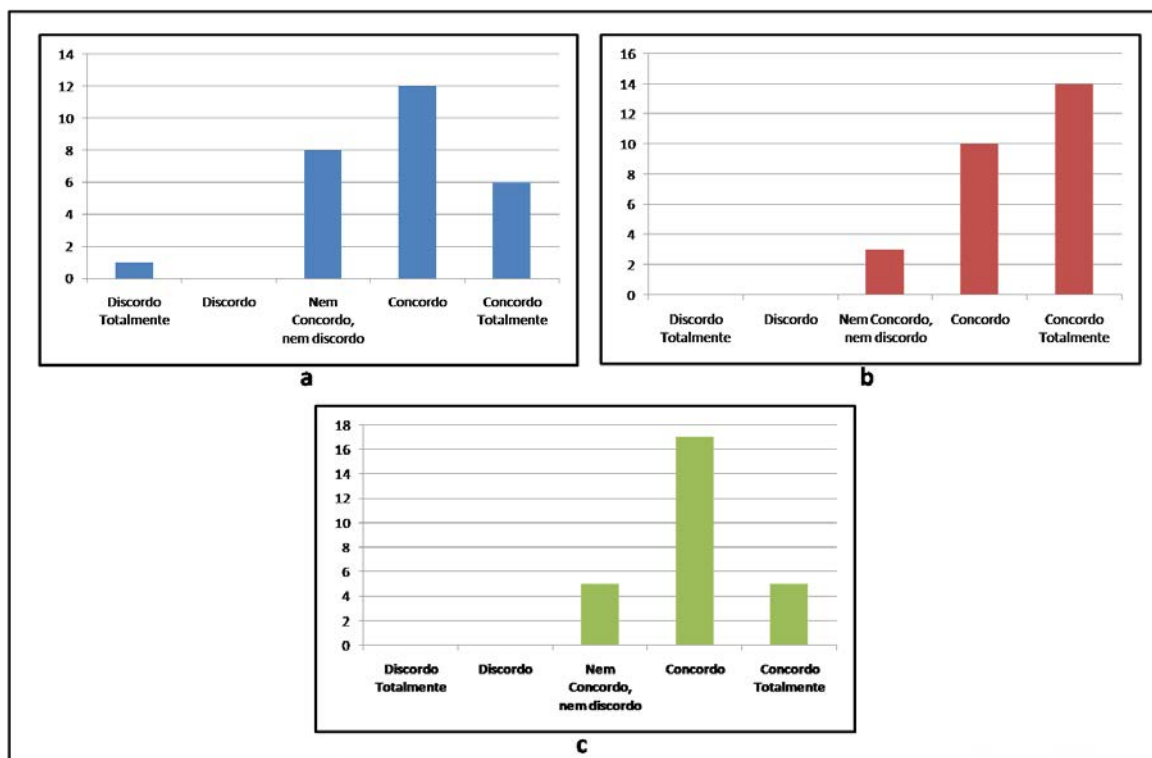


Figura 2. Respostas ao questionário: a) resposta a questão 1; b) resposta a questão 2; e c) resposta a questão 3.

5. Análise dos resultados

Observa-se que de forma geral a opinião dos alunos em relação ao procedimento metodológico da APS foi positiva, pois em todas as perguntas objetivas a maioria dos alunos

disse concordar ou concordar totalmente com as afirmações contidas nas questões objetivas.

Com relação as perguntas relacionadas a contribuição da APS para aquisição e fixação de novos conhecimentos (respectivamente as questões 2 e 3), o número de alunos que concordaram ou concordaram totalmente foi maior que 80%, sendo que com relação à aquisição de novos conhecimentos, aproximadamente 89% dos alunos tiveram uma opinião favorável.

Destaca-se que apesar de alguns dos alunos terem respondido que nem concordavam nem discordavam, com exceção a resposta de um aluno na questão 1, nenhum dos discentes discordaram ou discordaram totalmente, das perguntas contidas no questionário.

Com relação às críticas e sugestões dos alunos as seguintes medidas já estão sendo tomadas na aplicação atual da APS (primeiro semestre de 2016):

- Quanto às dificuldades encontradas na redação da proposta, atualmente na nova aplicação dessa APS foi desenvolvido um vídeo que explica aos alunos o que deve ser feito em cada atividade, e também como eles serão avaliados (no vídeo é utilizado como exemplo uma das APS's entregues no semestre anterior). Espera-se que tal vídeo diminua as dúvidas dos alunos.
- Quanto à interação entre grupos, esperava-se que todos os grupos interagissem entre si, porém se observou que alguns deles não se comunicaram. Sendo assim, na aplicação atual da APS, será dedicado formalmente um período de meia aula, ao início da atividade de preparação da proposta, para que os grupos que representam o adquirente e o fornecedor de cada um softwares possam conversar sobre os requisitos, por meio de uma entrevista, tal como, comumente ocorre em processos reais de desenvolvimento de software.

6. Conclusão

Conclui-se que foram obtidos bons resultados com a aplicação da APS apresentada nesse trabalho. A maioria dos alunos se sentiu motivada no desenvolvimento do trabalho, o que como é apresentado por [Paiva 2011], é um grande desafio em disciplinas maciçamente teóricas, como é o caso da disciplina de PDS analisada nesse trabalho, que é ofertada no segundo semestre de um curso de graduação em Engenharia de Software.

Além de motivadora, de acordo com a opinião dos alunos, a APS também serviu para fixação e aquisição de conhecimentos, a qual provavelmente contribuiu para que o índice de aprendizagem efetiva da disciplina aumentasse, o que também é um dos desafios apresentados em [Paiva 2011].

Por fim, relata-se que devido aos bons resultados obtidos com a aplicação da APS, pretende-se continuar a aplicá-la em próximas turmas da disciplina de PDS.

Referências

- Aguiar, Y. P. C., Saraiva, J., Dias, J., and Malheiros, Y. (2015). O ensino integrado de gestão de qualidade de software e engenharia de software aplicada como modelo de educação interdisciplinar: Um relato de experiência no ensino de graduação. In *Simpósio Brasileiro de Qualidade de Software*, pages 252–259.

- Albaum, G. (1997). The likert scale revisited: an alternate version. *Journal of the Market Research Society*, 39(2):331–348.
- Bavota, G., De Lucia, A., Fasano, F., Oliveto, R., and Zottoli, C. (2012). Teaching software engineering and software project management: an integrated and practical approach. In *Proceedings of the 34th International Conference on Software Engineering*, pages 1155–1164. IEEE Press.
- Cavalcante, N., Amâncio, F. D. d. S., Jucá, E. N., and Rodrigues, M. V. (2015). Uso de gamificação como auxílio para melhoria de processos: relato de experiência. In *Simpósio Brasileiro de Qualidade de Software*, pages 268–275.
- Gestal, P. R. E. and Barros, R. M. d. (2014). Proposta de um simulador para auxiliar no processo de ensino do scrum. In *Simpósio Brasileiro de Sistemas de Informação*, pages 723–736.
- ISO (2008). Iso/iec 12207 – tecnologia de informação-processos de ciclo de vida de software. *Rio de Janeiro: ABNT*.
- Kohwalter, T. C., Clua, E. W., and Murta, L. G. (2014). Reinforcing software engineering learning through provenance. In *Software Engineering (SBES), 2014 Brazilian Symposium on*, pages 131–140. IEEE.
- Kuhrmann, M., Fernández, D. M., and Münch, J. (2013). Teaching software process modeling. In *Proceedings of the 2013 International Conference on Software Engineering*, pages 1138–1147. IEEE Press.
- Monsalve, E. S., do Prado Leite, J. C. S., and Werneck, V. M. B. (2015). Transparently teaching in the context of game-based learning: the case of simules-w. In *Proceedings of the 37th International Conference on Software Engineering-Volume 2*, pages 343–352. IEEE Press.
- Paiva, S. R. (2011). Uma revisão sistemática das pesquisas realizadas sobre a melhoria no ensino de engenharia de software. *Relatório Técnico - UFPB/CCEN*.
- Prikladnicki, R., Albuquerque, A. B., von Wangenheim, C. G., and Cabral, R. (2009). Ensino de engenharia de software: desafios, estratégias de ensino e lições aprendidas. *FEES-Fórum de Educação em Engenharia de Software*.
- Rodriguez, G., Soria, Á., and Campo, M. (2015). Virtual scrum: A teaching aid to introduce undergraduate software engineering students to scrum. *Computer Applications in Engineering Education*, 23(1):147–156.
- SOFTEX (2012). Mps.br - melhoria de processo do software brasileiro: Guia geral de software.
- Werneck, M., Nelson, M. A. V., and Alonso, E. (2012). Experiências de um trabalho interdisciplinar orientado por um processo de gerência de projetos em um curso de sistemas de informação. *FEES-Fórum de Educação em Engenharia de Software*.
- White, S. A. (2008). *BPMN modeling and reference guide: understanding and using BPMN*. Future Strategies Inc.

Recursos Computacionais para Suporte ao Ensino de Teoria da Computação, Linguagens Formais e Autômatos

Ícaro Andrade Souza¹, Ecivaldo de Souza Matos¹, Débora Abdalla Santos¹,
Ranansamir Sousa¹

¹Departamento de Ciência da Computação – Universidade Federal da Bahia (UFBA)
40170-110 – Salvador – BA – Brazil

{icaro.andrade; ecivaldo; abdalla; ranansamir.silva}@ufba.br

Abstract. *According to the national curriculum guidelines, Theory of Computation (TC), Automata and Formal Languages (AFL), form a set of basic content to undergraduate courses of Computer Science. Students need to work cognitive aspects such as abstraction and theoretical and mathematical knowledge to learn that content. Over the years, it has been developed several computational resources to support teaching and learning TC and AFL. In this sense, this article presents a brief survey of software to support teaching strategies in Theory of Computation, Automata and Formal Languages, and an analysis of its mainly features and power of simulation.*

Resumo. *Teoria da Computação, Linguagens Formais e Autômatos, segundo as diretrizes curriculares nacionais (DCN), formam um conjunto de conteúdos fundamentais aos estudantes de cursos de graduação em Computação. Para aprendê-los, esses estudantes precisam trabalhar aspectos cognitivos como abstração e conhecimento teórico-matemático. Ao longo dos anos, foram desenvolvidos diversos recursos computacionais destinados a auxiliar o ensino desses conteúdos. Nesse sentido, este artigo apresenta um levantamento de softwares para o enriquecimento das estratégias de ensino de Teoria da Computação, Linguagens Formais e Autômatos, apresentando uma análise do desenvolvimento histórico e de suas funcionalidades para favorecimento da escolha/adoção desses recursos.*

1. Introdução

No Brasil os currículos de cursos superiores são normatizados pelas diretrizes curriculares nacionais (DCN), sob responsabilidade do Ministério da Educação (MEC) (Brasil, 1999). Além das DCN, há parâmetros que norteiam a implantação e revisão curriculares, como os Currículos de Referência da Sociedade Brasileira de Computação (SBC). Segundo as DCN em vigor, para todos os cursos de graduação em Computação (bacharelado/licenciatura), entre os conteúdos curriculares exigidos na formação tecnológica e básica estão Teoria da Computação (TC), Linguagens Formais e Autômatos (LFA).

Alguns autores relatam que os conteúdos curriculares referentes à Teoria da Computação, Linguagens Formais e Autômatos referem-se aos fundamentos matemáticos da Computação, os quais analisam problemas que podem ser computados por um dado modelo teórico de Computação e, de uma forma geral, respondem quais são as capacidades

e as limitações dos computadores (Hopcroft J. E., Motwani, R., Ullman, 2006; Sipser, 2005). Sipser (*ibid*) relata que esses conteúdos podem ser divididos em três partes, oriundas de três grandes teorias: Teoria das Linguagens Formais e dos Autômatos; Teoria da Computabilidade; e Teoria da Complexidade.

Pesquisadores descrevem a importância dos conteúdos curriculares referentes à TC e LFA para os cursos de Computação, enfatizando o quão essenciais são para o desenvolvimento dos estudantes nos cursos de Computação, mais especificadamente para o curso de Ciência da Computação (Chakraborty *et al.*, 2011; Dognini *et al.*, 2003; Sigman 2007). Esses pesquisadores relatam ainda que conteúdos curriculares existentes em TC e LFA dão suporte a outras disciplinas, também fundamentais aos cursos de Computação, a exemplo de Compiladores e Inteligência Artificial.

Desde meados do século passado os conteúdos dessa área já começaram a estabelecer-se como uma importante fundamentação da Ciência da Computação (Chakraborty, 2012). Entretanto, segundo Chakraborty, professores e pesquisadores já percebiam que a TC e LFA possuem temas difíceis de aprender e de ensinar.

Nesse sentido, Ezer e Trakhtenbrot (2005) sinalizam que uma das habilidades-chave necessária aos estudantes de disciplinas referentes aos tópicos de TC e LFA é aptidão para raciocínio matemático preciso, mas isso tem sido difícil e até mesmo assustador para os estudantes. Eles ainda mostram que os problemas ocorridos na tentativa de empregar métodos formais de raciocínio levam a vários equívocos e, por fim, a um conhecimento inadequado. Vijayalaskhmi e Karibasappa (2012) complementam ressaltando que um dos grandes desafios vivenciado por professores ao lecionar disciplinas dessas áreas é devido à natureza matemática do tema e à dificuldade dos estudantes em entender os conceitos de linguagem e teoria de autômatos.

Pirovani e Mataveli (2014) argumentam que os estudantes consideram os tópicos dessas disciplinas áridos, abstratos, complexos e desvinculados de suas atividades profissionais, o que contribui para reduzir o interesse e a motivação pelo seu aprendizado. Borges (2000) ressalta que o modo tradicional de ensino não consegue motivar os estudantes a se interessarem pela disciplina, entre outras razões, pois para esses estudantes não é clara a importância de certos conteúdos para a sua formação.

Outros pesquisadores, como Aguiar e Oeiras (2010), chamam atenção para os livros da disciplina, os quais afirmam que, em sua maioria, são direcionados a conceitos teóricos e matemáticos que acabam por distanciar os estudantes de uma aprendizagem por meio da prática. Os autores afirmam, ainda, que ocorre “baixo uso de ferramentas computacionais de visualização destinadas a superar dificuldades de aprendizagem em Teoria da Computação” (p.818), fato que, segundo eles, também pode contribuir para as dificuldades apresentadas por estudantes e professores.

Pirovani e Mataveli (2014) declaram ainda que “professores têm dificuldades de encontrar formas alternativas para ensinar o conteúdo e tornar a disciplina de Teoria da Computação mais interessante para os alunos” (p.53). O que, de acordo com esses autores, pode ser considerado como um ponto-chave ao desencadeamento de parte dos problemas existentes no ensino dos conteúdos de TC.

Tendo em vista os problemas relacionados ao ensino de conteúdos curriculares referentes à Teoria da Computação, Linguagens Formais e Autômatos, a pesquisa parcialmente relatada neste artigo visa identificar e analisar (i) recursos computacionais e (ii) metodologias educacionais para o ensino de conteúdos curriculares de Teoria da Computação, Linguagens Formais e Autômatos.

Todavia, neste artigo apresentamos os resultados parciais desta pesquisa: identificação e análise de recursos computacionais de apoio ao ensino de Teoria da Computação, Linguagens Formais e Autômatos em cursos de graduação em Computação.

2. Metodologia

Segundo Matos e Silva (2012) “a maioria dos artigos apresentados em conferências renomadas da área, como a conferência internacional do SIGCSE – *ACM Special Interest Group on Computer Science Education*, são reflexões acerca de experiências e introspecções dos seus autores” (p.4). Portanto, torna-se necessário conceber o Ensino de Computação (ou a Educação em Computação) enquanto área de pesquisa, fazendo com que essa nova área possua um diálogo constante e crítico com as ciências da educação (Lister, 2012).

Neste sentido, este artigo relata o processo e os resultados da identificação e análise de alguns recursos computacionais que visam auxiliar o ensino de conteúdos curriculares de Teoria da Computação, Linguagens Formais e Autômatos, cuja investigação foi realizada em duas fases: (i) levantamento de trabalhos sobre recursos computacionais de apoio ao ensino de TC e LFA e (ii) análise dos trabalhos/recursos mapeados. A primeira fase correspondeu a uma pesquisa exploratória com objetivo de mapear os trabalhos relacionados a recursos computacionais referentes a conteúdos curriculares de TC e LFA; a segunda fase tratou da análise dos trabalhos encontrados na primeira fase, buscando identificar recursos computacionais existentes para auxiliar o ensino da área.

Uma pesquisa exploratória foi desenvolvida no sentido de proporcionar uma visão geral acerca de determinado fato, procurando buscar “padrões, ideias ou hipóteses, em vez de testar ou confirmar uma hipótese” (Collis e Hussey, 2005, p. 77).

Na primeira fase da pesquisa, para o desenvolvimento da pesquisa exploratória, foram utilizados os seguintes termos de busca: ensino; Teoria da Computação; Linguagens Formais. Vale destacar que os termos foram buscados em português e em inglês. Foram analisados artigos científicos, trabalhos de dissertação e teses de mestrado/doutorado.

Foram consideradas seis fontes de dados: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD/IBICT); Biblioteca Digital Brasileira de Computação (BDBComp/SBC); Portal de Periódicos da Capes; Web of Science; ACM Digital Library; e a Revista Brasileira de Informática na Educação (RBIE). As bases foram selecionadas por indexarem a maioria dos veículos qualificados na área de Ciência da Computação, bem como pela disponibilidade e acessibilidade. A RBIE foi selecionada por ser o único periódico sobre Informática e Educação brasileiro indexado e sob a chancela da Sociedade Brasileira de Computação (SBC).

Para o desenvolvimento da pesquisa exploratória foram definidos os seguintes critérios de seleção: trabalhos de pesquisa, desenvolvimento tecnológico e/ou relato de

experiência publicados entre 2000 a 2015; escrito em português, inglês ou espanhol; acessíveis gratuitamente por pelo menos uma das fontes selecionadas; apresentar algum recurso computacional que possibilite auxiliar o ensino de tópicos referentes aos conteúdos curriculares de Teoria da Computação, Linguagens Formais e Autômatos.

Visando identificar os recursos computacionais disponíveis para auxiliar no ensino da área, foram analisados 47 trabalhos encontrados referentes a recurso(s) computacional(is) desenvolvidos para auxiliar o ensino de conteúdos relacionados a TC e LFA. Essa análise foi desenvolvida por meio da leitura completa dos trabalhos, identificando cada recurso computacional citado nos textos e identificando as seguintes informações: nome, data, tipos de máquinas que simula, se simula máquinas não determinísticas e se trabalha com outros conteúdos referentes à área. Caso o trabalho não apresentasse todas essas informações, foram realizadas buscas em outros trabalhos relacionados ao recurso com objetivo de completar as informações adquiridas.

3. Resultados

Nesta seção são apresentados os resultados obtidos a partir da primeira e segunda fase do desenvolvimento da pesquisa. Esses resultados referem-se à apresentação e análise dos trabalhos relacionados à área e dos recursos computacionais encontrados que possibilitam auxiliar o ensino de conteúdos curriculares de Teoria da Computação, Linguagens Formais e Autômatos.

3.1. Pesquisa Exploratória

A pesquisa exploratória resultou em 2.965 trabalhos referentes a recursos computacionais e/ou metodologias/métodos de ensino sobre a área. É importante salientar que foram utilizados, em algumas bases de dados, filtros secundários resultando em 2020 trabalhos para analisar. Após a análise dos respectivos títulos, resumos e palavras-chave, 86 artigos foram considerados relevantes para a pesquisa. Entretanto, verificou-se que dentre os 86 artigos selecionados, alguns estavam duplicados, a partir daí foram eliminados os trabalhos ambíguos, restando 63 artigos selecionados. Os dados preliminares dessa pesquisa exploratória podem ser visualizados na Tabela 1.

Tabela 1. Resultados obtidos com a pesquisa exploratória

Base de dados	Total de trabalhos encontrados	Trabalhos selecionados
BDTD (Dissertações)	94	0
BDTD (Teses)	66	0
Porta de Periódicos Capes	666	11
RBIE	16	4
Anais do CSBC/WEI	29	4
BDCOMP	10	10
Web of Science	42	0
ACM	1097	57
Totais	2020	63

Dentre os 63 trabalhos selecionados na pesquisa exploratória, 16 deles se referem a método(s)/metodologia(s) utilizado(s) para o ensino de conteúdo(s) curricular(es) pertencente(s) à Teoria da Computação, Linguagens Formais e Autômatos dos cursos de graduação. Os 47 trabalhos restantes descrevem recurso(s) computacional(is) que possibilita(m) auxiliar o ensino de conteúdo(s) curricular(es) pertencente(s) à área pesquisada.

3.2. Recursos Computacionais

No tocante aos recursos de apoio ao ensino de TC e LFA, foi possível identificar 80 recursos computacionais. Entretanto, para dois desses recursos (*Deus Ex Machina* e *Formal Languages and Automata Environment (FLUTE)*), não foi possível obter informações referente ao poder de simulação ou se simulavam máquinas não-determinismo; portanto, eles foram desconsiderados dos resultados finais da pesquisa, restando 78 recursos.

Para cada um dos 78 recursos computacionais então considerados, foram mapeadas as seguintes informações: *ano do último trabalho publicado referente à descrição do recurso; nome do recurso computacional; os tipos de máquinas reconhecidas pelos recursos (AF - Autômato Finito, AP – Autômato de Pilha e MT – Máquina de Turing); se o recurso computacional simula/trabalha de alguma forma conteúdos referentes a máquinas não-determinísticas; e, por fim, se o recurso apresenta outra(s) funcionalidade(s)/característica(s), dentre ela(s) esta(ão): trabalhar com algum tipo de linguagem ou gramática da hierarquia de Chomsky; bem como algoritmos de simplificação e/ou conversão de gramáticas e/ou conversão de autômatos; além de recursos que reconhecem máquinas de Moore; máquinas de Mealy; ou pumping lemas.*

A partir dessas informações, foi desenvolvido um panorama da criação/desenvolvimento desses recursos ao longo dos anos e gráficos que apresentam a relação dos recursos encontrados com o poder de simulação das máquinas referentes à hierarquia de Chomsky e a outros conteúdos da área. Esses trabalhos permitiram obter mais informações sobre recursos computacionais que possibilitam auxiliar o ensino de TC e LFA, apresentados resumidamente nas próximas subseções.

3.2.1 Panorama da criação/desenvolvimento de Recursos Computacionais

Um levantamento histórico da criação/desenvolvimento de recursos computacionais, destinados ao ensino de conteúdos referentes à TC e LFA, foi desenvolvido a partir dos dados referentes aos 78 recursos computacionais encontrados. Para tanto utilizou-se dos dados referentes ao ano de cada recurso, sendo que o ano considerado pelo levantamento refere-se ao último trabalho encontrado que descreve a criação/modificação do mesmo, uma vez que caso haja outros trabalhos referentes ao recurso, o último possivelmente trata-se da versão mais completa/estável.

O primeiro recurso computacional encontrado (*Simulation of Turing Machine on a Digital Computer*) foi desenvolvido na década de 1960. O panorama do desenvolvimento dos recursos computacionais encontrados até 2015, está apresentado no Gráfico 1.

Observou-se que nas décadas de 1960, 1970 e 1980 o desenvolvimento desses recursos foi pouco significativo com relação às décadas seguintes. Uma vez que, a partir da década de 1990 houve crescimento superior ao dobro dos recursos desenvolvidos

anteriormente. Segundo os dados obtidos na segunda fase da pesquisa, dos recursos encontrados o maior número situa-se entre os anos de 2000 e 2009. Possivelmente esse crescimento esteja relacionado à evolução tecnológica ocorrida na época, bem como ao surgimento/popularização de novos dispositivos eletrônicos, os quais possivelmente favoreceram a comunicação e a manipulação iterativa de informações.

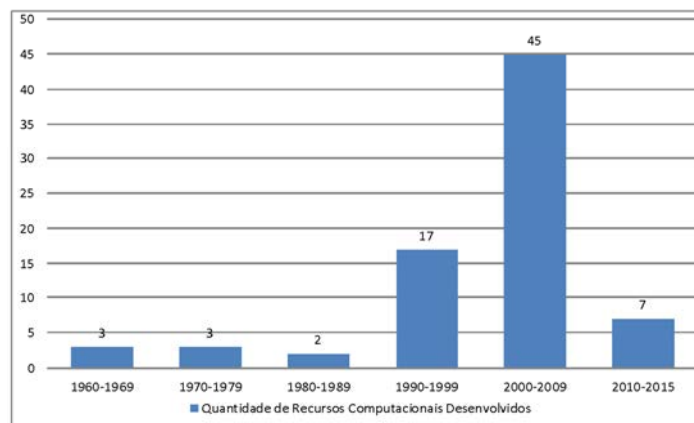


Gráfico 1. Desenvolvimento de recursos computacionais para TC e LFA ao longo dos anos

Conforme os resultados encontrados, a partir da década de 1990 houve crescimento no desenvolvimento desses recursos computacionais destinados a apoiar o ensino de TC e LFA, entretanto nos últimos cinco anos ocorreu considerável redução no número de recursos desenvolvidos.

As informações obtidas que se referem ao desenvolvimento de recursos computacionais entre os anos 2010 e 2015, *a priori* parecem contraditórias ao cenário atual, dado o desenvolvimento contínuo de novas tecnologias e consequente aumento na utilização de recursos tecnológicos em diversas áreas de ensino. Fato que suscitou os seguintes questionamentos: (i) *Por que no período entre 2000 e 2009 houve desenvolvimento desses recursos tão superior aos demais períodos?* (ii) *Por que no período de 2010 a 2015 houve uma queda significativa no desenvolvimento dos recursos computacionais para o ensino de TC e LFA?* Esses questionamentos não foram tratados nesta fase da pesquisa, entretanto poderão ser abordados em fases futuras.

3.2.2 Poder de Simulação dos Recursos Computacionais

Dentre os 78 recursos computacionais, cada um deles possui um tipo de simulação, alguns simulam estritamente autômatos finito (AF), outros autômatos de pilha (AP) ou máquinas de Turing (MT). Todavia, há recursos que simulam mais de um tipo de máquina e, por isso, podem ser utilizados para auxiliar o ensino de mais de um conteúdo curricular da área. Além disso, há recursos que auxiliam o ensino de outros conteúdos da área, como: linguagens, gramáticas, funções de conversão e simplificação, conforme Gráfico 2.

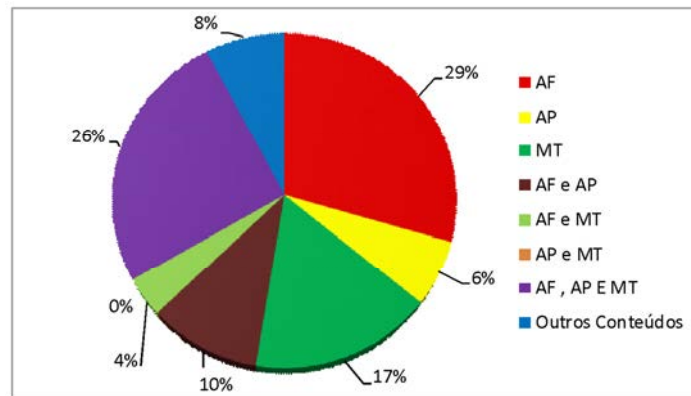


Gráfico 2. Capacidade de simulação dos recursos computacionais encontrados

Nota-se que há mais recursos computacionais destinados a simular os autômatos finitos, possivelmente por se tratar do reconhecedor das Linguagens Regulares que se refere à primeira classe da hierarquia de Chomsky, mais simples e geralmente a primeira a ser apresentada nas disciplinas de graduação relacionados à Teoria da Computação, Linguagens Formais e Autômatos.

Observou-se através dos recursos computacionais encontrados que inicialmente eles eram produzidos para simular uma máquina, seja ela AF, AP ou MT; entretanto, com o passar do tempo foram surgindo recursos que simulavam mais de um tipo de máquina. Segundo os dados obtidos nesta pesquisa temos o *Hypercard Automata Simulation*, no ano de 1992, como primeiro recurso a simular os três tipos de máquinas da hierarquia.

A partir de então é possível encontrar diversos outros recursos com poder de simulação equivalente. Conforme a busca realizada, atualmente há 20 recursos que possuem a capacidade de simular os três tipos de máquinas da hierarquia, totalizando o segundo maior grupo dos recursos encontrados. Possivelmente isso está relacionado ao fato de que quanto mais conteúdos o recurso computacional abranger, mais possibilidades de ele ser utilizado; o que, por outro lado, não pode ser utilizado como parâmetro de qualidade.

O não-determinismo é um conceito fundamental em Ciência da Computação; entretanto, por se tratar de um conceito muito abstrato é difícil de ensinar e aprender (Armoni et al., 2008). As máquinas/autômatos presentes na hierarquia de Chomsky tratam de dois conceitos: determinismo e não determinismo. Isso, segundo Armoni et al., é uma oportunidade para os alunos entenderem o significado e a importância desses conceitos.

Embora o conceito de não-determinismo seja importante para as disciplinas referentes a TC e LFA, diversos recursos computacionais não simulam máquinas não-determinísticas. Os recursos encontrados foram classificados conforme o(s) tipo(s) de máquina(s) que reconhecem (AF, AP e MT) e sua capacidade de simular as máquinas não-determinísticas (Gráfico 3).

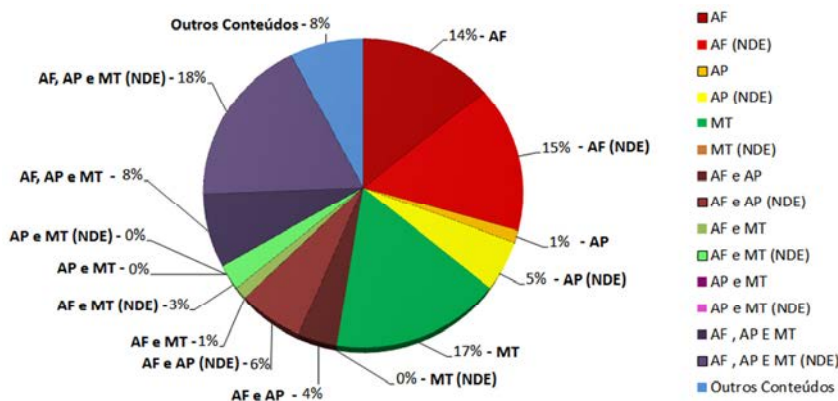


Gráfico 3. Poder de simulação dos recursos computacionais e sua relação com o não-determinismo

Observou-se que 18% dos recursos encontrados simulam as três máquinas de forma determinística e não-determinística, o que equivale a 14 recursos computacionais, dos quais sete podem ser destacados por conter outras funcionalidades referentes a conteúdos da área, são eles: JFLAP 4.0; Ambiente integrado (IE); Software para Criação e Teste de Modelos Formais (SCTMF); MTSolution; Ambiente MOODLE; Laboratório de Linguagens Formais (LabLF); e JFLAP (Adaptação Pirovani e Mataveli).

Entre outras informações pertinentes ao tema em discussão, levantadas nessa pesquisa, está o fato de nenhum dos 78 recursos computacionais encontrados ser destinado à(s) plataforma(s) de dispositivos *mobile*, ou seja, nenhum desses recursos possibilitam o ensino de conteúdos relacionados à TC e LFA por meio de mecanismos de *mobile learning* (*m-learning*). O *m-learning* é utilizado para denotar ensino através de aparelhos móveis, esse processo de ensino tem se desenvolvido muito nos últimos anos devido à popularização dos smartphones e a possibilidade de flexibilizar e otimizar o tempo de ensino/aprendizagem (Tarouco e Berch, 2009).

Através desses levantamentos buscou-se apresentar um panorama da criação/desenvolvimento de recursos computacionais que podem ser utilizados para o enriquecimento das estratégias de ensino de Teoria da Computação, Linguagens Formais e Autômatos. Chesñevar (2004) destaca que os simuladores de computador têm provado ser um excelente motivador para os alunos, ligando a teoria com a prática, incentivando descobertas e uma aprendizagem ativa. Entretanto alguns autores relatam que, embora a utilização de recursos computacionais possa trazer benefícios ao ensino de conteúdos referentes à TC e LFA, não se pode utilizar tais recursos como único método para o ensino (Elzer e Trakhlenbrot, 2006).

Schez (2009) resume relatando que os recursos de software educacionais devem ser utilizados para enriquecer estratégias de ensino, proporcionando um meio mais conveniente de exploração e *feedback* apropriado e instantâneo, sendo assim mais eficiente que os métodos tradicionais de ensino.

Como forma de difundir as informações levantadas e, conseqüentemente, facilitar a utilização de recursos computacionais no ensino de conteúdos da área, está sendo desenvolvido um espaço para disponibilização dos resumos e análises desenvolvidos nesta pesquisa. Dentre as informações disponibilizadas, para cada recurso, estarão: breve

descrição, tipos de conteúdos abrangidos, ano da última versão desenvolvida, *link* para acessar e manual para utilização.

4. Considerações Finais

Este artigo apresentou identificação e avaliação de recursos computacionais que possibilitam apoiar o ensino de conteúdos referentes à Teoria da Computação, Linguagens Formais e Autômatos em cursos de graduação em Computação. Dentre os resultados obtidos, destaca-se o resumo cronológico da criação/desenvolvimento dos recursos encontrados, o qual permitiu observar que nos últimos cinco anos houve uma queda significativa no desenvolvimento desses recursos, suscitando os seguintes questionamentos: (i) *Por que no período entre 2000 e 2009 houve desenvolvimento desses recursos tão superior aos demais períodos?* (ii) *Por que no período de 2010 a 2015 houve uma queda significativa no desenvolvimento dos recursos computacionais para o ensino de TC e LFA?*

Outro resultado a se destacar foi à avaliação da capacidade de simulação dos recursos computacionais, a qual constata-se que dos 78 recursos computacionais encontrados, 20 possuem capacidade de simular as máquinas (AF, AP e MT), dentre esses, 14 simulam as máquinas não-determinísticas. E destes, sete podem ser destacados por conter outras funcionalidades referentes a conteúdos da área. Embora os recursos computacionais possam apoiar o ensino, eles não podem ser considerados como substitutos ao professor nem a outros métodos, técnicas ou estratégias de ensino.

Como trabalho futuro, pretende-se aprofundar a análise por meio de experimentos controlados e avaliação da interação tecnológica com uma ou mais propostas metodológicas. Nesse processo serão identificados e analisados métodos/metodologias para ensino de TC e LFA. Com isso espera-se ao final desta pesquisa apresentar possibilidades metodológicas para o ensino de conteúdos curriculares de Teoria da Computação, Linguagens Formais e Autômatos.

5. Referências

- Aguiar, S. and Oeiras, J. Y. Y. (2010). Ambiente Moodle de auxílio ao ensino e aprendizagem em Linguagens Formais. XVIII Workshop sobre Educação em Informática. XXX Congresso da Sociedade Brasileira de Computação. p. 818–827.
- Armoni, M., Lewenstein, N. and Ari, M. Ben (2008). Teaching students to think nondeterministically. ACM SIGCSE Bulletin, p. 4–8.
- Borges, M. (2000). Avaliação de uma metodologia alternativa para a aprendizagem de programação. VIII Workshop de Educação em Computação (WEI). XX Congresso da Sociedade Brasileira de Computação.
- Brasil (1999). Diretrizes Curriculares Nacionais para os cursos de graduação em Computação. Disponível em: http://portal.mec.gov.br/index.php?option=com_content&id=12991:diretrizes-urriculares-cursos-de-graduacao. Acessado em: 20 de abril de 2016.
- Chakraborty, P., Saxena, P. C. and Katti, C. P. (2011). Fifty years of automata simulation: a review. ACM Inroads, v. 2, n. 4, p. 59–70.

- Chakraborty, P., Saxena, P. C. and Katti, C. P. (2012). Automata simulators: Classic tools for computer science education. *British Journal of Educational Technology*, v. 43, n. 1, p. 2011–2013.
- Chesñevar, C. I., González, M. P. and Maguitman, A. G. (2004). Didactic strategies for promoting significant learning in formal languages and automata theory. *ACM SIGCSE Bulletin*, v. 36, n. 3, p. 7–11.
- Collis, J., Hussey, R. (2005). *Pesquisa em administração: um guia prático para alunos de graduação e pós-graduação*. Porto Alegre: Bookman, p. 352.
- Dognini, M. J., Luís, A. and Raabe, A. (2003). EduLing - Software Educacional para Linguagens Regulares. XIV Simpósio Brasileiro de Informática na Educação – NCE – IM/UFRJ,
- Ezer, J. G. and Trakhtenbrot, M. (2005). Challenges in teaching the pumping lemma in automata theory course. *ACM SIGCSE Bulletin*, v. 37, n. 3, p. 369.
- Hopcroft J. E., Motwani, R., Ullman, J. D. (2006). *Introduction to Automata Theory, Languages and Computation*. International Edition.
- Lister, R. (2012). Teaching-oriented faculty and computing education research. *ACM Inroads*, v. 3, n. 1, p. 22–23.
- Matos, E. and Silva, G. Da (2012). Currículo de licenciatura em computação: uma reflexão sobre perfil de formação à luz dos referenciais curriculares da SBC. *Anais do XXXII Congresso da Sociedade Brasileira de Computação - XX Workshop sobre Educação em Computação (WEI)*.
- Pirovani, J. C. and Mataveli, G. V. (2014). Estudo e adaptação de software para o ensino de Linguagens Formais e Autômatos. *Revista Brasileira de Informática na Educação*, v. 21, p. 53–68.
- Schez, J. J. C., Castillo, E. Del, Hortolano, J. and Rodriguez, A. (2009). Designing and using software tools for educational purposes: FLAT, a case study. *IEEE Transactions on Education*, v. 52, n. 1, p. 66–74.
- Sigman, S. (2007). Engaging Students in Formal Language Theory and Theory of Computation. *SIGCSE '07: Proceedings of the 38th SIGCSE technical symposium on Computer science education*, p. 450–453.
- Sipser, M. (2005). *Introduction to the Theory of Computation*. Second Edition.
- Tarouco, L. and Berch, M. (2009). O uso de mobile learning no ensino de algoritmos. *Novas Tecnologias na Educação*, v. 7, n. 3, p. 4.
- Vijayalaskhmi, M. and Karibasappa, K. (2012). Activity based teaching learning in formal languages and automata theory-An experience. *Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on*

Sistema de Apoio à Avaliação de Atividades de Programação por Reconhecimento Automático de Modelos de Soluções

Márcia G. de Oliveira ¹, Leonardo Leal Reblin ², Elias Oliveira ³

¹ Centro de Referência em Formação e EaD (Cefor)
Instituto Federal do Espírito Santo

² Departamento de Engenharia Elétrica
Universidade Federal do Espírito Santo

³ Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal do Espírito Santo
Vitória – ES – Brasil

clickmarcia@gmail.com, elias@acm.org

Abstract. *Developing a computer program is a process of solving problems that results in several possibilities of solutions. Thus, the assessment of programming exercises demands a lot of efforts both in manual evaluation when several potential solutions are analyzed, as the automatic assessment, when many solutions should be provided as inputs. In order to assist teachers to identify solution models from programs developed by students, this paper proposes a clustering-based system to recognize solution models and to map them in scores assigned by teachers. The first application of this system experiments on two real bases of programs developed by programming students showed promising results.*

Keywords: *Solution Models, Programming, Clustering, Rubrics.*

Resumo. *O desenvolvimento de um programa de computador é um processo de resolução de problema que resulta em várias possibilidades de soluções. Dessa forma, a avaliação de exercícios de programação demanda muito esforço do professor tanto na avaliação manual, quando analisam-se várias possibilidades de soluções, quanto na avaliação automática, quando vários modelos de soluções devem ser fornecidos como entradas. Com o objetivo de auxiliar professores na identificação de modelos de soluções a partir de programas desenvolvidos por alunos, este trabalho propõe um sistema baseado em clustering para reconhecimento de modelos de soluções e para mapeamento dessas soluções em escores atribuídos por professores. Os primeiros experimentos de aplicação desse sistema em duas bases de programas desenvolvidos por estudantes de programação apresentaram resultados promissores.*

Palavras-chave: *Modelos de soluções, Programação, Clustering, Rúbricas.*

1. Introdução

O desenvolvimento de programas de computador como parte do processo de aprendizagem de programação é uma prática que resulta em várias possibilidades de soluções para

cada problema a ser resolvido. Dessa forma, o processo de avaliação de exercícios de programação demanda de professores um grande esforço de análise de diversas soluções para descobrir modelos de soluções, no caso de avaliações manuais; e de composição de modelos de soluções, no caso de avaliações automáticas que requerem entradas de gabaritos. Nesse último caso, o problema se amplia, uma vez que nem sempre o professor conhece antecipadamente todas as possíveis soluções para um exercício de programação.

Problemas similares ao problema de reconhecer modelos de soluções ou representações da diversidade a partir de um conjunto de soluções de exercícios são tratados em muitos trabalhos científicos como um processo de seleção de amostras representativas de um conjunto de padrões como, por exemplo, textos, imagens e pessoas.

Para resolver essa classe de problemas, têm-se desenvolvido, várias estratégias tecnológicas de amostragem seletiva ([Lindenbaum et al. 2004], de aprendizagem ativa [Tuia et al. 2011, Oliveira et al. 2014] e de rúbricas [Kwon and Jo 2005] para apoiar docentes na avaliação manual e automática de exercícios. Essas tecnologias, em geral, são desenvolvidas para sistemas de aprendizagem supervisionada, isto é, sistemas que aprendem através de exemplos para tomarem decisões de forma autônoma [Kotsiantis 2007].

Para o domínio da programação, uma sugestão para reconhecer a variedade de soluções é a aplicação de algoritmos de *clustering* [Naude et al. 2010]. Seguindo essa ideia, com o objetivo de auxiliar professores na avaliação de exercícios de programação, propomos um sistema de reconhecimento automático de possíveis soluções e de mapeamento dessas soluções em escores atribuídos por professores.

O reconhecimento de modelos de soluções e da diversidade destas é realizado a partir de uma coleção de programas em Linguagem C desenvolvidos por estudantes para um exercício de programação. Já o mapeamento em escores consiste em solicitar ao professor a atribuição de escores para os modelos de soluções reconhecidos. Esses modelos de soluções são, desse modo, utilizados como gabaritos de entrada de um avaliador automático ou como treino de avaliadores de aprendizagem supervisionada.

Este trabalho foca-se na seleção de modelos de soluções para composição de gabaritos de exercícios de programação. No entanto, esses modelos podem também ser utilizados por professores para gerar exemplos representativos de padrões de desempenhos informando critérios de avaliações no esquema de rúbricas [Mertler 2001].

As contribuições do sistema proposto são agilizar o processo de avaliação identificando automaticamente os potenciais gabaritos para cada exercício de programação e oferecer aos estudantes maior clareza do processo avaliativo fornecendo-lhes uma ampla variedade de soluções explicadas a partir dos escores atribuídos por professores.

Este trabalho está organizado conforme a ordem a seguir. Na Seção 2, apresentamos os trabalhos relacionados que oferecem mecanismos de seleção automática de modelos de soluções de exercícios. Na Seção 3, descrevemos a nossa proposta metodológica para reconhecimento de soluções e o algoritmo de *clustering* utilizado. Na Seção 4, apresentamos o primeiro experimento realizado em duas bases de soluções de exercícios desenvolvidas por estudantes de programação e os resultados alcançados. Na Seção 5, concluímos com as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

A estratégia tecnológica deste trabalho foi desenvolvida como uma extensão do sistema *PCodigo* [Oliveira et al. 2015b], visando reduzir esforço do especialista humano [Oliveira et al. 2014] através da seleção de amostras de treino para sistemas de aprendizagem supervisionada [Lindenbaum et al. 2004, Tuia et al. 2011, Oliveira et al. 2015a] e da apresentação de critérios de avaliação no esquema de rúbricas [Kwon and Jo 2005].

O *PCodigo* é um sistema de apoio à prática assistida de programação que, integrado ao *Moodle*, recebe soluções de atividades de programação submetidas por alunos, executa-as em massa, realiza análises e emite relatórios de avaliação para professores como, por exemplo, relatórios de plágio [Oliveira et al. 2015b].

A redução de esforço do especialista humano no fornecimento de exemplos para treino de sistemas de aprendizagem supervisionada tem sido uma temática abordada em muitos trabalhos de pesquisa. O trabalho de [Oliveira et al. 2014], por exemplo, propõe uma estratégia de aprendizagem ativa que visa selecionar uma quantidade mínima de exemplos de textos do *Twitter* para um especialista classificá-los e, a partir desses exemplos, classificar automaticamente uma grande quantidade de outros *tweets*.

Uma estratégia mais recente, que é aplicada para selecionar amostras representativas de exercícios de programação, é o método de seleção não-aleatória de amostras de treino para classificação de perfis proposto por [Oliveira et al. 2015a].

Nesse último método, visa-se reduzir o esforço humano na atribuição de escores aos exemplos selecionados e, da máquina, na classificação automática, uma vez que realiza-se redução de dimensionalidade da matriz de sessenta dimensões que representa soluções de exercícios em Linguagem C. Para a redução de dimensionalidade, utiliza-se a análise fatorial e, para a seleção de amostras que representem a diversidade de uma coleção de exercícios de programação, utiliza-se o *Clustering em Grafo*, [Oliveira et al. 2015a].

A estratégia de seleção de amostras representativas para treino de sistemas de aprendizagem supervisionada também pode ser uma opção para definir critérios de avaliação de um professor no esquema de rúbricas. Seguindo essa ideia, o trabalho de [Kwon and Jo 2005] aplica uma ferramenta que possibilita a análise das preferências dos professores e as características dos alunos para cada rúbrica, explorando a classificação e as regras de associação utilizando técnicas de *data mining*.

O diferencial da nossa estratégia em relação às estratégias apresentadas é que utilizamos processos de *clustering* ajustados pelo número de *clusters* e pelos índices de similaridade interna e externa desses *clusters* para selecionar modelos de soluções e representações da diversidade de uma coleção de exercícios.

3. Sistema de Reconhecimento de Soluções de Exercícios

O sistema de reconhecimento de soluções proposto neste trabalho é uma extensão do *PCodigo*, que é um sistema de apoio à prática assistida de programação através da execução em massa e análise de exercícios de programação em Linguagem C [Oliveira et al. 2015b].

O processamento do *PCodigo* se inicia após o professor disponibilizar via *Moodle*, na *Visão do Professor, Exercícios de Programação* e após os alunos realizarem

Submissões de soluções para esses exercícios [Oliveira et al. 2015b].

Em resumo, as etapas de processamento do *PCodigo* consistem em receber soluções de atividades de programação submetidas por alunos, executá-las, analisá-las para avaliação de indícios de plágios e emitir relatórios de avaliação para professores.

A Figura 1 apresenta a arquitetura do sistema de análise de soluções bem como a sua integração ao *Pcodigo*.

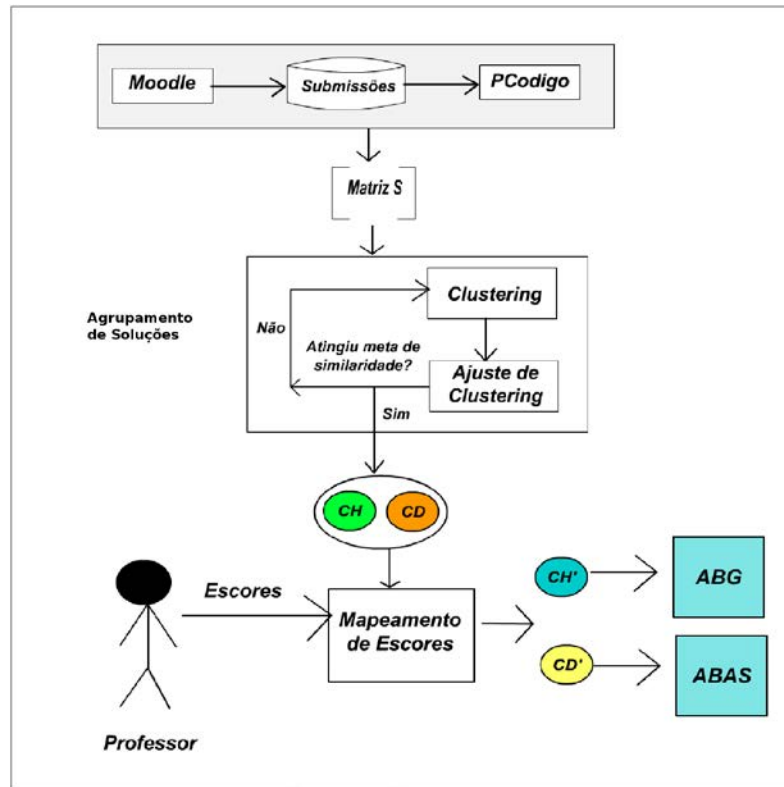


Figura 1. Sistema de análise de soluções

Conforme a Figura 1, o *Pcodigo* é integrado ao *Moodle* para receber as *Submissões* de exercícios e fornece como entrada ao sistema de reconhecimento de soluções uma *Matriz S* contendo as submissões vetorizadas em 60 dimensões. Nessa matriz, cada dimensão de um vetor representa a frequência de ocorrência de uma palavras-chave ou função da Linguagem C em código-fonte ou um valor lógico (1=Verdadeiro; 0= Falso) informando, por exemplo, se um programa compila e executa [Oliveira et al. 2015b].

No processamento de *Agrupamento de Soluções* da Figura 1, a *Matriz S* é submetida ao algoritmo de *Clustering Bissecting K-means* do software *Cluto* [Karypis 2003].

Sabendo que há mais possibilidades das melhores soluções de um exercício se assemelharem por aproximarem-se de um padrão de resolução ensinado por um professor [Naude et al. 2010], para obter um agrupamento com os melhores modelos de soluções, realizamos o *Ajuste de Clustering*. Para isso, repetimos os processos de *Formação de Agrupamentos* até que, pelo número de *clusters* escolhido, sejam obtidos *clusters* com maiores índices de similaridade interna (ISIM) e de similaridade externa (ESIM).

Após o processamento de *Agrupamento de Soluções*, são selecionados dois *clus-*

ters: os *clusters CH* e *CD*. O *cluster CH* representa o agrupamento com as melhores soluções e o *cluster CD*, o agrupamento com maior representação da diversidade.

No módulo de *Mapeamento de Escores*, os *clusters CH* e *CD* são apresentados a um professor para que este atribua escores para cada um dos padrões desses *clusters*. Em seguida, os *clusters CH'* e *CD'* contendo padrões avaliados pelo professor são enviados como entradas para os algoritmos de avaliação automática de exercícios *ABG (Avaliação Baseada em Gabaritos)* e *ABAS (Avaliação Baseada em Aprendizagem Supervisionada)*.

Os padrões de *CH'* enviados para o algoritmo de *ABG* podem ser utilizados como modelos de soluções para atribuição de escores a outras soluções muito semelhantes a estas. Já os padrões do *CD'* enviados para o algoritmo de *ABAS*, em um passo mais avançado, representando a diversidade de um conjunto de soluções composto de padrões com altos, baixos e médios escores, podem ser utilizados como treino do avaliador automático para que este aprenda o padrão de correção de um exercício e realize previsões de escores para outros exemplos de soluções desse mesmo exercício.

Uma outra aplicação da metodologia deste trabalho é utilizar os padrões avaliados dos *clusters CH'* e *CD'* como referências de avaliação para diferentes níveis de desempenhos em um esquema de rúbricas. Nesse caso, os professores apresentariam de forma mais clara seus critérios de avaliação de cada exercício de programação aplicado, dando possibilidades ao aluno de refletir sobre sua própria aprendizagem.

4. Experimentos e resultados

Para a experimentação do sistema de reconhecimento de soluções, utilizamos duas bases de soluções de um exercício de programação obtidas em turmas reais de programação de uma universidade e também utilizadas nos experimentos de [Oliveira et al. 2015a].

A primeira base, que chamaremos de *Base-A*, reúne 100 amostras de soluções de um exercício coletadas em diferentes turmas de programação. Já a segunda base, que chamaremos de *Base-B*, é uma base formada de 40 amostras de soluções do mesmo exercício utilizado na *Base-A*. No entanto, essa base é considerada mais difícil porque foi obtida em condições controladas de aplicação de prova. Desse modo, essa base apresenta uma maior diversidade de soluções e menor possibilidade de conter amostras plagiadas.

A *Base-A* e a *Base-B* foram mapeadas em vetores de 60 dimensões ou componentes de habilidades, onde cada dimensão, conforme já informamos, é quantificada pela frequência de ocorrência de palavras-chave, funções e indicadores de funcionamento (compila, executa) da Linguagem C [Oliveira et al. 2015b, Oliveira et al. 2015a]. Essas bases compostas de amostras em formato vetorial são geradas pelo *PCodigo* nos processos de normalização e indexação de submissões [Oliveira et al. 2015b].

No sistema de reconhecimento de soluções, essas bases são submetidas no formato da Matriz *S* (Figura 1) ao algoritmo de *clustering Bisecting K-means* [Karypis 2003]. O *Ajuste de Clustering* é realizado até que se alcance o melhor número de *clusters* com maiores valores de similaridade interna e externa (*ISIM* e *ESIM*, respectivamente).

Após escolher o número de *clusters* mais adequado, selecionamos o *cluster* mais provável de ter as melhores soluções (*CH*) e o *cluster* com a maior representação da diversidade de soluções (*CD*).

Para análise dos *clusters* formados, um *script* gera um relatório com a soma das dimensões de cada vetor, o que nos fornece um indicativo de esforço de programação do estudante. Em geral, as melhores soluções apresentam menos instruções e menos linhas de código, o que pode ser observado nos relatórios das figuras 2 e 3.

4.1. Resultados

Ao aumentarmos o número de *clusters* no processamento de *Ajuste de Clustering*, alcançamos valores maiores de ISIM e de ESIM para os *clusters* formados. Isso nos possibilita obter *clusters* mais homogêneos e identificar soluções atípicas, isto é, soluções que são únicas em um *cluster*.

A Tabela 1, relativa à *Base-A* de 100 amostras, apresenta como os valores de ISIM máximo e mínimo aumentam e o ESIM mínimo diminui, indicando que, aumentando o número de *clusters*, formam-se *clusters* mais homogêneos e distintos de outros *clusters*.

Análise de Similaridade ISIM-ESIM				
No. de clusters	ISim Máximo (%)	ISim Mínimo (%)	ESim Máximo(%)	ESim Mínimo (%)
1	69.8	69.8	100	100
2	80.8	76	59.5	59.5
3	87.6	76.1	67.9	59.5
4	87.6	72.2	67.9	49.2
5	92.6	72.2	70.2	49.2
6	92.6	72.2	70.5	49.2
7	92.6	72.2	70.5	49.2
8	92.6	83.8	70.5	42.0
9	92.6	79.1	72.6	42.0
10	100	85.0	72.6	30.5

Tabela 1. Análise de similaridade por número de clusters da Base-A

Para termos *clusters* mais homogêneos, conforme a Tabela 1, escolhemos dividir a *Base-A* em dez *clusters*. Na Tabela 2, apresentamos como esses dez *clusters* se organizaram conforme os seus valores de ISIM e ESIM.

Análise de similaridade dos clusters			
Cluster	Tamanho	ISim (%)	ESim (%)
0	1	100	30.5
1	2	85.4	42.0
2	4	86.8	53.1
3	5	92.2	63.4
4	21	85.2	61.8
5	22	92.6	70.3
6	8	89.9	67.7
7	10	89.1	70.2
8	14	91.3	72.6
9	13	85.0	70.5

Tabela 2. Análise de similaridade dos clusters da Base-A

Uma vez que os *clusters* são organizados de forma crescente de 0 a $N - 1$ (N = Número de *Clusters*), os maiores valores de ISIM-ESIM estão nos *clusters* identificados com valores maiores. Da mesma forma, os *clusters* mais apertados e mais distantes dos outros, isto é, com menor ESIM, são identificados com valores menores [Karypis 2003].

A Figura 2 apresenta os *clusters* 0, 1, 2, 3, 4 e 9 da *Base-A*. Nessas tabelas, *cluster* é o identificador do *cluster*, *índice* é o rótulo de cada amostra dentro dos *clusters*, *nota* é o escore atribuído pelo professor a uma solução e *somalinha* é a soma dos valores dos atributos, isto é, das dimensões de um vetor representando uma solução desenvolvida por um

aluno. Os escores altos são marcados de verde, os médios, de amarelo e os escores baixos, de lilás. Observa-se que o *Cluster 4* apresenta uma maior diversidade de escores médios e altos. Por outro lado, o *Cluster 9* apresenta-se mais homogêneo com predominância das soluções de altos escores.

Cluster	Indice	Nota	Somalinha
0	al21	3.0	7
1	al89	4.1	10
1	al60	3.75	6
2	al79	2.5	128
2	al61	4.8	67
2	al48	3.0	151
2	al64	2.0	84
Cluster	Indice	Nota	Somalinha
4	al77	3.7	39
4	al96	5.0	51
4	al52	3.25	52
4	al78	0.5	99
4	al43	5.0	57
4	al38	5.0	57
4	al53	4.75	88
4	al100	4.0	78
4	al49	4.0	51
4	al46	5.0	52
4	al25	3.25	70
4	al28	3.75	99
4	al74	5.0	56
4	al47	3.5	33
4	al72	4.0	80
4	al1	4.25	75
4	al22	5.0	51

Cluster	Indice	Nota	Somalinha
3	al12	3.75	66
3	al37	3.25	45
3	al8	4.0	37
3	al73	3.6	63
3	al19	3.5	79

Cluster	Indice	Nota	Somalinha
9	al39	4.5	46
9	al26	5.0	50
9	al34	4.75	56
9	al2	4.0	84
9	al83	4.3	44
9	al50	4.5	39
9	al62	2.5	37
9	al32	4.75	79
9	al56	1.5	113
9	al95	3.7	99
9	al69	4.75	75
9	al98	1.2	41
9	al15	3.75	104

Figura 2. Análise de clusters da Base-A

Para a *Base-A*, concluímos, portanto, que o *Cluster 9* é o mais indicado para ser o *cluster CH*, isto é, o *cluster* homogêneo com mais possibilidades de representar as soluções de altos escores, isto é, os modelos de soluções.

Já o *Cluster 4* é o mais indicado para ser o *cluster CD*, uma vez que apresenta maior variedade de soluções com escores altos e médios, que são os escores predominantes na *Base-A*.

Vale destacar que, assumindo que desconhecemos os escores das amostras reunidas nos *clusters*, nós identificamos o *cluster CH* pelo maior valor de identificação do *cluster*, isto é, pelo maior valor de ISIM-ESIM. Já, para identificar o *cluster CD*, selecionamos o *cluster* mais cheio com menor ISIM-ESIM.

A Tabela 3 apresenta como os valores de ISIM e ESIM dos *clusters* da *Base-B* variam à medida que o número de *clusters* aumenta. Para uma divisão de 10 *clusters* temos, portanto, maior ISIM máximo, maior ISIM mínimo, maior ESIM máximo e menor ESIM mínimo. Isso significa que a divisão em 10 *clusters* da *Base-A* oferece maior possibilidade de formar *clusters* mais homogêneos e bem distintos dos demais *clusters*.

Análise de Similaridade ISIM-ESIM

No. de clusters	ISim Máximo (%)	ISim Mínimo (%)	ESim Máximo(%)	ESim Mínimo (%)
1	82.8	82.8	—	—
2	90.0	78.9	73.0	73.0
3	94.7	78.9	79.4	73.0
4	100	85.1	79.4	51.7
5	100	87.3	79.4	51.7
6	100	87.3	79.7	51.7
7	100	91.6	79.7	51.7
8	100	91.6	81.8	51.7
9	100	92.9	81.8	51.7
10	100	95.0	82.8	51.7

Tabela 3. Análise de similaridade por número de clusters da Base-B

Da mesma forma que analisamos os *clusters* da *Base-A*, analisamos a *Base-B* identificando os *clusters CH* e *CD*. Mas vale destacar, conforme a Figura 3, que os *clusters* da *Base-B* apresentaram-se mais homogêneos e com as soluções inéditas isoladas nos primeiros *clusters*.

Cluster	Índice	Nota	Somalinha
0	al28	3.75	99
1	al3	3.25	101
2	al79	2.5	128
3	al22	5.0	51
4	al36	3.0	50
4	al84	3.0	82
4	al13	3.75	45
5	al87	3.2	80
5	al91	3.2	95
5	al80	3.7	98
5	al35	1.75	89
5	al63	1.5	108
6	al44	3.75	83
6	al82	2.7	99
7	al32	4.75	79
7	al33	4.5	55
7	al23	3.0	98
8	al2	4.0	84
8	al69	4.75	75
8	al40	3.5	133
8	al50	4.5	39
8	al20	3.5	88
8	al97	2.7	91
8	al26	5.0	50
8	al7	2.0	82
8	al67	3.75	78
8	al76	3.5	109
8	al94	2.7	88
8	al81	3.2	94
8	al14	3.25	112
8	al71	4.2	90
8	al9	2.75	85
8	al39	4.5	46
8	al60	3.75	6
8	al19	3.5	79
8	al29	3.5	93
9	al30	3.5	86
9	al92	5.0	46
9	al31	3.25	105

Figura 3. Análise de clusters da Base-B

Para a *Base-B*, selecionamos, conforme a Figura 3, o *Cluster 9* como o *cluster CH* e o *Cluster 8*, como o *cluster CD*. Para essa seleção, foram utilizados os mesmos critérios de seleção dos *clusters CH* e *CD* da *Base-A*.

Concluindo, os resultados da *Base-A* e da *Base-B* indicam que é possível selecionar modelos de soluções e amostras representando a diversidade de um conjunto de exercícios através de processos de *clustering*. No entanto, para gerar os conjuntos *CH'* e *CD'* com as amostras de *CH* e *CD*, respectivamente, pontuadas por um professor, demanda-se ainda esforço de correção de todas as amostras dos *clusters CH* e *CD*.

Para reduzir o esforço do professor de corrigir todas as amostras desses *clusters*, propomos, como próximos trabalhos, desenvolver estratégias para reduzir os conjuntos

CH' e CD' através da seleção das amostras mais representativas dos modelos de soluções e da diversidade de exercícios presentes nos *clusters* CH e CD , respectivamente.

5. Considerações Finais

Este trabalho de pesquisa apresentou uma proposta inicial de sistema de seleção de modelos de soluções e da diversidade de soluções de exercícios de programação. Uma vantagem desse sistema é que modelos de soluções reconhecidos podem ser utilizados como entradas de sistemas baseados em gabaritos e as soluções de representação da diversidade como treino de sistemas de avaliação automática de aprendizagem supervisionada.

Uma outra vantagem do sistema proposto é auxiliar professores na identificação de modelos de soluções de um exercício para apresentação de critérios de avaliação no esquema de rúbricas.

A principal limitação do sistema proposto é determinar o número mínimo de *clusters* que possibilita a melhor formação de agrupamentos dos modelos de soluções e *clusters* mais reduzidos.

Como trabalhos futuros a partir deste, propomos então desenvolver pesquisas para determinar esse menor número máximo de *clusters* e realizar transformações nas bases de exercícios com técnicas de redução de dimensionalidade como a análise fatorial [Oliveira et al. 2015a] de forma a melhorar o processo de *clustering*.

Com essas propostas de trabalhos futuros, teríamos boas possibilidades de formar *clusters* menores reunindo os modelos de soluções e as amostras de representação da diversidade, o que implicaria em menor esforço de correção dos professores e mais precisão na escolha dos modelos de soluções.

Concluindo, as principais contribuições do sistema proposto para a aprendizagem de programação é agilizar o processo de envio de *feedbacks* de professores, principalmente se estes utilizarem sistemas de avaliação automática, e possibilitar esclarecer critérios de avaliações a partir de exemplos corrigidos dinamicamente.

Referências

- Karypis, G. (2003). CLUTO - A Clustering Toolkit. Dept. of Computer Science, University of Minnesota.
- Kotsiantis, S. (2007). Supervised machine learning: A review of classification techniques. In *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 3–24. IOS Press.
- Kwon, H. and Jo, M. (2005). Design and implementation of the automatic rubric generation system for the neis based performance assessment using data mining technology. *Journal Of the Korean Association of information Education*, 9(1):113–126.
- Lindenbaum, M., Markovitch, S., and Rusakov, D. (2004). Selective sampling for nearest neighbor classifiers. *Machine Learning*, 54(2):125–152.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25):1–10.

- Naude, K. A., Greyling, J. H., and Vogts, D. (2010). Marking student programs using graph similarity. *Computers & Education*, 54(2):545 – 561.
- Oliveira, E., Basoni, H., Saúde, M. R., and Ciarelli, P. (2014). Combining clustering and classification approaches for reducing the effort of automatic tweets classification. In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (IC3K 2014)*, pages 465–472.
- Oliveira, M., Monroy, N., Daher, P., and Oliveira, E. (2015a). Representação da diversidade de componentes latentes em exercícios de programação para classificação de perfis. In *IV Congresso Brasileiro de Informática na Educação (CBIE 2015)*, Maceió. Anais do SBIE 2015.
- Oliveira, M., Nogueira, M. A., and Oliveira, E. (2015b). Sistema de Apoio à Prática Assistida de Programação por Execução em Massa e Análise de Programas. In *XXIII Workshop sobre Educação em Computação (WEI) - CSBC 2015*, Recife, PE. SBC.
- Tuia, D., Pasolli, E., and Emery, W. (2011). Using active learning to adapt remote sensing image classifiers. *Remote Sensing of Environment*, 115(9):2232–2242.

The teaching of functions as the first step to learn imperative programming

Carla A. D. M. Delgado¹, João C. P. da Silva¹,
Fabio Mascarenhas¹, Ana Luisa Duboc²

¹Departamento de Ciência da Computação – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brasil

²Departamento de Ciência da Computação – Colégio Pedro II
Rio de Janeiro – RJ – Brasil

{carla, jcps, fabio}@dcc.ufrj.br, alduboc@gmail.com

Abstract. *The literature on teaching programming covers a wide discussion regarding the approach to be adopted in an introductory course. While the object-oriented paradigm requires a high level of abstraction and can overshadow basic concepts, the structured imperative paradigm lacks a guideline to motivate and drive the process of building programs. In this paper we present a modularization based approach for teaching programming to the novices using the imperative paradigm. The construction of self contained modules of code - programming functions in Python - is worked with the students from the very first class until the end of the course. We describe our experience and report the results obtained from the implementation of this approach in a scenario of more than a dozen courses offered each semester over three years.*

1. Introduction

The teaching of algorithms and programming is an area of study and research that emerged in the 1970s and gained momentum in the 1980s [Robins et al. 2003]. Currently, many careers such as engineering, mathematics, meteorology and biology, require computing knowledge and programming skills. In higher education, an introductory programming course is of fundamental importance for any student in the area of exact sciences [McKeown and Farrell 1999], and in some countries, like Brazil, it is in the university that many students have their first contact with the computer, not only as users but also as developers.

The demand for professionals qualified in programming in several areas beyond computer science is growing, and should continue to grow in the coming years [ACM and IEEE 2013]. However, introductory programming courses (CS1 and CS2¹), have high failure rates [Nikula et al. 2011, Mcgettrick et al. 2005, Delgado et al. 2004]. Given the importance of the topic and the major challenges involved, the literature on teaching programming often addresses the restructuring of CS1 and CS2, mainly due to the choice of the programming paradigm and programming language to be adopted [Koulouri et al. 2014]. The most considered options of programming

¹CS1 and CS2 are the usual names for the Computer Science 1 and 2 courses, that cover introduction to programming in universities

paradigm are the Structured Imperative (SI) and the Object-Oriented (OO). The Functional paradigm appears less frequently.

The classical approach of programming teaching is based on the SI paradigm. The focus is on the teaching of the programming language structures and the mechanisms of storage and retrieval of information (reading and writing data to memory using variables). This approach allows an exhaustive exploration of these concepts, which are considered fundamental concepts for both SI programming as for OO programming (OOP). This is its greatest advantage. The OO approach, despite having the advantage of being very versatile and adopted in many software development projects, requires a higher level of abstraction and also a more extensive and elaborate syntax [Koulouri et al. 2014]. This approach, however, has the advantage of directly providing a program design technique.

For over more than 10 years teaching this course in different institutions, we accumulated experiences and intuitions about teaching programming for different profiles of novice students. During this time, we had the opportunity to evaluate students who took CS1 based both on SI and OO approaches², and with different lecturers. We observed that, in general, students who had good grades in introductory courses based on the SI paradigm progressed smoothly when faced with more advanced programming challenges, and even with OOP in subsequent courses. However, the percentage of failure and dropout in courses that adopt the SI approach is historically high.

On the other hand, we noticed high incidence of students who, despite having obtained good grades in CS1 based in the OO approach, not only had difficulties in understanding simple algorithms and programs written in the SI paradigm, but had also difficulty in developing more complex programs in the OO paradigm itself. The problem seems to get worse with the use of a "black-box" teaching approach, where the student uses pre-built packages and classes, with little opportunity to build his or her own code from the language primitives. Our hypothesis is that basic concepts are not well assimilated because they are overshadowed by the OO abstraction. Our perception from previous experience led us to conclude that a good understanding and mastery of the structures and mechanisms of SI programming is essential to build a foundation for the abstraction required by the OOP.

While driving a reformatting process for CS1 and CS2 with the aim to improve both learning and operationalization of these courses (about 20 classes each semester), we put up a didactic proposal based on the existing literature and the contributions of our colleagues from the Computer Science Department in Federal University of Rio de Janeiro (DCC/UFRJ). Our proposal was further refined using the feedback we received during the following three years of its implementation, when we had several lecturers working according to it in our institution. The SI paradigm is our choice for the novices. But our approach deviates from the classics: instead of starting the course with the writing of complete programs, we start with the construction of programming functions. Our goal is to provide the student a guideline for the more abstract cognitive tasks of design and construction of programs, from the very beginning of the learning process. OOP is left to CS2. Regarding the programming language, we choose Python, mainly because of its simplicity.

²Experiences with the Functional paradigm were fewer and are not worth documenting here.

In this paper we present our didactic proposal for introductory programming courses and the results obtained with its implementation over three years in several classes of different university courses of STEM³ at UFRJ. We also report our experience as coordinators of an initiative for the standardization of these course classes. Our proposal, although simple, allowed to give the students a more modern, cohesive and apparently more interesting format to CS1, without losing focus on conscious learning of the basics concepts. The adoption of the OO paradigm in CS2 also made up for a cohesive and motivating course, providing an appropriate scenario for tackling the use of graphical user interfaces and other more elaborate programming apparatus.

2. Related Works

[ACM and IEEE 2013] presents the three most common approaches to introductory programming courses: "programming first" - the classical approach, with the imperative paradigm; "Objects-first" - the object orientation is introduced early in a programming course; "functional-first" - programming concepts are introduced with a functional language. Arguments for and against each of these approaches have been raised. For example, a strength of the approach "objects-first" is that, from the beginning, students have contact with programming principles widely used in industry. In contrast, a point against it is that the object orientation requires an abstract design and syntactic overhead that can complicate the programming task. The discussion does not reach a final verdict in favor of any particular option.

[Bruce 2005] summarizes a discussion that took place on the SIGCSE mailing list in March 2004 on how to teach introductory programming using Java. The article mainly focuses on one question: when objects should be introduced in CS1. The options discussed were (1) OO concepts should be introduced at the beginning of the programming learning; (2) OO concepts should be introduced in the ending of an introductory course; and (3) OO concepts should be left to a subsequent programming course (CS2).

An argument in favor of option (1) is that the OO contributes to learn design techniques of programs and modularization. Another aspect concerning (1) is that people that are used to the SI paradigm need to change the way they think (and teach). Thus, an argument against option (1) would be the unpreparedness of some lecturers due to lack of experience with OOP. Another disadvantage pointed out is that the concepts of SI programming such as loops and conditional statements receive little attention. These are the strengths of approaches (2) and (3), and the main argument in favor of the approach (3) is that each paradigm receives adequate attention. A reported disadvantage of approaches (2) and (3) is that once another paradigm has been learned, learning the "way of thinking" required by OO becomes more difficult.

[Bruce 2005] concludes that approach (1) should preferably be adopted but in association with pedagogical tools for OOP teaching in order to allowing the reduction of the complexity of mental abstractions as well as easing the learning of the OOP language syntax. If this is not possible, option (3) would be preferable. We agree that option (3) is preferable with respect to option (2), but we can not assess whether option (1) would be our choice if a pedagogical platform for OOP was available. Never a platform of this

³An acronym for Science, Technology, Engineering and Mathematics.

kind was available at the institutions where we worked⁴.

Several authors have contributed to this debate (such as [Ehlert and Schulte 2009, Reges 2006, Vilner et al. 2007]). A consensus, however, was never reached. We believe that hardly one of the alternatives will show to be superior to others in all scenarios, considering courses-students-lecturers-institution-learning resources.

Our proposal is well aligned with [Reges 2006] where the central idea is going "back to basics" in introductory programming courses. [Reges 2006] emphasizes problem solving (although he admits this is a fuzzy concept), procedural decomposition and the mastery of basic skills. The differential of our proposal is to emphasize (even more) on building concise code modules, letting the mechanisms of interaction with the user to the end of the course. We believe that in this way we are working more properly cognitive tasks related to the development of modular programs, without losing focus on the basics.

Although our proposal initiates the learning of programming with the definition and use of functions, it is distinguished from approaches that emphasize the learning of functional programming paradigm, such as the approach called "*Program by Design*" [Felleisen et al. 2001, Felleisen et al. 2004, Bieniusa et al. 2008, Sperber and Crestani 2012] and the previous approach of [Abelson and Sussman 1996], that had already been adopted in Brazil in the introductory programming course at PUC-Rio [Ierusalimschy 1997], and was later replaced by a more traditional approach [Celes and Ierusalimschy 2012]. From here on, every time we use the term *function* regarding programming we are referring to imperative functions or procedures, not the more mathematically pure functions of functional programming.

The use of functions from the beginning of the course as a unit of decomposition and abstraction is the only point in common between our proposal and functional approaches. Functional approaches exchange the use of the building blocks of imperative programming, such as variable assignment, dynamic data structure, sequence of commands, repetition structures and input and output commands, for pure functional programming, such as immutable data structures, recursion, high order functions and combining functions. Our proposal builds on the traditional teaching of SI programming, but increases emphasis on the use of functions and procedures as modularization tool, restricting the use of input and output commands to the main function of the program.

Regarding the programming language to be adopted in an introductory course, a consensus seems to exist. According to [Mannila and de Raadt 2006] one of the most important criteria for the choice of the language is the simplicity of the syntax and the structure of the language. The Python language appears as a choice that meets these criteria very well [Downey 2007], as well as offers the possibility of adopting both the SI and OO paradigms. Several authors have reported positive experiences concerning enhancing motivation and student satisfaction, reduction in failure and dropout rates, and increase of the grades with the adoption of Python instead of languages like Java or C / C ++ [Grandell et al. , Agarwal et al. 2008, Goldwasser and Letscher 2008]. The use of Python in CS1 and CS2 in UFRJ preceded the proposal we present in this article. Python language was adopted in 2007, and our observations confirmed the reports of the

⁴Main reasons for this are the lack of resources invested in acquiring and maintaining both hardware and software for undergrad labs

aforementioned authors.

3. Our Proposal

This proposal stems from an initiative for dealing with the systematic problems of CS1 and CS2 in UFRJ. The high dropout and failure rates, together with the visible frustration of lecturers and students, lead a group of 8 teachers from DCC/UFRJ to meet weekly during 2012 in order to study, exchange experiences, discuss and propose strategies for improving introductory programming courses offered to students of several STEM areas. A coordinator for this initiative was officially established in 2014, with a mandate to align the current practices with what has been discussed by the group, and to manage and organize the tasks for that end.

3.1. Scenario

DCC/UFRJ offers in average 25 CS1+CS2 classes in the first semester, and 24 in the second, a total of 49 classes per year. These are service classes: the students are not computer science students, but students of STEM disciplines in other departments. There is considerable rotation among the lecturers of these classes, with most of them being temporary adjuncts. There was a lot of dissatisfaction with these courses among students, course coordinators and lecturers. The main sources of dissatisfaction were: high dropout rates, poor grades, and insecurity by the part of coordinators that the courses were being taught properly, because there was great variation among both the subject matter and teaching styles.

The objective of our initiative was to standardize these courses, in order to (1) guarantee a minimum quality for the courses and (2) make them more efficient, by reducing individual labor and by evolving the course through consideration of previous experiences in both managing and teaching those courses. We bet on the effectiveness of this standardization to address the dissatisfaction that we have identified. To reach this objective, the following goals were set:

1. Review the syllabus of both courses (CS1 and CS2)
2. Setting strategies for teaching programming in courses with this profile
3. Dissemination of the new syllabus among teachers, students and coordinators
4. Uniformization of teaching plans and course material
5. Unification of exams and grading criteria
6. Supervision of how the courses are being taught

3.2. Course objective and syllabus

We did an exercise among the teachers to encourage them to think about what skills they expected the students to learn on an introductory programming course. The first version of this list of skills did not mention modularization, and we ended up noticing a lack of program design techniques, or essentially a guideline for leading the students in the constructions of algorithmic solutions to problems. After several iterations of refining this list of fundamental skills, we ended up with the following list:

- Identifying the relevant information in a problem and its respective representation and manipulation by the program.

- Comprehending the syntactical and semantic aspects of the programming language.
- Articulating commands, data structures and control structures for the construction of solutions for simple problems.
- Modeling and implementing modularized code for non-trivial problems.
- Constructing organized, reusable and legible code, following good programming practices.

After establishing this list of skills we reviewed the syllabus. Just the basic material relating to imperative programming has been kept in CS1. Accessory concepts like programming graphical interfaces have been moved to CS2.

3.3. Teaching strategies

As specified in the previous section, the biggest objective of the course is to develop skills for constructing correct, legible and organized programs. Once we have chosen modularization as the key program design technique that we were going to focus on, it was necessary to map this technique to specific programming language structures. For the CS1, our units of modularization are Python functions.

Understanding the concept of functions, as well as their construction and use, is vital to reaching the objective of the course. In the classical approach for teaching SI programming, this concept is often only taught almost in the end of the course, after all the commands and control structures of imperative programming, as well as simple imperative data structures, have already been taught. But the lecturers involved in reviewing our course have been unanimous in noting that, by only teaching functions near the end of the course, their learning is compromised because there is not enough time to properly practice the concept.

To deal with this problem, we had the idea of bringing the teaching of functions to the very beginning of the course. Besides solving the problem of lack of time for practicing the concept, we noticed other advantages of this strategy: the concept of function is the most concrete, in the universe of the students, to something that relates information (data) with the sequences of operations that transform this input information in the desired (output) information. The students already have prior exposure to mathematical functions, and can bring their mathematical intuitions to the study of functions in imperative programming. The connection with existing skills and knowledge facilitates learning and decreases sources of fear and frustration.

The students are familiarized from the start with writing simple functions and using the functions that they already wrote as building blocks in the construction of more complex functions. During the course, we gradually leave behind the intuition brought from mathematics, where functions are just an expression involving inputs and constants, to imperative functions implemented by blocks of sequential commands that can affect local and global variables. The students are encouraged to begin viewing their functions as a sequence of steps that are executed to solve some problem.

By working from the start with the identification and building of small cohesive functions, important abstract concepts like modularization, code organization, and code reuse are brought to practice and consideration of the students. Other advantages of beginning the course with functions include:

1	What is programming; Functions	7	Loop statement - while
2	Functions	8	Loop statement - for
3	Data types, strings. Decision Statement - if	9	Nested loops and matrices
4	Variables and assignment. Strings	10	Dictionaries
5	String manipulation, tuples, lists.	11	Basic input and output
6	Lists	12	More complex programs

Table 1. Material covered in each week of the course. A midterm exam happens between weeks 6 and 7, and final exams are after week 12.

- Notions of input and output of an algorithm are treated conceptually and not confused with console input and output (during most of the course the students interact directly with the functions they define via the Python REPL⁵, and only in the final part of the course they are taught standard console input and output).
- The concept of parameters and return values precede the concept of variables. We noticed that students that were first taught to use variables struggle with the use of parameters as input values as well as the use of the return value from a function.
- We can practice developing and calling functions incrementally, without having to write a complete program. Short functions are also easier to test, and to debug.

The material covered in each week of the course can be seen in table 1. After covering functions and their use, we move on to simple data types (numbers, booleans and basic use of strings), and only after that we begin teaching conditional statements. Considerable time is spent on techniques for using conditional statements and testing code that uses them. Our view is that boolean logic is something new and non-intuitive (and thus error-prone) for students, and it takes time to properly build the formal knowledge to use these concepts correctly. Only after the student has mastered conditional statements we introduce variables and assignment. Local variables are shown as a way to organize and name intermediate results, and global variables as a way of keeping state.

Basic use of Python lists and tuples is seen after variables and assignment. Tuples and lists are the most simple structured data types in Python, having syntactic support in the Python language for their creation⁶. As the students have not yet been taught how to build loops, the focus is on basic list operations such as indexing, slicing, copying, mostly as a way of teaching how transforming and organizing data is the basis of algorithms.

Loops are introduced in the second half of the course, when we expect that the student has already internalized that a programming language is a formal language, unambiguous and deterministic. At this point, the student should be able to analyze and debug his or her own code. In order to develop this skill, besides assigning several laboratory exercises where the students are expected to build increasingly more complex programs, we also feature several exercises where the students "bench test" their programs⁷. Traversing and manipulating lists and matrices (encoded as lists of lists) are the

⁵Read-Eval-Print-Loop, a special console where the programmer can evaluate Python expressions and commands.

⁶Vectors at the C language style do exist in Python, but are not the conventional structured data types in Python, and its creation and manipulation are not as straightforward.

⁷A bench test is a step by step simulation of the program, either using pen and paper or a tool such as Python Tutor [Guo 2013].

main motivators for building functions that use loops.

In the end of the course we cover basic input and output or “interacting with the user”, consisting in the construction of a main program that uses console input and output commands to gather information from the user that is supplied to the program’s functions, and to show the results to the user. We stress that “algorithmic” functions should never directly interact with the user, allowing the same functions to be reused with other methods of interaction. The interaction should only be done either in the main program or in functions specifically built for that.

4. Results and Conclusion

The main sources of dissatisfaction identified in our scenario were the high dropout rates, failure, and uncertainty about the consistency and quality of the courses. When thinking about these issues, we realize that they are in fact not local issues, but issues related to changes in the higher education scenario. The amount of classes, courses and teachers involved grew an order of magnitude or more, making inappropriate the treatment given in the past. To operationalize courses with these new orders of magnitude, strategies that do not take into account the continuity of the teaching-learning process beyond an isolated discipline are not appropriate. Class changes, reproofs and the preservation of continuity of learning for a next discipline are important demands. Achieving quality in this scenario is far from what a lecturer can do alone focusing only on the class he teaches, regardless of the effort to do the best. To tackle the problems, we had to think back the core of the courses: objectives, course content, teaching strategies and syllabus.

Our main contribution is the proposed teaching methodology: start with structured imperative programming, focused on modularity in CS1, and then focus on teaching OOP in CS2. In this paper, we underlined the methodology used in CS1, addressing the basic contents of SI programming in an unconventional order as summarized in table 1. To the best of our knowledge, a SI function-centered teaching approach for CS1 is not documented in the literature. The benefits of this new approach were summarized in section 3.3. As the lack of a design strategy was pointed as a big disadvantage of classical SI-based CS1 courses, it is worth mentioning that this approach provides an underlying program design strategy - the modularization.

Besides reformatting CS1 and CS2, systematizing the provision of these courses was also a useful strategy to ensure the quality and increase efficiency. When you have a high rotation of lecturers and a considerable number of temporary adjuncts, every individual evolution is easily lost. The establishment of a coordinator to keep track of the classes and maintain contact with the lecturers was very important. Tasks to achieve the desired standardization of the classes, like the development of the course material and its continuous improvement could be centralized by someone who had a global picture in mind. The course material consists of slide presentations for theoretical lectures and lists of exercises for practical lectures, and is freely available at <http://ladybugcodingschool.com/> for viewing and download.

We consider that, after four years of work, from the goals presented in section 3.2 were satisfactorily met the review and wide dissemination of the syllabus(1 and 3), the establishment of strategies for teaching programming (2), and the uniformization of teaching plans and course material (4). The unification of exams and grading criteria (5)

and the supervision of the courses (6) were partially achieved.

Wide dissemination of the syllabus, despite looking like a detail, brought several benefits. This goal was achieved with the creation of a website, and the dissemination of this, every semester, to everyone involved (teachers, students and coordinators). The syllabus works as a contract, in a high level, establishing what to expect from the course, leaving those involved more confident that what is expected will actually be delivered.

The unification of the exams and grading criteria occurred during one year, but went no further because it requires the involvement of many people in different roles (secretary, heads, lecturers, and even higher councils). Perhaps, with more time, it could be possible to get engagement by all persons involved, and thus make the unified exam a reality in our institution. We are now experiencing the first year in which the unification of the exam did not happen, and we realized that it begins to affect the work: as each lecturer now has complete autonomy on the evaluation of their own class, the commitment to stick to the contents of the syllabus was weakened.

Our perception is that the emphasis on the construction and use of functions from the beginning of an introductory course reduces the time and effort required to develop the skills to build programs, in addition to providing the basis for the further development of this skills in future stages. Even teachers who had resistance to adopt this new approach, commented they were satisfied with the outcome. Students who have failed before in the course and who experienced both approaches, have reported the preference for the new one. As future work, we intend to conduct a detailed analysis of the grades of all the classes where this approach has been implemented, and to compare data collected before and after this implementation. Our preliminary analysis has revealed an interesting phenomenon: the polarization of the grades was attenuated, suggesting that the learning curve was smoothed. Our next goal is to consolidate these assumptions from the historical data.

References

- Abelson, H. and Sussman, G. J. (1996). *Structure and Interpretation of Computer Programs*. MIT Press, Cambridge, MA, USA, 2nd edition.
- ACM and IEEE (2013). Computer science curricula 2013: Curriculum guidelines for undergraduate degree programs in computer science. Technical report, ACM Joint Task Force on Comp. Curricula ; IEEE Comp. Soc.
- Agarwal, K. K., Agarwal, A., and Celebi, M. E. (2008). Python puts a squeeze on java for cs0 and beyond. *J. Comput. Sci. Coll.*, 23(6):49–57.
- Bieniusa, A., Degen, M., Heidegger, P., Thiemann, P., Wehr, S., Gasbichler, M., Sperber, M., Crestani, M., Klaeren, H., and Knauel, E. (2008). Htdp and dmda in the battlefield: A case study in first-year programming instruction. In *Proc. of the 2008 Int. Workshop on Func. and Decl. Prog. in Education*, FDPE '08, pages 1–12, NY, USA. ACM.
- Bruce, K. B. (2005). Controversy on how to teach cs 1: A discussion on the sigcse-members mailing list. *SIGCSE Bull.*, 37(2):111–117.
- Celes, W. and Ierusalimschy, R. (2012). Apostila de programação i.

- Delgado, C., Xexéo, J. A. M., Souza, I. F., Campos, M., and Rapkiewicz, C. E. (2004). Uma abordagem pedagógica para a iniciação ao estudo de algoritmos. In *Anais do XII Workshop de Educação em Comp. (WEI)*.
- Downey, A. B. (2007). Python as a first language: Pre-conference workshop. *J. Comput. Sci. Coll.*, 22(6):3–4.
- Ehlert, A. and Schulte, C. (2009). Empirical comparison of objects-first and objects-later. In *Proc. of the Fifth Int. Workshop on Comp. Education Res.*, ICER '09, pages 15–26, NY, USA. ACM.
- Felleisen, M., Findler, R. B., Flatt, M., and Krishnamurthi, S. (2001). *How to Design Programs: An Introduction to Programming and Computing*. MIT Press, Cambridge, MA, USA.
- Felleisen, M., Findler, R. B., Flatt, M., and Krishnamurthi, S. (2004). The structure and interpretation of the computer science curriculum. *J. Funct. Program.*, 14(4):365–378.
- Goldwasser, M. H. and Letscher, D. (2008). Teaching an object-oriented cs1 -: With python. *SIGCSE Bull.*, 40(3):42–46.
- Grandell, L., Peltomäki, M., Back, R.-J., and Salakoski, T. Why complicate things?: Introducing programming in high school using python. In *Proc. of the 8th Australasian Conf. on Comp. Education - Vol. 52*.
- Guo, P. J. (2013). Online python tutor: Embeddable web-based program visualization for cs education. In *Proc. 44th ACM Technical Symposium on Comp. Science Education*, SIGCSE '13, pages 579–584, NY, USA. ACM.
- Ierusalimschy, R. (1997). Apostila de introdução à ciência da computação.
- Koulouri, T., Lauria, S., and Macredie, R. (2014). Teaching introductory programming: A quantitative evaluation of different approaches. *Trans. Comput. Educ.*, 14(4):1:28.
- Mannila, L. and de Raadt, M. (2006). An objective comparison of languages for teaching introductory programming. In *Proc. of the 6th Baltic Sea Conf. on Comp. Education Research*, Baltic Sea '06, pages 32–37, NY, USA. ACM.
- Mcgettrick, A., Boyle, R., Ibbett, R., Lloyd, J., Lovegrove, G., and Mander, K. (2005). Grand challenges in computing: Education—a summary. *Comput. J.*, 48(1):42–48.
- McKeown, J. and Farrell, T. (1999). Why we need to develop success in introductory programming courses. *The Journal of Computing in Small Colleges*, 14(3):241–250.
- Nikula, U., Gotel, O., and Kasurinen, J. (2011). A motivation guided holistic rehabilitation of the first programming course. *Trans. Comput. Educ.*, 11(4):24:1–24:38.
- Reges, S. (2006). Back to basics in cs1 and cs2. *SIGCSE Bull.*, 38(1):293–297.
- Robins, A., Rountree, J., and Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Comp. Sci. Education*, 13(2):137–172.
- Sperber, M. and Crestani, M. (2012). Form over function: Teaching beginners how to construct programs. In *Proc. of the 2012 Annual Workshop on Scheme and Functional Programming*, Scheme '12, pages 81–89, NY, USA. ACM.
- Vilner, T., Zur, E., and Gal-Ezer, J. (2007). Fundamental concepts of cs1: Procedural vs. object oriented paradigm - a case study. *SIGCSE Bull.*, 39(3):171–175.

Torneios Baseados em Robocode para Incentivar Jovens a Aprender Programação

Matheus C. Meira¹, Marri S. S. Lima¹, Marcos A. F. Borges¹

¹Faculdade de Tecnologia – Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 456 – 13.484-332 – Limeira – SP – Brasil

m108479@dac.unicamp.br, {m118075, marcosborges}@ft.unicamp.br

***Abstract.** Students from the current generation, inserted in a context of connectivity in the digital world may consider demotivating the traditional learning environment. This article points a new approach to teaching programming through the development of scientific competitions based on the digital educational game Robocode. The study presents the experience of organizing scientific competitions involving contests and leagues based on the educational game Robocode over five years and it discuss how these can be used as promising tools to diversify playful and collaborative learning methods.*

***Resumo.** Alunos das gerações atuais, inseridos em um contexto de conectividade no universo digital, podem considerar ambientes de ensino tradicionais desmotivadores. Este artigo indica uma nova abordagem para ensino de programação, a partir do desenvolvimento de competições científicas baseadas no jogo digital educacional Robocode. O artigo apresenta a experiência de organização de competições científicas que envolvem torneios e ligas baseadas no jogo educacional Robocode ao longo de cinco anos e discute como eles podem ser usados como promissoras ferramentas de diversificação de métodos lúdicos e colaborativos de aprendizagem.*

1. Introdução

Atualmente, instituições de ensino desenvolvem ações com jogos educacionais digitais nos processos de ensino-aprendizagem para incentivo ao aprendizado de programação (Barnes, Richter, Powell, Chaffin, e Godwin, 2007; Chaffin, Doran, Hicks e Barnes, 2009). Conteúdos de lógica de programação ou programação em si não caracterizam apenas assuntos exclusivos aos cursos específicos das áreas relacionadas a informática ou engenharias. Estes assuntos tornaram-se tópicos essenciais para melhor entendimento de aplicativos de uso geral, ou mesmo como ferramentas de auxílio no ensino de variadas abordagens curriculares. A programação ganha o interesse dos jovens, envoltos em tecnologias cada vez mais presentes em suas vidas. A programação consegue despertar a criatividade das pessoas que a utilizam (Valente, 1999; Garcia, Correia e Shimabukuro, 2008). Alunos participantes de disciplinas de programação ou jovens admiradores de tecnologias, muitas vezes têm o anseio de entender a lógica aplicada nos aplicativos atuais.

A organização sem fins lucrativos Code.org¹ é dedicada a expandir o ensino de ciência da computação, incentivando o acesso ao conhecimento de programação entre os jovens. Conhecido no Brasil como “hora do código”, Code.org se enquadra nas ações que têm por objetivo permitir e incentivar a compreensão de conceitos de programação de modo lúdico. O presente artigo aborda ações para incentivar jovens programadores com a diversificação de métodos de aprendizagem com auxílio de torneios científicos.

Enquanto o “método tradicional de ensino é centrado no professor, que possui o papel de sujeito ativo, sendo o aluno passivo no processo de aprendizagem” (MEZZARI, 2011, p. 115), algumas iniciativas procuram disseminar a programação com propostas de ensino que diferem aos meios tradicionais.

Estudos empíricos apontados por Johnson (2010) e Adams (2012) demonstram que os alunos podem alcançar ganhos significativos de aprendizagem ao interagir com jogos educacionais em ambientes controlados. “Jogos educacionais digitais vão muito além do entretenimento e funcionam como eficientes ferramentas de aprendizagem” (GALLANT e MAHMOUD, 2008, p. 119). “Os jogos educacionais têm sido frequentemente sugeridos como promissoras ferramentas pedagógicas” (PRENSKY, 2012, p. 22).

Este artigo busca discutir abordagens interessantes para incentivar a aprendizagem de programação entre jovens. Na tentativa de responder a esse questionamento são levantadas hipóteses no sentido de adoção de jogos educacionais digitais que possam despertar o interesse em um ambiente que proporcione competições científicas. A proposta apresentada neste trabalho explora a adoção do jogo educacional digital denominado Robocode (2013) associado ao Aprendizado Baseado em Problemas (*Problem Based Learning* - PBL) no desenvolvimento de competições científicas. PBL é uma forma de aprender com a utilização de cenários que envolvem problemas da vida real. “É um método que desafia os alunos a aprender e trabalhar em grupos na busca de soluções para os problemas reais” (HOU, 2014).

O jogo Robocode permite o ensino de linguagem de programação Java ou C# (plataforma .NET). A partir do jogo Robocode é possível organizar competições científicas que possam favorecer acesso ao estudo de disciplinas relacionadas à lógica de programação e programação de computadores de modo lúdico. O objetivo da presente abordagem está centrado na apresentação do jogo educacional Robocode em competições para estímulo de aprendizagem de linguagem de programação. As seções seguintes apresentam o Robocode e estratégias para explorar esta plataforma em competições planejadas para um ambiente de ensino-aprendizagem de programação.

2. Framework de Simulação “Robocode”

O projeto Robocode é um software *open source*, com fontes abertas e sob os termos da *Eclipse Public License*² (EPL), que possibilita simulações de batalhas com robôs virtuais. Foi projetado para simular uma arena de batalha com robôs, cada qual

¹ Code.org. (2016) *Non-profit Dedicated to Expanding Access to Computer Science*. Disponível em: <<https://code.org/>>.

² EPL. (2016) *Eclipse Public License version 1.0, definitions homepage*. Disponível em: <<http://www.eclipse.org/legal/epl-v10.html>>.

com seus próprios algoritmos de ataque e defesa. O software pode ser executado em plataformas com suporte às versões recentes do Java Runtime (versão 2 ou superior) (Robocode, 2013).

Robocode foi desenvolvido por uma comunidade da IBM® denominada *AlphaWorks*, ativa desde 1996, com propostas de desenvolvimento de tecnologias emergentes. Robocode ganhou notoriedade nos laboratórios de pesquisa da IBM®, porém com incertezas de sua continuidade, o projeto foi levado em 2005 para o *SourceForge* (consiste em um repositório de código fonte que atua como centro para desenvolvedores gerenciarem projetos livres). Atualmente, o projeto Robocode é mantido por iniciativas de software livre em torno da comunidade Robocode.

Robocode (2013) é um jogo de programação com objetivo de codificar batalhas para competir em uma arena com outros robôs. O jogador é o programador do robô virtual, cuja função é desenvolver a lógica com informações de comportamento e reação a eventos que ocorrem na arena. O software é composto por um ambiente completo de desenvolvimento: instalador próprio, editor de robôs e compilador Java/C#.

O Robocode possui uma IDE (*Integrated Development Environment*) própria e uma interface formada por componentes GUI (*Graphical User Interface*), também conhecido por *widgets* (controles para simplificar o acesso a um outro programa ou sistema). A Figura 1 indica os componentes do sistema Robocode, onde é possível visualizar as janelas de campo de disputa (*Battlefield*) e editor do robô (*Robot Editor*) que representam a forma do IDE (Li, 2002).

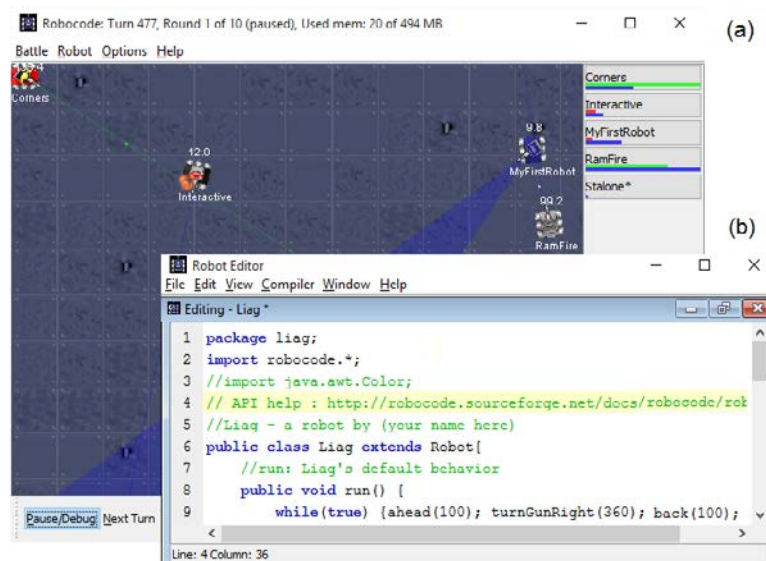


Figura 1 - Robocode (a) Campo de Batalha (b) Editor de Robô (Robocode, 2015).

Pode-se classificar Robocode como um jogo sério (*Serious Games* - SGs), por ser um jogo que proporciona um ambiente em que o agente de aprendizagem pode interagir e aprender. De acordo com Nidorf e Barone (2010), para este cenário, o agente tem a possibilidade de escolher uma variedade de comportamentos e medir a eficácia de cada um ao longo do desafio.

Com base em Recchia et al. (2014, p. 90) pode-se classificar o Robocode como um *framework* de simulação. A classificação do Robocode está centrada em um simulador de batalhas com base em Inteligência Artificial (IA). No simulador de

batalhas, é possível observar o comportamento dos robôs. Pelo fato das batalhas ocorrerem de modo *on-line*, na plataforma, é possível comparar a programação de robôs efetuada por outros desenvolvedores. No simulador, os robôs são entidades autônomas e os programadores não têm controle das ações, o que contribui para programação de agentes inteligentes. Um robô não possui total conhecimento do meio ambiente de batalha e suas informações estão restritas aos sensores do qual ele pode ler (ex. posição do robô, ângulo do canhão e/ou radar etc.) (Hong e Cho, 2004; Fayek e Farag, 2014).

3. Ambiente de Competição com Robocode

Um gerenciador da liga mundial Robocode denominado *RobotLeague* foi criado por Christian Schnell, em 2002. O gerenciador tinha funções de garantir o agrupamento dos participantes; conduzir partidas; gerenciar os resultados e produzir relatórios de status em páginas compartilhadas na internet. *RobotLeague* foi descontinuada um ano após sua criação, porém ganhou notoriedade por representar a primeira iniciativa de liga mundial do Robocode (Li, 2002). Atualmente, destaca-se a competição mundial chamada *Hat League*, mantida pela comunidade Robowiki³ associada a iniciativa colaborativa denominada *RoboRumble* para exibir as partidas e construir rankings.

Existem vários torneios Robocode no mundo, em países como: Tailândia, Bélgica, Austrália e Nova Zelândia. Dentre disputas ao redor do mundo, é possível citar como destaque o torneio da Irlanda, *GamesFleadh*⁴. O torneio da Irlanda foi iniciado pela ICS (*Irish Computer Society*), em 2006, sendo denominado “*Robocode Challenge Trophy*” e contando com a participação de estudantes de todo país. Cursos de informática (em nível médio e superior) e afins criam grupos de pesquisa e treinam suas equipes formadas por alunos dos primeiros anos das disciplinas de programação. O torneio da Irlanda proporciona uma emocionante competição com envolvimento das instituições acadêmicas do país. O torneio conta com importantes incentivadores dos eventos ao longo dos anos: Microsoft (atual); EA–*Electronic Arts* (atual); Dell; Pearson *Education*; Lenovo, etc. Originalmente, essa competição era chamada de “*Robocode Challenge Trophy*”, mas atualmente é denominado “*GamesFleadh*” (festival dos jogos na tradução da língua Irlandesa), organização mantenedora e coordenadora das disputas.

No Brasil, instituições como a ETEC (Escola Técnica Estadual - Centro Paula Souza) possuem torneios descentralizados e estudos para organização de liga em suas unidades espalhadas no estado de São Paulo. Institutos Federais possuem históricos de realizações de disputas de programação com Robocode. Várias universidades (públicas e privadas) espalhadas no Brasil já tiveram iniciativas com o torneio dentro das disciplinas de programação ou em semanas de ciência e tecnologia. A partir de pesquisas de informações registradas em domínios oficiais institucionais, a Tabela 1 exibe algumas instituições brasileiras que realizaram torneios Robocode ao longo dos últimos anos. São apresentadas apenas entidades com mais de um ano de incidência; aproximadamente 40 outras foram identificadas com apenas um ano de torneio. A classificação segue em ordem decrescente de incidências. A pesquisa foi realizada no

³ Robowiki. (2016) *Collecting Robocode Knowledge Since 2003*. Disponível em: <<http://robowiki.net>>. Código fonte aberto, com a promoção de competições ao redor do mundo.

⁴ GamesFleadh. (2016) *Digital Games Programming Festival - Robocode*. Disponível em: <<http://gamesfleadh.ie/robocode>>.

final do ano de 2015 no site de buscas do Google (2015) em domínios brasileiros com as seguintes expressões: “Torneio Robocode”; “Liga Robocode”; “Campeonato Robocode” e “Competição Robocode”.

Números apurados no presente artigo, indicam que competições disputadas entre algumas das primeiras instituições listadas na Tabela 1, Colégio Técnico de Limeira da Universidade Estadual de Campinas (COTIL-UNICAMP), Faculdade de Tecnologia da Universidade Estadual de Campinas (FT-UNICAMP) e Instituto Federal de Educação Ciência e Tecnologia de São Paulo (IFSP), caracterizam a maior competição Robocode oficial já realizada no Brasil. Esta competição ultrapassou a marca do Instituto Federal de Alagoas (IFAL) em número de participantes (60) equipes (16) e instituições envolvidas (1). Estas instituições deram origem a competição anual brasileira do Robocode organizada pelo laboratório de Informática Aprendizagem e Gestão (LIAG) da UNICAMP.

Tabela 1 - Ocorrências Torneio Robocode Brasil.

Instituição	Cidade	Ano(s) do Torneio(s)
COTIL-UNICAMP - Colégio Técnico de Limeira	Limeira-SP	2015, 2014, 2013, 2012, 2011, 2010
FT-UNICAMP - Faculdade de Tecnologia	Limeira-SP	2015, 2014, 2013, 2012, 2011, 2010
ETEC Centro Paula Souza - João Belarmino	Amparo-SP	2014, 2013, 2012, 2011
IFSP - Instituto Federal São Paulo	Capivari-SP	2015, 2014, 2011
IFAL - Instituto Federal de Alagoas	Arapiraca-AL	2015, 2014
IF Farroupilha	São Borja-RS	2015, 2014
UFU - Universidade Federal de Uberlândia	Patos de Minas-MG	2013, 2012
UFMA - Universidade Federal do Maranhão	São Luís-MA	2014, 2009
UFC - Universidade Federal do Ceará	Quixadá-CE	2011, 2007

3.1. Competições Robocode na UNICAMP e IFSP

O torneio Robocode do laboratório LIAG da UNICAMP ocorre desde de 2010. Os torneios Robocode LIAG são regidos por regras de modo estimular a competitividade, administrar as rodadas e garantir a organização. Algumas das principais regras são: (1) Equipes de no máximo quatro participantes; (2) Compostos por uma fase classificatória de pontos corridos e outra com enfrentamento direto em quadrangulares (quartas de final, semifinal e final); (3) Ao final de cada rodada, os códigos fontes dos robôs são disponibilizados e divulgados pela organização do evento; (4) Mudanças nos códigos devem ser explicadas para atender boas práticas de programação; (5) Grupos com códigos copiados são banidos da competição.

A partir de 2012, o LIAG passou a organizar Ligas. O conceito da liga é promover torneios descentralizados seguindo o mesmo padrão de regras e organizar uma final da liga entre os melhores colocados dos torneios.

No início do segundo semestre de 2015, o LIAG promoveu uma liga entre três instituições (FT-UNICAMP, COTIL-UNICAMP e IFSP) regidos sob as regras da liga. O LIAG promoveu os torneios na FT-UNICAMP e COTIL-UNICAMP com a condução e supervisão de integrantes e colaboradores do LIAG, sendo o primeiro aberto, contando com a participação de equipes da FT-UNICAMP e instituições externas e o segundo fechado apenas para alunos do COTIL-UNICAMP. Professores e alunos bolsistas de projetos relacionados ao Robocode do IFSP administraram um torneio próprio. Todas

informações das disputas (notícias dos torneios e liga, instituições participantes, resultados, ranking, disputas, premiações, cursos, fotos, vídeos, etc.) são centralizados e disponibilizados na internet em um canal oficial do evento, RobocodeBrasil⁵. Foram criados calendários distintos para cada um dos torneios locais onde cada torneio intermediário registra um campeão com direito a disputar a fase final da liga.

No final do segundo semestre de 2015, o LIAG recebeu as equipes campeãs dos torneios para organização e realização da liga entre os finalistas Robocode. Contou com a participação de equipes vencedoras do torneio do Instituto Federal Farroupilha (IF Farroupilha) do campus de São Borja-RS, disputado junto ao torneio da FT-UNICAMP. No ano de 2015, os eventos contabilizaram a participação de aproximadamente 30 equipes e mais de 100 competidores, em três torneios intermediários concomitantes e uma liga interinstitucional.

A liga dos campeões é composta por desafios diretos em nível de semifinal e final com os campeões de cada torneio intermediário. Os segundos melhores colocados de cada instituição tiveram a oportunidade de participar de uma “repescagem” e entrar no quadrangular final em busca do título de campeão.

Os torneios e a liga podem ser patrocinados e oferecer incentivo em forma de premiações. A liga LIAG 2015 contou com incentivos: (1) Destinados aos campeões dos torneios intermediários; (2) Para os três primeiros colocados da liga e; (3) Melhor código dentre todas equipes participantes. O LIAG promoveu a liga e incentivou as equipes intermediárias oferecendo cursos especializados para as mais bem colocadas.

4. Discussão

Foram respondidos 22 questionários individualmente por alunos participantes dos torneios e ligas Robocode nos anos de 2014 e 2015. O grupo foi composto por estudantes de cursos técnicos e graduação com idades que variaram de dezesseis a vinte e cinco anos e participação aproximada de dois terços do gênero masculino. A Figura 2 (a) exibe a porcentagem dos alunos regularmente matriculados que responderam questões referente as competições científicas nas instituições: FT-UNICAMP; COTIL-UNICAMP e IFSP, dentre os cursos: Técnico em Informática (médio); Técnico Integrado em Informática (médio integrado); Análise e Desenvolvimento de Sistemas (superior); Construção de Edifícios (superior) e Engenharia de Telecomunicações (superior). Algumas informações foram solicitadas aos alunos acerca do conhecimento ou entendimento da linguagem de programação Java, observado na Figura 2 (b). Os resultados apurados na Figura 2 puderam apresentar um público de alunos predominantemente da área de exatas e em sua maioria com conhecimento superficial ou apenas noção da existência da linguagem de programação Java. Alunos que indicaram o conhecimento prévio foram, em maioria, aqueles concluintes de disciplinas relacionadas a linguagem de programação.

As questões puderam envolver os alunos com objetivo de apresentar características das quais torneios científicos podem auxiliar na aprendizagem da linguagem de programação (Figura 3). Como resultado foi possível observar que cerca

⁵ RobocodeBrasil (2016) canal oficial competição. Disponível em <<http://www.robocodebrasil.com.br>>. Página acadêmica Robocode do LIAG, disponível em <<http://www.ft.unicamp.br/liag/robocode>>.

de metade das respostas focam: aprendizagem de maneira lúdica; torneio Robocode desperta a motivação e desafia os alunos a apresentar propostas para desenvolvimento dos robôs; a formação equipes colaborativas no desenvolvimento dos códigos e estratégias.

A pergunta final do questionário indica quais melhorias e iniciativas o evento Robocode deve oferecer para conquistar participação de novos alunos. Com esse foco em conquistar novos adeptos à competição Robocode, as principais sugestões foram apresentadas na Figura 3. Conforme respostas dos questionamentos, um fator considerado para o aumento do interesse por esse tipo de torneio é a mobilização de professores e alunos para disponibilizar minicursos ou tutorias para auxílio no desenvolvimento dos códigos entre as equipes. Outro fator está relacionado ao investimento e divulgação das instituições na realização das competições que envolvem jogos digitais e diversificação de métodos de ensino.

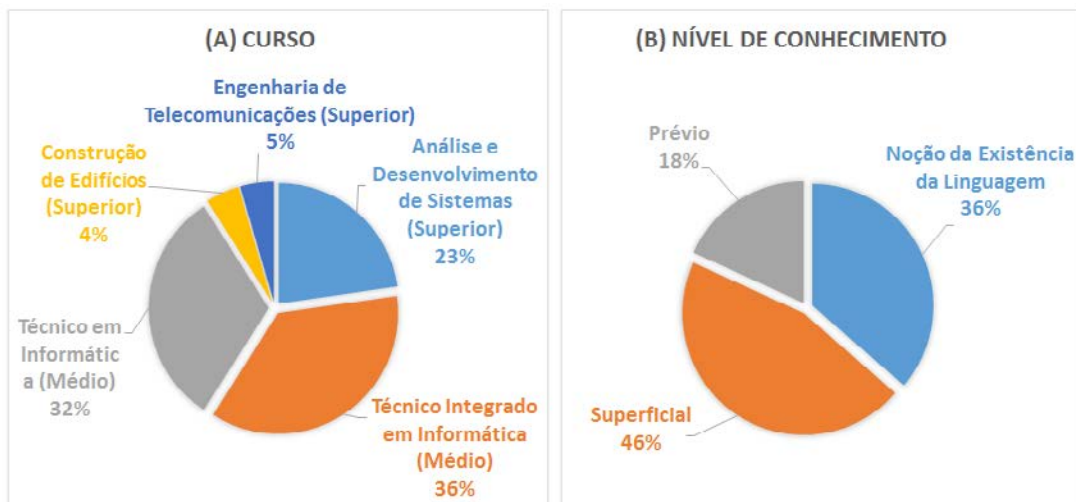


Figura 2 - (a) Participantes de Alunos por Curso (b) Nível de Conhecimento.

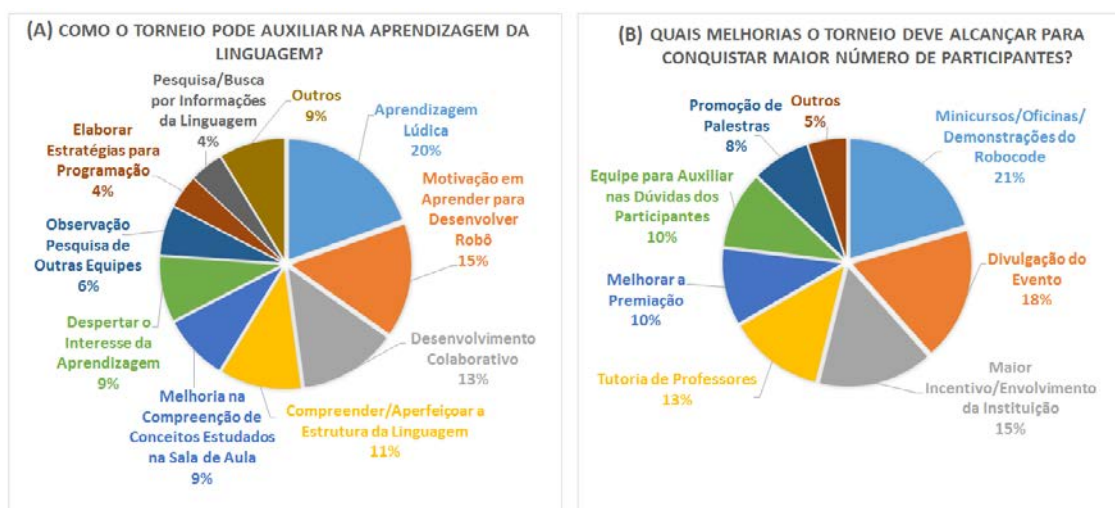


Figura 3 - (a) Torneio na Aprendizagem (b) Conquista de Participantes.

A realização dos torneios em caráter interinstitucional valorizou aspectos de colaboração entre instituições, pôde exibir que ao ser incentivado, estes eventos podem

resultar na diferenciação de metodologias de aprendizagem. Esse fato ficou evidente na equipe campeã do torneio de 2015, IF Farroupilha, que após a conquista do título gravou um vídeo e publicou nota em agradecimento e incentivo as iniciativas Robocode. Esse resultado obteve repercussão positiva interinstitucional apresentando motivações de novos alunos e instituições a conhecer os torneios Robocode.

Na Figura 4 (a), observa-se a aplicação de conceitos presentes nos planos de ensino da disciplina de linguagem de programação, plano de aula de conteúdo programático referente aos “comandos de decisões” *if-else* (se-senão). A Figura 4 (a) exibe o código de uma equipe do curso técnico em informática com noções ou conhecimento superficial da linguagem de programação Java. Essa equipe propôs “variáveis booleanas” ($\text{ângulo} == \text{true}$) com instruções para “tomada de decisão” (*if-else*) quanto à movimentação do robô. O exemplo indica *if* (se) o ângulo é verdadeiro, o robô deve avançar o valor de 100 e virar à direita em 90 graus, *else* (senão) ângulo é falso e o robô deve recuar o valor de 50 e virar para esquerda em 90 graus.

<pre> 5 public void run() { 6 boolean angulo = true; 7 while(true) { 8 if (angulo == true){ 9 ahead(100); turnRight(90); 10 angulo = false;} 11 else{ 12 back(50); turnLeft(90); 13 angulo = true;} 14 } } </pre>	(A)	<pre> 16 public void run() { 17 while(true) { 18 for (int virarDir=0;virarDir<4;virarDir++){ 19 ahead(100); 20 turnRight(90);} 21 for (int virarEsq=4;virarEsq>0;virarEsq--){ 22 ahead(100); 23 turnLeft(90);} 24 } 25 } </pre>	(B)
--	------------	---	------------

Figura 4 – (a) Comandos de Decisão (b) Estruturas de Repetição

Os códigos exibidos na Figura 4 (a) e (b) foram desenvolvidos pela mesma equipe em momentos distintos dentro da competição. Código (b) foi criado em rodadas posteriores ao (a). A Figura 4 (b) indica o uso do conceito de “estrutura de repetição” com comando *for* (para), apresentado na linha 18. Na mesma linha, observa-se o uso do comando *for* com sua “variável de inicialização” (int), as “condicionais” recebendo valores (inicial: $\text{virarDir}=0$; final: $\text{virarDir}<4$) e ao final seu “incremento” ($\text{virarDir}++$). Para entender o conceito a equipe pesquisou informações dos conceitos envolvidos acerca da “variável de inicialização” (inicia um controle de repetição), as “condições” (determina a condição inicial e final da repetição) e “incremento” (variável de controle alterada a cada repetição). Após o entendimento dos conceitos, a equipe pôde aferir o resultado em comandos de movimentação do robô: avançar no valor de 100 e virar à direita em ângulo de 90 graus por 4 vezes consecutivas; em seguida repetir o mesmo processo à esquerda.

5. Conclusão

Alunos das gerações atuais, inseridos em um contexto de conectividade no universo digital, estão sujeitos a encontrar ambientes de ensino tradicionais desmotivadores. Estes ambientes podem desestimular alunos acostumados a interagir com tecnologias digitais. Este trabalho apresentou experiências com a organização de competições científicas baseadas em Robocode. Foi apresentada uma experiência de condução de competições desde 2010.

As competições científicas que envolvem torneios e ligas do jogo educacional Robocode têm funcionado como promissoras ferramentas de diversificação de métodos de aprendizagem nas instituições que a realizam. Além de introduzir métodos diferenciados, as competições estimulam o aprendizado lúdico e incorporam aspectos do aprendizado colaborativo.

Com base nas experiências apresentadas, os torneios Robocode podem motivar participantes a buscar informações para estudo e desenvolvimento dos conceitos da linguagem de programação em ambientes coletivos. Nestas competições, os problemas apresentados são dinâmicos e geram estímulos para que as equipes tenham condições de trabalhar no desenvolvimento das soluções de acordo com situações visualizadas durante a execução das partidas.

As regras do torneio indicam que a cada nova rodada os códigos devem ser abertos. Essa característica garante as equipes com noções ou conhecimentos superficiais tenham a possibilidade de analisar códigos elaborados por outras equipes de alunos com prévio entendimento da linguagem. A partir de análises mútuas do desenvolvimento de códigos, equipes buscam entender e estudar novos conceitos com objetivo de aperfeiçoar as características dos robôs. A proposta com ciclo de abertura dos códigos fonte dos robôs durante o torneio favorece o nivelamento para estudos de conceitos como “comandos de decisões” ou “estruturas de repetição” discutidos no presente artigo. Existem ainda as discussões de estratégias de jogo e compartilhamento de experiência entre equipes ao longo da execução do torneio.

Para vencer o torneio, as equipes devem trabalhar o desenvolvimento contínuo de estratégias para suportar as diversas situações exigidas no jogo, como a melhoria da movimentação dos robôs. Em estágios avançados do torneio, equipes com movimentação na base do conceito de “comando de decisão” sentiram a necessidade de explorar novos conceitos ainda não estudados, como as “estruturas de repetição”. Observou-se a evolução das equipes durante o torneio, com o estudo de conceitos de linguagem de programação presentes nos planos de ensino dos cursos muitas vezes ainda não tratados nas disciplinas tradicionais. Torneios com envolvimento de jogos digitais, como Robocode, podem auxiliar na diversificação da aprendizagem. Os ambientes dos torneios geram situações-problemas que podem incentivar alunos a buscar soluções, na aprendizagem de novos conceitos, como aperfeiçoamento de estratégias de movimentação dos robôs. Equipes formadas por alunos de diferentes níveis de conhecimento, tiveram a oportunidade de estabelecer um nivelamento de conceitos da linguagem. Resultados positivos, enfatizadas pelas instituições e entrevistados, indicaram o jogo Robocode associado aos torneios podem auxiliar no estímulo do processo de aprendizagem de linguagem de programação de modo lúdico.

7. Referências Bibliográficas

- Adams, D. M.; Mayer, R. E.; Koenig, A. W. (2012) “Narrative Games for Learning: Testing the discovery and narrative hypotheses”. *Journal of Educational Psychology*, 104 (1), pp. 235-249, doi: 10.1037/a0025595.
- Barnes, T.; Richter, H.; Powell, E.; Chaffin, A.; Godwin, A. (2007) “Game2Learn: Building CS1 Learning Games for Retention”. *SIGCSE Bull.* 39, 3 (June 2007), 121-125, doi=<http://dx.doi.org/10.1145/1269900.1268821>.

- Chaffin, A.; Doran, K.; Hicks, D.; Barnes, T. (2009) “Experimental Evaluation of Teaching Recursion in a video game”. In Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games (Sandbox '09), Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 79-86, doi=<http://dx.doi.org/10.1145/1581073.1581086>.
- Fayek, M. B.; Farag, O. S. Hicma. (2014) “A Human Imitating Cognitive Modeling Agent Using Statistical Methods and Evolutionary Computation”. In Computational Intelligence for Human-like Intelligence (CIHLI), 2014 IEEE Symposium on, vol., no., pp.1-8, 9-12, doi: 10.1109/CIHLI.2014.7013383, Dec.
- Gallant, R. J.; Mahmoud, (2008) “Q. H. Using Greenfoot and a Moon Scenario to teach Java programming in CS1”. Proceedings of the 46th Annual Southeast Regional Conference on XX - ACM-SE 46 (p. 118). New York, New York, USA: ACM Press. doi: 10.1145/1593105.1593135.
- Garcia. R. E.; Correia, R. C. M.; Shimabukuro. M. H. (2008) Ensino de Lógica de Programação e Estruturas de Dados para Alunos do Ensino Médio. Anais XXVIII Congresso SBC. WEI–Workshop sobre Educação em Computação, Belém-PA, julho.
- Hong, J. H.; Cho, S. B. (2004) “Evolution of Emergent behaviors for shooting game characters in Robocode”. In Evolutionary Computation, 2004. CEC2004. Congress on, vol.1, no., pp.634-638 Vol.1, 19-23, doi: 10.1109/CEC.2004.1330917, June.
- Hou, S. I. (2014) “Integrating Problem-based Learning With Community-Engaged Learning in Teaching Program Development and Implementation”. Universal Journal of Educational Research 2(1): 1-9, doi: 10.13189/ujer.2014.020101.
- Johnson, W. L. (2010) “Serious Use of a Serious Game for Language Learning”. International Journal of Artificial Intelligence in Education, 20 (2), pp. 175-195, doi: 10.3233/JAI-2010-0006.
- Li, S. (2002) “Rock 'em, sock 'em Robocode!” Learning Java programming is more fun than ever with this advanced robot battle simulation engine. IBM developerWorks, <http://www.ibm.com/developerworks/java/library/j-robocode/>.
- Mezzari, A. (2011) O Uso da Aprendizagem Baseada em Problemas (ABP) como reforço ao ensino presencial utilizando o ambiente de aprendizagem Moodle. Revista Brasileira de Educação Médica. Rio de Janeiro. v. 35, n. 1, p. 114-121, jan./mar.
- Nidorf, D.G.; Barone, L.; French, T. (2010) “A comparative study of NEAT and XCS in Robocode”. In Evolutionary Computation (CEC), 2010 IEEE Congress on , vol., no., pp.1-8, 18-23, doi: 10.1109/CEC.2010.5586087, July.
- Prensky, M. (2012) Aprendizagem Baseada em Jogos. Tradução de Eric Yamagute; revisão técnica de Romero Tori e Denio di Lascio. Editora Senac, São Paulo.
- Recchia, T., Chung, J., and Pochiraju, K. (2014). “Performance of Heterogeneous Robot Teams With Personality Adjusted Learning”. Biologically Inspired Cognitive Architectures, vol. 7, pp. 87–97, doi=<http://dx.doi.org/10.1016/j.bica.2013.10.003>.
- Robocode. (2013) “Open Source Educational Game, ReadMe for Robocode”. Disponível em: <<http://robocode.sourceforge.net>>. Acesso em: agosto de 2015.
- Valente, J. A. (1999) O Computador na Sociedade do Conhecimento. São Paulo: UNICAMP/NIED.

Uma Análise da Evasão Escolar nos Cursos de Tecnologia da Informação: Um estudo de caso em Floresta/PE

Everton Eugênio de Oliveira Silva^{1,2}, Josilaine de Souza Silva^{1,2}, Cassiano Henrique de Albuquerque^{1,2}

¹Grupo de Pesquisa em Inovações Tecnológicas na Gestão de Hardware, Software e Peopleware – GITEG

²Instituto Federal de Educação Ciência e Tecnologia do Sertão Pernambucano
Caixa Postal 56400-000 – Caetano II – Floresta – PE – Brasil

evertonenew@gmail.com, josylayne.silva@gmail.com,
cassiano.henrique@ifsertao-pe.edu.br

Abstract: School dropout in information technology courses, directly affects economy and cultural development at any society. Interruption in the schooling process affects economy because it results in shortage of skilled labor, which hinders regional economic development. In order to address this issue, the Brazilian government has outlined goals to reduce this phenomenon, which is even more pronounced at Sertão Nordestino. This paper presents a qualitative and quantitative research on school dropout rates in information technology courses at Instituto Federal do Sertão Pernambucano no Campus Floresta, indicating possible solutions to this issue.

Resumo: A evasão escolar nos cursos de tecnologia da informação, afeta diretamente a economia e o desenvolvimento cultural de qualquer sociedade. A interrupção no ciclo escolar afeta a economia, pois resulta em escassez de mão-de-obra qualificada, o que dificulta o desenvolvimento econômico regional. Para tratar deste problema, o governo brasileiro tem traçado metas para atenuar este fenômeno, que é ainda mais acentuado no Sertão Nordestino. Este artigo expõe de forma qualitativa e quantitativa os índices de evasão escolar nos cursos de tecnologia da informação do Instituto Federal do Sertão Pernambucano no Campus Floresta, apontando possíveis soluções atenuadoras para esta problemática tão relevante.

1. Introdução

A evasão é um fenômeno social complexo, definido como interrupção no ciclo de estudos [GAIOSO, 2005]. A evasão escolar é um tema muito pertinente ao ambiente acadêmico, em vários países este tema requer maior atenção por ter um índice muito alto de evasão escolar. A evasão estudantil afeta os resultados dos sistemas educacionais e a economia pública e privada, pois resulta em escassez de mão-de-obra qualificada

impedindo assim o desenvolvimento econômico de forma mais efetiva. Outro problema bastante relevante que a evasão escolar traz é a ociosidade de professores, espaço físico e equipamentos.

O Relatório de Auditoria do Tribunal de Contas da União resultante do Processo nº 026.062/2011-9 [TCU, 2013], realizado com vistas a avaliar as ações de estruturação e expansão da Rede Federal de Educação Profissional, Científica e Tecnológica, encontrou uma taxa de evasão nos cursos Médios Subsequentes de 19%, utilizando os dados do Sistema Nacional de Informações da Educação Profissional e Tecnológica (SISTEC), para os ciclos de matrícula iniciados a partir de 2004 e encerrados até o primeiro semestre de 2011, nos Institutos Federais do Espírito Santo, Rio Grande do Sul, Rio de Janeiro, São Paulo, Minas Gerais, Pernambuco, Sertão Pernambucano, Goiás e Rio Grande do Norte. No Termo de Acordo de Metas e Compromissos que a Secretaria de Educação Profissional e Tecnológica (SETEC/MEC) celebrou com cada um dos Institutos Federais foi definido um Índice de Eficácia da Instituição, calculado pela média aritmética da eficácia de cada turma, medida pela relação entre o número de alunos concluintes e o número de vagas ofertadas no processo seletivo para cada uma dessas turmas, sendo a meta mínima de 80% de eficácia da Instituição no ano de 2016, com meta intermediária de, no mínimo, 70% no ano de 2013. Uma taxa elevada para esse índice significa que os investimentos realizados pelos Institutos Federais, em infraestruturas, professores e técnicos, estão gerando produtos (alunos concluindo cursos) em número satisfatório.

2. O Problema da Evasão no Ensino Superior

A evasão estudantil no ensino superior é um problema internacional que afeta o resultado dos sistemas educacionais. [SILVA FILHO, 2007]. Além de afetar os resultados dos sistemas educacionais, um índice de evasão escolar acima do esperado afetará também o desenvolvimento social.

Vários fatores podem ser elencados como condicionantes da crise da educação brasileira em relação a evasão escolar, tais como: crescimento desordenado no número de instituições de ensino; problemas de ordem orçamentária nas públicas; descompasso entre as pesquisas produzidas e as demandas da sociedade; descontinuidade e fragilidade dos programas de ensino e pesquisa; currículos fragmentados, defasados, sem interdisciplinaridade e pouco flexíveis; corporativismo muito forte em todas as classes/níveis nas instituições de ensino; a estrutura do poder ainda centralizadora e conservadora [KINPARA e DAN, 1995].

Assim, de acordo com Kinpara e Dan (1995), enquanto a técnica existir dissociada do homem que a usa, de um contexto histórico, entendida como sistema estático, que não se adapta, não se modifica, apenas reativo, e não ativo, não será possível compreender a realidade e muito menos se definir diante dela. Milléo et al. (2000) reforçam este pensamento, quando afirmam que o profissional formado dentro

de uma perspectiva tecnicista e pessimista, dentro e fora de um meio acadêmico, sofre os reflexos manifestando grande insegurança e dificuldade em tratar com diferentes segmentos da sociedade.

Sob esse aspecto, esta pesquisa tem a intenção de atender a duas esferas designadas: a primeira de ordem institucional e a segunda esfera decorrente do aspecto social. Como princípio, esta pesquisa atende ao aspecto institucional ao contribuir para a academia com o estudo na linha de pesquisa de economia regional, enfatizando uma dimensão social, política e econômica nas instâncias da educação técnica / superior , contribuindo, dessa forma, para o avanço dos estudos sobre fenômenos sociais e suas manifestações regionais. A segunda esfera deste projeto atende ao crescimento da demanda social por educação técnica / superior, demanda na qual os trabalhadores buscam refúgio para o enfrentamento de um problema de mudança social que configura mudança no aspecto social [CASTEL, 1998].

Silva Filho (2007) define dois aspectos similares, mas não idênticos: evasão anual, que verifica a diferença entre alunos matriculados de um ano para o outro, e a evasão total, ao comparar o número de alunos matriculados com o número final de alunos concluintes do curso.

2.1 O Problema da Evasão nos Institutos Federais

A Rede Federal de Educação Profissional, Científica e Tecnológica atua na capacitação de diferentes segmentos da mão de obra nacional. Seus cursos variam desde modalidades de curta duração, como os cursos de formação continuada e inicial, em que são requeridos apenas níveis fundamentais de educação, até cursos de mestrado. Conforme estabelecido no Art. 7º, incs. III e IV da Lei 11.892/2008, além de ministrar cursos profissionais, fazem parte dos objetivos da Rede Federal: realizar pesquisas aplicadas, estendendo seus benefícios à comunidade; e desenvolver atividades de extensão, em articulação com o mundo do trabalho e os segmentos sociais com ênfase na produção, desenvolvimento e difusão de conhecimentos científicos e tecnológicos [SILVA, 2009].

Muitos recursos são investidos para a manutenção e expansão da rede federal. O Ministério da Educação (MEC) conta, no Orçamento Geral da União, com o Programa Desenvolvimento da Educação Profissional e Tecnológica, que, no quadriênio 2008-2011, teve recursos empenhados da ordem de R\$ 11,5 bilhões, conforme explicitado na Figura 1. Isso mostra a necessidade de promover um serviço de qualidade a sociedade brasileira no tocante a educação, o que está sensivelmente interligado a diminuição dos índices de evasão.

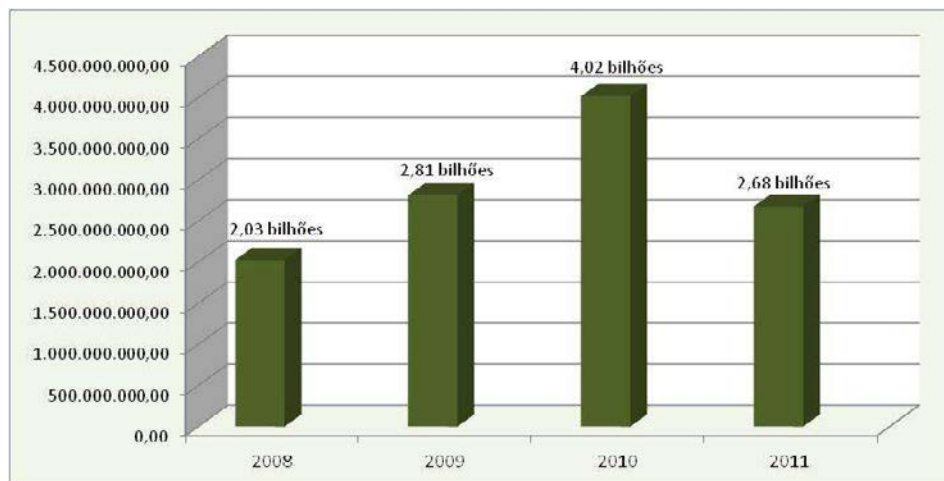


Figura 1 - Valores empenhados no Programa Desenvolvimento da Educação Profissional e Tecnológica, de 2008 a 2011. Fonte: [TCU, 2013].

A evasão representa um problema que alcança diferentes modalidades de ensino em maior ou menor medida. No Brasil, a educação profissional não foge a essa regra, sendo um importante vazamento que impede que boa parte dos alunos concluam seus respectivos cursos. A meta de 90% para a taxa de conclusão prevista no Projeto de Lei do Plano Nacional de Educação 2011-2020, ou mesmo da taxa de 80% para todas as modalidades de cursos ofertados pelos institutos prevista no Termo de Acordo de Metas, aparentemente, ainda é um ideal de longo-prazo. Quando se analisam as taxas de conclusão em nível nacional se situam em 46,8% para o médio integrado, 37,5% para o Proeja, 25,4% para a Licenciatura, 27,5% para o Bacharelado e 42,8% para os cursos de tecnólogo. Em termos de estratégias de combate à evasão, será observado que muitas vezes a atuação dos campi, de determinado Instituto Federal, dá-se de forma isolada em relação aos demais [TCU, 2013].

Frente à relevância desse tema, o presente artigo discorre sobre a caracterização da evasão nos cursos de tecnologia da informação ofertados pelo Instituto Federal do Sertão Pernambucano no Campus do município de Floresta/PE. Propõem-se analisar de modo conjugado, três tipos de indicadores – de evasão, retenção e conclusão – para corretamente monitorar essa evasão.

3. Metodologia

São inúmeros os conceitos sobre pesquisa, posto que os estudiosos ainda não chegaram a um consenso sobre o assunto. Os critérios de classificação dos tipos de pesquisa variam de acordo com o enfoque dado pelo autor. A divisão obedece a interesses, condições, campos, metodologia, situações, objetivos, objetos de estudo etc [LAKATOS e MARCONI, 2008].

Observa-se que para avaliar a evasão escolar faz-se necessário avaliar não somente os indicadores da evasão escolar, mas, também o índice de retenção e de conclusão.

Para a execução desta pesquisa serão utilizados três indicadores:

- Evasão escolar - Taxa de alunos que abandonaram o curso
- Retenção escolar - Atraso no tempo de permanência no curso
- Conclusão escolar - Taxa de conclusão dos alunos

Este artigo utiliza a técnica do estudo de caso como a melhor estratégia para obtenção dos dados. O local escolhido foi o Campus Floresta do Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano devido ao pioneirismo da pesquisa no campus que já possui 8 anos de funcionamento entre os anos de 2008 e 2016.

Coleta dos Dados - A coleta de dados compreende a aplicação dos instrumentos elaborados para esse fim, tendo em vista a importância do controle na sua execução visando à fidedignidade das informações [MARCONI e LAKATOS, 2008]. Para Yin (2004) o pesquisador que prima pela validade do constructo, deverá utilizar múltiplas fontes de evidência, a fim de se obter uma visão mais completa do contexto estudado. Para o autor podem ser consideradas fontes de evidência: a) documentos, b) registros em arquivos, c) entrevistas, d) observação direta, e) observação participante e; f) artefatos físicos. Suas vantagens de coleta, em geral, são: a) permite revisão; b) não foi criada como resultado do estudo de caso, e; c) contém referências e detalhes de um evento. Os registros em arquivos além dessas características são precisos e quantitativos.

Análise Qualitativa / Quantitativa - Os estudos estatísticos estão relacionados às situações que envolvem planejamentos, coleta de dados, organização de informações, análise das informações coletadas, interpretação e divulgação de forma clara e objetiva.

A coleta, a organização, a descrição dos dados, o cálculo e a interpretação de coeficientes pertencem à Estatística Descritiva, enquanto a análise e a interpretação dos dados, associados a uma margem de incerteza, ficam a cargo da Estatística Indutiva ou Inferencial, também chamada como a medida da incerteza ou métodos que se fundamentam na teoria da probabilidade [TRIOLA, 2011].

4. A Evasão Escolar nos Cursos de Tecnologia da Informação no Instituto Federal do Sertão Pernambucano – Campus Floresta

A evasão escolar nos cursos de tecnologia da informação no Instituto Federal do Sertão Pernambucano Campus Floresta/PE, apresenta índices muito elevados, chegando a

quase 50% de alunos evadidos desde o início do funcionamento da instituição em 2008, de acordo com dados coletados na base de dados do (SISTEC).

Um índice de evasão escolar de 49,78% é muito indesejável em qualquer curso que ocorra, merecendo ainda mais atenção quando se trata de cursos relacionados à tecnologia da informação, pois diante do desenvolvimento e do crescimento exponencial da tecnologia e da sua conseqüente necessidade de profissionais habilitados, este índice torna-se muito preocupante.

No gráfico constante na Figura 2, pode ser observada a diferença entre o percentual de alunos evadidos e de alunos concluídos ou em curso.

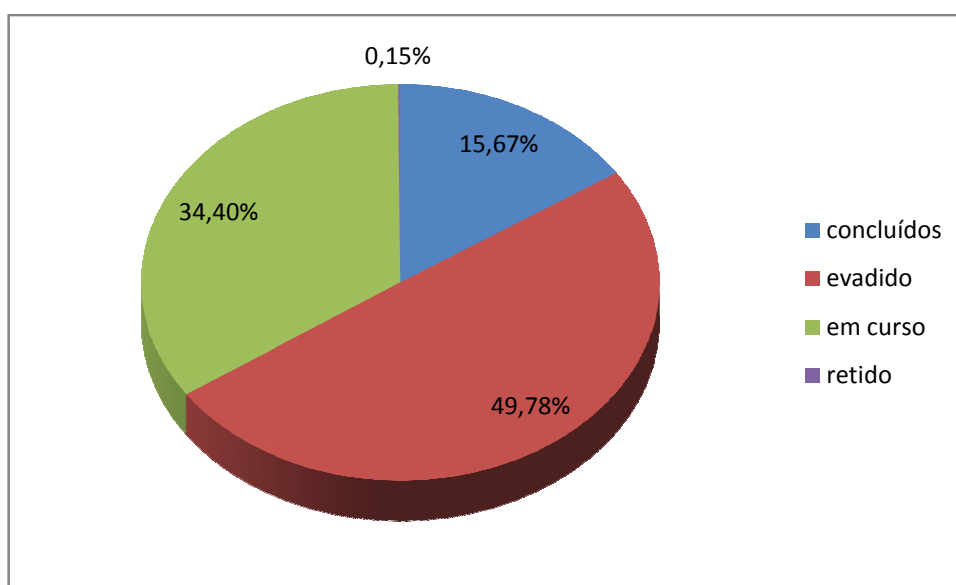


Figura 2 - Percentual de Alunos Evadidos.

De acordo com dados coletados na base de dados do (SISTEC), pode-se inferir que a taxa de conclusão é de 15,67% e a quantidade de alunos em curso é 34,40 %, caso todos os alunos em curso venham a concluir os seus respectivos cursos, o percentual de alunos concluintes será de 50,07% o que estará muito distante da meta estabelecida pelo Projeto de Lei do Plano Nacional de Educação 2011-2020. Estes números podem ser observados na Tabela 1.

O número total de evadidos (TV) nos cursos de tecnologia da informação no período considerado foi de 344 (TV = EV + RT).

Tabela 1 - Números da Evasão nos Cursos de Tecnologia da Informação no Instituto Federal do Sertão Pernambucano Campus Floresta

STATUS	ALUNOS
Concluídos (CO)	108
Evadido (EV)	343
Em Curso (EC)	237

Retido (RT)	1
TOTAL	689

Diante destes números a fica evidente a importância de este assunto ser cada vez mais abordado em estudos.

5. Resultados

De acordo com os formulários aplicados, infere-se que os alunos apresentam um perfil de indecisão, conforme pode ser observado na Tabela 3, quando questionados a dúvida na escolha do curso a resposta é negativa em sua maioria, porém, quando questionados a respeito de um possível curso novo a ser ofertado pela instituição os cursos com maior índice de escolha são de áreas bem díspares da área de tecnologia da informação, conforme pode ser observado no Figura 3.

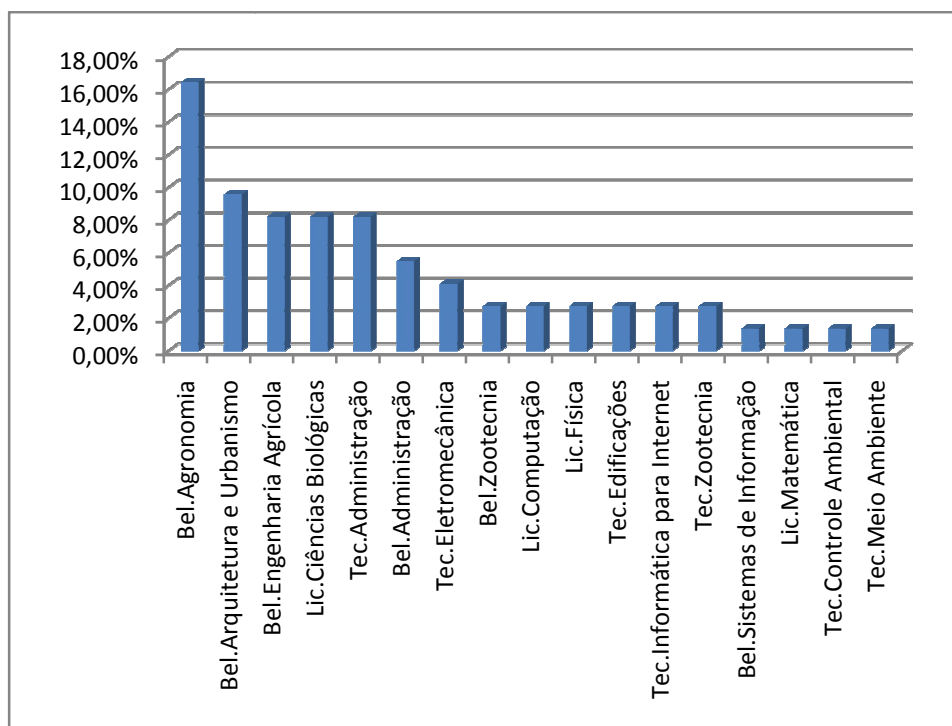


Figura 3 - Principais Cursos Escolhidos

Tabela 2 - Principais respostas escolhidas

PERGUNTA	Resposta mais escolhida
Teve dúvida na escolha do curso	Não (67,12%)
Se a instituição fosse ofertar um novo curso	Bel. Agronomia (16,44%)
Teve aula de nivelamento antes de iniciar o curso	Não (90,41%)
Faria um outro curso	Sim, em outra área (31,51%)
Qualidade do ensino ofertado como fator motivador	Concordo Totalmente (50,68%)

da permanência no curso	
Principal dificuldade para permanência no curso (Horário de início da aula)	Concordo Totalmente (61,64%)
A infraestrutura física do instituto deve ser melhorada	Concordo Totalmente (50,68%)

Com o avanço tecnológico que se vive nos dias atuais, um índice de evasão tão alto em cursos de tecnologia da informação torna-se muito preocupante, e medidas atenuadoras deste fenômeno devem ser colocadas em pauta com mais frequência. As novas tecnologias que aparecem no mercado de trabalho precisam de pessoas cada vez mais qualificadas e adaptadas as mesmas, para que possa haver um desenvolvimento socioeconômico mais efetivo.

6. Considerações Finais

De forma a enfrentar a evasão motivada por deficiências de formação, muitas instituições de Ensino Superior vêm implementando programas de nivelamento em conteúdos básicos para alunos ingressantes [SILVEIRA, et al. 2015].

Conforme pode ser observado na tabela 3, é possível inferir que a instituição não oferta cursos de nivelamento. Implantar programas de nivelamento para alunos ingressantes seria uma das soluções atenuadoras da evasão escolar, visto que os novos alunos ingressariam na instituição com um conhecimento mais aprimorado a respeito do curso em que tiver ingressando, evitando que os alunos tenham frustração logo no início da atividade acadêmica, levando-os posteriormente a evasão. A qualidade do curso oferecida pela instituição é positivamente avaliada pelos discentes, porém, a infraestrutura deve ser melhorada a fim de trazer um ambiente mais confortável.

Referências

- CASARTELLI, A. O. *et al.* A evasão da educação superior: uma análise da produção de conhecimento no Brasil. In: LEITE, D.; FERNANDES, C.B. (Org.). Qualidade da educação superior: avaliação e implicações para o futuro da universidade. Porto Alegre: EDIPUCRS, 2012. p. 75-86.
- CASTEL, R. As metamorfoses da questão social: uma crônica do salário. Petrópolis: Vozes, 1998.
- GAIOSO, N. P. L. O fenômeno da evasão escolar na educação superior no Brasil. Universidade Católica de Brasília, 2005. Disponível em: http://www.scielo.br/scielo.php?pid=S1414-40772011000200007&script=sci_arttext. Acesso em: 16/07/2015.
- KINPARA, Daniel I.; DAN, Edival. O papel da universidade na formação de educadores a partir de profissionais de Ciências Agrárias. Lavras, MG, 1995. 62 f.

- Dissertação (Mestrado em Administração Rural) - Universidade Federal de Lavras. Lavras, MG, 1995.
- MARCONI, Marina de Andrade; LAKATOS, Eva Maria. Pesquisa. In: Técnicas de Pesquisa: Planejamento, e Execução de Pesquisas, Amostras e Técnicas de Pesquisa, Elaboração, Análise e Interpretação de Dados. 7 ed. São Paulo: Atlas, 2008. p. 1-25.
- SILVA, Caetana Juracy (org.). Institutos Federais Lei 11.892, de 29/12/2008: Comentários e Reflexões. Natal: IFRN, 2009.
- SILVA FILHO, R L Lobo *et al.* Evasão no Ensino Superior Brasileiro, Instituto Lobo para o desenvolvimento da educação, ciência e tecnologia. Cadernos de Pesquisa, v.37, n. 132, p.641-659. Set/Dez, 2007.
- SILVEIRA, I. F. *et al.* O Desafio da Aprendizagem Adaptativa em Programas de Nivelamento para o Ensino Superior. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 35. 2015, Recife/PE.
- TCU, TC 026.062/2011-9, Relatório de Auditoria, 2013. Auditoria operacional. Fiscalização de orientação centralizada. Rede Federal de Educação Profissional. Necessidade de aprimoramentos nas atuações relacionadas à evasão escolar, à interação com os arranjos produtivos locais e ao apoio à inserção profissional dos alunos. Carência de professores e de profissionais de laboratório. Ausência de instalações físicas adequadas em alguns Institutos Federais. Recomendações. Determinação. Comunicações.
- TRIOLA, M. Introdução a estatística. 10. Ed. São Paulo: LTC, 2011.
- YIN, Robert K. Case study research: design and methods. Thousands Oaks: SAGE, 1994.

Uma Ferramenta para Ensinar Introdução à Teoria da Computação

Candido Ferreira Xavier de Mendonça¹, Luciano A. Digiampietri¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
Av. Arlindo Bértio, 1000 – 03828-000 – São Paulo – SP – Brasil

{cfxavier,digiampietri}@usp.br

Abstract. *Keeping students engaged from the beginning is one of the most important challenges in the undergraduate lectures. This challenge becomes more critical in theoretical subjects. This work reports the application of the concept of “proximal development zone” of the theory of Vygotsky, where a tool (“toy”) is applied to mediate student learning conflicts in the “Introduction to the Theory of Computation” subject.*

Resumo. *Manter a motivação e o interesse do aluno desde o primeiro dia de aula é um dos maiores desafios da educação superior. O qual fica mais acentuado nas disciplinas teóricas. Este trabalho reporta a aplicação do conceito de “zona de desenvolvimento proximal” da teoria de Vygotsky na construção de uma ferramenta (“brinquedo”) utilizada como proposta para mediar conflitos dos alunos na aprendizagem do conteúdo da disciplina “Introdução à Teoria da Computação”.*

1. Introdução

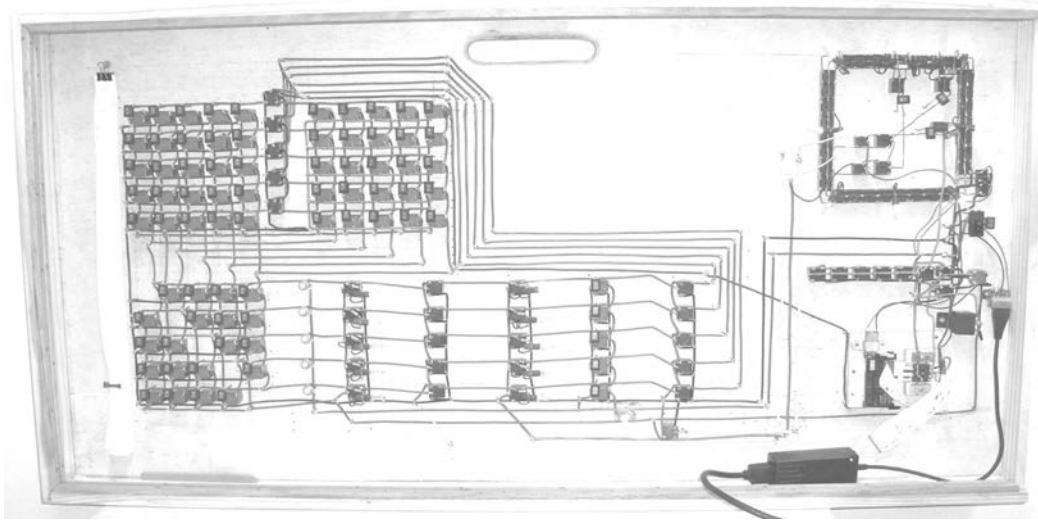
A disciplina Introdução à Teoria da Computação na maioria dos bacharelados da área de computação consiste no estudo da complexidade de três modelos computacionais teóricos: o autômato finito, o autômato à pilha e a máquina de Turing, com livros textos dentro dos quais destaca-se “Introdução à Teoria da Computação” [7].

Professores da disciplina “Introdução à Teoria da Computação” reclamam de um certo “desconforto” de alguns alunos na questão “modelos teóricos”. Este desconforto fica mais evidente no início da apresentação de cada um dos 3 modelos computacionais da disciplina. Naturalmente, observamos que este comportamento não aparece em disciplinas práticas. Assim, o desafio é apresentar a disciplina buscando uma mediação para os conflitos citados.

Uma das propostas de remediação para esse desafio é dada por Vygotsky [2] onde o “brinquedo/brincadeira” (no caso, a ferramenta construída) é utilizado para melhorar o potencial de aprendizagem (definido por Vygotsky como “zona do desenvolvimento proximal”). Assim, para aplicar a Teoria de Vygotsky na disciplina Introdução à Teoria da Computação, seria necessário encontrar uma ferramenta adequada.

Existem excelentes ferramentas, como JFLAP [1, 6] desenvolvida por Susan Rodger, que podem ser utilizadas na disciplina em questão, porém, não são de caráter introdutório. Portanto, a escolha foi criar um material próprio, construindo um autômato

Figura 1. Foto do AFNDP (dimensões: 120 × 60 cm).



finito não-determinístico programável, doravante denominado por AFNDP¹, como pode ser visto na Figura 1. Este protótipo consiste numa máquina física, permitindo a implementação de qualquer autômato finito seja ele “determinístico”, “não-determinístico” ou “não-determinístico com ε -transições”, contendo até 5 estados, reconhecendo entradas do alfabeto $\{0,1\}$ registradas em uma fita de papel perfurado de 25mm.

Ao longo deste trabalho definiremos o conjunto do AFNDP mais sua programação como *autômato finito*, explicando assim não apenas o funcionamento do AFNDP, mas também a sequência de sua utilização ao longo das aulas.

O restante deste trabalho é organizado como segue: a seção 2 descreve a máquina e seu funcionamento, incluindo a tabela de execução para uma dada entrada; a seção 3 descreve a máquina e a tabela de execução para mostrar a equivalência entre o autômato finito “determinístico” e o autômato finito “não-determinístico”; a seção 4 descreve o uso da tabela de execução da máquina na demonstração do “Pumping Lemma” e a seção 5 conclui este trabalho.

2. A Ferramenta AFNDP

O AFNDP é apresentado logo na primeira aula, iniciando com um breve histórico dos três modelos computacionais que serão apresentados dentro da disciplina, isto é, o autômato finito, o autômato à pilha e a máquina de Turing. Este histórico é apresentado a seguir numa forma resumida.

Os dois primeiros modelos computacionais: autômato finito e autômato à pilha, foram criados por McCulloch e Pitts [4] em 1943, devido ao fato dos autores estarem estudando o comportamento do sistema nervoso, eles aparecem na literatura com o nome de neurônios. Treze anos mais tarde (1956), no trabalho de Kleene [3], esses neurônios serão renomeados para *Autômatos Finitos*, no qual é mostrada a relação entre Autômatos Finitos e Expressões Regulares, isto é, modelos com um tamanho limitado de memória que independe do tamanho da entrada. Três anos mais tarde (1959), Rabin e Scott [5]

¹O diagrama esquemático do AFNDP está disponível em: 143.107.58.249/html/afnd-t2-r.eps

publicaram um estudo bastante completo sobre a Teoria dos Autômatos Finitos. Este último trabalho foi condecorado com “Turing Award”, a maior premiação dada a trabalhos em Ciência da Computação. O terceiro modelo computacional, a máquina de Turing, foi criado por Alan Turing em 1936 e apresentado em seminários nos anos subsequentes [8, 9]. Curiosamente, a criação do modelo mais complexo antecede em 7 anos a criação dos dois modelos mais simples.

Este histórico é utilizado com o propósito de contextualizar o AFNDP. Dentro deste contexto foi criada uma máquina em que foram utilizados componentes eletromecânicos encontrados no final do século XIX, a saber: 1 motor bipolar, 151 relés, 20 lâmpadas, motores bipolares, 80 chaves, 11 campainhas, condutores e um mecanismo de impressão de uma máquina de cartão de crédito (transformado em leitora de fita de papel perfurado). Além disso, o projeto foi executado de modo semelhante a um trabalho de adolescente apresentado em uma feira de ciências do segundo grau.

O componente fundamental neste projeto é o relé de 5 terminais – um dispositivo eletromecânico responsável pela comutação de contatos elétricos, como mostrado na Figura 2. Seu funcionamento é bem simples: quando não há corrente entre os terminais da bobina, a mola de rearme projeta a armadura para “baixo”, fazendo o contato entre os polos C e NF. Neste caso dizemos que o relé está *inativo*. Porém, na presença da corrente entre os terminais da bobina, um campo magnético é formado, atraindo a armadura, que se projeta para “cima”, fazendo o contato entre os polos C e NA. Neste caso dizemos que o relé está *ativo*.

Figura 2. Um relé de 5 terminais.

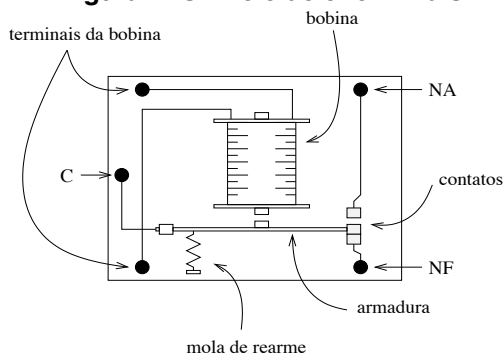
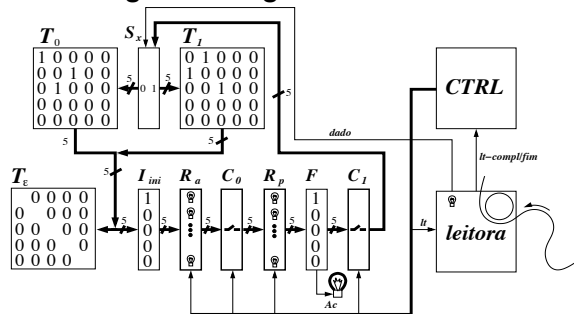


Figura 3. Diagrama de blocos.



2.1. O funcionamento do AFNDP

O diagrama de blocos funcionais do AFNDP pode ser visto na Figura 3. Os blocos funcionais do AFNDP são divididos em: tabelas, registradores e blocos de controle. As tabelas são caracterizadas pela presença de uma chave anexa a cada relé – exceto a tabela I_{ini} que consiste de uma coluna de campainhas. Os registradores são caracterizados pela presença de uma lâmpada anexa a cada relé do registrador, esta lâmpada serve para indicar se num dado momento, o relé correspondente está ativo ou não. Os blocos de controle não possuem uma característica definida, mesclando o uso de relés, lâmpadas e chaves de maneira não uniforme. Esses blocos funcionais estão dispostos em duas linhas: superior e inferior. A linha superior contém a tabela T_0 , o seletor S_X , a tabela T_1 e a unidade de controle $CTRL$. A linha inferior contém a tabela T_ϵ , a tabela de ativação inicial I_{ini} , o registrador auxiliar R_a , o bloco de conexão C_0 , o registrador principal R_p , a tabela F , a lâmpada Ac

para indicar estado de aceitação do AFNDP, o bloco de conexão C_1 e a leitora de papel perfurado.

A programação do autômato finito deve ser feita por meio das tabelas: I_{ini} , T_0 , T_1 , T_ε e F , onde um “1” significa que a chave (ou campainha) correspondente foi acionada, e um “0” significa que a chave ou campainha correspondente não foi acionada. A tabela I_{ini} é a única tabela com campainhas que devem ser acionadas somente no início da execução, indicando que o estado correspondente inicia “ativo”. Um “1” na linha i e coluna j da tabela T_0 (resp. T_1 ou T_ε) indica a presença, durante a execução, da transição do estado q_i para o estado q_j para a entrada “0” (resp. transição para a entrada “1” ou transição- ε). Naturalmente, um “0” indica, que durante a execução, a transição não estará presente. A tabela F contém uma coluna com 5 chaves, onde um “1” na linha i significa, que durante a execução, a chave da linha i estará ligada (o estado q_i é um estado final). Neste caso, se o estado q_i estiver ativo, a lâmpada Ac acende indicando estado de “aceitação”. Naturalmente, um “0” na linha i significa que a chave estará desligada (o estado q_i não é um estado final).

Assim, a programação exemplificada no AFNDP mostrada na Figura 3, a qual chamaremos de autômato finito M_1 , pode ser descrita pelo diagrama de estados mostrado na Figura 4, onde usamos a forma pictórica de círculo para cada estado e arco para cada transição entre estados, estas transições dependem do caractere obtido na leitora. Inicialmente, acionamos a campainha referente ao estado q_0 (a tabela I_{ini} possui um “1” na linha 0), isso é indicado no diagrama por uma seta vindo do nada apontando para o estado q_0 . Sempre que o símbolo “1” aparecer na linha i e coluna j na tabela T_0 (resp. tabela T_1 ou tabela T_ε) isso é indicado por um arco com rótulo “0” (resp. “1” ou “ ε ”) ligando o estado q_i ao estado q_j no diagrama de estados. Finalmente, sempre que o símbolo “1” aparece na linha i da tabela F , isso indica que a lâmpada Ac ficará acesa caso este estado esteja ativo, isto é indicado no diagrama marcando o estado q_0 com duas circunferências concêntricas. Note que o diagrama de estados define claramente a programação do AFNDP.

Figura 4. Diagrama de estados do AFD M_1 .

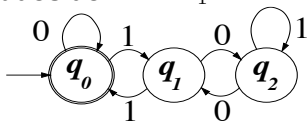
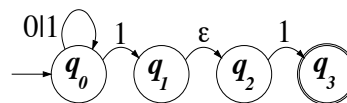


Figura 5. Diagrama de estados do AFND M_2 .



A seguir, mostramos o funcionamento do AFNDP para um autômato finito qualquer, onde a tabela T_ε contém apenas zeros, e, como não desempenha nenhuma função, será omitida.

Inicialização:

Quando o AFNDP é ligado a conexão C_0 está ativada, os dois registradores R_a e R_p estão habilitados e vazios, as tabelas: T_0 , T_1 e F estão desabilitadas e a unidade *leitora* está posicionada no início da fita de papel. Primeiramente, o estado inicial escolhido é ativado usando uma das 5 campainhas da tabela I_{ini} , que será registrado em R_a e R_p . Na sequência, ativamos a unidade de controle $CTRL$ para que execute ciclicamente as transições.

Execução de uma transição:

A unidade de controle $CTRL$ inicia cada transição com a conexão C_0 ativada, isto é,

o registrador auxiliar R_a terá um conteúdo idêntico ao do registrador principal R_p . Na sequência, a conexão C_0 é desativada e a conexão C_1 é ativada². Simultaneamente, um sinal de controle lt é passado para a unidade *leitora* para que obtenha o próximo caractere na fita de papel. Durante a leitura, as tabelas: T_0 e T_1 são desabilitadas e o registrador auxiliar R_a é apagado. Quando a unidade *leitora* completa o ciclo de leitura do próximo caractere (*dado*), este é enviado para o seletor S_X e um sinal de completude do ciclo de leitura $lt - compl$ é enviado para a unidade de controle $CTRL$, que irá habilitar as tabelas: T_0 e T_1 e o registrador auxiliar R_a . Na sequência, o seletor S_X , fazendo uso do *dado* enviado pela unidade *leitora*, seleciona a tabela que deve receber o conteúdo do registrador principal R_p . Independentemente de qual seja a tabela usada, a saída é registrada no registrador auxiliar R_a . Na sequência, a unidade de controle $CTRL$ apaga o conteúdo do registrador principal R_p , ativa a conexão C_0 e habilita o registrador principal R_p , que irá copiar o conteúdo do registrador auxiliar R_a .

Final da execução:

No caso em que a fita de papel termine durante a execução do ciclo de leitura, no meio da execução de uma transição, um sinal *fim* é enviado à unidade de controle $CTRL$, que interrompe o processamento. Neste momento, caso a lâmpada Ac esteja acesa, o AFNDP termina em estado de “aceitação”, caso contrário (se a lâmpada Ac estiver apagada) o AFNDP termina em estado de “rejeição”.

A título de exemplo, vamos acompanhar a execução do autômato M_1 para a palavra “1001”, que pode ser visto na Tabela 1. Inicialmente, pressionamos a campainha da linha 0 de I_{ini} , ativando a unidade de controle $CTRL$ para que execute as transições. O primeiro caractere lido é um “1”. Logo, S_X irá colocar o conteúdo de R_p na entrada da tabela T_1 , onde q_0 ativa q_1 . No final desta transição, o estado q_1 é o único estado ativo. O próximo caractere lido é um “0”. Logo, S_X irá colocar o conteúdo de R_p na entrada da tabela T_0 , onde q_1 ativa q_2 . O próximo caractere lido é um “0”. Logo, S_X irá colocar o conteúdo de R_p na entrada da tabela T_0 , onde q_2 ativa q_1 . O próximo caractere lido é um “1”. Logo, S_X irá colocar o conteúdo de R_p na entrada da tabela T_1 , onde q_1 ativa q_0 . Por fim, como a cadeia de entrada terminou, o AFNDP finaliza a execução. Entretanto, como o estado q_0 está ativo e este estado é final (a chave 0 da tabela F está ligada), a lâmpada Ac está acesa, indicando que a cadeia “1001” foi aceita. De fato, essa cadeia é a representação binária do número 9 e o autômato implementado no exemplo aceita uma representação binária de um múltiplo de 3, incluindo a cadeia vazia.

Tabela 1. Exec. para a entrada “1001”

	I_{ini}	1	0	0	1
q_0	x				x
q_1		x		x	
q_2			x		

Tabela 2. Definição de ψ para M_1

ψ	0	1
q_0	q_0	q_1
q_1	q_2	q_0
q_2	q_1	q_2

O autômato M_1 pode ser definido formalmente como: $M_1 = (\{q_0, q_1, q_2\}, \{0, 1\}, q_0, \{q_0\}, \psi)$, onde ψ é definida pela Tabela 2.

Observamos que o AFNDP aqui mostrado somente permite a implementações de autômatos de no máximo 5 estados. Porém, caso o AFNDP não possuísse tal limitação, perguntamos: Que tipo de linguagens o autômato finito reconhece? Que tipo de linguagens o autômato finito não reconhece?

²as duas conexões C_0 e C_1 são complementares, quando uma é ativada a outra é desativada e vice-versa

Figura 6. Programação de $T_0, T_1, T_\varepsilon, I_{ini}$ e F para o AFND M_2 .

$\begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$	$\begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$	$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix}$	$\begin{matrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix}$	$\begin{matrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{matrix}$
T_0	T_1	T_ε	I_{ini}	F

3. AFD e AFND são equivalentes

Embora a máquina apresentada tenha o nome de AFNDP (Autômato Finito Não-Determinístico Programável). O modelo de programação e funcionamento mostrados, como veremos a seguir, é “determinístico”. Porém, como desejamos responder a pergunta “Que tipo de linguagens o AFNDP reconhece?”, precisamos explorar todas as possibilidades. Para isso devemos fazer as seguintes definições.

Dizemos que um autômato finito é *determinístico* se existe uma única instância do AFNDP em execução, onde um e somente um estado está ativo no final de cada transição.

Dizemos que um autômato finito é *não-determinístico* se durante a execução do AFNDP é possível haver mais de uma instância em funcionamento sincronizado na mesma posição para a mesma cadeia de entrada onde cada instância possui estados ativos q_i e q_j diferentes. Porém, é fácil demonstrar que as múltiplas instâncias citadas são equivalentes a ter uma única instância do autômato finito com ambos os estados q_i e q_j ativos.

Um exemplo de diagrama de estados de um autômato não-determinístico (autômato M_2) pode ser visto Figura 5 e a programação de sua implementação no AFNDP pode ser vista na Figura 6.

O autômato M_2 pode ser definido formalmente como: $M_2 = (\{q_0, q_1, q_2, q_3\}, \{0, 1\}, \{q_0\}, \{q_3\}, \psi)$, onde ψ é definida pela Tabela 3.

Tabela 3. Def. de ψ para M_2

ψ	0	1	ε
q_0	$\{q_0\}$	$\{q_0, q_1\}$	\emptyset
q_1	\emptyset	\emptyset	\emptyset
q_2	\emptyset	$\{q_3\}$	\emptyset
q_3	\emptyset	\emptyset	\emptyset

Tabela 4. Execução de M_2 para “1011”

	I_{ini}	E	1	E	0	E	1	E	1	E
q_0	x	x	x	x	x	x	x	x	x	x
q_1			x	x			x	x	x	x
q_2				x				x		x
q_3									x	x

Durante a execução do AFNDP para um autômato não-determinístico, caso este contenha transições- ε , haverá uma mudança na execução de cada transição, pois a tabela T_ε pode acrescentar novos estados ativos ao registrador auxiliar R_a antes deste ser copiado no registrador principal R_p . Deste modo, o diagrama de execução passa a ter uma nova coluna E para registrar estes acréscimos. Um exemplo do diagrama de execução do autômato finito não determinístico M_2 para a cadeia de entrada “1011” pode ser visto na Tabela 4. Note que toda vez que aparece um caractere “1” no estado q_1 , o estado q_2 é ativado na coluna E correspondente.

A execução da Tabela 4 (o AFNDP para M_2) pode ser explicada como segue. Inicialmente ativamos o estado q_0 , em seguida, copiamos este estado ativo na coluna E . Como não existe nenhuma transição- ε saindo de nenhum estado ativo para um estado

inativo, então, esta coluna E é uma cópia da coluna anterior. Em seguida, o caractere “1” é lido, para este caractere o estado q_0 ativa os estados q_0 e q_1 . Estes dois estados são copiados na próxima coluna E . Como existe uma transição- ε de q_1 (ativo) para q_2 (inativo), este estado q_2 é marcado como ativo. Em seguida o caractere “0” é lido, para este caractere, o estado q_0 é o único estado a ser ativado. Esta coluna é copiada na coluna E seguinte. Novamente, não existe nenhuma transição- ε saindo de nenhum estado ativo para um estado inativo, logo esta coluna E é uma cópia da coluna anterior. Em seguida, o caractere “1” é lido, para este caractere e o estado q_0 ativa os estados q_0 e q_1 . Estes dois estados são copiados na próxima coluna E . Como existe uma transição- ε saindo de q_1 (ativo) para q_2 (inativo), este estado q_2 é marcado como ativo. Em seguida, o caractere “1” é lido, para este caractere o estado q_0 ativa os estados q_0 e q_1 e o estado q_2 ativa o estado q_3 . Estes três estados são copiados na próxima coluna E . Como existe uma transição- ε saindo de q_1 (ativo) para q_2 (inativo), este estado q_2 é marcado como ativo. Neste momento o AFNDP termina o processamento em estado de aceitação, pois o estado final q_3 está ativo e a lâmpada Ac está acesa.

Agora, vamos construir a Tabela 5 que consiste no conjunto de estados ativos durante a execução do AFNDP, onde a linha Q_i indica os estados ativos no final de cada transição i , para o autômato finito não-determinístico M_2 , com uma entrada qualquer.

Iniciamos a Tabela 5, linha dois, coluna um, com o conjunto que pode ser obtido por meio de uma nova tabela (Tabela 6). Essa tabela é construída da seguinte forma. Na primeira coluna colocamos todos os estados do autônomo M_2 e na última linha da segunda coluna, incluímos o conjunto unitário contendo o estado inicial de M_2 . Em seguida marcamos como ativo esse estado. Essa coluna é copiada na coluna seguinte (E). Como não existe nenhuma transição- ε (em M_2) de q_0 para um estado inativo, a coluna E é uma cópia da coluna I_{ini} . Por fim, criamos um conjunto na última linha da coluna E contendo todos os estados marcados nesta coluna (no exemplo, $\{q_0\}$). Esse valor é transportado para a Tabela 5. Feito isto, é necessário completar as duas últimas colunas desta segunda linha. O procedimento para o preenchimento destas colunas é feito como segue. Cria-se uma tabela contendo quatro colunas. Na primeira coluna colocamos todos os estados do autônomo M_2 e na última linha da segunda coluna, incluímos o conjunto indicado na primeira coluna da Tabela 5. Marcamos nesta coluna como ativos todos os estados que aparecem nesse conjunto. Colocamos na primeira linha, segunda coluna, o caractere “0” (resp. “1”) que seria lido na cadeia de entrada. Nessa coluna é indicado que estados seriam ativados por M_2 para o caractere “0” (resp. “1”). A terceira coluna é copiada na coluna quatro (E) e verifica-se possíveis transições- ε completando o conjunto de estados ativos. Todos os estados marcados nesta coluna são copiados no conjunto representado na última linha. Este conjunto é copiado na coluna indicada por “0” (resp. “1”) da Tabela 5.

Tabela 5. Tabela de estados ativos do AFND M_2 .

$\{Q_i\}$	0	1
$\{q_0\}$	$\{q_0\}$	$\{q_0, q_1, q_2\}$
$\{q_0, q_1, q_2\}$	$\{q_0\}$	$\{q_0, q_1, q_2, q_3\}$
$\{q_0, q_1, q_2, q_3\}$	$\{q_0\}$	$\{q_0, q_1, q_2, q_3\}$

Tabela 6. Estados ativos, no início e no final da transição- ε .

	I_{ini}	E
q_0	x	x
q_1		
q_2		
q_3		
	$\{q_0\}$	$\{q_0\}$

Usando esse procedimento, completamos a linha dois da Tabela 5, obtendo os conjuntos $\{q_0\}$ e $\{q_0, q_1, q_2\}$ conforme Tabela 7. Note que o novo conjunto $\{q_0, q_1, q_2\}$ foi transportado para a coluna um linha três da Tabela 5.

Tabela 7. Simulação do comportamento do AFNDP para as entradas “0” e “1”

		0	E
q_0	x	x	x
q_1			
q_2			
q_3			
	$\{q_0\}$		$\{q_0\}$

		1	E
q_0	x	x	x
q_1		x	x
q_2			x
q_3			
	$\{q_0\}$		$\{q_0, q_1, q_2\}$

Usando o mesmo procedimento, completamos a linha três da Tabela 5, obtendo os conjuntos $\{q_0\}$ e $\{q_0, q_1, q_2, q_3\}$ conforme Tabela 8. Note que o novo conjunto $\{q_0, q_1, q_2, q_3\}$ foi transportado para a coluna um linha quatro da Tabela 5.

Tabela 8. Simulação do comportamento do AFNDP para as entradas “0” e “1”

		0	E
q_0	x	x	x
q_1	x		
q_2	x		
q_3			
	$\{q_0, q_1, q_2\}$		$\{q_0\}$

		1	E
q_0	x	x	x
q_1	x	x	x
q_2	x		x
q_3		x	x
	$\{q_0, q_1, q_2\}$		$\{q_0, q_1, q_2, q_3\}$

Usando o mesmo procedimento, completamos a linha quatro da Tabela 5, obtendo os conjuntos $\{q_0\}$ e $\{q_0, q_1, q_2, q_3\}$ conforme Tabela 8. Note que, nenhum novo subconjunto do conjunto de estados Q do autômato M_2 aparece nas colunas dois e três da Tabela 5, e, portanto ela está completa³.

Tabela 9. Simulação da execução do AFNDP para o autômato M_2 - estados ativos

		0	E
q_0	x	x	x
q_1	x		
q_2	x		
q_3			
	$\{q_0, q_1, q_2, q_3\}$		$\{q_0\}$

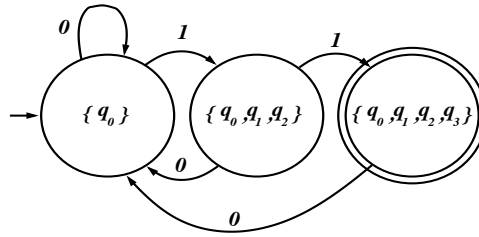
		1	E
q_0	x	x	x
q_1	x	x	x
q_2	x		x
q_3		x	x
	$\{q_0, q_1, q_2, q_3\}$		$\{q_0, q_1, q_2, q_3\}$

O comportamento dos estados ativos no AFNDP na simulação da execução do autômato finito não-determinístico M_2 , para uma entrada genérica, pode ser exemplificado pelo diagrama de estados ativos mostrado na Figura 7. Porém, este diagrama de estados é o diagrama de um autômato finito determinístico. Não é difícil mostrar que isto é verdade para qualquer AFND, logo o AFD e o AFND são equivalentes.

Embora o não-determinismo não acrescente nenhum poder novo ao AFNDP, ele nos ajuda a responder a primeira pergunta. Neste ponto, seguindo os melhores livros de Linguagens Formais, introduzindo as Expressões Regulares e mostrando a sua equivalência com autômatos não-determinísticos. Com isto, respondemos “Que tipo de linguagens o AFNDP reconhece?”.

³Observamos que, no pior caso, o número de conjuntos que pode aparecer nessa tabela é $2^{|Q|}$, isto é, no máximo o número de subconjuntos do conjunto Q .

Figura 7. Diagrama de estados ativos durante a execução do AFNDP para autômato M_2 .



4. Pumping Lemma

O “Pumping Lemma” é peça chave para responder: “O que não é uma linguagem regular?” (não existe um autômato que reconheça a linguagem).

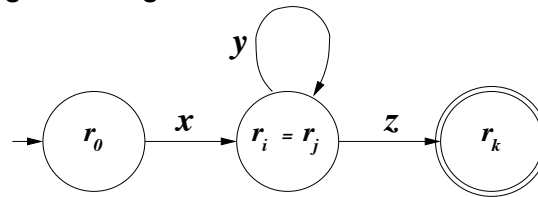
Lema 1. *Se L é uma linguagem regular então existe um número n tal que toda palavra s pertencente à linguagem L , onde $|s| \geq n$, pode ser dividida em três sub-palavras x , y e z , tal que:*

- (i) $|xy| \leq n$,
- (ii) $|y| > 0$ e
- (iii) $xy^*z \in L$.

Demonstração. Seja L uma linguagem regular e $M' = (Q, \Sigma, q_0, F, \psi)$ um autômato finito determinístico que reconhece L . Seja n o número de estados do autômato M' , isto é, $n = |Q|$. Seja $s = s_1s_2s_3\dots s_k$ uma palavra qualquer de L tal que $|s| = k \geq n$. Agora, executamos o AFNDP para os primeiros n caracteres de s , conforme a Tabela 10. Note que renomeamos os estados no final de cada transição (última linha da tabela), renomeando para r_0 o estado ativo inicial (coluna I_{ini}), para r_i o estado ativo na transição onde o caractere s_i foi lido. Então, contamos o número de estados que ficaram ativos durante esse processo. Essa contagem dá $n + 1$, pois temos $n + 1$, contando as colunas de 0 à n . Porém, existem apenas n estados no autômato M' , logo, pelo menos um destes estados é repetido durante essa execução. Sejam i o menor valor tal que o estado r_i se repete durante a execução. Seja r_j a primeira repetição de r_i . Sejam x a sub-palavra formada pelos caracteres $s_1s_2\dots s_i$, y a sub-palavra formada pelos caracteres $s_{i+1}s_{i+2}\dots s_j$ e z a sub-palavra formada pelos caracteres $s_{j+1}s_{j+2}\dots s_k$. Naturalmente, pela escolha de i e j as sub-palavras x e y satisfazem (i) e (ii), pois existe uma repetição de estado garantida dentro das primeiras n transições. Ora, a palavra s pertence à linguagem, logo, o estado r_k pertence à F . Além disso r_j é uma repetição de r_i , logo, quando a execução chega na transição i (no final da sub-palavra x) podemos repetir qualquer quantidade de vezes a palavra y (inclusive zero vezes) sempre terminando com o estado $r_i = r_j$ ativo. Assim, se executarmos M' para a sub-palavra z chegaremos no final com r_k ativo (veja Figura 8). Logo, $xy^*z \in L$ o que satisfaz (iii). □

Tabela 10. Execução do AFNDP para os primeiros n caracteres de s .

	I_{ini}	s_1	s_2	s_3	...	s_n	s_{n+1}	...	s_k
q_0	x				
q_1		x			
q_2			x		
...	x		...	
q_{n-1}				x	
	r_0	r_1	r_2	r_3	...	r_n	r_{n+1}	...	r_k

Figura 8. Diagrama de estados do autômato M' .

5. Conclusão

Por se tratar de um protótipo, a máquina foi apenas apresentada aos alunos. A apresentação do AFNDP na primeira aula⁴ de Introdução à Teoria da Computação tem se mostrado uma excelente estratégia de ensino nos últimos oito anos. Não poucas vezes, após o término da aula às 22:30h, alguns alunos se detinham, curiosos sobre a máquina.

O diagrama de execução aqui descrito facilita a introdução tanto da forma determinística como da forma não-determinística dos autômatos finitos e ajuda na demonstração de sua equivalência. Também, simplifica a demonstração do “Pumping Lemma”.

A experiência têm mostrado que a disciplina iniciada com o AFNDP consegue manter um número bastante elevado de alunos concentrados até o último dia de aula.

Referências

- [1] D. Caugherty and S. H. Rodger. NPDA: A tool for visualizing and simulating nondeterministic pushdown automata. In *DIMACS Workshop*, pages 365–377, March 1992.
- [2] M. K. de Oliveira. *Pensamento e Ação no Magistério: Vygotsky, Aprendizado e desenvolvimento, Um processo sócio-histórico*, chapter 4, pages 55–79. Editora Scipione, São Paulo, 2006.
- [3] S. C. Kleene. Representation of events in nerve nets and finite automata. *Automata Studies*, 1956.
- [4] W. S. McCulloch and W. H. P. Jr. A logical calculus of ideas immanent in nervous activity. *Bull. Math. BioPhys.*, 5:115–133, 1943.
- [5] M. O. Rabin and D. Scott. Finite automata and their decision problems. *IBM Journal of Research and Development archive*, 3 Issue 2:114–125, 1959.
- [6] S. H. Rodger, E. Wiebe, K. M. Lee, C. Morgan, K. Omar, and J. Su. Increasing engagement in automata theory with jflap. In *SIGCSE '09 Proceedings of the 40th ACM technical symposium on Computer science education*, pages 403–407. ACM New York, NY, USA, March 2009.
- [7] M. Sipser. *Introdução à Teoria da Computação*. Thompson Pioneira, Brasil, 2007.
- [8] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. In *Proc. London Math. Soc. Ser. 2* 42, pages 230–265, 1937.
- [9] A. M. Turing. Correction to: On computable numbers, with an application to the entscheidungsproblem. In *Proc. London Math. Soc. Ser. 2* 43, pages 544–546, 1938.

⁴O conteúdo da Seção 2 é coberto na primeira aula (aprox. 1:45h).

Uma Plataforma Ubíqua e Pervasiva para Distribuição de Conteúdo em Ambientes Educacionais

André Felipe Monteiro¹, Dalbert Matos Mascarenhas¹, Laura Silva de Assis¹

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ)
Rua do Imperador, 971 - Centro – Petrópolis – RJ – Brasil

Abstract. *In most educational environments, the process of communication with students is performed on ad hoc basis, with low standardization and a little interaction of side participation. However, a more dynamic communication is demanded by the young profile of most students. In this way, mobile devices are presented as an efficient and flexible channel to reach this public. In this scenario, we present a ubiquitous and pervasive platform to provide the distribution of educational content using a sensitive model context. Thus, the platform can satisfy the specific demands informed by the user, and also respond to the current context, sending personalized content according to predefined aspects such as location, profile, among others.*

Resumo. *Nos dias atuais, em grande parte dos ambientes educacionais, o processo de comunicação com os alunos é realizada de forma ad hoc, com baixa padronização e pouca interação entre as partes. Entretanto, uma comunicação mais dinâmica é demandada pelo perfil jovem da maior parte dos alunos. Desta forma, os dispositivos móveis apresentam-se como um canal eficiente e flexível de atingir esse público. Nesse cenário, apresentamos uma plataforma ubíqua e pervasiva para prover a distribuição de conteúdo educacional utilizando um modelo sensível a contexto. Assim, a plataforma pode atender às demandas específicas informadas pelo usuário, e também reagir ao contexto atual do mesmo, enviando conteúdos personalizados conforme aspectos pré-definidos como localização, perfil, dentre outros.*

1. Introdução

O processo de comunicação entre o público interno em ambientes educacionais (professores, técnicos administrativos, alunos, colaboradores, etc.) é realizado em geral de forma *ad hoc*, onde múltiplas plataformas de comunicação são utilizadas de acordo com a conveniência dos envolvidos. Desta forma, podemos observar o uso de e-mails, página pessoal de professores e instituições, grupos fechados em redes sociais, dentre outras alternativas, para implementar um canal de comunicação em ambientes educacionais.

Entretanto, com o avanço tecnológico e as múltiplas plataformas de informação disponíveis, o processo de comunicação principalmente entre professores e alunos necessita ser dinâmico e eficiente, de forma a complementar o conteúdo apresentado e discutido em sala de aula. Sob a perspectiva dos alunos, há um ambiente heterogêneo de comunicação, visto que cada professor pode optar por um canal de comunicação diferente. Este cenário pode acarretar em eventuais falhas de comunicação e desinteresse dos alunos em obter conteúdo educacional, como material didático complementar, listas de exercício, etc., conforme analisado em [Silva et al. 2011].

Para contornar este gargalo de comunicação em ambientes educacionais, apresentamos uma plataforma ubíqua e pervasiva para distribuição de conteúdo utilizando a rede de dados sem fio da instituição. Os alunos receberão conteúdo de seus professores, ou até mesmo comunicações institucionais de departamentos e setores, como biblioteca, laboratórios, etc., por meio de dispositivos móveis (*smart phones* e *tablets*) utilizando um modelo sensível a contexto. Portanto, a plataforma pode atender ao interesse em conteúdos específicos previamente informados pelo usuário, e também reagir ao contexto atual do mesmo, enviando conteúdos personalizados de acordo com diversos aspectos como localização, perfil de acesso, dentre outros.

Além de prover uma solução eficiente para comunicação interna, esta plataforma também se mostra um ambiente multidisciplinar, com aplicação de diversas áreas da Computação como Redes, Programação, Otimização, dentre outras, que necessitam se relacionar de forma integrada e consistente. Ou seja, um ambiente ideal para aplicação prática de conhecimento dos alunos do curso de Engenharia de Computação do CEFET/RJ *Campus* Petrópolis, que participam do projeto como usuários finais da plataforma, ou em conjunto com os autores deste trabalho na modelagem e implementação da plataforma, por meio de bolsas de projetos PIBIC e de Extensão.

Optamos por utilizar dispositivos móveis como único canal de comunicação do usuário final com a plataforma em virtude da melhor usabilidade e da maior familiaridade dos jovens com esta tecnologia, conforme apontado em [Charland and Leroux 2011]. Além disso, uma avaliação da nossa plataforma junto ao público alvo do projeto é apresentada na Seção 5. Os resultados indicam que há um grande interesse dos alunos em participar da especificação e implementação da plataforma, ratificando a característica interdisciplinar e integradora da mesma.

Este trabalho está organizado da seguinte forma: na Seção 2 apresentamos os trabalhos relacionados, comparando a abordagem utilizada e as funcionalidades da plataforma junto à literatura. Na Seção 3 analisamos a rede em malha sem fio utilizada como ambiente de comunicação da plataforma. Já na Seção 4, discutimos os detalhes e a implementação da aplicação ubíqua e pervasiva projetada para uso dos alunos em seus dispositivos móveis. Na Seção 5 apresentamos os resultados de uma pesquisa de interesse realizada com o público alvo da plataforma, e finalmente na Seção 6 descrevemos as conclusões e os apontamentos para trabalhos futuros.

2. Trabalhos relacionados

Diversos trabalhos presentes na literatura abordam a implementação de uma plataforma para compartilhamento de conteúdo educacional. Em [Orlandi and Isotani 2012] é proposto um sistema para correção *online* de listas de exercício enviadas pelos professores de disciplinas através dos dispositivos móveis de alunos. Já em [Rabello et al. 2012] é apresentada uma plataforma para colaboração entre alunos, que trocam informações e experiências sobre as atividades designadas pelos professores.

Nossa solução possui similaridades com as ferramentas mencionadas, como a utilização de dispositivos móveis como canal de comunicação com os alunos, e o envio de material educacional através da plataforma. Entretanto, nosso trabalho se diferencia pela segmentação e personalização do conteúdo a ser enviado para cada aluno de acordo com o contexto associado ao mesmo. Desta forma, priorizamos o envio de um conteúdo

mais seletivo e representativo sob o ponto de vista do usuário, evitando um processo de saturação de conteúdo e informações que poderia prejudicar a usabilidade dos usuários da plataforma.

Sob o ponto de vista do ambiente de aprendizado multidisciplinar viabilizado pelo processo de concepção e implementação da nossa plataforma, há trabalhos relevantes na literatura que analisam e propõe ambientes de aprendizagem e de troca de conteúdo educacional. Em geral, as disciplinas relacionadas à área de Programação são as mais frequentes na utilização de ambientes complementares para auxiliar o processo de aprendizagem. Em [Campos et al. 2015] e [Vahldick et al. 2015] um ambiente de jogos é utilizado para complementar o conteúdo destinado a disciplinas introdutórias de Programação. Já os trabalhos presentes em [Chaves et al. 2013] e [Santos et al. 2013] não abordam o aspecto da *gamificação*, mas também apresentam ambientes para apoiar o conteúdo de disciplinas de Programação.

Conforme pode ser observado, esses trabalhos não abrangem um escopo multidisciplinar, já que são destinados prioritariamente a disciplinas específicas da Computação. Por outro lado, nossa plataforma é abrangente no aspecto multidisciplinar, já que a mesma aborda de forma integrada várias áreas da Computação durante todo o processo de modelagem e implementação, conforme já discutido na Seção 1. Desta forma, podemos observar que a nossa plataforma é capaz de prover conteúdo e informações em ambientes educacionais baseados no contexto de cada usuário, além de se consolidar como um ambiente multidisciplinar para atuação de alunos e professores.

3. A rede de comunicação proposta

A utilização de redes em malha sem fio (*wireless mesh networks*) para prover acesso à internet em locais públicos ou privados mostra-se uma solução eficiente quando há necessidade de cobertura de uma área geograficamente grande. Isso ocorre em virtude da facilidade de implantação das redes em malha em áreas onde o uso de infraestrutura cabeada seria inviável.

Atualmente, muitas universidades utilizam as redes em malha sem fio para interligar prédios de seus *campi* e prover acesso à internet para seus alunos e colaboradores, além de a utilizarem como uma poderosa plataforma de teste e pesquisa. Algumas cidades também começam a utilizar esta solução para prover conexão entre pontos relevantes de seu território, como prédios públicos, museus, praças, viabilizando o acesso à internet para a comunidade em diversos pontos da cidade e a criação de cidades digitais, contribuindo assim para o avanço do processo de inclusão digital [Ahuja et al. 1993].

Uma das principais características das redes em malha está no uso de roteadores sem fio, geralmente instalados em locais fixos e com conexão à rede elétrica, que possuem um maior poder de processamento em relação aos roteadores móveis, que em geral apresentam suprimento de energia limitado. Uma arquitetura de uma rede em malha sem fio pode ser observada na Figura 1, onde cada nó representa um roteador e conseqüentemente um ponto de acesso distinto na rede.

Os roteadores presentes na rede em malha têm duas funções básicas: (i) encaminhar tráfego para os outros roteadores que fazem parte da rota de destino e (ii) receber e encaminhar tráfego para os nós clientes (dispositivos móveis de usuários). A auto-

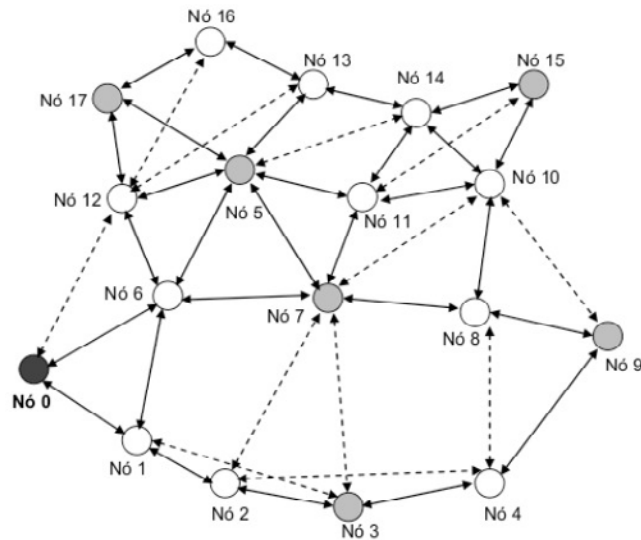


Figura 1. Arquitetura de uma rede em malha sem fio

organização e a auto-configuração devem ser características presentes nas redes em malha, de forma que a manutenção das conexões entre os roteadores seja automática. Tais características facilitam a inclusão de novos nós roteadores na rede, ampliando sua cobertura, além de as tornarem redes gerenciáveis e tolerantes a falhas, devido aos múltiplos caminhos possíveis criados entre nós roteadores e nós clientes [Mascarenhas et al. 2008].

Uma das vantagens de se utilizar as redes em malha é que elas oferecem um menor custo de infra-estrutura, dado que requerem menor número de pontos de acesso em comparação às redes cabeadas, proporcionando assim conectividade a uma área muito maior que as redes sem fio tradicionais. As redes em malha, ao contrário das redes ad-hoc nas quais se baseiam, não enfrentam restrições referentes à mobilidade e ao consumo de energia, pois normalmente os nós têm localização fixa e podem ser facilmente alimentados, tornando-os capazes de prover maior largura de banda e confiabilidade.

Uma infraestrutura de uma rede em malha sem fio permite que aplicações destinadas a dispositivos móveis utilizem a localização do ponto de acesso (roteador) para prover conteúdo específico destinado àquele local. Um possível exemplo é um usuário utilizando seu dispositivo móvel (smart phone, tablet, notebook) e conectado à rede em malha a partir do roteador *A* que está instalado ao lado da biblioteca do campus. Este usuário pode receber em seu dispositivo informações relacionadas à biblioteca, como horário de funcionamento, espaços ainda disponíveis para estudo, quantidade de computadores livres para uso, etc. Da mesma forma, caso este mesmo usuário se desloque para o prédio onde está situado o laboratório de redes, que é atendido na rede em malha pelo roteador *B*, as informações enviadas serão outras, como horário de aulas, recados, quantidade de máquinas disponíveis, etc.

Conhecendo o local de onde o usuário faz o acesso na rede em malha sem fio, é possível personalizar o comportamento de uma aplicação para que ela se adeque à localização do usuário. Este conceito é utilizado em larga escala para aplicações ubíquas e pervasivas. De uma forma geral, este tipo de aplicação utiliza o paradigma cliente-

servidor em sua arquitetura. Assim, o cliente é representado pelos dispositivos móveis dos usuários que enviarão requisições para os servidores de conteúdo, conforme indicado na Figura 2. Dado a possibilidade de identificar o local de origem da requisição, utilizando a identificação do roteador de acesso, o servidor que atender à requisição poderá enviar um conteúdo específico para o cliente.

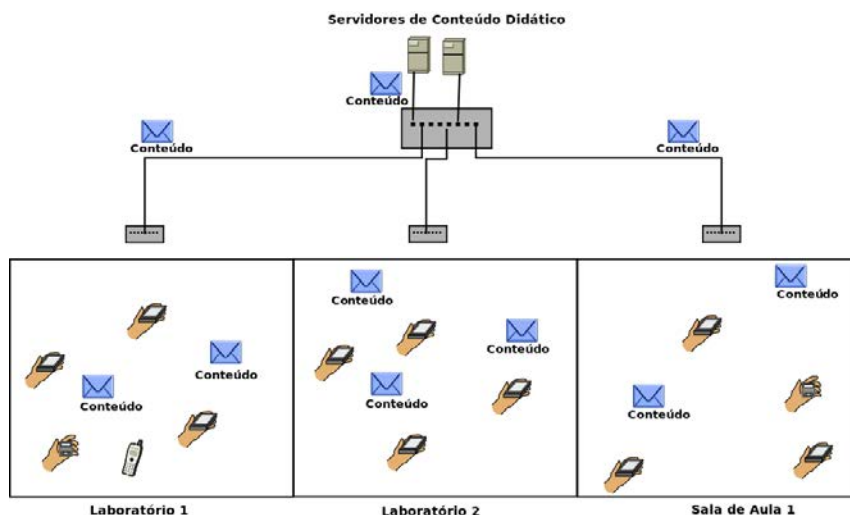


Figura 2. Modulo da arquitetura da rede

O projeto visa também restringir o impacto causado pela redundância de informações armazenadas e conseqüentemente trafegadas na rede de computadores. A Figura 3 exemplifica parte do escopo de armazenamento e distribuição do projeto. Para isto foram criadas estratégias de contenção para os recursos didáticos em redundância. Uma vez que os detalhes da rede de dados estão apresentados, descrevemos a seguir a aplicação cliente-servidor que irá distribuir os conteúdos educacionais aos usuários da plataforma.

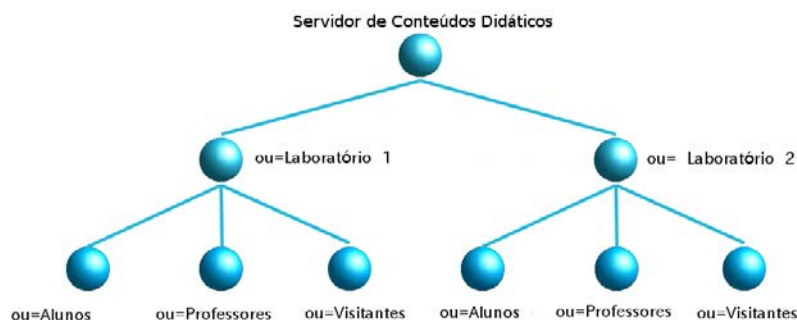


Figura 3. Escopo de Segmentação dos Conteúdos

4. A plataforma de conteúdo

A plataforma de conteúdo proposta utiliza uma arquitetura cliente-servidor, e se baseia em 3 componentes básicos. O primeiro consiste em uma Interface HTML onde é inserido

o conteúdo que será distribuído aos usuários. Posteriormente, os conteúdos inseridos em formato texto, (recados, exercícios, etc.) e seus meta-dados de controle são convertidos em um arquivo `Json` antes do envio para o servidor. Caso o conteúdo inserido esteja na forma de arquivos de imagem (figuras, arquivos PDF, etc.), os mesmos são enviados para o servidor sem necessidade de conversão.

Um servidor `Apache` é o segundo componente do sistema, sendo responsável por armazenar em um banco de dados SQL o conteúdo enviado pela `Interface HTML`, e disponibilizá-lo ao `App`. O terceiro componente é o `App` utilizado pelo aluno em seu dispositivo móvel para obter atualização de conteúdo através da rede em malha sem fio. Para esta ação, o `App` envia requisições HTTP ao servidor, que identifica a origem da requisição por meio do endereço IP dos pacotes recebidos. Essa identificação é verificada no servidor por meio de funções implementadas em PHP que manipulam uma tabela `hash` onde os endereços IP dos dispositivos móveis estão associados aos roteadores presentes na rede em malha. Ou seja, com base nos endereços IP das requisições podemos identificar o local onde o dispositivo móvel está presente e segmentar o conteúdo a ser enviado em resposta àquela requisição.

Utilizamos arquivos `Json` para realizar a comunicação entre os componentes do sistema em virtude de sua portabilidade entre plataformas, e pelo seu tamanho reduzido em relação a outras alternativas similares como o XML, por exemplo. Os trabalhos apresentados em [Nurseitov et al. 2009] e [Wang 2011] ratificam que o `Json` é mais eficiente que outras abordagens para a comunicação entre componentes cliente-servidor quando as plataformas Java e PHP são utilizadas. Em virtude de sua menor complexidade e tamanho, o arquivo `Json` requer menos processamento do cliente Java, provendo inclusive um menor consumo de energia do dispositivo móvel, que é um dos principais gargalos dos *smart phones* e *tablets* atuais. A Figura 4 exemplifica um arquivo `Json`.

```

{
  "lista de recados": [
    {
      "disciplina": "Aplicativos",
      "recados": [
        {
          "mensagem": "Prova remarcada para o dia 19/04",
          "data": "04/03/2016"
        },
        {
          "mensagem": "Hoje não teremos aula",
          "data": "29/02/2016"
        },
        {
          "mensagem": "Livros disponiveis na biblioteca",
          "data": "24/02/2016"
        },
        {
          "mensagem": "Aula no laboratório",
          "data": "16/02/2016"
        }
      ]
    }
  ]
}

```

Figura 4. Arquivo `Json` confeccionado pelo servidor de conteúdo

A título de exemplo, o arquivo `Json` apresentado na Figura 4 foi confeccionado pelo servidor de conteúdo com os eventos de recados de uma disciplina fictícia de “Aplicativos” e enviado para processamento no `App`, possuindo um tamanho de 476 By-

tes. Caso essas informações fossem enviadas pelo servidor em um arquivo imagem PDF para apresentação no *App*, esse arquivo teria o tamanho de 15,3 kB. Ou seja, o tamanho do arquivo *Json* é aproximadamente 30 vezes menor que o arquivo PDF.

Para a aplicação cliente implementamos um *App* na plataforma Android. Optamos por implementar um *App*, em detrimento a um cliente *web* a ser executado em um *browser* do dispositivo móvel, em virtude da melhor usabilidade e da maior familiaridade dos jovens com esta plataforma, conforme apontado em [Charland and Leroux 2011]. Utilizamos a linguagem Java para construir o *App*, já que a mesma oferece um suporte mais eficiente à API do Android, de acordo com [Gavalas and Economou 2011].

Na Figura 5 são apresentadas algumas telas do *App* implementado para a plataforma Android. A Figura 5(a) ilustra um *menu* do *App* onde o usuário pode verificar os recados, listas de exercícios e o material didático associados à localização do usuário dentro da rede em malha sem fio. Já a Figura 5(b) mostra os recados cadastrados da disciplina “Aplicativos” apresentados em ordenação cronológica no *App*, considerando que o usuário encontra-se localizado na sala de aula destinada à referida disciplina. Assim, caso o usuário se deslocasse para uma outra sala de aula que está associada a outro roteador na rede em malha, o conteúdo provido estaria relacionado à disciplina desta nova sala de aula, e não mais com a disciplina “Aplicativos”.

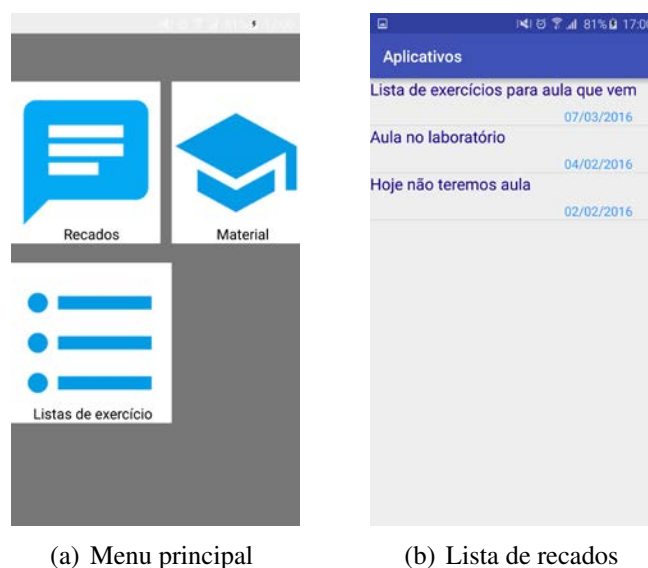


Figura 5. Exemplos de telas do *App*

5. Receptividade da plataforma proposta

Realizamos uma pesquisa com os alunos do curso de Engenharia de Computação a fim de obter a receptividade da plataforma proposta neste trabalho. Foi distribuído um questionário com três questões de múltipla escolha e tivemos um total de 58 questionários respondidos pelos alunos. O questionário fornecido contém as seguintes perguntas e suas respectivas alternativas de resposta:

1. Como você considera a comunicação entre o professor da disciplina e os alunos? Tal como envio de lista de exercícios, recados e avisos, material didático, etc.

- (a) Ótimo
 - (b) Bom
 - (c) Regular
 - (d) Ruim
 - (e) Péssimo
 - (f) Nenhuma das alternativas
2. Você considera que um aplicativo para dispositivos móveis onde fosse possível receber informações dos professores e da instituição de ensino por meio da rede sem fio interna do *campus* seria:
- (a) Imprescindível
 - (b) Muito útil
 - (c) Útil
 - (d) Pouco útil
 - (e) Sem utilidade
 - (f) Nenhuma das alternativas
3. Como você gostaria de colaborar no projeto de construção desta plataforma?
- (a) No processo de desenvolvimento do *App* (dispositivo móvel e servidor)
 - (b) No processo de configuração e instalação da rede em malha sem fio
 - (c) No processo de teste, utilizando *App* como um usuário e provendo o *feedback* de suas experiências
 - (d) Não tenho interesse em colaborar

Os resultados do questionário estão apresentados na Figura 6. A Figura 6(a) indica que há demanda para a plataforma proposta neste trabalho, visto que aproximadamente 42% dos alunos consideram a atual forma de comunicação professor-aluno como regular, ruim ou péssima. Além disso, a Figura 6(b) indica que aproximadamente 80% dos alunos entrevistados consideram que um *App* para prover conteúdo educacional complementar seria imprescindível, muito útil ou útil.

Por fim, apesar da implementação da plataforma ainda estar em caráter inicial, como apenas alguns protótipos funcionais em execução, os alunos se interessaram de forma significativa pelo projeto. Conforme apresentado na Figura 6(c), em torno de 30% dos entrevistados manifestaram interesse em participar de atividades relacionadas à especificação e implementação do *App* e dos servidores de conteúdo. Já para participar de tarefas associadas à configuração e instalação da rede em malha sem fio, 15% dos entrevistados manifestaram interesse.

Em outras palavras, 45% do público entrevistado indicou o desejo de participar diretamente dos processos de modelagem, especificação, implementação e instalação da plataforma proposta neste trabalho. Esse relevante quantitativo ratifica o caráter catalisador e integrador do projeto, abrindo ótimos horizontes para atividades de pesquisa e extensão relacionadas a áreas abordadas pela plataforma.

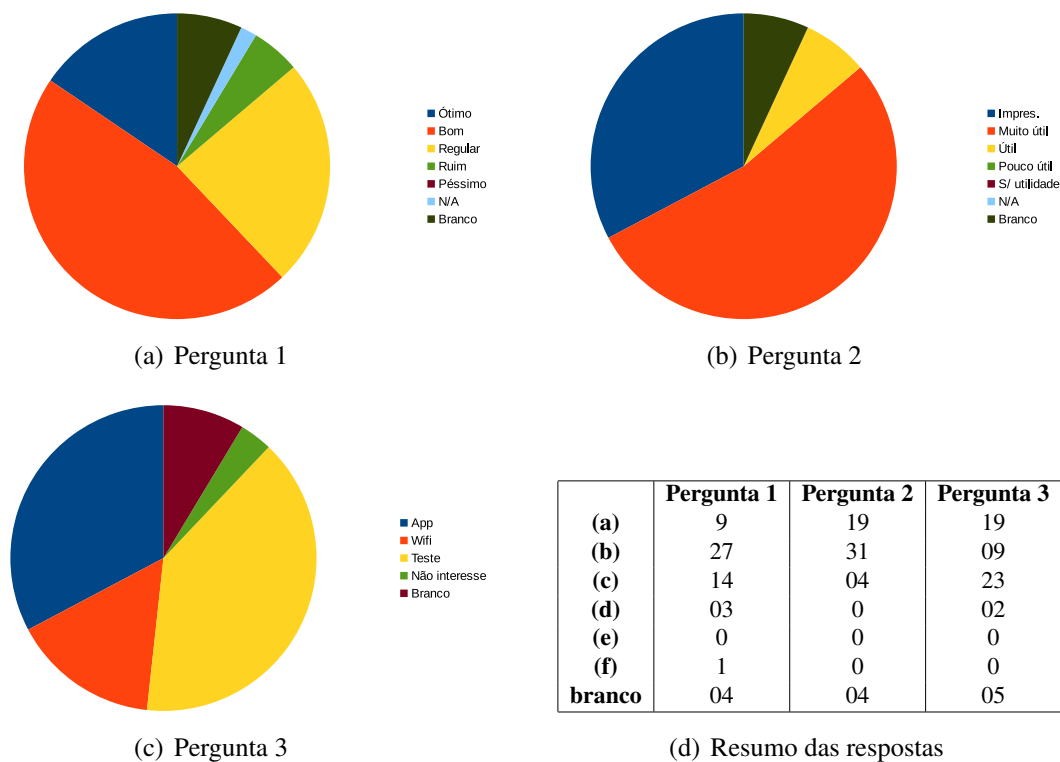


Figura 6. Resultados dos questionários de avaliação

6. Considerações finais

Este trabalho apresentou uma plataforma ubíqua e pervasiva para distribuição de conteúdo em ambientes educacionais. As primeiras implementações mostraram a viabilidade da plataforma, ratificando seu aspecto multidisciplinar. Cabe ressaltar o potencial catalizador da solução proposta neste trabalho, atraindo interesse de alunos e professores para colaboração em todas as fases do projeto, com ótimas perspectivas para futuras atividades de pesquisa e extensão.

Dentre as próximas atividades previstas no projeto, pretendemos viabilizar um melhor escoamento dos dados trafegados na plataforma utilizando processos de otimização. Esses processos deverão otimizar o fluxo de dados na rede, evitando gargalos e sobrecargas em pontos específicos. Também desejamos implementar um *cluster* de servidores virtuais de conteúdo utilizando ferramentas de virtualização, alcançando uma maior flexibilidade no gerenciamento de recursos dos servidores, e viabilizando outras características relevantes como tolerância a falhas, economia de energia, etc. Por fim, agradecemos ao CNPq e ao CEFET/RJ pelas bolsas de iniciação científica e extensão, utilizadas pelos alunos de graduação em Engenharia de Computação que participaram deste projeto.

Referências

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network flows: theory, algorithms, and applications*. Prentice Hall.

- Campos, A., Gardiman, R., and Madeira, C. (2015). Gamificação de uma disciplina introdutória de programação. *35º Congresso da Sociedade Brasileira de Computação (CSBC'15), 23º Workshop sobre Educação em Computação (WEI' 15)*.
- Charland, A. and Leroux, B. (2011). Mobile application development: web vs. native. *Communications of the ACM*, 54(5):49–53.
- Chaves, J. O. M., Castro, A. F., Lima, R. W., Lima, M. V. A., and Ferreira, K. H. (2013). Mojo: Uma ferramenta de auxílio à elaboração, submissão e correção de atividades em disciplinas de programação. *33º Congresso da Sociedade Brasileira de Computação (CSBC'13), 21º Workshop sobre Educação em Computação (WEI' 13)*.
- Gavalas, D. and Economou, D. (2011). Development platforms for mobile applications: Status and trends. *Software, IEEE*, 28(1):77–86.
- Mascarenhas, D., Rubinstein, M., and Sztajnberg, A. (2008). Uma nova métrica para protocolos de roteamento em redes em malha sem fio. *XXVI Simpósio Brasileiro de Redes de Computadores-SBrT*.
- Nurseitov, N., Paulson, M., Reynolds, R., and Izurieta, C. (2009). Comparison of json and xml data interchange formats: A case study. *Caine*, 2009:157–162.
- Orlandi, B. H. and Isotani, S. (2012). Uma ferramenta para distribuição de conteúdo educacional interativo em dispositivos móveis. *23º Simpósio Brasileiro de Informática na Educação*.
- Rabello, S., Barbosa, J. L., Oliveira, J., Wagner, A., Barbosa, D. N., and Bassani, P. B. (2012). Um modelo para colaboração em ambientes descentralizados de educação ubíqua. In *Anais do 23º Simpósio Brasileiro de Informática na Educação (CBIE'12)*, volume 23.
- Santos, C. J., Franco, M. E., and Corsini, H. C. (2013). Algbr: Uma nova ferramenta para apoio ao ensino/aprendizagem de lógica computacional por meio da construção e testes de algoritmos. *33º Congresso da Sociedade Brasileira de Computação (CSBC'13), 21º Workshop sobre Educação em Computação (WEI' 13)*.
- Silva, L. C. N., Neto, F. M. M., and Júnior, L. J. (2011). Mobile: Um ambiente multi-agente de aprendizagem móvel para apoiar a recomendação sensível ao contexto de objetos de aprendizagem. In *Anais do 22º Simpósio Brasileiro de Informática na Educação (CBIE'11), 17º Workshop de Informática na Escola (WIE'11)*, volume 1.
- Vahldick, A., Mendes, A. J., Marcelino, M. J., Hogenn, M., and Schoeffel, P. (2015). Testando a diversão em um jogo sério para o aprendizado introdutório de programação. *35º Congresso da Sociedade Brasileira de Computação (CSBC'15), 23º Workshop sobre Educação em Computação (WEI' 15)*.
- Wang, G. (2011). Improving data transmission in web applications via the translation between xml and json. In *Communications and Mobile Computing (CMC), 2011 Third International Conference on*, pages 182–185. IEEE.

Utilização do Mastery Learning em sala de aula: uma abordagem integrada aos Mundos Virtuais

Felipe Becker Nunes¹, Fabrício Herpich¹, Manuel Constantino Zunguze¹, José Valdeni de Lima¹

¹ Universidade Federal do Rio Grande do Sul (UFRGS) Av. Paulo Gama, 110, Porto Alegre, RS, Brasil - CEP 90040-060

{nunesfb, fabricio.herpich, manuelczunguze}@gmail.com,
valdeni@inf.ufrgs.br

Abstract. *This paper describes the general theory of Mastery Learning, with the proposed use jointly with the Virtual Worlds (OpenSim) to assess student performance. Students of an Accounting Sciences course in a discipline of general informatics learned topics related to the use of HTML. Preliminary results obtained with this proposal were positive and instigators, as well as the difficulties in the application of this research were clarified.*

Resumo. *Este artigo descreve de forma geral a teoria do Mastery Learning, tendo como proposta sua utilização de forma conjunta com os Mundos Virtuais (OpenSim) para averiguar o desempenho dos estudantes. Estudantes do curso de Ciências Contábeis em uma disciplina de informática geral aprenderam tópicos referentes ao uso de HTML. Os resultados preliminares obtidos com esta proposta foram positivos e instigadores, assim como foram esclarecidas as dificuldades existentes na aplicação desta pesquisa.*

1. Introdução

No cenário tradicional de aprendizagem, cada indivíduo que compõe um grupo de alunos está sujeito às políticas e métodos de aprendizagem propostos pela instituição de ensino e/ou docentes. Um ponto crucial neste tipo de arquitetura funcional está situado no fato de que cada aluno possui características e estilos individuais, visto que não é possível afirmar que cada grupo de alunos irá possuir do mesmo conjunto de particularidades e reagir de forma igualitária às atividades didáticas propostas.

Lin et al. (2013) entendem que para estreitar estas diferenças entre os alunos que atingem percentuais altos e baixos, se configura necessário auxiliar os sujeitos que obtiveram um desempenho inferior ao desejado, por meio de atividades de reforço que levem em consideração as características individuais de cada um. Uma linha de pensamento raciocinada por Bloom (1968) no artigo “Learning for Mastery”, converge para a crença de que todos alunos tem o potencial necessário para aprender conteúdos igualmente complexos, ressaltando que alguns poderão despende mais tempo para concluir tal procedimento.

É neste formato de sistema educacional, no qual o que importa é o conteúdo ser ensinado a todos, sendo fornecido mais tempo aos alunos que necessitarem, e, atividades de enriquecimento aos mais avançados, que se encaixa a teoria educacional denominada *Mastery Learning*. Desta forma, o modelo proposto no *Mastery Learning*

defende a crença de que todos os estudantes podem alcançar um nível elevado em seu desempenho acadêmico e suas diferenças de aprendizagem podem ser minimizadas.

Por se tratar de um modelo criado em 1968, aliado ao fato do fenômeno evolucionário ocorrido nas últimas décadas no que concerne à introdução das Tecnologias da Informação e Comunicação (TIC) no âmbito educacional, se torna factível a reflexão sobre o modo de aplicação do *Mastery Learning* atualmente em sala de aula e a distância utilizando as TIC. Do ponto de vista defendido por Purbohadi et al. (2013), esta teoria pode ser introduzida facilmente tanto no formato presencial quanto na educação a distância, caso o suporte tecnológico seja apropriado.

Diante deste contexto, se torna oportuno introduzir como solução os ambientes virtuais de aprendizagem (Moodle), assim como a proposta do uso dos Mundos Virtuais. Orgaz et al. (2012) explica que os Mundos Virtuais tem como objetivo disponibilizar espaços 3D onde o estudante pode transitar e vivenciar experiências em um ambiente altamente interativo. Conforme os aspectos dissertados anteriormente, que reforçam a interligação das TIC à abordagem *Mastery Learning* no ensino presencial e a distância, se cria a possibilidade de uma possível integração envolvendo a abordagem de Mundos Virtuais e o emprego de atividades educacionais neste ambiente, baseadas nos preceitos do ML.

Este artigo visa apresentar uma visão geral da aplicação da teoria *Mastery Learning* no contexto escolar, cuja integração aos Mundos Virtuais é proposta para a realização das atividades educacionais. Para isto, um estudo de caso foi desenvolvido em uma disciplina denominada “Computador e Sistemas de Informação”, com o uso do mundo virtual OpenSim e a implementação de diversas atividades educacionais neste ambiente de acordo com os preceitos do *Mastery Learning*, com o intuito de analisar a viabilidade e os resultados da aplicação desta proposta.

2. Trabalhos Relacionados

Ao longo de vários anos, pesquisas têm sido realizadas com base na teoria educacional *Mastery Learning*, as quais têm variado em termos da população (do ensino básico até ao universitário), e, em alguns casos usando tecnologias educacionais para suporte na realização das atividades. Na literatura podem ser encontrados centenas de trabalhos que utilizam de alguma forma a teoria do ML no decorrer dos últimos 60 anos, em que Leonard, Hollot, & Gerace (2008) identificaram pelo menos cinco *surveys* que foram escritos por diferentes pesquisadores neste período, sendo que quatro deles concluíram que grande parte dos trabalhos analisados teve diversos efeitos positivos nas escolas.

Marteleira (2010) explica que o modelo de Bloom teve um grande impacto durante as décadas de 60 e 70, nas quais houve um aumento significativo no número de estudos realizados, se estendendo até o início da década de 80. Entretanto, conforme esclarece Guskey (2010), devido às más interpretações e o surgimento de novas teorias, o ML caiu em esquecimento na década de 80. Corroborando tais asserções, López (2006) também considera que a exigência e algumas dificuldades na operacionalização do ML, como a exigência de um nível elevado de compromisso por parte do docente e a gestão em sala de aula, podem ser considerados como algumas das razões para que o modelo tenha sido progressivamente ignorado, apesar dos seus significativos efeitos na aprendizagem dos alunos.

As últimas décadas foram marcadas por uma ação de revalorização do *Mastery Learning*, motivado em grande parte pela evolução tecnológica alavancada pelas TIC e a sua acessibilidade nas escolas, o que cria maiores condições para a implantação deste tipo de teoria educacional. Com relação à aplicabilidade em diferentes áreas, similarmente aos Mundos Virtuais, esta abordagem contempla uma diversificada gama de extensões. Esta afirmação é corroborada por Guskey (2010) em seu trabalho, no qual ele afirma que o ML se mostrou efetivo em uma grande variedade de áreas, como matemática, ciências, língua estrangeira, treinamento e educação médica.

A análise efetuada acerca dos diferentes domínios de aplicação que o *Mastery Learning* tem sido referenciado e utilizado, demonstra a sua evolução e posterior decadência no decorrer dos anos, assim como a presente recuperação desta teoria educacional, principalmente devido ao uso das TIC, permite esclarecer seu cenário atual. Com base nisso, em relação à sua aplicação nesta proposta, o aporte tecnológico para aplicar esta teoria pode vir a ser preenchido com a utilização dos Mundos Virtuais, o qual possibilita a construção de uma diversificada quantidade de recursos educacionais e atividades didáticas, além da realização de atividades avaliativas, tornando esta proposta diferenciada das demais por utilizar tal ferramenta tecnológica, que dispõe de aspectos de imersividade e objetos tridimensionais.

3. Mastery Learning

O ML é um quadro pensado para planejar sequências de instrução com o objetivo de que todos os estudantes possam atingir um nível de desempenho razoável em um determinado conteúdo (MARTELEIRA, 2010). Se caracteriza necessário efetuar de forma detalhada a sua forma de operação e aplicação. Assim, são listados os princípios básicos que norteiam o ML (Kazu et al., 2005).

- Definir claramente os objetivos que representam os propósitos do curso;
- Dividir o currículo em unidades de aprendizagem relativamente pequenas, cada uma com seus próprios objetivos e avaliações;
- Identificar materiais e estratégias de aprendizagem; ensino, modelagem, prática formativa, avaliações, tutorias e reforço devem ser incluídas;
- Cada unidade é precedida por testes de diagnóstico breves (pré-teste) ou avaliações formativas;
- Os resultados dos testes formativos devem ser usados para fornecer instruções suplementares, ou atividades corretivas para ajudar o aluno a superar os problemas.

Portanto, no *Mastery Learning* os conteúdos são segmentados em pequenas unidades ordenadas por uma sequência lógica com objetivos específicos pré-determinados, que são seguidos até serem alcançados. Os alunos trabalham sozinhos ou em grupos dentro de cada bloco de conteúdo em uma série de etapas sequenciais, nas quais os estudantes devem demonstrar um alto nível de sucesso em testes, geralmente cerca de 80%, antes de avançar para novos conteúdos (MARTELEIRA, 2010).

Os professores procuram evitar a repetição desnecessária de conteúdos, avaliando regularmente os conhecimentos e habilidades. Para isto, um conjunto de

exames são aplicados ao decorrer de um módulo, no qual deverá ser aplicado um pré-teste para avaliar a situação inicial de cada indivíduo. Em seguida, um teste avaliativo deve ser aplicado para mensurar se o aluno detém o conhecimento necessário para que possa avançar no módulo ou atividades de reforço devem ser aplicadas.

Os alunos que não atingirem o grau de domínio requisitado, devem receber um *feedback* imediato com a aplicação de atividades de reforço, enquanto que os demais poderão ser encaixados em atividades de enriquecimento. Para aqueles que não atingem o nível exigido, são fornecidas aulas e/ou conteúdos suplementares, discussões em pequenos grupos, ou lição de casa para que eles possam alcançar o nível esperado. Após terem sido aplicadas as ações corretivas, um novo teste formativo deve ser realizado, a fim de, mensurar o grau de evolução do aluno.

Com relação às desvantagens apontadas, um dos principais pontos criticados no ML está relacionado ao custo de tempo que esta abordagem envolve ao ser aplicada. Docentes e gestores podem considerar inconveniente o tempo adicional que precisa ser fornecido aos estudantes que detêm maiores dificuldades para aprender, devido à grade de horários e formulação de atividades e materiais adicionais.

Entretanto, autores como Guskey (2010) e Marteleira (2010) defendem estas críticas esclarecendo que o tempo adicional necessário não é permanente, visto que Bloom ressaltou em sua proposta que inicialmente os alunos com dificuldades maiores podem necessitar de um tempo maior para o seu aprendizado, mas que no decorrer do curso isto provavelmente iria diminuir com a adequação à abordagem do ML e o aprendizado dos conceitos anteriores que servem como base para novos aprendizados.

Com relação ao maior número de avaliações realizadas nesta teoria, de acordo com Correia et al. (2015), na abordagem tradicional o professor, mesmo quando assessorado por monitores, não consegue fornecer um *feedback* adequado e rápido para cada uma das soluções propostas pelos alunos, pois a correção de cada solução demanda uma quantidade de tempo considerável. Portanto, se considera necessário que o docente planeje adequadamente as avaliações a serem realizadas e a forma de correção para que o esforço necessário não seja demasiado.

Marteleira (2010) explica que as TIC têm, no que se refere à operacionalização do *Mastery Learning*, um grande potencial de elaboração de atividades, facilitando assim ao professor o desenvolvimento das unidades de aprendizagem seguindo os princípios e a sequência subjacentes ao ML. Com o uso das TIC no *Mastery Learning*, atividades que devem ser desenvolvidas pelos alunos podem ser trabalhadas de diferentes maneiras para atender às suas preferências.

4. Mundos Virtuais

Griol et al. (2014) tem o entendimento de que os Mundos Virtuais podem ser considerados ambientes gráficos simulados por computador, no qual os seres humanos convivem com outros usuários através de seus *avatares*. Já Dalsasso et al. (2014) explica que a possibilidade de imersão em ambientes como Second Life e OpenSimulator tem sido utilizada para o desenvolvimento de ambientes colaborativos para uso educacional. Para a visualização e interação com o mundo virtual, é necessária a instalação de uma aplicação conhecida como *viewer*, no qual o usuário pode interagir e

realizar suas atividades. De acordo com Nunes et al. (2013), essa aplicação fornece suporte para a visualização, importação e exportação de objetos, e demais recursos relacionados ao mundo virtual.

Johnson et al. (2009) defendem a idéia de que os ambientes imersivos podem oferecer um novo tipo de abordagem, na qual os usuários passam de um estado considerado passivo e observatório para se tornarem sujeitos mais ativos dentro do mundo virtual. Essa linha de pensamento pode ser corroborada pelos diversos recursos presentes neste tipo de ambiente, tendo como exemplo o uso de *chat* via texto ou voz, navegar pelos cenários dispostos e interagir com os elementos presentes no MV, o que pode gerar um cenário propício para que este tipo de transformação de atitude ocorra.

A criação de ambientes imersivos voltados à educação necessita que diversos fatores sejam considerados, e.g., objetivos pedagógicos e estratégias de ensino bem definidas com base em teorias de aprendizagem, *design* amigável e objetos capazes de incentivar a interação e colaboração entre os usuários (HERPICH et al., 2014). A natureza imersiva deste tipo de ambiente incentiva os alunos (como *avatares*) a executarem atividades de aprendizagem que não estavam inicialmente previstas pelos professores, como interagir com outros alunos através de *chats*, procurar novos conteúdos multimídia ou utilizar novos artefatos físicos 3D para aprender sobre um determinado tópico (FERNÁNDEZ-GALLEGO et al., 2013).

A autonomia fornecida aos estudantes pode auxiliar no desenvolvimento de suas ações próprias e investigações acerca de novos conhecimentos calcados nos recursos presentes nestes ambientes 3D. É importante ressaltar as possibilidades inerentes ao desenvolvimento de atividades de cunho prático e simulações, que demonstrem experimentos que corriqueiramente são difíceis de serem visualizados em laboratórios reais, devido aos custos e/ou perigos intrínsecos à execução destes junto aos alunos.

5. Metodologia

A metodologia empregada nesta pesquisa pode ser dividida em diferentes etapas de aplicação, como são detalhadas a seguir. A primeira etapa consistiu em um levantamento teórico acerca dos tópicos Mastery Learning e Mundos Virtuais, assim como foi realizada uma análise do cenário geral e os direcionamentos das pesquisas que envolvem o uso desta teoria educacional.

Na segunda etapa foi definida a infraestrutura tecnológica utilizada para aporte desta pesquisa. O OpenSim 0.8.1 foi selecionado por ser gratuito, *open source* e com uma vasta documentação no meio acadêmico; para a visualização do mundo virtual foi selecionado o *viewer* Singularity, que possui recursos apropriados para atender às necessidades identificadas na construção do mundo virtual. Para a hospedagem e funcionamento das ferramentas selecionadas usadas, o Wamp Server foi escolhido por ser gratuito e conter os recursos necessários (PHP, MySQL e Apache). Todos os *softwares* foram instalados em um servidor no qual os clientes acessaram por meio do *viewer* instalado nos seus computadores.

Na terceira etapa foi realizado o detalhamento do experimento executado nesta pesquisa, cujo público alvo abrangeu alunos do curso de graduação em Ciências Contábeis da Universidade Federal do Rio Grande do Sul. Um total de 12 participantes

completaram todas etapas requeridas no experimento, cuja disciplina denominada “Computador e Sistemas de Informação” foi utilizada para a realização dos testes, sendo o experimento composto por dois encontros presenciais.

A formulação das atividades educacionais a serem aplicadas no Mundo Virtual abarcaram o tópico de ensino de recursos multimídias em HTML, que está presente em diferentes áreas de domínio. Desta forma, este tópico foi subdividido em duas unidades para a realização das atividades no mundo virtual, sendo o primeiro voltado para o ensino de inserção de vídeos em páginas *Web*, e o segundo compreende o estudo acerca da inserção de áudios em páginas *Web*.

O primeiro encontro presencial ficou caracterizado pela apresentação da proposta, realização do pré-teste e apresentação do mundo virtual aos estudantes para poderem se ambientar e conhecer o modo de funcionamento do ambiente. O segundo encontro presencial foi constituído pelas duas unidades, as quais possuíam atividades de reforço, tanto para os alunos que tiveram um desempenho satisfatório (enriquecimento do conteúdo), como para os alunos que não atingiram a maestria durante sua interação no mundo virtual (reforço do conteúdo), sendo abordada desta forma a ação de *feedback* por parte do professor junto aos alunos para suprir as dificuldades durante o período de realização das atividades, e, não somente ao final. Após o aluno percorrer as subunidades propostas e atingir a maestria, o mesmo deveria responder ao pós-teste (idêntico ao pré-teste) para avaliação do seu desempenho.

Para o estudo de caso aplicado neste primeiro experimento, um laboratório virtual foi desenvolvido dentro do mundo virtual focalizado no ensino de recursos multimídias em páginas HTML, no qual diferentes tipos de atividades educacionais foram criados, como a apresentação de *slides*, vídeos, textos, questões e a realização de atividades de cunho prático, como o uso de ferramentas *online* para a geração de páginas HTML em tempo real.

A quarta etapa abrangeu a revisão preliminar dos resultados obtidos, que foram mensurados tendo como base a teoria Mastery Learning, a qual explicita que para o estudante atingir o nível de maestria exigido, o mesmo deve obter um percentual mínimo, geralmente 80% na média final do tópico abordado. O desempenho final de cada estudante foi composto pelas notas obtidas ao longo de quatro avaliações realizadas durante este primeiro processo experimental conduzido, sendo os seguintes:

- Pré-teste = possui 9 questões sobre o ensino de recursos multimídias em HTML (vídeo e áudio) para avaliar o conhecimento inicial de cada aluno.
- Avaliação intermediária 01 = composta por 5 questões pertencentes ao subtópico referente à inserção de vídeos em páginas *Web*.
- Avaliação intermediária 02 = composta por 5 questões pertencentes ao subtópico referente à inserção de áudios em páginas *Web*.
- Pós-teste = este teste será o mesmo realizado no pré-teste, com o intuito de avaliar se ao final do curso o aluno atingiu o nível de maestria exigido na abordagem Mastery Learning.

Desta forma, o desempenho geral de cada estudante por meio das avaliações sugeridas na teoria do Mastery Learning pode ser obtido, e, analisado para ser usado na

composição das trajetórias de aprendizagem desenvolvidas por cada aluno durante as suas interações com o mundo virtual.

6. Estudo de Caso

O experimento desenvolvido contemplou avaliar o desempenho dos alunos que realizaram as atividades educacionais formuladas seguindo os preceitos definidos pelo Mastery Learning, dentro do mundo virtual. No primeiro encontro com a turma, o pré-teste composto por 9 questões relacionadas aos tópicos de inserção de vídeos e áudios em páginas HTML foi aplicado junto ao grupo de alunos presentes.

Como atividade inicial foi proposto utilizar o OpenSim com o intuito de proporcionar uma fluência destes estudantes com mundo virtual. Entretanto, a adaptabilidade por parte dos 12 alunos ocorreu naturalmente, em um processo que pôde ser considerado avançado, visto que pouquíssimas dúvidas surgiram sobre a forma de utilização do ambiente e do *avatar*. Findado o processo de ambientação dos alunos ao OpenSim, o docente realizou a descrição do experimento e atividades que seriam desempenhadas por eles, tomando o cuidado de esclarecer os objetivos de cada tópico e da proposta em geral, assim como funcionariam as avaliações no decorrer desta etapa.

Foram desenvolvidos no mundo virtual quatro laboratórios para atender os requisitos exigidos pela abordagem Mastery Learning, sendo os dois primeiros laboratórios, denominados respectivamente de “Laboratório 01” e “Laboratório 01 - Reforço”, o foco dos materiais estava centrado no ensino da unidade 01 (vídeos no HTML). Os demais laboratórios (“Laboratório 02” e “Laboratório 02 - Reforço”) tiveram seu foco situado no ensino da unidade 02 (áudios no HTML).

A figura 1 apresenta a sala de *slides*, na qual os usuários podiam efetuar a leitura dos conteúdos trabalhados no tópico. Já na figura 2 é demonstrada a interação dos usuários com a ferramenta “Online HTML Editor”, disponibiliza diretamente no mundo virtual e que fornecia em tempo real o resultado dos códigos criados.



Figura 1 e 2. Sala de *slides* e Sala com a ferramenta *online* de HTML

O início das atividades ocorreu no Laboratório 01 com a participação de todos os usuários, que adotaram uma postura ativa e independente, o que resultou em uma fácil utilização dos conteúdos dispostos no laboratório e a existência de um número pequeno de dúvidas durante a interação. O primeiro encontro contou com a realização de um teste intermediário composto por 5 questões relacionadas ao conteúdo discorrido, para avaliar o desempenho dos estudantes após realizarem as interações com os materiais no Mundo Virtual.

As avaliações foram corrigidas e aqueles alunos que não obtiveram uma nota superior à 80% foram guiados para o Laboratório 01 – Reforço, com o objetivo de realizar atividades complementares, e os alunos que atingiram a maestria estavam liberados para transitar neste laboratório para ler novos conteúdos e interagir, no intuito de agregar um enriquecimento maior a seu conhecimento. Para finalizar a primeira unidade, os alunos que não haviam atingido a média, realizaram o teste avaliativo novamente, com o objetivo de mensurar o seu desempenho após a realização das atividades de reforço.

Na segunda unidade, as atividades ocorreram no Laboratório 02, que contou com a participação ativa e caracterizada pela motivação dos alunos em utilizarem o ambiente novamente para aprender os conteúdos abordados. Nesta segunda unidade, o tema de inserção de áudios em HTML foi estudado pelos alunos no laboratório, que seguia exatamente o mesmo formato e estrutura utilizada no primeiro encontro.

Após terem interagido com os recursos disponibilizados nas cinco salas do laboratório, os alunos foram orientados a responderem a um teste intermediário composto por 5 questões relacionadas ao conteúdo discorrido, para avaliar o seu desempenho. Passada a aplicação do teste, as avaliações foram corrigidas e aqueles alunos que não obtiveram uma nota superior à 80% foram guiados para o Laboratório 02 – Reforço, com o objetivo de realizarem atividades complementares, e os alunos que atingiram o grau de maestria estavam liberados para transitar neste laboratório para ler novos conteúdos, no intuito de agregar um enriquecimento maior a seu conhecimento.

Os alunos que não haviam atingido a nota realizaram o teste avaliativo novamente, com o objetivo de mensurar o seu desempenho após a realização das atividades de reforço. Para finalizar o segundo encontro, o pós-teste foi respondido por eles com o objetivo de fornecer um comparativo de desempenho entre o conhecimento inicial que detinham antes de iniciar a aprendizagem das duas unidades, em relação aos novos conceitos adquiridos ao término dos dois encontros.

Com relação aos resultados obtidos nas avaliações, a figura 3 apresenta a média dos alunos em cada avaliação. Relacionando primeiramente os resultados do pré-teste com os pós-teste, notadamente observa-se uma evolução nas notas obtidas, sendo que no primeiro teste o intervalo ficou entre 45 e 89, enquanto que no último teste as notas variaram de 78 a 100.

No que diz respeito às notas dos testes intermediários, na unidade 01 a média resultante foi positiva (87,14), tanto no primeiro teste quanto na recuperação, da qual participaram três alunos. Dos 12 estudantes, três realizaram o reforço e obtiveram uma evolução significativa com uma nota de 93,3, praticamente atingindo a média geral dos que não estavam em recuperação. Na unidade 02, a média geral (75,71) ficou abaixo do esperado devido a concentração de notas no percentual de 80% e também das quatro notas abaixo, que ficaram entre 20 e 60, diminuindo a média geral. Os quatro alunos que necessitaram de reforço, assim como na unidade 01, demonstraram uma evolução e obtiveram uma média positiva (90).

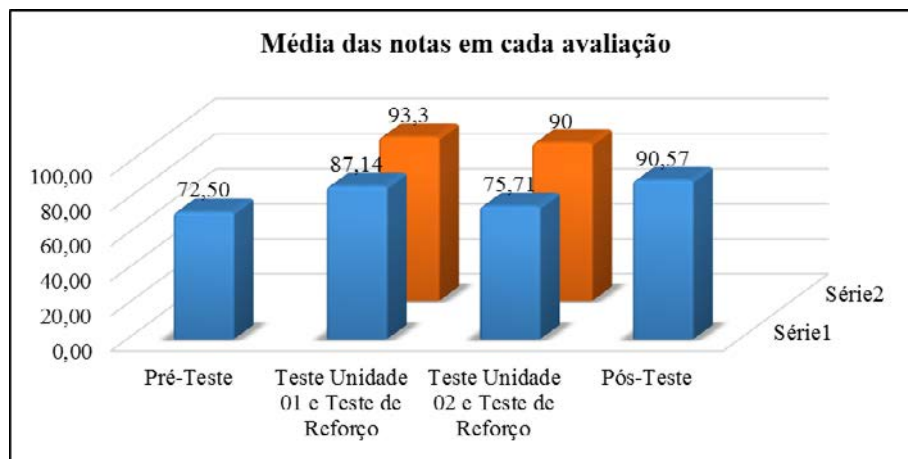


Figura 3. Resultados gerais das avaliações realizadas pelos participantes

Desta forma, é possível notar o progresso nos resultados obtidos nos testes realizados, demonstrando que poucos alunos necessitaram realizar a recuperação, e, aqueles que a efetuaram obtiveram notas iguais ou superiores à média da turma dos aprovados. Associado a isto, o desempenho crescente evidenciado no gráfico, tomando como comparativo o pré-teste e o pós-teste, provê indícios positivos de que a abordagem envolvendo Mundos Virtuais e Mastery Learning operou de forma satisfatória.

7. Considerações Finais

O uso do Mastery Learning no contexto escolar tem sido retomado no decorrer dos últimos anos, gerando resultados positivos no que concerne ao desempenho dos alunos. É importante ressaltar que a sua metodologia de aplicação requer um esforço maior por parte do professor, assim como demanda um tempo extra.

Nesta presente proposta, que está voltada para a aplicação do Mastery Learning utilizando os Mundos Virtuais como aporte tecnológico, por meio da análise dos resultados preliminares deste estudo foi possível identificar notar indícios positivos na melhoria de desempenho dos alunos. As atividades de reforço aplicadas e de especialização podem ser consideradas como um dos principais elementos que interferiram para que houvesse esta melhoria, assim como, as constantes avaliações que reproduzem para o docente o cenário geral e individual da turma.

Apesar das melhorias, o tempo disposto para a realização das atividades educacionais foi mais extenso do que o utilizado corriqueiramente em uma aula tradicional, o que pode ser considerado um fator negativo devido às exigências presentes em uma instituição de ensino quanto aos horários. Como abordagem futura, está a realização de testes mais robustos, levando em consideração um semestre ou mais, assim como a apresentação de uma alternativa à problemática do tempo, que está focada na realização das atividades de reforço e enriquecimento fora do período normal de aula.

Referências

BLOOM, B. S. Learning for Mastery. **Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints**, v. 1, n. 2, 1968.

CORREIA, A. L.; COSTA, D. V. B. DA; BARBOSA, A. DE A.; COSTA, E. DE B.

Uso de avaliação por pares em disciplinas introdutórias de programação. Workshop sobre Educação em Computação, p.1–10, 2015.

DALSASSO, P.; BOS, A. S.; BERNARDI, G.; MÜLLER, F. M.; CORDENONSI, A. Z. Resgatando A História da Computação através de um Museu Virtual 3D Interativo. Workshop sobre Educação em Computação, p.1–10, 2014.

FERNÁNDEZ-GALLEGU, B.; LAMA, M.; VIDAL, J. C.; MUCIENTES, M. Learning analytics framework for educational virtual worlds. **Procedia Computer Science**, v. 25, p. 443–447, 2013.

GRIOL, D.; MOLINA, J.; CALLEJAS, Z. An approach to develop intelligent learning environments by means of immersive virtual worlds. **Journal of Ambient Intelligence and Smart Environments**, v. 6, n. 2, p. 237–255, 2014.

GUSKEY, T. Lessons of Mastery Learning. **Educational Leadership**, v. 2, n. 68, p. 52–57, 2010.

HERPICH, F.; NUNES, F. B.; VOSS, G. B.; JARDIM, R. R.; MEDINA, R. D. Ambiente Virtual Imersivo para ensino em Redes de Computadores: uma proposta usando Agentes Inteligentes. 25º Simpósio Brasileiro de Informática na Educação (SBIE) - 3º Congresso Brasileiro de Informática na Educação (CBIE), p.65–69, 2014.

JOHNSON, C. M.; VORDERSTRASSE, A.; SHAW, R. Virtual Worlds in Health Care Higher Education. **Journal of Virtual Worlds Research**, v. 2, n. 2, p. 1–12, 2009.

KAZU, I. Y.; KAZU, H.; OZDEMIR, O. The Effects of Mastery Learning Model on the Success of the Students Who Attended “Usage of Basic Information Technologies” Course. **Educational Technology & Society**, v. 8, n. 4, p. 233–243, 2005.

LEONARD, W. J.; HOLLOT, C. V.; GERACE, W. J. Mastering Circuit Analysis: An innovative approach to a foundational sequence. 38th ASEE/IEEE Frontiers in Education Conference, p.1–6, 2008.

LIN, C. H.; LIU, E. Z.; CHEN, Y.; et al. Game-Based Remedial Instruction in Mastery Learning for Upper-Primary School Students. **Educational Technology & Society**, v. 16, n. 2, p. 271–281, 2013.

LÓPEZ, E. El mastery learning a la luz de la investigación educativa. **Revista Educación**, v. 340, p. 625–665, 2006.

MARTELEIRA, C. P. **Mastery Learning – a revalorização de um modelo de ensino-aprendizagem em cursos profissionais?**, Dissertação de Mestrado, Portugal, 2010.

NUNES, F. B.; VOSS, G. B.; HERPICH, F.; et al. Viewers para ambientes virtuais imersivos: uma análise comparativa teórico-prática. **RENOTE - Revista Novas Tecnologias na Educação**, v. 11, n. 1, p. 1–10, 2013.

ORGAZ, G. B.; R-MORENO, M. D.; CAMACHO, D.; BARRERO, D. F. Clustering avatars behaviours from virtual worlds interactions. Proceedings of the 4th International Workshop on Web Intelligence & Communities, 2012.

PURBOHADI, D.; NUGROHO, L.; SANTOSA, I.; KUMARA, A. GaMa Feedback Learning Model: Basic Concept and Design. **Journal of e-Learning and Knowledge Society**, v. 9, n. 3, p. 67–77, 2013.

WIM – XVI Workshop de Informática Médica

Apresentação

O XVI Workshop de Informática Médica (WIM 2016) tem como objetivo reunir pesquisadores, estudantes, professores, empresários e profissionais interessados em Computação aplicada à Saúde. O WIM é o evento anual da Comissão Especial de Computação Aplicada à Saúde (CE-CAS) da Sociedade Brasileira de Computação (SBC).

As atividades no WIM compreendem apresentação de trabalhos completos e trabalhos em andamento, além de palestras convidadas e mesas-redondas. Na edição de 2016, foram submetidos 76 artigos, dos quais 52 foram trabalhos completos e 24 trabalhos em andamento. Foram aceitos 14 trabalhos completos, com uma taxa de aceitação de 26,9%, e 8 trabalhos em andamento, com taxa de aceitação de 33%.

Os artigos abordam pesquisas originais e aplicações inovadoras nas áreas de Computação e de Sistemas de Informação aplicadas à Saúde, destacando-se os seguintes tópicos: ensino à distância em saúde (*e-learning*), telessaúde, uso de mídias sociais aplicadas à saúde, gestão de informação em saúde, segurança em sistemas de informação em saúde, recuperação da informação para aplicações clínicas, sistemas de informação em saúde, sistemas de apoio à decisão, realidade aumentada e realidade virtual aplicadas à saúde humana, reconhecimento de padrões, processamento e análise de sinais biológicos, processamento e análise de imagens médicas, visualização de dados biomédicos, modelos de otimização para planejamento e recomendação de terapias, modelos de diagnóstico, prognóstico ou propagação de doenças, redes de sensores e sistemas pervasivos em saúde, redes corporais sem fio (*Wireless Body Area Networks*), Internet das Coisas para saúde, sistemas computacionais de alto desempenho aplicados à saúde, web semântica e ontologias, acessibilidade e tecnologia web aplicadas à saúde humana e jogos sérios em saúde.

Débora Christina Muchaluat Saade (UFF)

Coordenadora Geral do WIM 2016

Lucas Ferrari de Oliveira (UFPR)

Coordenador do Comitê de Programa do WIM 2016

Comitê de Organização

Coordenação Geral

Débora Christina Muchaluat Saade
Universidade Federal Fluminense (UFF)

Coordenação do Comitê de Programa

Lucas Ferrari de Oliveira
Universidade Federal do Paraná (UFPR)

Comissão Especial de Computação Aplicada à Saúde (CE-CAS)

Marcia Ito - IBM Research Brazil (Coordenadora)
Sergio Miranda Freire – UERJ (Vice-coordenador)
Alessandra Alaniz Macedo – FFCLRP/USP
Débora Christina Muchaluat Saade – UFF
Denise Guliato – UFU
Luciana Tricai Cavalini – UERJ/UFF
Marco Antonio Gutierrez – INCOR/USP
Vera Werneck – UERJ
Lucas Ferrari de Oliveira - UFPR

Comitê de Programa

Adriano Andrade (UFU)
Alécio Binotto (IBM Research – Brazil)
Alessandra Macedo (FFCLRP/USP)
Alexandre Sztajnberg (UERJ)
Anderson Maciel (UFRGS)
André Backes (UFU)
Anselmo Paiva (UFMA)
Antonio Tadeu Azevedo Gomes (LNCC)
Aristófanes Silva (UFMA)
Arlindo da Conceição (ICT-UNIFESP)
Artur Ziviani (LNCC)
Carine Webber (UCS)
Celso Alberto Saibel Santos (UFES)
Claudia Moro (PUCPR)
Cristine Gusmão (UFPE)
Daniel Cotrim (FTEC)
Daniel Weingaertner (UFPR)
Daniela Trevisan (UFF)
David Menotti (UFOP)

Débora Christina Muchaluat-Saade (UFF)
Deborah Carvalho (PUC-PR)
Eduardo Albuquerque (UFG)
Eduardo Todt (UFPR)
Fatima Nunes (EACH-USP)
Flávio Seixas (UNESA)
Flávio Silva (UFU)
Gilson Giraldi (LNCC)
Gustavo Motta (UFPB)
Hélio Guardia (UFSCar)
Henrique Amaral-Silva (UT Health Science Center at Houston)
Humberto Razente (UFU)
Hyggo Almeida (UFCEG)
Iara Augustin (UFSM)
Ildeberto Rodello (FEARP – USP)
Iwens Sene Jr. (UFG)
Jones Albuquerque (UFRPE)
Jose Remo Brega (UNESP – Bauru)
Leandro Fernandes (UFF)
Leandro Oliveira (UFG)
Leandro Silva (UFAL)
Leila Silva (UFAL)
Lucas Ferrari de Oliveira (UFPR)
Luciano Romero Soares de Lima (Rede Sarah de Hospitais de Reabilitação)
Luis Eduardo Oliveira (UFPR)
Luiz Antonio Pereira Neves (UFPR)
Luiz Eduardo Silva (USP)
Marcelo Honda (UESC)
Marcelo Lobosco (UFJF)
Marcelo Oliveira (UFAL)
Marco Gubitoso (USP)
Maria Camila Nardini Barioni (UFU)
Maria Luiza Campos (UFRJ)
Mario Dantas (UFSC)
Marly Guimarães Costa (UFAM)
Mauricio Cunha Escarpinati (UFU)
Márcia Ito (IBM Research – Brazil)
Munir Gariba (PUCPR)
Paulo de Azevedo-Marques (USP)
Paulo Eduardo Ambrósio (UESC)
Rafael Pasquini (UFU)
Rômulo Souza (UERJ)
Renato Bulcão Neto (UFG)
Renato Tinos (USP)
Rita Lima (UFPE)
Rodrigo Veras (UFPI)
Romero Tori (Centro Universitário Senac / USP)
Rosa Maria Costa (UERJ)
Sílvio Cazella (Unisinos)

Selan dos Santos (UFRN)
Sergio Freire (UERJ)
Vinicius Tragante do Ó (University Medical Center Utrecht)
Vivian Motti (George Mason University)
Wanderley Lopes de Souza (UFSCar)
Wellington dos Santos (UFPE)

Painéis / Palestras

A Medicina do Futuro e sua Integração com a Informática

Valter Roesler (UFRGS)

A área da medicina vem incorporando inovações tecnológicas e computacionais, e essa tendência deve seguir no futuro. A palestra tem por objetivo apresentar conceitos envolvendo a integração de diferentes áreas para apoiar a medicina nos próximos anos. Serão abordados aspectos práticos de tecnologias emergentes na área, como: integração de smartphones e dispositivos sem fio na medicina; tecnologias disruptivas, como novas próteses robóticas, integração com o corpo humano para controle, regeneração de membros, 3D bio-printers, genética, exoesqueletos, entre outros; sistemas de tele-laudo, tele-monitoramento, tele-consulta, tele-consultoria, salas cirúrgicas inteligentes. Alguns dos tópicos possuem relação direta também com melhorias nos aspectos educacionais e de formação na área médica. Alguns desses pontos serão comentados durante os exemplos em vídeo apresentados.

Colaboração entre Equipes da Computação e Saúde em Projetos Interdisciplinares

Alexandre Sztajnberg (UERJ)

No final de 2014, o Núcleo de Atenção ao Idoso (NAI) da Universidade Aberta da Terceira Idade da UERJ (UnATI/UERJ) e o Laboratório de Ciência da Computação da UERJ (LCC/UERJ) iniciaram uma colaboração para o desenvolvimento do projeto “Sistema de Acompanhamento Remoto de Idosos, em Situação de Dependência, em Atendimento Ambulatorial e Domiciliar, e seus Cuidadores”. São duas equipes, a primeira multidisciplinar, composta de Pesquisadores e Médicos (com especialização em Geriatria), Residentes em Medicina, Enfermeiros, Psicólogos, Fonoaudiólogos e Fisioterapeutas, e a outra, composta por Pesquisadores e Alunos em Ciência da Computação. No âmbito do projeto, foi desenvolvido um sistema móvel para dispositivos *Android*. O sistema é composto de duas aplicações: uma dedicada ao profissional de saúde e otimizada para a utilização em um *tablet*, e a outra desenvolvida para *smartphone*, dedicada ao cuidador ou paciente. No final de 2015, um teste clínico foi iniciado, com o acompanhamento de um grupo de 30 cuidadores sob intervenção do sistema e outro grupo de controle de 30 cuidadores acompanhado com questionários por telefone. Neste tutorial, serão discutidos vários aspectos da colaboração entre as duas equipes e as lições aprendidas, desde a elaboração da proposta, passando pelas decisões técnicas, de logística e a interação entre os atores (os dois grupos de pesquisa, pacientes e cuidadores), até a execução do teste clínico e o teste de usabilidade.

Mesa Redonda

A Importância da Submissão de Projetos de Pesquisa em Saúde aos Comitês de Ética das Instituições

Paulo Ricardo Gazzola Zen (UFCSPA)

Márcia Ito (IBM Research Brasil)

Trabalhos aceitos

Trabalhos Completos

Elaboração de Próteses Auriculares Individualizadas por Meio de Manufatura Auxiliada por Computador

Emerson G. Moretto (USP)

Neide P. Coto (USP)

Roseli Lopes (USP)

Reinaldo Dias (USP)

Marcelo Zuffo (USP)

Usability Engineering of a Radiological Workstation for Mobile Devices: An Experience Report

Pedro van Rooij Costa (UFSC)

Aldo von Wangenheim (UFSC)

Christiane G. von Wangenheim (UFSC)

Andrei de Souza Inácio (UFSC)

Douglas D. J. de Macedo (UFSC)

Classification of Malignancy of Lung Nodules in CT Images Using Convolutional Neural Network

Giovanni L. F. da Silva (UFMA)

Aristófanés C. Silva (UFMA)

Anselmo C. de Paiva (UFMA)

Marcelo Gattass (PUC-Rio)

Redução de Falsos Positivos em Imagens de Mamografias Digitais Usando os Índices Phylogenetic Species Variability, Phylogenetic Species Richness e Múltiplos Classificadores

Laércio N. Mesquita (UFPI)

Antônio O. de C. Filho (UFPI)

Alcilene D. de Sousa (UFPI)

Patrícia M. L. de L. Drumond (UFPI)

Segmentação Automática de Imagens Térmicas das Mamas Utilizando Limiarização com Refinamento Adaptativo

Matheus F. O. Baffa (IFSEMG)

Deivison J. M. Cheloni (IFSEMG)

Lucas G. Lattari (IFSEMG)

Um Estudo Sobre Características Extraídas de Imagens Infravermelhas Para Detecção de Anormalidades de Mama

Lincoln F. Silva (UFF)
Roger Resmini (UFF)
Aristófanés C. Silva (UFMA)
Anselmo C. Paiva (UFMA)
Renato S. Bravo (UFF)
Alair Augusto S. M. D. Santos (UFF)
Débora C. Muchaluat-Saade (UFF)
Aura Conci (UFF)

Qualidade de Software do Maior Sistema de Teletoxicologia

João M. Alves (UFSC)
Christiane G. von Wangenheim (UFSC)
Alexandre Savaris (INCoD)
Aldo von Wangenheim (UFSC)

SMAI - Sistema Móvel de Assistência ao Idoso

Matheus Costa Stutzel (UERJ)
Michel Filippo (UERJ)
Luciana Branco da Motta (UERJ)
Alexandre Sztajnberg (UERJ)
André Brittes (UERJ)

Avaliação da Aceitação do Game Motion Rehab por Idosos com Sequelas de AVE

Ana Carolina Bertoletti De Marchi (UPF)
Rafael Rieder (UPF)
Eliane Lucia Colussi (UPF)
Magliani Reis Fiorin Martel (UPF)
Mateus Trombetta (UPF)
Nathália Pinto Cechetti (UPF)

Sistema Integrado de Análise e Visualização Epidemiológica Georreferenciada Online para Telemedicina

Andrei de Souza Inácio (UFSC)
Alexandre Savaris (INCoD)
João Marcus Alves (UFSC)
Aldo von Wangenheim (UFSC)

Relational Databases Versus Search Engines: A Performance Comparison for Storing and Querying DICOM Metadata

Alexandre Savaris (INCoD)
Gabriela Bussolo Colonetti (UFSC)
Rodrigo Rodrigues Pires de Mello (UFSC)
Aldo von Wangenheim (UFSC)

Vigilância Entomológica da Dengue, Zika e Chikunguya: Uma Solução Baseada em Redes Sociais e Dispositivos Móveis

Soeli T. Fiorini (PUC-Rio)
Leonardo Sousa (PUC-Rio)
Diego Cedrim (PUC-Rio)
Alessandro Garcia (PUC-Rio)
Débora Christina Muchaluat-Saade (UFF)
Igor Monteiro Moraes (UFF)
Leonardo Frajhof (UNIRIO)

Implementação da IoT para o Monitoramento das Variáveis Meteorológicas num AAL

Madalena Pereira da Silva (UFSC)
Débora Cabral Nazário (UDESC)
Mário A.R. Dantas (UFSC)
Alexandre Leopoldo Gonçalves (UFSC)
Alex R. Pinto (UFSC)
Guilherme Manerichi (UFSC)
Bruno Vanelli (UFSC)

Sistema de Informação para Perguntas e Respostas em Doenças Crônicas

Luciana F. Almansa (USP)
Alessandra A. Macedo (USP)

Trabalhos em Andamento

Abordagem Deep Learning para Classificação de Lesões Mamárias

Roberto M. P. Pereira (UFMA)
Caio E.F. Matos (UFMA)
João O.B. Diniz (UFMA)
Geraldo Braz Junior (UFMA)
João D. S. De Almeida (UFMA)
Aristófanês Corrêa Silva (UFMA)
Anselmo Cardoso de Paiva (UFMA)

Diagnóstico de Câncer de Mama Através de Vetores de Descritores Localmente Agregados

Ricardo Costa da Silva Marques (UFMA)
Geovane Menezes Ramos Neto (UFMA)
Geraldo Braz Júnior (UFMA)
João Dallyson de Sousa Almeida (UFMA)

Classificação de Doenças Intersticiais Pulmonares Difusas através de Tomografia Computadorizada de Alta-Resolução

Isadora Cardoso (UFAL)
Heitor Ramos (UFAL)
Eliana Almeida (UFAL)

Detecção de Anomalias em Termografia Dinâmica das Mamas

Jessica P. S. Cardoso (UFMA)
Aristófanés C. Silva (UFMA)
Anselmo C. Paiva (UFMA)
Stelmo M. B. Netto (UFMA)
Aura Conci (UFF)

**The use of Artificial Neural Network for Lipid and Glycaemic Profiles
Quantification Through Infrared Spectroscopy**

Henrique Hesse (UNISC)
Rejane Frozza (UNISC)
Valeriano Corbellini (UNISC)
Cézane Reuter (UNISC)
Miria Burgos (UNISC)

**Apreensibilidade e Qualidade da Informação: Bases de uma Avaliação Textual
Automática na Área da Saúde**

Asdrubal Falavigna (UCS)
Carine G. Webber (UCS)
Fernando A. Abel (UCS)
Marco A. Koff (UCS)
Maurício K. Santos (UCS)
Natália Lisboa (UCS)

**Identificação de Situações de Risco para Pacientes em Reabilitação Cardíaca
Explorando uma Arquitetura de Software na Internet das Coisas**

Douglas Scheunemann (UCPEL)
Adenauer Yamin (UCPEL)
João Lopes (UFRGS)
Cláudio Geyer (UFRGS)

Aplicativo Multiplataforma para Planejar Cirurgias de Estrabismo

Igor T. C. Furtado (UFMA)
João D. S. Almeida (UFMA)
Jorge A. Meireles-Teixeira (UFMA)
Geraldo B. Júnior (UFMA)

Elaboração de Próteses Auriculares Individualizadas por Meio de Manufatura Auxiliada por Computador

Emerson G. Moretto¹, Neide P. Coto², Roseli Lopes¹, Reinaldo Dias², Marcelo Zuffo¹

¹Laboratório de Sistemas Integráveis – Escola Politécnica – Universidade de São Paulo – Brazil

²Departamento de Cirurgia, Prótese e Traumatologia Maxilo Faciais – Faculdade de Odontologia – Universidade de São Paulo – Brazil

{emoretto, npcoto, roseli.lopes, rbdias, mkzuffo}@usp.br

Abstract. *The standard procedure of auricular prosthesis design involves the work of an expert using manual processes to recreate the complex shape of an ear. This is a slow and complex process that naturally lead to imperfections. This study presents a procedure that, from imaging exam uses computational techniques of reconstruction, mirroring and inversion of three-dimensional models for additive manufacturing (3D printing) of molds. The results show that the designed prosthesis using the process has a high level of realism, with 97.8% of similarity compared to healthy ear of the patient.*

Resumo. *O procedimento convencional de elaboração de próteses auriculares envolve o trabalho de um especialista que utiliza processos manuais para recriar a forma singular de uma orelha. Este é um processo lento, complexo e naturalmente pode ocasionar imperfeições. Este estudo apresenta um procedimento que, a partir de exames de imagem, utiliza técnicas computacionais de reconstrução, espelhamento e inversão de modelos tridimensionais para manufatura aditiva (impressão 3D) de moldes. Os resultados mostram que a prótese elaborada utilizando o processo exibe alto grau de realismo, apresentando 97,8% de similaridade dimensional comparada com a orelha sã do paciente.*

1. Introdução

A qualidade estética das próteses auriculares, sob os aspectos geométricos e de detalhamento superficial é fator importante para a reabilitação, pois propicia a dissimulação da perda e por consequência uma maior confiança ao paciente, aumento da autoestima e devolução menos traumática ao convívio da vida em sociedade proporcionando qualidade de vida [Watson e Muhanad 2014; Carvalho et al. 2013]. Atualmente as próteses auriculares são confeccionadas por meio de procedimentos que envolvem em sua maior parte processos manuais. O procedimento convencional de confecção de próteses auriculares utilizado pelo Departamento de Cirurgia, Prótese e Traumatologia Maxilo Faciais da Faculdade de Odontologia da Universidade de São Paulo (USP), e de acordo com Carvalho et al. (2013), inicia-se a partir da moldagem da orelha sã do paciente e da região acometida, elaboração de modelos em gesso, escultura

em cera ou modelina do formato da orelha são espelhada e muflagem para obtenção de um molde. A partir da prensagem de silicone no molde é obtida a prótese auricular.

Por envolver procedimentos manuais é naturalmente esperado a obtenção de imperfeições, principalmente pela etapa de escultura em cera a partir de observação visual, fotográfica e mesmo usando dispositivos que auxiliam na captação de dimensões. Pelo mesmo motivo, o tempo investido na elaboração de uma prótese é um fator a ser considerado, pois pode levar até mais que 2 semanas e por consequência ter um alto custo [Tanner e Mobley 2006].

Este trabalho apresenta um procedimento por meio da combinação de técnicas computacionais e de manufatura aditiva para a elaboração de próteses auriculares em menor tempo e com uma grande semelhança física se comparada a orelha saudável do paciente. As próteses produzidas por manufatura auxiliada por computador tendem a ser mais realistas e por consequência oferecem uma maior harmonia facial [Karatas 2011].

Recentemente, a manufatura aditiva tem sido explorada para auxiliar na produção de próteses com o objetivo de reduzir algumas desvantagens dos métodos convencionais, como apresentado por Karatas et al. (2011), Watson e Hatamleh (2014) e Subburaj et al. (2007).

Assim como nos trabalhos de Jiao et al. (2004), Karayazgan-Saracoglu et al. (2009) e Karatas et al. (2011) também são utilizadas imagens de Tomografia Computadorizada (TC) para a segmentação e reconstrução da região de interesse. A opção por imagens de TC se dá pela alta resolução espacial, permitindo que os detalhes geométricos do pavilhão auricular sejam propagados para o modelo virtual. Diferentemente dos trabalhos anteriormente citados, o procedimento apresentado neste estudo realiza os processos de espelhamento, filtragem de imagem, segmentação, reconstrução tridimensional e suavização de malhas de forma automática através de algoritmos de processamento de imagens e de computação gráfica. O resultado do processo é um molde para a elaboração da prótese.

Mardini et al. (2005) e Ciocca e Scotti (2004) utilizaram de scanners 3D para a captura da superfície da orelha saudável. Essa abordagem resulta em oclusões no modelo tridimensional inerentes ao processo de digitalização e geralmente demanda de um especialista de CAD para a modelagem de regiões oclusas e conversão da nuvem de pontos para modelo tridimensional.

Grande parte dos trabalhos correlatos observados [Karatas et al. (2011), Watson e Hatamleh (2014), Subburaj et al. (2007), Jiao et al. (2004), Karayazgan-Saracoglu et al. (2009), Karatas et al. (2011)] utilizam a manufatura aditiva para a fabricação do modelo da orelha sã. Esta abordagem ainda demanda do procedimento convencional de muflagem para obtenção do molde a ser utilizado para confecção da prótese.

Dada a característica investigativa e experimental envolvendo um procedimento clínico de elaboração de próteses, o presente estudo foi submetido para análise e aprovação do Comitê de Ética em Pesquisa da Faculdade de Odontologia da USP. A análise e aprovação do estudo está formalizada sob o parecer número 864.391 datado de 10 de novembro de 2014.

2. Método

As imagens obtidas da tomografia em formato DICOM são utilizadas como entrada em um software desenvolvido nesta pesquisa que tem como saída um arquivo do molde a ser enviado para uma impressora 3D. Inicialmente, as imagens de TC são pré-processadas pelo software com o objetivo filtrar ruídos provenientes do processo de imageamento e remoção de regiões indesejadas. Uma segmentação automática da região da orelha sã é realizada a partir da indicação do usuário.

Com as imagens segmentadas, é feita a reconstrução tridimensional da região inversa da segmentada, ou seja, a região que compreenderá um molde. Um algoritmo de suavização de malhas tridimensionais é aplicado com o objetivo de remover imperfeições causadas principalmente pela distância entre as camadas das imagens de tomografia.

O objeto tridimensional representando um molde é então convertido para o formato padrão de manufatura aditiva, o STL (*Stereolithography*), e é produzido pelo processo de Fabricação por Filamento Fundido (FFF). O objeto físico obtido é então submetido a um processo químico de suavização de superfície com solvente não tóxico de forma a eliminar a rugosidade proveniente do processo de fabricação.

Ao final do tratamento superficial, o molde possui aspecto superficial homogêneo e está apto para a elaboração da prótese através da prensagem de silicone pigmentado pelo protesiólogista.

Uma visão geral do processo apresentado por este trabalho está ilustrado na figura 1. As seções seguintes detalham cada uma das etapas.



Figura 1. Processo proposto

2.1. Filtragem

Para uma segmentação acurada do pavilhão auricular é preciso que as bordas da imagem estejam bem definidas e que os ruídos sejam suprimidos ou minimizados. A filtragem por Difusão Anisotrópica proposta por Perona e Malik (1990) possui um comportamento característico, suaviza regiões homogêneas na imagem e mantém o contraste em regiões de bordas. Por este comportamento, o filtro é amplamente utilizado em imagens médicas [Raut e Ruikar, 2013] e também é utilizado por este trabalho.

Além do filtro por Difusão Anisotrópica, um segundo filtro foi desenvolvido e aplicado com o propósito de classificar e remover artefatos na imagem que possam causar uma segmentação errada da região de interesse. A solução utilizada para esta segunda etapa de filtragem foi a criação de um classificador probabilístico *Naïve Bayes*, que determina se uma região pertence ou não a região da orelha.

O classificador probabilístico é aplicado para cada região encontrada na imagem e são rotuladas por uma classe K , obtêm-se a probabilidade condicional dada suas

características independentes: $p(K|c_0, c_1, \dots, c_n)$, para c_0, c_1, \dots, c_n sendo $c \in \{0,1\}$. As classes são divididas em regiões de interesse ($K=1$) e não interesse ($K=0$).

A tabela 1 é utilizada como base de conhecimento para o classificador, a qual foi definida através de mensuração e classificação manual sobre um conjunto de 132 imagens de tomografia contendo a região da orelha.

Tabela 1. Tabela de conhecimento utilizada no classificador.

<i>Classe (K)</i>	<i>Área</i>	<i>Escala Hounsfield</i>	<i>Distância da centroide</i>
0	$1 < x < 5000$	$-1000 < y < 3000$	*
0	$1 < x < 100$	$200 < y < 600$	$250 < d$
1	$4000 < x$	$200 < y < 600$	*
1	$x < 1500$	$100 < y < 800$	$d < 50$

A figura 2 ilustra um exemplo da aplicação da técnica onde em *a* a imagem original, *b* aplicação do cálculo de probabilidade e *c* resultado da supressão dos falsos candidatos.

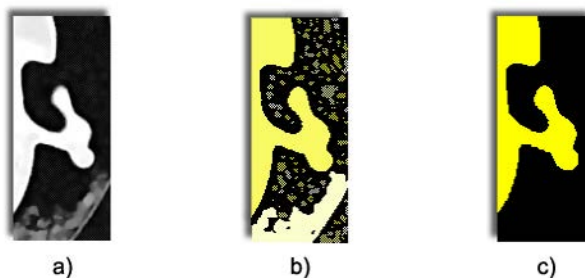


Figura 2. Aplicação da filtragem pelo classificador *Naïve Bayes*: Imagem original (a), aplicação do cálculo (b) e supressão dos falsos candidatos (c).

2.2. Segmentação

A segmentação de imagens por contornos ativos “*Snake*” [Kass et al. 1988] é um método comumente utilizado para segmentação em imagens médicas pela sua capacidade de generalização e tolerância a ruídos, o que é comum em imagens de TC. O método de contorno ativo por Fluxo de Vetor Gradiente, ou *Gradient Vector Flow* (GVF), proposto por Prince (1997) apresenta uma maior capacidade de contornar regiões côncavas.

A técnica GVF é utilizada pelo presente trabalho para segmentar de forma automática e iterativa a região da orelha. A técnica se baseia em um cálculo de uma energia externa, obtida a partir de uma difusão do vetor gradiente nas duas direções (x , y). O contorno ativo utiliza essa difusão para encontrar as regiões de bordas na imagem através de um cálculo de minimização de energia. Isso permite que a convergência do contorno seja melhor em regiões côncavas e que o contorno possa ser inicializado em qualquer região da imagem [Xu e Prince 1998]. Sua formulação original é apresentada pela equação 1:

$$f(x,y) = | \nabla [G_{\sigma}(x,y) * I(x,y)] | \quad (1)$$

Como observado na equação é utilizada uma convolução gaussiana com o objetivo de suprimir ruídos granulares na imagem. Na etapa anterior foi utilizada uma filtragem por difusão anisotrópica com o mesmo objetivo, e, portanto, esta convolução não é aplicada, resultando apenas na difusão dos vetores gradiente da imagem. A figura 3 apresenta em etapas o processo de convergência iterativo do contorno ativo por GVF.

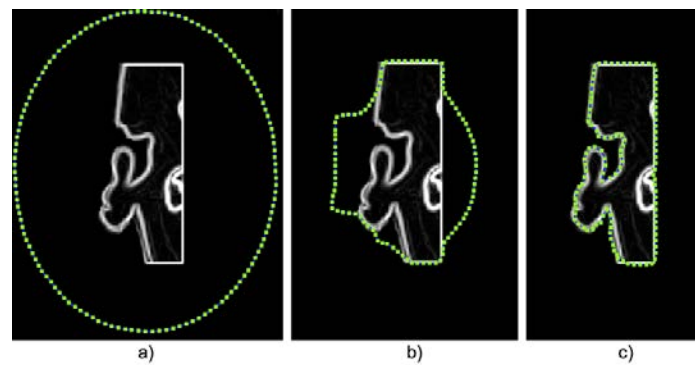


Figura 3. Contorno ativo por GVF: Contorno inicial arbitrário (a), contorno ativo na iteração $i=30$ (b) e convergência sobre a região da orelha (c).

2.3. Reconstrução Tridimensional

O resultado obtido da etapa de processamento de imagens é um conjunto de imagens binárias bidimensionais representando o pavilhão auricular em múltiplas camadas axiais. As camadas, ou secções horizontais, são então combinadas entre si através dos valores dos *pixels* entre as camadas de forma a encontrar um triângulo que intercepte os *pixels* de acordo com sua intensidade. A técnica utilizada para a reconstrução foi o *Marching Cubes*, proposto por Lorensen e Cline (1987), a qual é amplamente utilizada para este propósito.

A técnica foi aplicada no conjunto de imagens contendo a orelha segmentada com sua definição de *pixels* invertida. Dessa forma, o modelo tridimensional obtido é a representação do molde a ser fabricado.

2.4. Suavização de Malha Tridimensional

A técnica de reconstrução tridimensional utiliza-se de n imagens bidimensionais equidistantes para a composição do modelo. Essa estratégia resulta em determinados relevos na malha de triângulos relacionados à distância entre as imagens 2D, conforme ilustrado no item *a* da figura 4. Dessa forma, o modelo 3D do molde pode propagar esses artefatos indesejados para a prótese auricular.

De forma a suavizar esses relevos foi empregada uma técnica de suavização laplaciana dependente de escala. Esta aproximação do operador laplaciano utiliza pesos inversamente proporcionais à distância entre os vértices. Estes pesos são conhecidos como operador de Fujiwara [Fujiwara 1995] e pode ser definido pela equação 2:

$$w_{ij} = 1 / |e_{ij}| \quad (2)$$

O principal benefício desse operador é a preservação da distribuição das dimensões dos triângulos, minimizando potenciais deformações no modelo. Por ser um operador laplaciano combinado com uma característica local de cálculo de peso o

operador não é linear. A cada iteração o operador é recalculado conforme os vértices se movem e a iteração chega ao final quando não há mais aproximações a serem realizadas.

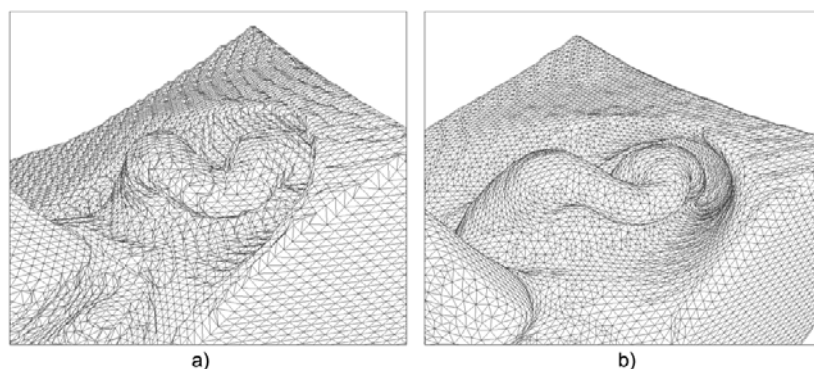


Figura 4. Modelo tridimensional do molde apresentando relevos (a) e resultado da suavização pelo operador laplaciano (b).

2.5. Manufatura Aditiva

Dentre as estratégias de Manufatura Aditiva atualmente disponíveis, a Fabricação de Filamento (FFF) foi escolhida para uso neste trabalho por ser bastante acessível e geralmente apresentar baixo custo de material (Guio e Leu, 2013).

Foram utilizadas 2 modelos de impressoras 3D para a fabricação dos moldes, a primeira 3D Clonner, Microbras[™], Brasil; a segunda RepRap Prusa i3, projeto de hardware aberto. Os moldes foram fabricados com sucesso em ambos os modelos.

2.6. Material

Dentre as opções de materiais para uso em FFF, o Acrilonitrila Butadieno Estireno (ABS) foi escolhido pelas suas características de acessibilidade e toxicidade. O ABS é amplamente utilizado por FFF, sendo um dos filamentos mais acessíveis e disponíveis no mercado. De acordo com Rutkowsky e Levint (1986) o ABS não apresenta níveis toxicológicos em temperatura ambiente e de derretimento.

2.7. Suavização de Superfície

A superfície dos objetos produzidos através da Manufatura Aditiva geralmente apresenta uma determinada rugosidade inerente ao processo de fabricação por camadas. De forma a inibir a propagação desta rugosidade para a prótese, um processo de suavização superficial é aplicado ao molde. A utilização de solventes reativos oferece um método flexível para alteração das características físicas de polímeros termoplásticos [Meijer et al. 1996]. Os solventes reativos atuam enfraquecendo as forças intermoleculares dos polímeros, permitindo que as moléculas se movimentem conforme incidência de outras forças, como a tensão superficial. Esta tensão permite que as moléculas se reposicionem de forma a eliminar regiões de micro vales.

A acetona (C_3H_6O) em concentração 100% foi adotada pelo processo de suavização de superfície por ser um solvente acessível e pertencer a Classe 3 - solventes sem perigo a saúde humana em níveis normalmente aceitos - do *Guidance for Industry*, do *Food and Drug Administration* (FDA) do Departamento de Saúde e Serviços Humanos dos EUA.

O molde fabricado pelo processo de FFF é exposto em uma câmara aquecida sem controle de pressão e sem contato direto com o solvente no estado líquido.

A proporção de solvente necessário para suavizar um molde foi determinada através de um cálculo de regressão linear sobre uma análise experimental. Foi obtida a proporção de 0,435 ml de acetona em concentração 100% para cada cm² de área de superfície do molde. A análise teve como objetivo encontrar a menor quantidade de solvente capaz de suavizar a peça cujo o índice de rugosidade - média aritmética das alturas da superfície (*Sa*) do *Geometric Product Specifications* (GPS) do ISO 25178 - fosse menor que 7,88 µm. A escolha do valor determinado para *Sa* foi baseada em Htsuki et al. (2013), onde foi realizado um estudo de sobre a rugosidade da pele humana e a região amostral com menor rugosidade obtida para *Sa* foi 7,88 µm. Para os testes de rugosidade foi utilizado um perfilômetro ótico digital CCI MP-HS Taylor Hobsontm, Leicester, Inglaterra. A figura 5 apresenta o resultado da aplicação do processo de suavização em um molde.

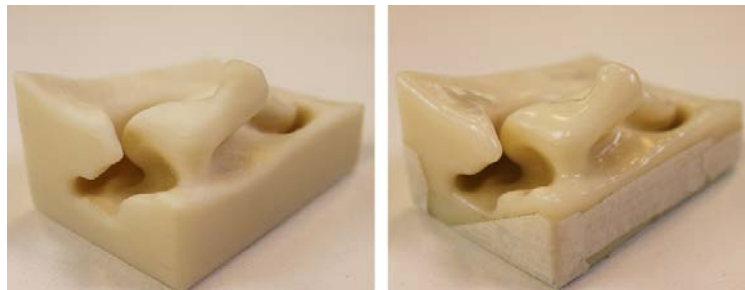


Figura 5. Molde obtido pela impressão 3d (esquerda) e após após a suavização (direita).

Ao final do processo de suavização, o molde está pronto para receber a prensagem de silicone para a concepção da prótese. O silicone utilizado foi o A2186, Factor IItm, Lakeside, EUA, cujo o tempo de cura é de 1 hora a 100°C.

3. Resultados

Todos os processos computacionais aqui apresentados foram implementados em um software de código aberto, desenvolvido em linguagem de programação *Java* e disponível em <http://bitbucket.org/emoretto/reear>. O software é multi-plataforma e foi testado em computadores com os sistemas operacionais Ubuntu 14.04 e Mac OS X 10.10. Em um computador Intel Core i7 com 8GB de memória RAM o tempo total de processamento do molde foi de 56 segundos. Um molde, dependendo de sua geometria e tamanho, leva entre 2 a 3 horas para ser fabricado utilizando uma impressora 3D com velocidade de 60mm/s.

A técnica apresentada foi aplicada em 2 casos clínicos de pacientes da Faculdade de Odontologia da USP. Ambos pacientes e/ou responsáveis assinaram um termo de consentimento para a participação experimental no projeto de pesquisa, permitindo que seus exames de tomografia computadorizada fossem utilizados.

A aplicação do procedimento foi realizada nos seguintes casos clínicos: ARF, 14 anos, sexo feminino, com perda total do pavilhão auricular direito por decorrência de acidente; KSS, 6 anos, sexo masculino, com má formação congênita no pavilhão auricular esquerdo. Em ambos os casos foram utilizadas imagens de TC com resolução de 512x512 *pixels* e espessura de camada de 1mm.

Os materiais utilizados para a fabricação de um molde são o ABS em forma de filamento para impressoras 3D ao custo de 20 USD por kg e acetona em solução 100% ao custo de 17,00 USD por litro, Sigma-Aldrichtm, Brasil.

O molde do caso clínico ARF utilizou 36 gramas de ABS para sua fabricação e 55 ml de acetona, ao custo total de materiais de 1,67 USD. O segundo caso, KSS, teve um custo de material de 1,51 USD.

O processo de suavização de superfície foi avaliado através de um perfilômetro ótico digital de forma a verificar a efetividade do processo. A figura 6 apresenta a renderização tridimensional de uma determinada região do molde antes e depois do processo. Foi possível observar que a amplitude dos vales e picos da rugosidade foi diminuída por toda a área.

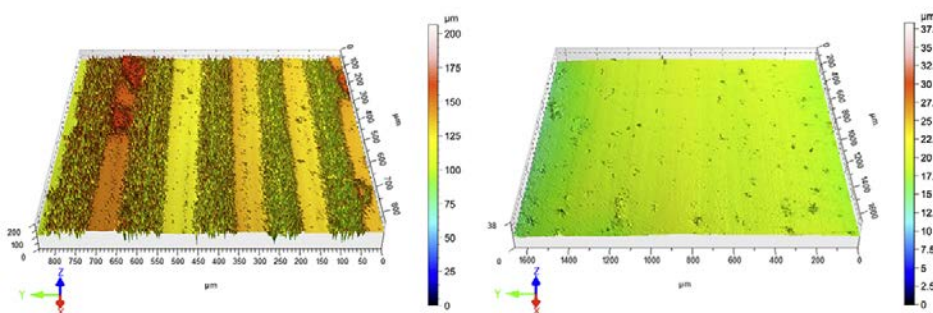


Figura 6. Superfície antes (esquerda) e após (direita) tratamento de suavização.

A avaliação de superfície seguiu os padrões de medidas do *Geometric Product Specifications* (GPS) da ISO 25178 e foram obtidos os valores para *Sa* de 18,4 μm e 0,79 μm para antes e depois da suavização respectivamente.

Foi também realizada uma análise comparativa das dimensões da prótese obtida (caso ARF) com a orelha do paciente utilizando o software Osirix, Pixmeotm, Suíça. As medidas seguiram os padrões morfológicos auriculares, conforme Subburaj et al. (2007). A tabela 2 apresenta o resultado desta análise.

Tabela 2. Medidas e diferenças dos padrões morfológicos auriculares entre as imagens da orelha saudável e a prótese obtida.

Região	TC (mm)	Prótese (mm)	Diferença(%)
Altura total	63.05	63.85	1.25
Largura total	23.50	23.90	1.67
Tragus a anti-hélice	21.50	22.32	3.66
Tragus a hélice	28.90	29.22	1.10
Largura do lóbulo	13.80	13.96	1.15
Altura do lóbulo	21.70	21.72	0.09
Projeção auricular	15.70	16.95	7.96
Inclinação auricular	6.60	6.55	-0.76

A diferença média das medidas observadas foi de 2,2%, o que representa um indicador de semelhança aceitável (97,8% de similaridade) sob o aspecto estético

considerando a geometria singular do pavilhão auricular. O item *b* da figura 7 apresenta a prótese obtida do caso ARF após a desmoldagem ilustrada pelo item *a* da mesma figura.

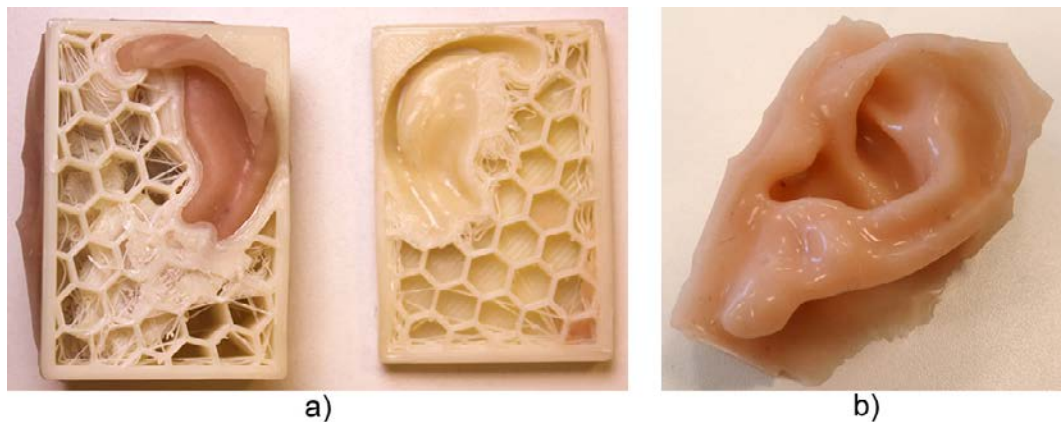


Figura 7. Molde com silicone injetado (a) e prótese obtida do caso ARF (b).

4. Discussão

O processo proposto apresentou um procedimento de elaboração de próteses alternativo ao convencional com potencial de expansão para demais próteses buco-maxilo-faciais. As próteses obtidas pelo processo apresentaram grande similaridade (aproximadamente 98%) em relação a orelha saudável do paciente, utilizando menos recurso manual e a baixo custo.

A necessidade de um exame de TC para a obtenção do modelo tridimensional do molde é uma desvantagem do procedimento apresentado pois expõe o paciente a uma dose de radiação onde eventualmente não seria necessário. Entretanto, em determinados casos clínicos principalmente decorrentes de dano causados por acidentes, o exame de tomografia geralmente é exigido para de análise do canal auditivo e/ou eventual fratura óssea.

A elaboração de uma prótese auricular pelo procedimento proposto possui um grande potencial de ser realizado em menor tempo em relação ao método convencional utilizado pelo Departamento de Cirurgia, Prótese e Traumatologia Maxilofaciais da Faculdade de Odontologia da USP, pois um molde pode ser fabricado em até 4 horas a partir do exame de TC.

O uso de impressão 3D na área médica vem sendo explorado amplamente nos últimos anos e deve ser um processo inerente a inovação na área, reduzindo custos e principalmente oferecendo um melhor tratamento ao paciente.

5. Agradecimentos

Os autores agradecem o Departamento de Cirurgia, Prótese e Traumatologia Maxilofaciais da Faculdade de Odontologia da USP pelo apoio clínico, científico e experimental ao projeto, ao Laboratório de Fenômenos de Superfície da Escola Politécnica da USP pelas análises de superfícies e o Departamento de Química Fundamental do Instituto de Química da USP pela revisão dos processos químicos envolvidos.

Referências

- Carvalho, J.C.M., Dias, R.B., Mattos, B.S., André, M. (2013), Reabilitação Protética Craniomaxilofacial, Ed. Santos, 1^o edição.
- Watson, J., Muhanad, M.H. (2014) “Complete integration of technology for improved reproduction of auricular prostheses”, *J Pros. Dentistry*, p. 430–436.
- Jiao, T., Zhang, F., Huang, X., Wang, C. (2004) “Design and fabrication of auricular prostheses by CAD/CAM system”, *Int J Prosthodont*, p. 460-463.
- Karayazgan-Saracoglu, B., Gunay, Y., Atay, A. (2009) “Fabrication of an auricular prosthesis using CT and rapid prototyping technique”, *J Cran. Surg*, p. 1169-1172.
- Mardini, M. Al., Ercoli, C., Graser, GN. (2005) “A technique to produce a mirror-image wax pattern of an ear using rapid prototyping technology”, *J Pros. Dent*, p. 195-198.
- Ciocca, L., Scotti, R. (2004) “CAD-CAM generated ear cast by means of a laser scanner and rapid prototyping machine”, *J Prosthet Dent.*, p. 591-595.
- Ciocca, L., Mingucci, R., Gassino, G., Scotti, R. (2007) “CAD/CAM ear model and virtual construction of the mold”, *J Prosthet Dent.*, p. 339-343.
- Perona, P., Malik, J. (1990) “Scale-space and edge detection using anisotropic diffusion”, *IEEE Trans PAMI.*, p. 629-639.
- Raut, V.N., Ruikar, S.D. (2013) “Image quality improvement in computed tomography using anisotropic diffusion”, *IEEE 2013 ICECCN*, p. 601-605.
- Kass, M., Witkin, A., Terzopoulos, D. (1988) “Snakes: Active contour models”. *Int J Comput Vis.*, p. 321-331.
- Prince, J.L. (1997) “Gradient vector flow: a new external force for snakes”, *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.*, p. 66-71.
- Xu, C., Prince, J.L. (1998) “Snakes, shapes, and gradient vector flow” *IEEE Trans Image Process.*, p. 359-369.
- Lorensen, W.E., Cline, H.E. (1987) “Marching cubes: A high resolution 3D surface construction algorithm”, *ACM SIGGRAPH Comput Graph.*, p. 163-169.
- Fujiwara, K. (1995) “Eigenvalues of Laplacians on a Closed Riemannian Manifold and Its Nets”, *Proc Am Math Soc.*, p. 123-131.
- Guo, N., Leu, M.C. (2013) “Additive manufacturing: Technology, applications and research needs”, *Front Mech Eng.*, p. 215-243.
- Rutkowski, J.V., Levint, B.C. (1986) “Acrylonitrile-Butadiene-Styrene Copolymers (ABS): Pyrolysis and Combustion Products and their Toxicity”, p. 93-105.
- Meijer, H., Venderbosch, R.W., Goossens, P., Lemstra, J. (1996) “Processing of Thermoplastic Polymers using Reactive Solvents”, *High Perform Polym*, p. 133-167.
- Htsuki, O., Akamaki, S., Ominaga, T. (2013) “Analysis of Skin Surface Roughness by Visual Assessment and Surface Measurement”, p. 94-101.
- Subburaj, K., Nair, C., Rajesh, S., Meshram, M., Ravi, B. (2007) “Rapid development of auricular prosthesis using CAD and rapid prototyping technologies”, *J. Oral Maxillo Surg.*, p. 938-943.

Usability Engineering of a Radiological Workstation for Mobile Devices: An Experience Report

Pedro van Rooij Costa¹, Aldo von Wangenheim¹, Christiane G. von Wangenheim¹,
Andrei de Souza Inácio¹, Douglas D. J. de Macedo²

¹Department of Informatics and Statistics (INE)

²Department of Information Science (CIN)

Universidade Federal de Santa Catarina (UFSC)

Caixa Postal 476 – 88.010-970 – Florianópolis – SC - Brazil

{pedro, andrei}@incod.ufsc.br, {aldo.vw, c.wangenheim, douglas.macedo}@ufsc.br

Abstract. *Medical imaging is commonly used to diagnose many emergent conditions as well as plan treatments. Digital images can be visualized on almost any computing device that has a screen. These days most devices used to visualize digital images are still desktops. Despite that, the world that surrounds us is evolving to a more mobile environment, therefore, mobile devices are increasingly being incorporated onto Picture Archiving and Communication Systems (PACS). In this paper we describe the usability engineering of a new medical image viewer for iPads as a Teleradiology tool for evaluating CTs, X-Rays, Ultrasounds, etc. First we gathered requirements, after we created a series of prototypes and finally the development of the actual application. The experience resulted in insights into how to develop an appropriate viewer for the specific target end user group. Our results suggest that emergency conditions commonly encountered on CT and MRI can be diagnosed using tablet computers with good agreement with dedicated PACS workstations, accordingly to ACR (American College of Radiology). Shortcomings in application design should be addressed if the potential of tablet computers for mobile Teleradiology is to be fully realized.*

1. Introduction

The use of mobile devices such as smartphones or tablets is revolutionizing the way we communicate and interact [Kim 2011]. This has also created opportunities for Teleradiology applications that allow an on-the-go interpretation of medical images as opposed to the use of an office-based workstation, optimizing referring physicians' access to images and reports from anywhere, anytime. With growing demands on radiology services worldwide, there are many scenarios in which such mobile teleradiology devices can be used to support rapid medical and imaging decision making, for instance, in an emergency situation when doctors are away from the hospital. Enabling rapid expert review from an experienced radiologist anywhere anytime facilitates interventions that [John 2012] allow safe and timely communication of important radiological findings to the clinician that may affect the patient's treatment, while at the same time, removing restrictions in physical location [Avitzur 2010]. Moreover, by removing geographical and temporal boundaries, these technologies may reduce economic disparities, lessen healthcare costs, and promote more personalized healthcare [Weiss 2010].

Thus, mobile radiological diagnosis support employing wireless communication devices can be a desirable extension of imaging services, particularly as an additional feature of an existing Picture Archive and Communication System (PACS), normally processing DICOM

images, which are a standard format associated with PACS. In this context, portable tablet computers, which have a large high-resolution display and touchscreen interface, may have potential to successfully allow the remote image review of radiological studies. Among current tablet computers, Apple's iPad is one of the most prominent hand-held devices, with diverse health care applications in the teleradiology field, also holding the largest market share of this kind of device and a high adoption rate among physicians.

Compared to medical workstations, tablets have several limitations: small screen size, inherent low resolution of display, low contrast ratio, poor connectivity, slow data transfer, security issues, and minimal inherent memory [Avitzur 2010]. Additionally, they most probably will not be used in an environment with controlled ambient lighting, as is expected from radiological workstations used in accordance to recommendations such as AAPM TG-18 or norms such as DIN V 6868-57 [Choudhri 2012]. Therefore, most mobile DICOM-viewing applications come with a disclaimer of "not for primary diagnosis", and there is a paucity of published literature documenting the accuracy of primary radiological diagnosis made on tablet computers [Wasserman 2010]. Nevertheless, there are several studies that indicate the potential usefulness of the iPad for radiological diagnosis of multi-image CT and MRI studies in after-hours conditions, where a diagnosing physician can view the images from home, another office, his car or any other place, using a tablet computer and then send his diagnosis and treatment instructions. Recently, the US Food and Drug Administration has also issued clearance for the use of a specific software application (Mobile MIM, [Chen 2009]) on the iPhone and iPad for viewing and diagnosis of CT, MR imaging, and nuclear medicine images where access to primary-class displays is an issue, even though it accesses DICOM images it has a drawback of needing to have specific DICOM servers also provided by Mobile MIM.

Another issue in this context is the required degree of software quality, as these applications often fall short on usability. Medical software manufacturers have been providing applications with Graphic User Interfaces (GUI) that are not necessarily compliant with standard usability criteria [Drnasin 2010]. On the other hand, there are several DICOM viewers available to download and purchase that seem to be accepted such as OsirixMD, MobileMIM, VueMe and ResolutionMD.

The need to ensure that medical applications and devices are designed, manufactured and also used in a way that does not lead to unnecessary risks to patients and users is not new. We need to observe specific recommendations from devices until the conditions about the environment in radiological images [Choudhri 2012]. As the design for patient safety initiatives play an increasing role in public health policy, it becomes necessary to consider ergonomic design in the essential requirements.

In this way, usability is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [ISO 2002]. In essence, a system with good usability is easy to use and effective. It is intuitive, forgiving of mistakes and allows one to perform necessary tasks quickly, efficiently and with a minimum of mental effort. Usability represents an important, although often overlooked, factor that impacts the adoption and meaningful use of PACS and mobile DICOM viewers. In mobile applications, usability is really important because it is a different environment than traditional desktop systems [Yen 2011]. If an application cannot be correctly usable, the accomplishment of simple tasks becomes more difficult and error-prone.

Being mobile devices, tablets and smartphones can be used in different places and contexts, e.g. on the move anywhere and anytime. Therefore, many common desktop design patterns and usability principles do not apply in a mobile environment [Siau 2004]. Small screens and limited input methods create extra challenges for interaction design. For example, small screen size implies less visible content and options that can make almost all interactions with such devices harder [Yen 2011]. It is therefore important how health information is displayed for stakeholders without losing important parts or how to design navigation so that people will not have difficulties to access the required information. Navigation mechanisms

(including searching and browsing) for applications need to be designed for the small screen, and direct data entry carried out via touch and multi-touch interactions. For example, it is important to provide a logical path for the user to follow, making a navigation control consistent and obvious. In addition, in the mobile domain, users will find it difficult if an application relies heavily on text input, so it is important to use alternatives, for instance table views or dropdown menus in order to minimize the effort required for user input [Choi 2012].

In this paper we describe the development of a DICOM viewer application for the iPad, created to be used in the context of a public statewide telemedicine network, the STT/SC in Southern Brazil [STT 2016]. This application was explicitly developed accordingly to usability guidelines and through context analysis and prototypes made before the creation of the actual application. After the application was developed, we performed usability tests with potential users of the visualizer in order to evaluate the application, in order to validate applicability of this DICOM visualizer for the iPad as an auxiliary device in the process of performing and issuing medical reports in a Teleradiology environment. In this study we analyzed if the application is well developed from the users' point of view and if all the required tools are present and easy to use, following all common guidelines of usability for mobile devices and medical image viewers.

The application development started with studies of guidelines and usability for mobile applications, especially image visualizers. Following this study we gathered user requirements interviewing users of our Telemedicine network, and then started to prototype the interfaces using the knowledge gathered through the study and the requirements. After that, the actual application was developed following usability and software engineering. With the application ready we did evaluations with potential users and analyzed all the data and results gathered.

2. Related Work

So far few specific usability guidelines for mobile medical apps have been proposed. Nielsen (1994), on the other side, proposes a number of generic guidelines for increasing usability of mobile devices as well as touch-screen application usability. So far, there are basically design guidelines created by platform providers and/or handset manufacturers, e.g. iOS design guidelines, UI guidelines for Android, and UI guidelines for Windows Mobile. There are also evaluations of Nielsen heuristics (1994) by Chan (2011) and Yen (2011).

Especially research regarding the usability of a medical image viewer is very sparse. One exception is the work by Chen (2009), which, however, does not focus on mobile devices. On the other hand, Cannella (2009) only focus on technical issues regarding the automatic creation of interfaces.

Therefore, we present in this article an experience report on the usability engineering of the development of a mobile DICOM viewer application for the iPad for the visualization of radiological images in the context of a large-scale Telemedicine network. The application is being developed and integrated into the STT/SC - Santa Catarina State Integrated Telemedicine and Telehealth System – a public asynchronous web-based telemedicine network in the state of Santa Catarina/Brazil [STT 2016], that performs store-and-forward telemedicine and findings reports delivery in the fields of: clinical laboratory analysis, radiology (MR, US, CT, SPECT, bone densitometry), endoscopy and colonoscopy, and ECG and tele dermatology, besides asynchronous emergency assessment, mainly on trauma cases. Today, the network interconnects more than 400 hospitals and primary health care facilities in 295 cities, performing, on average, more than 70,000 examinations/month.

3. Material and Methods

We performed this case study through a systematic development of a DICOM viewer to answer the questions raised previously. The development stages are shown in Figure 1. Figure 1 shows all the stages of the application development; firstly there is a literature analysis, after, prototyping of user interface, application development and finally the evaluation of the

application and interfaces created.

The research was executed as a cooperation between the GQS – Software Quality Group and the LabTelemed – Telemedicine Laboratory of the Brazilian Institute for Digital Convergence (INCoD) at the Federal University of Santa Catarina and the Santa Catarina State Health Department – SES/SC, involving researchers from the area of usability engineering, design and software engineering, as well as from the area of telemedicine.

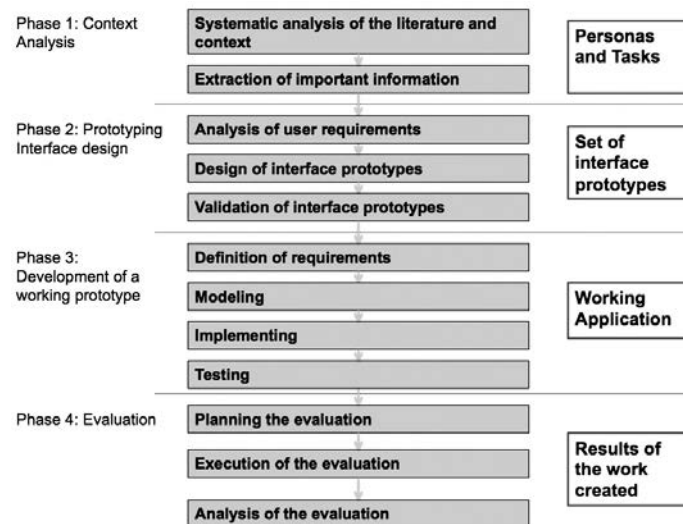


Figure.1. Stages of the application development

In order to achieve high usability, interface design was a priority from the very beginning of the project. Design input, design verification and design validation were conducted throughout the design and development of the medical application. Design input calls for manufacturers to establish and maintain procedures that ensure design requirements are appropriate and address the intended use of the device, user needs and patient needs.

A specific iterative design process model is presented by ISO 9241-210. This "human-centered design" approach focuses on the needs of the future user in order to develop a usable human-machine-interface. Principles are, among others, an explicit understanding and integration of users, their tasks and realistic environment, iterative processes and multidisciplinary skills and perspectives within the design teams.

3.1 Phases

In this phase (Phase 1 – Context Analysis) General requirements were obtained from the ACR standard for teleradiology. It states that a radiological viewer must be capable, among other features, to select image sequences, accurately associate the patient and study, adjust window and level, pan and zoom image, rotate or flip the images provided, calculate and display accurate linear measurements and pixel value determinations in appropriate values for the modality (e.g., Hounsfield units for CT images), and display prior image compression ratio, processing or cropping.

In order to elicit characteristics, preferences and needs of the target users of the application, we performed a series of interviews with radiologists and domain experts – all professionals working for the public telemedicine system in Santa Catarina/Brazil and familiar with the concept of Teleradiology. Based on the analysis of the collected data, the users have been characterized and illustrated through the definition of personas. Based on the analysis results, the following usability requirements have been defined:

- UR1. All commonly used functions (measuring, zoom, windowing and leveling) have to be accessible by at most 2 clicks;
- UR2. The use of two fingers simultaneously should be possible;
- UR3. Zoom should be done using the pincer movement;
- UR4. The task “visualize and interpret exams” should be completed in at least 70% cases;
- UR5. The task “visualize and interpret exams” should take less than 10 minutes;
- UR6. The degree of user satisfaction should be at least 60 points using the SUS (Unified Health System in Portuguese language) - satisfaction questionnaire.

During the next phase (Phase 2 – Interface Design Prototyping), we produced design solutions, trying to meet user requirements as well as possible. This is done by iteratively prototyping designs starting with sketching low-fi paper prototypes. These first prototypes have been evaluated through cognitive walkthroughs. Based on the feedback, we improved the design, e.g. reducing the size of the inferior toolbar and re-positioning of the return button. For this prototyping iteration we used a wireframe tool (Figure 2). Based on the feedback, we included an additional logon screen for the login. Another issue we observed was the need for a more direct and quick visualization of images rather than first presenting the patient and examination data, which the physician may have already accessed beforehand. Therefore, we eliminated in the sequence of screens the screen that visualized patient and examination data only. Instead, after selecting a patient and examination, the system directly starts to visualize images – allowing the physician, if necessary, to push down an overlay with the patient and examination information.

During the next phase (Phase 3 – Development of a Prototype), a working prototype for studying the interaction with end users was created. Here, we concentrated on the Hi-Fi prototype and adapted the choice of software tools to technological requirements. We initially performed a requirement analysis, covering the identification of functional and non-functional requirements, including the usability requirements as identified in the context analysis. Using this analysis as a starting point, the modeling of the MVC (Model-View-Controller) architecture was performed. Immediately after this, a class diagram was created with the objective to define the application. After this first step, the app interfaces were created and integrated with the model using the previously created controllers as interfaces between them. All of these functionalities were created in Mac OS using XCode as the app development program.

In next phase (Phase 4) his evaluation had the objective to validate the usability of the user interface we designed and developed and the functionalities implemented from the viewpoint of the doctors who perform Teleradiology examination reports for the STT/SC system. The objectives of the evaluation are:

- Analyze the effectiveness, efficiency and satisfaction of the interfaces and functionalities;
- Analyze if the functionalities developed are enough;
- Validate the application as an aid on the process of creating an examination report;
- Identify the strong and weak points of the design created with the goal of improving it and
- Verify if the requirements were met.

To simulate the situation during exam analysis, the participants had as much time as they desired to review the data set using the working prototype. The test was performed on an iPad 2, running iOS 6.1.4 as the operating system. Observers were allowed to use all functions of the DICOM viewer on both systems, including zooming, panning, and control of radiological window width and level.

At the start of each session, the study procedures were explained to each participant and

written assent was obtained. The participants were then asked to fill out a short demographic questionnaire. Each participant performed the defined task on images related with their expertise. Participants were asked to think aloud and to verbalize the steps taken to carry out each task as they used the devices. Participants' interactions with the devices were video and audio-recorded. Data collection was performed via video recording, post-test questionnaire, and unstructured interview. Equipment used in the usability test to collect all data were two cameras, one notebook and a support to fix the iPad and the cameras (figure 3). See the video on our website: http://site.telemedicina.ufsc.br/?page_id=410.

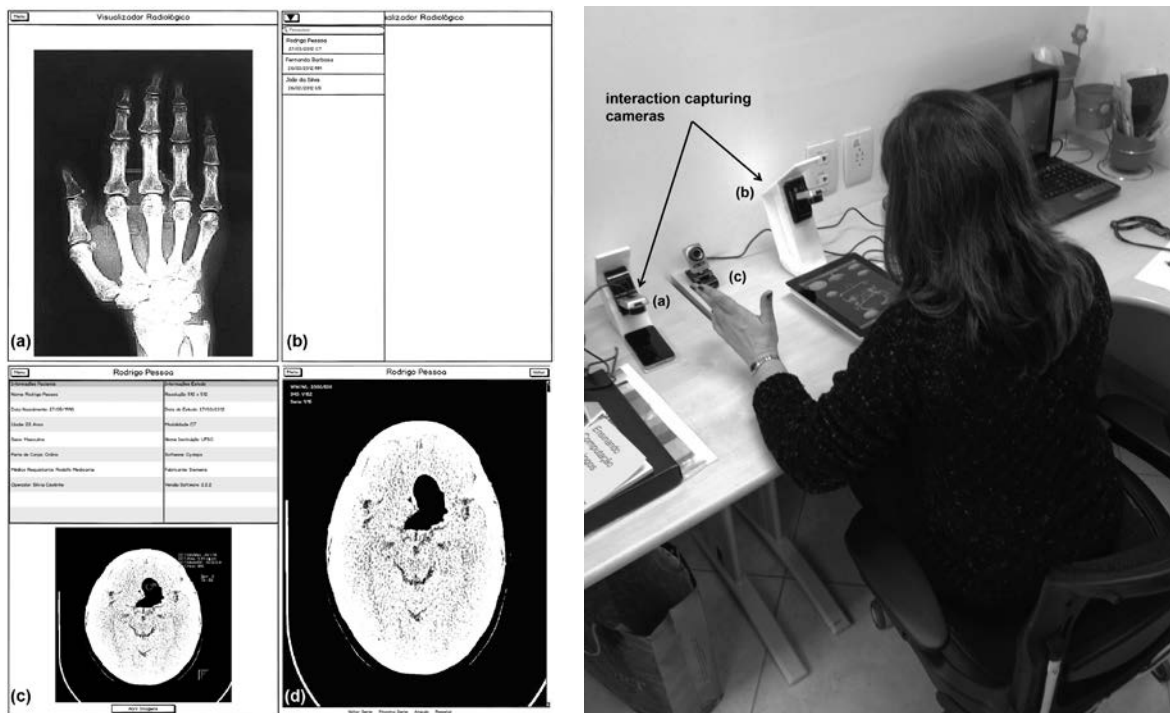


Figure.2. Left: Screen mockups of the application created with: (a) start screen, (b) patient list, (c) patient and study information and (d) study visualization. Right: Usability test set showing an interaction situation on the iPad tablet. Interactions on the device were captured through specially developed camera mounts for the cellphone (a) and the tablet (b). User attention, gazing and eye movements were captured through a second camera (c).

The notebook works as a server to capture images from both cameras simultaneously. One camera is directed to the participant's face to capture facial expressions, gazing and eye movements (see fig.3 (c)). The second camera is directed to the mobile device screen to capture the participant's interaction with the device (see fig.3 (a) and (b)). The audio is captured through the microphone of both cameras. During each session, one member of the study team moderated the usability test. After performing the task participants completed a post-test questionnaire based on SUS satisfaction questionnaire with additional questions with respect to the evaluation objectives. Demographic data of the participants is given in the table I below.

TABLE I
DEMOGRAPHIC DATA

Participant	Specialty	STT User	Has an iPad
1	Radiology	Yes	Yes
2	Stomatology	Yes	Yes
3	Neurosurgery	Yes	Yes
4	Obstetrics	Yes	Yes
5	Radiology	Yes	Yes
6	Radiology	Yes	Yes

Each item in the SUS questionnaire is measured on a 7-point Likert scale. Responses range from 1 (strongly agree) to 7 (strongly disagree). The other part of the questionnaire is made of specific question to evaluate all the other objectives that are not contemplated by the SUS questionnaire. All the questions have to be answered with “yes” or “no” and if it has a negative answer the participant should explain why.

The tests were taken between 6 and 10 of may of 2013. They were taken in the Governador Celso Ramos Hospital (HGCR), children’s hospital and INE-CTC at UFSC (Universidade Federal de Santa Catarina). In total there were 5 participants, all doctors of the SUS-SC (Unified health system of the Santa Catarina state). These participants are representatives of the users of the application created. The participants were chosen through phone contact with the help from the technical support of telemedicine in Santa Catarina.

The data were analyzed through observation of the participants during the test and manual video analysis (see Figure 3 (a) and (b)). On the videos, errors were manually identified when a participant executed an incorrect action in carrying out task steps. When the task was completed, it was coded as meeting a task goal. With the use of questionnaires, it was also possible to detect more specific problems and acknowledgments about the use of the application that the video couldn’t detect. The answers were analyzed and used to form data tables making it easier to compare all the results and get an overall impression of the performance of the application.

According to the ACR (American College of Radiology) images of computerized tomography (CT), magnetic resonance (MR) and ultrasound (US) require a monitor with at least 512 x 512 pixels and 8 pixels of depth (providing 256 shades of gray) so that there is no loss of processing and image manipulation [Drnasin 2010]. The iPad provides a screen resolution of 2048 x 1536 pixels and also enough space for visualization for almost all radiological images, as seen on table II, which shows information about the typical resolution, average quantity of images per exam and the space occupied in megabytes for common modalities of imaging study. For the study we randomly selected CT and MR imaging cases from the STT/SC database. CT cases were selected from studies of different anatomies, including dental and torso.

TABLE II
AVERAGE SIZE OF IMAGES PER MODALITY [DRNASIN 2010]

Modality	Image Size (px)	Images per Study	Space (MB)
Tomography	512 x 512	40 – 3000	20+
Ultrasound	512 x 512	15 – 240	5 – 60
Magnetic Resonance	256 x 256	50 – 2000	8+

4. Experimental Results

Having the interface prototypes defined, the process of implementing the software was performed.

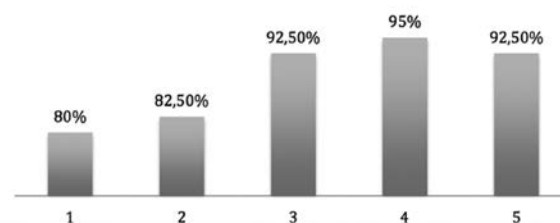


Fig.4.SUS evaluation results per participant. Scores have a minimum value of 0% and maximum value of 100%.

To achieve the goals defined a usability test was performed to identify the strong and weak spots and to verify if all functionalities implemented are intuitive and sufficient. The stages performed during the tests were:

- 1) The tester greets the participant and invites him to sit in front of the device.
- 2) The tester explains the motives for the test and also the research context. It is explained that will be made recordings during the test duration and that the objective of the test is to evaluate the interface and not the user.
- 3) It is presented to the participant the free informed consent term and it is asked to be read and signed by the participant.
- 4) The tester fills the demographic questionnaire making questions to the participant when necessary.
- 5) The tester then, delivers to the participant the task to be accomplished and the recording starts.
- 6) After the conclusion of the task it is delivered to the participant a satisfaction and acceptance questionnaire that must be filled by him
- 7) After the completion of this questionnaire the tester asks the participant if he has any doubts or commentaries and if there are it is noted and answered.
- 8) When it is over the tester thanks the presence and participation of the participant and makes final compliments.

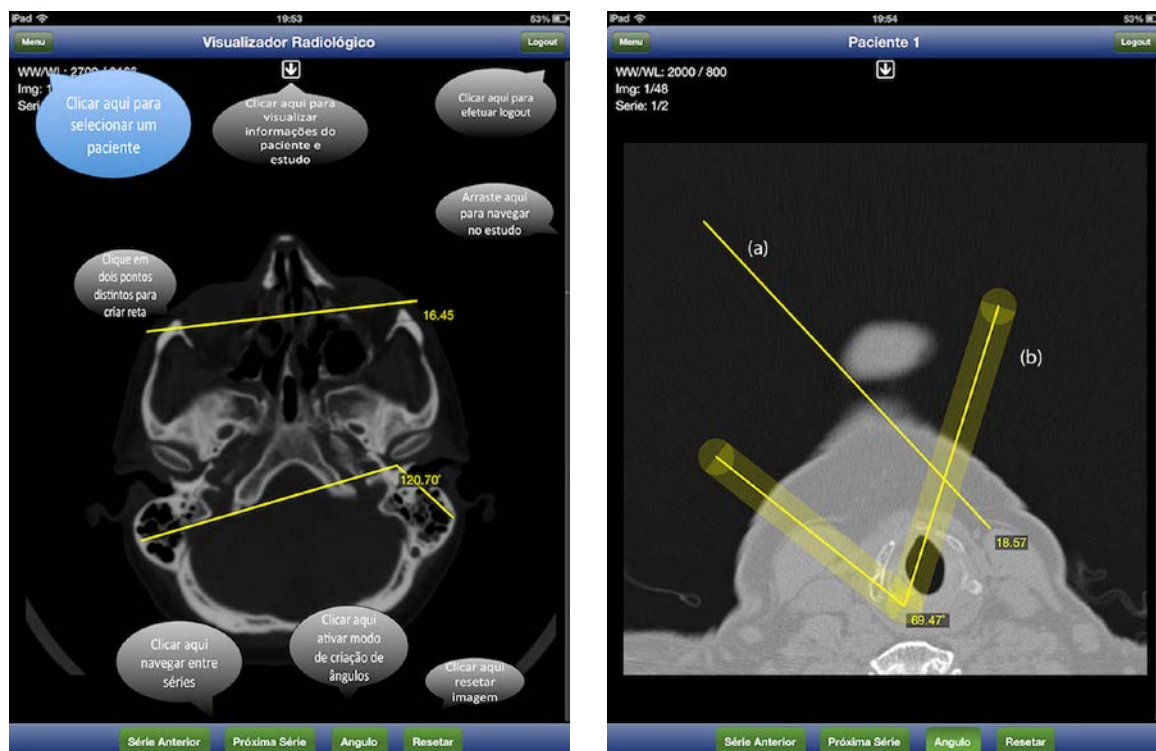


Fig.4. Left: Image visualization and analysis screen showing length and angle measurements and tooltips explaining the functionality of each screen tool. Right: The image measurement screen showing (a) an already performed length measure and (b) an angle measurement in the process of being executed.

After finishing the execution of the tasks the users answered a SUS satisfaction questionnaire. It has the objective to verify the satisfaction of the users in relation to the app.

It can be observed that the minimum score of the SUS questionnaire was 80% of satisfaction. It was expected a value above 75% so the result is considered really good, considering that the SUS has a minimum value of 0% and a maximum value of 100%. Excellent values were also

observed in the mean values for each question. With this data and verbal confirmations done at the end of the tests, it could be inferred that the app was well accepted by the participants.

It was reported by the majority of doctors that there is the necessity of more tools including:

- Point and region density (ROI);
- Measurement of small structures;
- Use of the scale in millimeters;
- Preferences for standard leveling and windowing of different tissues;
- MPR (*Multi Planar Reconstruction*);
- Dental reconstruction;
- At least 2 different visualizations simultaneously in the screen with scrolling in both of them;

So it could be inferred that there still is a need for improvements in order to all these functionalities be available. The main strong points shown by the participants where:

- Design has colors similar to the Report app (STT Mobile Version for the iPad), making it more easy for participants to get used to it;
- Login interface is similar to the Report App app (STT Mobile Version for the iPad) so is also familiar to the participants;
- Scroll bar works really well and was adopted almost instantly;
- Menu with the patient list has easy access and was considered practical;
- Image zoom with both fingers in a pinch movement was considered intuitive;

There were also weak spots raised by the participants:

- Handling the image (windowing and leveling) with 2 fingers simultaneously was not adopted rapidly;
- In ultrasound images the leveling and windowing must have better precision, it was hard for some participants to reach desired values;
- There was a difficulty in perform measurements in tiny body parts;
- Inversion of image scrolling, because now it is performed in ascending order, but all participants opted to start the scrolling from the top of the screen;
- In the patient list it was noted the need of date filters and requesting doctors associated with the DICOM studies.
- Measurements in centimeters must be converted to millimeters

5. Conclusions and Future Works

The work described in this paper had as main objective to identify critic questions related to the interface design of a teleradiology client on a tablet computer. To achieve this objective we performed a case study, creating user interfaces for a DICOM visualizer on iPad 2 that was incorporated into the STT following a pragmatic process of usability engineering. It was implemented also and functional application following a software engineering process.

This application was then evaluated by a series of usability tests with doctors of the unified health system in Santa Catarina (SUS). The application was considered practical and efficient and was well accepted by the participants of the study. We hope that this application will be used to improve the current process of generating reports in the Telemedicine System through portability, this way doctors can access these exams in an easier and faster way, without requiring a workstation or even being in a hospital.

In general, the usability tests results demonstrate a first indication that the application can meet the requirements of it's main objective, which is function as an auxiliary in the process of making and emitting a medical report for the STT system, through detailed analysis of DICOM images. According to the results the application was considered useful and objective.

The design of interfaces was considered very good by the participants, maintaining a pattern with the design of the already existing application for emission of medical reports called "STT

Laudos”. The access to images worked in an intuitive and fast way, it was praised by the participants mostly because the context where they are inserted depends on how fast they can access these images and make the report for them. The patient list was considered ideal, because it has almost all the tools necessary, there were some suggestions like the filter of patients by date.

For future works it is suggested the perfecting of the application implementing functionalities request by the test participants and integrating this application in the current context of telemedicine and Santa Catarina’s health system (SUS). There is also the suggestion to validate the application as a tool to help the creation of reports.

References

- Choi, HJ, Lee, JH, and Kang, BS. Remote CT reading using an ultramobile PC and web-based remote viewing over a wireless network. *J Telemed Telecare*, 1, 18 (Jan 2012), 26-31.
- Kim, DK, Kim, EY, Yang, KH, Lee, CK, and Yoo, SK. A mobile teleradiology imaging system with JPEG2000 for an emergency care. *J Digit Imaging*, 4, 24 (Aug 2011), 709–718.
- John, S, Poh, AC, Lim, TC, Chan, EH, and Chong, R le. The iPad tablet computer for mobile on-call radiology diagnosis? Auditing discrepancy in CT and MRI reporting. *J Digit Imaging*, 5, 25 (Oct 2012), 628-634.
- Avitzur, O. In Practice: How Neurologists are Using the Newest Tablets — In and Out of the Clinic. *Neurology Today*, 10, 24 (Dec 2010), 22-24.
- Weiss, FE, Jeudy, J, and Siegel, EL. TB or not TB? Answering an age old question using a modern device: The iPad. (2010), RSNA.
- Choudhri, AF, Carr, TM, Ho, CP, Stone, JR, Gay, SB, and Lambert, DL. Handheld device review of abdominal CT for the evaluation of acute appendicitis. *J Digit Imaging*, 25, 4 (Aug 2012), 492-6.
- Wasserman, A I. Software Engineering Issues for Mobile Application Development. (Santa Clara 2010), Proceedings of Workshop on Mobile Software Engineering, 397-400.
- Chen, Chufeng, Abdelnour-Nocera, Jose, Wells, Stephen, and Pan, Nora. Usability Practice in Medical Imaging Application Development. *Lecture Notes in Computer Science*, 5889 (2009), 405-415.
- Drnasin, Ivan and Grgic, Mislav. "The use of mobile phones in radiology," *ELMAR, 2010 PROCEEDINGS*, Zadar, 2010, pp. 17-21.
- Cannella, V.; Orazio, G.; Pirrone, R.; Vitabile, S. GUI Usability in Medical Imaging. *CISIS 2009: 778-782*
- ISO. International Organization for Standardization. ISO 9241 Requisitos Ergonômicos para Trabalho de Escritórios com Computadores Parte 11 – Orientações sobre Usabilidade, 2002.
- Chan,. Usability evaluation of order sets in a computerized provider order entry system. *BMJ Qual Saf*. v. 20, n. 11, p. 932-40, 2011.
- Yen, P..Y. Health Information Tecnology Usability Evaluation: Methods, Models and Measures. Ann Arbor: ProQuest, UMI Dissertation Publishing, 162 p., 2011.
- Siau, K. Mobile communications and mobile services. *International Journal of Mobile Communications*. v. 1, n. 1-2, p. 3-14, 2004.
- STT. Disponível em: <https://telemedicina.saude.sc.gov.br/rctm/>. Acessado em: 02 mar. 2016.
- Nielsen, Jakob. "Enhancing the explanatory power of usability heuristics." *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 1994.

Classification of malignancy of lung nodules in CT images using Convolutional Neural Network

Giovanni L. F. da Silva¹, Aristófanés C. Silva¹, Anselmo C. de Paiva¹,
Marcelo Gattass²

¹Núcleo de Computação Aplicada – Federal University of Maranhão (UFMA)
Caixa Postal 65085-580 - São Luís - MA - Brazil

²Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
Caixa Postal 22453-900 - Rio de Janeiro - RJ - Brazil

{gioh.lucca, aricsilva, anselmo.c.paiva}@gmail.com

mgattass@tecgraf.puc-rio.br

Abstract. Lung cancer presents the highest mortality rate, besides being one of the smallest survival rates after diagnosis. Thereby, early detection is extremely important for the diagnosis and treatment. This paper proposes three different architectures of Convolutional Neural Network (CNN), which is a deep learning technique, for classification of malignancy of lung nodules without computing the morphology and texture features. The methodology was tested onto the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), with the best accuracy of 82.3%, sensitivity of 79.4% and specificity 83.8%.

1. Introduction

Lung cancer is the most frequent malignant tumor and has an increase of 2% per year in its worldwide incidence. In 90% of cases, lung cancer is associated with the consumption of tobacco products. In Brazil, it is estimated an average of 28,220 cases of lung cancer in 2016, where 17,330 male and 10,890 female [INCA 2016].

One of the best opportunities to diagnose lung cancer is when an asymptomatic patient, normally a smoker, undergoes a computerized tomography (CT) exam [Srichai et al. 2007]. The detection of such nodules using CT is not a simple task, since they can have contrasts similar to other structures, low density, and a small size in an area of complex anatomy (can be connected to blood vessels or in the borders of the lung), among other issues [Leef and Klein 2002].

A variety of computer-aided detection techniques have been proposed for the detection and characterization of tumors. The development of such techniques involves two main categories: as computer-aided detection (CAD) and computer-aided diagnosis (CADx). The construction of a CADx system would reduce the number of unnecessary biopsies in patients with benign tumors, preventing physical and mental depression in patients. Thus, CADx acts as a second reader and aids experts for accurate and efficient diagnosis of cancer cells in the earlier stages of disease [Parveen and Kavitha 2014].

Deep learning techniques, in recent years, have highlighted the possibility of automatically uncovering features from the training images and exploiting the interaction,

even hierarchy, among features within the deep structure of a neural network. Accordingly, the issues of feature computing, selection, and integration can potentially be addressed by this new learning framework without a complicated pipeline of image processing and pattern recognition steps [Hua et al. 2015]. Thus, this paper proposes a methodology for classifying lung nodules in whether benign or malignant, using Convolutional Neural Network as features extractor and classifier, which can represent a second opinion to the expert.

The paper is organized as follows. Section 2 presents the related work to this research. Section 3 presents the proposed methodology for the classification of lung nodules. Section 4 presents and discusses the results. Finally, in Section 5 presents final remarks about this paper.

2. Related Work

Various initiatives are frequently developed aiming at increasing the accuracy of lung cancer diagnosis using CADx systems. Hua et al. (2015) proposed a methodology using deep learning techniques, Deep Belief Network (DBN) and Convolutional Neural Network (CNN), with a sensitivity rate of 73.40% and specificity rate of 82.20% using DBN and a sensitivity rate of 73.30% and specificity rate of 78.70% using CNN. Kumar et al. (2015) proposed a methodology using Stacked Autoencoder (SAE), deep learning technique, with an accuracy rate of 75.01%.

The work published by Kuruvilla and Gunavathi (2014) proposed a methodology based on texture features using Artificial Neural Network (ANN), with an accuracy rate of 93.30%. In their part, Gupta and Tiwari (2014) proposed a methodology based on shape features using ANN, with an accuracy rate of 90.00%. Dandil et al. (2014) proposed a methodology based on texture features using Principal Component Analysis (PCA) and ANN, with an accuracy rate of 90.63%. Parveen and Kavitha (2014) proposed a methodology based on texture features using Support Vector Machine (SVM), with a sensitivity rate of 91.38% and specificity rate of 89.56%.

The methodology proposed by Nascimento et al. (2012) based on texture features, uses Linear Discriminant Analysis (LDA) and SVM, with an accuracy rate of 92.78%. Orozco et al. (2013) proposed a methodology based on texture features using Correlation-based Feature Subset Selection, k-Nearest Neighbor (KNN) and SVM, with an accuracy rate of 82.66% and Krewer et al. (2013) proposed a methodology based in a combination of texture and shape features using Correlation-based Feature Subset Selection and KNN, with an accuracy rate of 90.91%.

3. Materials and Methods

The proposed methodology is divided into three steps as described in Figure 1. In summary, the first step is the image acquisition from the images database of CT exams in the LIDC-IDRI database. In the second step, the nodules segmentation is conducted through markings of experts. Finally, the diagnostic is completed using Convolutional Neural Network following of the results evaluation.

3.1. Image Acquisition

The image database used in this work is the LIDC-IDRI [Armato III et al. 2011], which is available online as a result of an association between the Lung Image Database Consor-

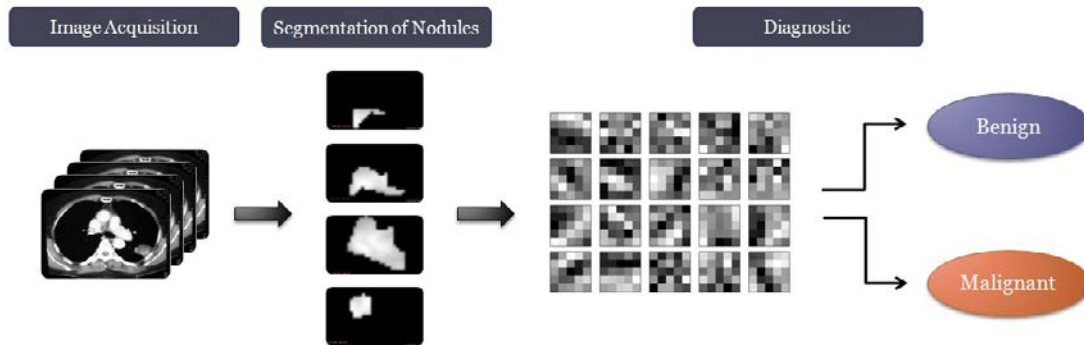


Figure 1. Proposed Methodology.

tium and the Image Database Resource Initiative, including 1,018 CT exams. However, two factors made some of them (185 exams) inappropriate for this methodology. The first factor is related to exams that do not present nodules equal to or larger than 3 mm. The second factor is the divergence of information found in the marking file of an exam versus the information present in the DICOM header of the same exam, which invalidates the marking [Carvalho Filho et al. 2014]. Therefore, the proposed methodology was applied to 833 exams.

The database supplies an XML file with contour information for the slices, and several features including calcification, texture and malignancy with values ranging from 1 to 5 for lung nodules larger than 3 mm. The process of annotating the nodules of the LIDC-IDRI database was performed by four experts in two stages. In the first stage, each expert analyzed the exams individually. In the second stage, the results of the four analyses of the first stage were presented together to the four experts. During this stage, each of the experts re-analyzed the exams and remade their annotations independently [Nascimento et al. 2012].

There is no consensus imposition, all nodules indicated by the revision of experts are taken into account and recorded. Thus, it is possible to have a different diagnosis for the same nodule. This paper considers only one instance per nodule, with the objective of minimizing the impact of subjectivity in exams. The classification of malignant or benign is obtained first with computation as presented in Jabon et al. (2009), which summarizes the features of each nodule as determined by the four experts by computing the mode or the median into a single value. According to the result of this summary, this paper considers that malignant nodules are those cases that present malignancy semantic values of moderately suspicious or highly suspicious and benign nodules are those cases that present characteristics that are highly or moderately indicative of a benign tumor. Regarding contour, the value that contains larger bounds among the four markings made during the annotating process was used. As a total, 1,405 nodules three-dimensional (1,011 benign and 394 malignant) were obtained.

3.2. Nodules Segmentation

With respect to nodules segmentation, information was obtained from a XML file containing the coordinates of the nodules with analysis criterion of each expert. However, the segmentation used in this work is summarized and presented in the Subsection 3.1, in

which only the larger bound was used to represent the instance of the nodules described by up to four markings by experts.

3.3. Diagnostic of Lung Nodules

According to Hua et al. (2015), the physical resolution of CT images on the x and y axes is higher, and three-dimensional features may sometimes be inaccurate due to low resolution on the z axis. Thus, each two-dimensional CT slice is served as individual training sample, resulting in 8,296 nodules (4,329 benign and 3,967 malignant).

To facilitate the training on the CNN architecture, each CT slice is resized to 28×28 , as illustrated in Figure 2. Nodules with smaller dimensions, Figure 2 (a), were extracted and put into background images of 28×28 and nodules with larger dimensions, Figure 2 (b), were rescaled by setting the higher axis in the dimension 28 and reducing the less axis proportionally, finally, they were extracted and put into background images of 28×28 . Thereon, each CT slice is submitted to the CNN for classification of the malignant or benign nature of lung nodules without computing the morphology and texture features.

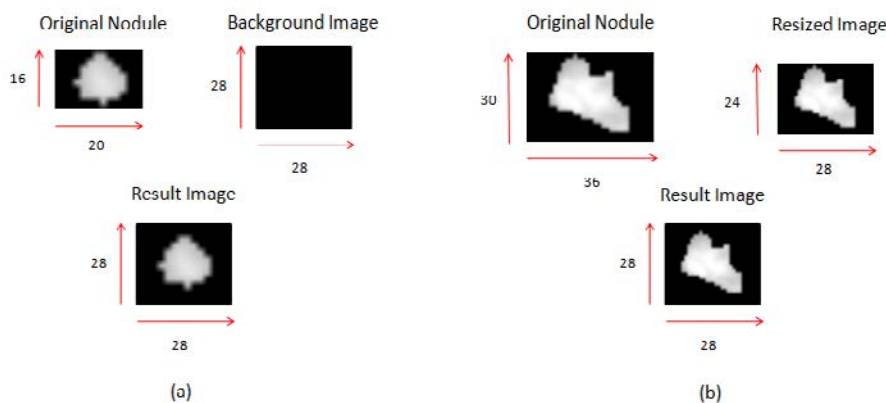


Figure 2. Image Resizing.

3.3.1. Convolutional Neural Network

Convolutional Neural Networks are a biologically-inspired trainable architecture that can learn multi-level hierarchies of features [LeCun et al. 2010]. The network extracts implicitly features of visual patterns presented as input and classify patterns from the extracted features. Typical CNN usually consist of convolutional layers, pooling layers, neuron layers and fully-connected layers [Kang and Wang 2014].

Convolutional layers have trainable filters (also called features maps) that are applied across the entire input [LeCun et al. 1989]. Each filter detects a particular feature that occurs in any part of the input [LeCun et al. 2010]. Pooling layers are non-linear down-sampling layers that yield maximum or average values in each sub-region of input image or feature maps. Pooling layers increase the robustness of translation and reduce the number of network parameters [Kang and Wang 2014]. Neuron layers apply non-linear activations on input neurons. Common activations are softmax function, sigmoid function, rectified linear unit, etc. Fully-connected layers are responsible for classifying patterns presented to CNN.

Supervised training is performed using a form of stochastic gradient descent to minimize the discrepancy between the desired output and the current output of the network [LeCun et al. 2010]. This paper uses cross entropy as the loss in the objective function [Rubinstein and Kroese 2013]. All the coefficient of all the filters in all the layers are updated simultaneously with the back-propagation method [Hecht-Nielsen 1989].

Concerning classification of the malignant or benign nature of lung nodules, three different architectures were built with the purpose of analyzing the performance of each one over the same parameters such as number of epochs, learning rate, filter size, training and testing samples, etc. One used a simple annotation to indicate the layer parameters: C(N) indicates convolutional layer with N filters, P(S) denotes pooling layer with kernel size S, CC(K) indicates fully-connected layer with K neurons, R and S represent rectified linear unit (ReLU) and softmax function, and D indicates dropout layer to help preventing overfitting, one of the problems that occur during neural network training.

Figure 3 shows architecture 1 consisting of a convolutional layer with 32 filters and ReLU activation, another convolutional layer with 16 filters and ReLU activation, a pooling layer with kernel size of 2 and a dropout layer, followed by a fully-connected layer with 16 neurons, a ReLU activation and a dropout layer, on top, a fully-connected layer with 2 neurons and a softmax function.

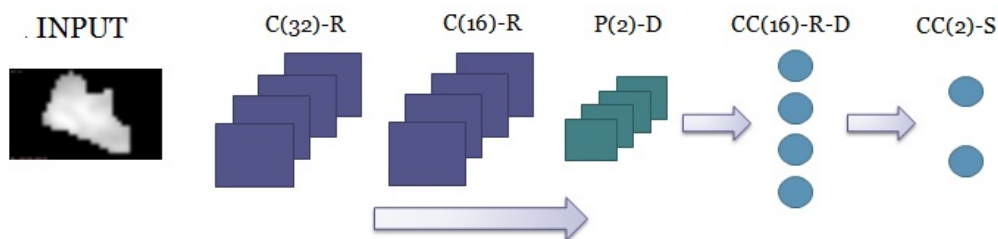


Figure 3. Architecture 1 of CNN.

Architecture 2, described in Figure 4, consists of a convolutional layer with 16 filters and ReLU activation, a pooling layer with kernel size of 2, another convolutional layer with 32 filters and ReLU activation, a pooling layer with kernel size of 2 and a dropout layer, followed by fully-connected layer with 16 neurons, a ReLU activation and a dropout layer, on top, a fully-connected layer with 2 neurons and a softmax function.

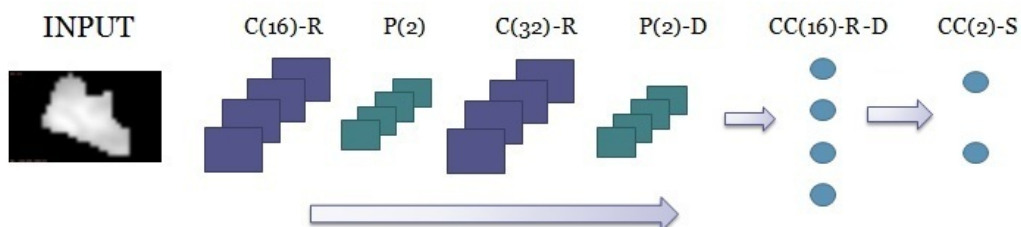


Figure 4. Architecture 2 of CNN.

Finally, architecture 3, presented in Figure 5, consists of a convolutional layer with 16 filters and ReLU activation, a pooling layer with kernel size of 2, another convolutional

layer with 32 and ReLU activation, a convolution layer with 8 filters and ReLU activation, followed by fully-connected layer with 16 neurons, a ReLU activation and a dropout layer, another fully-connected layer with 32 neurons and ReLU activation, on top, a fully-connected layer with 2 neurons and a softmax function.

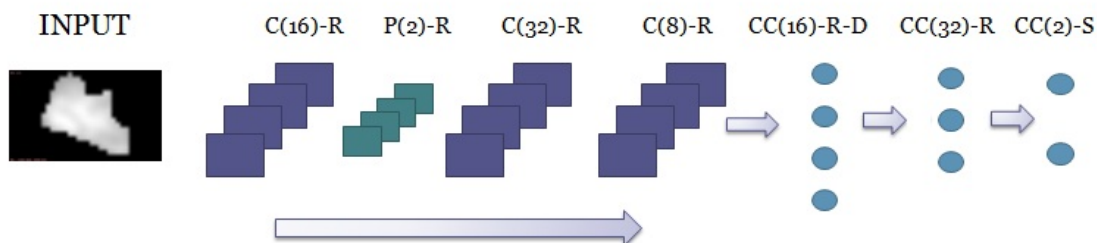


Figure 5. Architecture 3 of CNN.

4. Results and Discussion

In order to evaluate the proposed architectures, one used 7,000 nodules (3,500 benign and 3,500 malignant) for training and 1,296 nodules (829 benign and 467 malignant) for testing. All architectures were trained and tested with the same data set, with learning rate of 0.01, kernel size of 5 in convolutional layer and kernel size of 2 in pooling layer, with training in batch of 30 by 30, for 30 epochs. Table 1 summarizes the performance of accuracy, sensitivity and specificity for each architecture.

Table 1. Results for all architectures.

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)
Architecture 1	80.2	75.2	83.1
Architecture 2	80.2	82.9	78.6
Architecture 3	82.3	79.4	83.8

Architecture 1 obtained accuracy of 80.2%, sensitivity of 75.2% and specificity of 83.1%, architecture 2 achieved the same accuracy of architecture 1, sensitivity of 82.9% and specificity of 78.6%. The best accuracy was obtained by architecture 3 with 82.3%, sensitivity of 79.4% and specificity of 83.8%, we can attribute this to the fact contain more convolution layer, thus generating more features maps, and the presence of a hidden layer in the fully-connected layer.

Regarding sensitivity, success rate of malignant nodules, the architecture 2 is better than other architectures with rate of 82.9%, showing the importance of pooling layer after each convolutional layer. According to Browne and Ghidary (2003), a CNN with convolutional layer followed by pooling layer, the outcome is a gradual increase in the number of features used to describe the image, combined with a gradual decrease in spatial resolution.

In the Figure 6 are presented cases in which all the three architectures of CNN succeeded and failed to the classification. Figure 6 (a) presents a benign nodule that networks classified as benign, Figure 6 (b) presents a benign nodule that networks classified

as malign, Figure 6 (c) presents a malignant nodule that networks classified as malignant and Figure 6 (d) presents a malignant nodule that networks classified as benign. It shows that all the architectures have learned how to classify the nodules in malignant and benign considering most the morphology instead of the texture, thereby causing some false positives and false negatives in the classification.



Figure 6. Examples of Nodules.

Table 2 shows a brief comparison between the results found on this paper and some of the related works. It is important to emphasize that to perform a reliable comparison with these previous works, it should be necessary to use the same image database, same training and testing set, and same settings for the classifiers, among other parameters. Notwithstanding, comparing with the works that used deep learning techniques such as Hua et al. (2015) and Kumar et al. (2015), the present work presents better results. So, it can be attributed to the large amount of images used in training, this is, 7,000 nodules (3,500 benign and 3,500 malignant).

Table 2. Comparison with other publications with respect to the classification of lung nodules in benign and malignant. Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe).

Work	Database (Samples)	Acc (%)	Sen (%)	Spe (%)
Nascimento, 2012	LIDC (73)	92.78	85.64	97.89
Orozco et al., 2012	NBIA-ELCAP (113)	-	96.15	52.17
Krewer et al., 2013	LIDC-IDRI (33)	90.91	85.71	94.74
Dandil et al., 2014	Private (128)	90.63	92.30	89.47
Parveen and Kavitha, 2014	Private (3,278)	-	91.38	89.56
Kuruvilla and Gunavathi, 2014	LIDC (110)	93.30	91.40	100
Gupta and Tiwari, 2014	Private (120)	90	86.66	93.33
Hua et al., 2015	LIDC (2,545)	-	73.30	78.70
Kumar et al., 2015	LIDC (4,323)	75.01	83.35	-
Proposed Architecture 1	LIDC-IDRI (8,296)	80.2	75.2	83.1
Proposed Architecture 2	LIDC-IDRI (8,296)	80.2	82.9	78.6
Proposed Architecture 3	LIDC-IDRI (8,296)	82.3	79.4	83.8

5. Conclusion

Lung cancer remains the leading cause of cancer mortality worldwide, with one of the lowest survival rates after diagnosis. Therefore, early detection greatly increases the chances of improving patient survival. The major contribution of this paper lies in the

application and analysis of three different architectures of CNN, a deep learning technique, in the classification of lung nodules as benign or malignant without computing the morphology and texture features.

The methodology was tested on the LIDC-IDRI database and the best result was achieved by means of the architecture 3 with accuracy of 82.3%, sensitivity of 79.4% and specificity of 83.8%, outperforming the results obtained by other deep learning techniques. Finally, it is important to highlight the possibility of automatically uncovering features from the training images using deep learning technique and still maintain promising results, encouraging new applications of deep learning technique in CADx systems.

References

- Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., et al. (2011). The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931.
- Browne, M. and Ghidary, S. S. (2003). Convolutional neural networks for image processing: an application in robot vision. In *AI 2003: Advances in Artificial Intelligence*, pages 641–652. Springer.
- Carvalho Filho, A. O. d., de Sampaio, W. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., and Gattass, M. (2014). Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial intelligence in medicine*, 60(3):165–177.
- Dandil, E., Cakiroglu, M., Eksi, Z., Ozkan, M., Kurt, O. K., and Canan, A. (2014). Artificial neural network-based classification system for lung nodules on computed tomography scans. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 382–386. IEEE.
- Gupta, B. and Tiwari, S. (2014). Lung cancer detection using curvelet transform and neural network. *International Journal of Computer Applications*, 86(1).
- Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 593–605. IEEE.
- Hua, K.-L., Hsu, C.-H., Hidayati, S. C., Cheng, W.-H., and Chen, Y.-J. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8:2015–2022.
- INCA (2016). Instituto nacional do câncer. tipos de câncer: Pulmão. <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pulmao>. Accessed: 2016-02-29.
- Jabon, S. A., Raicu, D. S., and Furst, J. D. (2009). Content-based versus semantic-based retrieval: an lidc case study. In *SPIE Medical Imaging*, pages 72631L–72631L. International Society for Optics and Photonics.
- Kang, K. and Wang, X. (2014). Fully convolutional neural networks for crowd segmentation. *arXiv preprint arXiv:1411.4464*.

- Krewer, H., Geiger, B., Hall, L. O., Goldgof, D. B., Gu, Y., Tockman, M., and Gillies, R. J. (2013). Effect of texture features in computer aided diagnosis of pulmonary nodules in low-dose computed tomography. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 3887–3891. IEEE.
- Kumar, D., Wong, A., and Clausi, D. A. (2015). Lung nodule classification using deep features in ct images. In *Computer and Robot Vision (CRV), 2015 12th Conference on*, pages 133–138. IEEE.
- Kuruville, J. and Gunavathi, K. (2014). Lung cancer classification using neural networks for ct images. *Computer methods and programs in biomedicine*, 113(1):202–209.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.
- LeCun, Y., Kavukcuoglu, K., Farabet, C., et al. (2010). Convolutional networks and applications in vision. In *ISCAS*, pages 253–256.
- Leef, J. L. and Klein, J. S. (2002). The solitary pulmonary nodule. *Radiologic Clinics of North America*, 40(1):123–143.
- Nascimento, L. B., de Paiva, A. C., and Silva, A. C. (2012). Lung nodules classification in ct images using shannon and simpson diversity indices and svm. In *Machine Learning and Data Mining in Pattern Recognition*, pages 454–466. Springer.
- Orozco, H. M., Villegas, O. O. V., Dominguez, H. d. J. O., and Sanchez, V. G. C. (2013). Lung nodule classification in ct thorax images using support vector machines. *The 11th International Conference on Information Sciences, Signal Processing and their Applications: Main Tracks*, pages 277–283.
- Parveen, S. S. and Kavitha, C. (2014). Classification of lung cancer nodules using svm kernels. *International Journal of Computer Applications*, 95(25).
- Rubinstein, R. Y. and Kroese, D. P. (2013). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer Science & Business Media.
- Srichai, M. B., Naidich, D. P., Müller, N. L., and Webb, W. R. (2007). *Computed tomography and magnetic resonance of the thorax*. Lippincott Williams & Wilkins.

Redução de falsos positivos em imagens de mamografias digitais usando os índices *phylogenetic species variability*, *phylogenetic species richness* e múltiplos classificadores

Laércio N. Mesquita¹, Antônio O. de C. Filho¹, Alcilene D. de Sousa¹, Patrícia M. L. de L. Drumond¹

¹Campus Senador Helvídio Nunes de Barros (CSHNB) – Universidade Federal do Piauí (UFPI) – Picos – PI – Brasil

{laerciomesquita90, antoniooseas, alcileneluzsousa}@gmail.com,
patymedy@hotmail.com

Abstract: *The CADx systems has gained more attention because of his importance in the medicinal area, making the diagnosis by the experts more accurate. For the creation of these tools are used computational methods, digital image processing and knowledge about the disease. In this work we are using indices of phylogenetic diversity for extraction of features based on texture. Such indexes are used as features for the classifiers: Support Vector Machine, Random Forest, Random Basis Function and MultiLayer Perceptron, identifying in mammograms tissues the presence of mass and not mass, making thus part of a Computer-Aided Diagnosis. For the validation of the methodology, was used 200 mammography images, where 100 contain mass and the others not mass. The results showed promising, because the best results reach an accuracy of 91.5%, sensitivity of 89.5%, specificity of 94% and a false positives rate of 0.085 per examination.*

Resumo: *Os sistemas CADx têm ganhado cada vez mais atenção devido sua importância na área médica, tornando o diagnóstico por parte dos especialistas mais preciso. Para criação dessas ferramentas são utilizados métodos computacionais, processamento digital de imagens e conhecimentos sobre a doença. Neste trabalho utilizam-se índices de diversidade filogenética para extração de características baseada na textura. Tais índices são utilizados como características para os classificadores: Support Vector Machine, Random Forest, Random Basis Function e MultiLayer Perceptron identificando em tecidos de mamografias a presença de massa e não massa, perfazendo assim parte de um Computer-Aided Diagnosis. Para validação da metodologia, foram utilizadas 200 imagens de mamografia, onde 100 contém massa e as demais não massa. Os resultados mostram-se promissores, pois os melhores resultados alcançam uma acurácia de 91,5%, sensibilidade de 89,5%, especificidade de 94% e uma taxa de falsos positivos de 0,085 por exame.*

1. Introdução

Segundo o INCA (2015), o câncer de mama é um tumor maligno formado por um grupo de células cancerosas que crescem rapidamente em tecidos e órgãos, atingindo-os ao se espalharem desordenadamente por todo o corpo. O câncer de mama é mais comum entre

as mulheres, depois do câncer de pele não melanoma, respondendo por cerca de 25% dos novos casos a cada ano. O câncer de mama também acomete homens, porém é raro, representando apenas 1% do total de casos da doença. Para o ano de 2016 a estimativa é de 57.960 novos casos.

Este tipo de câncer é relativamente raro antes dos 35 anos, acima desta idade sua incidência cresce progressivamente, especialmente após os 50 anos. Estatísticas indicam um aumento da sua incidência tanto nos países desenvolvidos quanto nos em desenvolvimento. A detecção precoce do câncer de mama pode também ser feita pela mamografia, quando realizada em mulheres sem sinais e sintomas da doença, numa faixa etária em que haja um balanço favorável entre benefícios e riscos. A recomendação no Brasil, atualizada em 2015, é que mulheres entre 50 e 69 anos façam uma mamografia a cada dois anos. Os benefícios da mamografia de rastreamento incluem a possibilidade de encontrar o câncer em fase inicial e ter um tratamento menos agressivo, como também uma menor chance de morrer pela doença [INCA 2016].

Para que esse rastreamento ocorra de forma mais eficiente, os profissionais da área saúde utilizam de ferramentas que darão uma segunda opinião quando usadas em exames de mamografias. Essas ferramentas são conhecidas como *Computer-Aided Diagnosis* (CADx), cujos sistemas são responsáveis por diminuir um diagnóstico errôneo, aumentando, assim, as chances de detecção do câncer de mama em estágios iniciais.

Esse trabalho tem como objetivo descrever uma nova abordagem para extração de características baseado em índices de diversidade filogenéticos, que serão responsáveis por caracterizar as regiões de interesse em um sistema CADx, outra contribuição desse trabalho é a redução da taxa de falsos positivos nesse tipo de sistema. Serão usados descritores de textura baseados em índices filogenéticos, em seguida, será feita uma classificação através dos classificadores: *MultiLayer Perceptron* [Haykin 2008], *Radial Basis Function* [Haykin 2008], *Random Forest* [Breiman 2001], *Support Vector Machine* (SVM) [Vapnik 1998] com intuito diferenciar massas e não massas em exames de mamografias e, por fim, reduzir a quantidade de falsos positivos para o pré-diagnóstico do câncer de mama.

Este trabalho está dividido da seguinte forma: na Seção 2, são apresentados trabalhos relacionados; na Seção 3 descreve-se a metodologia empregada; na Seção 4, são descritos os resultados da execução das fases do projeto; e por fim, são apresentadas as conclusões e trabalhos futuros na Seção 5.

2. Trabalhos Relacionados

Na literatura especializada, existem diversos trabalhos relacionados à diminuição de falsos positivos em exames de mamografia e classificação de massas e não massas. Para este propósito, utilizam-se características extraídas de imagens médicas, as quais servem como vetores de entrada para os classificadores.

Sampaio (2015) apresenta uma metodologia computacional que ajuda o especialista na descoberta de massas mamárias, baseando-se na densidade da mama. Para a segmentação foi utilizado um Micro Algoritmo Genético (μ AG) objetivando criar uma máscara de proximidade de textura e selecionar regiões suspeitas em conter lesão. Para reduzir o número de regiões suspeitas erroneamente segmentadas, utilizaram-se duas etapas de redução de falsos positivos. A primeira redução de falsos

positivos usa o *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) e um ranking de proximidade de textura extraídos das regiões de interesse. Na segunda, as regiões resultantes têm as suas texturas e formas analisadas pela combinação de Árvores Filogenéticas e descritores geométricos, Padrões Binários Locais e SVM. Um μ AG foi utilizado para escolher as regiões suspeitas que gerem os melhores modelos de treinamento e maximizem a classificação de massas e não massas usadas na SVM. Os melhores resultados obtidos produziram uma sensibilidade de 94,02%, especificidade de 82,28% e acurácia de 84,08%, com uma taxa de 0,85 falsos positivos por imagem e uma área sob a *Free-Response ROC Curve* (curva FROC) de 1,13 nas análises de mamas não densas. Para mamas densas, obteve-se uma sensibilidade de 89,13%, especificidade de 88,61% e acurácia de 88,69%, com uma taxa de 0,71 falsos positivos por imagem e uma Área sob a Curva FROC de 1,47.

Oliveira (2013) propõe uma metodologia de discriminação e classificação de regiões extraídas da mama em massa e não massa. O banco de imagens *Digital Database for Screening Mammography* (DDSM) é usado neste trabalho para a aquisição das mamografias, onde são extraídas as regiões de massa e não massa. Na descrição da textura da região de interesse são usados os Índices de Diversidade Taxonômica (Δ) e Distinção Taxonômica (Δ^*), provenientes da ecologia. O cálculo destes índices é baseado nas árvores filogenéticas, sendo aplicados neste trabalho na descrição de padrões em regiões das imagens da mama com duas abordagens de regiões delimitadoras para análise da textura: círculo com anéis e máscaras internas com externas. Para classificação das regiões em massa e não massa, utilizou-se o classificador SVM. A metodologia apresenta resultados promissores para a classificação de massas e não massas, alcançando uma acurácia média de 99,67%.

Carvalho (2012) também propõe uma metodologia de discriminação e classificação de regiões extraídas de mamografias em massa e não massa. Neste estudo, a *Digital Database for Screening Mammography* (DDSM) é usada. Para descrever a textura da região de interesse, é aplicado o Índice de Diversidade de *McIntosh*, comumente usado em ecologia. O cálculo deste índice é proposto em quatro abordagens: através do Histograma, da Matriz de Co-ocorrência de Níveis de Cinza, da Matriz de Comprimentos de Corrida de Cinza e da Matriz de Comprimentos de Lacuna de Cinza. Para classificação das regiões em massa e não massa foi utilizado o classificador supervisionado SVM. A metodologia apresenta resultados eficazes para a classificação de massas e não massas, alcançando uma acurácia de 93,68%.

Sousa (2011) apresenta uma metodologia de discriminação e classificação de regiões de tecidos de mamografias em massa e não massa. Para este propósito utiliza-se o Índice de Diversidade de *Shannon-Wiener*, comumente aplicado para medir a biodiversidade em um ecossistema, o qual descreve padrões de regiões de imagens de mama com quatro abordagens: global, em círculos, em anéis e direcional. Em seguida, utiliza-se o classificador SVM para classificar estas regiões em massa e não massa. A metodologia apresenta resultados promissores para a classificação de regiões de tecidos de mamografia em massa e não massa, obtendo uma acurácia máxima de 99,85%.

Esse trabalho tem como objetivo a redução de falsos positivos em exames de mamografias, usando índices de diversidade filogenéticos *Phylogenetic Species Variability* (PSV) e *Phylogenetic Species Richness* (PSR) e múltiplos classificadores. O presente estudo apresenta uma abordagem inovadora, em que são utilizadas

características que baseiam-se em comportamentos presentes em comunidades, como o parentesco entre espécies e sua riqueza. Essa metodologia foi aplicada em imagens de mamografia para identificação de massa e não massa de forma automática.

3. Materiais e métodos

Para que fosse possível identificar a presença de massa e não massa em exames de mamografia, foi empregada a seguinte metodologia: utilização da base de imagens pública *Digital Database for Screening Mammography* (DDSM); para a extração de características foi utilizado o descritor baseado na textura utilizando os índices de diversidade filogenética *Phylogenetic Species Variability* e *Phylogenetic Species Richness* em [Helmus *et al.* 2007]; e para a classificação utilizamos a Máquina de Vetores de Suporte em [Vapnik 1998], *Random Forest* [Breiman 2001], RBF [Haykin 2008], MLP [Haykin 2008] como está demonstrado na Figura 1.

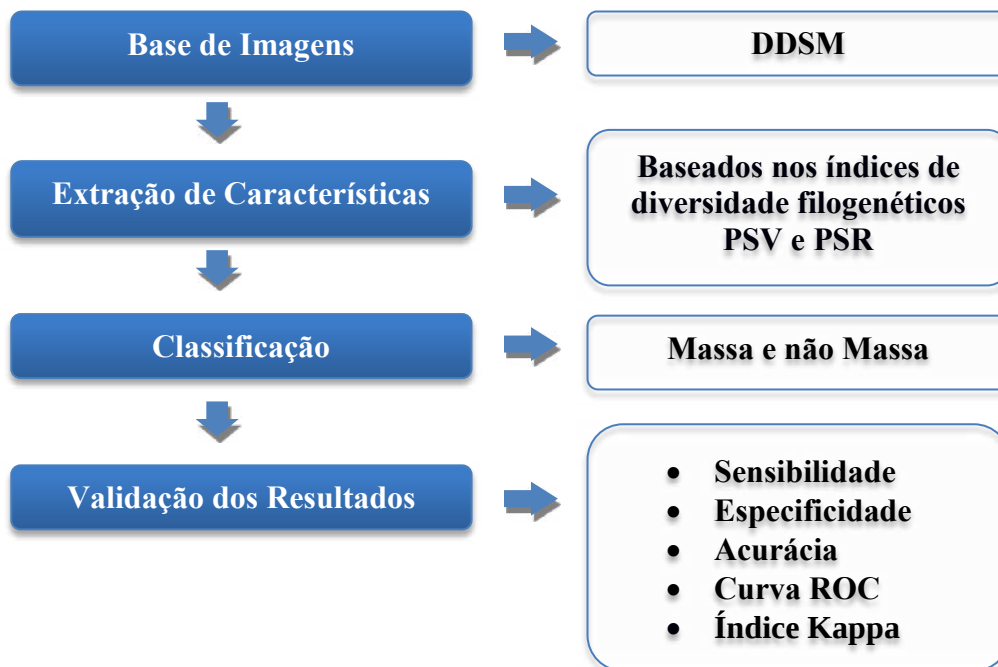


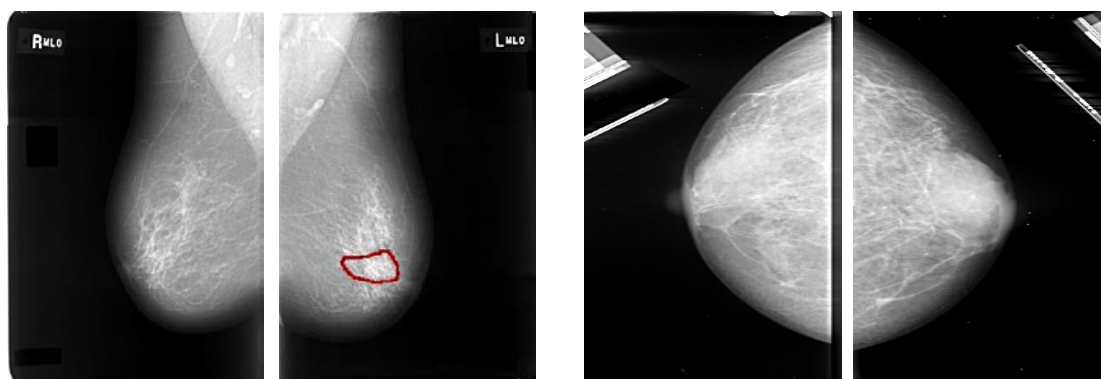
Figura 1. Metodologia proposta para redução de falsos positivos

3.1. Referencial Teórico

A mamografia é a radiografia da mama que permite a detecção precoce do câncer, sendo capaz de mostrar lesões em fase inicial, muito pequenas (de milímetros). Este exame é realizado em um aparelho de raio-X apropriado, chamado mamógrafo. Nele, a mama é comprimida de forma a fornecer melhores imagens e, portanto, melhor capacidade de diagnóstico. Estudos sobre a efetividade da mamografia sempre utilizam o exame clínico como exame adicional, o que torna difícil distinguir a sensibilidade do método como estratégia isolada de rastreamento. A sensibilidade varia de 46% a 88% e depende de fatores tais como: tamanho e localização da lesão, densidade do tecido mamário (mulheres mais jovens apresentam mamas mais densas), qualidade dos recursos técnicos e habilidade de interpretação do radiologista. A especificidade varia entre 82% e 99% e depende igualmente da qualidade do exame [INCA 2016].

Normalmente são feitas duas radiografias, sendo uma para cada mama, em duas

projeções: Médio Lateral Oblíqua (MLO) e uma Crânio-Caudal (CC). Na Figura 2, podemos ver dois exemplos de mamografia.



a) MLO esquerdo b) MLO direito c) CC esquerda d) CC direito

Figura 2. Exemplos de Mamografia

Pode ser observado na Figura 2. a) e b) Mamografias com incidência Médio-Lateral Oblíquo em ambas as mamas, sendo que em (b) mostra delimitado em vermelho uma massa maligna [Heath *et al.* 2000]; e temos em (c) e (d) Mamografias com incidência Crânio-Caudal em ambas as mamas e que não apresentam região suspeita de anormalidade [Heath *et al.* 2000].

3.2 Base de Imagens

A *Digital Database for Screening Mammography (DDSM)* [Heath *et al.* 2000] é uma base pública de imagens de mamografias, a qual foi uma doação do *Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Comman.* O objetivo desse banco é facilitar a pesquisa e desenvolvimento de algoritmos para ajudar no diagnóstico, além de ser uma ferramenta de ensino para formação profissional. O banco possui 2500 estudos. Cada estudo contém duas imagens de cada mama, juntamente com informações do paciente, tais como classificação da densidade da mama, sutileza para anormalidades e informações de imagens, como resolução espacial e imagens contendo áreas suspeitas. Para a concretização desse trabalho, foram utilizadas 200 imagens de mamografias; sendo 100 imagens com a presença de massa e 100 de não massa.

3.3 Índices de Diversidade Filogenética

Segundo Oliveira (2003), a diversidade é um termo muito utilizado na área da ecologia. O seu objetivo é informar a variedade de espécies presentes em uma comunidade ou área. Para [Webb *et al.* 2000] a filogenia pode ser utilizada para investigar de forma contemporânea os processos ecológicos e a composição das espécies nas comunidades. Tendo isto em mente, foram utilizados os índices PSV e PSR com intuito de encontrar padrões em regiões de imagens de mamografias. Nesse sentido os índices foram utilizados para descrever a textura das regiões de interesse do inglês *Region of Interest (ROIs)* de massa e não massa extraída de exames de mamografias. Para a obtenção dos índices foram usadas apenas as informações das espécies (níveis de cinza) presentes na ROI analisada e a quantidade de indivíduos de cada espécie.

3.4 Phylogenetic Species Variability

Segundo [Helmus *et al.* 2007] a variabilidade quantifica o parentesco filogenético, diminuindo a variação das características compartilhadas por todas as espécies da comunidade. A variabilidade de espécies filogenéticas resume o grau em que as espécies de uma comunidade são filogeneticamente relacionadas. Na fórmula 1 do PSV, o trC representa a soma dos valores da diagonal de uma matriz C , $\sum C$ é o somatório de todos os valores da matriz, n é o número de espécies e \bar{c} é a média dos elementos da diagonal de C .

$$PSV = \frac{ntrC - \sum C}{n(n-1)} = 1 - \bar{c} \quad (1)$$

3.5 Phylogenetic Species Richness

A riqueza de espécies quantifica o número de espécies em uma comunidade. Podemos encontrar o valor do PSR através da fórmula 2, onde multiplicamos o número espécies n pela variabilidade da comunidade [Helmus *et al.* 2007].

$$PSR = nPSV \quad (2)$$

3.6 Classificação e Validação

A classificação foi realizada pelo *software WEKA*, que é um conjunto de algoritmos de aprendizado de máquina para tarefas de mineração de dados. Contém ferramentas para pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização. Ele foi desenvolvido por um grupo de pesquisadores da Universidade de Waikato, Nova Zelândia [WEKA 2016].

A fim de considerarmos a presença ou ausência de massa em imagens de mamografia, para validação dos resultados, utilizou-se de métricas de avaliação baseadas em estatísticas como, a Área sob a Curva ROC, que mensura o quanto o algoritmo é eficiente, usualmente referida como AUC do inglês *Area Under the Curve* é referida na tabela como (AUC) [Metz 1986], Sensibilidade (SE) [Martinez *et al.* 2003], Especificidade (EP) [Martinez *et al.* 2003], Acurácia (AC) [Metz 1986] e índice Kappa (K) [Landis e Koch 1977]. Segundo [Martinez *et al.* 2003] a sensibilidade (SE) é definida como a probabilidade do teste sob investigação fornecer um resultado positivo, dado que o indivíduo é realmente portador da enfermidade. Ainda em [Martinez *et al.* 2003] a especificidade (EP) é definida como a probabilidade do teste fornecer um resultado negativo, dado que o indivíduo está livre da enfermidade. Em [Metz 1986] acurácia é a métrica que calcula o total de acertos em relação a todas as instâncias classificadas corretamente. Essa medida mede o quão bem um classificador reconhece instâncias de diversas classes. Foram obtidos também os Falsos Positivos por imagem (FP/i), para uma análise mais detalhada. O índice Kappa (K) de acordo com [Landis e Koch 1977] é uma medida de concordância que pode ser formulada para medir o desacordo de um conjunto de respostas, baseada em pesos, medindo a concordância entre um número de resposta baseado em observadores, chegando a um consenso.

Os classificadores testados utilizaram o método *k-fold cross validation* para a obtenção dos resultados. Os dados foram divididos em 10 conjuntos, sendo 9 deles para treinamento e 1 para testes. Este processo é repetido 10 vezes, de forma que o conjunto escolhido para o teste será diferente do anterior e no final é gerada uma média dos resultados.

4. Resultados e Discussões

Para os testes realizados neste trabalho, utilizou-se a base DDSM [Heath *et al.* 2000]. As características foram extraídas a partir dos índices de diversidade filogenéticos de variabilidade e riqueza de espécies. Os classificadores *MultiLayer Perceptron* (MLP) [Haykin 2008], *Radial Basis Function* (RBF) [Haykin 2008], *Random Forest* (RF) [Breiman 2001], *Support Vector Machine* (SVM) [Vapnik 1998] foram usados para classificação de massa e não massa em imagens de mamografias, utilizou-se vários parâmetros tendo em vista melhores resultados, os melhores parâmetros para o MLP foi utilizando 6 camadas ocultas, para o RBF foi usando 11 *clusters* e para o RF foi utilizando 155 árvores e para o SVM foi utilizando o *kernel* linear e como parâmetros os valores padrões, através do *software WEKA* [WEKA 2016]. A classificação foi realizada com validação cruzada de *k-folds*, sendo $k = 10$.

Tabela 1. Resultados da classificação usando PSV

Classificador	SE	EP	FP/i	AUC	AC	K
SVM	61,6%	77%	0,355	0,645	64,5%	0,29
RF	56,2%	59%	0,435	0,632	56,5%	0,13
RBF	62,2%	74%	0,355	0,621	64,5%	0,29
MLP	63,7%	65%	0,360	0,661	64%	0,28

Nos testes realizados, o melhor resultado foi obtido através do classificador SVM, com uma acurácia de 64,5%, sensibilidade de 77% e especificidade de 61,6%, a Área sob a Curva ROC foi de 0,645 é um índice Kappa de 0,29, sendo considerado razoável de acordo com [Landis e Koch 1977]. O resultado menos significativo foi obtido através do classificador RF com uma acurácia de 56,5%, sensibilidade de 56,2%, especificidade de 59% e um índice Kappa de 0,13.

Tabela 2. Resultados da classificação usando PSR

Classificador	SE	EP	FP/i	AUC	AC	K
SVM	75,9%	82%	0,220	0,780	78%	0,56
RF	72,3%	68%	0,290	0,820	71%	0,42
RBF	73,2%	90%	0,215	0,836	78,5%	0,57
MLP	75%	84%	0,220	0,845	78%	0,56

Podemos observar na Tabela 2 que, para o índice PSR, o melhor resultado obtido foi com o classificador *RBF*, o qual atingiu uma acurácia de 78,5%, sensibilidade de 73,2% e especificidade de 90%, a Área sob a Curva ROC foi de 0,836 e um índice Kappa de 0,57, sendo considerado bom de acordo com [Landis e Koch 1977]. O resultado menos relevante foi obtido pelo classificador RF com uma acurácia de 71%, sensibilidade de 72,3%, especificidade de 68% e um índice Kappa de 0,42.

Tabela 3. Resultados da classificação usando PSV e PSR

Classificador	SE	EP	FP/i	AUC	AC	K
SVM	82,9%	92%	0,135	0,865	86,5%	0,73
RF	88,3%	91%	0,105	0,934	89,5%	0,79
RBF	88,5%	92%	0,100	0,957	90%	0,80
MLP	89,5%	94%	0,085	0,950	91,5%	0,83

Como descrito na Tabela 3, a junção dos índices PSV e PSR conseguiu atingir resultados muito promissores, no melhor resultado obteve-se 91,5% de acurácia, sensibilidade de 89,5% e especificidade de 94%. A taxa de falsos positivos foi de 0,085

por imagem, com uma Área sob a Curva ROC de 0,952 e índice Kappa de 0,83, sendo considerado excelente de acordo com a tabela do índice Kappa [Landis e Koch 1977]. O resultado menos significativo para análise foi conseguido através do SVM, como uma acurácia de 86,5% e um índice Kappa de 0,73.

Tabela 4. Comparação da metodologia com os trabalhos relacionados

Trabalhos	Base	SE	EP	FP/i	AC
Sampaio (2015)	DDSM	94,02%	82,28	0,85	84,08%
Oliveira (2013)	DDSM	89,5%	-	-	99,67%
Carvalho (2012)	DDSM	90,10%	-	-	93,68%
Sousa (2011)	DDSM	91,50%	-	-	99,85%
Metodologia proposta	DDSM	89,5%	94%	0,085	91,5%

Como podemos observar na Tabela 4, a metodologia proposta utilizando os índices PSV e PSR apresentou uma pequena melhoria em relação ao trabalho de Sampaio (2015) com uma acurácia de 91,5%, especificidade de 94% e uma taxa de falsos positivos similar. Sabendo-se que os resultados de falsos positivos retornados se assemelham aos de Sampaio (2015), a utilização dos índices PSV e PSR teria uma maior eficácia na sua metodologia, reduzindo ainda mais essa taxa. Nas tabelas 1 vemos que o classificador SVM apresenta os melhores resultados, já nas tabelas 2 e 3, o classificador SVM não atingiu os resultados esperados, tendo em vista que todos os trabalhos relacionados atingiram resultados muito expressivos. Na tabela 3 era esperado que o SVM fosse o mais eficiente dentre os classificadores utilizados usando a combinação dos dois índices de diversidade. Os resultados mostraram-se promissores em relação aos trabalhos recentes na literatura relacionados à classificação de tecidos de mamografias nas classes de massa e não massa.

5. Conclusão e Trabalhos Futuros

A partir dos resultados obtidos, pode-se inferir que a utilização de descritores baseados na textura apresentam resultados eficazes. O uso de índices filogenéticos para descrever padrões em regiões de imagens mostrou-se eficiente para o que foi proposto. Para os trabalhos futuros, pretende-se utilizar novos índices filogenéticos a fim de uma melhor redução da taxa de falsos positivos, como também garantir uma maior eficiência nesse método.

Referências

- Breiman, L. (2001) “Random forests”. *Machine Learning*, v. 45, n. 1, p. 5-32.
- Carvalho, P. M. S. (2012) “Classificação de Tecidos da Mama a partir de imagens mamográficas em massa e não massa usando índices de diversidade de McIntosh e Máquina de Vetores de Suporte”. Dissertação de Mestrado na área de Ciência da Computação, (Programa de Pós-Graduação em Engenharia de Eletricidade), Universidade Federal do Maranhão, São Luís.
- Digital Database for Screening Mammography. DDSM (2016). Disponível em: <<http://marathon.csee.usf.edu/Mammography/Database.html>>. Acesso em 29 de fevereiro de 2016.
- Helmus MR, Bland TJ, Williams CK, Ives AR (2007) “Phylogenetic measures of

- biodiversity”. *American Naturalist*, 169, E68–E83.
- Heath, M.; Bowyer, K.; Kopans, D.; et al. The digital database for screening mammography. Citeseer, Proceedings of the 5th international workshop on digital mammography. p. 212–218, 2000.
- Haykin, S. (2008) “Neural Networks and Learning Machines”. 3ª ed. New Jersey: Prentice Hall, 936 p.
- Instituto Nacional do Câncer – INCA (2015), Ministério da Saúde, Estimativa 2016. Incidência do Câncer no Brasil, <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama>, março, 2016.
- Instituto Nacional do Câncer – INCA (2016), Ministério da Saúde, Câncer de mama, http://www.inca.gov.br/conteudo_view.asp?id=1932, março, 2016.
- Landis, J. Richard, and Gary G. Koch. "The measurement of observer agreement for categorical data." *biometrics* (1977): 159-174.
- Lima, B. V. A. (2014) “Rotulação de Dados com Aprendizado Semi-Supervisionado”, Dissertação de Mestrado na área de Ciência da Computação, (Programa de Pós-Graduação em Ciência da Computação), Universidade Federal do Piauí, Teresina.
- Martinez E. Z; Louzada-Neto F; Pereira B. B (2003). “A curva ROC para testes diagnósticos”. *Cadernos Saúde Coletiva* 11 (1): 7-31.
- Metz C. E. (1986). ” ROC methodology in radiologic imaging”. *Invest. Radiol*, v. 21(9), p. 720-33.
- Oliveira, F. S. S. (2013) “Classificação de Tecidos da Mama em Massa e Não-Massa usando Índice de Diversidade Taxonômico e Máquina de Vetores de Suporte”. Dissertação de Mestrado na área de Ciência da Computação, (Programa de Pós-Graduação em Engenharia de Eletricidade), Universidade Federal do Maranhão, São Luís.
- Sampaio, W. B. (2015) “Detecção de massas de imagens mamográficas usando uma metodologia adaptada à densidade da mama”. Tese de Doutorado na área de Ciência da Computação, (Programa de Pós-Graduação em Engenharia de Eletricidade), Universidade Federal do Maranhão, São Luís.
- Sousa, U. S. (2011) “Classificação de massas na mama a partir de imagens mamográficas usando o índice de diversidade de shannon-wiener”. Dissertação de Mestrado na área de Ciência da Computação, (Programa de Pós-Graduação em Engenharia de Eletricidade), Universidade Federal do Maranhão, São Luís.
- Vapnik, N, “Statistical Learning Theory”. New York: John Wiley & Sons, 1998.
- Webb, Campbell O. "Exploring the phylogenetic structure of ecological communities: an example for rain forest trees." *The American Naturalist* 156.2 (2000): 145-155.
- WEKA - Machine Learning Group at the University of Waikato. Disponível em:<<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em 26 de fev. de 2016

Segmentação Automática de Imagens Térmicas das Mamas Utilizando Limiarização com Refinamento Adaptativo

Matheus F. O. Baffa¹, Deivison J. M. Cheloni¹, Lucas G. Lattari¹

¹Departamento Acadêmico da Ciência da Computação – Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais (IFSEMG)

{mfreitas826, deivison.junior}@gmail.com, lucas.lattari@ifsestmg.edu.br

Abstract. *Breast cancer is the most common type of cancer in the world. New cases of cancer are detected every year, and 25% of them are breast cancer. It is known that early diagnosis is critical for the prognosis of patients, and image based techniques has been developed to guide an effective, minimally invasive and cheaper way to diagnose breast cancer. In this work, we propose an automatic segmentation method based on adaptive thresholding. The experimental results show that our method is competitive with other works on both solution quality and runtime performance, achieving 96% of accuracy and 98% of sensibility. Furthermore, it is simpler to implement, computationally efficient and suitable for real time applications.*

Resumo. *O câncer de mama é o tipo mais comum de câncer no mundo. Todo ano são detectados milhares de casos de câncer, e destes, 25% são de mama. Sabendo que o diagnóstico precoce é crítico para o prognóstico do paciente, novas tecnologias à base de análise de imagens são desenvolvidas para guiar um diagnóstico eficaz e menos invasivo. Nesse artigo, é desenvolvido um novo método de segmentação de imagens das mamas em imagens termográficas utilizando limiarização com refinamento adaptativo. Esse método se mostrou eficaz com aproximadamente 96% de acurácia e 98% de sensibilidade. Além disso, a abordagem proposta é simples de ser implementada computacionalmente, é eficiente e apropriada para aplicações em tempo real.*

1. Introdução

Para o ano de 2015, o Instituto Nacional do Câncer (INCA) previu um total de 500 mil novos casos de câncer somente no Brasil. Dos novos casos, estima-se que 25% seriam de câncer de mama. Essas doenças afetam homens e mulheres das mais diversas idades, sendo mais frequente em mulheres acima dos 50 anos. De acordo com uma pesquisa divulgada pelo INCA, estima-se que o câncer de mama foi responsável pela morte de aproximadamente 14 mil mulheres no ano de 2013 [INCA 2015].

A detecção precoce do câncer pode aumentar as chances de cura do paciente. Em média, os pacientes diagnosticados no início do desenvolvimento do tumor podem ter até 95% de chance de cura [HCB 2015]. O INCA divulga que não há apenas uma causa para a doença, mas diversos fatores que desencadeiam o câncer, como fatores genéticos, hormonais e ambientais.

Atualmente, existem diversas formas de se diagnosticar o câncer de mama. Um exemplo a ser considerado é o autoexame, em que a própria pessoa verifica se há alguma anomalia. Outra forma é através de exames de imagens como a mamografia, o

ultrassom e a ressonância magnética, em que através da análise de imagens, o médico irá verificar a existência de tumores. Além destas, as imagens termográficas aplicadas a saúde têm se mostrado promissoras, sendo utilizadas por vários pesquisadores para se detectar anomalias nas mamas [Arora et al. 2008, Motta 2010]. Segundo Gore e Xu [Gore and Xu 2014], o corpo humano possui alta simetria em relação a sua distribuição térmica, que se mantém estável mesmo após longos períodos de observação. Qualquer alteração na simetria térmica pode indicar a existência de doenças.

Para se diagnosticar câncer em imagens termográficas, torna-se necessário identificar a região de interesse (ROI) que compõe as mamas, removendo o fundo da cena que não é desejado para a análise. Somente a ROI é avaliada posteriormente para a identificação de doenças.

O trabalho de [Motta 2010] realiza uma segmentação automática das mamas em imagens térmicas. Seu algoritmo, representado na Figura 1, realiza a segmentação utilizando morfologia matemática, detecção de contornos usando filtros de Canny, detecção de formas através da transformada de Hough e limiarização adaptativa. A acurácia média da avaliação quantitativa foi de aproximadamente 96% em relação ao *ground truth*.

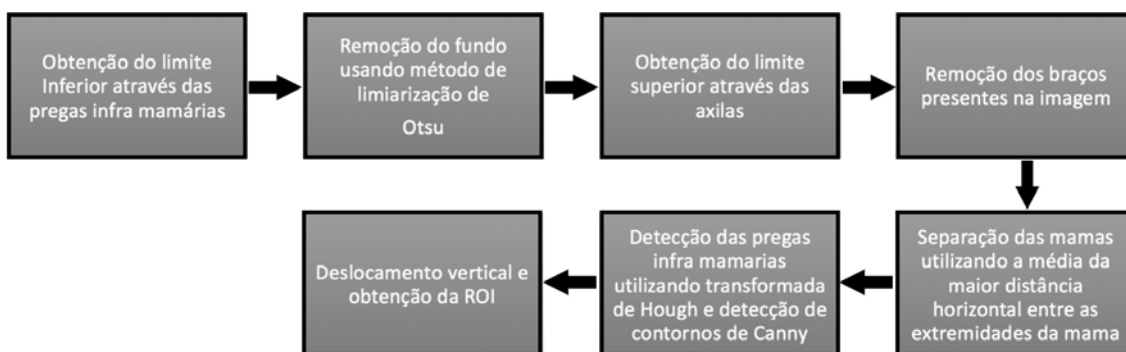


Figura 1. Fluxograma do método de segmentação proposto em [Motta 2010].

O trabalho de [Marques 2012] também apresenta uma abordagem diferente à segmentação automática das mamas. Seu método, descrito na Figura 2, inclui detecção de contornos, limiarização, crescimento de regiões, ajuste de curvas por mínimo quadrado e *B-Splines*.

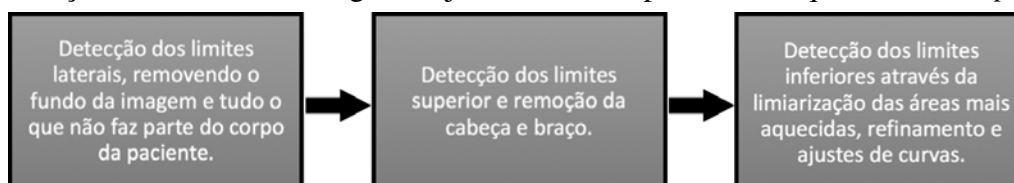


Figura 2. Fluxograma do método de segmentação proposto em [Marques 2012].

O método de [Marques 2012] apresenta uma acurácia de 97%. As imagens da ROI (*Region of Interest*) segmentadas são comparadas com um *ground truth* gerado a partir de um aplicativo usando tablets para a segmentação manual das imagens.

Nesse artigo, é desenvolvido um novo método para a segmentação da região das mamas em imagens termográficas, usando o conceito de limiarização com refinamento adaptativo. Os resultados desse método se mostraram muito promissores, obtendo taxas médias de 96% de acurácia e 98% de sensibilidade. Algumas contribuições do presente trabalho são:

- A criação de uma nova técnica de limiarização com refinamento adaptativo, que a partir da simetria térmica da imagem, identifica a região que potencialmente representa as pregas inframamárias, fundamental para a identificação das mamas. Essa descrição se mostra de alta qualidade, pois apresenta resultados competitivos e superiores a outros trabalhos da literatura, mesmo após a avaliação de resultados experimentais com centenas de imagens de teste obtidas em bases de dados científicas.
- A descrição de uma aplicação computacional que é mais simples de se implementar e portar em diversas arquiteturas computacionais, mesmo em dispositivos que dispõem de pouca memória. Além disso, o tempo de execução médio permite a possibilidade de usar o método proposto para aplicações em tempo real, como vídeos.

2. Método Proposto

As imagens termográficas das mamas são introduzidas às etapas do método proposto que é apresentado na Figura 3.

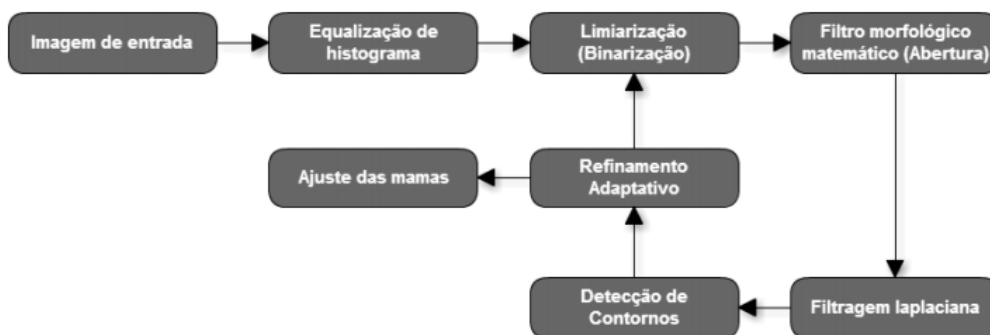


Figura 3. Fluxograma do método proposto neste trabalho.

2.1. Limiarização por Refinamento Adaptativo

Como mencionado em Marques [Marques 2012], as pregas inframamárias são indicadores dos limites inferiores das mamas. Além disso, esta região possui maior temperatura devido a sobreposição entre a parte inferior da mama e a parede torácica. Esta característica facilita a identificação computacional da prega inframamária e, posteriormente, das mamas.

Dessa forma, utilizando a limiarização com limiar L elevado, torna-se possível realizar a segmentação das pregas inframamárias. Entretanto, experimentalmente, foi observado que o valor de L ideal para a correta segmentação das pregas inframamárias é variável, dependendo do conteúdo da imagem térmica I . Para o cálculo desse valor ideal, este trabalho propõe uma abordagem de refinamento adaptativo, que tem por objetivo calcular o melhor valor de L para qualquer I considerado.

No início do método, I é limiarizado com limiar elevado ($L = 254$). Após essa etapa, a imagem térmica é filtrada usando um operador morfológico de abertura. O objetivo da aplicação desse filtro é remover pequenos artefatos de uma imagem que deterioram o resultado final.

Na próxima etapa, é realizada a filtragem laplaciana de I , após a limiarização utilizada para a obtenção da prega inframamária. Isso é feito para a obtenção dos contornos das pregas e, conseqüentemente, permitir uma análise geométrica para a identificação da região mamária.

A partir dos contornos obtidos, a área A_i de cada segmento conexo $i \in I$ é calculada e comparada a um conjunto de limiares A_{l1} e A_{l2} fornecidos a priori. Essa avaliação é realizada da seguinte maneira:

- É procurado um segmento $i_1 \in I$, tal que sua área A_{i1} seja $A_{i1} > A_{l1}$.
- São procurados dois segmentos $i_1, i_2 \in I$, tal que suas áreas A_{i1} e A_{i2} sejam, respectivamente, $A_{i1} > A_{l2}$ e $A_{i2} > A_{l2}$.

Se qualquer uma das duas condições listadas anteriormente forem atendidas, então as pregas inframamárias foram identificadas e a etapa de refinamento adaptativo é encerrada. Caso contrário, a etapa de refinamento adaptativo é repetida a partir da limiarização. No entanto, o valor de L é decrementado em uma unidade ($L = 253$). A ideia é repetir iterativamente esse procedimento até que a região das pregas inframamárias seja identificada.

Logo, a redução iterativa desse limiar acaba por maximizar a área pertencente às pregas inframamárias, minimizando o efeito causado pela introdução de artefatos indesejados.

Dois limiares associados às áreas são necessários devido a uma avaliação empírica observada durante a criação do método. Em uma série de imagens de teste consideradas, observou-se que a região de alta temperatura associada às pregas inframamárias é composta por uma única componente conexa. No entanto, para um conjunto de imagens, notou-se que a região das pregas inframamárias é composta por duas componentes conexas de menor tamanho. Dessa forma, com a introdução das duas condições mencionadas, foram obtidas às regiões de prega inframamária para um número elevado de imagens advindas das bases de dados utilizadas na literatura.

Uma vez obtida a região das pregas inframamárias, a etapa a seguir, denominada ajuste de curvas, tem por intuito representar de maneira mais fidedigna às curvas que compõem as mamas. Essa representação é feita a partir da modelagem realizada por um polinômio cúbico, que interpola todos os pontos da região inferior das pregas obtidas na etapa anterior.

Finalmente, é utilizado um operador de filtro laplaciano na imagem térmica original, após a conversão de suas cores para tons de cinza. A combinação dos segmentos de ambas as imagens determina a detecção final da mama.

3. Desenvolvimento

3.1. Metodologia

A metodologia apresentada nesse trabalho foi desenvolvida utilizando o Python 2.7.8 e a biblioteca OpenCV 3.1. Para a geração de resultados e análise de tempos computacionais, foi utilizado um notebook Asus N46V com processador Intel I7, 8 GB de RAM e sistema operacional Windows 10.

Para a análise de resultados é empregada uma base de dados pública conhecida [Silva et al. 2014], totalizando 283 imagens térmicas. As imagens possuem resolução espacial de 320 x 240 pixels, além de 256 níveis de resolução tonal. Maiores informações sobre a captura das imagens utilizadas podem ser vistas no mesmo trabalho. Alguns exemplos de imagens da base podem ser vistos na Figura 4.

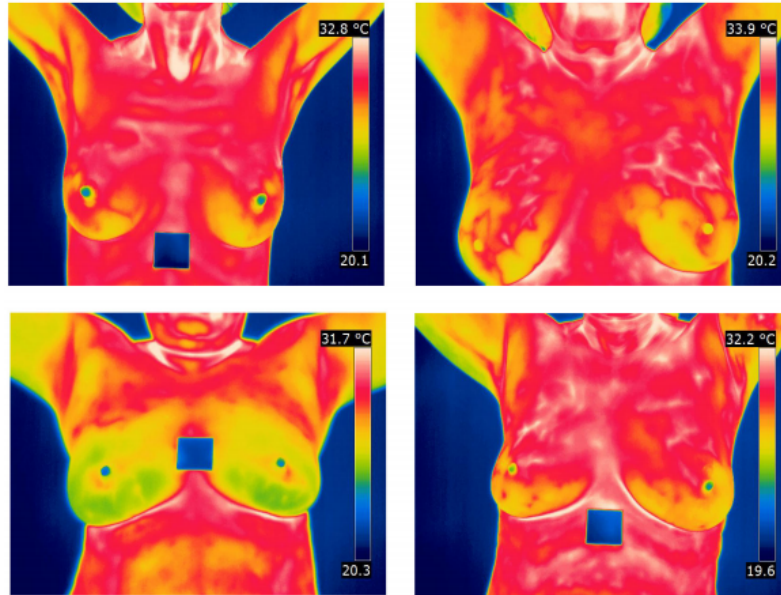


Figura 4. Algumas imagens advindas da base de dados apresentada em [Silva et al. 2014].

Para cada imagem de teste disponível na base de dados, é gerada sua versão *ground truth*. Essas imagens são construídas a partir de um método de política de votos apresentado em Li et al. [Li et al. 2011], que utiliza diferentes imagens *ground truths* obtidas manualmente. Essas imagens são comparadas com as obtidas pelo método proposto nesse artigo.

Para avaliar a eficiência desse método, diferentes métricas estatísticas serão consideradas: acurácia, sensibilidade, especificidade, preditividade positiva e negativa. Tais medidas são definidas pelas equações:

$$Acurácia(ACC) = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

$$Sensibilidade(SEN) = \frac{VP}{VP + FN} \quad (2)$$

$$Especificidade(ESP) = \frac{VN}{VN + FP} \quad (3)$$

$$PreditividadePositiva(PDP) = \frac{VP}{VP + FP} \quad (4)$$

$$PreditividadeNegativa(PDN) = \frac{VN}{VN + FN}, \quad (5)$$

em que VP representa as amostras de verdadeiros positivos (pixel que for identificado como parte da região da mama, tanto na imagem teste quanto em sua versão *ground truth*), VN representa os verdadeiros negativos considerados (pixel que for identificado como parte da região do fundo, tanto na imagem teste quanto em sua versão *ground truth*), FP denota os falsos positivos avaliados (pixel identificado incorretamente como parte da mama na imagem teste) e, finalmente, FN assinala as amostras de falsos negativos obtidas (pixel erradamente reconhecido como parte do fundo).

Para maximizar os resultados obtidos, foram introduzidos três parâmetros avaliados experimentalmente e empiricamente. A utilização desses parâmetros é necessária pois foi constatado, durante a análise do método, que as pregas inframamárias identificadas podem variar significativamente, dependendo das mamas consideradas. Esses parâmetros estão diretamente relacionados com o procedimento de limiarização com refinamento adaptativo.

O primeiro parâmetro verifica se existe uma única componente conexa que compõe a região das pregas inframamárias, sendo denominado la_1 . Esse valor determina um limiar mínimo de área medida em pixels que a prega inframamária deve ter, antes da etapa de ajustes das mamas.

Analogamente, o segundo parâmetro analisa se existem duas componentes conexas que definem as pregas inframamárias, sendo esse chamado la_2 . Essa informação determina uma área mínima a ser medida em pixels para as duas componentes distintas, uma para cada mama.

Finalmente, o último parâmetro se baseia no conceito de que a prega inframamária se localiza nos quadrantes inferiores da imagem térmica. Ele é denominado la_r . Dessa maneira, após a etapa de limiarização, todos os pixels $I(x, y)$, tal que

$$I(x, y) = \begin{cases} 0, & \text{se } x, y \geq 0 \text{ e } x, y \leq la_r, \\ 1, & \text{se } x, y > la_r, \end{cases} \quad (6)$$

assumindo que $I(x, y) = 0$ define um pixel de cor preta na posição x, y , e $I(x, y) = 1$ torna o pixel daquela posição na cor branca.

No presente trabalho foram realizados diversos experimentos avaliando empiricamente cada parâmetro apresentado. Para os parâmetros apresentados, os seguintes intervalos foram considerados: $la_1 = \{1000, 1100, 1200, 1300, 1400, \dots, 2500, 2600\}$, $la_2 = \{500, 600, 700, 800, 900, \dots, 1500, 1600\}$, $la_r = \{80, 90, 100, 110, 120, \dots, 140, 150\}$. Isso foi feito para que fosse possível determinar qual o conjunto de parâmetros que maximiza as métricas estatísticas apresentadas.

Após computar o método proposto realizando a combinação de todos os valores possíveis dos parâmetros selecionados, foi observado que o conjunto de parâmetros apresentado é similar ao apresentado na Tabela 1.

3.2. Resultados e Discussão

A partir do conjunto de parâmetros da Tabela 1 e do método proposto, o presente trabalho será comparado com outros trabalhos importantes apresentados na literatura.

Na Tabela 2, são apresentados os resultados médios de métodos de segmentação automática das mamas, considerando as taxas estatísticas apresentadas em todas as ima-

Tabela 1. Conjunto de parâmetros que maximiza o resultado das métricas estatísticas consideradas.

Parâmetro	Valor
la_1	2200
la_2	150
la_r	130

gens disponibilizadas na base de dados. Observando os resultados obtidos, é possível concluir que o método apresentado obteve taxas médias altas para todas as medidas estatísticas analisadas. Alguns resultados dos trabalhos referidos são apresentados nas Figuras 5 e 6.

Tabela 2. Resultados normalizados obtidos pelos trabalhos apresentados na literatura e o método proposto.

Trabalhos Considerados	ACC	SEN	ESP	PDP	PDN
[Motta 2010]	0.96	0.88	0.99	0.96	0.96
[Marques 2012]	0.97	0.97	0.97	0.97	0.98
Método Proposto	0.96	0.98	0.95	0.95	0.98

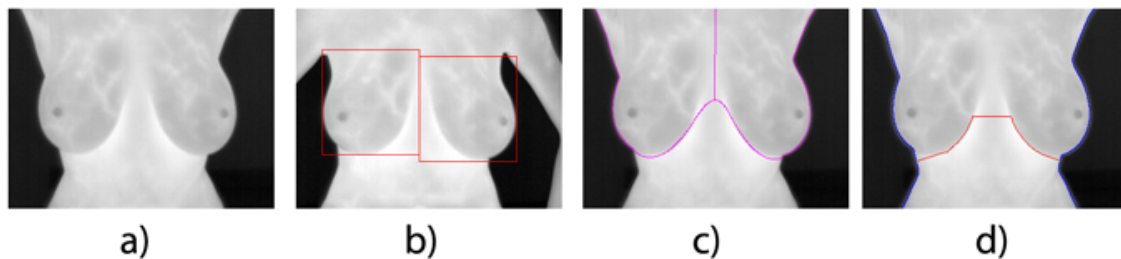


Figura 5. Resultados obtidos para a imagem IR_0100: a) Imagem original de entrada; b) Resultado obtido por [Motta 2010]; c) Resultado obtido por [Marques 2012]; d) Resultado adquirido pelo método proposto nesse trabalho.

Dos três trabalhos considerados, o método apresentado possui as maiores taxas médias de sensibilidade. Isso significa que, dos trabalhos avaliados, a solução proposta possui maior capacidade de identificação de amostras positivas (pertencentes a região de interesse). Essa diferença é superior ao trabalho de [Motta 2010] em uma margem de 10%.

A diferença de acurácia entre [Marques 2012] e o método proposto é de aproximadamente 0,5%, devido a introdução de amostras de falsos positivos na ROI. No entanto, nosso método é bem mais simples de ser implementado, possuindo menos etapas que os trabalhos anteriores. Essa simplicidade sem acarretar em discrepâncias nas métricas estatísticas torna o método muito competitivo e vantajoso para utilização na literatura e na sua utilização em diversas arquiteturas computacionais.

O método de [Marques 2012] possui tempos computacionais de execução muito similares ao trabalho proposto. Isso é visível na Tabela 3, em que os tempos médios em segundos de todos os métodos são apresentados. Se comparado a [Motta 2010], a

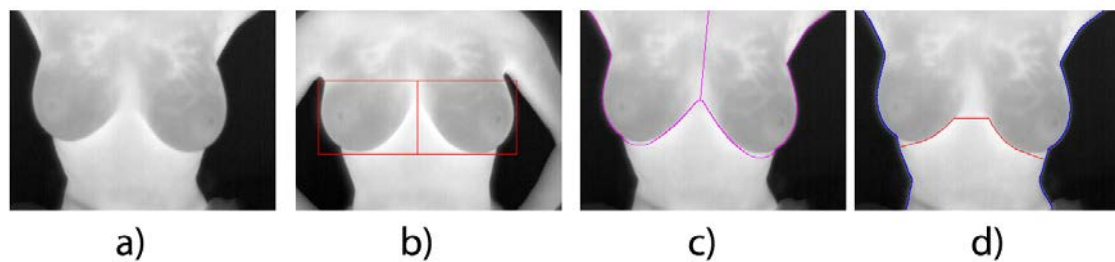


Figura 6. Resultados obtidos para a imagem IR_5986: a) Imagem original de entrada; b) Resultado obtido por [Motta 2010]; c) Resultado obtido por [Marques 2012]; d) Resultado adquirido pelo método proposto nesse trabalho.

diferença torna-se muito mais significativa, sendo aproximadamente 18 vezes mais lento do que o método apresentado nesse trabalho.

Tabela 3. Comparativo de tempo de execução entre os métodos comparados e o trabalho proposto.

Trabalhos Considerados	Tempo de Execução Médio (s)
[Motta 2010]	36,63
[Marques 2012]	1,55
Método Proposto	1,94

Para algumas imagens da base de dados, o método não apresenta resultados de maior qualidade. Isso é visível, por exemplo, quando as pacientes possuem pouco volume nas mamas. Isso torna a identificação da região das pregas inframamárias mais complexa. No entanto, como é visível na Figura 7, o impacto desse problema é menos significativo do que em alguns resultados de [Marques 2012] devido a abordagem de refinamento adaptativo, que iterativamente busca uma região de prega inframamária com maior contraste de temperatura, de acordo com a imagem de teste considerada.

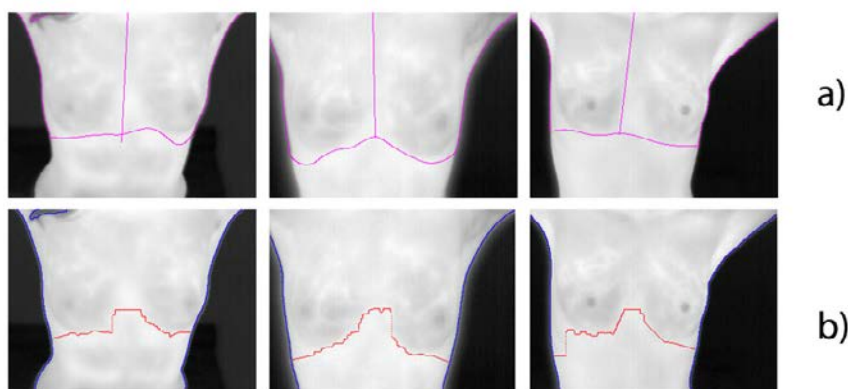


Figura 7. Imagens em que as pacientes possuem mamas de menor volume (respectivamente: IR_5451, IR_7446 e IR_5506): a) Resultados obtidos em [Marques 2012]; b) Resultados obtidos pelo método proposto nesse trabalho.

Outros problemas comuns são: a identificação inapropriada de uma dobra abdominal como parte da região das pregas inframamárias (Figura 8) e a assimetria en-

tre as mamas (Figura 9). No primeiro caso, a identificação de uma dobra do abdômen como parte das pregas inframamárias introduz uma grande quantidade de falsos positivos, que poderiam ser corrigidos se o ponto de interseção entre os seios fosse melhor assinalado. Com relação à assimetria, o resultado é superior a alguns exemplos apresentados em [Marques 2012], devido a premissa de se assumir que as pregas inframamárias localizam-se nos quadrantes inferiores da imagem e pelo fato de se extrair uma região de prega inframamária de maior densidade de pixels. No entanto, erros ainda ocorrem com certa frequência. É possível corrigi-los adicionando restrições que impeçam que a distância entre as regiões inferiores da mama seja tão contrastante.

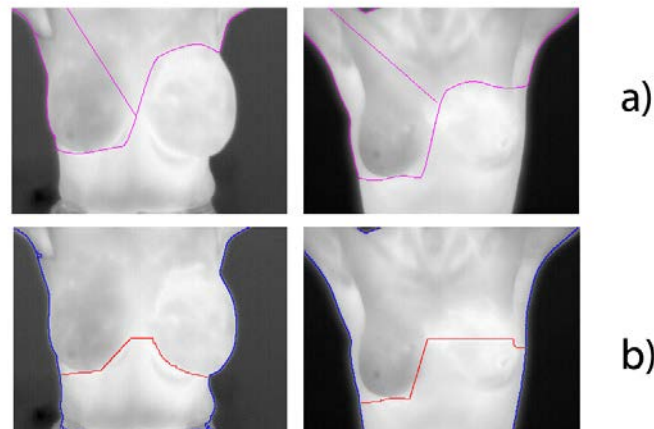


Figura 8. Imagens em que as pacientes possuem dobras abdominais passíveis de confusão com as pregas inframamárias (respectivamente: IR_0889 e IR_1330): a) Resultados obtidos em [Marques 2012]; b) Resultados obtidos pelo método proposto nesse trabalho.

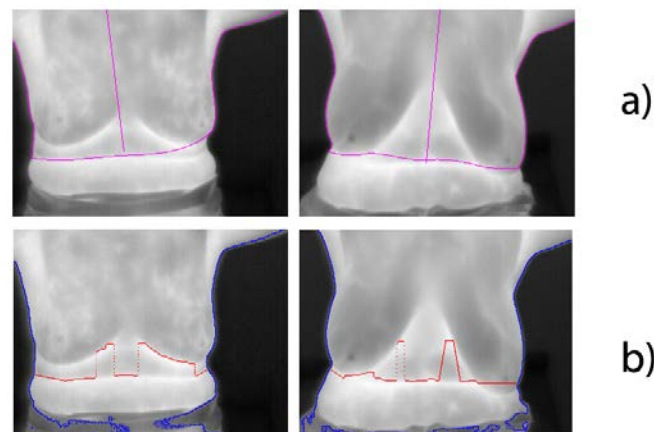


Figura 9. Imagens em que as pacientes possuem mamas com significativa assimetria (respectivamente: IR_8304 e IR_3786): a) Resultados obtidos em [Marques 2012]; b) Resultados obtidos pelo método proposto nesse trabalho.

4. Considerações Finais

A detecção precoce do câncer de mama é essencial para melhorar os prognósticos de cura da paciente. Sistemas CAD baseados na simetria da distribuição térmica das mamas se

mostram eficazes, baratos e pouco invasivos, podendo se tornar relevantes no diagnóstico de doenças da mama. No entanto, para que isso seja feito, torna-se necessário extrair a região de interesse que será avaliado por sistemas de diagnósticaçã de doenças.

A partir dos resultados apresentados, é visível que o método apresentado torna-se bastante competitivo para uso em situações reais, devido a acurácia do resultado final, bem como o tempo de execução médio que é relativamente baixo. Além disso, a metodologia é fácil de ser desenvolvida computacionalmente, pois possui muito menos etapas que os outros trabalhos da literatura e requer poucos recursos computacionais (tanto de hardware quanto de software) para o seu funcionamento.

Para trabalhos futuros, espera-se:

- Adaptar o método para imagens com câmeras em ângulos distintos;
- Utilizar técnicas para extração de atributos e características da mama para diagnóstico de doenças. Esses atributos podem ser geométricos ou de textura;
- Desenvolver sistemas CAD para diagnóstico de doenças diversas da mama a partir do algoritmo de segmentação proposto.

Referências

- Arora, N. M., Martins, D., Ruggerio, D., Tousimis, E., Swistel, A. J., Osborne, M. P., and Simmons, R. M. (2008). Effectiveness of a noninvasive digital infrared thermal imaging systems in the detection of breast cancer. *American Journal of Surgery*, 196(4):523–526.
- Gore, J. P. and Xu, L. X. (2014). *Thermal Imaging for Biological and Medical Diagnostics*. CRC Press.
- HCB (2015). Informação: Saiba quais são os tipos de câncer mais comuns no Brasil. <http://www.hcancerbarretos.com.br/82-institucional/noticias-institucional/1300-informacao-saiba-quais-sao-os-tipos-de-cancer-mais-comuns-no-brasil>. Acesso em: 29 de outubro de 2015.
- INCA (2015). Tipos de câncer. <http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama>. Acesso em: 07 de novembro de 2015.
- Li, X., Aldridge, B., Fisher, R. B., and Rees, J. (2011). Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to skin lesion segmentation. In *ISBI*, pages 1438–1441. IEEE.
- Marques, R. S. (2012). Segmentação automática das mamas em imagens térmicas. Master's thesis, Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil.
- Motta, L. S. (2010). Obtenção automática da região de interesse em termogramas frontais da mama para o auxílio à detecção precoce de doenças. Master's thesis, Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil.
- Silva, L. F., Saade, D. C. M., Sequeiros-Oliveira, G. O., Silva, A. C., Bravo, R. S., and Conci, A. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100.

Um Estudo sobre Características Extraídas de Imagens Infravermelhas para Detecção de Anormalidades de Mama

Lincoln F. Silva¹, Roger Resmini¹, Aristófanés C. Silva³, Anselmo C. Paiva⁴,
Renato S. Bravo², Alair Augusto S. M. D. Santos²,
Débora C. Muchaluat-Saade¹ e Aura Conci¹

¹Instituto de Computação – Universidade Federal Fluminense (UFF)

²Centro de Ciências Médicas – Universidade Federal Fluminense (UFF)

³Departamento de Informática – Universidade Federal do Maranhão (UFMA)

⁴Departamento de Engenharia de Eletricidade – Universidade Federal do Maranhão (UFMA)

{lsilva, rresmini, debora, aconci}@ic.uff.br

rsbravo@predialnet.com.br, alairsarmet@globo.com

paiva@deinf.ufma.br, ari@dee.ufma.br

Abstract. *Breast cancer is the second most common type of cancer worldwide. Diagnosis and treatment in early stages increase the chances of cure of the patient. The temperature of cancerous tissue is generally higher than that of healthy surrounding tissues, making thermography an option to be considered in screening strategies of this cancer type. This paper proposes a set of features extracted from images obtained by Dynamic Infrared Thermography in order to detect breast abnormalities, including cancer, using unsupervised machine learning techniques. Clustering validation Indexes are applied to evaluate formed groups of temperature time series for various k value applied in the algorithm k -means, generating values treated as features. The analysis reveals that the Strehl index generates the most significant features.*

Resumo. *O câncer de mama é o segundo tipo de câncer mais comum no mundo. O diagnóstico e o tratamento em estágios iniciais aumentam as chances de cura da paciente. A temperatura de tecidos cancerosos é geralmente mais alta do que a de tecidos vizinhos saudáveis, tornando a termografia uma opção a ser considerada em estratégias de rastreamento deste tipo de câncer. Este artigo propõe um conjunto de características extraídas de imagens obtidas por Termografia Infravermelha Dinâmica com o objetivo de detectar anomalias da mama, entre elas o câncer, utilizando técnicas de aprendizagem de máquina não supervisionada. Índices de validação de agrupamento são aplicados para avaliar os grupos formados de séries temporais de temperatura para vários valores de k aplicados no algoritmo k -means, gerando valores tratados como características. As análises revelam que o índice Strehl gera as características mais significativas.*

1. Introdução

O câncer de mama é o câncer mais comum entre as mulheres no mundo. Porém, quando identificado e tratado em estágios iniciais, esse tipo de câncer possui relativamente um

bom prognóstico. Rastreamento é uma estratégia adotada por autoridades de saúde com o intuito de identificar mulheres que estão em estágios iniciais dessa doença de mama. Porém, todos os tipos de exames, incluindo a mamografia, considerada o padrão ouro na detecção de câncer, possuem suas limitações. A mamografia, por exemplo, possui alta taxa de classificação de falsos positivos, efetividade insuficiente em mamas densas e usa radiação ionizante para formar as imagens. Visto que a temperatura de tecidos cancerosos é geralmente maior do que a de tecidos saudáveis, a termografia tem sido considerada um método de rastreamento promissor para a detecção do câncer de mama, por gerar imagens que revelam a distribuição de calor sobre a superfície de ambas as mamas [Silva 2015].

A termografia é classificada basicamente em dois tipos: *Termografia Infravermelha Estática* (TIE) e *Termografia Infravermelha Dinâmica* (TID). As séries temporais usadas neste trabalho são obtidas a partir da TID, um método para monitorar a variação da temperatura da pele no tempo, após estresse térmico. O tipo de estresse térmico mais utilizado, em vários protocolos de execução da TID, é a aplicação de fluxo de ar direcionado às mamas utilizando um ventilador elétrico [Silva 2015]. O resfriamento das mamas, teoricamente, melhora o contraste térmico entre tecidos saudáveis e doentes na imagem, pois vasos sanguíneos gerados em função do tumor canceroso não possuem camada muscular e nem regulação neural como vasos embrionários [Amalu 2004]. Esses vasos são somente tubos endoteliais e portanto não contraem em resposta à estimulação simpática. Por essa razão, as regiões da mama com tumores cancerosos permanecem com temperatura quase inalterada, enquanto que a parte saudável da mama é resfriada [Amalu 2004]. A TID é mais rápida e mais robusta quando comparada à TIE, que requer condições rígidas de ambiente e tempo significativamente longo para aclimação da paciente às condições da sala de exame. A TIE apenas registra a distribuição da temperatura na superfície da pele, não considerando sua variação no tempo. Por outro lado, a TID é menos dependente das condições e temperatura da sala de exame [Herman 2013]. Além disso, a Equação 1 [Gavrioloia et al. 2009] torna nítida a importância da variação da temperatura no tempo, onde k é a condutividade térmica do tecido, ρc e c_b são a densidade e o calor específico do sangue, w_b é a taxa de perfusão do sangue ($ml/s/ml$), P é a taxa de geração de calor metabólico (w/m^3), T_b é a temperatura do sangue arterial, e T é a temperatura local do tecido. Observa-se nessa equação que existem termos de primeira derivada participial em relação ao tempo e de segunda derivada em relação a posição, mas não temperaturas constantes. Isso sugere a importância da variação da temperatura ao logo do tempo e, conseqüentemente, da análise dinâmica.

$$\rho c \frac{\partial T}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 k \frac{\partial T}{\partial r} \right) + P + w_b c_b (T_b - T) \quad (1)$$

Em um trabalho anterior, aplicamos classificadores tendo como entrada características extraídas de séries temporais de temperatura para identificar pacientes com câncer de mama. Os detalhes da metodologia estão em [Silva et al. 2015a]. Resumidamente, as imagens obtidas por TID de uma determinada paciente são primeiramente registradas. Em seguida, a região das mamas é dividida em quadrados pequenos e a temperatura máxima de cada um desses quadrados é observada em todas as imagens da paciente gerando, assim, as séries temporais de temperatura. Após, o algoritmo *k-means* é aplicado sobre tais séries formando k grupos ($2 \leq k \leq 10$). Índices de avaliação de agrupamento (ou validade de grupo) são aplicados para avaliar o resultado do agru-

pamento produzido pelo algoritmo *k-mean* [MacQueen 1967] para cada valor de k . Os valores obtidos são tratados como características e submetidos à fase de seleção de características. Por fim, técnicas de aprendizagem de máquina são aplicadas para gerar o modelo de classificação tendo como entrada as características selecionadas. A evolução dessa metodologia nos últimos anos pode ser acompanhada em outros trabalhos do grupo em [Galvão et al. 2013] [Silva et al. 2014a][Silva et al. 2015b][Silva et al. 2016].

Neste artigo discutimos etapas específicas da metodologia para investigar os fatores que determinam a permanência de apenas alguns índices de avaliação de agrupamento no conjunto de características selecionadas. Para isso, cada índice foi analisado separadamente com o auxílio de gráficos para verificar o quanto o mesmo é discriminativo em relação a pacientes com e sem anormalidades de mama. Na análise de cada índice, todos os valores de k ($2 \leq k \leq 10$) foram considerados, um por vez.

O restante deste artigo está organizado como segue. Na segunda seção, os principais trabalhos relacionados da literatura são discutidos. O metodologia proposta para análise das características extraídas é detalhado na terceira seção. Conclusões e trabalhos futuros são apresentados na última seção.

2. Trabalhos relacionados

Como uma revisão atualizada em TID de mama é apresentada em [Borchardt et al. 2013], esta seção focará somente em trabalhos computacionais mais recentes utilizando essa modalidade de exame.

No trabalho de Gerasimova *et al.* [Gerasimova et al. 2013], uma análise multifractal das séries temporais, geradas a partir da TID, foi executada para verificar a diferença de comportamento entre tecido com tumor maligno e tecido saudável. O método de máximo módulo da transformada de wavelet foi aplicado para caracterizar as propriedades multifractais de séries temporais oriundas de mamas cancerosas e saudáveis. Os autores concluíram que as propriedades escalares multifractais complexas, observadas em séries temporais sobre regulação automática (mamas saudáveis) são drasticamente alteradas na existência de doenças (mamas cancerosas). Nesse estudo, as duas mamas de 9 mulheres foram examinadas, 6 com câncer e 3 saudáveis. A metodologia conseguiu distinguir região com tumor e região saudável. Para mamas saudáveis, foi encontrada uma dimensão multifractal como característica de uma contínua mudança na forma da função de densidade de probabilidade de variação de temperatura através do tempo. Entretanto, séries temporais de temperatura da mama com tumor maligno mostraram estatísticas de variação de temperatura monofractal homogênea como evidência da perda de complexidade. As análises foram realizadas visualmente por meio de gráficos e tabelas.

Recentemente, Gerasimova *et al.* publicou outro trabalho [Gerasimova et al. 2014] usando uma base de dados maior, 33 pacientes com câncer de mama histopatologicamente confirmado e 14 voluntárias saudáveis para controle. Os achados reafirmaram os resultados de [Gerasimova et al. 2013].

Saniei *et al.* [Saniei et al. 2015] desenvolveram uma metodologia para análise de termografias obtidas por TID onde apenas duas imagens são usadas. Características são extraídas dos padrões vasculares e técnicas semelhantes às usadas para reconhecimento de impressões digitais são aplicadas para comparar tais padrões vasculares, dos termogramas

antes e após o estresse térmico. Para quantificar o grau de similaridade entre os padrões vasculares nas duas imagens, um valor de pontuação de relacionamento é gerado por meio de uma expressão matemática e esses padrões serão considerados correspondentes se o valor de pontuação estiver abaixo de um determinado valor de limiarização, que é definido empiricamente. Segundo os autores, a metodologia alcançou sensibilidade de 86% e especificidade de 61%. Foram usadas imagens de 50 pacientes, 25 com e 25 sem câncer de mama.

3. Metodologia proposta

O fluxograma da metodologia proposta está na Figura 1. Primeiramente as séries temporais de temperatura são construídas, em seguida as características são extraídas e por último o estudo dessas características é realizado.

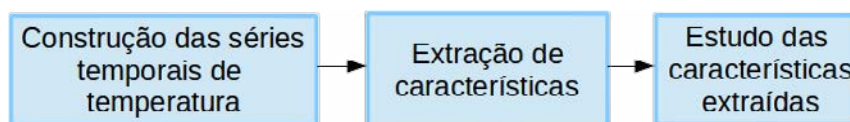


Figura 1. Fluxograma das etapas da metodologia proposta.

3.1. Construção das séries temporais de temperatura

Logo após a etapa de registro das imagens térmicas adquiridas por TID de uma paciente, detalhado em [Silva et al. 2015a], a construção das séries temporais de temperatura segue os seguintes passos:

1. a ROI (Region of Interest) no termograma, correspondendo aos pixels de cor branca na máscara (Figura 2 (a)), é dividida em uma “malha” de quadrados R_k de tamanho 11x11 pixels (Figura 2 (b)), com $k = 1, 2, \dots, p$, onde p é a quantidade de quadrados formados;
2. a temperatura mais alta de cada quadrado R_k é observada em todos os vinte termogramas da paciente (Figura 2 (c)), produzindo a série temporal $S_k = (t_{k,1}; t_{k,2}; \dots; t_{k,20})$, onde $t_{k,1}$ é a temperatura mais alta da região quadrada R_k no primeiro termograma da sequência, $t_{k,2}$ é a temperatura mais alta da região quadrada R_k na segunda imagem da sequência, e assim sucessivamente.

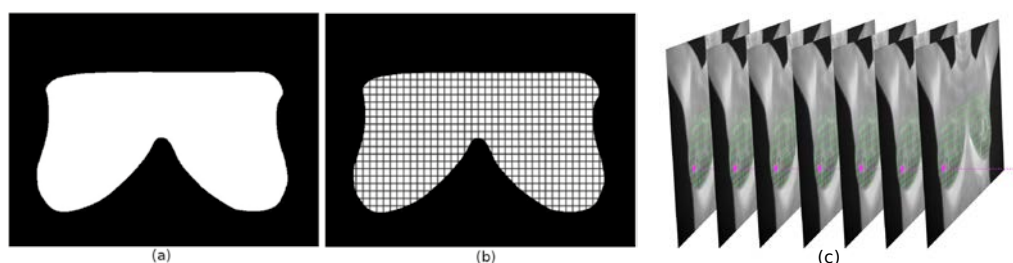


Figura 2. Máscara dividida em uma “malha” de quadrados de tamanho 11x11 pixels e a construção de uma série temporal de temperatura.

A Figura 3 mostra as séries temporais de temperatura de uma paciente saudável em (a) e as séries temporais do mesmo tipo de uma paciente com câncer de mama em

(b). Em cada gráfico: o eixo x representa o tempo (os 20 pontos de cada série); o eixo y contém o índice de cada um das séries temporais de temperatura; e o eixo z indica a temperatura em graus Celsius. É possível observar que existe um grupo de séries temporais com temperaturas maiores (cor vermelha) e com inclinação maior nos momentos iniciais da recuperação da temperatura, após o estresse térmico, para a paciente doente. O mesmo não é verdadeiro para as séries temporais da paciente saudável. As características extraídas buscam detectar as séries temporais de temperaturas que se destacam das demais para que seja possível indicar pacientes com anormalidade de mama.

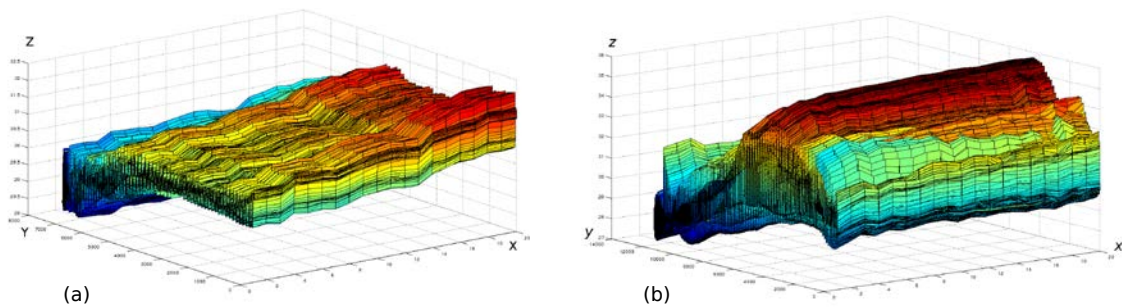


Figura 3. Em (a), sinais térmicos de uma paciente saudável, e em (b), sinais térmicos de uma paciente com câncer.

3.2. Extração de características

Dada uma paciente qualquer, as suposições são seguintes:

1. Se essa paciente for saudável, as séries temporais construídas a partir de cada ponto da mama são semelhantes (como na Figura 3 à esquerda), ou seja, possuem um grau alto de similaridade e então um agrupamento formado por tais séries temporais possui as seguintes características:
 - i) os grupos formados não são compactos;
 - ii) os grupos formados são próximos (são semelhantes) uns dos outros, em relação a uma medida de similaridade.
2. Por outro lado, se essa paciente for doente, para as séries temporais construídas a partir de cada ponto da mama, existe um grupo que se destaca dos demais por possuir séries temporais com comportamento diferente (como na Figura 3 à direita), e então um agrupamento formado por tais séries temporais possui as seguintes características:
 - i) os grupos formados são mais compactos, se comparados aos da suposição 1.;
 - ii) os grupos formados são menos próximos (são menos semelhantes) uns dos outros, em relação a uma medida de similaridade, se comparados aos da suposição 1.

Para medir a compacidade de cada grupo formado e a distância entre eles, dentro de um agrupamento formado pelo algoritmo *k-means*, medidas (índices) de validação de agrupamento foram aplicados na esperança de que essas medidas retornassem valores diferentes entre pacientes doentes e saudáveis.

Na literatura existem vários índices de validação, porém neste trabalho foram aplicados apenas os direcionados a resultados de algoritmos de agrupamento rígido dos dados (cada elemento deve pertencer a um, e somente um, grupo). São eles: Silhouette [Rousseeuw 1987]; Davies-Bouldin [Davies and Bouldin 1979]; Calinski-Harabasz [Caliński and Harabasz 1974]; Dunn [Bolshakova and Azuaje 2003]; Krzanowski-Lai [Krzanowski and Lai 1988]; Hartigan [Hartigan 1985]; Homogeneity [Chen et al. 2002]; Separation [Chen et al. 2002]; Hubert-Levin (C-index) [Bolshakova and Azuaje 2006]; e Strehl [Strehl and Ghosh 2000].

Para uma determinada paciente i , as séries temporais construídas na etapa anterior formam o conjunto, aqui denominado, X_i . Para separar tais grupos (de regiões saudáveis e de regiões com anomalias), uma técnica de aprendizagem de máquina não-supervisionada é aplicada por meio de um algoritmo de agrupamento. O algoritmo escolhido para essa tarefa é o k -means [Han and Kamber 2006]. O k -means é executado, tendo como entrada os elementos do conjunto X_i , nove vezes, ou seja, para cada valor de $k \in \{2, 3, \dots, 10\}$, gerando o conjunto de resultados de agrupamento $P_i = \{P_{i,2}, P_{i,3}, \dots, P_{i,10}\}$, onde $P_{i,2}$ contém 2 grupos de séries temporais, $P_{i,3}$ contém 3 grupos de séries temporais e assim sucessivamente até $P_{i,10}$, que contém 10 grupos. A medida de proximidade utilizada no k -means é a *distância Euclidiana*.

O processamento ocorre na seguinte sequência:

1. o algoritmo k -means é executado para $k = 2$ sobre o conjunto X_i , gerando o agrupamento $P_{i,2}$;
2. os 10 índices de validação de agrupamento, citados anteriormente, são aplicados para medir a qualidade dos grupos formados no agrupamento $P_{i,2}$, ou seja, para medir o quão compactos e bem separados os grupos são, e os resultados são armazenados no vetor V_i ;
3. o valor de k é atualizado para $k + 1$ e os passos 1. e 2. são executados novamente.

O processo descrito ocorre até o valor de k igual a 10 ($k = 10$), e ao final o vetor V_i contém 90 valores, ou seja, $V_i = (v_{i,1}, v_{i,2}, \dots, v_{i,90})$. Todos os valores do vetor V_i estão no intervalo $[0, 1] \subset \mathbb{R}$, ou seja, $v_{i,j} \in [0, 1] \subset \mathbb{R}$, $\forall j \in [2, 10] \subset \mathbb{N}$.

3.3. Estudo das características extraídas

Em relação aos índices de validação de agrupamento, o mais discriminativo, ou seja, o que mais esteve presente entre as características selecionadas, na fase de seleção de características, foi o *Strehl*. Isso aconteceu para $k \in \{3, 4, 6, 7, 8, 9, 10\}$, onde k é o número de grupos formados pelo k -means. Além dele, o índice *Hubert-Levin* (para $k \in \{3, 8, 10\}$), o índice *Homogeneidade* (para $k \in \{2, 7\}$) e o índice *Silhueta* (para $k = 3$) também estiveram presentes entre as características selecionadas. A seleção de características foi auxiliada pelo Auto-Weka [Thornton et al. 2013], que é uma ferramenta de mineração de dados usada para resolver o problema de seleção de algoritmos e otimização de parâmetros em tarefas de classificação. O método de seleção de características e seus parâmetros, recomendados pelo Auto-Weka, foram usados no WEKA para a seleção das características.

Uma investigação foi realizada para averiguar o motivo da superioridade do índice Strehl sobre os outros índices na tarefa de separar as pacientes doentes das saudáveis, o motivo que fez com que ele aparecesse tantas vezes no vetor das características selecionadas. Cada índice, para cada valor de k , foi analisado através de gráficos de linha

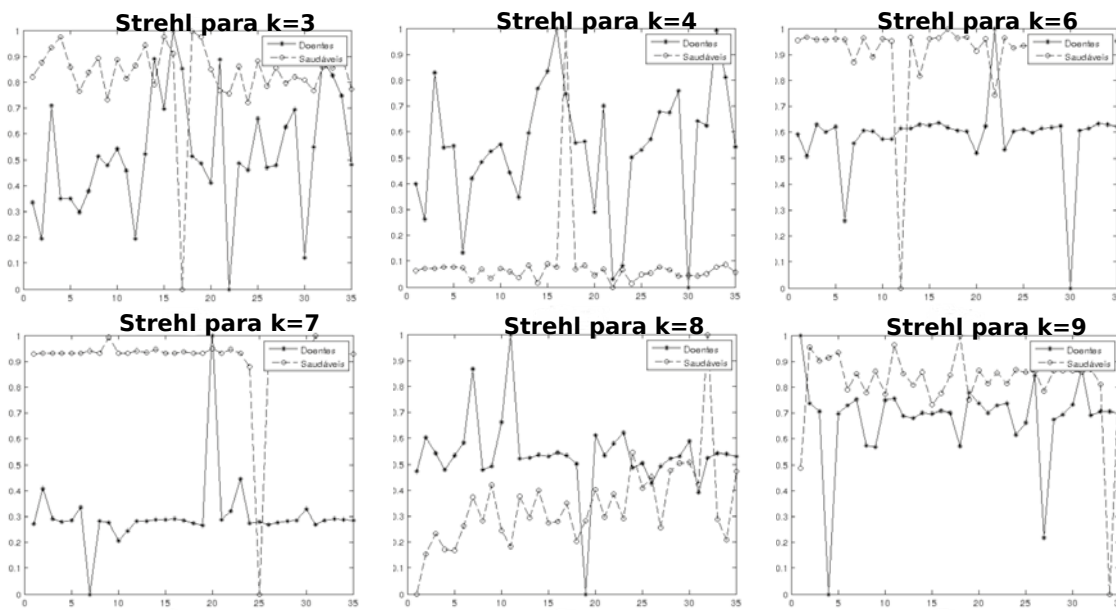


Figura 4. Cálculo do índice Strehl para $k = 3, 4, 6, 7, 8, 9$.

construídos a partir de pontos que representavam os valores do índice, calculado para cada paciente. Observando esses gráficos foi possível perceber que o índice Strehl de fato consegue praticamente separar os dois tipos de pacientes (doente e saudável), para quase todos os valores de k . Algo semelhante não aconteceu para os outros índices testados. Para ilustrar, a Figura 4 contém os gráficos com o cálculo do índice Strehl para $k \in \{3, 4, 6, 7, 8, 9\}$, ou seja, características que estiveram entre as selecionadas. Nos gráficos, o eixo horizontal representa o número (índice) das pacientes doentes e saudáveis, o eixo vertical é a escala de valores reais do índice Strehl, a linha contínua representa as pacientes doentes e a linha tracejada, as pacientes saudáveis. Assim, utilizando os valores desse índice quando ele é usado para avaliar os agrupamentos formados por 3, 4, 6, 7, 8 e 9 grupos de séries temporais é possível praticamente separar as pacientes doentes das saudáveis.

O índice Strehl utiliza as similaridades intragrupo e intergrupo para medir a qualidade dos k grupos formados com os n elementos de um conjunto de dados X . A similaridade intragrupo representa a similaridade média entre os elementos de um grupo $G_i \subset G$ (G é o resultado do agrupamento dos elementos de X e $i \in 1, 2, \dots, k$) e é definida pela Equação 2:

$$\text{intra}(G_i) = \frac{1}{(g_i - 1)g_i} \sum_{x,y \in G_i} d(x,y) \quad (2)$$

onde g_i é o número de elementos do grupo G_i e $d(x,y)$ é a distância entre dois elementos de G_i . Essa equação não define a similaridade intragrupo para grupos vazios ou com um único elemento. Por definição, a similaridade intragrupo para grupo com um único elemento é igual a 1 (um).

A similaridade intergrupo entre grupos não vazios G_i e G_j , com $i \neq j$ é definida

pela Equação 3:

$$inter(G_i, G_j) = \frac{1}{g_i g_j} \sum_{x \in G_i, y \in G_j} d(x, y) \quad (3)$$

onde g_i e g_j são o número de elementos do grupo G_i e do grupo G_j , respectivamente.

O objetivo é maximizar a similaridade intragrupo e minimizar a similaridade intergrupo. Assim, a qualidade de agrupamento G é definida como a razão entre o total das similaridades intragrupo e o total das similaridades intergrupo. O total das similaridades intragrupo é o somatório de cada uma das similaridades intragrupo ponderada por $g_i - 1$, pois a auto-similaridade (o qual é sempre 1) não deve ser contabilizada. O total das similaridades intergrupo é obtido pelo somatório de cada similaridade $inter(G_i, G_j)$ ponderada por g_i .

$$\frac{\sum_{i=1}^k \frac{g_i - 1}{n - k} intra(G_i)}{\sum_{i=1}^k \sum_{j=i+1}^k \frac{g_i}{n} inter(G_i, G_j)} \quad (4)$$

O valor da razão 4 está dentro do intervalo $[0, \infty)$.

Sendo G o resultado de um agrupamento dos elementos em X , a qualidade de G é definida pela Equação 5:

$$\Gamma(G) = 1 - \frac{(n - k) \sum_{i=1}^k \sum_{j=i+1}^k g_i inter(G_i, G_j)}{n \sum_{i=1}^k (g_i - 1) intra(G_i)} \quad (5)$$

onde $\Gamma(G) \in [0, 1]$

A eficiência do índice Strehl é consequência de sua formulação, que pondera as similaridades intragrupo e intergrupos pelo tamanho de cada grupo envolvido no cálculo de tais similaridades, visto que, dependendo do tamanho das mamas da paciente, para cada uma (paciente) teremos mais ou menos séries temporais construídas. Isso não acontece para os demais índices, ou seja, as similaridades intragrupo e intergrupo são calculadas sem levar em conta o tamanho de cada grupo.

As imagens infravermelhas usadas neste trabalho são da *Database for Mastology Research with Infrared Image - DMR-IR* (<http://visual.ic.uff.br/dmi/>). Essa base de dados e o protocolo de aquisição das imagens são descritos em trabalhos anteriores do grupo [Silva et al. 2014b][Silva et al. 2013]. A aquisição das imagens e o uso delas em pesquisas está registrado no Ministério da Saúde sob o número CAAE: 01042812.0.0000.5243.

4. Conclusões

Regiões da mama com câncer produzem séries temporais de temperatura com alteração de comportamento quando observados em função do tempo. Este trabalho realizou um estudo de características presentes na detecção dessas alterações usando aprendizagem de máquina não supervisionada. Os resultados mostraram que o índice Strehl, para vários valores de k aplicado no algoritmo *k-means*, gera as características mais significativas na discriminação de mamas saudáveis e mamas com alguma anomalia. Como trabalho futuro, métodos de classificação serão testados apenas com as características geradas a partir do índice Strehl para indicar pacientes com alguma anormalidade de mama. Isso gerará uma diminuição de tempo e de processamento computacional, quando comparado ao nosso trabalho anterior, para realizar tal tarefa.

Referências

- Amalu, W. C. (2004). Nondestructive testing of the human breast: The validity of dynamic stress testing in medical infrared breast imaging. *Engineering in Medicine and Biology Society*, 1:1174–1177.
- Bolshakova, N. and Azuaje, F. (2003). Cluster validation techniques for genome expression data. *Signal Processing*, 83(4):825–833.
- Bolshakova, N. and Azuaje, F. (2006). Estimating the number of clusters in dna microarray data. *Methods of Information in Medicine*, 45(2):153–157.
- Borchardt, T. B., Conci, A., Lima, R. C. F., Resmini, R., and Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, 93:2785–2803.
- Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27.
- Chen, G., Jaradat, S. A., Banerjee, N., Tanaka, T. S., Ko, M. S. H., and Zhang, M. Q. (2002). Evaluation and comparison of clustering algorithms in analyzing es cell gene expression data. *Statistica Sinica*, 12:241–262.
- Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-1(2):224–227.
- Galvão, S. S. L., Conci, A., Galvão, S. S. L., and Silva, L. F. (2013). Registro de imagens afim para o protocolo dinâmico de aquisição de imagens térmicas da mama. In *IV Encontro Nacional de Engenharia Biomecânica: ENEBI 2013*, pages 158–159, Vitória, ES, Brazil.
- Gavriloaia, G. V., Hurduc, A., Ghimigean, A.-M., and Fumarel, R. (2009). Spatial-temperature high resolution map for early cancer diagnosis. *Proc. SPIE*, 7171:71710W–71710W–8.
- Gerasimova, E., Audit, B., Roux, S., Khalil, A., Gileva, O., Argoul, F., Naimark, O., and Arneodo, A. (2014). Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. *Frontiers in Physiology*, 5(176).
- Gerasimova, E., Audit, B., Roux, S. G., Khalil, A., Argoul, F., Naimark, O., and Arneodo, A. (2013). Multifractal analysis of dynamic infrared imaging of breast cancer. *EPL (Europhysics Letters)*, 104(6):68001–p1–68001–p6.
- Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, Burlington, MA 01803, USA, 2nd ed edition.
- Hartigan, J. (1985). Statistical theory in clustering. *Journal of Classification*, 2(1):63–76.
- Herman, C. (2013). The role of dynamic infrared imaging in melanoma diagnosis. *Expert Review of Dermatology*, 8(2):177–184.
- Krzanowski, W. J. and Lai, Y. T. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44(1):23–34.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics*

- and Probability, Volume 1: Statistics, pages 281–297, Berkeley, Calif. University of California Press.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0):53–65.
- Saniei, E., Setayeshi, S., Akbari, M. E., and Navid, M. (2015). A vascular network matching in dynamic thermography for breast cancer detection. *Quantitative InfraRed Thermography Journal*, 12(1):24–36.
- Silva, L. F. (2015). *Uma Análise Híbrida para Detecção de Anomalias da Mama usando Séries Temporais de Temperatura*. PhD thesis, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.
- Silva, L. F., Marques, R. S., Carvalho, G. S., Santos, M. L. O., Fontes, C. A. P., Santos, A. A. S. M. D., and Conci, A. (2013). Protocolo de captura de imagens térmicas da mama para construção de um banco público de exames. In *IV Encontro Nacional de Engenharia Biomecânica: ENEBI 2013*, pages 104–105, Vitória, ES, Brazil.
- Silva, L. F., Olivera, G. O. S., Borchardt, T. B., Resmini, R., Santos, A. A. S. M. D., Fontes, C. A. P., Muchaluat-Saade, D. C., and Conci, A. (2015a). Uma análise híbrida para identificação de cancer de mama usando sinais térmicos. *WIM-XV Workshop de Informática Médica-Anais CSBC*.
- Silva, L. F., Olivera, G. O. S., Galvão, S. S. L., Silva, J. B., Santos, A. A. S. M. D., Muchaluat-Saade, D. C., and Conci, A. (2014a). Análise de séries temporais de sinais térmicos da mama para detecção de anomalias. *WIM-XIV Workshop de Informática Médica-Anais CSBC*, pages 1818–1827.
- Silva, L. F., Saade, D. C. M., Sequeiros, G. O., Silva, A. C., Paiva, A. C., Bravo, R. S., and Conci, A. (2014b). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100.
- Silva, L. F., Santos, A. A. S., Bravo, R. S., Silva, A. C., Muchaluat-Saade, D. C., and Conci, A. (2016). Hybrid analysis for indicating patients with breast cancer using temperature time series. *Computer Methods and Programs in Biomedicine*, 130:142–153.
- Silva, L. F., Sequeiros, G., Santos, M. L., Fontes, C., Muchaluat-Saade, D. C., and Conci, A. (2015b). Thermal signal analysis for breast cancer risk verification. *MEDINFO 2015: eHealth-enabled Health, Ebook Series: Studies in Health Technology and Informatics*, 216:746–750.
- Strehl, A. and Ghosh, J. (2000). Value-based customer grouping from large retail datasets. *Proc. SPIE Conf. Data Mining Knowl. Discov.*, pages 33–42.
- Thornton, C., Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2013). Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 847–855.

Qualidade de Software do Maior Sistema de Teletoxicologia Utilizado no Brasil

João M. Alves^{1,2}, Christiane G. von Wangenheim¹, Alexandre Savaris²,
Aldo von Wangenheim^{1,2}

¹ Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Santa Catarina (UFSC)
Florianópolis, Santa Catarina, Brasil

²Laboratório de Telemedicina
Instituto Nacional para Convergência Digital (INCoD)
Florianópolis, Santa Catarina, Brasil

{joao,savaris}@telemedicina.ufsc.br, {gresse,awangenh}@inf.ufsc.br

Abstract. *Teletoxicology Systems are information systems that help physicians and toxicologists to attend cases of patient exposure to poisonous agents. They provide information about poisoning and poisoning case management and support toxicovigilance activities. There is still no consensus on how to assess the quality of such systems. This work performs a case study applying a customized software quality evaluation model – the AdEQUATE model – in order to assess the software quality degree of a teletoxicology system widely used in Brazil, the DATATOX information system. Analysing the specificities of DATATOX we customized AdEQUATE, a model for evaluating the quality of telemedicine systems based on the ISO/IEC 25010 standard through the use of a questionnaire. The acquired data were analysed employing descriptive statistics. Overall, the DATATOX information system has been perceived as having a good quality, specifically for effectiveness, efficiency, and satisfaction. Nevertheless, few characteristics have been considered of low quality, including reliability and compatibility. This work also attests the feasibility on using AdEQUATE as a tool for software quality evaluation.*

Resumo. *Sistemas de teletoxicologia são sistemas de informação que auxiliam médicos e toxicologistas a registrar casos de exposição de pacientes a agentes tóxicos. Tais sistemas fornecem informações sobre intoxicações, gerenciam casos de intoxicações e providenciam atividades de vigilância toxicológica. Contudo, ainda não há um consenso em como avaliar a qualidade de tais sistemas. Este trabalho realiza um estudo de caso aplicando um modelo de avaliação de qualidade customizado – o AdEQUATE – com o objetivo de determinar o grau de qualidade de software de um sistema de teletoxicologia largamente usado no Brasil, o sistema DATATOX. Analisando as especificidades do DATATOX, customizou-se o AdEQUATE, um modelo para avaliação de qualidade de sistemas de telemedicina baseado na norma ISO/IEC 25010 através do uso de um questionário. Os dados coletados foram analisados usando estatística descritiva. No geral, o sistema DATATOX foi considerado como tendo uma boa qualidade, especificamente em termos de eficácia, eficiência e satisfação. Contudo, algumas características foram consideradas como de baixa qualidade,*

incluindo confiança e compatibilidade. Este trabalho também comprova a viabilidade de usar o AdEQUATE como ferramenta para avaliação de qualidade de software.

1. Introdução

O número de intoxicações reportadas aos Centros de Informação Toxicológica (CITs) do Rio Grande do Sul e de Santa Catarina aumentou em 46% entre 2005 e 2013, sendo que mais de 147.000 casos foram causados por drogas de abuso ou animais venenosos [CIT/RS 2015, CIT/SC 2015]. No Reino Unido, anualmente, aproximadamente 170.000 das entradas no *National Health Service* são relacionadas a intoxicações. Esse volume é bastante expressivo e sobrecarrega o pessoal de hospitais e de centros de emergência [Perry et al. 2014].

Para auxiliar pacientes e centros hospitalares, os CITs têm a responsabilidade de fornecer informações e aconselhamento, gerenciamento de casos de intoxicação ou envenenamento, serviços de análise clínica, vigilância, pesquisa, educação e treinamento na prevenção de intoxicações [World Health Organization et al. 1997]. Muitos países além do Brasil mantêm CITs que funcionam em regime 24/7, como os Estados Unidos [Mowry et al. 2015] e as Ilhas Britânicas [Tormey and Moore 2013]. Fornecendo serviços para uma vasta gama de *stakeholders* (de um médico em uma unidade de tratamento intensivo a um fazendeiro que acabou de ser picado por uma cobra em seu sítio), a maior parte das interações ocorre por telefone; contudo, alguns CITs providenciam interação *on-line*, como o *Web Poison Control* [Poison Control Centers 2015] que fornece um auxílio limitado para vítimas de intoxicações via Internet.

Na ausência de um auxílio toxicológico imediato, teletoxicologia – uma especialização da telemedicina que trata do fornecimento de serviços toxicológicos à distância – é uma opção válida [Skolnik 2013]. Contudo, não existe um consenso sobre o conjunto de funcionalidades que constitui um sistema de teletoxicologia. Alguns sistemas como o TOXBASE [Bateman et al. 2002] fornecem ontologias para a classificação de agentes tóxicos; outros, como o *American National Poison Data System* [Seifert et al. 2013], registram casos de exposição de pacientes a agentes tóxicos.

Nesse contexto, avaliar a qualidade de sistemas de teletoxicologia é de extrema importância. Contudo, uma avaliação usando modelos tradicionais que consideram qualidade como um todo (como a norma ISO/IEC 25010) é complexa; a mesma é genérica e pode não considerar necessidades específicas de sistemas de toxicologia. Ademais, métricas de avaliação de qualidade como as da norma supracitada demandam muito tempo e intervenção direta no sistema. Por exemplo, a coleta via *logs* para contabilizar quantos erros por tarefa um usuário comete geraria um *overhead* para a equipe de desenvolvimento de um *software* que pode já estar em uso.

De outro modo, o modelo AdEQUATE (*questionnaire for Evaluation of Quality in Telemedicine systems*) é uma abordagem customizada da norma ISO/IEC 25010 [Alves et al. 2015] que foi criado considerando especificidades de sistemas de telemedicina, decompondo o objetivo de avaliação em métricas e sendo operacionalizado por um questionário para permitir a análise da qualidade segundo a percepção dos usuários. Por usar um questionário, esse modelo de avaliação de qualidade de *software* é mais rápido e facilmente aplicável que modelos tradicionais, e como customização de uma norma

reconhecida mantém uma abrangente gama de características de qualidade.

Este trabalho tem como objetivo avaliar a qualidade de *software* do sistema DATATOX – um sistema de teletoxicologia desenvolvido sob a observação da Associação Brasileira de Centros de Informação e Atenção Toxicológica (ABRACIT), utilizado por diversos CITs em nove estados do Brasil – usando o modelo AdEQUATE em 11 dimensões. Além disso, pretende-se mostrar a viabilidade de se adotar modelos de avaliação de qualidade sistematicamente criados a partir de normas internacionais de qualidade aplicados a sistemas de teletoxicologia, focando em suas especificidades enquanto mantendo a generalidade dos modelos internacionais usados como base.

A seção seguinte descreve sistemas de informação toxicológica, o processo de customização do modelo AdEQUATE e suas aplicações via questionário. Os dados coletados são analisados e discutidos nas sessões 3 e 4, respectivamente. A última seção apresenta a conclusão e a validade do estudo.

2. Materiais e Métodos

As subseções seguintes descrevem o objeto de estudo – o sistema DATATOX – e o modelo de avaliação proposto – o AdEQUATE.

2.1. O sistema DATATOX

O sistema DATATOX é um exemplo de sistema de informações toxicológicas implantado em 2013, baseado nas necessidades da ABRACIT de um sistema interativo, *on-line* e robusto. Usado por múltiplos CITs em nove estados do Brasil em um regime 24/7, o sistema já gerenciou mais de 81.000 ocorrências de intoxicações por drogas de abuso, pesticidas (tanto domésticos quanto industriais), venenos de rato, produtos de limpeza, produtos cosméticos e de higiene, produtos industriais, metais, plantas e fungos e animais peçonhentos. Médicos são os requisitantes de informação mais comuns, seguidos pelos parentes dos intoxicados e enfermeiras; residentes de medicina e toxicologistas de plantão coordenados por um médico são os atendentes das ocorrências.

A Figura 1 ilustra uma interação usando o sistema DATATOX. Um requisitante em uma localidade A qualquer inicia o contato com uma ligação telefônica, recebida pelo atendente (1) em um CIT de uma localidade B. Durante essa etapa o atendente colhe informações sobre o paciente e o agente de intoxicação, e paralelamente as registra no sistema (2). Dependendo do caso o atendente pode recorrer a outras fontes de conhecimento para aconselhar o requisitante, seja utilizando bibliografia local ou bases de dados externas (3.a); outra opção é consultar o médico supervisor (3.b), que também pode consultar material externo (4), dando *feedback* ao atendente (5). O final da interação se dá com a instrução do atendente ao requisitante (6), podendo incluir requisições de exames ou mesmo o encaminhamento para tratamentos específicos.

2.2. O modelo AdEQUATE

AdEQUATE é um modelo de qualidade desenvolvido com base na norma ISO/IEC 25010, cujo objetivo é avaliar a qualidade de *software* de sistemas de telemedicina e telessaúde sob a perspectiva dos seus usuários finais [Alves et al. 2015]. Como [Behkamal et al. 2009], o modelo sistematicamente decompõe as características de qualidade de uma norma internacional e a estende criando medidas, e pré-define um questionário para operacionalizar a coleta de dados. O questionário¹ é composto de 70

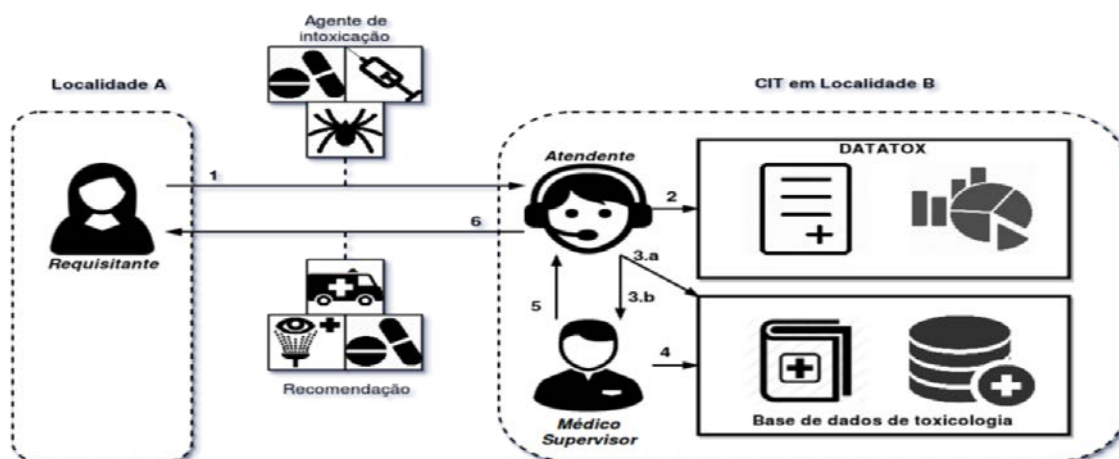


Figura 1. *Workflow* de um atendimento no DATATOX.

itens: dois itens demográficos sobre o papel dos usuários e a experiência dos mesmos, e 68 itens sistematicamente derivados das características e subcaracterísticas da referida norma. Todos esses 68 itens são estruturados (não permitindo respostas em texto livre), e são formulados de maneira positiva para evitar erros de entendimento e/ou interpretação [Sauro and Lewis 2011]. As opções de resposta seguem uma escala Likert com quatro pontos, mais três opções adicionais: "Não se aplica", "Não sei" e "Não entendi o item". Todos os itens são acompanhados por contra-exemplos.

Contudo, embora não seja convencional avaliar qualidade de software por meio de questionários – dados possíveis vieses estatísticos – o anexo C da norma ISO/IEC 25010 define que todas as subcaracterísticas e características de qualidade podem ser avaliadas através do uso de medidas externas. Ademais, a norma ISO/IEC 25022 propõe o uso de questionários como método de verificar qualidade [International Standard Organization 2003].

Para avaliar a qualidade de *software* do sistema DATATOX no contexto dos CITs e da ABRACIT, este trabalho apresenta um estudo de caso; o mesmo é caracterizado na seção 2.3, com os resultados obtidos sendo analisados na seção 3.

2.3. Definição do Estudo de Caso

Baseado no estado da arte referente a sistemas de teletoxicologia e na experiência de mais de 10 anos do Laboratório de Telemedicina no desenvolvimento de sistemas de telemedicina, foram consideradas relevantes todas as características e subcaracterísticas de qualidade propostas pelo modelo AdeQUATE¹. Para coletar os dados com maior precisão, todos os contra-exemplos foram revisados e adaptados para o contexto da rotina de sistemas de toxicologia.

2.4. Execução do Estudo de Caso

De novembro a dezembro de 2015 foram coletados dados usando o questionário AdeQUATE disponível *on-line* pelo sistema *web* LimeSurvey². Durante esse período

¹O conjunto completo das características, subcaracterísticas e itens do questionário pode ser acessado em <https://arquivos.ufsc.br/f/61d4095f6a/>

usuários do sistema DATATOX de diversos turnos de trabalho de várias localidades responderam o questionário; no total, foram obtidas 49 respostas, sendo 65% de atendentes e 35% de médicos. Desse número, 55% atestam terem usado o sistema por mais de um ano, 31% por mais de seis meses e apenas 14% por menos de 6 meses. A seção a seguir resume os dados coletados e os apresenta utilizando-se de estatística descritiva.

3. Resultados e Análise de Dados

Os dados coletados durante a execução do estudo de caso foram analisados usando duas métricas para cada característica de qualidade do modelo AdeQUATE, e uma terceira métrica para cada resposta dos participantes. As três métricas são definidas e analisadas na sequência.

Métrica #1 Distribuição da soma das opções de resposta de cada item do questionário relacionado à qualidade de software.

O gráfico da Figura 2 mostra a distribuição de cada opção de resposta agrupada pelas características de qualidade do modelo AdeQUATE. Pôde-se observar que na maioria das características, as opções de resposta “Concordo totalmente” ou “Concordo” (barras verde e verde claro) correspondem a mais de 50% das respostas, com algumas exceções: “Compatibilidade”, “Confiança” e “Cobertura de contexto”. Como a opção “Não entendi o item” representa no máximo 2% das respostas, considera-se que todas as características foram bem compreendidas pelos participantes do estudo, atestando a validade do mesmo.

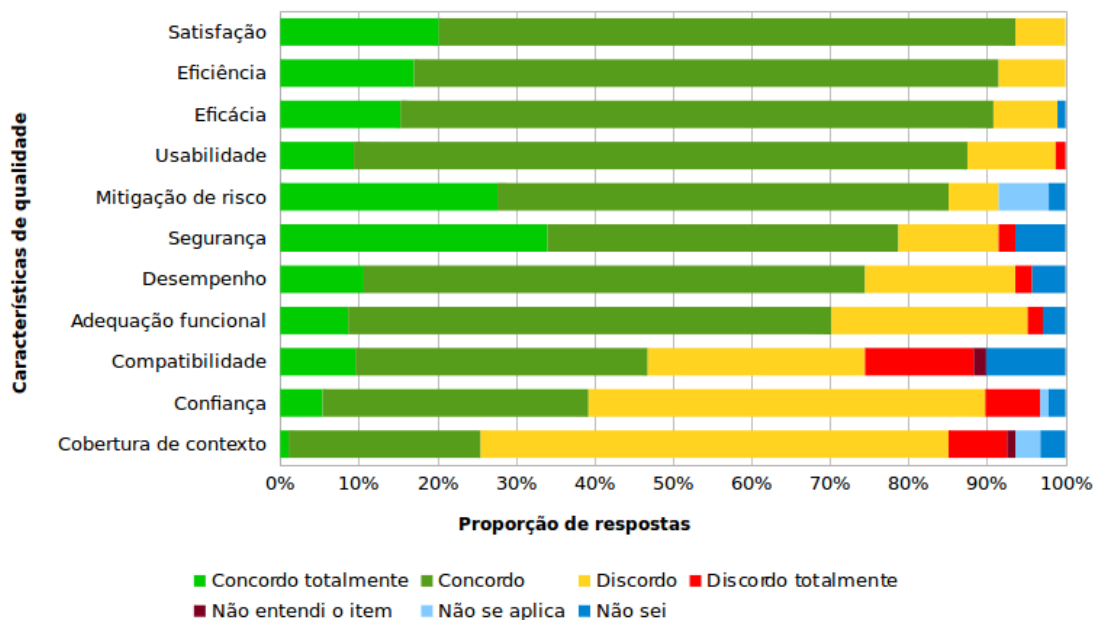


Figura 2. Distribuição das opções de resposta no questionário entre as características.

Para analisar mais profundamente o grau de qualidade de cada característica percebido pelos usuários participantes, foi calculada uma pontuação para cada característica

²<https://www.limesurvey.org/>

usando (1). Na equação, \tilde{R}_i corresponde à mediana das respostas do questionário para cada característica n , por participante i , enquanto $W(\tilde{R}_i)$ corresponde a uma função de atribuição de peso cujo domínio é o conjunto de opções de resposta e o contradomínio um peso associado: 3 para “Concordo totalmente”, 2 para “Concordo”, 1 para “Discordo” e 0 para “Discordo totalmente” (as demais opções não são avaliadas). O último componente da equação garante que cada S_n varie de 0 (grau mais pobre em qualidade – isto é, se todos os participantes respondessem “Discordo totalmente”) a 100 (grau de qualidade excepcional – isto é, se todos os participantes respondessem “Concordo totalmente”).

Esse sistema de pontuação é adaptado da pontuação definida pelo *System Usability Scale* (SUS) [Brooke 1996], por estar relacionado à qualidade de software e por ser uma ferramenta muito bem conceituada na indústria [Brooke 2013].

$$S_n = \left(\sum_{i=1}^{49} W(\tilde{R}_i) \right) * \frac{100}{147} \quad (1)$$

Usando (1), a segunda métrica é definida como segue:

Métrica #2 A mediana das pontuações para cada item do questionário em sua característica.

Os dados obtidos que ilustram a análise da métrica #2 são exibidos no gráfico da Figura 3.

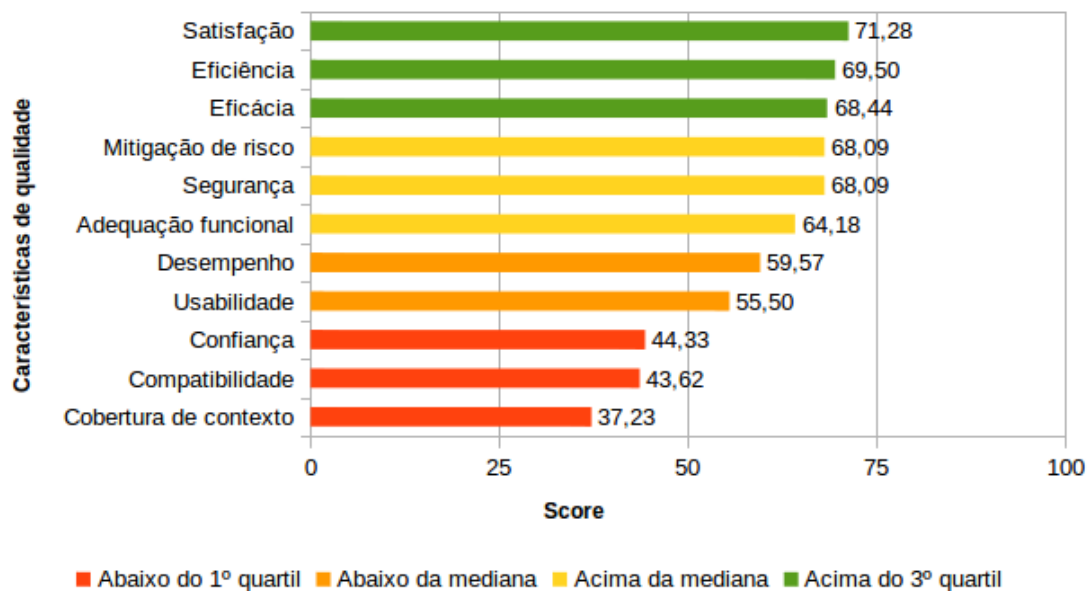


Figura 3. Pontuação calculada para cada característica.

Diferenças entre o gráfico da Figura 2 e o gráfico da Figura 3 mostram a influência do peso aplicado às opções de resposta no resultado final. Enquanto a Figura 2 mostra “Eficiência” e “Eficácia” como tendo aproximadamente o mesmo grau de qualidade, a Figura 3 mostra que a percepção dos usuários finais do grau de qualidade de “Eficiência” é maior, pois obteve mais respostas com “Concordo totalmente”. Essa diferença é ainda

mais evidente no caso das características “Confiança” e “Compatibilidade”. “Compatibilidade” obteve mais respostas com a opção “Concordo totalmente”, mas também obteve mais “Discordo totalmente” – mostrando uma certa divergência de opiniões entre os usuários; como resultado, “Confiança” pôde ser percebida como tendo um grau de qualidade maior que “Compatibilidade”. Por outro lado, outras características não foram afetadas pela mudança de métrica: “Satisfação” e “Cobertura de contexto” ainda são percebidas como tendo o grau de qualidade mais alto e mais baixo, respectivamente.

Todas as características obtiveram pontuações dentro de duas distâncias interquartis (IQR), ou seja, os resultados não se afastam muito da mediana. De forma complementar, não foram identificados *outliers* para a métrica #2.

Métrica #3 Pontuação de qualidade por participante do questionário.

Enquanto as métricas #1 e #2 mostram no geral a percepção dos usuários com relação a cada característica de qualidade, a métrica #3 mostra a percepção de qualidade como um todo por participante, ilustrada pela Figura 4. Utilizando-se das referências usadas na escala SUS [Bangor et al. 2009], podemos ver que nenhum participante tem a percepção de que o sistema DATATOX possui um grau pobre de qualidade, e que a maioria das pontuações mostra uma percepção voltada para um grau de qualidade aceitável; contudo, apenas 16% dos participantes consideram o sistema como tendo um bom grau de qualidade.

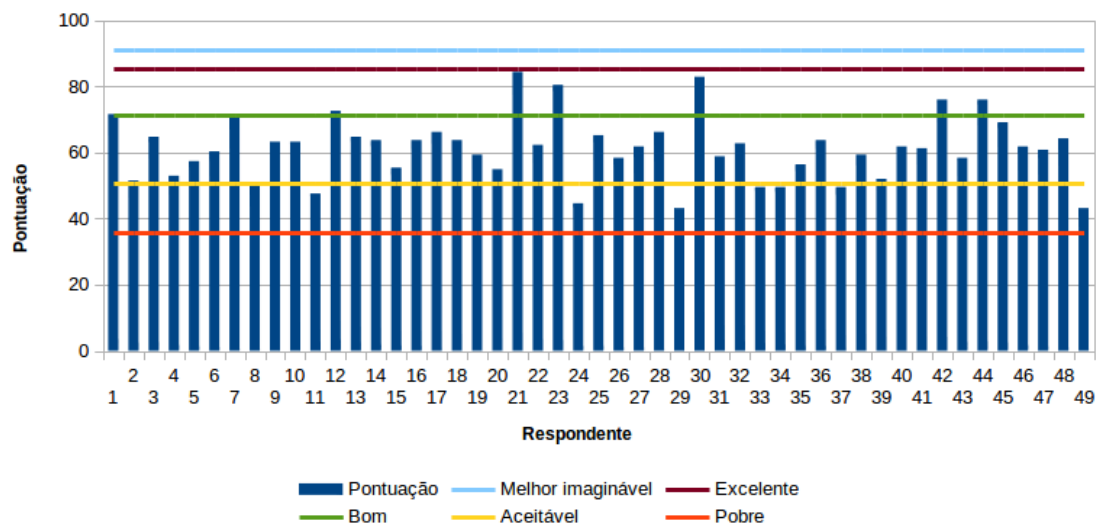


Figura 4. Pontuação calculada para cada característica.

A seção seguinte discute e contextualiza os dados analisados, ilustrando as possíveis razões para as percepções de baixo grau de qualidade em algumas características.

4. Discussão

Os dados obtidos na execução do estudo de caso mostram que, no geral, o sistema DATATOX foi percebido como tendo um grau aceitável de qualidade, como mostra a

Figura 4. As divisões em quartis exibidas na Figura 3 mostram como os usuários percebem o sistema relacionado a suas características de qualidade: “Satisfação”, “Eficiência” e “Eficácia” têm os graus de qualidade mais altos, seguidos por “Mitigação de risco”, “Segurança” e “Adequação funcional”. “Desempenho” e “Usabilidade” foram percebidos com um grau de qualidade moderadamente menor.

O baixo grau de qualidade percebido para “Cobertura de contexto” é dado pelo baixo grau de flexibilidade do sistema: o mesmo só pode ser acessado usando-se o navegador Mozilla Firefox, apenas em computadores *desktop* ou *laptop* e não pode ser acessado pelos requisitantes. Além disso, os usuários finais consideraram que o conjunto de funcionalidades não cobre todo o contexto de seus trabalhos, isto é, há funcionalidades desejadas pelos usuários que o sistema ainda não disponibiliza.

Como o DATATOX é um sistema *web* de teletoxicologia, ele está suscetível a flutuações na qualidade da conexão com a Internet. Tais instabilidades contribuem para um baixo grau de tolerância a falhas e disponibilidade, o que é percebido nos gráficos pela característica “Confiança”. O sistema também não é integrado com nenhuma base de dados toxicológicos, forçando seus usuários a depender de bases de dados externas, contribuindo para um baixo grau de “Interoperabilidade”.

4.1. Ameaças à validade

Para aumentar a validade interna e evitar um viés de seleção, durante o planejamento do estudo de caso foi definido que seriam convidados usuários do sistema DATATOX de todas as unidades disponíveis, de todos os turnos de trabalho, pessoalmente ou via telefone. Pode-se assim garantir uma maior uniformidade dos dados, uma vez que cada CIT tem suas particularidades e flutuações na qualidade da conexão de Internet (o que pode gerar visões de qualidade diferentes). Para minimizar o viés de projeto foi adotado o modelo AdeQUATE no lugar de um questionário construído *ad-hoc*. Embora esse modelo e suas escalas de avaliação ainda não tenham sido validados por um estudo maior, assumiu-se que a derivação sistemática dos itens baseados em normas de qualidade internacionais como a ISO/IEC 25010 reduz ameaças à validade da construção do estudo de caso.

Durante este trabalho, o foco se manteve em um sistema específico de teletoxicologia – o sistema DATATOX. Logo, os resultados deste estudo de caso têm uma validade externa limitada. Contudo oferece valiosas informações para trabalhos futuros ligados à validação do modelo AdeQUATE.

5. Conclusão

Sistemas de controle toxicológico auxiliam toxicologistas a gerenciar casos de intoxicações por vários agentes, manipulando informações sobre o caso, informações sobre o paciente e acompanhando os casos quando necessário. Alguns sistemas auxiliam ainda mais os toxicologistas, oferecendo bases de dados toxicológicos especializadas. Nesse contexto, qualidade de software é um tópico delicado porque não há um padrão específico para guiar o desenvolvimento de tais sistemas, tampouco um modelo de avaliação de qualidade. Enquanto a norma ISO/IEC 25010 apresenta características e subcaracterísticas que permitem a engenheiros de *software* avaliar qualidade geral, as especificações do domínio de atenção à saúde são ignoradas.

Neste trabalho foi analisada a qualidade de *software* do sistema DATATOX – um sistema de teletoxicologia usado por vários CITs no Brasil – através de um estudo de caso que adota o modelo AdEQUATE. A análise dos dados mostra que o sistema foi percebido como tendo um bom grau de qualidade. Seus pontos fortes estão relacionados à satisfação, eficiência e eficácia, mostrando um sistema que satisfaz seus usuários e permite-os executar tarefas com acuracidade e completude, gastando poucos recursos. Por outro lado, o sistema DATATOX foi considerado como tendo um baixo grau de confiança, compatibilidade e cobertura de contexto, significando que o sistema algumas vezes não realiza algumas funções, não troca informações com outros sistemas e não é flexível o suficiente para cobrir todos os contextos de uso.

Adicionalmente, este estudo mostra a viabilidade de se avaliar qualidade de *software* usando a perspectiva dos usuários finais de sistemas de informação toxicológica adotando o questionário sistematicamente derivado do modelo AdEQUATE. Através dessa avaliação a ABRACIT tem a compreensão da qualidade atual do sistema, podendo investir em otimizações específicas para garantir um serviço de qualidade ainda melhor.

Após uma extensa e profunda validação do modelo AdEQUATE, um trabalho futuro pretende rever a percepção dos usuários da qualidade do sistema usando o modelo validado, e possivelmente com mais participantes, compará-lo com outros sistemas de teletoxicologia. Nesse ínterim, futuros trabalhos na área de banco de dados e inteligência artificial podem aumentar a qualidade do sistema, integrando o mesmo a uma base de dados toxicológicos (aumentando assim sua compatibilidade).

Agradecimentos

Este trabalho teve o apoio da Secretaria de Estado da Saúde de Santa Catarina - SES/SC, do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e da Fundação de Amparo à Pesquisa de Santa Catarina - FAPESC.

Os autores também gostariam de agradecer a todos os usuários do sistema DATATOX que dedicaram seu tempo e esforço para responder o questionário.

Referências

- Alves, J. M., Wangenheim, C., Lacerda, T., Savaris, A., and Wangenheim, A. (2015). Adequate software quality evaluation model v1.0. Technical report, Instituto Nacional para Convergência Digital – INCoD.
- Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.
- Bateman, D. N., Good, A. M., Laing, W., and Kelly, C. (2002). Toxbase: poisons information on the internet. *Emergency medicine journal*, 19(1):31–34.
- Behkamal, B., Kahani, M., and Akbari, M. K. (2009). Customizing iso 9126 quality model for evaluation of b2b applications. *Information and software technology*, 51(3):599–609.
- Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.
- Brooke, J. (2013). Sus a retrospective. *Journal of Usability Studies*, 8(2):29–40.

- CIT/RS, C. (2015). Estatísticas dinâmicas. Date last accessed 2015-12-11.
- CIT/SC, C. (2015). Estatísticas anuais. Date last accessed 2015-12-11.
- International Standard Organization (2003). ISO/IEC 9126-2:2003 Software engineering – Product quality – Part 2: External metrics. Technical report.
- Mowry, J. et al. (2015). 2014 annual report of the american association of poison control centers' national poison data system (npds): 32nd annual report. *Clinical toxicology*, 53(10).
- Perry, L. et al. (2014). National toxicovigilance for pesticide exposures resulting in health care contact—an example from the uk's national poisons information service. *Clinical toxicology*, 52(5):549–555.
- Poison Control Centers (2015). Date last accessed 2015-12-14.
- Sauro, J. and Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2215–2224. ACM.
- Seifert, S. M. et al. (2013). An analysis of energy-drink toxicity in the national poison data system. *Clinical toxicology*, 51(7):566–574.
- Skolnik, A. (2013). Telemedicine and toxicology: back to the future? *Journal of Medical Toxicology*, 9(3):217–219.
- Tormey, W. and Moore, T. (2013). Poisonings and clinical toxicology: a template for ireland. *Irish journal of medical science*, 182(1):17–23.
- World Health Organization et al. (1997). Guidelines for poison control.

SMAI - Sistema Móvel de Assistência ao Idoso

Matheus Costa Stutzel¹, Michel Filippo¹, Luciana Branco da Motta³, Alexandre Sztajnberg^{1,2,4}, André Brittes³

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ

¹Bacharelado em Ciência da Computação - Instituto de Matemática e Estatística

²Pós-Graduação em Ciências Computacionais - Instituto de Matemática e Estatística

³Núcleo de Apoio ao Idoso – Universidade Aberta da Terceira Idade

⁴Pós-Graduação em Eletrônica - Faculdade de Engenharia

matheusstutzel@yahoo.com.br, michelpf13@gmail.com, lubmotta@gmail.com,
alexsz@ime.uerj.br, brites84@gmail.com

Abstract. *SMAI uses Android applications as the infrastructure to monitor elderly patients with chronic degenerative disease that present functional loss. The objectives are: (i) make the caregiver communication with the health team more agile; (ii) reduce caregiver stress; (iii) provide the health team information about the patient's condition more often in an organized manner, facilitating decision-making. Applications are developed under the guidance of the health team and are being assessed in a clinical trial with 30 +30 patients accompanied by NAI/UERJ.*

Resumo. *O SMAI utiliza aplicações Android como infraestrutura de monitoramento de pacientes idosos com doença crônico-degenerativa que apresentem perda funcional. Os objetivos são: (i) tornar a comunicação do cuidador com a equipe de saúde mais ágil; (ii) reduzir o estresse do cuidador; (iii) prover à equipe de saúde informações sobre o estado do paciente com maior frequência, de forma organizada, facilitando a tomada de decisões. As aplicações são desenvolvidas segundo a orientação da equipe de saúde e estão sendo avaliadas em um ensaio clínico com 30 + 30 pacientes acompanhados pelo NAI/UERJ.*

1. Introdução

As patologias crônicas têm potencial para causar incapacidades, somando prejuízos na autonomia e independência para a vida diária e aumentando a necessidade de cuidados prolongados prestados por familiares, amigos e comunidade [MinS, 2006]. Estima-se que entre 2010 e 2050 o número de idosos que necessitam de cuidados, em todo o mundo, quase triplicará de 101 para 277 milhões [Prince, 2013].

A família é a principal forma de suporte social disponível para o cuidado. A legislação e as políticas públicas afirmam e a própria sociedade acredita que os idosos devem ser cuidados por sua família [Caldas, 2003]. No seio familiar, a presença do idoso dependente implica em sobrecarga emocional, física e financeira [UNFPA, 2012]. Desta forma, é importante desenvolver ações que permitam que a equipe de saúde apoie os responsáveis pelo cuidado, aliviando o estresse de cuidar.

Na Universidade Aberta da Terceira Idade da UERJ (UnATI/UERJ), o Núcleo de Atenção ao Idoso (NAI) oferece um serviço de monitoramento contínuo para pacientes idosos com vários estágios de desordem neurocognitiva (demência). Uma

equipe de saúde multidisciplinar acompanha 250 pacientes e seus cuidadores. O serviço é gratuito. Consultas médicas ambulatoriais são agendadas periodicamente para cada paciente, acompanhados por seus cuidadores. São seguidos protocolos e procedimentos estabelecidos, e as recomendações são enfatizadas com os cuidadores em situações de emergência. O serviço é oferecido tendo em conta as características do grupo de pacientes monitorados e seus cuidadores: (i) o cuidador é geralmente um membro da família e constantemente está estressado, (ii) a família tem, em média, baixo rendimento e a maioria reside em bairros periféricos e (iii) necessitam de transporte público para chegar ao NAI. Assim, cada encontro impõe sacrifício para o cuidador e paciente.

A tecnologia da informação tem despontado como recurso de apoio, atuando como facilitador da comunicação e monitoramento. O grande avanço nas áreas de dispositivos de hardware e telecomunicações pavimentou o caminho para aplicações de telemedicina [Wootton, 2006]. Entre estas, as aplicações de telessaúde, onde dispositivos de monitoramento permitem o telemonitoramento e telediagnóstico vem sendo pesquisadas. Assim, a telemedicina pode integrar a rede de cuidado do idoso dependente, para complementar o atendimento presencial pela equipe de saúde.

Este artigo apresenta o SMAI - Sistema Móvel de Assistência ao Idoso, cujos objetivos são (i) tornar a comunicação do cuidador com a equipe de saúde mais ágil; (ii) reduzir o estresse do cuidador; (iii) facilitar o cuidado e suporte do idoso nas suas atividades de vida diária; (iv) prover à equipe de saúde informações sobre o estado do paciente com maior frequência, de forma organizada, facilitando a tomada de decisões. Além disso, são avaliados os impactos do uso do sistema na melhoria da qualidade de vida dos pacientes e cuidadores, nas intercorrências e na aderência aos tratamentos.

Para alcançar tais objetivos foi desenvolvido um sistema composto de 2 aplicações: uma dedicada ao profissional de saúde e otimizada para a utilização em um *tablet*, e a outra desenvolvida para *smartphone*, dedicada ao cuidador ou paciente. Elas foram projetadas para o contexto do grupo de pacientes e cuidadores monitorados pelo NAI, considerando as características e os condicionalismos deste grupo, as atuais práticas de saúde adotadas e a experiência da equipe no cuidado deste grupo, portanto as aplicações possuem interfaces desenvolvidas de modo a facilitar a visualização dos dados gerados pelos pacientes, bem como simplificar a entrada de novas informações no sistema.

O restante do texto está estruturado da seguinte forma. A Seção 2 apresenta trabalhos relacionados que consideram tecnologia da informação aplicada à telessaúde. A Seção 3 apresenta o contexto do projeto e o estudo preliminar realizado para nortear a definição do sistema. A Seção 4 apresenta uma visão geral do sistema e a Seção 5 aborda o protótipo desenvolvido e alguns detalhes da implementação. A Seção 6 discute uma primeira avaliação realizada e a Seção 7 traz as últimas considerações.

2. Trabalhos relacionados e experiências anteriores

O uso da telemedicina no controle de doenças crônicas tem sido discutido nos últimos anos, demonstrando benefícios para o paciente e para os sistemas de saúde [Dorsey, 2013], [Celler, 2006], [Jaglal, 2013], [Finkelstein, 2006]. Alguns deles se concentram em sistemas de notificação e lembretes aos pacientes, com o objetivo de melhorar a aderência ao tratamento [Fisher, 2006], [Keranen, 2013].

O uso de soluções móveis e tecnologia da informação na área da telemedicina no Brasil está se consolidando. Considerando aplicações de telessaúde, algumas propostas já podem ser elencadas, como [Sanches, 2012].

Em [Nicholas, 2007] são avaliados vários aspectos da tecnologia para e-Health, incluindo o impacto para o paciente, para o médico e para os serviços de saúde já estabelecidos. Um dos problemas recorrentes é o abandono do uso do sistema, por várias razões: falta de incentivo, falta de interesse e a não percepção de benefícios.

A revisão apresentada em [Mosa, 2012] classifica aplicações para smartphones apresentadas em 2894 artigos. A maior parte das aplicações está relacionada à diagnóstico de doenças ou calculadoras médicas. Apenas 15 das aplicações pesquisadas estão relacionadas ao monitoramento de pacientes com doenças crônicas. Destas 15 aplicações poucas se preocupam em tornar lúdico o uso da aplicação. No SMAI procuramos adicionar aspectos lúdicos e características de *gamificação* ao aplicativo destinado aos pacientes/cuidadores como forma de atrair o usuário ao uso contínuo do sistema (mais detalhes são discutidos em [Stutzel, 2016]).

3. Preparação

O projeto do SMAI envolve duas equipes: o LCC - Laboratório de Ciência da Computação e pelo NAI – Núcleo de Atenção ao Idoso/ UnATI, ambos da UERJ. O público alvo do sistema é composto de idosos, com algum grau de dependência, e seus cuidadores. A equipe do NAI tem experiência no monitoramento e acompanhamento deste grupo empregando os procedimentos e as práticas disponíveis.

O SMAI é proposto com foco em aplicações simples, para serem utilizadas em um equipamento do tipo *smartphone* que já faz parte do dia-a-dia do paciente ou cuidador. O entendimento é o de que o uso de um aparelho de telefonia celular não introduza mudanças significativas na rotina das pessoas.

O acesso à Internet tornou-se ubíquo e economicamente mais acessível. Equipamentos que permitem acesso à Internet também se tornaram mais baratos. Ainda assim, devido ao baixo poder aquisitivo da maioria dos pacientes e cuidadores monitorados, para a avaliação do sistema é fornecido um *smartphone* com acesso à Internet para transmissão de dados e recebimento de alertas para cada paciente. Como preparação para o desenvolvimento do SMAI, foi realizado um levantamento preliminar sobre o uso de recursos de computação e telefonia móvel, aplicando-se um questionário ao grupo acompanhado pelo NAI. Ao todo, 42 (quarenta e dois) pacientes ou seus acompanhantes, responderam às questões do formulário. Do total dos entrevistados:

- 76.60% têm computador e 93.62% celular. 72.34% sabem enviar SMS (torpedo);
- 80.85% fazem uso de correio eletrônico e/ou redes sociais como o Facebook;
- 72.34% tem acesso à Internet em casa, sendo que destes 73% acha o acesso bom;
- 86% acessa a Internet algumas vezes por semana, sendo que 40% todos os dias.

Os resultados obtidos constituiriam importante informação para direcionar a abordagem proposta para o SMAI. O uso de um aparelho celular é comum para a maioria dos pacientes, familiares ou cuidadores. Destes, 25.53% já utilizam o sistema Android [Google, 2003]. Com base nestas informações o sistema Android foi adotado.

4. SMAI - Sistema Móvel de Assistência ao Idoso

O SMAI é composto de duas aplicações: paciente/cuidador e médico, e um serviço na nuvem. Os usuários têm acesso ao sistema através de um aparelho do tipo *smartphone* ou *tablet*, tanto em suas residências, como quando em mobilidade, para realizar seus afazeres fora de casa. A Figura 1 apresenta uma visão geral do SMAI.

Nos aparelhos *smartphone* uma aplicação (*app*) permite ao paciente, ou ao cuidador/familiar enviar e receber informações e notificações da equipe de saúde:

- lembretes sobre uso de medicação e consultas agendadas;
- informações contendo medidas fisiológicas como a pressão arterial;
- relatório diário com informações sobre o comportamento, rotina alimentar, evacuação e micção do paciente; e relatório semanal sobre o próprio cuidador;
- recebimento de mensagens de suporte ao cuidador nas orientações de cuidado;
- envio periódico da localização do paciente, usando uma combinação de possibilidades como o GPS e sistemas de localização baseado na rede.

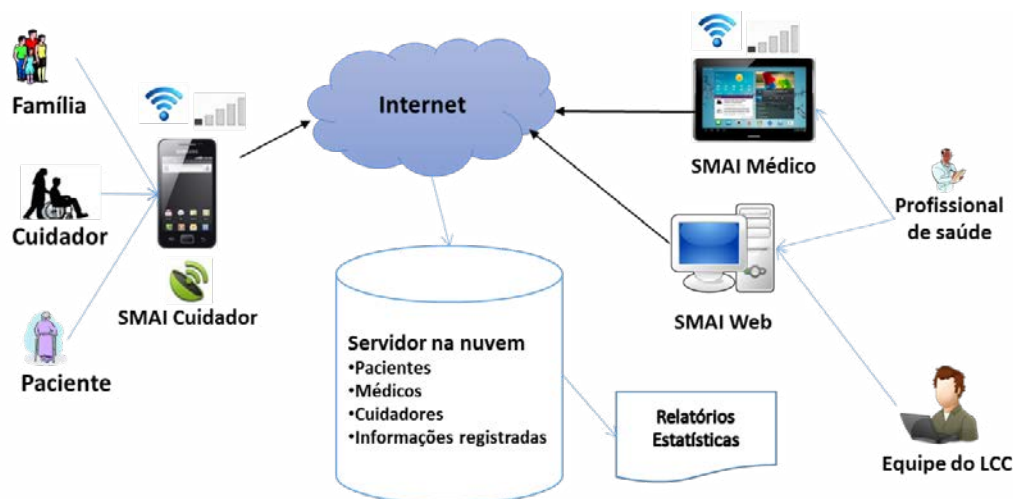


Figura 1. Esquema geral do SMAI

A aplicação utilizada pelo cuidador/familiar tem o objetivo de auxiliar seu dia a dia e facilitar o envio de informações para o NAI. A aplicação quando utilizada pelo paciente tem um número reduzido de opções, cujo objetivo é receber deste paciente informações sobre seu estado geral, mantendo uma conexão “virtual” com o NAI.

Os profissionais da saúde têm acesso ao sistema através de um *tablet* onde uma aplicação permite o monitoramento e a comunicação com o grupo de pacientes, com uma visualização estruturada das informações. Entre as funções estão:

- visualizar alarmes e informações atualizadas dos pacientes, em grupo;
- visualizar detalhes de cada paciente, incluindo a sua localização atual e dos últimos dias, permitindo verificar sua mobilidade;
- envio de agendamento de consulta e lembretes de medicamentos, além de mensagens para usuários individuais ou grupos de pacientes;
- visualizar os relatórios relacionados aos pacientes e o relatório do cuidador.

Todas as informações são enviadas para um servidor de aplicação na nuvem que oferece um conjunto de serviços específico de uma API. Da mesma forma, deste serviço

são recebidas as mensagens, notificações e alarmes. Este serviço possui os requisitos de disponibilidade e segurança necessários para aplicações de telessaúde.

Privacidade e Segurança. A comunicação de cada ator do sistema, equipe de saúde e cuidadores, é feita a partir de uma identificação única no serviço na nuvem, permitindo a autenticação, o controle de acesso e privacidade. Antes de serem transmitidos, os dados são criptografados. Com isso é garantida a autenticidade e confidencialidade da informação sem exigir a identificação do usuário a cada acesso. O acesso aos dados dos pacientes é controlado por mecanismos de permissão e compartilhamento autorizado. Os agendamentos, lembretes de medicações e notificações para os pacientes tem associada a informação do profissional de saúde responsável.

Transmissão e Tolerância a Falhas. Os dados de todas as aplicações são transmitidos e recebidos pela Internet. Na ausência de conexão com a internet, o SMAI disponibiliza a visualização das informações já carregadas e os dados gerados localmente pelo usuário são persistidos localmente, até que uma conexão seja estabelecida.

5. Protótipo

As funcionalidades, painéis de informação e notificações foram implementadas de acordo com a orientação e requisitos da equipe de saúde do NAI, seguindo as práticas e padrões de projeto recomendados pela Google. A Figura 2 apresenta a estrutura geral adotada para as duas aplicações. Cada módulo herda um conjunto de classes **Activity*, que transmitem/recebem os dados, em XML, criptografados, para/do serviço da nuvem.

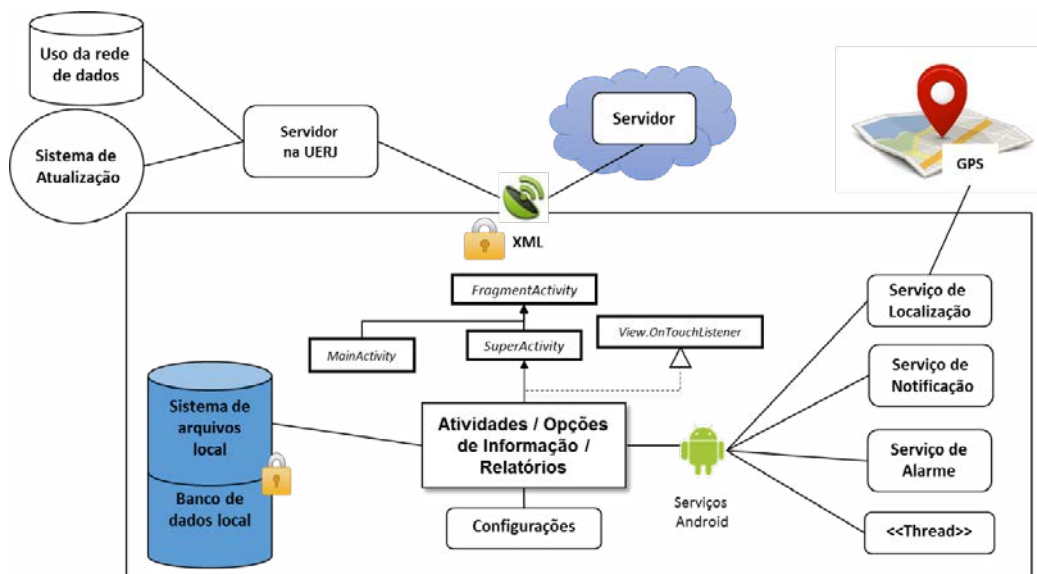


Figura 2. Estrutura do SMAI

As notificações no SMAI utilizam um mecanismo especialmente adaptado que permite, por exemplo, criar lembretes que demandam do cuidador uma confirmação e que não se acumulam evitando sobrecarga para o cuidador. Quando geradas, as notificações são acompanhadas de um alerta sonoro e visual no dispositivo.

Todos os dados gerados são salvos imediatamente na memória interna do equipamento para que não haja perda de dados. Após esse processo, é iniciado o envio para o servidor na nuvem. O procedimento de transmissão criptografa os dados a serem enviados e assim que o aplicativo recebe a confirmação a cópia local é apagada (no caso

do *smartphone*), ou persistida no banco de dados local (*tablet*). Em caso de erro na transmissão, o processo é reiniciado após um tempo, que varia de acordo com o erro.

SMAI Cuidador. A aplicação do cuidador possui um menu principal, com módulos de entrada de informações (Figura 3). Ao ser carregado, o programa inicia todas as atividades de plano de fundo, como por exemplo, o envio da localização do paciente e a verificação de atualizações. A Figura 3 apresenta exemplos de entrada de dados de *Dor*, *Pressão Arterial* e *Temperatura*. Alguns módulos são cientes de contexto, acionando ações específicas. Por exemplo, se a temperatura ultrapassar 37,8° é solicitada a informação sobre tosse.

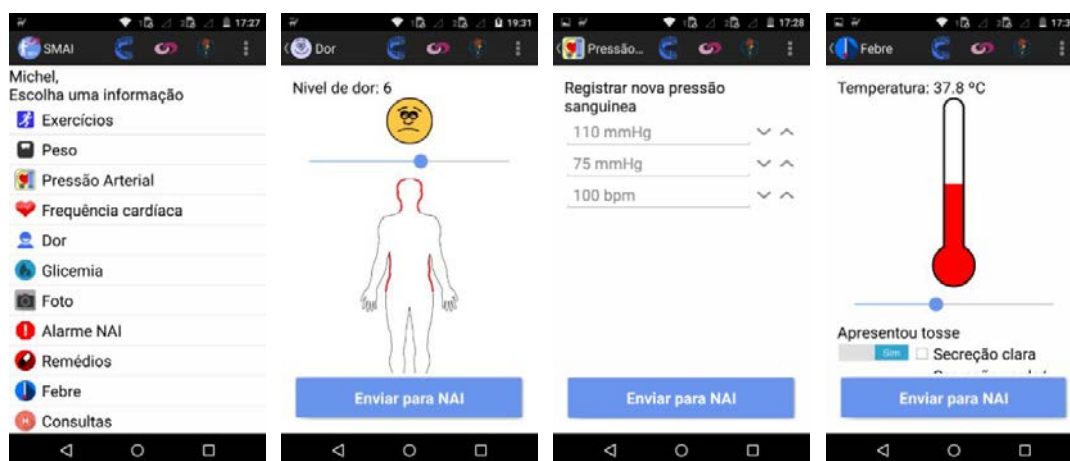


Figura 3. Tela principal e envio de informações

Diariamente, uma notificação é gerada como lembrete para o cuidador preencher um relatório (Figura 4). Ele pode preencher os itens de acordo com sua percepção de importância ou indicação do profissional de saúde. Por exemplo, na Figura 4 temos o painel de Informações Gerais, onde o cuidador anota a qualidade do sono, se houve - ou não distúrbios comportamentais e a hidratação do paciente.

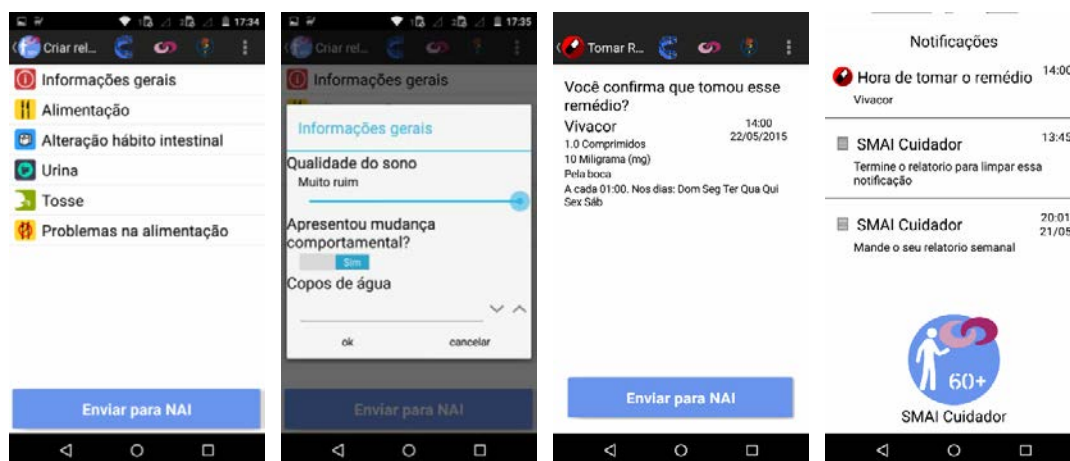


Figura 4. Relatórios, notificações e lembretes

Este é um dos módulos desenvolvidos especificamente para o conjunto de pacientes monitorados e aderentes à prática adotada pelo cuidador. Já existe uma rotina para que informações do dia a dia do paciente sejam anotadas em uma planilha em papel, de forma que a equipe de saúde possa acompanhar a evolução do paciente e intercorrências. O relatório através do SMAI objetiva tornar esta rotina mais ágil para o

cuidador e proporcionar um acompanhamento mais próximo da equipe de saúde, já que os relatórios ficam disponíveis imediatamente.

SMAI Médico. A estrutura da aplicação do profissional de saúde se assemelha ao SMAI Cuidador (Figura 2), tendo como diferença a utilização da API de mapas da Google, além do gerenciamento de vários pacientes ao mesmo tempo.

A aplicação tem como base uma lista de pacientes monitorados e um mecanismo automático de atualização de informações que traz do serviço na nuvem as últimas informações enviadas pelos pacientes. Esta lista é persistida em um banco de dados local, e exibida no Painel de Controle inicial. Desta forma, é minimizada a necessidade de se obter os dados do servidor na nuvem a cada operação do profissional de saúde.

Pacientes ou cuidadores que acionaram o Alarme são exibidos com a borda em vermelho no Painel de Controle (Figura 5a). Quando o profissional de saúde seleciona um dos pacientes, um menu similar ao do cuidador também é apresentado (Figura 5b), com opções adicionais, como o gráfico das medidas fisiológicas transmitidas (Figura 6).



Figura 5 (a) Painel de controle, (b) detalhe de um paciente



Figura 6 Detalhes do histórico de dor

Algumas opções permitem que se listem os itens enviados pelo paciente, na forma de relatórios, ou que se adicionem elementos para serem notificados para o paciente. Este é o caso dos Medicamentos, que podem ser adicionados/apagados para cada paciente. Além disso, é possível acompanhar o registro dos medicamentos que tiveram sua notificação confirmada por um cuidador. Por exemplo, o registro de localização do usuário e relatórios diários preenchidos (Figura 7) ficam à disposição da equipe de saúde para acompanhamento e tomada de decisões.

O Alarme NAI é atualizado a cada 2 minutos, com isso ele é recebido rapidamente pela equipe de saúde, tornando menor o tempo de resposta ao paciente. As demais informações também são atualizadas periodicamente de forma automática, porém em um intervalo de tempo maior. A qualquer momento podem ser solicitados os dados mais recentes, seja de um determinado paciente ou de todo o grupo.

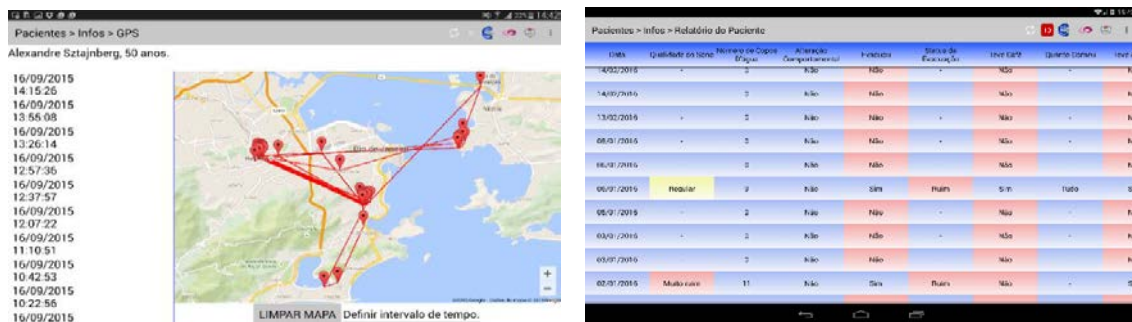


Figura 7. (a) Localização do paciente e (b) relatórios diários

6. Avaliação

Estudo piloto. No início de 2015 o protótipo do SMAI passou por uma primeira avaliação. Um grupo de 5 cuidadores foi selecionado. Cada um recebeu um aparelho Motorola Moto G, com o SMAI Cuidador pré-instalado e créditos suficientes. O objetivo era entender como o cuidador perceberia a nova ferramenta, facilidade de uso, intuitividade e a disposição em continuar utilizando o sistema. Para isso cada cuidador utilizaria o sistema por 15 dias, da melhor forma, sem o compromisso de fornecer dados reais. Uma reunião coletiva foi realizada para a entrega dos aparelhos e um treinamento de 40 minutos, *hands-on*, foi oferecido pela equipe do LCC para cada cuidador.

Durante os 15 dias da avaliação o uso do sistema foi monitorado através do SMAI Médico. Também foram monitorados os contadores de dados transmitidos e recebidos, pelo chip e pela rede WiFi (se habilitada) e qual o consumo específico da aplicação SMAI e outras aplicações. Deste monitoramento destacamos o seguinte:

- Todos os cuidadores transmitiram dados, com média variável por dia;
- O uso de dados do SMAI é baixo. Média de 87,93 kilobytes com download e de 614,04 kilobytes com upload por dia;
- As atualizações automáticas e outros mecanismos do Android representam 90% do consumo registrado nas duas semanas de avaliação.

Durante a reunião marcada para a devolução dos aparelhos, foi realizada uma pesquisa informal de opinião com os cuidadores e as respostas foram gravadas. Foram questionados sobre a sua experiência ao usar o sistema, com destaque para os seguintes pontos: (i) todos os cuidadores fariam uso contínuo do sistema; (ii) alguns cuidadores não entenderam completamente algumas opções; (iii) sugestões foram registradas no sentido de mudar algumas opções ou acrescentar a possibilidade de informar detalhes.

Teste clínico. No final de 2015 a equipe do NAI planejou um teste clínico com o SMAI. Para isso foram desenvolvidos protocolos complementares incorporando o uso do sistema aos procedimentos e rotinas já adotados. O teste é randomizado (método de minimização) acompanhando um grupo de 30 cuidadores sob intervenção do sistema e um grupo de controle de 30 cuidadores é acompanhado com questionários por telefone. A duração prevista é de nove meses. Cada cuidador recebe um Motorola Moto G e um plano de dados de 50 Mbytes. Os gráficos da Figura 8 apresentam informações sobre (a) a quantidade de dados enviados, permitindo verificar quais são os módulos mais usados e (b) o consumo médio do plano de dados.

Ainda é prematuro fazer correlações e observações sobre os resultados. O consumo de dados, por exemplo, pode aumentar significativamente ao se lançar uma nova versão do

sistema, ou quando a Google insiste em fazer atualização de alguma aplicação. Por outro lado, é possível identificar que a transmissão de dados pelos módulos iniciados a partir das notificações é maior (Relatórios, Remédios tomados).

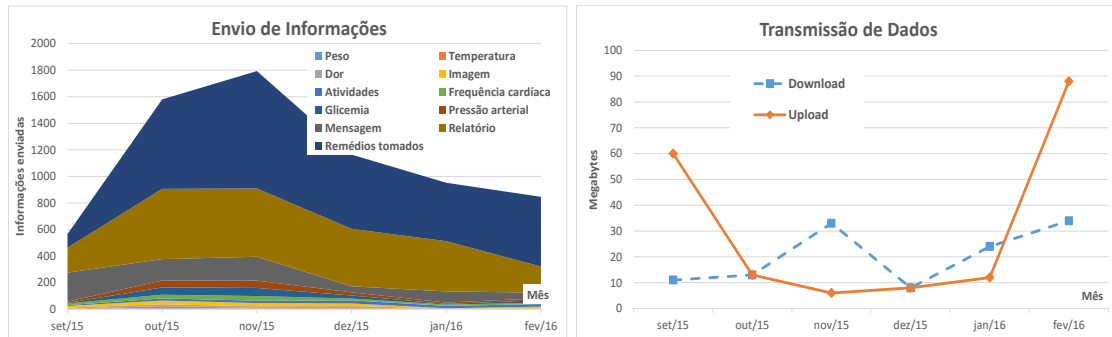


Figura 8. (a) Quantidade de informações enviadas (b) consumo de dados

O estudo tem o objetivo de avaliar a efetividade do aplicativo SMAI no monitoramento de idosos dependentes a partir da melhoria das alterações clínicas mais frequentes na população estudada. Também objetiva-se avaliar como a qualidade de vida dos cuidadores e familiares é afetada e qual a aderência dos cuidadores ao uso do sistema proposto, correlacionada aos outros aspectos avaliados.

7. Conclusão

A abordagem do SMAI é a simplicidade, ajudar o cuidador em sua atividade, sem ser mais uma tarefa ou preocupação, dentre tantas outras que já fazem parte do seu dia. O suporte oferecido pelo aplicativo móvel é complementar ao atendimento ambulatorial, não substituindo a avaliação clínica do paciente pela equipe de saúde.

As avaliações realizadas indicam que as opções de projeto adotadas são consistentes. O retorno dos usuários em consultas informais mostra que a usabilidade do SMAI é boa. A maioria das entradas não exige que o cuidador use o teclado virtual, mas isso não seria um problema, como verificamos em nosso levantamento preliminar. O desenvolvimento das aplicações, adaptados ao contexto e às possibilidades reais das pessoas envolvidas, e orientadas pelo conhecimento da equipe de saúde tem mostrado bons resultados. Além disso, a partir dos relatos de sucesso e insucesso de experimentos semelhantes na literatura [Celler, 2006], foi adotada a estratégia de tornar as aplicações atraentes ao uso, evitando o desinteresse do cuidador ou familiar.

Uma das lições aprendidas até este ponto, observado informalmente durante as reuniões presenciais com os usuários do SMAI, é o efeito de fortalecer a relação cuidador-profissional de saúde. Por exemplo, o sistema de mensagens simples do SMAI permite que a equipe NAI envie pequenos textos de orientação. Assim, além das notificações e lembretes, o cuidador pode receber da equipe NAI, mensagens personalizadas. O acompanhamento diário das informações enviadas pelos cuidadores permite que as pequenas intervenções por mensagens de texto, ou até mesmo por telefone, sejam feitas em um tempo mais curto quando comparado com procedimentos padrão. A percepção do cuidador é que o suporte é mais eficaz, o que aumenta a adesão ao uso do sistema.

Agradecimentos. Agradecemos o auxílio financeiro da FAPERJ e CNPq o apoio do InovUERJ no processo de registro do software junto ao INPI e nos pedidos de patente.

8. Referências

- Caldas, C. P. (2003) “Envelhecimento com dependência: responsabilidades e demandas da família”, *Cadernos de Saúde Pública*, vol.19, no. 3, pp. 773-81.
- Celler, B. G., Basilakis, J., Budge, M. e Lovel, N. H. (2006) “A clinical monitoring and management system for residential aged care facilities”, *28th IEEE EMBS Annual International Conferences*, pp. 3301-3304, New York, EUA.
- Dorsey, E. R. (2013) “Randomized controlled trial of “virtual houses calls” for Parkinson disease”, *JAMA Neurology*, vol. 70, no. 5, pp. 565- 70.
- Finkelstein, S. M., Speedie, S. M. e Potthoff, S. (2006) “Home telehealth improves clinical outcomes at lower cost for home healthcare”, *Telemed. Journal and E-health*, vol. 12, no. 2, pp. 128-38.
- Fisher, G. S. (2006) “Brief Communication: Electronic Reminds to Patients Within an Interactive Patients”, *Telemed. Journal and e-Health*, vol.12, no. 2, pp. 128-38.
- Google, “Android”, <http://www.android.com/about/>
- Jaglal, S.B., Haroun, V.A., Salbach, N.M., Hawker, G., Voth, J. e Lou, W. (2013) “Increasing access to chronic disease self-management programs in rural and remote communities using telehealth”, *Telemed. Jnl. and e-Health*, vol.19, no.6, pp.467-74.
- Keranen, T. e Liikkanen, S. (2013) “Medication reminder service for mobile phones: an open feasibility study in patients with Parkinson's disease”, *Telemed. Journal and e-Health*, vol.19, no. 11, pp. 888-90.
- MinS-Ministério da Saúde, Brasil (2006) “Envelhecimento e saúde da pessoa idosa”, *Cadernos de Atenção Básica*, n. 19. Ministério da Saúde. <http://dab.saude.gov.br/portaldab/biblioteca.php?conteudo=publicacoes/cab19>
- Mosa, A. S. M., Yoo, I. e Sheets, L. (2012) “A Systematic Review of Healthcare Applications for Smartphone”, *BioMed Central – Medical Informatics and Decision Making*, vol. 12, pp. 67.
- Nicholas, D., Huntington, P. e Jamali, H. (2007) *Digital Health Information for the Consumer: Evidence and Policy Implications*, Ashgate Publishing, Ltd., England.
- Prince, M., Prina, M. e Guerchet, M. (2013) “World Alzheimer Report 2013. Journey of Caring An analysis of long-term care for dementia”, *Alzheimer’s Disease International (ADI)*, London. www.alz.co.uk/worldreport2013.
- Sanches, L. M., Alves, D. S., Lopes, M. H. e Novaes, M. A. (2012) “The practice of telehealth by nurses: an experience in primary healthcare in Brazil”, *Telemed J E Health*, vol. 18, No 9, pp. 679-83.
- Stutzel, M. C, Filippo, M. P., Sztajnberg, A., Brittes, A., Motta, L. B. (2016) “SMAI – Mobile System for Elderly Monitoring”, *4th IEEE SeGAH International Conference on Serious Games and Applications for Health*, Orlando, EUA.
- UNFPA-United Nations Population Funds (2012) “Ageing in the Twenty-First Century: A Celebration and A Challenge”, Madrid. <http://www.unfpa.org/webdav/site/global/shared/documents/publications/2012/UNFPA-Exec-Summary.pdf>.
- Wootton, R., Craig, J. e Patterson, V. (2006) *Introduction to telemedicine*, Second Edition. London: The Royal Society of Medicine Press Ltd. 206p.

Avaliação da aceitação do *game* Motion Rehab por idosos com sequelas de AVE

Ana Carolina Bertoletti De Marchi^{1,2,3}, Rafael Rieder^{1,3}, Eliane Lucia Colussi²,
Magliani Reis Fiorin Martel², Mateus Trombetta³, Nathália Pinto Cechetti¹

¹Programa de Pós-Graduação em Computação Aplicada – Universidade de Passo Fundo (UPF) - BR 285 – Bairro São José – CEP: 99052-900 - Passo Fundo – RS – Brasil

²Programa de Pós-Graduação em Envelhecimento Humano - Universidade de Passo Fundo (UPF)

³Curso de Ciência da Computação – Universidade de Passo Fundo (UPF)

{carolina,rieder,colussi}@upf.br, maglifiorin@yahoo.com.br,
mateus.trombetta@hotmail.com, nathalia.cechetti@gmail.com

Abstract. *The aim of the research was to evaluate the acceptance of the Motion Rehab game in elderly patients with stroke during a rehabilitation process. Ten men and women participated of the experiment, aged of 67.7 ± 7.1 years, predominantly ischemic stroke and frequenters of a Physical Rehabilitation Center. For our approach, we used a questionnaire based on the Technology Acceptance Model (TAM). The results show that the women elderly tend to better acceptance the use of game as a rehabilitation tool. The evaluation also points out that the game is a motivating factor, and stimulate performing exercises during each physiotherapy session.*

Keywords. *Activities of daily living, Aging, Attention, Games, Stroke.*

Resumo. *O objetivo da pesquisa foi avaliar a aceitação do game Motion Rehab por idosos com sequelas de AVE que o utilizaram em um processo de reabilitação. Participaram 10 idosos de ambos os sexos, com idade média de $67,7 \pm 7,1$ anos, com predominância do AVE isquêmico e frequentadores de um Centro de Reabilitação Física. Para a avaliação foi utilizado um questionário construído com base no Modelo de Aceitação de Tecnologia (Technology Acceptance Model - TAM). Os resultados mostram que as idosas têm uma tendência maior a aceitar o game como ferramenta auxiliar de reabilitação. A avaliação também aponta que o game é um elemento motivador, servindo de estímulo à realização dos exercícios durante cada sessão de fisioterapia.*

Palavras-chave. *Acidente Vascular Encefálico, Atenção, Atividades Cotidianas, Envelhecimento, Jogos.*

1. Introdução

O Acidente Vascular Encefálico (AVE) é uma das patologias neurológicas cada vez mais recorrentes em adultos e idosos, sendo que quando não resulta em óbito, deixa

sequelas persistentes por toda a vida. De acordo com a *World Health Organization* (2013), trata-se de uma doença decorrente de uma interrupção da circulação sanguínea cerebral, classificada pela presença de hemorragia (rompimento de vasos sanguíneos) ou por isquemia (bloqueio da artéria por um coágulo), resultando em lesão do tecido cerebral.

As sequelas motoras do AVE são caracterizadas a partir da sua extensão e região afetada [Langhorne, *et al.*, 2011] e podem atingir membro superior, tronco e membro inferior [Coupar *et al.*, 2012], prejudicando, entre outros aspectos, a mobilidade, o equilíbrio, a marcha e a coordenação motora [Barbosa, 2012]. A presença de tais alterações compromete a execução das atividades de vida diária com interferência direta na qualidade de vida, o que torna fundamental a inserção do paciente em um processo de reabilitação o mais breve possível.

Dentre os processos de reabilitação atualmente disponíveis, pesquisadores estão inserindo em seus protocolos os *games* [Kim *et al.*, 2011; Pompeu *et al.*, 2014 e Vogiatzaki e Krukowski, 2014], que se caracterizam como uma ferramenta complementar, motivadora e diferenciada para o tratamento terapêutico. No entanto, a prova de sua eficácia entre os idosos ainda necessita de maiores pesquisas [Molina *et al.*, 2014], principalmente no que se refere a sua aceitação.

Estudos como o de Hanif *et al.* (2011) sobre a aceitação de tecnologias em processo de reabilitação, comprovam que o seu uso como complemento em tratamentos fisioterapêuticos apresenta resultados positivos aos pacientes. Os autores mencionam facilidade de uso, acessibilidade, baixo custo da tecnologia, incentivo ao aumento da intensidade dos exercícios e motivação, como fatores desencadeadores desta aceitação. Da mesma forma, Nawaz *et al.* (2015), em uma revisão que avaliou a aceitação do uso de tecnologia em processos de reabilitação com idosos, concluíram que os participantes da maioria dos estudos incluídos aprovaram a experiência e sentiram-se satisfeitos.

Adicionalmente, Liu *et al.* (2015) apontam que esta aceitação também ocorre pelos profissionais terapêuticos. Os resultados do estudo indicam que em centros de reabilitação onde o uso de tecnologia nos tratamentos não é obrigatório, os profissionais, quando incentivados a incluí-la, passam a aceitá-la percebendo-a como um facilitador no tratamento. No entanto, ainda existe a necessidade de uma maior variedade de jogos e da diversificação de treinamentos [Hanif, Niaz e Khan, 2011].

Diante deste contexto, este trabalho tem como objetivo avaliar a aceitação do *game* Motion Rehab por idosos com sequelas de AVE que o utilizaram em um processo de reabilitação. Na sequência, o artigo apresenta a metodologia, os resultados, a discussão e a conclusão sobre o experimento.

2. Metodologia

Participaram do estudo 10 idosos com sequelas de AVE, de ambos os sexos, que realizaram um programa de reabilitação física e cognitiva proposto por Martel (2015), com o uso do *game* Motion Rehab. A intervenção foi realizada em 15 sessões, com duração de 45 minutos cada, executadas individualmente, uma vez por semana, em um Centro Regional de Reabilitação Física do Noroeste do Estado do Rio Grande do Sul.

O *game* Motion Rehab (registrado no INPI sob número BR 51 2014 001087-2) foi desenvolvido por Fiorin *et al.* (2014) com base na fisioterapia convencional. O *game* possui uma interface 2D e é composto por quatro fases divididas em exercícios de flexão e abdução de ombro; extensão de cotovelo; extensão e flexão de quadril e joelho (Figura 1). Para a interação com o *game* os idosos realizavam os exercícios de frente para uma TV de 32 polegadas conectada a um computador e ao sensor de movimento Kinect, localizado abaixo da TV.



Figura 1. Fase 3 do Motion Rehab

Ao término do programa de intervenção, os idosos responderam ao questionário de avaliação de aceitação, construído com base no Modelo de Aceitação de Tecnologia (*Technology Acceptance Model* - TAM) proposto por Davis, Bagozzi e Warshaw (1989). O modelo é utilizado por pesquisadores para descrever a aceitação de tecnologias da informação após o seu uso [Dias *et al.*, 2011]. A avaliação sustenta-se em três categorias: **utilidade percebida**, que determina o grau em que se acredita que o uso de uma tecnologia pode melhorar o desempenho e a produtividade; **facilidade de uso**, que corresponde ao grau em que se acredita que o uso do sistema de informação será livre de esforço, ou seja, fácil de aprender e interagir; e **variáveis externas**, que fornecem uma melhor compreensão do que influencia a utilidade percebida e a facilidade de uso.

O questionário utilizado conta com 16 questões agrupadas nas três categorias, conforme mostra a Tabela 1. As respostas foram classificadas em uma escala de Likert de cinco pontos, sendo cinco o maior valor: concordo totalmente (5), concordo parcialmente (4), indiferente (3), discordo parcialmente (2) e discordo totalmente (1).

O estudo seguiu as Diretrizes e Normas Regulamentadoras de Pesquisas Envolvendo Seres Humanos, Resolução no 466/2012, e foi aprovado pelo Comitê de Ética em Pesquisa da Universidade de Passo Fundo sob parecer de nº 605.652/2014. Todos os participantes receberam informações acerca dos objetivos e procedimentos da pesquisa e assinaram o Termo de Consentimento Livre e Esclarecido.

Os dados foram tabulados e analisados estatisticamente no software SPSS 17. Na análise exploratória foram utilizadas as medidas de média e de desvio padrão (DP),

e o teste Shapiro-Wilk (W) para avaliar a normalidade dos dados ($\alpha = 0,05$, $W_{crítico} = 0,842$).

Tabela 1. Questionário de avaliação da aceitação da tecnologia utilizado

<i>Categoria</i>	
ID	Questão
<i>Utilidade Percebida (U)</i>	
U1	O Motion Rehab é importante e adiciona valor à minha reabilitação.
U2	O Motion Rehab é útil em minha reabilitação.
U3	O Motion Rehab torna minha reabilitação mais interessante.
U4	O Motion Rehab melhora meu desempenho na reabilitação.
U5	O Motion Rehab produz o efeito desejado em minha reabilitação
U6	O Motion Rehab aumenta a minha produtividade na reabilitação.
U7	O Motion Rehab motiva meu envolvimento na reabilitação.
<i>Facilidade de Uso (F)</i>	
F1	Usaria o Motion Rehab frequentemente se fosse fácil de utilizá-lo.
F2	Usaria o Motion Rehab frequentemente se tivesse o equipamento em casa.
F3	A interface do Motion Rehab facilita a realização da reabilitação.
F4	Usar o Motion Rehab é uma boa ideia.
F5	Usar o Motion Rehab é agradável.
<i>Variáveis Externas (V)</i>	
V1	Houve treinamento para usar o Motion Rehab.
V2	O fisioterapeuta tem bom nível de conhecimento sobre o Motion Rehab e ajudou-me a entendê-lo.
V3	Aprender a usar o Motion Rehab foi difícil pra mim.
V4	Frequentemente, confundo-me durante o uso do Motion Rehab.

Nas comparações entre grupos Masculino (GM) e Feminino (GF), utilizaram-se os testes t de Student (para distribuições normais) e Mann-Whitney (para distribuições não normais, $U_{crítico} = 2$), com $p < 0,05$. A hipótese alternativa é de que a avaliação realizada pelos homens é diferente das mulheres.

3. Resultados e Discussão

A Tabela 2 apresenta os resultados da análise exploratória, e a Tabela 3 mostra os resultados da comparação entre grupos Masculino e Feminino. Já a Tabela 4 e a Tabela 5 consideram o agrupamento das respostas obtidas por categoria (Ugeral, Fgeral e Vgeral).

Com relação aos resultados da análise exploratória, nota-se na Tabela 2 e na Tabela 4 que a ampla maioria dos dados se apresenta com distribuição não normal. A exceção se faz para as amostras das questões U7 e F1, que estão normais.

Referente a significância estatística dos resultados, também observa-se na Tabela 3 e na Tabela 5 que somente dois itens apresentam relação significativa (U2 e Ugeral). Em virtude do uso do teste Mann-Whitney para ambos os casos, é preciso também considerar o valor U para comprovar a significância.

Para U2, o valor U calculado é 2,5 e o U crítico é 2 para $p < 0,05$. Logo, o resultado não é estatisticamente significativo.

Já no caso de Ugeral, o valor U calculado é 391,5, e a distribuição é aproximadamente normal – o que confirma a significância do resultado. Nesta situação, pode-se dizer que existe diferença de percepção entre os grupos Masculino e Feminino para a categoria de avaliação “Utilidade Percebida”. Considerando os dados da análise exploratória, relacionados a essa categoria, nota-se que as mulheres acham o uso do *game* Motion Rehab mais útil no processo de reabilitação do que os homens.

Tabela 2. Resultados da análise exploratória

ID Questão	Média ± DP	W calculado
U1	4,5 ± 0,527	0,655
U2	4,4 ± 0,516	0,640
U3	4,2 ± 0,632	0,794
U4	4,1 ± 0,876	0,805
U5	4,2 ± 0,632	0,794
U6	3,9 ± 0,876	0,821
U7	3,9 ± 0,994	0,886*
F1	3,4 ± 1,174	0,873*
F2	4,7 ± 0,483	0,594
F3	4,1 ± 0,738	0,833
F4	4,7 ± 0,483	0,594
F5	4,5 ± 0,527	0,655
V1	4,5 ± 0,527	0,655
V2	4,6 ± 0,516	0,640
V3	2,2 ± 0,789	0,751
V4	2,6 ± 1,265	0,794

Tabela 3. Resultados da comparação entre grupos Masculino e Feminino

ID Questão	GM – Média ± DP	GF – Média ± DP	p(value)
U1	4,4 ± 0,548	4,6 ± 0,548	0,674
U2	4,0 ± 0,000	4,8 ± 0,447	0,048*
U3	4,0 ± 0,707	4,4 ± 0,548	0,465
U4	3,8 ± 1,095	4,4 ± 0,548	0,401
U5	3,8 ± 0,447	4,6 ± 0,548	0,095
U6	3,8 ± 1,095	4,0 ± 0,707	1,000
U7	3,6 ± 1,140	4,2 ± 0,837	0,374
F1	2,8 ± 1,304	4,0 ± 0,707	0,120
F2	4,4 ± 0,548	5,0 ± 0,000	0,144
F3	3,8 ± 0,837	4,4 ± 0,548	0,298
F4	4,6 ± 0,548	4,8 ± 0,447	0,674
F5	4,6 ± 0,548	4,4 ± 0,548	0,674
V1	4,4 ± 0,548	4,6 ± 0,548	0,674
V2	4,6 ± 0,548	4,6 ± 0,548	0,920
V3	2,0 ± 0,707	2,4 ± 0,894	0,674
V4	2,2 ± 1,095	3,0 ± 1,414	0,401

Tabela 4. Resultados da análise exploratória, agrupados por categoria

ID	Média ± DP	W calculado > W crítico
Ugeral	4,171 ± 0,742	0,802 > 0,947
Fgeral	4,280 ± 0,858	0,747 > 0,947
Vgeral	3,475 ± 1,358	0,825 > 0,940

Tabela 5. Resultados da comparação entre grupos Masculino e Feminino, agrupados por categoria

ID	GM – Média ± DP	GF – Média ± DP	p(value)
Ugeral	3,914 ± 0,781	4,429 ± 0,608	0,010*
Fgeral	4,040 ± 1,020	4,520 ± 0,586	0,110
Vgeral	3,300 ± 1,418	3,300 ± 1,418	0,435

De maneira geral, analisando os resultados das três categorias de avaliação, percebe-se que, em média, o público feminino tem uma aceitação melhor do uso de tecnologias como ferramentas de intervenção auxiliares – o que era esperado, dado que as mulheres são mais abertas às novidades e mais engajadas em reabilitações diferenciadas. Tal fator também é percebido nos estudos de Kim *et al.* (2011) e Molina *et al.* (2014), por meio de uma proporção maior de mulheres nas amostras das pesquisas apontadas. Também nota-se que as mulheres apresentam, em média, um pouco mais de dificuldade para aprender a interagir com a ferramenta, e que, eventualmente, se confundem mais com os elementos virtuais do *game* do que os homens – também esperado, pois os homens tendem a ter melhor noção de espaço [Halpern, 2013].

Observa-se também que o público masculino, em média, sente que a intervenção é mais agradável com o uso do *game*. Isso pode demonstrar que os homens idosos tem interesse pela ferramenta, porém, precisam sentir mais segurança de movimentos com o *game*, para ter o mesmo resultado que imaginam ter com as ações cotidianas do fisioterapeuta.

Relacionado ao contexto do uso do *game* Motion Rehab observou-se inicialmente que os idosos tinham certa dificuldade na interação com o computador. Isso se justifica pois, para a grande parcela dos idosos, o momento da intervenção foi o primeiro contato com esta tecnologia.

Com o decorrer das sessões, também identificou-se que por meio da repetição semanal utilizando o *game*, os idosos obtiveram maior aceitação frente ao uso desta tecnologia. Durante a intervenção, era possível visualizar a motivação e o entusiasmo dos participantes, corroborando com os achados de Molina *et al.* (2014) e Hanif, Niaz e Khan (2011). Este último apontou alta aceitação do uso do *game* como uma ferramenta nova e interessante na reabilitação da deficiência motora.

Percebeu-se, ainda, que ao mesmo tempo em que eram exigidos movimentos corporais e manutenção de atenção, os estímulos que surgiam durante o jogo tornavam a sessão animada, sendo este um diferencial das sessões convencionais de reabilitação.

4. Conclusão

Este trabalho apresentou os resultados de uma avaliação de aceitação do *game* Motion Rehab por idosos com sequelas de AVE, realizada por profissionais da Saúde e da Computação. A análise estatística apontou, significativamente, que o grupo de pacientes

idosas apresentou uma aceitação melhor ao *game*. Além disso, os observadores do experimento destacaram que, a cada sessão de reabilitação, o uso do *game* estimulava os pacientes a realizarem as atividades propostas pelo fisioterapeuta.

Adicionalmente, a partir dos resultados obtidos, uma nova versão do *game* foi desenvolvida para melhorar sua usabilidade. Na versão 2.0 foram implementadas melhorias quanto, por exemplo, acessibilidade, orientação e *feedback*. Além da versão 2.0, está sendo desenvolvida uma versão 3D do *game* com um cenário imersivo que explora a interação com elementos virtuais espalhados dentro do raio de alcance do paciente. Esta última versão tende motivar ainda mais os pacientes no processo de reabilitação, bem como dar suporte às atividades de domínio espacial.

Como trabalhos futuros, recomenda-se a aplicação do experimento com um grupo maior de pacientes junto ao Centro de Reabilitação Física, visando obter resultados mais significativos. Também sugere-se a disponibilização do *game* para os pacientes utilizarem em seu domicílio. Nesse caso, recomenda-se que uma nova versão do Motion Rehab tenha disponível um módulo para coleta e envio de dados pela *web*, para posterior análise e acompanhamento dos idosos em atividades de reabilitação.

5. Agradecimentos

Os autores agradecem ao CNPq, pelo apoio por meio do Edital MCTI/CNPQ/Universal 14/2014 e pela concessão da bolsa de Iniciação Científica, e à FAPERGS, pela concessão da bolsa de mestrado.

Referências

- Barbosa, M. T. L. M. J. Custos e Efectividade da Reabilitação após Acidente Vascular Cerebral Uma Revisão Sistemática. 2012. 153 p. Dissertação (Mestrado em Gestão e Economia da Saúde) – Universidade de Coimbra, Coimbra.
- Coupar, F.; Pollock, A.; Rowe, P.; Weir, C.; Langhorne, P. Predictors of upper limb recovery after stroke: a systematic review and meta-analysis. *Clinical rehabilitation*, v. 26, n. 4, p. 291–313, 2012.
- Davis, F. D., Bagozzi, R. P., Warshaw, P. R. User Acceptance of computer technology: a comparison of two theoretical models. *Manage Sci.* New York. 1989; 35(8): 982-1003.
- Dias, G. A., Silva, P. M., Junior, J. B. D., Almeida, J. R. Technology Acceptance Model (TAM): avaliando a aceitação tecnológica do Open Journal Systems (OJS). *Informação & Sociedade: Estudos*. 2011; 21(2):133-149.
- Fiorin, M. R.; De Marchi, A. C. B. ; Rieder, R. ; Colussi, E. L.; Trombetta, M. Motion Rehab: um jogo sério para idosos com sequelas de Acidente Vascular Encefálico. In: XIV Workshop de Informática Médica, 2014, Brasília. Anais do CSBS. Brasília, 2014.
- Halpern, D. F. Sex differences in cognitive abilities. Psychology press, 2013.
- Hanif, M., Niaz, H., & Khan, M. A. (2011, June). Investigating the possible role and usefulness of video capture virtual reality in motor impairment rehabilitation. In Next

- Generation Information Technology (ICNIT), 2011 The 2nd International Conference on (pp. 23-30). IEEE.
- Kim, B. R., Chun, M. H., Kim, L. S., Park, Y. Effect of virtual reality on cognition in stroke patients. *Annals of Rehabilitation Medicine*, 2011; 35(4): 450-459.
- Langhorne, P.; Bernhardt, J.; Kwakkel, Gert. Stroke rehabilitation. *Lancet*, v. 377, n. 9778, p. 1693-702, 2011. Elsevier.
- Liu, L., Miguel Cruz, A., Rios Rincon, A., Buttar, V., Ranson, Q., & Goertzen, D. (2015). What factors determine therapists' acceptance of new technologies for rehabilitation—a study using the Unified Theory of Acceptance and Use of Technology (UTAUT). *Disability and rehabilitation*, 37(5), 447-455.
- Martel, M. R. F., Efeitos da intervenção com game na atenção e na independência funcional em idosos após Acidente Vascular Encefálico. Dissertação (Mestrado em Envelhecimento Humano) - Universidade de Passo Fundo, Passo Fundo, 2015.
- Molina, K. I., Ricci, N. A., Moraes, A. S., Perracini, M. R. Virtual reality using games for improving physical functioning in older adults: a systematic review. *Journal of NeuroEngineering and Rehabilitation*, 2014; 11(156).
- Moraes, E. N; Moraes, F. L., Lima, S. P. P. Características biológicas e psicológicas do envelhecimento. *Revista Médica de Minas Gerais, Minas Gerais*, v. 20, n. 1, p. 67-73, 2010.
- Nawaz, A., Skjæret, N., Helbostad, J. L., Vereijken, B., Boulton, E., & Svanaes, D. (2015). Usability and acceptability of balance exergames in older adults: A scoping review. *Health informatics journal*, 1460458215598638.
- Pompeu, J. E., Alonso, T. H., Bordello, I. M., Alvarega, A. P. S. M., Torriani-Pasin, C. Os efeitos da realidade virtual na reabilitação do acidente vascular encefálico: Uma revisão sistemática. *Motricidade*, 2014; 10 (4): 111-122.
- Vogiatzaki, E., Krukowski, A. Serious Games for Stroke Rehabilitation Employing Immersive User Interfaces in 3D Virtual Environment. *Journal of Health Informatics*. 2014; 6: 105-13.
- World Health Organization. Stroke, Cerebrovascular accident: Health topics. Geneva, 2013. Disponível em: < http://www.who.int/topics/cerebrovascular_accident/en/>. Acesso em: 22 jun. 2013.

Sistema Integrado de Análise e Visualização Epidemiológica Georreferenciada Online para Telemedicina

Andrei de Souza Inácio^{1,2}, Alexandre Savaris², João Marcus Alves^{1,2}, Aldo von Wangenheim^{1,2}

¹Programa de Pós-Graduação em Ciência da Computação (PPGCC)
Universidade Federal de Santa Catarina (UFSC)
Florianópolis, SC – Brasil

²Laboratório de Telemedicina
Instituto Nacional para Convergência Digital (INCoD)
Florianópolis, SC – Brazil

{andrei,savaris,joao}@telemedicina.ufsc.br, aldo.vw@ufsc.br

Abstract. *This paper presents a georeferenced epidemiological analysis tool that is integrated into the infrastructure of the Santa Catarina State Integrated Telemedicine and Telehealth System (STT/SC). GISTeMed integrates STT/SC with the Geographic Information System of Santa Catarina (SIG@SC) and supports recovery, georeferenced visualization and epidemiological analysis on structured and semi-structured data. Results from a case study indicate a positive perception of the usefulness of GISTeMed by experts and health managers.*

Resumo. *Este trabalho apresenta o GISTeMed, uma ferramenta de análise epidemiológica georreferenciada que é integrada à infraestrutura do Sistema Catarinense Integrado de Telemedicina e Telessaúde (STT/SC). O GISTeMed integra o STT/SC com o Sistema de Informações Geográficas de Santa Catarina (SIG@SC) e permite a recuperação, visualização georreferenciada e análise epidemiológica em dados estruturados e semiestruturados. Resultados obtidos através de um estudo de caso indicam uma percepção positiva em relação à utilidade do GISTeMed por especialistas e gestores de saúde.*

1. Introdução

Avanços na análise de dados relativos à mortalidade e sobrevivência populacional têm permitido aos pesquisadores examinar outras dimensões da saúde, medidas pela morbidade, incapacidade, acesso a serviços, qualidade de atendimento, condições de vida e fatores ambientais [Rede Interagencial de Informação para a Saúde 2008], contribuindo para uma observação mais eficaz do estado de saúde da população.

A epidemiologia (uma importante área da saúde pública [Friis and Sellers 2013]) se beneficia da fácil e rápida visualização de dados, pois esta auxilia os epidemiologistas a entenderem as relações entre os potenciais fatores de risco de doenças e suas morbidades e mortalidades associadas [Shi et al. 2007]. Como ferramenta de suporte para aquisição de dados, apesar de não ser muito considerada em situações epidemiológicas, a telemedicina oferece a capacidade de coletar e organizar os dados de pacientes, auxiliando a vigilância epidemiológica através da identificação e acompanhamento de questões de saúde pública [World Health Organization 2010]. O uso da telemedicina também permite o aprimoramento de investigações epidemiológicas, o controle de doenças e a gestão de casos

clínicos [Ohannessian 2015], e eficientemente ajuda os gestores de políticas públicas a direcionar recursos de saúde e profissionais com base em cenários e relações concretas de morbidade [de Souza Inacio et al. 2014].

Um exemplo de sistema de telemedicina amplamente utilizado é o Sistema Catarinense Integrado de Telemedicina e Telessaúde (STT/SC). Desenvolvido no ano de 2005 através de uma parceria entre a Secretaria de Estado da Saúde de Santa Catarina (SES/SC) e a Universidade Federal de Santa Catarina (UFSC) [Maia et al. 2006], o STT/SC fornece serviços síncronos e assíncronos de telemedicina e telessaúde via *web* e dispositivos móveis [de Souza Inacio et al. 2014] que suportam radiologia, eletrocardiografia, dermatologia, análises clínicas e educação em saúde. Disponível em todos os municípios do Estado de Santa Catarina, o STT/SC armazena dados estruturados e semiestruturados de mais de quatro milhões de exames, demandando estratégias para a coleta, organização e visualização dos dados adquiridos por um grande número de profissionais (por exemplo, médicos, profissionais de cuidados primários, administradores de sistema, gestores de saúde) previamente cadastrados.

O georreferenciamento de dados, que consiste na associação de um dado à uma coordenada na superfície terrestre [Moreira et al. 2007], permite a análise da distribuição espacial de incidência de eventos e fornece informações importantes para compreensão, prevenção e monitoramento de doenças [Cavicchioli Neto et al. 2014]. Para o processamento desses dados são usados os Sistemas de Informações Geográficas (SIGs), um conjunto de sistemas de computador com capacidades de coletar, armazenar, gerenciar, analisar e apresentar informações georreferenciadas em mapas temáticos [de Pina and Santos 2010].

Neste trabalho apresentamos o GISTelemed, uma abordagem para a organização, recuperação e análise epidemiológica georreferenciada online dos dados armazenados no STT/SC. Focado em telecardiologia, teledermatologia e doenças infecciosas (hepatite B, HIV, toxoplasmose, dengue, vírus zika e febre chikungunya), o GISTelemed coleta dados estruturados e semiestruturados de diversas fontes disponíveis (Sistemas Gerenciadores de Banco de Dados Relacionais - SGBDRs e Sistemas de Comunicação e Arquivamento de Imagens - PACSs), empregando vocabulários controlados para a catalogação e indexação dos dados e provendo uma interface customizada para busca e visualização.

O restante deste trabalho está organizado como segue: a Seção 2 apresenta os trabalhos relacionados em sistemas de informações epidemiológicas; a Seção 3 descreve a arquitetura do GISTelemed e o modelo de qualidade customizado para sua avaliação. Os dados coletados por meio dos questionários de avaliação de qualidade são analisados e discutidos na Seção 4, orientando as conclusões sobre a viabilidade da abordagem de avaliação apresentadas na Seção 5.

2. Trabalhos Relacionados

Várias abordagens têm sido desenvolvidas com foco em análise de doenças infecciosas, tais como cólera, gripe H1N1 e leishmaniose [Rasam et al. 2011, Jiang et al. 2011, Jian-rong and Yan-jun 2010, Amoli 2011], usando um SIG como uma forma de melhorar a qualidade, acesso e uso das informações de saúde. As ferramentas baseadas em SIG podem ser úteis também para pesquisadores e gestores de políticas públicas de saúde, por permitir a exploração de mapas para entender a dinâmica espacial das doenças

[Kandwal et al. 2009]. Com um enfoque diferente, outras abordagens concentram-se no monitoramento das condições de saúde e alerta de doenças através de dados coletados a partir de mídias e redes sociais [Freifeld et al. 2008, Widener and Li 2014, Parker et al. 2013], explorando técnicas de recuperação de informação e classificação como uma alternativa para identificação de doenças.

No contexto da saúde pública, o CDC Wonder¹ e OASIS² fornecem recursos de informação históricos para pesquisadores e profissionais de saúde pública, a fim de auxiliar o processo de tomada de decisão. A população em geral pode acessar informações detalhadas sobre saúde pública, especificando critérios de pesquisa e visualizando os resultados em tabelas ou mapas.

No contexto da telemedicina, algumas abordagens têm sido desenvolvidas com o intuito de coletar, processar e indexar dados estruturados e semiestruturados [Farruggia et al. 2014, Costa et al. 2009]. Essas soluções são integradas a sistemas PACS, permitindo a busca e recuperação de imagens e laudos de exames; contudo, não há integração com SIGs para visualização georreferenciada dos dados.

O GISTelemed (apresentado neste trabalho) difere das ferramentas acima mencionadas por integrar no mesmo sistema a capacidade de indexar dados estruturados (relacionais) e semiestruturados (laudos armazenados no padrão DICOM SR – *Digital Imaging and Communications in Medicine - Structured Reports*), buscar e recuperar informações a partir de pesquisas estruturadas e em texto livre, e apresentar os resultados em tabelas, gráficos e mapas, todos integrados à uma aplicação de telemedicina (STT/SC).

3. Materiais e Métodos

Na infraestrutura do STT/SC, o GISTelemed é integrado como um módulo ao Sistema de Informações STT, usando tecnologias relacionadas à extração, indexação, busca, recuperação e visualização de dados. Através da interface do usuário, é possível avaliá-lo de acordo com a perspectiva de seus usuários finais em termos de qualidade percebida.

3.1. A Arquitetura do GISTelemed

O módulo GISTelemed, apresentado na Figura 1, estrutura-se em cinco componentes, distribuindo as tecnologias adotadas para atender aos processos ETL (*Extract Transform Load*), o componente motor de busca, as operações de pré-processamento e pós-processamento e a interface do usuário. O processo ETL é responsável por acessar as fontes de dados estruturadas (bancos de dados relacionais) e semiestruturadas (PACSs), coletando dados para indexação. Processos individuais foram criados para atender às características específicas de cada conjunto de dados (exames de HIV, dengue, hepatite B, toxoplasmose, vírus Zika e febre Chikungunya, extraídos do sistema do Laboratório Central de Saúde Pública de Santa Catarina - LACEN/SC; laudos de exames de cardiologia e dermatologia extraídos do Sistema de Informações do STT e do Servidor DICOM - PACS). Esses processos são executados via Mirth[®] Connect³ v.3.2.0 (uma ferramenta *open-source* para envio bi-direcional de mensagens por meio de diversos protocolos). Os

¹<http://wonder.cdc.gov/>

²<https://oasis.state.ga.us/>

³<https://www.mirth.com/Products-and-Services/Mirth-Connect>

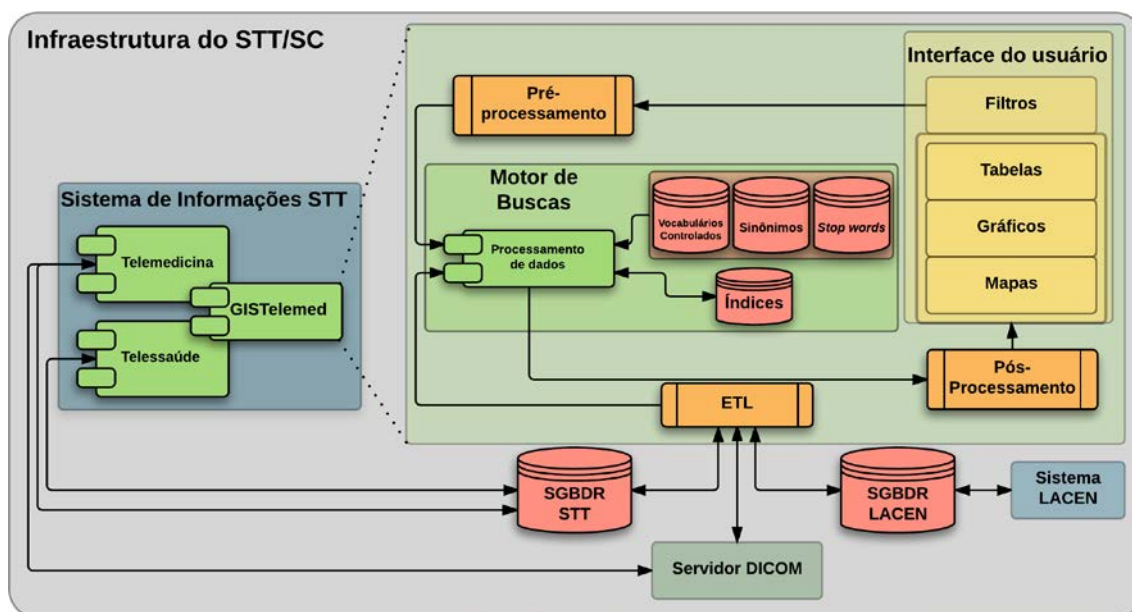


Figura 1. Arquitetura do módulo GISTelemed integrado ao STT/SC.

dados coletados são processados e convertidos em documentos XML (*eXtensible Markup Language*) e enviados via método HTTP POST (*HyperText Transfer Protocol*) para o componente motor de busca.

O componente motor de busca, baseado em Apache Solr⁴ v.4.10 (um servidor de indexação e buscas *open-source*), é responsável por receber e indexar os dados provenientes do processo ETL; ele responde, também, pela execução de buscas e recuperação de documentos de acordo com os termos presentes na pesquisa recebida via interface do usuário. Ambos os passos de indexação e recuperação empregam as técnicas de recuperação de informação *stemming* e *stop words* para melhorar o desempenho das consultas em texto livre.

As operações de pré-processamento e pós-processamento são responsáveis, respectivamente, por padronizar os filtros e os campos de filtro selecionados antes da submissão das instruções de pesquisa ao motor de busca, e pela normalização dos resultados que são apresentados em tabelas, gráficos e mapas.

A interface do usuário (apresentada na Figura 2) foi desenvolvida para permitir que usuários experientes e inexperientes possam realizar consultas e análises epidemiológicas. Através da interface de filtros (a) é possível selecionar os parâmetros de busca a partir de um conjunto pré-definido de filtros, bem como usar texto livre; a interface também permite escolher os dados que serão apresentados como resultado, usando um mecanismo de arrastar-e-soltar. Através da interface de resultados é possível visualizar o resultado das buscas em gráficos (b), tabelas (c), ou em três tipos de mapas (ocorrências (d), calor (e), e coroplético (f)) usando o formato GeoJSON⁵ e a biblioteca OpenLayers⁶.

O mapa do Estado de Santa Catarina é carregado usando o WMS (*Web Map Ser-*

⁴<http://lucene.apache.org/solr/>

⁵<http://geojson.org/>

⁶<http://openlayers.org/>

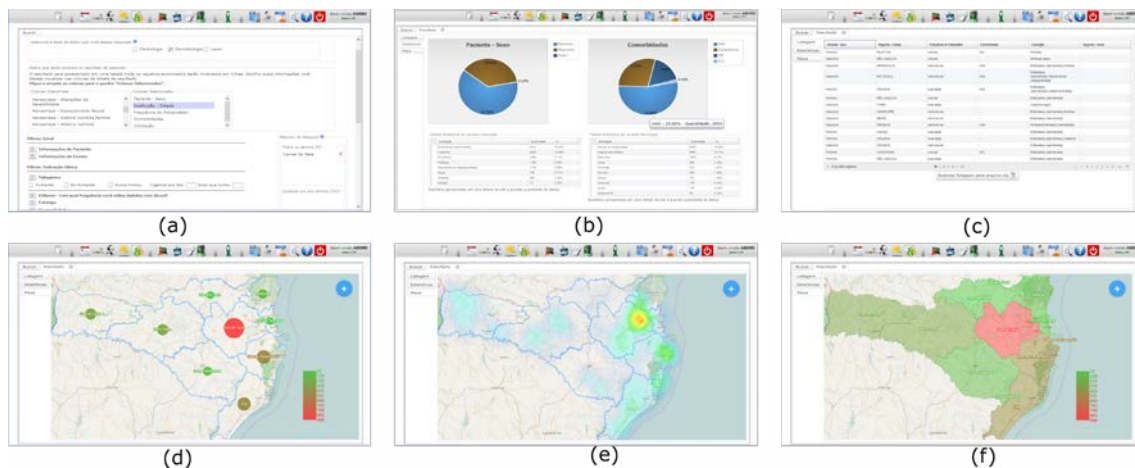


Figura 2. Interface do módulo GISTelemed – filtros (a), resultados em gráficos (b), tabelas (c), mapas: ocorrências (d), calor (e) e coroplético (f).

vice) disponibilizado pelo SIG@SC⁷, um serviço desenvolvido através de uma parceria entre a UFSC e o Centro de Informática e Automação do Estado de Santa Catarina (CIASC) que oferece acesso público aos dados geográficos vetoriais e matriciais (*raster*). Os dados plotados em mapas podem ser visualizados por estabelecimento de saúde, município ou região, permitindo um acesso rápido à distribuição espacial de doenças específicas ou exames realizados.

3.2. Estratégia de Validação

Com o objetivo de avaliar a qualidade de *software* do módulo GISTelemed foi realizado um estudo de caso com os usuários finais do STT/SC (incluindo médicos, gestores de políticas de saúde pública e pesquisadores). Os participantes foram convidados a seguir um roteiro predefinido, utilizando o módulo GISTelemed para executar tarefas de pesquisa e recuperação de dados. A tarefa solicitada consiste na pesquisa do número de ocorrências de uma doença específica (de acordo com a especialidade do usuário), a análise estatística das variáveis idade, sexo e cidade do paciente, e a distribuição dos dados no mapa; sua avaliação foi realizada por meio de um questionário.

De um total de 37 usuários convidados, 26 aceitaram participar da avaliação (taxa de aceitação de 70,3%). Concluída a tarefa, os usuários responderam ao questionário com a finalidade de colher dados sobre a sua percepção de qualidade relativa ao módulo GISTelemed. O questionário foi personalizado de acordo com o AdeQUATE (*questionnaire for Evaluation of QUALity in TElemedicine*) [Alves et al. 2015], um modelo de avaliação para atender às especificidades e necessidades dos sistemas de telemedicina baseado na norma ISO/IEC 25010 [Iso.org 2011], no TAM (*Technology Acceptance Model*) [Davis 1989], e no SUS (*System Usability Scale*) [Brooke 1996]. O modelo decompõe conceitos abstratos de qualidade através dos fatores de qualidade da norma ISO/IEC 25010 com respeito à sua relevância para sistemas de telemedicina. Todos os itens do questionário foram formulados como afirmações, com o objetivo de eliminar erros de interpretação e reduzir a probabilidade de erros nas respostas [Sauro and Lewis 2011]; uma escala Likert de 4 pontos [Boone and Boone 2012] foi usada, incluindo as alterna-

⁷<http://sigsc.sds.sc.gov.br/>

tivas “Não se aplica”, “Não sei responder” e “Não entendi o item”. As respostas aos questionários foram usadas para computar a distribuição de cada opção respondida para cada item de avaliação, usando uma abordagem semelhante ao SUS para sumarização dos dados.

4. Resultados e Discussões

Os dados adquiridos através do questionário foram analisados usando estatísticas descritivas, adotando a mediana como um representante para o meio da distribuição, e quartis para agrupamento [Boone and Boone 2012]. As pontuações das respostas dadas para cada item do questionário foram calculadas de maneira similar ao método usado pelo SUS [Finstad 2010], conforme fórmula apresentada em (1).

$$S_n = \left(\sum_{i=1}^{26} W(\tilde{R}_i) \right) * \frac{100}{78} \quad (1)$$

Em (1), \tilde{R}_i corresponde à média das respostas obtidas no questionário para a subcaracterística n pelo participante i , e $W(\tilde{R}_i)$ é uma função-peso com base em uma opção de resposta assumindo os seguintes valores: 3 para “Concordo totalmente”, 2 para “Concordo”, 1 para “Discordo” e 0 para “Discordo totalmente”. As opções de respostas “Não sei responder”, “Não se aplica” e “Não entendi o item” são ignoradas. O último componente em (1) garante que S_n varie de 0 (pior grau de qualidade – todos os 26 participantes escolheram a opção “Discordo totalmente”) até 100 (grau excepcional de qualidade – todos os 26 participantes escolheram a opção “Concordo totalmente”).

As pontuações de cada subcaracterística são apresentadas na Figura 3 (b), variando de 0 (pior grau de qualidade) até 100 (melhor grau de qualidade). Nós podemos observar que “Aprendizado”, “Eficiência”, “Conforto”, “Utilidade” e “Corretude Funcional” foram as subcaracterísticas que receberam as melhores avaliações, indicando a boa percepção e aceitação do sistema do ponto de vista do usuário final. Algumas subcaracterísticas, principalmente as presentes no quartil inferior, receberam baixa pontuação, pois as opções de resposta “Não se aplica”, “Não entendi o item” e “Não sei responder” não foram consideradas para o cálculo. Além disso, as subcaracterísticas com pontuação menor que 50 tiveram menos de 60% das respostas respondidas positivamente ou negativamente, conforme apresentado na Figura 3 (a). Nós também podemos observar na Figura 3 (b) que, de acordo com os usuários, as piores subcaracterísticas avaliadas foram “Acessibilidade”, “Interoperabilidade”, “Recuperabilidade” e “Tolerância a falhas”. No entanto, uma vez que a média é maior que 50, pode-se notar uma percepção positiva da qualidade do GISTeamed. O intervalo interquartil (IQR) é de 16, demonstrando uma dispersão baixa.

A Figura 3 (a) apresenta a proporção das respostas de acordo com as subcaracterísticas do padrão ISO/IEC 25010. Nós podemos observar diferenças na posição de algumas subcaracterísticas quando comparadas com a Figura 3 (b); isso ocorre pois o gráfico apresentado na Figura 3 (a) apresenta a distribuição das respostas, e ele é apresentado em ordem decrescente de pontuação. “Eficiência”, por exemplo, apesar de ter a mesma pontuação da subcaracterística “Aprendizado” calculada por (1), tem uma pior avaliação devido ao maior número de respostas negativas. Também é possível constatar que mais de 40% dos usuários responderam “Não se aplica” ou “Não sei responder” para as subcaracterísticas “Tolerância a falhas”, “Recuperabilidade”, “Interoperabilidade”, e

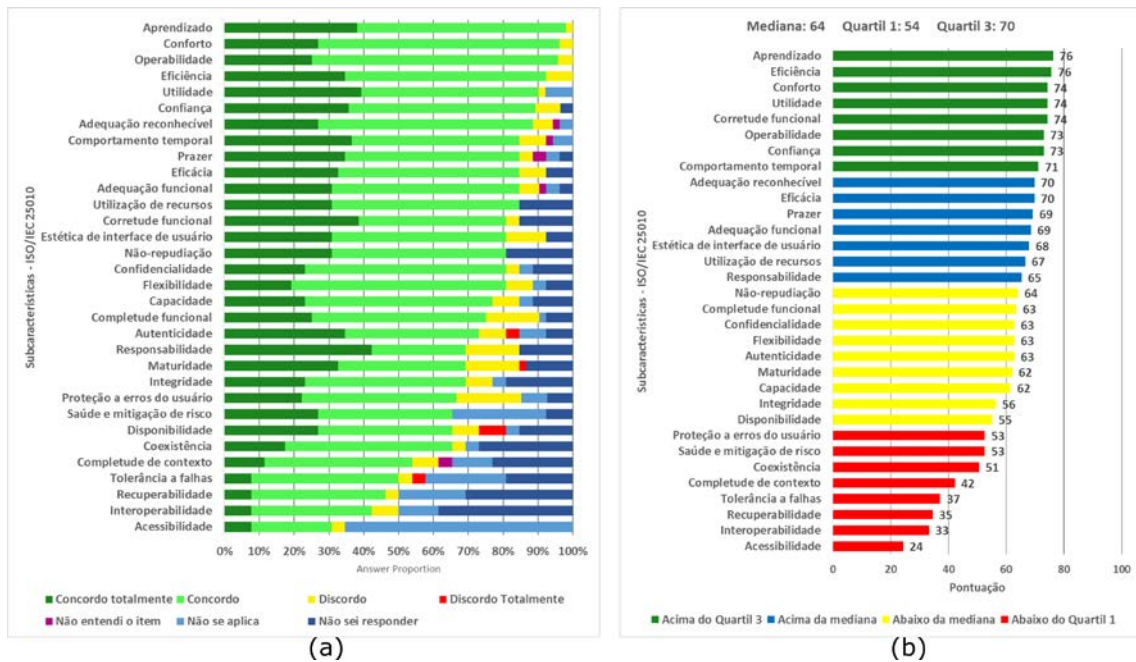


Figura 3. Resultados do estudo de caso de acordo com as subcaracterísticas de qualidade da norma ISO/IEC 25010 - (a) Proporção das respostas, (b) Pontuação da qualidade.

“Acessibilidade”; isso pode indicar que os usuários não conseguem avaliar o item através de um questionário, ou que os usuários não possuem experiência em usar o GISTeamed. “Acessibilidade” teve mais de 60% das respostas como “Não se aplica”, podendo indicar um problema no GISTeamed referente a essa subcaracterística, ou um viés de seleção dos participantes – de todos os participantes, nenhum possuía necessidades especiais.

A norma ISO/IEC 25010 define, também, duas dimensões: *Qualidade em uso*, associada com o grau com que o sistema está em conformidade com os objetivos dos usuários em suas atividades; e *Qualidade do Produto*, relacionada com as propriedades estáticas do software e as propriedades dinâmicas do sistema de computador [Iso.org 2011]. De acordo com a Figura 4, o GISTeamed foi avaliado positivamente por 80,31% dos participantes para a qualidade em uso – apenas 5,41% responderam negativamente, e 14,29% responderam as opções “Não sei responder” e “Não se aplica”. Em relação à qualidade do produto, o GISTeamed foi avaliado positivamente por 74,71% dos participantes; apenas 7,34% responderam negativamente, e 17,94% responderam as opções “Não sei responder” e “Não se aplica”. Considerando que a maioria dos itens

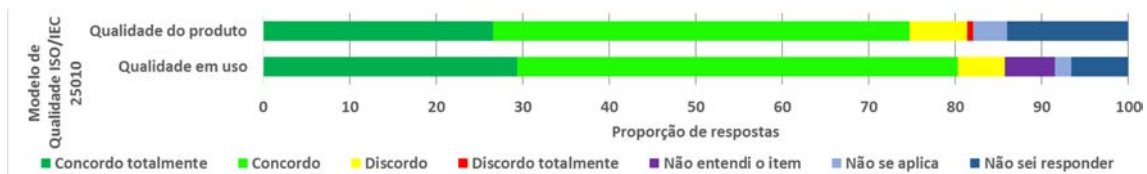


Figura 4. Proporção das respostas do modelo de qualidade ISO/IEC 25010.

foram avaliados positivamente, a mediana calculada para as subcaracterísticas foi 64, e 84,4% das subcaracterísticas tiveram a pontuação maior que 50, nós podemos atestar a

boa percepção da qualidade e aceitação do módulo GISTeamed pelos usuários finais do STT/SC.

4.1. Ameaças à Validade

Uma das ameaças relacionadas com a validade interna é o viés de seleção. Foram convidados apenas médicos pesquisadores, os gestores de políticas de saúde pública e epidemiologistas a responderem o questionário; isso é devido a razões práticas: GISTeamed foi desenvolvido principalmente como um módulo para análise epidemiológica, e estes são os seus potenciais utilizadores.

Uma possível ameaça relacionada ao viés de construção é o uso do modelo de avaliação AdEQUATE. O AdEQUATE não foi extensivamente validado por um grande número de usuários, e no contexto de outros sistemas; no entanto, acreditamos que esta ameaça é minimizada porque o questionário adotado teve seus itens sistematicamente derivados da norma ISO/IEC 25010, TAM e SUS, que são métodos largamente utilizados.

Outra ameaça identificada relacionada com um viés de seleção é que alguns entrevistados tiveram seu primeiro contato com GISTeamed durante o processo de avaliação, o que pode ter influenciado as respostas para alguns itens. Para minimizar esta ameaça, as opções “Não se aplica”, “Não sei responder”, e “Não entendi o item” foram incluídas como respostas válidas.

5. Discussão e Conclusões

Este trabalho apresenta o GISTeamed, uma ferramenta de análise epidemiológica georreferenciada integrada ao sistema de telemedicina STT/SC. Com o objetivo de ser usado por médicos, gestores de políticas de saúde pública e pesquisadores, o módulo tem como objetivo contribuir para o estudo das condições crônicas e infecciosas de doenças, em orientar a tomada de decisões de cuidados de saúde, e em contribuir para o desenvolvimento e avaliação das ações intervencionistas para controle e prevenção de doenças.

Os resultados do nosso estudo de caso utilizando o questionário AdEQUATE mostram que o GISTeamed tem uma boa avaliação em relação à qualidade do software percebida pelos usuários, de acordo com medidas que incluem “Eficácia”, “Eficiência”, “Satisfação”, “Usabilidade”, “Confiabilidade” e “Utilidade”. Entretanto, alguns itens avaliados não tiveram um bom desempenho, como “Acessibilidade”, “Interoperabilidade”, “Recuperabilidade” e “Tolerância a falhas”. Além disso, este trabalho mostra a viabilidade na avaliação de softwares de telemedicina por meio de questionários sistematicamente derivados e personalizados de acordo com o modelo de avaliação AdEQUATE.

A incorporação do georreferenciamento ao sistema, além de possibilitar a visualização dos dados em mapas, permitiu também analisar a evolução de surtos e ocorrências das doenças no tempo. Uma limitação do uso do SIG neste trabalho é a dificuldade de manter frequentemente a base de dados atualizada com as coordenadas dos estabelecimentos de saúde que registram as informações no STT/SC.

Como trabalhos futuros, buscaremos a inclusão de dados de Telessaúde (teleconsultorias), capacidades melhoradas de cruzamento de dados entre modalidades de exames e melhorias na interface do mapa para permitir análise temporal sobre os dados.

Agradecimentos

Este trabalho teve o apoio da Secretaria de Estado da Saúde de Santa Catarina - SES/SC, do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e da Fundação de Amparo à Pesquisa de Santa Catarina - FAPESC.

Referências

- Alves, J. M., Wangenheim, C. G. v., Savaris, A., and Wangenheim, A. v. (2015). *AdEQUATE Software Quality Evaluation Model v1.0*. Brazilian Institute for Digital Convergence, 1 edition.
- Amoli, G. M. (2011). Gis-based risk map analysis of leishmaniasis disease in isfahan, iran. In *IEEE Symposium on Business, Engineering and Industrial Applications (IS-BEIA 2011)*, pages 275–280.
- Boone, H. N. and Boone, D. A. (2012). Analyzing likert data. *Journal of Extension*, 50:1 – 5.
- Brooke, J. (1996). *SUS: A quick and dirty usability scale*. Usability evaluation in industry.
- Caviccholi Neto, V., Chiari, N. S., Pisa, I. T., and Alvez, D. (2014). Desenvolvimento e integração de mapas dinâmicos georreferenciados para o gerenciamento e vigilância em saúde. *Journal of health informatics*, 6(1):4.
- Costa, C., Freitas, F., Pereira, M., Silva, A., and Oliveira, J. L. (2009). Indexing and retrieving dicom data in disperse and unstructured archives. *International Journal of Computer Assisted Radiology and Surgery*, 4(1):71–77.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340.
- de Pina, M. F. and Santos, S. M. (2010). *Conceitos Básicos de Sistemas de Informação Geográfica e Cartografia Aplicados à Saúde*. OPAS Ministério da Saúde.
- de Souza Inacio, A., de Macedo, D. D. J., Andrade, R., and von Wangenheim, A. (2014). Designing an information retrieval system for the stt/sc. In *IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom 2014)*, pages 500–505.
- Farruggia, A., Magro, R., and Vitabile, S. (2014). A text based indexing system for mammographic image retrieval and classification. *Future Generation Computer Systems*, 37(0):243–251.
- Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, 22(5):323–327.
- Freifeld, C. C., Mandl, K. D., Reis, B. Y., and Brownstein, J. S. (2008). Healthmap: Global infectious disease monitoring through automated classification and visualization of internet media reports. *Journal of the American Medical Informatics Association: JAMIA*, 15(2):150–157.
- Friis, R. H. and Sellers, T. (2013). *Epidemiology for public health practice*. Jones & Bartlett Learning.

- Iso.org (2011). Iso/iec 25010:2011 - systems and software engineering – systems and software quality requirements and evaluation (square) – system and software quality models.
- Jian-rong, C. and Yan-jun, L. (2010). Analysis on spatial heterogeneity of h1n1 flu based on gis spatial analysis technology. In *International Conference on Multimedia Technology (ICMT 2010)*, pages 1–4.
- Jiang, S., Zhang, J., Jin, J., Zhang, D., and Wang, T. (2011). Spatial analysis of influenza a h1n1 and evaluation of risk factors in changsha city based on gis. In *19th International Conference on Geoinformatics*, pages 1–5.
- Kandwal, R., Garg, P. K., and Garg, R. D. (2009). Health GIS and hiv/aids studies: perspective and retrospective. *Journal of Biomedical Informatics*, 42(4):748–755.
- Maia, R. S., von Wangenheim, A., and Nobre, L. F. (2006). A statewide telemedicine network for public health in brazil. In *IEEE 19th International Symposium on Computer-Based Medical Systems (CBMS 2006)*, pages 495–500.
- Moreira, R. d. S., Nico, L. S., and Tomita, N. E. (2007). A relação entre o espaço e a saúde bucal coletiva: por uma epidemiologia georreferenciada. *Ciência e Saúde Coletiva*, 12:275 – 284.
- Ohannessian, R. (2015). Telemedicine: potential applications in epidemic situations. *European Research in Telemedicine / La Recherche Européenne en Télémédecine*, 4(3):95–98.
- Parker, J., Wei, Y., Yates, A., Frieder, O., and Goharian, N. (2013). A framework for detecting public health trends with twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASO-NAM '13*, pages 556–563, New York, NY, USA. ACM.
- Rasam, A. R. A., Noor, A. M. M., Ahmad, N., and Ghazali, R. (2011). Mygeohealth: Gis-based cholera transmission risk system in sabah, malaysia. In *IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA 2011)*, pages 474–479.
- Rede Interagencial de Informação para a Saúde (2008). *Indicadores básicos para a saúde no Brasil: conceitos e aplicações*. Organização Pan-Americana da Saúde.
- Sauro, J. and Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 2215–2224, New York, NY, USA. ACM.
- Shi, H., Zhang, Y., Zhang, J., Wan, P., and Shaw, K. (2007). Development of web-based epidemiological reporting system for tasmania utilizing a google maps add-on. In *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on*, pages 118–123.
- Widener, M. J. and Li, W. (2014). Using geolocated twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography*, 54:189–197.
- World Health Organization (2010). Telemedicine: opportunities and developments in member states: report on the second global survey on ehealth 2009. http://www.who.int/goe/publications/goe_telemedicine_2010.pdf.

Relational Databases versus Search Engines: A Performance Comparison for Storing and Querying DICOM Metadata

Alexandre Savaris¹, Gabriela Bussolo Colonetti^{1,2},
Rodrigo Rodrigues Pires de Mello^{1,3}, Aldo von Wangenheim^{1,3}

¹ Telemedicine Laboratory
Brazilian Institute for Digital Convergence – INCoD
Florianópolis, Santa Catarina, Brazil

²Graduate Program in Information Systems – SIN
Federal University of Santa Catarina – UFSC
Florianópolis, Santa Catarina, Brazil

³Graduate Program in Computer Science – CCO
Federal University of Santa Catarina – UFSC
Florianópolis, Santa Catarina, Brazil

{savaris, gabriela, rmello}@telemedicina.ufsc.br, aldo.vw@ufsc.br

Abstract. *The Digital Imaging and Communications in Medicine (DICOM) standard adopts files as individual, self-contained repositories for the storage of a mixed of alphanumeric and binary content regarding radiological images. Usually, groups of DICOM files are hierarchically organized in studies and series, physically disposed into filesystem directory trees. Despite its simplicity in storing content, ordinary filesystems do not provide index capabilities allowing searches by content – restricting access by directory names and file names. To surpass such limitation, Picture Archiving and Communication Systems (PACSs) often adopt Relational Database Management Systems (RDBMSs) as metadata repositories, benefiting from its general-purposed index structures. An alternative approach, not quite explored, considers the adoption of search engines as metadata catalogs, aiming to minimize the search time by exploring the engine's index optimizations. In order to evaluate the performance on managing DICOM metadata, this work compares relational database instances to a search engine in terms of storage space, storage time, and query time. Results show that, in the best case, the search engine is slightly slower in storing content; however, it requires 69% less disk space for the same dataset. For queries, in turn, the search engine performs up to 8.3 times faster in retrieving groups of tags.*

1. Introduction

Picture Archiving and Communication Systems (PACSs) have been developed over three decades [Huang 2011, Lemke 2011], incorporating new technologies and evolving into a complete integrated system far beyond radiology [Faggioni et al. 2011]. Nevertheless, the management of digital images continue to be its most appealing characteristic – resulting in improvements on productivity and efficiency, and in attending new or combined examination modalities [Mansoori et al. 2012, Singh et al. 2011].

Digital images stored and managed by PACSs are structured according to the precepts of the Digital Imaging and Communications in Medicine (DICOM) standard [Pianykh 2012]. Each image is stored as a file, being persisted in ordinary or specialized filesystems (according to the PACS implementation); usually, metadata extracted from images are persisted as well – aiming to provide search capabilities without involving the original files. Relational Database Management Systems (RDBMSs) are commonly adopted as repositories for DICOM metadata. It is possible to define database schemas lined up with the structure of DICOM tags, and the provided index structures contribute on reducing query execution time [Savaris et al. 2014]. Alternatives to RDBMSs have been used, exploring NoSQL implementations as a replacement for relational database instances [Rascovsky et al. 2012, Bastiao Silva et al. 2014]; these alternatives provide schemaless storage capabilities, aligned with native scalability and data partitioning.

This work investigates the behavior of search engines – tools not quite explored in managing DICOM metadata – acting as a replacement for RDBMSs in indexing and responding to queries. Experiments are performed in order to evaluate the performance on storing partial- and full-content metadata extracted from DICOM images, as well as in querying the stored content using different filter options and returning different sets of tags.

The remainder of this work is organized as follows: Section 2 presents background information about the DICOM standard, RDBMSs, RDB (Relational Database) instances, and search engines; Section 3 describes the experimental environment and experiments performed to measure performance on indexing DICOM metadata; Section 4 presents the obtained results, discussed in Section 5; Section 6 concludes the work, including future directions.

2. Background

2.1. The DICOM standard

Originally developed by the National Electrical Manufacturers Association (NEMA) in a partnership with the American College of Radiology (ACR), the DICOM standard covers a set of non-proprietary specifications including structure, format, and exchange protocols for digital-based medical images. Since the release of its first version in 1985 as ACR/NEMA 300, the standard evolves according to deliberations of academies, medical device manufacturers, and scientific societies organized as work-groups [Mildenberger et al. 2002, Bidgood Jr et al. 1997].

According to the standard a DICOM image is stored as a self-contained file, grouping basic data elements (tags). Each tag is characterized by a Value Representation (VR) and a Value Multiplicity (VM), specifying the content supported by the tag, formatting rules applicable to the content, and the number of allowed occurrences for the content inside the tag [National Electrical Manufacturers Association 2015b]. Tags with numerical and textual data types are candidates to be included in search and retrieval operations, being used as filters or as returning values for query expressions.

2.2. Relational Database Management Systems and Relational Database Instances

A Database Management System (DBMS) is a collection of interrelated and persistent data, together with a set of programs responsible for accessing that data and by guaran-

teeing availability, integrity, security, and independence [Sumathi and Esakkirajan 2007]. DBMSs implementing the relational model defined in [Codd 1970] are specified as RDBMSs. RDB instances, in turn, can be defined as a set of physical structures organized according to the relational model, responsible for the effective data storage and managed by RDBMSs.

The physical structure of an RDB instance follows a well-defined database schema, which describes and specifies its component objects (e.g. tables, fields, indexes, integrity constraints) [Elmasri and Navathe 2010]. To be inserted into an RDB instance, data must respect the underlying schema restrictions, which establishes a set of boundaries. Changes in data storage requirements for an RDB instance demand adjustments in its database schema definition.

2.3. Search engines

A search engine can be characterized as a practical application of Information Retrieval (IR) techniques, delivering performance, scalability and adaptability to scenarios demanding ranking, evaluation, and information storage, search and retrieval [Croft et al. 2009]. Commonly organized for storage and management of documents, it has loose restrictions when compared to RDB instances in terms of data schemas.

In performing transformations to the original stored content, including tokenization, stopping, and stemming [Baeza-Yates and Ribeiro-Neto 2011], search engines build high-performance indexes capable of answering to heterogeneous queries, with expressive results when compared to b-tree and hash indexes available on RDBMSs.

3. Materials and Methods

In order to evaluate the behavior of RDB instances and the search engine in managing DICOM content, this work adapts a series of experiments from [Savaris et al. 2014] for measuring storage space, storage time, and query time.

3.1. Experimental environment

The setup adopted for the experiments is based on the client-server model, with two nodes communicating to each other at a time through a 1Gbps LAN as depicted in Fig. 1. Server configuration: Intel® Xeon® X7460 2.66GHz, 1GB DDR3 RAM, 536.9GB SATA-HD, Debian 7.9 wheezy. Client configuration: Intel® Core™ i7-3610QM 2.30GHz, 8GB DDR3 RAM, 500GB SATA-HD, Linux Mint 17.2 rafaella.

For storage experiments, the client side is similar for both RDB instances and the search engine. Using dcm4che¹ v.2.0.28 (an open source implementation of the DICOM standard), a dataset composed by radiological images is read from an ordinary filesystem and sent to the server according to the DICOM network protocol via C-STORE [National Electrical Manufacturers Association 2015a]. The server side, in turn, has particularities for each evaluated implementation. RDB instances (Fig. 1, top) receive the radiological images through dcm4chee² v.2.18.1 (an open source DICOM clinical data management system, acting as a DICOM archive). For each received image, alphanumeric tags are extracted and inserted into a relational database instance managed by PostgreSQL³ v.9.1.18 (an open source ORDBMS – Object-Relational Database Management System); next, the content from the image file is archived into a filesystem directory

tree. The search engine (Fig. 1, bottom), implemented using Elasticsearch⁴ v.2.1.1, receives the radiological images through Mirth[®] Connect⁵ v.3.3.1.7856.b91 (an open source interconnection engine). For each received image, its DICOM content is converted to an XML (eXtensible Markup Language) representation followed by the extraction of its alphanumeric tags, which compose a JSON (JavaScript Object Notation) document whose key-value pairs are fully indexed.

For queries, both client and server sides differ. RDB instances (Fig. 1, top) respond do C-FIND [National Electrical Manufacturers Association 2015a] operations fired from the client using dcm4che. Once received by dcm4chee on the server, the search request is converted to a single or multiple SQL instructions used to access the underlying relational database instance. Tag values who satisfy the search criteria compose a resultset, sent from the ORDBMS to dcm4chee; the archive, then, converts the resultset to a proper format according to the DICOM network protocol, returning it to the requester. The search engine (Fig. 1, bottom) is accessed through HTTP (Hypertext Transfer Protocol) GET requests sent from the client using cURL⁶ v.7.35.0, an open source multi-protocol library and command line tool. Each request (received as a JSON document) is processed by Elasticsearch, who returns another JSON document including the found resultset.

¹<https://dcm4che.atlassian.net/wiki/display/lib/>
²<https://dcm4che.atlassian.net/wiki/display/ee2/>
³<http://www.postgresql.org/>
⁴<https://www.elastic.co/products/elasticsearch>
⁵<https://www.mirth.com/Products-and-Services/Mirth-Connect>
⁶<https://curl.haxx.se/>

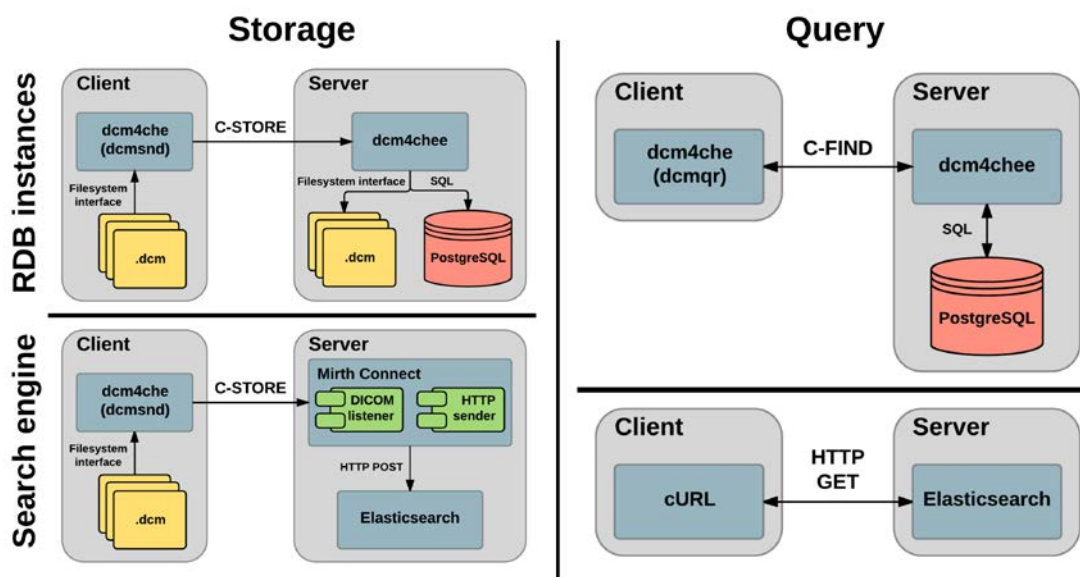


Figure 1. Experimental environment for storage and query measurements using RDB instances and the search engine.

3.2. Experiments

Both storage and query experiments are performed using computed tomography (CT) images from the public available LIDC-IDRI DICOM dataset [The Cancer Imaging Archive Team 2015], an integrating part of The Cancer Imaging Archive (TCIA) [Clark et al. 2013]. The adopted dataset includes 243,954 images with a resolution of 512x512, distributed in 1,010 series and in 1,010 studies belonging to 1,010 patients, occupying 128.4GB of disk space.

Experiments performed for storage measurements send the whole dataset to both RDB instances and to the search engine. Particularly to the relational database instances, two configurations made in dcm4chee are evaluated: the *original* configuration and the *extended* configuration. In the original configuration (depicted in the result charts as OC), only tags that physically exist on the database schema as fields are persisted; remainder tags are ignored. In the extended configuration (depicted in the result charts as EC), tags that physically exist are stored into its respective fields; remainder tags are concatenated into BLOB (Binary Large Object) fields, and become available for filtering and retrieval.

Query experiments are performed on a hierarchical-level basis; for each level defined in the DICOM standard (patient, study, series, and image) a set of tags is retrieved (if available) from the stored dataset. Tags are retrieved individually or as groups of 5, 10, 15, 20, 25, and 30 tags, and are filtered by the unique identifiers of each level (*patientid*, *studyinstanceuid*, *seriesinstanceuid*, and *sopinstanceuid*) up to the top of the hierarchy. All queries are executed 10 times each, using filter values randomly selected from the dataset.

4. Results

The results acquired through experiments described in Section 3 depict the behavior of RDB instances and the search engine in terms of the storage space needed to persist the DICOM dataset, the time spent for the complete persistence, and the time needed to search and retrieve single or multiple tags at once.

4.1. Storage space

The storage space needed for the persistence of the complete DICOM dataset is divided into three volumes: *data*, *indexes*, and *others*. Data comprehend alphanumerical tag values, being stored according to the specificities of each implementation. Indexes are built and updated based on data values; its organization, i.e., storage structure and order, is implementation-dependent. Other structures store complementary data (e.g., data dictionaries, lock control files) used by RDBMSs and search engines for its initial configuration and runtime management. Fig. 2 (left) shows the required space for each volume, for each evaluated implementation.

In observing Fig. 2, it is possible to perceive the lack of the data volume for the search engine. In this work, the search engine indexes JSON documents, each document including the complete metadata set extracted from a DICOM image; therefore, it is assumed that the index volume includes the data volume. RDB instances, in turn, adopt separate physical structures for data and indexes, using the former as tag repositories and the latter as structures for reducing query time. By indexing entire documents, the search

engine requires approximately 51% and 69% less disk space when compared, respectively, to the originally and extended configured RDB instances; when compared alone, the index volume generated by the search engine is approximately 29% and 34% smaller than its counterparts.

The storage space needed for the data volume increases considerably when an extended configuration is adopted for RDB instances: selecting all alphanumerical tags for persistence, the volume grows up approximately three times. Index volumes, in turn, present a slight increasing of 7.73% on size. This behavior indicates that the underlying database schema supports the insertion of new, heterogeneous sets of tags; however, it does not indicate that such tags are indexed.

Despite its importance in regulating RDBMSs and search engine runtimes, the storage space needed by the “others” volume is negligible: it corresponds to only 0.24% of the total space required by all evaluated implementations.

4.2. Storage time

The time spent to persist the whole DICOM dataset is divided into the following operations: *C-STORE*, which includes the transmission of all DICOM images from a client to the server hosting the evaluated implementations respecting the DICOM network protocol; conversions from *DICOM to XML*, performed automatically by the interconnection engine at the reception of each DICOM image; conversions from *XML to JSON*, performed by the interconnection engine to provide content supported by the search engine; and requests and responses via *HTTP POSTs*, performed by the interconnection engine to send data to the search engine for indexing. Individual times for each operation can be seen in Fig. 2 (right).

Common to all evaluated implementations, the C-STORE is also the most time-consuming operation. According to the chart in Fig. 2, storing and indexing all metadata tags contribute to increase the storage time in approximately 5.9 and 5.5 times when comparing the originally configured RDB instance to, respectively, the extended configured RDB instance and the search engine. When comparing similar configurations, i.e., those who store the full set of tags, the search engine performs 7% faster.

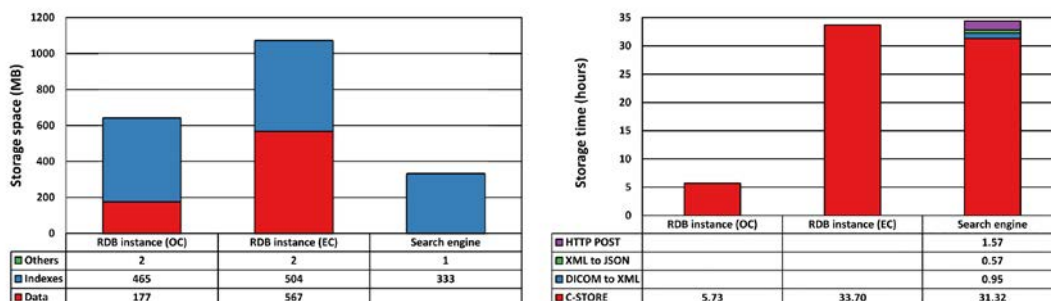


Figure 2. Left: Storage space, in MB, used to persist the whole DICOM dataset. For both RDB instances and the search engine, the stored content is divided into data, indexes, and other structures. Right: Storage time, in hours, used to persist the whole DICOM dataset. For the search engine, the cumulative time includes conversions between data formats and HTTP POST operations.

For both RDB instances, the C-STORE operation encompasses the transmission of DICOM content from client to server, the parsing and extraction of alphanumeric tags, and the insertion of the extracted tags into the underlying database. For the search engine, in turn, the C-STORE operation performs only the transmission of DICOM content from client to server – parsing, extraction, and indexing are performed by conversion operations and HTTP POST operations. In numbers, converting from XML to JSON correspond to 1.65% of the total storage time, followed by the conversion from DICOM to XML (2.76%) and HTTP POST executions (4.55%). Together with the C-STORE operation, conversions and HTTP POSTs contribute to classify the search engine as the worst option for storage, being six times slower than the originally configured RDB instance and approximately 2% slower than the extended configured RDB instance.

4.3. Query time

The time spent on searching and retrieving single or multiple tags at once is a sum of individual times for the following steps: a *query*, sent from a client to the server specifying filtering parameters and tags to be retrieved; the *processing*, including the access to the underlying databases and/or indexes to effectively perform the search; and the *response*, sent from the server to the client with the resulting dataset. These steps are equivalent to a DICOM C-FIND operation [National Electrical Manufacturers Association 2015a].

Fig. 3 depicts the behavior of the evaluated implementations in searching for tags related to the four hierarchical levels defined by the DICOM standard (*patient*, *study*, *series*, and *image*). According to the chart, the greater the number of tags per level, the greater the time needed to search and retrieve such tags. This behavior is consistent in both RDB instances, not considering some exceptions (e.g., individual tags, extended RDB instance). Being approximately 10% slower than the originally configured RDB instance, the extended configuration is penalized by the overhead on managing tag values stored into BLOB fields. Despite its flexibility in storing heterogeneous sets of tags, this approach compromises the search performance due to the parsing step needed to identify and extract tags that are not persisted as individual, schema-defined fields.

When compared to both RDB instances, the search engine performs better; it is approximately 4.1 times faster than the original configuration and 4.5 times faster than the extended configuration, summarizing all results. Executing a level-by-level comparison, performance gains on adopting the search engine vary from 2.3 times (related to the originally RDB instance, querying tags for the patient level) to 6.1 times (querying tags for the study level, in both RDB instances). The search engine surpasses both RDB instances, also, in comparing results for queries searching for individual tags or groups of tags: reductions in time vary from 2.3 times (querying groups of 10 tags, in both RDB instances) to 8.3 times (querying groups of 15 tags, in the originally configured RDB instance).

5. Discussion

When dealing with DICOM indexing, it is paramount to observe the final objectives to be accomplished. An uncontextualized analysis of the chart presented in Fig. 2 (right) indicates a clear advantage in adopting the originally configured RDB instance, due to the reduced time spent on storing content. For scenarios with limited query possibilities, restricted to the available physical structures, this assumption is valid; however, the referred instance does not provide a direct solution for search and retrieval of tags that are

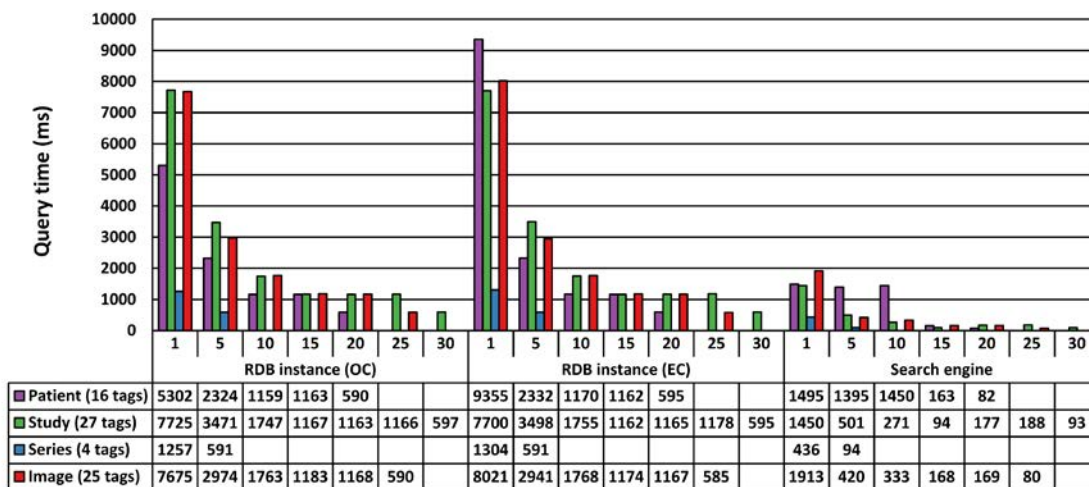


Figure 3. Query time, in milliseconds, used to search and retrieve tags. Queries were executed for each DICOM hierarchical level, retrieving individual tags and groups of tags.

not stored as individual database fields. The economy in storage space and in storage time may be insufficient to justify the lack of query flexibility.

Built over the same underlying database schema than the originally configured RDB instance, the extended configured RDB instance uses the available BLOB fields as a workaround for the physical limitations on storing DICOM tags. It is effective in the sense of guaranteeing full-content storage; however, the approach does not reduce search and retrieval times due to the lack in indexing (as can be seen in Fig. 2 - left, when comparing the size of the index volumes between RDB instances). Assuming that the storage of a tag is performed once, and the search and retrieval of this same tag is performed whenever necessary, performance boosts for queries are at least desirable.

The adoption of schema-free or schema-flexible solutions is an alternative to the limitations of both RDB instances. Using an approach based on JSON documents, the search engine provides flexibility for both storage and query operations. Metadata acquired from radiological images are persisted and indexed in full, requiring a fraction of the space for storage and reducing significantly the time needed for content search. Another characteristic of such solutions, not explored in this work, is its native scalability; in scenarios like federated PACSs, high-demanding by nature, the adoption of extensible infrastructures – malleable enough to be reconfigured and to grow as needed – is differential. Here, the replacement of RDBMSs by less-constrained search engines turns out to be a strong option.

6. Conclusion

This work compares a search engine to relational database instances in the context of DICOM metadata management, observing its behavior in terms of storage space, storage time, and query time. RDB instances are configured to store partial- and full-content metadata extracted from a DICOM dataset; the search engine, in turn, indexes the same full-content metadata as JSON documents. The storage and query operations are executed similarly to DICOM C-STORE and C-FIND.

The acquired results attest the viability in adopting search engines as alternatives to relational databases in support for PACSs. Requiring less storage space and demanding a similar amount of time to store all alphanumeric tags when compared to an RDB instance, the evaluated search engine performed better in searching and retrieving individual tags and groups of tags for all DICOM hierarchical levels. Demanding 69% less storage space to a gain of 8.3 times in query response times, in the best case, the search engine outperforms the RDB instances in the majority of the experiments.

As future work, it is suggested to extend queries aiming to verify the index responses, both in relational database instances and in the search engine, to predicates involving intervals, lists of values, and patterns. It is suggested, also, the evaluation of the search engine deployed in a cluster, measuring the impact of distribution in storing and querying DICOM tags.

Acknowledgment

This work was supported by Santa Catarina's State Health Department (SES/SC), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina).

References

- Baeza-Yates, R. and Ribeiro-Neto, B. (2011). *Modern Information Retrieval - the concepts and technology behind search*. Pearson.
- Bastiao Silva, L. A., Beroud, L., Costa, C., and Oliveira, J. L. (2014). Medical imaging archiving: A comparison between several nosql solutions. In *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, pages 65–68.
- Bidgood Jr, W. D., Horii, S. C., Prior, F. W., and Van Syckle, D. E. (1997). Understanding and using dicom, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association*, 4(3):199–212.
- Clark, K. et al. (2013). The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26:1045–1057.
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.
- Croft, B., Metzler, D., and Strohman, T. (2009). *Search Engines: Information Retrieval in Practice*. Pearson.
- Elmasri, R. and Navathe, S. B. (2010). *Fundamentals of Database Systems*. Pearson.
- Faggioni, L., Neri, E., Castellana, C., Caramella, D., and Bartolozzi, C. (2011). The future of pacs in healthcare enterprises. *European Journal of Radiology*, 78:253–258.
- Huang, H. K. (2011). Short history of pacs. part i: Usa. *European Journal of Radiology*, 78:163–176.
- Lemke, H. U. (2011). Short history of pacs. part ii: Europe. *European Journal of Radiology*, 78:177–183.
- Mansoori, B., Erhard, K. K., and Sunshine, J. L. (2012). Picture archiving and communication system (pacs) - implementation, integration and benefits in an integrated health system. *Academic Radiology*, 19(2):229–235.

- Mildenberger, P., Eichelberg, M., and Martin, E. (2002). Introduction to the dicom standard. *European Radiology*, 12(4):920–927.
- National Electrical Manufacturers Association (2015a). Dicom ps3.4 2015c - service class specifications. Date last accessed 2016-01-31.
- National Electrical Manufacturers Association (2015b). Dicom ps3.5 2015c - data structures and encoding. Date last accessed 2016-02-03.
- Pianykh, O. S. (2012). *Digital Imaging and Communications in Medicine (DICOM) – A Practical Introduction and Survival Guide*. Springer.
- Rascovsky, S. J., Delgado, J. A., Sanz, A., Calvo, V. D., and Castrillón, G. (2012). Informatics in radiology: Use of couchdb for document-based storage of dicom objects. *RadioGraphics*, 32(3):913–927.
- Savaris, A., Härder, T., and Wangenheim, A. v. (2014). Dcm DSM: a dicom decomposed storage model. *Journal of the American Medical Informatics Association*, 21(5):917–924.
- Singh, R., Chubb, L., Pantanowitz, L., and Parwani, A. (2011). Standardization in digital pathology: Supplement 145 of the dicom standards. *Journal of Pathology Informatics*, 2(1):23.
- Sumathi, S. and Esakkirajan, S. (2007). *Fundamentals of Relational Database Management Systems*. Springer.
- The Cancer Imaging Archive Team (2015). Lidc-idri. Date last accessed 2016-02-02.

Vigilância Entomológica da Dengue, Zika e Chikungunya: Uma Solução Baseada em Redes Sociais e Dispositivos Móveis

Soeli T. Fiorini¹, Leonardo Sousa^{1,2}, Diego Cedrim^{1,2}, Alessandro Garcia^{1,2}, Débora Christina Muchaluat Saade³, Igor Monteiro Moraes³, Leonardo Frajhof⁴

¹Laboratório de Engenharia de Software da PUC-Rio

²Pontifícia Universidade Católica do Rio de Janeiro

³Laboratório MídiaCom, Universidade Federal Fluminense – UFF

⁴Núcleo de Telemedicina UNIRIO

Rio de Janeiro, Brasil

soeli@les.inf.puc-rio.br, {afgarcia, leonardo, dcgrego}@inf.puc-rio.br, debora@midia.com.uff.br, igor@ic.uff.br, leonardo.frajhof@gmail.com

Abstract. *Dengue, Chikungunya and Zika are one of the major public health problems in the world (OMS). The use of mobile, Internet and social networks grows day by day. Thus, they offer a simple interaction way that enables a faster society mobilization to combat the vector of those diseases (Aedes Aegypti mosquito). This paper describes a technology solution (Vaza Dengue) developed to allow the society to notify and see in real time cases of Dengue outbreaks. Doing social media data mining, the system shows the situation of the outbreaks and other information correlated to the mosquito.*

Resumo: *A Dengue, Chikungunya e a Zika estão entre os principais problemas de saúde pública no mundo. O uso de dispositivos móveis, Internet e redes sociais cresce a cada dia. Sendo assim, oferecem formas de interação simples que possibilitam uma mobilização mais rápida da sociedade no combate ao vetor dessas doenças: o mosquito Aedes Aegypti. O presente artigo discute as limitações das alternativas existentes, bem como descreve uma solução tecnológica (Vaza Dengue) para notificação e visualização em tempo real de incidências de foco da Dengue em regiões geográficas. Através da mineração de dados das redes sociais, o sistema mostra a situação dos focos e outras informações correlacionadas ao mosquito.*

1. Introdução

Segundo o Boletim Epidemiológico da Secretaria de Vigilância em Saúde – Ministério da Saúde [BOLETIM 2016], foram registrados 170.103 casos prováveis de Dengue no Brasil apenas nas primeiras 5 semanas de 2016. Mais recentemente, tem se observado um número crescente de transmissões da febre de Chikungunya e Zika pelo mosquito. Essa última, inclusive, tem a agravante associação com os casos de microcefalia e outras malformações neonatais e, a Síndrome de Guillain-Barré. Segundo relatório da Organização Mundial da Saúde (OMS) divulgado recentemente [OPAN 2016], atualmente 47 países e territórios registram transmissão de Zika.

A Dengue é uma doença viral aguda e de rápida disseminação, assim como a Zika e Chikungunya. A notificação oportuna dos casos é medida essencial para que a vigilância seja capaz de acompanhar o padrão de transmissão da doença na área e a curva endêmica. O Ministério da Saúde tem iniciativas como o Plano Nacional de Enfrentamento ao Aedes Aegypti e à Microcefalia [PLANO 2016] [BOLETIM 2016]. No entanto, como forma de amplificar a vigilância entomológica (mosquito), uma série de aplicativos em dispositivos móveis estão sendo disponibilizados como um meio para a sociedade notificar e visualizar em tempo real (mapa) os casos de focos da Dengue, Zika e Chikungunya.

O uso de dispositivos móveis, internet e redes sociais cresce a cada dia e oferecem formas de interação simples que possibilitem uma mobilização mais rápida da sociedade no combate ao vetor (mosquito). Com o uso das informações georreferenciadas, a coleta das informações por parte do usuário e a notificação de focos passa a ser simplificada. Além disso, as informações que são registradas naturalmente nas redes sociais, são organizadas através de modelos para que a sociedade possa tomar rapidamente o conhecimento sobre fatos, podendo a isso serem associadas medidas educativas ou de alerta, auxiliando também o estado na tomada de ações preventivas. O setor da saúde, por si só, não tem como resolver a complexidade dos fatores que favorecem a proliferação do vetor. No entanto, se o setor de saúde prover a sociedade com instrumentos adequados isso pode amplificar significativamente a vigilância entomológica da Dengue colocando a sociedade no seu papel de corresponsabilidade do combate desta doença.

Este artigo propõe o Vaza Dengue (Seção 2), uma solução tecnológica (portal e aplicativo) que tem forte ênfase na mineração de dados das redes sociais, mostrando a situação das infestações e outras informações correlacionadas ao mosquito. Para um melhor entendimento, apresenta-se uma análise de trabalhos relacionados (Seção 3) com estudos referentes ao uso das redes sociais e pesquisas de alguns aplicativos existentes e suas características. Para finalizar (Seção 4) apresenta-se as contribuições do artigo com o desenvolvimento do portal e do aplicativo proposto.

2. Vaza Dengue - Uma Solução Tecnológica Baseada em Dispositivos Móveis e Redes Sociais

O Vaza Dengue é um sistema que tem como objetivo principal o fortalecimento da vigilância entomológica do mosquito transmissor da Dengue, Zika e Chikungunya. Esse sistema de vigilância entomológica é articulado em ações sociais a partir da implementação de um portal Web e de um aplicativo para dispositivos com o sistema operacional Android (Google Play). Ele permite que os usuários notifiquem as ocorrências do mosquito da Dengue ou casos de pessoas doentes, bem como visualizar tais notificações. A Figura 1 ilustra algumas funcionalidades do aplicativo Vaza Dengue.

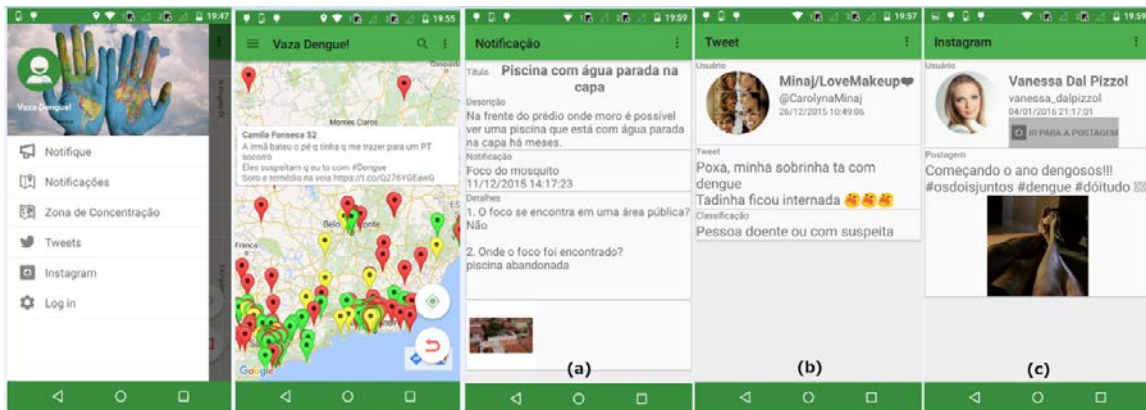


Figura 1 – Telas do Aplicativo Vaza Dengue.

2.1 Visão Geral

O Vaza Dengue é dividido em três grandes sistemas: Servidor da Aplicação, Twitter Crawler e aplicativos Web e Android. A Figura 2 apresenta uma visão geral da arquitetura do Vaza Dengue. Nessa visão, é apresentada (i) a organização dos componentes do sistema de acordo com o dispositivo físico que o mesmo executará e (ii) como os componentes se comunicam entre si.

Android-APP é o serviço implantado em dispositivos móveis Android, enquanto o componente WEB-APP é o portal web que é acessado por meio do navegador web. Esses 2 componentes provêm o papel de clientes e trocam mensagens com o Servidor da Aplicação. O Servidor da Aplicação é implantado em um servidor Tomcat e gerencia e armazena as notificações realizadas pelos usuários. O Data Crawler, por sua vez, recupera das redes sociais Instagram e Twitter os dados importantes sobre a Dengue e os armazena no banco de dados da aplicação.

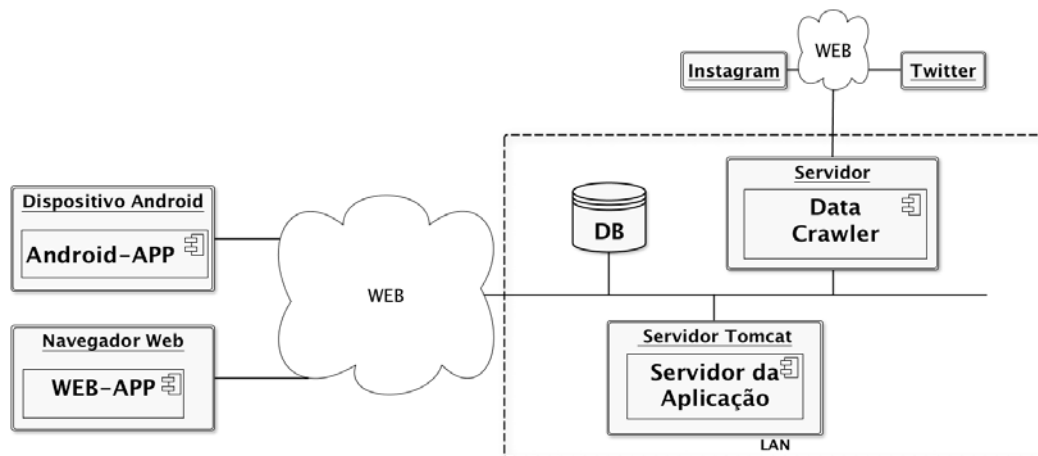


Figura 2 – Visão Geral da Arquitetura do Vaza Dengue.

2.2 Servidor da Aplicação

O Servidor da Aplicação é o responsável por prover serviços para os aplicativos Vaza Dengue. Ele é o componente responsável por responder às requisições HTTP, através do padrão de arquitetura e de transferência de objetos REST. Esse componente é

responsável por implementar as regras de negócio, de autenticação e de criação de dados publicáveis.

O Servidor da Aplicação é responsável por expor uma interface para as regras de negócio associadas ao domínio da aplicação. Tanto o aplicativo Android quanto o portal Web são os clientes que usufruem dos serviços do Servidor da Aplicação. Os clientes se comunicam com o servidor para a recuperação das notificações realizadas pelos usuários. Além do mais, o Servidor da aplicação é responsável por receber as notificações que os usuários realizaram por meio do Vaza Dengue.

2.3 Twitter Crawler

O Twitter Crawler é o componente responsável por obter tweets a partir dos serviços providos pela Twitter Streaming API. Desta forma, cada novo tweet que atenda aos critérios de busca é imediatamente inserido no banco de dados gerenciado pelo Servidor da Aplicação. O Twitter Crawler é responsável também por catalogar os tweets que estão relacionados ao mosquito da Dengue. Os tweets podem ser classificados em quatro categorias (Informativo, Doença, Foco do Mosquito e Piada) de acordo com o conteúdo do tweet. Uma vez que os tweets foram recuperados e classificados, o Vaza Dengue faz uma requisição ao servidor da aplicação, pedindo pelos tweets classificados. Dessa maneira o aplicativo apresenta aos usuários os tweets categorizados de acordo com o seu conteúdo.

A categorização dos tweets acontece por meio de um classificador supervisionado tradicional [KOTSIANTIS 2007]. Esse classificador foi treinado usando um conjunto inicial de tweets relacionados ao mosquito da Dengue. Esse conjunto foi analisado e categorizado manualmente. A fase de classificação manual conta com uma grande carga de trabalho manual e, por isso, é difícil de escalar. Felizmente, essa fase só é realizada uma única vez, durante o treinamento do classificador. Com base no treinamento usando esse conjunto de tweets, o classificador é capaz de categorizar novos tweets dos usuários de acordo com as quatro categorias supracitadas. Após a classificação manual, o classificador ocorre instantaneamente. Ou seja, o classificador atua rapidamente a cada novo tweet selecionado. O processo é altamente escalável, sendo possível instanciar tantos classificadores quantos forem necessários. Também por meio do treinamento, é possível selecionar tweets que são falsos positivos. Para tal, o classificador foi treinado com tweets que não agregavam informação ou eram de cunho puramente cômico. Dessa maneira, o classificador é capaz de capturar e filtrar tais “ruídos”.

2.4 Aplicativo e Portal Web

O aplicativo e portal web correspondem ao serviço com o qual o usuário irá interagir. Ambos serviços permitem que os usuários notifiquem as ocorrências do mosquito da Dengue ou casos de pessoas doentes, bem como visualizar tais notificações. Eles se comunicam com o Servidor de Aplicação, enviando as notificações realizadas pelos usuários, e recebe do Servidor de Aplicação as notificações que já foram cadastradas por outros usuários. Além dessas notificações, eles também recebem do Servidor de Aplicação outros dados armazenados no banco de dados, tais como os tweets coletados usando o Twitter Crawler e as postagens do Instagram, ambos relacionadas ao mosquito da Dengue. O Vaza Dengue pode ser dividido em duas principais funcionalidades:

Cadastrar Notificações

O principal objetivo do Vaza Dengue é oferecer uma interface de comunicação entre os usuários, de tal maneira que os mesmos ajam como agentes etimológicos. Diante disso, o usuário pode usar o aplicativo para notificar ocorrências relacionadas à Dengue, Zika ou Chikungunya. O usuário possui três tipos de notificações:

- **Foco do Mosquito:** esse tipo de notificação permite que o usuário relate um possível foco do mosquito da Dengue. Junto a essa notificação o usuário pode informar onde o foco se encontra e se esse lugar se encontra em uma área pública ou privada.
- **Pessoa Doente:** esse tipo de notificação permite que o usuário (nesse caso o cidadão usando o aplicativo ou portal web) informe o caso de uma pessoa que esteja doente devido à interação com o mosquito da Dengue. Ele pode escolher entre três tipos de enfermidades: Dengue, Zika ou Chikungunya. Junto a essa notificação o usuário pode informar a idade do paciente e se os agentes de saúde visitam a região onde o paciente mora.
- **Suspeita de Doença:** esse tipo de notificação permite que o usuário informe o caso de uma pessoa que está apenas com suspeita de Dengue, Zika ou Chikungunya. Junto a essa notificação o usuário pode informar quais os sintomas que a pessoa está sentindo.

Junto com os dados informados pelo o usuário, o aplicativo e portal web também adicionam às notificações informações sobre a localização do usuário, a data em que a notificação está sendo cadastrada e um foto, caso o usuário deseje.

Visualizar Notificações

Além de permitir que os usuários capturem notificações, o aplicativo e portal web também permitem que os usuários busquem as notificações mais recentes que foram cadastradas por outros usuários. Para tal, eles fazem uma requisição para o Servidor da Aplicação, que por sua vez envia as 500 notificações mais recentes. O aplicativo recebe essas notificações e as plota no mapa de acordo com as suas coordenadas.

Uma vez que as notificações foram plotadas no mapa, o usuário pode clicar em uma notificação (representada por um marcador) e acessar os dados referentes à notificação. Cada notificação é representada por um marcador colorido. A cor do marcador varia de acordo com o tipo de notificação definido quando o usuário cadastrou a notificação. Os marcadores amarelos representam notificações do tipo “Foco do Mosquito”, os marcadores verdes são notificações do tipo “Informativo” e os marcadores vermelhos representam as notificações relacionadas às doenças: “Pessoa Doente” e “Suspeita de Doença”.

As notificações recuperadas do Twitter seguem a mesma categorização de cores para os marcadores (amarelo = Foco do Mosquito, verde = Informativo e vermelho = Pessoa Doente ou Suspeita de Doença). Entretanto, as notificações que representam piadas não são exibidas. As notificações recuperadas do Instagram são plotadas com marcadores azuis, uma vez que as postagens não estão categorizadas.

A seguir descreve-se trabalhos relacionados, apresentando primeiramente o uso das redes sociais, e em seguida, analisando-se alguns aplicativos e portais relacionados, focando em seus pontos positivos e oportunidades de melhoria. Para complementar, realizou-se uma pesquisa para caracterizar os aplicativos relacionados, comparando os resultados com a solução proposta.

3. Trabalhos Relacionados

Uma estratégia promissora para ajudar a engajar a população no combate às doenças causadas pelo mosquito *Aedes Aegypti* é o uso das redes sociais. Este artigo faz contribuições originais justamente nesse campo, no qual se desenvolve pesquisas sobre monitoramento de redes sociais para vigilância em saúde. Redes sociais, tais como o Twitter, foram e são fontes que trazem informações valiosas consigo. Para o usuário dessas redes, é natural enviar uma mensagem no Twitter que esteja relacionada à Dengue, por exemplo. Tais mensagens, chamadas de *tweets*, contém informações que variam desde denúncias de possíveis criadouros de mosquito, até os próprios usuários relatando estarem doentes. Além das redes sociais serem um canal natural para a população, existe estudos que indicam o benefício da análise de redes sociais em diferentes domínios. Por exemplo, o próprio Twitter já foi utilizado para prever crimes. No artigo de M. S. Gerber [GERBER 2014], o autor monitorou os tweets vindo da cidade de Chicago, Illinois. Ao incorporar os tweets, o autor mostrou que para 19 dos 25 crimes estudados, os tweets aumentaram a taxa de predição de crimes.

Da mesma forma que o nosso trabalho, os dados do Twitter foram usados por [GOMIDE et al 2011] para monitorar o fluxo de tweets relacionados a Dengue e filtrar sinais relevantes a partir deles. Uma vez que os autores apenas usaram um classificador supervisionado para filtragem de conteúdo, a abordagem deles ficou limitada à quantidade de tópicos pré-definidos. Um classificador supervisionado é treinado usando um conjunto de exemplos que são analisados e categorizados manualmente, chamado de conjunto de treino. Nosso trabalho, em contra partida usa um **classificador não-supervisionado**, ou seja, sem a categorização manual de exemplos. O classificador não-supervisionado se baseia em uma abordagem de modelo de tópicos (do inglês *Topic Model*) e no algoritmo LDA [BLEI et al 2003], um algoritmo que tem apresentado bons resultados para classificar *tweets* em tópicos, isto é, em subconjuntos [RAMAGE et al 2010]. Uma abordagem de modelo de tópicos se baseia em modelos estatísticos para encontrar tópicos que ocorrem com frequência em uma coleção de dados. A desvantagem do classificador usado por [GOMIDE et al 2011] diz respeito à limitação de prever novas informações no início de cada onda epidemia de Dengue. Em nosso trabalho, nós não usamos apenas um classificador supervisionado, mas também usamos um classificador não-supervisionado como meio de identificar informações relevantes tais como prever inícios de surtos epidêmicos.

Em [LAMPOS E CRISTIANINI 2010] e [ACHREKAR et al 2010], os autores mostraram que a frequência de *tweets* que contém palavras-chave simples podem ser um bom indicador de uma epidemia de gripe. Os autores mostraram que há uma forte correlação entre o número de visitas ao médico (suspeita de gripe) e o número de tweets que citam a gripe. Esta abordagem para rastrear epidemias é complementar à nossa. Enquanto os autores mencionados anteriormente mediram a atividade dos *tweets* em um conjunto inteiro de *tweets*, nós usamos a aprendizagem de máquina para automaticamente descobrir ainda mais sub-sinais de epidemias específicas. Nossa abordagem permite explorar e estudar com mais detalhes os *tweets* relevantes, podendo categorizá-los de acordo com seu conteúdo.

Estratégias similares de monitoramento de dados do Twitter têm sido aplicadas para detecção de eventos em geral, como foi feito, por exemplo, por [BECKER et al 2011]. No entanto, a obtenção de um oráculo é reconhecida como um sério gargalo no

classificador supervisionado e os esforços para reduzir o esforço manual para gerar o conjunto de treino ainda é tema de pesquisa em andamento. Por exemplo, [GO et al 2009] identificaram automaticamente oráculos para a classificação sentimento (descobrir se a mensagem tem um conteúdo positivo ou negativo) a partir de *emojicons*. No entanto, como mencionado anteriormente, mesmo se oráculos foram de alguma forma identificados, o uso de classificadores supervisionados pode não ser suficiente para lidar com cenários onde o conteúdo pode mudar com o tempo ou em cenários onde diferentes epidemias surgem.

Analisando-se alguns aplicativos e portais verifica-se o pouco uso das redes sociais. As Tabelas 1 e 2 que seguem descrevem essa análise, relatando os aspectos positivos e oportunidades de melhoria de cada aplicativo e portal mencionado. Os aplicativos listados na Tabela 1 podem ser instalados diretamente da loja oficial de cada dispositivo móvel. No caso de dispositivos Android, os aplicativos estão disponíveis no Google Play. No caso de dispositivos iOS, os aplicativos estão disponíveis no Apple Store.

Tabela 1 – Aspectos positivos e oportunidade de melhoria dos aplicativos analisados.

Aplicativo	Aplicativo/ Aspectos positivos	Oportunidade de melhoria
ContraDengue	Permite que os usuários enviem dados de foco de Dengue, através do Twitter ou por e-mail através do aplicativo.	Não apresenta os dados coletados para os usuários e não usa o GPS para identificar o endereço (cidade, estado, rua) do foco.
Dengue Report	Facilita o monitoramento para os agentes de saúde. Criado com o intuito de substituir as planilhas impressas utilizadas de forma tradicional pelas secretarias de saúde dos municípios	Não é um sistema de coleta colaborativo. O foco dele é gerencial.
Dengue Activity Tracking System	Permite que os usuários enviem dados de foco de Dengue. Usa GPS para apresentar os dados coletados em um mapa e em um portal. Permite envio de imagens.	Interface disponível apenas em Árabe. Por não ser focado apenas em Dengue, possui interface mais complexa que o desejável.
Dengue Radar	Apresenta casos reais de Dengue que aconteceram próximo ao usuário reportados por órgãos da saúde. Cada caso é apresentado como um marcador em um mapa.	Não permite envio de nenhuma informação por parte dos usuários. Aplicativo usado apenas para visualização.
UNA-SUS Dengue	Apresenta ações que devem ser feitas em caso de suspeita de Dengue e outros problemas. Usa informações sobre o perfil do paciente para automaticamente propor a conduta a ser adotada de acordo com o perfil de cada um.	Aplicativo apenas informativo. Não há envio de dados por parte dos usuários. Informações sempre estáticas.
Dengue Brasil	Apresenta informações e notícias relacionadas ao mosquito e a doença da Dengue.	Aplicativo apenas informativo. Não há envio de dados por parte dos usuários.
Radar Dengue	Permite que os usuários enviem dados de foco de Dengue. Apresenta os focos em um mapa.	Não apresenta dados estatísticos sobre a Dengue para os usuários. Não permite enviar imagem do foco.

Tabela 2 – Aspectos positivos e oportunidade de melhoria dos portais analisados.

Portais	Aspectos positivos	Oportunidade de melhoria
Observatório Dengue	Agrega vários dados da internet no portal (e.g. vídeos do Youtube, matérias de sites e blogs, citações de redes sociais); apresenta duas visões dos dados no mapa: (i) uma visão global com todos os dados, (ii) uma visão reduzida com dados em tempo real; apresenta conteúdo informativo sobre formas de prevenção, e tratamento da Dengue e apresenta os dados (vídeos e links de notícias) de acordo com sua relevância.	Não apresenta dados estatísticos sobre as citações; não apresenta a possibilidade de filtrar os dados por critérios definidos (e.g. por cidade, por período de tempo); utiliza a rede Facebook de forma limitada (apenas comentários); não realiza cruzamento dos dados do sistema com dados dos sistemas de vigilância tradicionais e não possui nenhuma aplicação para dispositivo móvel associada.
DengueApp	Possui uma aplicação de dispositivo móvel associada ao portal; apresenta dados estatísticos das citações por bairro e cidade; existe uma política de uso do sistema; para uso do DengueApp deverá ser feito um cadastro prévio no sistema e o usuário pode fazer denúncias e alterar o status da denúncia (e.g. “resolvido”, “não-resolvido”).	Não apresenta a possibilidade de filtrar os dados por critérios definidos (e.g. por cidade, por período de tempo); não utiliza as redes sociais e não realiza cruzamento dos dados do sistema com dados dos sistemas de vigilância tradicionais.
Dengue Trends	Agrega dados de várias fontes distintas; correlaciona os dados do portal com dados dos sistemas de vigilância tradicionais; agrega dados de vários anos (desde 2008 até os dias atuais); utiliza tags e gradação em cores nos mapas para indicar a tendência de ocorrência de Dengue e possui apresentação de dados estatísticos através de planilhas e gráficos e possibilidade de filtrar os dados (por período, por região etc).	Não possui aplicação de dispositivos móveis associada ao portal e não utiliza dados de redes sociais.

Para corroborar com a questão do pouco uso das redes sociais no desenvolvimento de aplicativos, este artigo fez a análise da característica de 49 aplicativos móveis. A pesquisa foi realizada em março de 2016, utilizando-se a palavra “Dengue” na busca e considerando apenas os aplicativos que tinham sua descrição em inglês ou português. A pesquisa também analisou o uso do GPS, mapas e imagens, pois são questões que promovem facilidades no uso dos aplicativos e na visualização das informações e foram pontos incluídos nas funcionalidades do aplicativo e portal do Vaza Dengue.

Os aplicativos, na sua maioria (65%), são para realizar denúncias de foco do mosquito ou para relatar doenças e sintomas relacionados. O restante dos aplicativos se dividem em jogos (23%), formas para repelir mosquitos através de som emitido pelo celular (8%) e uma pequena parcela (4%) é destinada a profissionais da saúde descrevendo protocolos para enfrentamento das doenças transmitidas pelo mosquito *Aedes Aegypti*. Os gráficos que seguem dizem respeito aos aplicativos relacionados à denúncia e informativos.

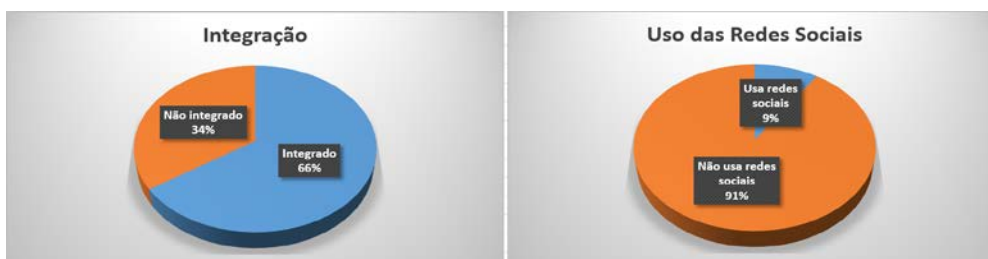


Figura 3 – Integração dos aplicativos e uso das redes sociais.

A Figura 3 mostra que mais da metade dos aplicativos possuem uma integração (66%) com os órgãos públicos como secretaria da saúde dos estados e prefeituras enviando as denúncias e notificações. No entanto, apenas 9% desses aplicativos possuem alguma integração com as redes sociais como twitter e blogs (App *Contra Dengue*) e link para facebook (Apps *Komunti Bebas Denggi* e *Zika Out*). Nenhum relatou ter um portal.



Figura 4 – Uso de GPS, mapas e imagens.

Sob a ótica do uso do GPS, mapas e captura de imagens (Figura 4) para notificar focos do mosquito verificou-se que pouco mais da metade dos aplicativos (53%) faz uso desses recursos. Muitos o fazem ainda de forma limitada, no que diz respeito ao GPS, sendo utilizado apenas para plotar dados no mapa e poucos para pegar de forma automática o endereço para facilitar a notificação.

O aplicativo *Vaza Dengue* (Seção 2) tem como grande diferencial em relação aos aplicativos pesquisados, a questão de utilizar e explorar de fato as redes sociais (existem trabalhos com redes sociais mas não estão integrados aos aplicativos). Além disso, possui associado ao aplicativo um portal, utiliza georreferenciamento identificando o endereço do usuário, mapas e realiza denúncias com a captura de imagens.

4. Considerações Finais

A solução *Vaza Dengue* apresentada neste artigo promove benefícios, conhecimentos e inovações aos temas de Dengue, Zika e Chikungunya, gestão da saúde e exposições ambientais e impactos na saúde humana. Apresenta um portal e aplicativo de vigilância epidemiológica com mineração de dados em redes sociais, com possibilidades para a definição de indicadores da vigilância entomológica e seus relacionamentos. Do ponto de vista econômico e gestão, pode reduzir custos através da tomada de ações preventivas e melhora na gestão através da organização da rede assistencial ou campanhas publicitárias e educativas. Na questão ambiental tem reconhecimento geográfico automático dos pontos estratégicos (focos) para o planejamento das atividades de controle vetorial e, contribui com informações para a organização das operações de campo. Do lado social, tem instrumentalização das entidades da sociedade, atuando no enfrentamento das doenças. Além desses resultados, existe a possibilidade de incorporação do portal e aplicativo na gestão e/ou atenção em saúde pública. O governo lançou o aplicativo “0800 Saúde” que não possui as funções do *Vaza Dengue*.

Como trabalhos futuros, a solução proposta pode incorporar novas funcionalidades, como por exemplo, informações educativas sobre as doenças, protocolos de atendimento, integrações com software de ERP para prefeituras ou integrações com sistemas e aplicativos de unidades de saúde do governo em suas diferentes esferas. Ainda como trabalho futuro, espera-se usar explorar os dados fornecidos pelo usuário por meio do aplicativo e portal web. Uma vez que tanto o aplicativo quanto o portal web esteja sendo usado pela população é possível, mediante amostra suficiente inserida pela população, uma análise dos resultados. Exemplos dessa análise incluem apresentar os dados coletados e compará-los com dados consolidados pelo Ministério da Saúde. Embora o vetor de Dengue, Zika e Chikungunya seja o mesmo, ainda pretendemos como trabalho futuro diferenciar esses três tipos de notificação. Assim cada usuário pode escolher com maior precisão que informações ele está interessado, além do mais, essa subclassificação permitirá que os dados coletados para os três tipos de notificação possam ser cruzados estabelecendo correlação entre os dados para cada vetor.

5. Referências

- H. Achrekar, R. Lazarus, and W. C. Park, “Predicting Flu Trends using Twitter Data.” 2010.
- H. Becker, M. Naaman, and L. Gravano, “Beyond Trending Topics: Real-World Event Identification on Twitter,” *ICWSM*, pp. 1–17, 2011.
- Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3:993–1022.
- Boletim Epidemiológico Secretaria de Vigilância em Saúde – Ministério da Saúde – Brasil, consultado em 08/03/2016, <http://portalsaude.saude.gov.br/images/pdf/2016/fevereiro/29/2016-006---Dengue-SE5-publica---o.pdf>
- M. S. Gerber, “Predicting crime using Twitter and kernel density estimation,” *Decis. Support Syst.*, vol. 61, pp. 115–125, 2014.
- A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N Proj. Report*, Stanford, vol. 1, no. 12, 2009.
- J. Gomide, A. Veloso, W. Meira, V. Almeida, F. Benevenuto, F. Ferraz, and M. Teixeira, “Dengue surveillance based on a computational model of spatio-temporal locality of Twitter,” *Proc. ACM WebSci’11*, June 14–17 2011, Koblenz, Ger., pp. 1–8, 2011.
- S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques”. *Proc. Emerging Artificial Intelligence Applications in Computer Engineering*, pp 3–24, 2007
- V. Lampos and N. Cristianini, “Tracking the flu pandemic by monitoring the social web,” 2010 2nd Int. Work. Cogn. Inf. Process. *CIP2010*, 2010.
- Organização Pan-Americana de Saúde, consultado em 08/03/2016 http://www.paho.org/bra/index.php?option=com_content&view=article&id=5015:relatorio-de-situacao-sobre-Zika-da-oms-aponta-transmissao-local-do-virus-em-47-paises&Itemid=816
- Plano Nacional de Enfrentamento ao Aedes e à Microcefalia – Ministério da Saúde – Brasil, consultado em 08/03/2016. <http://combateaedes.saude.gov.br/plano-nacional>
- Ramage, D.; Dumais, S. T.; and Liebling, D. J. 2010. Characterizing Microblogs with Topic Models. *ICWSM* 10:1.

Implementação da IoT para o Monitoramento das Variáveis Meteorológicas num AAL

Madalena Pereira da Silva¹, Débora Cabral Nazário¹, M.A.R. Dantas^{1,2}, Alexandre Leopoldo Gonçalves¹, A.R. Pinto³, Guilherme Manerichi³, Bruno Vanelli³

¹Engineering and Knowledge Management Department, ²Computer Science and Statistics Department, ³Control and Automation Engineering Department

Federal University of Santa Catarina, ^{1,2}Florianópolis, ³Blumenau, Brazil

madalena.silva@posgrad.ufsc.br, debora.nazario@udesc.br,
{m.a.r.dantas, a.l.goncalves, a.r.pinto}@ufsc.br,
{guilherme.manerichi, brunno.v}@grad.ufsc.br

Abstract. *Most research on Human Biometeorology use information from health and weather data providers to correlate and infer the impacts of weather variables on some morbidities. Although research generates indicators for the health management, it needs to be refined and consider weather variables in the indoor environment and provide these data for evaluation in conjunction with the patient's clinical data. In order to contribute to human biometeorology, this paper proposes implementing the IoT (Internet of Things) using the ZigBee to monitor the environmental conditions of a ubiquitous AAL (Ambient Assisted Living) setting, and to store data in the cloud in order to be consumed by caregiver health.*

Resumo. *A maioria das pesquisas sobre a Biometeorologia Humana usa dados de provedores de dados da saúde e do clima, para correlacionar e inferir os impactos das variáveis climáticas sobre as morbidades. Embora possam gerar indicadores para a gestão da saúde, as pesquisas precisam ser refinadas no sentido de considerar as leituras das variáveis meteorológicas no ambiente interno e, disponibilizar esses dados a serem avaliados juntamente com os dados clínicos dos pacientes. Visando contribuir com a Biometeorologia Humana, neste artigo é proposta a implementação da Internet das Coisas (IoT), com o uso da ZigBee, para monitorar as condições ambientais de um AAL (Ambient Assisted Living) ubíquo, e armazenar os dados na nuvem, para serem consumidos por cuidadores da saúde.*

1. Introdução

A Biometeorologia Humana é a ciência que estuda o impacto da influência atmosférica sobre a saúde e bem estar do homem. São muitas as propostas existentes na literatura que correlacionam as condições climáticas com a saúde, incluindo o conforto térmico [Thom 1959], as morbidades [Quinn et. al., 2014], a mortalidade [Quinn et al., 2014, Zhang et al., 2014], a saúde fetal [Ngo e Horton, 2016], entre tantas outras. A grande maioria usa dados abertos de provedores de dados da saúde e do clima, para correlacionar e fazer inferências sobre os impactos das variáveis climáticas sobre algumas morbidades de um país, região ou grupo de pessoas [Critchfield 1983 e Azevedo et al., 2015].

Embora as pesquisas citadas possam gerar indicadores para a gestão da saúde, as mesmas precisam ser refinadas, uma vez que existe diferença na precisão dos valores das variáveis meteorológicas lidas no ambiente externo versus interno [Quinn et al., 2014]. Outro aspecto a considerar é a provisão desses dados, junto ao PEP (Prontuário Eletrônico do Paciente) para auxiliar nas análises médicas.

Uma das morbidades que tem sido fortemente associada com alterações climáticas são as Infecções Respiratórias Agudas (IRA) [Gonçalves e Colho 2010, Araújo et al. 2010, Azevedo et al. 2015]. Como as pessoas vivem a maior parte do tempo em ambientes internos e muitas delas têm propensão a desencadear algum tipo de doença respiratória, torna-se extremamente recomendado a monitoração das condições meteorológicas no ambiente interno. Os valores das variáveis meteorológicas, associados ao contexto do utente (i.e. estado clínico, histórico do paciente/familiar) tornam-se indicadores valiosos para a tomada de decisões dos cuidadores da saúde.

Usando a tecnologia a serviço da Biometeorologia Humana, neste artigo é proposta a implementação da Internet das Coisas (IoT), com o uso das Redes de Sensores Sem Fio (RSSF) ZigBee, para monitorar as condições ambientais (temperatura do bulbo seco, temperatura do ponto de orvalho, umidade relativa do ar) de um AAL (*Ambient Assisted Living*) ubíquo, comunicando objetos físicos (sensores, dispositivos e máquinas) e armazenando dados provenientes destes na nuvem.

Para validar a proposta, foi idealizado um cenário de um paciente idoso, com pré disposição a desencadear crises asmáticas. O mesmo é assistido por cuidadores formais e informais da saúde. Nos experimentos, configuramos uma infraestrutura física para gerar, armazenar e consumir os dados persistidos na nuvem, com diferentes visões do sistema. Uma análise sobre os dados gerados foi realizada e o sistema demonstrou um percentual de 94% de confiabilidade.

O artigo está organizado em seções. Na seção 2 são apresentadas as pesquisas sobre o uso das RSSF e IoT em aplicações de Biometeorologia Humana. A seção 3 apresenta a proposta e a implementação do sistema. A seção 4 apresenta os experimentos e os resultados. Finalizando, a seção 5 apresenta as considerações finais.

2. Revisão da Literatura

A seção investiga o uso da RSSF aplicada na área da saúde, sobretudo a ZigBee com IoT. Também são apresentados alguns trabalhos que correlacionam fatores meteorológicos do ambiente interno a saúde das pessoas.

2.1. Uso das Redes de Sensores Sem Fio na Biometeorologia Humana

A Figura 1 sumariza a quantidade de publicações que usam RSSF na área da saúde. Dos 594 documentos publicados (2005 a 15 de Julho de 2015), a grande maioria utilizou a ZigBee para o monitoramento de sinais biomédicos, usando redes de sensores do corpo; monitoramento de atividades de idosos em AAL, entre outras. Destas, apenas 6 utilizaram a ZigBee para monitoramento de variáveis meteorológicas do ambiente interno [Yang et al. 2014a, Sun et al. 2013, Nam et al. 2011, Bader et al. 2009, Jayakumar et al. 2015, Sung et al. 2012]. Essa evidência motivou a escolha de usar e avaliar a confiabilidade da ZigBee num cenário de Biometeorologia Humana (seção 4).

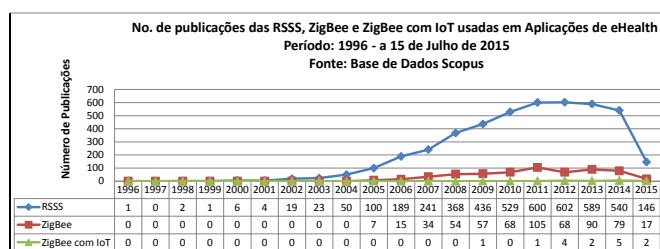


Figura 1. Número de publicações RSSF, RSSF e ZigBee, ZigBee em IoT.

As 15 publicações que utilizaram a IoT com ZigBee, demonstram que a IoT e ZigBee é o conjunto de tecnologia mais adequado a ser usado no monitoramento de sinais biomédicos. O

trabalho descrito em [Yang et al. 2014b] utiliza a IoT e ZigBee para estruturar um ambiente de monitoramento inteligente das condições físicas e qualidade do ar de um hospital.

2.2. Correlações da Temperatura e Umidade Relativa do Ar com Crises Asmáticas

Segundo os autores [Teixeira et al. 2012], desde os primórdios a umidade relativa (UR) do ar (alta ou baixa) tem sido relacionada a problemas de saúde, devendo ficar entre 40% a 70%. Em seus estudos, os autores [Collins 1986] evidenciaram que tanto a UR baixa quanto a alta podem contribuir para o aparecimento de doenças respiratórias.

Segundo a OMS (Organização Mundial da Saúde), a asma é uma das doenças respiratórias que tem sido apontada como problema mundial de saúde pública. De acordo com GIMA os fatores que podem desencadear ou agravar os sintomas da doença incluem infecções virais, alérgicos domésticos ou profissionais (por exemplo, ácaros, pólenes, barata), o fumo do tabaco, exercício e stress. Além destes, os parâmetros meteorológicos (i.e. umidade relativa do ar, precipitação, temperatura do ponto de orvalho, horas de sol e temperatura de bulbo seco) têm sido estudados como indicadores da frequência de asma brônquica [Fleischer 1978].

Outros estudos correlacionam às crises asmáticas com as variáveis, temperatura e umidade [Ayres Sampaio et al. 2014, Teixeira e Postolache 2014, Ezequiel et al. 2007, Santic' et al. 2002, Zaninovic e Raos 2001, Shimizu et al. 2001, Strannegard and Strannegard 1990, Ito et al. 1989, Fleischer 1978, Paulus e Smith 1967].

2.3. Correlação da Temperatura do Ponto de Orvalho, como Indicador de Desconforto Térmico Humano

O ponto de orvalho pode ser usado como indicador de desconforto térmico, fazendo-se uma relação com a temperatura, pressão e a umidade. O ponto de orvalho refere-se à temperatura na qual o vapor de água presente no ar ambiente passa ao estado líquido na forma de pequenas gotas por via da condensação. A Tabela 1 [Horstmeyer 2006] apresenta os níveis de sensação no ser humano, em função do cálculo da temperatura de orvalho, levando em consideração a UR a uma temperatura ambiente de 32° C.

Tabela 1. Sensação Térmica em função da Temperatura do Ponto de Orvalho

Ponto de Orvalho em °C	Sensação Térmica no ser Humano	UR a 32°C	Sensação Térmica dos 32°C com UR
> 29 °C	Extremamente sufocante. Casos de morte por hipertermia	> = 85%	54°C-47°C
26-28 °C	Altíssima. Mortal para doenças relacionadas à asma	84-73%	46°C-42°C
24-26 °C	Extremamente desconfortável, consideravelmente sufocante	72-62%	41°C-38°C
21-24 °C	Muito úmido, desconfortável	61-52%	37°C-35°C
18-21 °C	Desconfortável para a maioria das pessoas no limite superior	51-44%	35°C-33°C
16-18 °C	Aceitável para a maioria, porém todos percebem a umidade no limite superior	43-37%	32°C
13-16 °C	Confortável	36-31%	32°C-31°C
10-12 °C	Muito confortável	30-26%	31°C-30°C
< 10 °C	Um pouco seco para algumas pessoas	<=25%	30°C-27°C

A Tabela 2 apresenta as equações necessárias para obter o índice de desconforto térmico e demais parâmetros usados no sistema proposto (seção 3).

Tabela 2. Equações para Obtenção dos Valores das Variáveis Meteorológicas

Cálculo	Equações	Variáveis	Autores
Índice de Desconforto Térmico (ID)	$ID = 0.99 \times Ta + 0.36 \times To + 41.5$ (1)	ID = índice de desconforto Ta = temperatura em graus °C To = temperatura de orvalho em °C	[Horstmeyer 2006], adaptada de [Thom 1959]
Temperatura do Ponto de Orvalho (To) em °C	$To = \frac{B_1 \times \left[\ln\left(\frac{Ur}{100}\right) + \frac{A_1 \times Ta}{B_1 + Ta} \right]}{A_1 - \ln\left(\frac{Ur}{100}\right) - \frac{A_1 \times Ta}{B_1 + Ta}}$ (2)	A1 = 17,625 B1 = 243,04 Ln = log2,718 Ur = umidade relativa	[Lawrence 2005]

Umidade Relativa (UR) em %	$Ur = 100 \times \left(\frac{e}{e_s}\right) \quad (3)$	e = pressão atual de vapor e_s = pressão saturada de vapor	[Lawrence 2005]
Pressão Saturada do Vapor (es)	$e_s = E \times 10^{\left(\frac{7,5 \times Ta}{237,3 + Ta}\right)} \quad (4)$	E = parâmetro equivalente a 0,6108 para resultados em kilopascal (kPa) e Ta em graus °C.	[Lier et al. 2008].

3. Implementação da IoT para Monitoramento das Condições Ambientais do AAL

A Figura 2 apresenta uma visão geral da proposta. O sistema funciona da seguinte forma. Os dados lidos dos sensores biomédicos são enviados via conexão USB para o sistema de gerenciamento de contexto. Os dados capturados dos sensores de ambiente são enviados ao Gateway IoT e deste, são encaminhados para o sistema de gerenciamento de contexto, via comunicação wireless (802.11). No sistema de gerenciamento de contexto, ocorre a gestão da qualidade do contexto, ou seja, é realizado o filtro dos valores lidos e somente os valores válidos, acurados, precisos e atuais são enviados para a nuvem [Nazário et al., 2014]. Uma vez filtrados, os dados são enviados para o gateway doméstico usando conexão ethernet (802.3). O transporte dos dados para a nuvem é feito através de túneis criptografados, usando VPN (*Virtual Private Network*) e somente pessoas autorizadas (cuidadores formais, informais, grupos familiares e serviços de manutenção), com diferentes visões do sistema, podem acessar os dados.

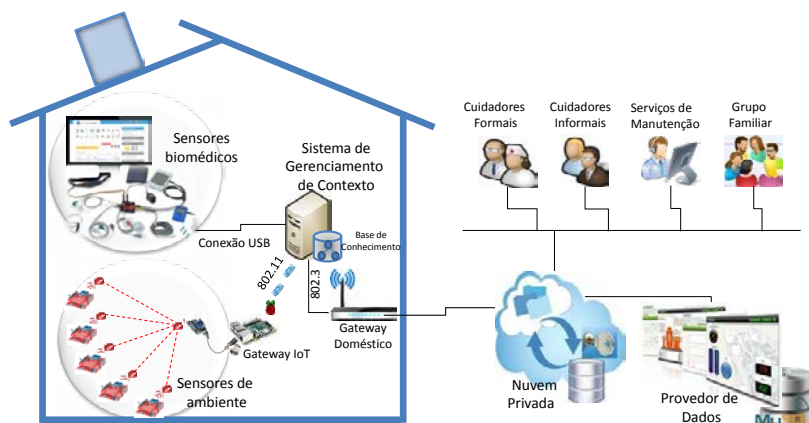


Figura 2. Visão Geral do Sistema Proposto

3.1. Configuração dos Sensores Biomédicos e Distribuição do Contexto

Inicialmente foi proposta uma arquitetura para o gerenciamento da QoC (*Quality of Context*) dos dados dos sensores biomédicos [Nazário et al., 2014] usando a plataforma *eHealth* [Cooking-hacks s.d.], num AAL ubíquo. Para a distribuição do contexto do AAL ubíquo (sinais biomédicos e dados do utente) para a nuvem, uma infraestrutura de transporte usando SDN (*Software Defined Network*) foi proposta e validada [Da Silva et al., 2015]. Portanto, já realizamos os experimentos e validações da captura, processamento e distribuição dos sinais biomédicos do AAL ubíquo para a nuvem. Agora, iremos nos concentrar em agregar a ZigBee com IoT para a monitoração das variáveis meteorológicas, armazenando-as numa base de dados na nuvem para serem consumidas por cuidadores da saúde.

3.2. Configuração dos Sensores de Ambiente

A rede ZigBee foi configurada com topologia em estrela, composta por 5 nodos escravos e 1 nodo controlador. Na topologia foram usados 5 módulos XBee Antenna 1Mw Serie 1. Cada módulo é conectado na Xbee shield, que por sua vez é encaixada na placa Arduino Uno. Em cada placa Arduino de um nodo escravo está plugado o sensor de temperatura e umidade, DHT11.

O elemento sensor do DHT11 é um termistor do tipo NTC, capaz de fazer leituras de temperaturas entre 0 a 50°C com $\pm 2^\circ\text{C}$ de precisão. O elemento sensor de umidade, do tipo HR20, é capaz de ler a umidade entre 20 a 90% com $\pm 5\%$ de precisão.

O DHT11 possui um circuito interno programado para fazer a leitura dos sensores e enviar ao microcontrolador Arduino, via comunicação serial. O Arduino envia um pulso negativo de 18ms para o sensor DHT11, seguido de um pulso positivo de 20 μS a 40 μS . Logo após, o DHT11 envia uma resposta para o Arduino, seguido dos bits de dados provenientes das medições. O sensor DHT11 envia 40 bits (5 bytes) mais os bits de inicialização.

3.3. Implementação da IoT para a Comunicação entre os Dispositivos da RSSF ZigBee e Raspberry PI

Para obter as leituras das medições, via troca de mensagens, entre o coordenador da rede e os nodos ZigBee foram programados dois códigos na linguagem C, um para os nodos escravos e outro para o controlador. Para o nosso experimento foram usadas as bibliotecas XBee e DHT. A biblioteca XBee é responsável pela comunicação entre os nodos da rede ZigBee. A biblioteca DHT possui funções prontas para leitura e escrita dos dados da temperatura em graus Celsius e umidade em percentual.

A. Leitura da Temperatura e Umidade pelos Nodos ZigBee

Ao integrar a rede, cada nodo escravo recebe uma identificação e fica no aguardo das solicitações originadas pelo nodo controlador. Ao receber uma mensagem do controlador cujo ID seja o seu identificador, o nodo escravo coleta e envia a leitura dos dados do sensor ao controlador.

B. Leitura de Temperatura e Umidade pelo Controlador ZigBee

O controlador inicia o processo de comunicação, enviando requisições em sequência, para cada nodo da rede. As mensagens geradas pelo controlador contém o número de identificação do nodo e são enviadas a cada intervalo de tempo de 1500ms. As mensagens são enviadas sempre a cada 1500ms, independente do tempo da resposta recebida. Ao receber a resposta e realizar o cálculo do ponto do orvalho, o controlador coloca a mensagem no formato de endereço e envia para a porta de comunicação serial do Arduino, que está conectada diretamente na porta serial do Raspberry PI.

A biblioteca DHT não possui um método para cálculo do ponto de orvalho, no entanto, a partir dos valores lidos (temperatura e umidade) é possível criar um método para estimar este valor. Nós programamos o método “dewPoint”, aplicando a Equação 2. As Equações 3 e 4, não foram necessárias, pois o método de escrita da umidade, nativo da biblioteca DHT, já fornece a umidade formatada em percentual.

3) Implementação do Gateway IoT

O Raspberry PI atua como gateway IoT. Ele recebe os dados de sensoriamento do coordenador ZigBee via comunicação serial. Os dados recebidos são processados no Raspberry PI e enviados para o sistema de gerenciamento do contexto. Após a etapa do filtro [Nazário et al., 2014], os dados são enviados para a nuvem usando a VPN. A transmissão dos dados do Raspberry PI para o sistema de gerenciamento de contexto ocorre via interface wireless, com o roteador wireless.

3.4. Persistência dos Dados

Para implementar a persistência de dados foi instalado o sistema supervisor ScadaBR num servidor Web, na nuvem. Os dados são enviados via protocolo HTTP e persistidos na base de dados MySQL do sistema supervisor.

A. Data Source e Data Points do AAL

Para enviar os dados ao ScadaBR, foi criada uma fonte de dados do tipo HTTP. A fonte de dados recebe um nome, está associada a um endereço IP. No nosso experimento a fonte de dados representa um AAL e pode ter múltiplos usuários autorizados no seu gerenciamento. Para cada nodo da rede ZigBee foram criados três pontos de dados para armazenar os valores das variáveis temperatura ambiente, umidade e temperatura do ponto de orvalho. Cada ponto de dado está associado a um cômodo do AAL.

B. Monitoramento dos Dados dos Sensores do AAL

Para monitorar os pontos da fonte de dados foi utilizado o recurso *watch list* do ScadaBR. Esse recurso permite criar listas dinâmicas dos pontos com seus valores associados, visualizar os últimos tempos de atualização e gerar gráficos de informações históricas.

C. Tratamento de Eventos

Este recurso foi usado para gerar alertas por tratadores de eventos, sempre que os valores das variáveis monitoradas estejam fora ou próxima dos limites mínimos e máximos aceitáveis. Foi utilizado o alarme como tratador de eventos.

4. Ambiente e Resultados Experimentais

Nossos experimentos consistiram em gerar dados em tempo real de sinais biomédicos, juntamente com as variáveis meteorológicas do AAL, enviá-los para a nuvem e analisar o comportamento do sistema, no consumo dos dados pelo cuidador formal e informal e, avaliar a confiabilidade do sistema referente aos pacotes enviados em relação aos entregues.

4.1. Cenário Experimental

Idealizamos e criamos um cenário real, com todos os componentes e dispositivos descritos na seção 3. Para a realização dos experimentos, as três entidades (utente no AAL, provedor de dados e cuidador da saúde) foram distribuídas em diferentes cidades de um mesmo estado. O utente monitorado é um paciente idoso, capacitado e que possui histórico de crises asmáticas.

Dois usuários, com diferentes visões do sistema foram criados, representando, um cuidador formal e um cuidador informal da saúde. O cuidador informal está autorizado em visualizar a monitoração das condições ambientais do AAL, diretamente da IHC do ScadaBR. E o cuidador formal pode consumir a fonte de dados do AAL a partir de uma aplicação externa, pois, pode haver a necessidade de correlacionar os valores das variáveis meteorológicas monitoradas com os dados clínicos do utente.

Com a plataforma de eHealth [Cooking-hacks s.d.] foram usados os sensores de pulso, oxigenação no sangue, temperatura corporal, pressão sanguínea, posição e queda do paciente. O tempo entre leituras de 5 minutos foi usado para os sensores de pulso, oxigênio, temperatura corporal e posição do paciente. E para a pressão arterial foi feita a leitura a cada 2 horas. Foram realizadas 9 baterias de testes, totalizando 807 leituras dos sensores biomédicos.

Para monitorar a temperatura ambiente, temperatura do ponto de orvalho e umidade relativa do ar, foram usados os sensores DHT11, com 5 nodos ZigBee e 1 nodo coordenador. Os nodos ZigBee, foram configurados para enviar amostras das leituras dos sensores DHT11 a cada 15 segundos. As monitorações foram avaliadas, por um período de 30 dias.

4.2. Monitoração das Variáveis Meteorológicas do AAL pelo Cuidador Informal

O cuidador informal é um usuário autorizado para monitorar a temperatura, umidade relativa e ponto de orvalho em qualquer cômodo do AAL e tomar algumas ações de sua competência quando observar anormalidades. Alguns tratadores de eventos foram configurados para gerar alertas em situações anormais ou que exigem atenção.

No caso da temperatura, foram usados os detectores de eventos para limites superior a 32°C e inferior a 10°C. No período monitorado as temperaturas medidas ficaram dentro dos limites estabelecidos (Figura 3). O ScadaBR foi configurado para armazenar os valores da temperatura apenas quando o valor da mesma sofrer alteração.

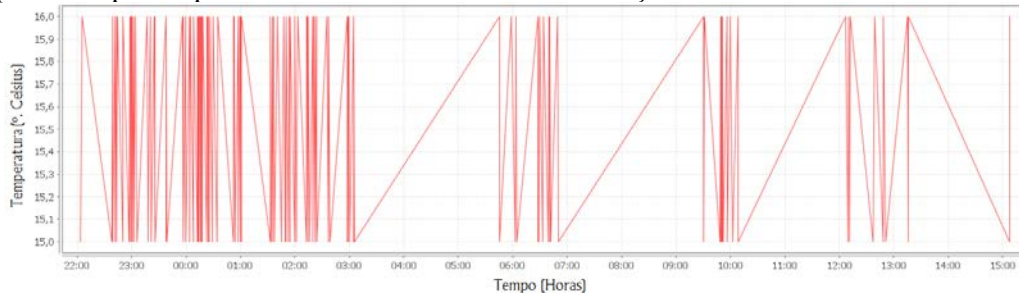


Figura 3. Temperatura Ambiente em um dos Cômodos do AAL

A OMS preconiza que os valores ideais da umidade relativa devem ficar na faixa de 40% a 70%. Sendo que de 20% a 30% exige-se atenção e alguns cuidados; de 12% a 20% deve-se entrar em estado de alerta e abaixo de 12% a situação torna-se emergencial. Os detectores de alertas e níveis de alarmes para umidade relativa foram definidos com base nessas recomendações. No período das monitorações, a umidade relativa ficou entre 48 a 51% (Figura 4).

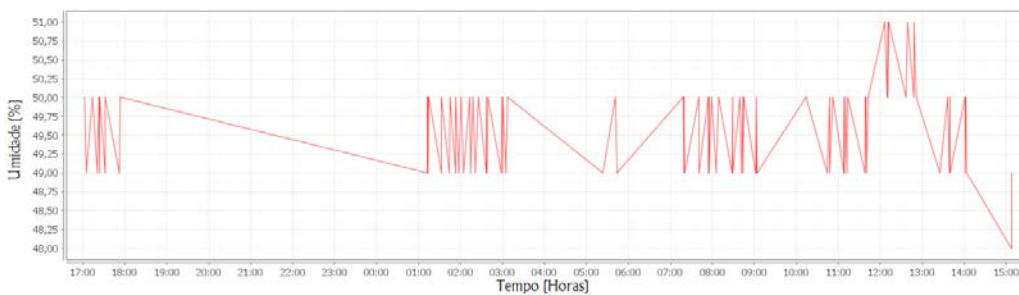


Figura 4. Umidade Relativa em um dos Cômodos do AAL

4.3. Monitoração do AAL pelo Cuidador Formal

O cuidador formal realiza as análises das medições diretamente no PEP (Prontuário Eletrônico do Paciente), pois este necessita avaliar os dados clínicos do utente, o histórico das crises asmáticas e se for o caso correlacionar os fatores meteorológicos, com as crises asmáticas, fazer inferências e descobrir novos fatos.

Note que nosso objetivo não consiste em programar um PEP e sim, disponibilizar essas informações para serem consumidas por uma aplicação de terceiros e avaliar como o sistema proposto se comporta. Por outro lado, o fato das informações do AAL estarem disponíveis na mesma aplicação facilita e pode agilizar a avaliação do cuidador. Para isso prototipamos uma aplicação em Java (Figura 5), capaz de ler as informações do AAL, persistidas no MySQL do ScadaBR. Opcionalmente, o cuidador formal, pode salvar no PEP a data, a hora e os valores das variáveis monitoradas, para fazer o registro, juntamente com a ocorrência da crise asmática.

Na Figura 5 pode-se observar que a monitoração das condições ambientais do AAL pode ser realizada por cômodo, por período de tempo e pelo filtro de valores e operadores relacionais para buscar a temperatura e a umidade. Observa-se também que a aplicação exibe o índice de conforto térmico, de acordo com os valores medidos, ou seja, a uma temperatura de 27°C, com umidade relativa a 41% e temperatura do ponto de orvalho de 12°C, a sensação de conforto para o utente é “muito confortável”.

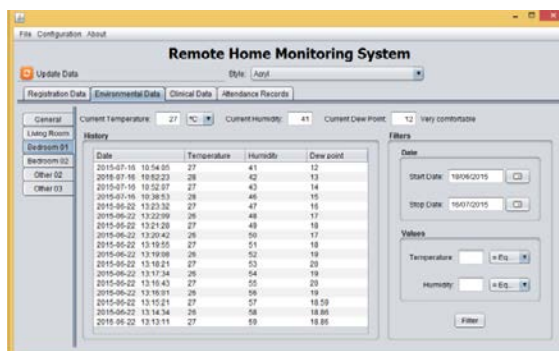


Figura 5. Monitoramento Remoto Usada pelo Cuidador Formal

Para o cálculo do índice de conforto, foi utilizada a Equação 1. A mesma foi calculada, com os valores existentes na base de dados do ScadBR. Para o cálculo, além do uso da temperatura e umidade, usou-se o valor da temperatura do ponto de orvalho (Equação 2).

4.4. Confiabilidade do Sistema Proposto

Para determinar a confiabilidade do sistema, os dados dos sensores de ambiente foram monitorados por um período de um mês. Um registro de *log* foi gerado e analisado no Raspberry PI para manter informações de pacotes recebidos com sucesso dos nodos ZigBee. O *log* mantém o ID do nodo, o tempo e o contador de pacotes. E com isso, foi possível obter o percentual de confiabilidade no envio dos dados dos nodos ZigBee para o Gateway IoT, conforme ilustrado na Tabela 3.

Tabela 3. Confiabilidade do Sistema em Função da Entrega dos Pacotes

ID do nodo	Tempo médio entre chegadas	Número de pacotes esperados	Número de pacotes recebidos	Número de pacotes perdidos	Confiabilidade (%)
1	15,2	176211	166791	9420	94,65412186
2	15,8	169519	159619	9900	94,15994624
3	16,2	165333	153993	11340	93,14112903
4	15,6	171692	161896	9796	94,29440508
5	15,5	172800	163006	9794	94,33224537

Para o envio dos dados do Gateway IoT para o ScadaBR, usou-se o protocolo de transporte TCP, neste caso, o próprio protocolo possui mecanismos de controle de erro, garantindo confiabilidade entre origem e destino.

5. Considerações e Continuidade do Trabalho

Através das pesquisas evidenciamos que as variáveis meteorológicas têm sido usadas para avaliar a qualidade do ar, no ambiente interno, bem como fazer correlações com crises asmáticas. Embora, a ZigBee [Yang et al. 2014b] seja considerada bastante flexível e mais adaptável às RSSF, observamos que existem pouquíssimas propostas de IoT com ZigBee na área da saúde. Isso demonstra que a aplicabilidade da IoT e ZigBee na aquisição e transmissão de sinais em AAL ubíquo, é carente de investigação.

Evidenciando a necessidade descrita acima, foi proposto e implementado a IoT com a RSSF ZigBee para o monitoramento das condições do ar ambiente de um AAL, visando os cuidados do utente com alguma pré-disposição a crises asmáticas. Para validar nossa proposta realizamos os experimentos, com as três principais entidades (utente no AAL, provedor de dados e cuidador da saúde) situadas em diferentes cidades. Os resultados foram satisfatórios, uma vez que foi possível gerar gráficos e alertas em tempo quase real; disponibilizar os dados das variáveis climáticas do AAL tanto no IHC do sistema supervisorio, quanto para serem consumidos por uma aplicação externa. Além disso, o sistema demonstrou uma confiabilidade de 94%. Contudo, convém implementar a tolerância a falhas para melhorar este percentual.

Como propostas futuras, pretende-se: - configurar a ZigBee tolerante a falhas; escalar o número de AAL; migrar para uma base de dados NoSQL e criar um motor semântico para suportar a análise humana pelo cuidador da saúde.

Referências

- Araújo, R. A. F., Gomes Filho, M. F. and Dantas, R.T. “Morbidades em Função das Variáveis Meteorológicas em Campina Grande – PB”, In: Jacques Servain; José Nilson B. Campos; Eduardo Savio P.R. Martins; Dirceu Silveira Reis Junior. (Org.). *Climat de'l Atlantique Tropical et Impacts sur le Nordeste (CATIN)*. 01 ed. Fortaleza - CE: Gráfica e Editora Ronda Ltda, 2010.
- Alduchov, O. A. and Eskridge, R. E. (1996) “Improved Magnus form approximation of saturation vapor pressure”. *Journal of Applied Meteorology*, v. 35, n. 4, p. 601-609.
- Azevedo, J. V. V., Santos, A. C., Alves, T. L. B., Azevedo, P. V. and Olinda, R. A. (2015) “Influência do Clima na Incidência de Infecção Respiratória Aguda em Crianças nos Municípios de Campina Grande e Monteiro, Paraíba, Brasil”, *Revista Brasileira de Meteorologia*, v. 30, n. 4, p. 467-477.
- Ayres-Sampaio, D., Teodoro, A.C., Sillero, N., Santos, C., Fonseca and J., Freitas, A. (2014) “An investigation of the environmental determinants of asthma hospitalizations: An applied spatial approach”, *Geography*, v. 47, p. 10-19, ISSN 0143-6228.
- Bader, E., Al-Shamma'a, A. and Shaw, A. (2009) “Design and implementation of low-power and wireless multi-sensor devices for industrial applications”, *Developments in eSystems Engineering (DESE), Second International Conference on, Abu Dhabi*, p. 418-424.
- Collins, KJ. (1986) “Low indoor temperatures and morbidity in the elderly”, *Oxford Journals Medicine Age and Ageing*, v. 15, p. 212-220.
- Cooking-hacks. e-health sensor platform v2.0 for arduino and raspberry pi [biometric/medical applications].
- Critchfield, H. J. *General Climatology*. Englewood Cliffs: Prentice-Hall, 1983.
- Nazário, D.C, Tromel, I.V.B., Dantas M.A.R., and Todesco, J.L. (2014) “Toward Assessing Quality of Context Parameters in a Ubiquitous Assisted Environment,” *IEEE Symposium on Computers and Communications – ISCC*, p.1-6.
- Da Silva, M.P., Dantas, M.A.R., Gonçalves, A.L., Pinto, A.R. (2015) “A managing QoE approach for provisioning user experience aware services using SDN,” *11th ACM Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet)*, p. 51-58.
- Ezequiel, O.D.S., Gazeta, G.S. and Freire, N.M.D.S (2007) “Prevalência dos atendimentos por crises de asma nos serviços públicos do Município de Juiz de Fora (MG)”, *J. bras. pneumol.* v. 33, n.1.
- Gonçalves, F. L. T. and Coelho, M. S. Z. S. (2010) "Variação da morbidade de doenças respiratórias em função da variação da temperatura entre os meses de abril e maio em São Paulo", *Ciência e Natura*, UFSM, 32 (1): 103-118.
- Horstmeyer, S. (2006). *Relative Humidity Relative to What? The Dew Point Temperature a better approach*, in <http://www.shorstmeyer.com/wxfaq/humidity/humidity.html>. Access in July 2015.
- Ito, S., Kondo, H., Kawaoi, T., Hiruma, F., Takashima, H., Kim, B., Togasaki, K., Togo, T. and Abe, T. (1989) “Outbreaks of asthma attacks and meteorological parameters-multivariate analysis”, *Japanese Journal of Allergology*.
- Jayakumar, D., Omana, J., Sivakumar and M. and Senthil, B. (2015) “A safe guard system for mine workers using wireless sensor networks”, *International Journal of Applied Engineering Research* 10(8):21429-21441.
- Fleischer, S.L.M. and Asnani, G.C (1978) “The influence of weather on asthma in Nairobi”, *International Journal of Biometeorology*, v. 22, p. 263-270.
- GIMA, Global Initiative for Asthma. Available in <http://www.ginasthma.org/>. Access in July 2015.

- Lawrence, M. G. (2005) “The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air”, American Meteorological Society, [s.l].
- Lier, Q. D. V., Dam, J. C. V., Metselaar, K., Jong R. and Duijnsveld, W. H. M. (2008) “Macroscopic root water uptake distribution using a matric flux potential approach”, *Vadose Zone Journal*, v.7, n.3, p.1065-1078.
- Nam, J.-W., Kim, H.-T., Min, B.-B., Kim, K.-H., Kim, G.-S. and Kim, J.-C. (2011) “Ventilation control of subway station using USN environmental sensor monitoring system”, *International Conference on Control, Automation and Systems*, pp.305-308.
- Ngo, N. S. and Horton, R. M. (2016) “Climate change and fetal health: The impacts of exposure to extreme temperatures in New York City”, *Environmental Research*, v. 144, Part A, p. 158-164.
- Paulus, H.J. and Smith, T.J. “Association of allergic bronchial asthma with certain air pollutants and weather parameters” (1967) *International Journal of Biometeorology*, v. 11, p. 119-127.
- Quinn, A., Tamerius, J. D., Perzanowski, M., Jacobson, J. S., Goldstein, I., Acosta, L. and Shaman, J. (2014) “Predicting indoor heat exposure risk during extreme heat events”, *Science of The Total Environment*, v. 490, p. 686–693.
- Santić, Z., Santić, Z., Santić, K., Kondza, D. and Bogut, S. (2002) ”The relationships between the asthma and weather”, *Medicinski Arhiv*, v. 56, n. 3, p. 155-157.
- SCADABR. Available in <http://www.scadabr.com.br/>. Access in May 2015.
- Sun, F.M., Fang, Z., Zhao, Z., Xu, Z.H., Tan, J., Chen, D.L., Du, L.D., Qian, Y.M., Hui, H.Y. and Tian, L.L. (2013) “A wireless ZigBee router with PHT sensing for health monitoring”, *IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCoM*, pp. 1773-1778.
- Shimizu, S., Kagawa, J. and Ishiguro, M. (2001) ”The association between emergency clinic visits for asthmatic attacks and fluctuating environmental factors”, *Japanese Journal of Allergology*, v. 50, n. 7, p.612-620.
- Strannegard, I.-L. and Strannegard, O. (1990) “Childhood bronchial asthma in a desert country”, *Allergy: European Journal of Allergy and Clinical Immunology*, v. 45, n. 5, p. 327-33.
- Sung, W.-T., Chen, J.-H. and Wang, H.-C. (2012) “Wisdom health care environment systems for monitoring and automated control via RBF function”, *Applied Mechanics and Materials*, v. 157-158, p. 315-318.
- Teixeira G., F.L., Spohr Nedel, A. and Cardoso Alves, M.R (2012) “An analysis of the air relative humidity in internal and external environment in the city of São Paulo”, *Brazil, Revista Brasileira de Medicina*, v. 69, n. 7.
- Teixeira, F. A. and Postolache, O. (2014) “Wireless sensor network and web based information system for asthma trigger factors monitoring”, *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pp. 1388-1393.
- Thom EC (1959) “The discomfort index”, *Weatherwise*, v. 12, p. 57-61.
- Zhang, K., Li Y., Schwartz, J. D. and O'Neill, M. S. (2014) “What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods”, *Environmental Research*, v. 132, p. 350–359.
- Yang, C.-T., Liao, C.-J., Liu, J.-C., Den, W., Chou, Y.-C. and Tsai, J.-J. (2014a) “Construction and application of an intelligent air quality monitoring system for healthcare environment”, *Journal of Medical Systems*, v. 38, n. 2, p. 1-10.
- Yang, C.-T., Liu, J.-C., Liao, C.-J., Wu, C.-C. and Le, F.-Y. (2014b) “On construction of an intelligent environmental monitoring system for healthcare”, *Parallel and Distributed Computing, Applications and Technologies, PDCAT Proceedings*, art. no. 6904262, p. 246-253.
- Zaninović, K. and Raos, M. (2001) “The influence of meteorological parameters on asthmatic attacks in children”, *Hrvatski Meteoroloski Casopis*, v. 37, p. 89-94.

Sistema de Informação para Perguntas e Respostas em Doenças Crônicas

Luciana F. Almansa¹, Alessandra A. Macedo¹

¹Departamento de Computação e Matemática – Universidade de São Paulo (USP)
Caixa Postal 14.027-175 – Ribeirão Preto – SP – Brazil

{luciana.f.almansa, ale.alaniz}@usp.br

Abstract. *The search for relevant information considering medical scientific papers is a complex task due to the lack of time and the complexity of creation of queries. Therefore, this work presents a Question Answering (QA) architecture that helps healthcare professionals find related answers quickly. This framework is supported mainly by Text Mining and Information Retrieval techniques and the evaluation is using a chronic disease reference collection and performance measures. This work intends to contribute with a general QA framework architecture to be used in different medical fields.*

Resumo. *No ambiente médico, buscar informações relevantes em artigos científicos é uma tarefa que exige tempo e experiência dos profissionais. Assim, o objetivo deste trabalho é apresentar uma arquitetura de um sistema do tipo Pergunta e Respostas (PR) que auxilie profissionais da área da saúde na busca rápida por respostas. O sistema foi desenvolvido principalmente por técnicas de Mineração de Texto e Recuperação de Informação. A avaliação do sistema está sendo realizada por meio de uma coleção de referência do domínio de doenças crônicas e epigenética e com o uso de medidas de avaliação de desempenho. Este trabalho pretende contribuir com uma arquitetura genérica de sistemas de PR que pode ser adaptada por vários domínios de informação médica.*

1. Introdução

No ambiente médico e de saúde, especificamente no tratamento clínico do paciente, o papel da informação descrita nos prontuários médicos é registrar o estado de saúde do paciente e auxiliar os profissionais diretamente ligados ao tratamento [Bhat et al. 2016]. Os pesquisadores Shortliffe e Cimino vislumbram o futuro de sistemas computacionais médicos como “*Learning Healthcare Systems*” (LHS). Segundo esses pesquisadores, sistemas LHS funcionarão de acordo com um ciclo de informação cujos dados distribuídos de prontuários eletrônicos do paciente serão submetidos a base de dados de registros médicos e de pesquisa científica para serem processados e gerarem conhecimento [Shortliffe and Cimino 2013].

Os artigos científicos são uma importante fonte de informações e de descobertas científicas relevantes na área da saúde. Geralmente, os artigos científicos são armazenados em bibliotecas digitais como, por exemplo, o PubMed¹. À medida que mais artigos

¹O Pubmed é uma repositório de informações online com mais de 25 milhões de citações, periódicos e livros da literatura biomédica (<http://www.ncbi.nlm.nih.gov/pubmed>).

são inseridos, a rede de informação vai aumentando e se tornando cada vez mais rica e complexa para a busca da informação desejada [Rzhetsky et al. 2009]. O crescimento acelerado do número de publicações anuais e a falta de tempo dos profissionais da saúde faz com que as informações dos artigos científicos sejam pouco aproveitadas pela sociedade, em geral, e pela comunidade científica e profissional [Shortliffe and Cimino 2013]. Ler artigos científicos é uma tarefa que exige tempo e disposição dos pesquisadores. A busca por informações desejadas tende a ser complexa e minuciosa devido à quantidade de artigos escritos em idiomas diferente da língua nativa e em diferentes estruturas e formatos. Neste contexto, os sistemas computacionais de Recuperação de Informação (RI) e de Perguntas e Respostas (PR) foram desenvolvidos.

Os sistemas de RI são responsáveis pela representação e organização das informações de modo que o acesso à informação seja facilitado para o usuário [Baeza-Yates and Ribeiro-Neto 1999]. Uma das desvantagens dos sistemas de RI é a quantidade de documentos retornados como resultado do processo de recuperação de informação [Kolomiyets and Moens 2011]. Em sistemas de PR, a informação que o usuário necessita é apresentada em linguagem natural e, como resultado do processo, respostas curtas e diretas são retornadas ao usuário [Kolomiyets and Moens 2011].

O objetivo deste trabalho é apresentar a arquitetura de desenvolvimento de um sistema de informação do tipo Perguntas e Respostas, o qual coleta perguntas, processa e apresenta respostas a partir da análise de artigos científicos. A arquitetura denominada QASF (*Question Answering Surveillance Framework*) pode ser utilizada em diferentes domínios de informação médica, considerando que os artefatos linguísticos que apoiam o framework como, por exemplo, as ontologias sejam alterados, conforme o domínio de informação escolhido. Inicialmente, o sistema está sendo validado no domínio de doenças crônicas relacionadas a fatores epigenéticos ². Este trabalho pretende auxiliar profissionais da área da saúde na busca rápida e direta por informações de interesse.

Este artigo está estruturado na seguinte forma: a Seção 2 apresenta a arquitetura genérica de um sistema de PR e a arquitetura do QASF com a descrição das técnicas utilizadas na construção de cada módulo. A Seção 3 apresenta a avaliação do QASF. Na Seção 4 são detalhados os trabalhos que se relacionam à proposta do QASF e, por fim, a Seção 5 relata as conclusões do trabalho desenvolvido e trabalhos futuros.

2. Arquitetura do Question Answering Surveillance Framework

Os sistemas de informação do tipo PR tem como objetivo fornecer informações diretas e precisas sobre uma pergunta proposta pelo usuário. Para realizar esta tarefa, estes sistemas utilizam técnicas computacionais, principalmente, das áreas de Extração de Informação (EI), Mineração de Texto (MT) e Recuperação de Informação [Allam and Haggag 2012].

O QASF é uma arquitetura do tipo PR que tem como objetivo auxiliar profissionais da área da saúde respondendo questões escritas em linguagem natural realizadas por profissionais dessa área em linguagem natural. O QASF é composto pelos três módulos tradicionais de PR: (i) processamento da questão, (ii) processamento da resposta e (iii)

²Estudos em medicina genômica sugerem que seres humanos expostos a fatores de risco no início da vida como, por exemplo, a escassez de alimentos, podem sofrer influências na expressão do gene e, como resultado, na vida adulta, desenvolverem doenças crônicas [Barker 2001]. A área da biologia que estuda estas influências na expressão do gene é conhecida como epigenética [Miyake et al. 2012].

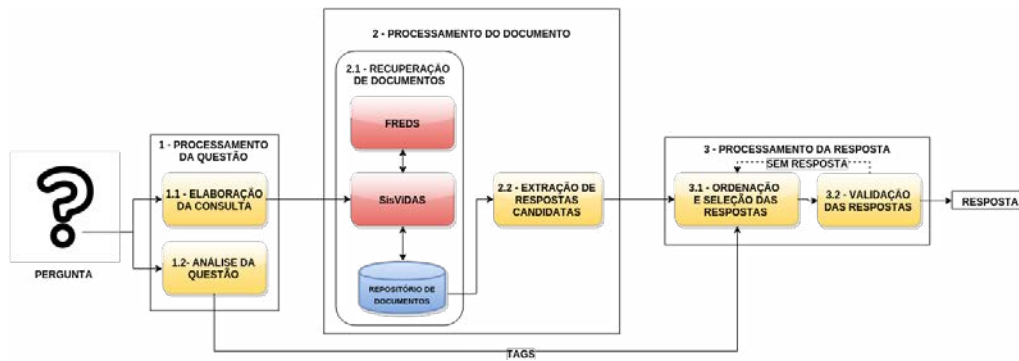


Figura 1. Arquitetura geral de um sistema de PR junto aos submódulos propostos especificamente para o QASF.

processamento do documento; esse último módulo agrega um sistema de suporte que será descrito na Subseção 2.2. A Figura 1 apresenta a arquitetura do QASF que será detalhada nas próximas subseções.

Inicialmente, o domínio de informações em epigenética e doenças crônicas foi escolhido para ser utilizado no desenvolvimento e avaliação deste framework. Contudo, a arquitetura do QASF foi projetada, em sua grande maioria, para ser genérica e suportar diferentes domínios de informações médicas. Porém, o módulo de “Processamento de Documentos” necessita de alguns artefatos linguísticos como, neste caso, uma ontologia do domínio de epigenética e doenças crônicas para realizar uma busca mais apurada pelos artigos que possam conter as respostas corretas. Portanto, caso estes artefatos sejam substituídos, o QASF se torna capaz de responder perguntas em outro domínio médico. Todos os módulos do QASF foram desenvolvidos utilizando a linguagem de computação Python³, apoiada, principalmente, pelas bibliotecas de Processamento de Linguagem Natural (PLN) *NLTK*⁴ e de Aprendizado de Máquina (AM) *scikit-learn*⁵.

2.1. Processamento da Questão

O módulo de “Processamento da Questão” objetiva extrair informações da questão realizada pelo usuário em linguagem natural e utilizar as informações na seleção das respostas candidatas. A etapa de “Processamento da Questão” é dividida em dois submódulos: “Análise da Questão” e “Elaboração da Consulta”.

A “Análise da Questão” busca informações que auxiliam na localização e na verificação das respostas candidatas. Por exemplo, a pergunta “*O que é doenças crônicas?*” pode extrair informações, por exemplo, o assunto da questão (doenças crônicas). Este tipo de informação (doenças crônicas) pode excluir respostas candidatas cujo assunto seja “câncer”. Na literatura, foram encontradas algumas abordagens para “Análise da Questão”: correspondência de padrões [Er and Cicekli 2013], Máquina de Vetores de Suporte (do inglês, *Support Vector Machine* - SVM) e análise morfosintática [Monz 2003].

O submódulo de “Análise da Questão” do QASF é suportado por abordagens de

³<https://wiki.python.org/moin/>

⁴<http://www.nltk.org/>

⁵<http://scikit-learn.org/stable/>

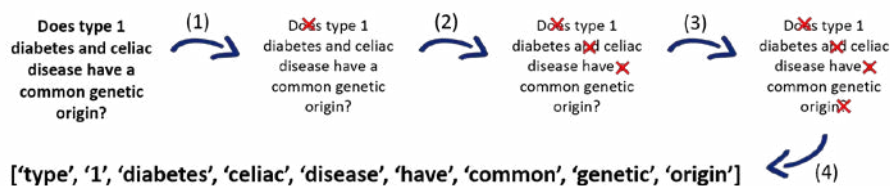


Figura 2. Representação do processo de construção da consulta computacional baseada na pergunta do usuário.

Aprendizado de Máquina, especificamente por SVM e Naive Bayes, para construção de classificadores que categorizam as questões de acordo com o seu assunto. Os classificadores foram construídos e validados utilizando um conjunto de perguntas subdividido de acordo com o assunto da pergunta. Os resultados obtidos foram apresentados e comparados entre si na Seção 3.

A “Elaboração da Consulta” extraí um conjunto de palavras-chave da pergunta feita pelo usuário que será transmitido para o módulo de “Processamento do Documento”, especificamente na etapa de “Recuperação de Documentos”. No QASF, a construção do vetor de palavras é feita desconsiderando os pronomes interrogativos, as “*stopwords*” e as pontuações. Na Figura 2, o modelo de construção da consulta computacional do QASF a partir de uma pergunta inserida pelo usuário é apresentado. Para iniciar o processo, o usuário insere uma pergunta e como resultado final é construído um vetor com as palavras-chave da pergunta. Na primeira etapa, os pronomes interrogativos são desconsiderados da frase inserida pelo usuário. Em seguida, as “*stopwords*” e, por último, as pontuações da frase são eliminadas. O resultado final é um vetor de palavras-chave que representa a pergunta do usuário.

Os resultados alcançados pelos submódulos “Análise da Questão” e “Elaboração da Consulta” não se relacionam nesta etapa, contudo, eles auxiliam o submódulo de “Extração das Respostas Candidatas” na busca pelas respostas candidatas.

2.2. Processamento dos Documentos

O módulo de “Processamento dos Documentos” engloba os submódulos de “Recuperação dos Documentos” candidatos e de “Extração das Respostas Candidatas”. A tarefa de “Recuperação de Documentos” objetiva a recuperação de documentos que contenham possíveis respostas para a pergunta elaborada pelo usuário. Como entrada de dados desta etapa, o vetor de palavras-chave da etapa de “Processamento da Questão” é inserido.

No QASF, o submódulo de “Recuperação de Documentos” foi projetado a partir do SisViDAS [Pollettini et al. 2014]. O SisViDAS é um sistema que relaciona artigos científicos a fatores de riscos descritos em prontuários médico do paciente e alerta profissionais da área da saúde sobre problemas no desenvolvimento humano. Este submódulo foi construído para ampliar o leque das informações manipuladas para a busca de informações relevantes no contexto de um sistema de PR para obter mais precisão das respostas retornadas ao usuário. O SisViDAS compreende os processos de busca da informação em artigos científicos e prontuários médico do paciente nos domínios de epigenética e doenças crônicas.

O submódulo de “Extração das Respostas Candidatas” é um dos mais complexos,

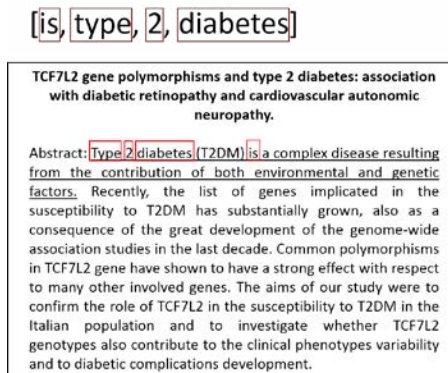


Figura 3. Esquema que mostra a busca por parágrafos candidatos.

pois o usuário pode fazer perguntas em diferentes níveis de complexidade. Na literatura, vários autores propõem diferentes abordagens como, por exemplo, utilizar as palavras encontradas no submódulo de “Análise da Questão”, o uso de árvores de análise sintática ou de grafos de dependência [Monz 2003] e a proximidade linear [Gupta and Gupta 2012].

No QASF, a extração de respostas candidatas utiliza os documentos retornados pelo submódulo de “Processamento do Documento”. Inicialmente, o conteúdo de cada artigo retornado é segmentado em parágrafos. Em seguida, uma comparação entre o vetor de palavras que representa a pergunta inserida pelo usuário e cada parágrafo segmentado dos artigos retornados pelo submódulo de “Recuperação de Documentos” é realizada. Neste sentido, as palavras contidas no vetor de palavras-chave e no parágrafo candidato ao mesmo tempo são contadas. Caso a quantidade de palavras seja igual ou maior a um limiar l , que pode ser definido manualmente, o parágrafo é considerado um parágrafo candidato. A Figura 3 apresenta um exemplo de resumo de artigo cuja frase sublinhada ilustra o processo de seleção desse parágrafo como resposta candidata.

2.3. Processamento da Resposta

O módulo de “Processamento da Resposta” é a última etapa do processo de PR. Neste módulo, as respostas candidatas são selecionadas e ordenadas de acordo com o grau de similaridade entre a questão e as respostas candidatas. A etapa se divide entre os submódulos de “Seleção e Ordenação das Respostas” e “Validação das Respostas”. Como resultado deste módulo, as respostas para a pergunta inserida pelo usuário no início do processo de PR são retornadas. Os sistemas de PR não precisam, necessariamente, retornar uma única resposta, eles podem retornar várias respostas curtas ordenadas de acordo com o grau de relevância. Atualmente, o QASF tem retornado as cinco respostas mais semelhantes à pergunta inserida pelo usuário, contudo, esse valor pode ser reajustado.

O submódulo de “Ordenação e Seleção das Respostas” realiza a ordenação das respostas candidatas, por meio de medidas de similaridade, e apresenta ao usuário um número n de respostas mais semelhantes. Na ordenação das respostas candidatas, alguns critérios podem ser adotados: (i) a quantidade de palavras da questão do usuário que é reconhecida nas respostas candidatas na mesma ordem de posição, (ii) o número de palavras que separam as palavras-chave mais distantes entre si no parágrafo analisado ou (iii) o número de palavras incompatíveis com as palavras-chave [Allam and Haggag 2012].

No QASF, o submódulo de “Ordenação e Seleção das Respostas” realiza o cálculo da similaridade do cosseno. Nesse contexto, a pergunta realizada pelo usuário e as respostas candidatas são convertidas para o espaço vetorial. Desse modo, a similaridade representada pelo cosseno entre a pergunta do usuário e cada resposta candidata é calculada. Desse modo, as respostas são ordenadas e apresentadas ao usuário. A Figura 4 apresenta um exemplo da ordenação realizada pelo QASF.

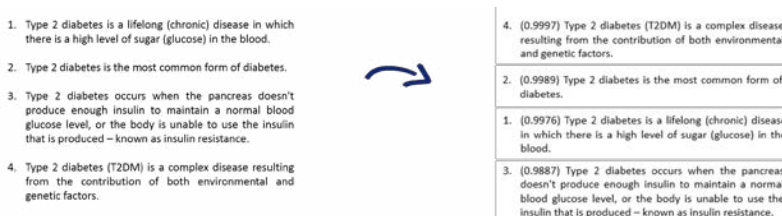


Figura 4. Esquema de ordenação e seleção das respostas candidatas.

3. Avaliação do QASF: Módulo de Análise da Questão

Uma das formas mais conhecidas de avaliação e validação das respostas retornadas pelos sistemas de PR é realizada com o auxílio de *corpora* compostos por tuplas de perguntas e respostas. Os *corpora* podem ser criados especificamente para a avaliação do sistema em questão ou disponibilizados por conferências como a TREC [Bilotti and Nyberg 2006]. Outras formas de avaliar os sistemas de PR são pela comparação dos resultados de avaliações efetuadas pelo sistema e por um colaborador e pelo uso de FAQs (do inglês, *Frequently Asked Questions*). No contexto das perguntas FAQs, as respostas das perguntas são inseridas em partes aleatórias do documento e o sistema deve retornar a resposta esperada [Lin and Demner-Fushman 2005].

Após a construção dos módulos de processamento da questão, dos documentos e das respostas, a etapa seguinte foi iniciada com as primeiras avaliações do sistema. O primeiro módulo a ser avaliado foi o módulo de análise da questão. Este módulo foi construído com técnicas de SVM e Naive Bayes. *Cao et al.* apresentam uma coleção de referência em seus trabalhos que foi utilizada para a construção e avaliações dos classificadores do QASF [Cao et al. 2011]. Na Subseção 3.1, os detalhes desta coleção de referência são apresentados e na Subseção 3.2, as avaliações e os resultados obtidos são resumidos.

3.1. Coleção de Referência

A coleção de referência usada para avaliar o módulo de “Análise da Questão” do QASF foi construída para avaliação do sistema AskHERMES [Cao et al. 2011] e disponibilizada pelos autores para *download* na web. A coleção de perguntas é composta por 4654 questões médicas sobre doenças crônicas. As questões são separadas em 12 tópicos: dispositivo, diagnóstico, epidemiologia, etiologia, história, gestão, farmacologia, achado físico, procedimento, prognóstico, teste e tratamento e prevenção. As questões foram classificadas pelos próprios clínicos do projeto AskHERMES que fizeram as perguntas. Os tópicos não possuem a mesma quantidade de perguntas. Alguns exemplos de perguntas são: “*What is the protein you can test to see if someone has type 1 or type 2?*”. Esta pergunta é do tipo teste. Outro exemplo é “*Is diabetes a risk factor for carpal tunnel syndrome?*” do tipo diagnóstico.

Tabela 1. Classificação de uma coleção de referência em classes pelo QASF.

	Corpus Completo		4258 perguntas – 7 classes	
	Naive Bayes	SVM	Naive Bayes	SVM
Precisão	0.48	0.51	0.48	0.59
Revocação	0.42	0.53	0.46	0.58
F-Measure	0.32	0.51	0.38	0.56
Acurácia	0.43 +- 0.04	0.55+-0.04	0.47 +-0.02	0.59 +-0.03

3.2. Resultados

Os testes realizados com os classificadores e a coleção de referência foram separados em dois cenários para avaliação. No primeiro cenário, toda a coleção de referência é utilizada para avaliar os classificadores Naive Bayes e SVM. No segundo cenário, os tópicos que continham menos de cem perguntas foram desconsiderados. Consequentemente a coleção de referência foi reduzida a 4258 questões distribuídas nos seguintes sete tópicos: diagnóstico (475 perguntas), teste (578 perguntas), tratamento e prevenção (319 perguntas), achados patológicos (189 perguntas), etiologia (172 perguntas), gestão (1.380 perguntas) e farmacologia (1.145 perguntas).

Na Tabela 1, os resultados dos experimentos são apresentados. As duas primeiras colunas compõem o primeiro cenário com toda a coleção de referência para classificação. As duas últimas colunas apresentam o segundo cenário cujos tópicos com menos de cem perguntas são desconsiderados. A tarefa de classificação é avaliada a partir da execução de dois algoritmos de aprendizado de máquina, SVM e Naive Bayes. Em ambos os cenários, o QASF com o classificador SVM apresentou melhor desempenho com valores acima de 0.50 para as medidas de Precisão, Revocação, F-Measure e Acurácia. Especificamente, no segundo cenário, os valores da precisão e acurácia alcançaram valores próximos a 0.6.

Como já apontado em estudos anteriores [Cao et al. 2011, Yen et al. 2013], o classificador SVM apresentou melhor desempenho em ambos os cenários. Além disso, é possível perceber a importância do balanceamento das classes na tarefa de classificação da questão. Esta avaliação indicou performance moderada do QASF. Contudo, um ponto a ser destacado é que as perguntas foram classificadas em seu formato original sem a aplicação de etapas de tratamento da questão como, por exemplo, a eliminação das *stopwords* ou a adição de termos similares por meio da UMLS. Este desempenho moderado pode sobrecarregar ou até atrapalhar os demais módulos do sistema, uma vez que as informações deste módulo auxiliam na escolha das respostas candidatas. Atualmente, os demais módulos estão em avaliação.

4. Trabalhos Relacionados

Na literatura científica atual, diferentes tentativas de construção de sistemas de PR foram identificadas. O primeiro trabalho relacionado apresenta um sistema de PR baseado em abordagens de Aprendizado de Máquina para integração das etapas de classificação da questão e da seleção das respostas. O classificador categoriza a questão inserida pelo usuário e repassa esta informação para a etapa de ordenação das passagens que reorganiza as passagens recuperadas da etapa de processamento dos documentos [Yen et al. 2013]. No domínio médico há dois sistemas que se destacam. O primeiro é o AskHERMES que processa semanticamente questões e busca as respostas em

Tabela 2. Organização das arquiteturas encontradas.

Sistemas de PR	Domínio Informação	Artefatos Linguísticos	Método Classificação
QASF	saúde	sim	SVM e Naive Bayes
[Yen et al. 2013]	genérico	não	SVM
AskHERMES [Cao et al. 2011]	saúde	sim	SVM
MEANS [Ben Abacha and Zweigenbaum 2015]	saúde	sim	NER ⁶
[Amorim et al. 2012]	genérico	sim	AIML ⁷ e Técnicas de RI
[Prestes 2011]	genérico	sim	Análise de Padrões
[Arrigo et al. 2014]	genérico	não	Análise de Padrões

resumos de artigos científicos do MEDLINE, PubMed, eMedicine e Wikipédia. O sistema integra UMLS [Bodenreider 2004, Brin 1999] na expansão dos termos da consulta do usuário, técnicas de PLN [Liddy 2001], RI [Baeza-Yates and Ribeiro-Neto 1999] e EI [Cowie and Lehnert 1996]. Segundo os autores, o sistema apresentou uma boa habilidade na solução de questões longas e complexas [Cao et al. 2011]. O segundo sistema do domínio médico é o MEANS, apoiado por abordagens semânticas de PLN e Web Semântica. As abordagens semânticas são utilizadas para análise dos dados em dois níveis de complexidade: busca por entidades médicas (drogas, sintomas e doenças) e análise em nível de relacionamento (tratamentos, prevenções e causas). As questões são focadas nos tipos factuais e booleanas [Ben Abacha and Zweigenbaum 2015].

No Brasil, o desenvolvimento de sistemas de PR é pouco explorado. Um dos sistemas encontrados foi desenvolvido para domínios de informações genéricas. Segundo os autores, o sistema possui uma arquitetura flexível que independe do conteúdo da pergunta inserida [Amorim et al. 2012]. Para que o sistema suporte esta abrangência, os autores utilizaram um banco de ontologias junto a uma gama de softwares de apoio. Caso a resposta não seja encontrada neste banco de dados, o sistema busca a resposta na Web.

Neste trabalho apresentado, o autor comparou os seguintes sistemas de PR, com foco no módulo de processamento da resposta levando em consideração a complexidade da linguagem a qual foram criados: (i) um sistema para o idioma português, (ii) um para o idioma inglês e (iii) um em idioma português, porém estendido para o inglês. O autor verificou o uso de regras semânticas e sintáticas e técnicas PLN [Prestes 2011]. Por fim, o último sistema de PR apresentado utiliza a Web para criar um corpus anotado de informações em Português. O sistema faz análise semântica da pergunta e utiliza métodos de aprendizado automático de padrões para classificação da informação e para a análise da precisão das respostas [Arrigo et al. 2014]. A Tabela 2 apresenta uma organização sistematizada das arquiteturas encontradas na literatura. As principais semelhanças entre o QASF e as demais arquiteturas são o uso de artefatos linguísticos e SVM, quando o método de classificação é por AM.

O QASF quando comparado com os demais sistemas de informação do tipo PR possui alguns diferenciais. O QASF foi inicialmente projetado para atuar no domínio médico, cuja busca por informações clínicas relevantes pode ser mais complexa quando

comparada a outros domínios da informação. Por essa razão, o módulo de recuperação de documentos foi construído baseado em um sistema especialista em recuperação de informações médicas, o SisVIDAS. A arquitetura do sistema QASF é ampla e pode ser estendida para vários outros domínios de informação, variando apenas os artefatos linguísticos como as ontologias. Outro diferencial é que o sistema está sendo desenvolvido para processar perguntas com diferentes níveis de complexidade, ou seja, perguntas que exigem diferentes níveis de processamento do QASF para encontrar a resposta mais adequada.

5. Conclusões

No ambiente médico, buscar informações relevantes em artigos científicos é uma tarefa que exige tempo e experiência dos profissionais. Portanto, o objetivo deste trabalho é apresentar uma arquitetura de um sistema do tipo Pergunta e Respostas (PR) que auxilie profissionais da área da saúde na busca rápida por respostas. Atualmente o sistema está focado em doenças crônicas e os módulos de processamento da questão, documento e resposta estão em fase final de avaliações. No submódulo de “Análise da Questão”, avaliações apontam que SVM apresenta melhores resultados na classificação de perguntas. Como trabalhos futuros, pretende-se: (i) avaliar o QASF em outros domínios de informação como, por exemplo, o câncer de tireoide e (ii) transformar o QASF em aplicação composta por *interfaces* de usuário web e para dispositivos móveis que facilite a interação com o sistema.

6. Agradecimento

Os autores gostariam de agradecer a CAPES pelo apoio financeiro concedido para o desenvolvimento do projeto.

Referências

- Allam, A. M. N. and Haggag, M. H. (2012). The question answering systems: A survey. *Int. Journal of Research and Reviews in Information Sciences (IJRRIS)*, 2(3):211–220.
- Amorim, M. T. C. F. d., Cury, D., and Menezes, C. S. (2012). Um Sistema Inteligente Baseado em Ontologia para Apoio ao esclarecimento de Dúvidas.
- Arrigo, A. J. S., Silva, E. G., Martins, H. P., and Silva, P. P. (2014). Desenvolvimento de um Sistema de Pergunta e Resposta Baseado em Corpus. In *14o Congresso Nacional de Iniciação Científica (CONIC-SEMESP)*, pages 1–6, São Paulo, SP.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern information retrieval*, volume 463. ACM Press New York, 1nd edition.
- Barker, D. J. P. (2001). Fetal and infant origins of adult disease. *Monatsschrift Kinderheilkunde*, 149(1):S2–S6.
- Ben Abacha, A. and Zweigenbaum, P. (2015). MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies. *Information Processing & Management*, 51(5):570–594.
- Bhat, S., Gijo, E., and Jnanesh, N. (2016). Productivity and performance improvement in the medical records department of a hospital: An application of lean six sigma. *International Journal of Productivity and Performance Management*, 65(1):98–125.

- Bilotti, M. W. and Nyberg, E. (2006). Evaluation for scenario question answering systems. In *Proc. of the Int. Conference on Language Resources and Evaluation*, pages 1–6.
- Bodenreider, O. (2004). The unified medical language system (umls): integrating biomedical terminology. *Nucleic acids research*, 32(suppl 1):D267–D270.
- Brin, S. (1999). Extracting patterns and relations from the world wide web. In *Selected Papers from the International Workshop on The World Wide Web and Databases, WebDB '98*, pages 172–183, London, UK, UK. Springer-Verlag.
- Cao, Y., Liu, F., Simpson, P., Antieau, L., Bennett, A., Cimino, J. J., Ely, J., and Yu, H. (2011). AskHERMES: An online question answering system for complex clinical questions. *Journal of biomedical informatics*, 44(2):277–88.
- Cowie, J. and Lehnert, W. (1996). Information extraction. *Comm. of the ACM*, 39(1):80–91.
- Er, N. P. and Cicekli, I. (2013). A Factoid Question Answering System Using Answer Pattern Matching. In *Int. Joint Conf. on Natural Language Processing*, pages 854–858.
- Gupta, P. and Gupta, V. (2012). A survey of text question answering techniques. *International Journal of Computer Applications*, 53(4):1–8.
- Kolomiyets, O. and Moens, M.-F. (2011). A survey on question answering technology from an information retrieval perspective. *Information Sciences*, 181(24):5412 – 5434.
- Liddy, E. D. (2001). Natural Language Processing. In Decker, M., editor, *In Encyclopedia of Library and Information Science*, pages 1–15. New York, New York, USA.
- Lin, J. and Demner-Fushman, D. (2005). Automatically evaluating answers to definition questions. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 931–938.
- Miyake, K., Hirasawa, T., Koide, T., and Kubota, T. (2012). *Neurodegenerative Diseases*, chapter Epigenetics in Autism and Other Neurodevelopmental Diseases, pages 91–98.
- Monz, C. (2003). *From document retrieval to question answering*. Institute for Logic, Language and Computation.
- Pollettini, J. T., Baranauskas, J. A., Ruiz, E. S., da Graça Pimentel, M., and Macedo, A. A. (2014). Surveillance for the prevention of chronic diseases through information association. *BMC Medical Genomics*, 7(1):1–11.
- Prestes, K. V. (2011). Avaliação de métodos de seleção da resposta de um sistema de perguntas e respostas. Technical report.
- Rzhetsky, A., Seringhaus, M., and Gerstein, M. (2009). Getting Started in Text Mining: Part Two. *PLoS Comput Biol*, 5(7):e1000411+.
- Shortliffe, E. H. and Cimino, J. J. (2013). *Biomedical informatics: computer applications in health care and biomedicine*. Springer Science & Business Media.
- Yen, S.-J., Wu, Y.-C., Yang, J.-C., Lee, Y.-S., Lee, C.-J., and Liu, J.-J. (2013). A support vector machine-based context-ranking model for question answering. *Information Sciences*, 224:77 – 87.

Abordagem Deep Learning para Classificação de Lesões Mamárias

Roberto M. P. Pereira¹, Caio E.F. Matos¹, João O.B. Diniz¹,
Geraldo Braz Junior¹, João D. S. De Almeida, Aristófaes Corrêa Silva¹,
Anselmo Cardoso de Paiva¹

¹Núcleo de Computação Aplicada (NCA) – Universidade Federal do Maranhão (UFMA)
Caixa Postal 15.064 – 91.501-970 – São Luís – MA – Brasil

{robertomatheuspp, caioefalcao, joao.obd}@gmail.com
{geraldobraz, joao.dallyson, anselmo.paiva}@ufma.br
ari@dee.ufma.br

Abstract. *Female breast cancer is a major cause of death in western countries. Several computer aid techniques have been developed seeking to help radiologists in task of detection and diagnosis of breast abnormalities. Recently, deep learning techniques have shown good results in the image classification issues. In this paper, we present a methodology to distinguish mass and normal tissue extracted from mammograms using deep convolutional neural networks.*

Resumo. *O câncer de mama é uma das principais causas de morte de mulheres ocidentais. Várias técnicas têm sido desenvolvidas para auxiliar radiologistas na tarefa de detecção e diagnóstico de lesões mamárias. Recentemente, a técnica de deep learning tem apresentado resultados eficientes na classificação de imagens. Este artigo propõe uma metodologia para classificar tecidos de mamografias em massa e não-massa com o uso de redes neurais convolucionais profundas.*

1. Introdução

O câncer de mama se manifesta de diferentes maneiras, sendo uma delas caracterizada pelo crescimento de forma anormal de massas. Uma massa é um aglomerado de células que se unem de forma mais densa em relação ao tecido que a envolve. Este aglomerado pode ser causado por câncer de mama, assim como também por condições benignas. Algumas características das massas são determinantes para estabelecer suas probabilidades de malignidade, como por exemplo tamanho, forma e disposição de suas margens.

Medidas de prevenção, como o auto exame ou a utilização de exames de mamografias e ultrassom de rastreamento têm contribuindo para a redução dos índices. A detecção de massas (não se limitando a apenas estas) suspeitas nesses exames é o primeiro passo para o tratamento eficaz de câncer.

Este trabalho tem como objetivo apresentar uma nova metodologia de classificação de regiões extraídas de mamografias em massa e não massa usando técnicas de *deep learning*. Nessas técnicas, as características são extraídas automaticamente de acordo com a base de treinamento. A contribuição consiste em propôr um método eficiente, de alta sensibilidade e baixa taxa de falso positivos através de *deep learning* que permite que características diversas sejam extraídas da própria imagem gerando um número maior de peculiaridades relevantes durante a classificação.

2. Metodologia

A metodologia proposta é apresentada na Figura 1, e possui quatro etapas: aquisição, aplicação de filtros, treinamento e resultados.

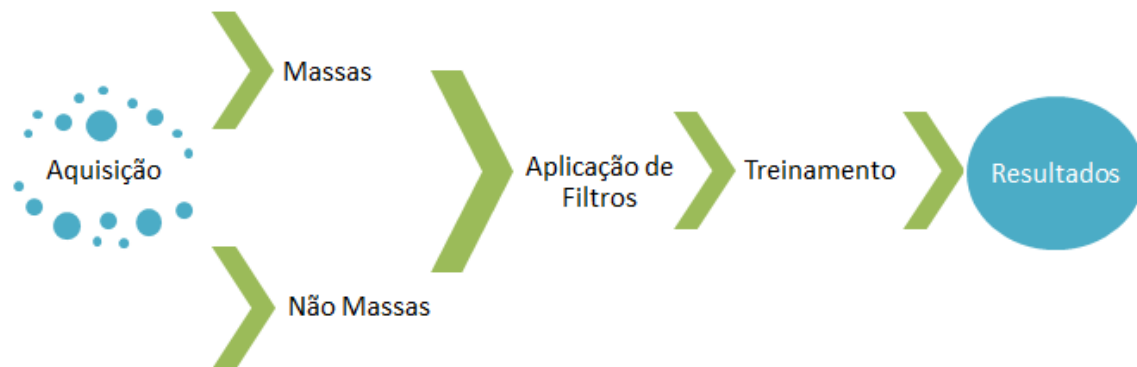


Figura 1. Fluxograma básico da metodologia proposta

2.1. Aquisição

A etapa de aquisição foi dedicada à obtenção das imagens mamográficas. Utilizou-se a base pública, de mamografias digitalizadas, *Digital Database for Screening Mammography* (DDSM) [Heath et al. 1998], disponível na internet.

A seleção das amostras foi realizada com base nas marcações realizadas por especialistas para o caso de massas. Regiões normais foram aleatoriamente retiradas de mamografias que continham massas. O tamanho dos *bounding boxes* para estas foi definido com base no tamanho médio das regiões que continham massa. Para este trabalho foram utilizados 6.181 regiões, sendo 1.156 diagnosticadas como massas e 5.025 não massas.

2.2. Aplicação de Filtros

Com o intuito de verificar o comportamento do resultado das demais etapas, algumas técnicas de pré-processamento foram aplicadas nas imagens. Foram testados realce logarítmico e equalização do histograma. Ainda foram utilizadas nos testes as regiões sem nenhum tipo de pré-processamento.

2.3. Treinamento

Neste trabalho, foram utilizadas duas arquiteturas de Redes Neurais de Convolução, a AlexNet [Krizhevsky et al. 2012] e a GoogLeNet [Szegedy et al. 2015]. Ambas implementam a ideia de *deep learning* e mostram resultados promissores na tarefa de classificação de diferentes classes de imagens.

Ao todo, cada rede neural foi treinada com 3 instâncias da base de dados original. O primeiro grupo, representa a base de dados aplicando-se o realce logarítmico sobre todas as imagens. O segundo grupo, constitui a base resultante da equalização de histograma das imagens originais e por último, temos a base de dados natural, ou seja, sem a aplicação de nenhum filtro.

No processo de treinamento, alguns parâmetros das redes AlexNet e GoogLeNet foram modificados, contudo, a arquitetura física das mesmas é mantida durante todos os treinos executados. Os parâmetros alterados foram o número de interações que foi fixado em 90 para ambas as redes, a taxa de aprendizagem foi inicializada em 0,01 e reduzida a cada 30 interações a uma taxa gama de 0,1, por fim, temos que ambos as redes foram treinadas utilizando o método de gradiente descendente estocástico.

Durante o treino, foram reservados 25% das imagens da base de dados para teste, 50% para treino e os 25% restantes para a validação do treino. Todas as separações treino/teste/validação foram realizadas de maneira aleatória.

2.4. Validação dos Resultados

Os resultados são analisados com base nas métricas de Especificidade (Es), Sensibilidade (Se) e Acurácia (Ac). A especificidade mostra quão bem o método é em reconhecer tecidos que não possuem massa e é definida por $Es = \frac{VN}{VN+FP}$.

A sensibilidade é definida por $Se = \frac{VP}{VP+FN}$ e mostra a relação de acurácia do método para acerto em massas. Finalmente, a acurácia demonstra a relação global de acerto, definida por $Ac = \frac{VP+VN}{VP+VN+FP+FN}$.

3. Resultados Parciais

Os experimentos foram conduzidos de acordo com a metodologia apresentada na Seção 2. Os resultados dos testes, obtidos pelas duas arquiteturas de redes neurais são apresentados nas Tabelas 1 e 2 para a AlexNet e a GoogLeNet, respectivamente.

Os melhores resultados foram obtidos utilizando-se a rede neural AlexNet (Tabela 1) sem a aplicação de filtros sobre a base de dados. As transformações de contraste, em todos os testes realizados, apresentaram resultados inferiores à utilização das imagens originais para o teste com a AlexNet.

A estrutura GoogLeNet, contudo mostrou melhoria tanto em relação a acurácia quanto em especificidade quando aplicou-se a equalização de histograma (Tabela 2). Nesse caso, a especificidade é superior aos outros valores obtidos pelas outras arquiteturas, contudo, os valores de sensibilidade e acurácia ainda são inferiores aos obtidos pela AlexNet.

Tabela 1. Resultados encontrados quando os tecidos foram treinados e classificados utilizando-se a AlexNet

	<i>Sensibilidade</i>	<i>Especificidade</i>	<i>Acurácia</i>
<i>Histograma Equalizado</i>	0,7024	0,965	0,9159
<i>Realce Logarítmico</i>	0,6886	0,9586	0,9081
<i>Base de Dados Original</i>	0,7993	0,9745	0,9417

Através dos resultados podemos observar que as redes tiveram alta taxa de reconhecimento de regiões normais, indicando um melhor aprendizado das amostras que tinham a maior prevalência na base. Regiões normais são também regiões de fundo bem definida. Regiões de massa, embora mais homogêneas possuem larga variação de forma e textura. As arquiteturas representaram um menor poder de aprendizado de massas, embora ainda significativo.

Tabela 2. Resultados encontrados quando os tecidos foram treinados e classificados utilizando-se a GoogLeNet

	<i>Sensibilidade</i>	<i>Especificidade</i>	<i>Acurácia</i>
<i>Histograma Equalizado</i>	0,5986	0,9705	0,901
<i>Realce Logarítmico</i>	0,2845	0,9299	0,7427
<i>Base de Dados Original</i>	0.6021	0,9315	0,8699

4. Conclusão e Próximos Passos

Neste trabalho foi apresentada uma aplicação de *CNN*, para a classificação de tecidos de mamografia, em massa e normal.

As mamografias são imagens de pouca informação de textura. As massas presentes nesse tipo de imagem apresentam em sua maioria grande semelhança com regiões próximas, tornando o trabalho de detecção dependente do ajuste de inúmeros parâmetros. Pensando nisso, este trabalho apresentou a classificação de massas e regiões normais usando arquiteturas *deep learning* como uma alternativa da aplicação de técnicas de grande parametrização. Arquiteturas deste tipo são recentes e sua grande vantagem consiste no autoaprendizado de características e na robustez de classificação.

A intenção é a construção de uma abordagem altamente adaptável e automatizada na detecção de massas em mamografias. Os resultados apresentados, utilizando-se as arquiteturas AlexNet e GoogLeNet foram promissores embora sejam necessários ajustes para o correto aprendizado de regiões de massa.

Um dos problemas encontrados até o momento consiste no desbalanceamento da base que pretendemos corrigir utilizando a técnica SMOTE [Chawla et al. 2002]. Ainda como continuação, pretendemos ajustar corretamente o erro retropropagado para ajuste de sensibilidade e a construção das camadas de convolução de acordo com textura de baixa informação, como no caso de regiões de mamografia.

5. Agradecimento

Os autores desejam agradecer o apoio financeiro concedido pela FAPEMA e CNPq para a execução desta pesquisa.

Referências

- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pages 321–357.
- Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., and Munnishkumaran, S. (1998). Current status of the digital database for screening mammography. In *Digital mammography*, pages 457–460. Springer.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9.

Diagnóstico de Câncer de Mama Através de Vetores de Descritores Localmente Agregados

Ricardo Costa da Silva Marques¹, Geovane Menezes Ramos Neto¹,
Geraldo Braz Júnior¹, João Dallyson de Sousa Almeida¹

¹Núcleo de Computação Aplicada – Laboratório de Mídias Interativas
Universidade Federal do Maranhão (UFMA) Campus do Bacanga – São Luís – MA

{rikardomarques, geovane, geraldo, dallyson}@nca.ufma.br

Abstract. *An approach to early detect such breast anomalies is the mammography image. However, complex image patterns and the different organization of the breast tissues requires skill and experience by a trained physician to avoid faults in the mammograms interpretation. The main goal of this work is reduces the number of faults associateds to exam. For this, we propose a feature extraction of texture using Vector of Locally Aggregated Descriptors. Initial tests have achieved promising results , the better values obtained with 600 samples of DDSM base are: 90.18 (accuracy), 91.83 (sensitivity) and 94.02 (specificity).*

Resumo. *Uma abordagem para detectar precocemente anomalias na mama é a imagem de mamografia. No entanto, os padrões de imagem complexos e a diferente organização dos tecidos mamários requer habilidade e experiência de médicos treinados para evitar falhas na interpretação de mamografias. O objetivo principal deste trabalho é diminuir esse número de falhas associadas ao exame. Para isso, propomos uma extração de características de textura usando Vetores de Descritores Localmente Agregados. Os testes iniciais alcançaram resultados promissores, os melhores valores obtidos com 600 amostras da base DDSM são: 90.18 (acurácia), 91.83 (sensibilidade) e 94.02 (especificidade).*

1. Introdução

É o tipo de câncer mais comum entre as mulheres no mundo e no Brasil, depois do de pele não melanoma, respondendo por cerca de 25% dos casos novos a cada ano. O câncer de mama é raramente detectado em homens, totaliza apenas 1% do total de casos da doença. Possui estimativa de 57.960 novas ocorrências [INCA 2016].

Por sua vez, a atividade de rastreamento de anomalias através de mamografias, demonstra redução na taxa de mortalidade associada a essa patologia. Entretanto, a sensibilidade desse exame varia diretamente proporcional a qualidade, implicando em falhas.

Nesse raciocínio, baseando-se em algoritmos descritores de pontos chave, a metodologia deste trabalho realiza uma extração de características capaz de prover alto poder discriminatório, ao reconhecer regiões de massa encontradas em ROIs (*Region of Interest*). Tendo como objetivo principal, a classificação dessas regiões em benignas e malignas, fornecendo ao especialista na área uma segunda opinião a respeito do diagnóstico.

2. Metodologia

A metodologia está subdividida em 4 passos principais (Figura-1): aquisição das imagens, pré-processamento com realce logarítmico para evidenciar regiões suspeitas na ROI, a penúltima etapa, faz a detecção de descritores de textura (SURF - Speeded Up Robust Features), a partir desses descritores criamos o dicionário de palavras visuais (BOW+KMeans), resultando na obtenção do vetor de características localmente agregadas (VLAD - Vector of Locally Aggregated Descriptors) e com a ferramenta AutoWeka é feita a classificação.

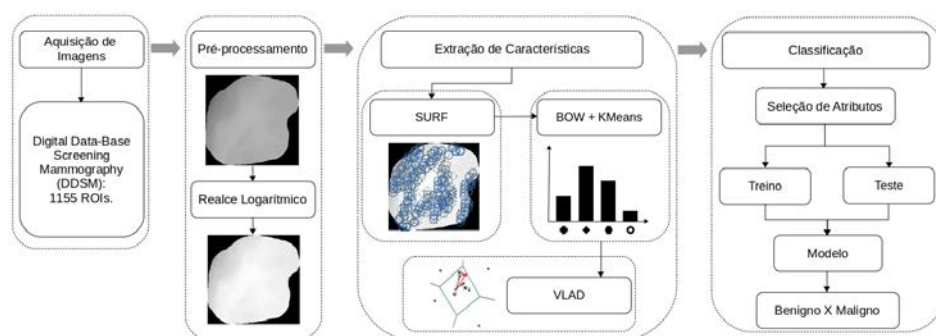


Figura 1. Etapas da metodologia: aquisição das imagens, pré-processamento (Realce Logarítmico), extração de características detectando pontos chave e gerando descritores (SURF), para criar o vocabulário de palavras visuais (BOW+KMeans), gerar o vetor de características localmente agregadas (VLAD) e fazer a classificação.

Para pré-processamento, as imagens são submetidas ao Realce Logarítmico [Gonzalez and Woods 2010]. Esta técnica de realce tem como objetivo evidenciar os valores de pixels mais escuros em tons mais claros de cinza, propiciando uma melhor obtenção dos diferentes níveis de quantização.

Após o realce logarítmico, o algoritmo SURF [Bay et al. 2008] identifica pontos chaves e gera descritores de textura. Por via de experimentos, foi possível notar que o algoritmo se comporta de forma mais otimizada diante do problema em questão. Seus parâmetros foram ajustados para gerar descritores de 128 elementos, com 4 oitavas, 2 pirâmides gaussianas iguais dentro de cada oitava, limitado em 400 pontos chave e computando rotações.

Com os descritores das classes adquiridos é gerado o dicionário de palavras visuais (Figura 2). Nessa perspectiva, o KMEANS distingue dentre milhares de características quais irão definir de forma mais compacta cada ROI, para a criação do dicionário.

A partir de agora, o VLAD é aplicado com a finalidade de criar um vetor de tamanho fixo contendo a representação das palavras visuais de uma ROI. Sua criação é feita atribuindo cada descritor da imagem ao elemento mais similar do vocabulário e calculando o valor da diferença de ambos.

Baseando-se no método Fisher Kernel, juntamente com o BOW para a criação do vocabulário [Jégou et al. 2010]. Lembre do passo anterior que usou o Kmeans para agrupar os descritores em clusters e construir um vocabulário de k itens: $C = [c_1, c_2, \dots, c_k]$. Subtraindo cada descritor x do seu respectivo índice no vocabulário e localizando os vizi-

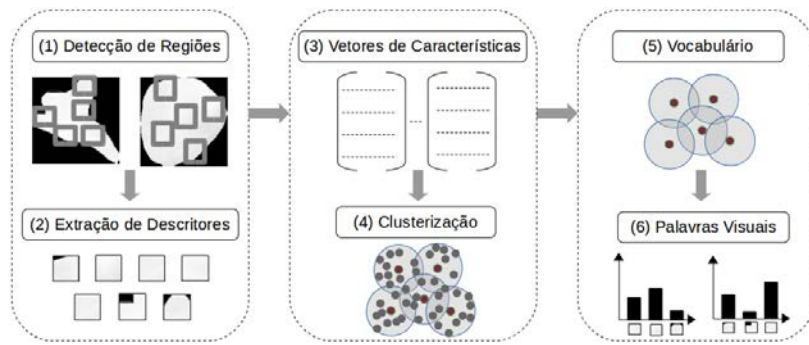


Figura 2. Passos para a obtenção do dicionário de palavras visuais.

nhos mais próximos através do método Nearest Neighbors, foi definido o grau de similaridade.

Portanto, usamos o maior valor c_j para calcular as diferenças. Tal que, c_j e x possuam número de elementos igual, essa diferença é feita objeto a objeto, e armazenada na posição c_j (objeto do vocabulário). Logo, a dimensão D de um VLAD é de $D = (d \times k)$ onde d agrupa todas as características do descritor e k os elementos do vocabulário. Totalizando $k \times d$ características para cada ROI.

Na classificação, utilizamos a ferramenta AUTOWEKA ([Thornton et al. 2013]), para criação de 6 experimentos, que elegeram como seletor de atributos e classificador, respectivamente, os algoritmos Best First e Random Forest.

3. Resultados

A metodologia está sendo validada com o uso de ROIs da base DDSM [Heath et al. 2000] que contém 1155 amostras, sendo 530 benignas e 625 malignas. Para a realização dos testes iniciais a base foi dividida em duas parcelas com 200 e 600 amostras. Em cada etapa foram utilizados três tamanhos de dicionário: 100, 300 e 500. Por fim, foi feita a separação de 60%, 70% e 80% dos experimentos para treino, tendo como resultado a média de 5 rounds executados para cada percentual, como mostram as Figuras 3 e 4.

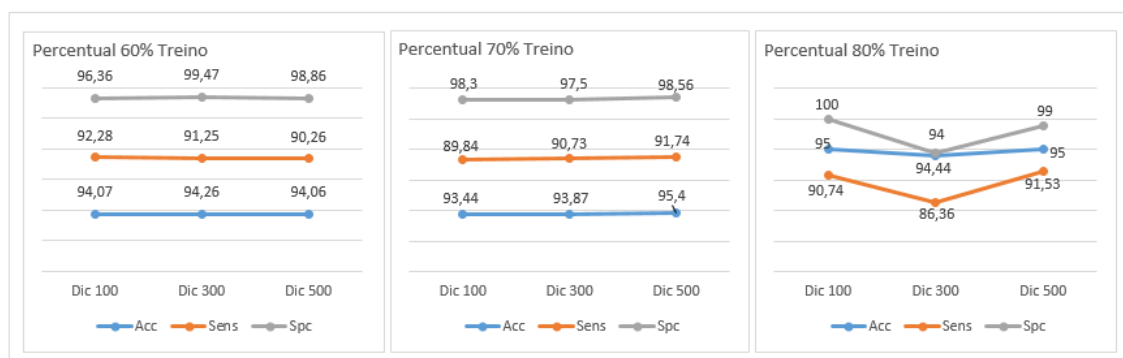


Figura 3. Resultados obtidos para 100 amostras benignas e 100 malignas.

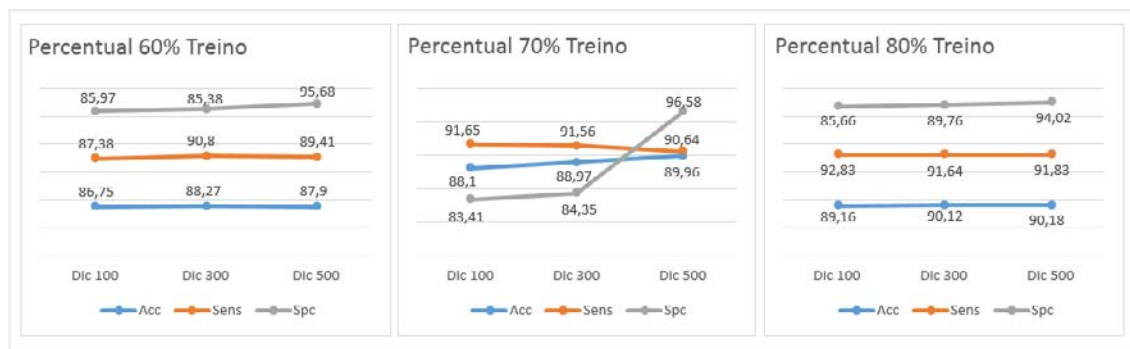


Figura 4. Resultados obtidos para 300 amostras benignas e 300 malignas.

4. Conclusão e Próximos Passos

A metodologia aqui mostrada desempenhou um papel bastante aceitável, demonstrando alto poder discriminatório e grande sensibilidade para discernir a respeito de padrões extraídos de regiões de massa. Para dar continuidade ao trabalho, toda a base será submetida a testes para complementar o estudo e para que tenhamos uma visão mais ampla sobre os resultados obtidos. De modo que, o uso de técnicas baseadas em Bag of Words seja analisado com mais clareza, pois notamos que o tamanho do dicionário influencia diretamente na natureza dos resultados.

5. Agradecimentos

Os autores desejam agradecer o apoio financeiro concedido pela FAPEMA e CNPq para a execução desta pesquisa.

Referências

- Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.
- Gonzalez, R. and Woods, R. (2010). *Processamento Digital de Imagens*. Pearson Prentice Hall, São Paulo, 3 edition.
- Heath, M., Bowyer, K., Kopans, D., and Moore, R. (2000). The digital database for screening mammography. Citeseer.
- INCA (2016). Estimativas 2016: Incidência de Câncer no Brasil. <http://http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/>.
- Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3304–3311. IEEE.
- Thornton, C., Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2013). Auto-weka: Combined selection and hyperparameter optimization of classification algorithms. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 847–855. ACM.

Classificação de Doenças Intersticiais Pulmonares Difusas através de Tomografia Computadorizada de Alta-Resolução

Isadora Cardoso¹, Heitor Ramos¹, Eliana Almeida¹

¹LACCAN/Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Av. Lourival Melo Mota, s/n – Tabuleiro do Martins – CEP 57072-900
Maceió – AL – Brasil

{icps, heitor, eliana.almeida}@ic.ufal.br

Abstract. *The goal of this work is to help the development of a computer-aided diagnosis of lung diseases. In this first stage we used principal component analysis (PCA), linear discriminant analysis (LDA) and k-nearest neighbors algorithm (KNN) to classify 3252 regions of interest (ROI) of High-resolution computed tomography of the chest into 6 lung patterns. From each ROI we extracted 28 features, which were used to evaluate the performance of two dimensionality reduction techniques (PCA and LDA). We further applied KNN ($K = 5$) to classify the ROI into the correspondent lung pattern. We achieved 80,82% correct classification rate using 13 dimensions with PCA and 83,74% using 5 dimensions with LDA.*

Resumo. *O objetivo deste trabalho é auxiliar no desenvolvimento de uma ferramenta de diagnóstico de doenças pulmonares auxiliado por computador. Nessa primeira etapa utilizamos análise de componentes principais (PCA), análise do discriminante linear (LDA) e o algoritmo de k-vizinhos mais próximos (KNN) para classificar 3252 regiões de interesse (ROI) de Tomografias Computadorizadas de Alta-Resolução de tórax em relação à 6 padrões pulmonares. Cada ROI possui um total de 28 dimensões que foram reduzidas por PCA e LDA e então classificadas por KNN ($k = 5$). Obtivemos uma taxa de classificação correta de 80,82% em 13 dimensões com PCA e 83,74% em 5 dimensões com LDA.*

1. Introdução

Doenças pulmonares intersticiais difusas (DPIDs) são patologias que causam disfunção respiratória. Combinando dados clínicos com dados tomográficos, é possível um alto grau de acurácia no diagnóstico correto. Dentre as técnicas de imagens radiológicas, principalmente na avaliação de DPIDs, destaca-se a Tomografia Computadorizada de Alta-Resolução (TCAR), uma vez que melhora significativamente a sensibilidade e a especificidade do diagnóstico clínico [Elicker et al. 2008], além de reduzir a exposição da estrutura torácica [Pereyra et al. 2014].

De uma base de dados com 3252 regiões de interesse (ROI) obtidas de TCAR do tórax de pacientes, [Pereyra et al. 2014] extraíram 28 atributos relativos à textura para classificar seis padrões pulmonares, a saber: pneumonia (PC), áreas enfisematosas (AE), espessamento de septo (ESI), favo de mel (FM), opacidade em vidro fosco (OVF) e tecido pulmonar saudável (TS). Após, utilizaram o algoritmo k-vizinhos mais próximos (*k-nearest neighbor* - KNN) ($k = 5$) e obtiveram acurácia média de 80%. Com a mesma

base de dados, [Almeida et al. 2015b] utilizaram o modelo estatístico de mistura de Gaussianas (MMG) e obtiveram uma classificação correta mínima de 60%. Em sequência, [Almeida et al. 2015a] aplicaram MMG nos cinco atributos mais significativos para obter funções de pertinência *fuzzy* e obtiveram uma média de 63% de classificação correta.

Ainda com esse conjunto de dados, o presente projeto tem a intenção de ajudar no desenvolvimento de um sistema de diagnóstico auxiliado por computador (CAD). Nossa primeira etapa é a identificação de técnicas que possam ajudar na redução de dimensionalidade, pois, com todas as 28 dimensões extraídas, é possível ter um *overfitting* e a chamada maldição da dimensionalidade, que gera erros devido à distância em que os pontos se encontram espalhados nas dimensões, gerando uma escassez (pode haver escassez de pontos em algumas partes). Embora reduzindo, é importante atentar para não perder informações importantes presentes nos dados [Tan et al. 2005]. Para isso, utilizamos análise de componentes principais (PCA) e análise do discriminante linear (LDA), por serem duas técnicas já consolidadas e muito utilizadas em vários aspectos de imagens médicas.

2. Materiais e métodos

Foram extraídas de 3252 ROI um conjunto de 28 atributos, a saber: estatísticas de primeira ordem como média, mediana, desvio padrão, assimetria e curtose; 14 atributos de textura de [Haralick et al. 1973] (energia, momento da diferença, correlação, variância, momento da diferença inversa, soma dos quadrados: média, soma dos quadrados: variância, soma dos quadrados: entropia, momento da variância, momento da entropia, medida de informação de correlação 1, medida de informação de correlação 2 e máximo coeficiente de correlação); Medidas de energia de textura de Laws (*Laws' wave measures (rotation-invariant)*, *Laws' ripple measures (rotation-invariant)*, e *Laws' level measures*); medidas estatísticas da densidade spectral de potência (DSP) (média, desvio padrão e mediana); e dimensão fractal. Detalhes desses atributos estão disponíveis em [Rangayyan 2004].

Utilizamos estes atributos e aplicamos as técnicas PCA e LDA. Em sequência, utilizamos um classificador KNN, com $k = 5$. Para validar, utilizamos o método *holdout* de validação cruzada, com cerca de 80% das 3252 ROIs para treino do modelo e cerca de 20% para testes, comparando-a à classificação dos médicos especialistas feitas anteriormente. Em todo o procedimento foi utilizado a linguagem R (versão 3.2.2)¹.

2.1. Análise de Componentes Principais

PCA é uma técnica não supervisionada que, em um conjunto de d dimensões, procura uma transformação de menor dimensão que mantenha o máximo possível da variância dos dados [James et al. 2013].

2.2. Análise do Discriminante Linear

Ao lidar com dimensões paralelas, PCA escolherá uma dimensão que junte todas, o que, claramente, levará a erros de classificação. Nessas situações, LDA é uma melhor escolha. LDA é uma técnica supervisionada que deseja encontrar o hiperplano que maximize a separação entre as médias de cada classe e minimize a dispersão total das mesmas, o que evita sobreposições que podem piorar a separação [Zaki and Wagner Meira 2014].

¹<http://www.r-project.org>

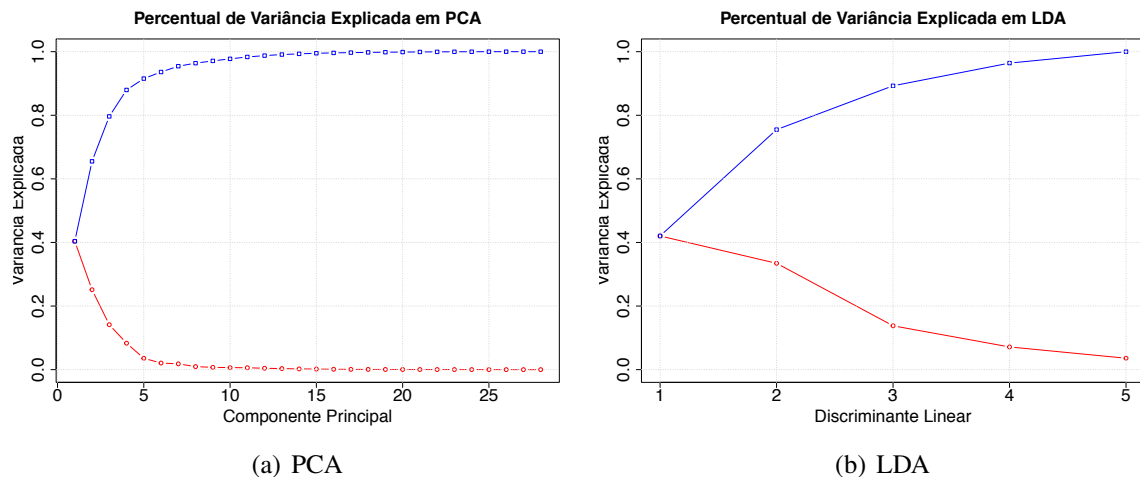


Figura 1. Quantidade de variância explicada através das 1(a) componentes principais e 1(b) discriminante linear ideais.

2.3. K-vizinhos mais próximos

Um classificador k-vizinhos mais próximos analisa a localização de um ponto no espaço e determina, por seus k vizinhos, qual sua classificação. Isso é feito medindo-se a distância (ex: Euclidiana) do ponto pros seus vizinhos [Tan et al. 2005]. Por simplicidade e para facilitar na comparação das técnicas de redução de dimensionalidade aqui tratadas, utilizamos o KNN para classificar os padrões pulmonares.

3. Resultados

Na figura 1, a linha vermelha representa o quanto de variância cada componente principal e discriminante linear ideais explicam em relação aos dados. A linha azul representa o acumulado de variância que representam ao total. É possível notar pela figura 1 que com cerca de 13 componentes principais já se pode explicar quase 100% da variância dos dados. LDA consegue uma redução ainda maior, 5 dimensões. O lado negativo dessas transformações é que temos que utilizar todo o conjunto de dados para obtê-la.

Na tabela 1 podemos ver a comparação do resultados aplicando as técnicas PCA e LDA e dos dados sem técnica alguma de redução de dimensionalidade, classificados pelo KNN. Na tabela, TP (*true positive*) representa a classificação correta, enquanto FP (*false positive*) representa o erro do tipo I. O Espessamento do septo (ESI) é a doença mais difícil de classificar, obtendo um número mais baixo de TP em todas as técnicas. A classificação dos dados originais e dos obtidos pelas técnicas são próximas, o que mostra quão bom são os conjuntos conseguidos. Em negrito, estão destacados os melhores valores de TP e FP para cada classe. Observe que a técnica LDA apresenta sistematicamente melhores valores quando comparada com PCA ou com os dados originais, ficando pouco abaixo em apenas alguns poucos casos.

4. Conclusão

Neste trabalho utilizou-se PCA e LDA para redução de dimensionalidade. Obtivemos uma taxa de classificação correta de 81,13% com PCA (13 dimensões) e 84,38% com LDA (5 dimensões). Ambas são semelhantes ou superiores à taxa de acerto do conjunto

	Original		PCA		LDA	
	TP	FP	TP	FP	TP	FP
PC	0.8686869	0.02355072	0.7979798	0.03558719	0.8484848	0.02669039
AE	0.8111111	0.03014184	0.8111111	0.03019538	0.8777778	0.01950355
ESI	0.6796875	0.07663551	0.6718750	0.07865169	0.6796875	0.07706767
FM	0.8240741	0.03531599	0.8333333	0.03358209	0.8333333	0.03321033
TS	0.8596491	0.03018868	0.8596491	0.03024575	0.9122807	0.01893939
OVF	0.8672566	0.02772643	0.8938053	0.02238806	0.9115044	0.01879699
Total	0.8184109	0.03725986	0.8112923	0.03844169	0.8438448	0.03236805

Tabela 1. Resultado das classificações do conjunto de dados original, PCA e LDA, com KNN (k = 5)

original, o que cumpre o objetivo. Trabalhos futuros visam utilizar técnicas que explicitem quais atributos são os mais importantes, descartando os que não contribuem tanto. Também visa-se a utilização de mais classificadores a fim de obter melhores resultados.

Referências

- Almeida, E., Rangayyan, R. M., and Azevedo-Marques, P. M. (2015a). Fuzzy membership functions for analysis of high-resolution CT images of diffuse pulmonary diseases. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2015, Milan, Italy, August 25-29, 2015*, pages 719–722.
- Almeida, E., Rangayyan, R. M., and Azevedo-Marques, P. M. (2015b). Gaussian mixture modeling for statistical analysis of features of high-resolution CT images of diffuse pulmonary diseases. In *2015 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2015, Torino, Italy, May 7-9, 2015*, pages 1–5.
- Elicker, B., de Castro Pereira, C. A., Webb, R., and Leslie, K. O. (2008). Padrões tomográficos das doenças intersticiais pulmonares difusas com correlação clínica e patológica. In *Jornal Brasileiro de Pneumologia, Volume 34*, number 9, pages 715–744.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610–621.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer.
- Pereyra, L. C., Rangayyan, R. M., Ponciano-Silva, M., and de Azevedo Marques, P. M. (2014). Fractal analysis for computer-aided diagnosis of diffuse pulmonary diseases in HRCT images. In *2014 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2014, Lisboa, Portugal, June 11-12, 2014*, pages 455–460.
- Rangayyan, R. (2004). *Biomedical Image Analysis*. Biomedical Engineering. CRC Press.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Zaki, M. J. and Wagner Meira, J. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.

Detecção de Anomalias em Termografia Dinâmica das Mamas*

Jessica P. S. Cardoso¹, Aristófanés C. Silva¹, Anselmo C. Paiva², Stelmo M. B. Netto¹,

Aura Conci²

¹Núcleo de Computação Aplicada – Universidade Federal do Maranhão (UFMA)
Av. dos Portugueses, 1966 – Bacanga, São Luís – MA, 65080-805

²Instituto de Computação
Universidade Federal Fluminense (UFF) – Niterói, RJ – Brasil

{anselmo.c.paiva, palomaousa15, snobnet}@gmail.com

ari@dee.ufma.br, aconci@ic.uff.br

Abstract. *Thermography is a test that can be used for early diagnosis of breast cancer. Regions with abnormalities usually present disturbances in the gain temperature. Based on this idea, this paper aims to analyse multiples temporal series in order to identify behaviour patterns among healthy and non-healthy regions. The method was able to correctly identify 10 of 13 exams with anomalies, pointing that the detection method based on amplitude can be promising on the identification of suspicious regions with anomalies.*

Resumo. *A termografia é um exame que pode ser utilizada para diagnóstico precoce de câncer de mama. Regiões com anomalias geralmente apresentam perturbações no ganho de temperatura. Com base nessa ideia, o presente trabalho propõe uma metodologia de análise de séries temporais a fim de investigar esse padrão comportamental. O método foi capaz de identificar corretamente 10 dos 13 exames que possuíam anomalias, configurando assim que o método de detecção baseado em amplitude pode ser promissor na identificação de regiões suspeitas de lesão.*

1. Introdução

O câncer de mama é o mais comum entre as mulheres no mundo, representando cerca 25% de todos os tipos de cânceres diagnosticados no sexo feminino de acordo com o Instituto Nacional do Câncer [INCA 2012]. Quando detectado em estágios iniciais, amplia-se as chances de tratamento e cura da doença. Além da mamografia, existem outros exames de imagens, tais como a ultrassonografia, ressonância e a termografia que servem de auxílio à mamografia. O exame de termografia consiste em um tipo de técnica não invasiva que produz uma imagem através da radiação infravermelha emitida pela superfície corpórea.

Segundo Lavinsky Wolff et al.[2012], uma região onde se encontra tumor geralmente apresenta maior temperatura, pelo processo de angiogênese, se comparada ao restante da mama. Silva[2015] observou que pacientes doentes apresentavam um grupo com temperaturas e inclinação maiores nos momentos iniciais se comparada a pacientes saudáveis.

*Projetos aprovados no Comitê de Ética, registrados da Plataforma Brasil do Ministério da Saúde para pesquisa sob o CAE 01042812.0.0000.5243, aprovado pelo parecer número 31646.

Um método que se mostrou bem promissor para tal finalidade foi proposto por Izakian e Pedryez[2013]. Tal método é interessante, pois com ele é possível identificar a presença de anomalias dentro de um sinal a partir da análise de suas subsequências. Séries temporais com comportamento irregular seriam identificadas como anômalas comparadas a tecidos saudáveis. O objetivo deste artigo é apresentar uma metodologia de detecção de anomalias em termografia dinâmica das mamas utilizando conceitos de séries temporais a fim de verificar padrões.

2. Metodologia

Esta seção apresenta alguns dos procedimentos utilizados no desenvolvimento desse trabalho a fim de alcançar os objetivos esperados pelo mesmo, como o processo de aquisição das imagens térmicas, e seu processo de aquisição, a técnica de registro de imagens e os métodos de anomalias em séries utilizados neste trabalho.

2.1. Aquisição

Os exames utilizados no processo de produção desse artigo provém da *Database for Mastology Research with Infrared Image (DMR-IR)*, mesma base utilizada na tese de Silva[2015]. As imagens foram adquiridas de pacientes no Hospital Universitário Antônio Pedro da Universidade Federal Fluminense (HUAPE-UFF), constituindo 70 exames, sendo 35 saudáveis e 35 não saudáveis. No presente artigo foram utilizados apenas os exames não saudáveis a fim de testar a eficácia da técnica em distinguir regiões que possuem alterações do restante. Devido ao número reduzido de exames que possuem a localização das lesões, somente 13 exames, do total de 35 exames não saudáveis, foram utilizados neste trabalho.

O processo de aquisição das imagens é feito através do resfriamento das mamas fazendo uso de um ventilador elétrico. Este resfriamento é realizado sobre a região torácica até atingir a temperatura de $30,5^{\circ}$ ou após 5 minutos do início do mesmo, o equipamento é desligado, iniciando-se o processo de captura das imagens. O intervalo de captura das imagens ocorre a cada 15 segundos.

2.2. Registro

A aquisição das imagens pode sofrer com interferências do meio, como pequenas movimentações das pacientes, que dificultariam no processo de construção das séries temporais. Desta forma, a interpretação de uma certa posição no espaço no decorrer do tempo não seria possível. A fim de evitar esse problema, é aplicada uma técnica chamada de registro que efetua correções geométricas com base em um sistema de referência[INPE 2016].

O registro utilizado neste artigo é o deformável, proposto por Silva[2015], no qual escolhe o primeiro termograma da sequência como referência e todos os demais são registrados com base neste.

2.3. Construção das séries

Uma série temporal representa um conjunto de observações sequencialmente feitas ao longo do tempo. Para cada pixel no ponto $P(x, y)$ da imagem (Figura 1), é construída

uma série temporal gerando a série $P(x, y) = t(1), t(2), t(3), \dots, t(20)$, com $t(1)$ representando a temperatura no ponto $P(x, y)$ no tempo inicial.

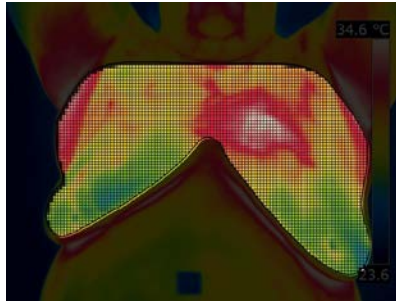


Figura 1. Região de interesse

A Figura 2 mostra as séries temporais de uma paciente saudável e uma paciente com carcinoma. A temperatura está indicada no eixo z , o eixo y indica cada série temporal e o x representa o tempo de 1 a 5. A partir dessas imagens é perceptível que as séries da paciente saudável (Figura 2(a)) apresentam um comportamento homogêneo ao decorrer do tempo se comparada a paciente doente (Figura 2(b)).

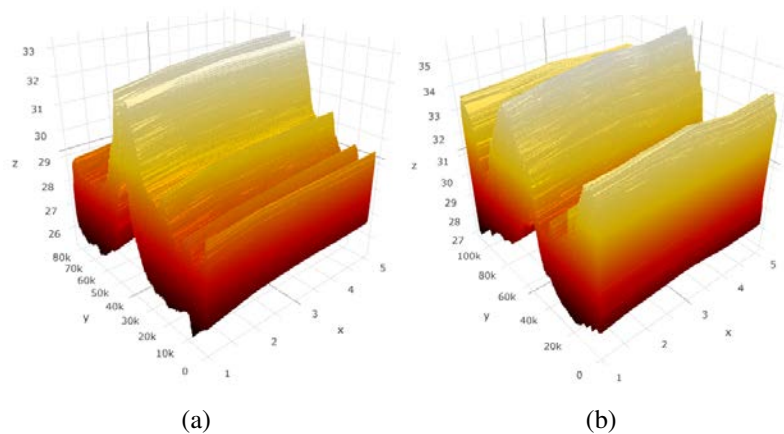


Figura 2. Séries Temporais

2.4. Detecção de anomalias usando fuzzy c-means

A metodologia proposta no presente artigo consiste em extrair N séries temporais a partir de um exame contendo 20 termogramas. A partir das séries encontradas, detecta-se regiões anômalas em amplitude de temperatura.

Seja $X = \{x_1, x_2, x_3, \dots, x_{20}\}$ o conjunto de séries temporais x de tamanho 20. Sobre esse conjunto de dados é aplicado o método de clusterização Fuzzy C-Means, proposto por Bezdek[1984], no qual é calculada uma pseudo partição fuzzy. A partir dessa pseudo partição gerada, é reconstruída um novo conjunto de séries.

Segundo Izakian e Pedryez[2013] uma subsequência normal dos dados é similar a uma ou mais partições ou semelhante a uma combinação de partições e, quanto maior a similaridade entre a sequência e os protótipos, menor é a anomalia associada a ela.

Após o processo de reconstrução das sequências, é calculado o erro de reconstrução, que consiste em uma distância euclidiana ao quadrado entre o conjunto de séries originais e as séries reconstruídas, que mede o quanto uma sequência é similar a seu protótipo. Desta forma, as séries de temperaturas que apresentam comportamento anormal possuem suas pontuações de anomalias altas se comparada as demais séries do conjunto.

3. Resultados

Das 13 imagens que possuíam alguma alteração, 3 pacientes não tiveram a região da lesão marcadas pelo método proposto. Os parâmetros usados no Fuzzy c-means foram número de grupos e índice fuzzy iguais a 2, representando as séries normais e séries anômalas. Na Figura 3, a região em branco representa o conjunto de séries que apresentou maior variação em amplitude no decorrer do tempo.

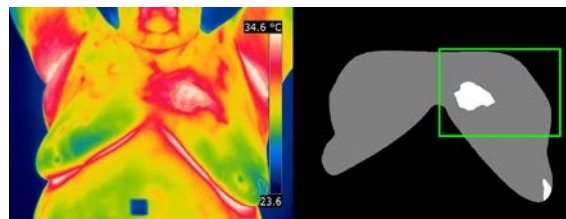


Figura 3. Região detectada pela metodologia proposta

Referências

- Bezdek, J. C., Ehrlich, R., and Full, W. (1984). Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2):191–203.
- INCA (2012). Controle do câncer de mama conceito e magnitude. http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_controle_cancer_mama/conceito_magnitude. Accessed: 2014-09-20.
- INPE (1991-2016). SPRING: Tutorial de Geoprocessamento aula 3 – registro de imagem. <http://www.dpi.inpe.br/spring/teoria/>. Accessed: 2014-09-20.
- Izakian, H. and Pedrycz, W. (2013). Anomaly detection in time series data using a fuzzy c-means clustering. In *IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint*, pages 1513–1518. IEEE.
- LavinskyWolff, M., Rohde, L., Polanczyk, C. A., et al. (2012). Avaliação de Tecnologias em Saúde sumário das evidências e recomendações sobre o uso da termografia no diagnóstico de câncer de mama. <http://www.unimed.coop.br/pct/servlet/ServletDownload?id=MjE0NzUxMTE3NA>. Accessed: 2014-09-19.
- Silva, L. F. (2015). *Uma Análise Híbrida para Detecção de Anomalias da Mama usando Séries Temporais de Temperatura*. PhD thesis, Universidade Federal Fluminense.

The use of Artificial Neural Network for lipid and glycaemic profiles quantification through infrared spectroscopy

Henrique Hesse¹, Rejane Frozza¹, Valeriano Corbellini², Cézane Reuter², Miria Burgos²

¹Mestrado em Sistemas e Processos Industriais – Universidade de Santa Cruz do Sul (UNISC) – Santa Cruz do Sul, RS – Brazil

²Mestrado em Promoção a Saúde – Universidade de Santa Cruz do Sul (UNISC) – Santa Cruz do Sul, RS - Brazil

henriquehesse@me.com, frozza@unisc.br, valer@unisc.br,
cpreuter@hotmail.com, mburgos@unisc.br

Abstract. *This paper aims to look at the viability of the use of artificial neural networks to solve nonlinear correlations between infrared spectra and biochemical quantification tests, to build a computational system to predict the levels of glycaemic and lipid profiles using infrared spectroscopy. The studies of one of the parameters was modelled and showed signs of viability to quantify all parameters with the suggested methodology. Therefore, more complex and larger data sets are going to be tested with this technique.*

1. Introduction

The monitoring of the glycaemic and lipid profiles are important, since the human vascular system plays a variety of roles, mainly providing nutrients and oxygen to tissues and cells and removing waste from the human system. Basically, each blood component can be quantified, through biochemical analysis. Thresholds are useful to diagnostics and to monitor the subjects' health. Despite the fact that there is equipment to do such analysis, they still rely on the usage of chemicals and they need at least 10 mL of blood to perform glycaemic and lipid analysis [Low-Yinga et al., 2002].

Infrared spectroscopy is a technique that can be used for most organic liquids, solids and gases. Providing information related to the atoms and its bonds, revealing what type of organic function it is and indirectly its quantification [Barbosa, 2007]. This method does not require any chemicals to be executed and the sample's volume is about 5 μ L. Both of these factors are imperative when another factor is introduced to the equation of analysis routines, the analysis time [Irudayaraj, 2002]. For infrared technique the analysis time is less than a minute, covering all substances, while the standard biochemical quantification has to be done for each parameter one at the time, with specific reactions [Jessen et al., 2014; Song, Lee e Kim, 2015].

One infrared spectrum from 5 μ L of blood provides a great amount of information, since each wavelength from the spectrum is a variable. The mid infrared works between 4,000 and 400 cm^{-1} , therefore every spectrum generates 3,500 variables [Mohd et al., 2015]. Normally the quantification using infrared data is done through multivariate analysis, providing good results for linear variables. The blood spectra are not linear to the amount of cholesterol, HDL (high-density lipoprotein), LDL (low-density

lipoprotein), triglycerides and glucose. Hence the non-linear information requires more advanced set of tools for its quantification [Yadav et al., 2015; Shahin, 2014].

An artificial neural network (ANN) is capable to resolve non-linear problems due its different functions. The ANN is built over a group of sample that the values for each parameter of the glycaemic and lipid profiles were already measured. Once the network topology was established, the neural network is trained to reach the minimum error associated with the real values. After the training the ANN can ready new samples and quantify all parameter at once and absorb the information of this new sample to increase its precision [Lima, Pinheiro & Santos, 2014; Piotrowski *et al.*, 2015].

2. Experimental

For the analyses were collected 92 blood samples from subjects in Santa Cruz do Sul, Rio Grande do Sul, Brazil, with open consent and well informed about the research, which is registered in the ethics committee in research with human beings of Universidade de Santa Cruz do Sul (reg. number 469916).

The mid-infrared spectra of the total peripheral blood samples were analysed fresh or lyophilized and acquired in real triplicate using diffuse reflectance accessory (DRIFTS) using a spectrophotometer Perkin Elmer (Spectrum 400) with resolution of 4 cm^{-1} and 16 scans.

For the first trial was chosen the total cholesterol, due its importance. The total cholesterol gives a general idea of the ratio between the HDL and LDL. In fact most of the quantification of the LDL is done by the equation of Friedewald, which uses the amount of total cholesterol, HDL and triglyceride to calculate the LDL amount.

The infrared spectrum of the cholesterol in blood samples are hard to distinguish, because of the densitiy variations are very distinct and not very perceptible. The others parameters from the profiles are more easily correlated to certain organic functional groups.

It was used a matrix with size of 95×2331 and a vector of 95×1 in a feed-forward backpropagation ANN type, configured with four layers containing 10, 5, 20 and 1 neurons respectively containing 94 input attributes (due the dimensionality reduction using principal component analysis algorithm, PCA) and 1 output (total cholesterol) (Figure 1). The MATLAB 2010 (Mathworks Inc., Natick, MA) software was used for the implementation of PCA and RNA.

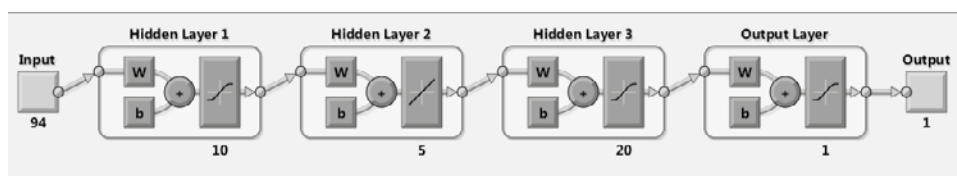


Figure 1. Topology for the ANN.

3. Preliminary results

The initial results of ANN showed a good response to nonlinearity of the data set, achieving an overall correlation coefficient (r) of 0.80985 for this preliminary model and the root-mean-square error (RMSE) of 30.14, (Figure 2).

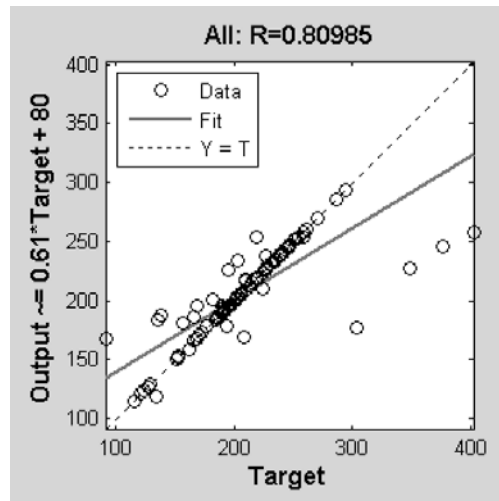


Figure 2. ANN overall regression, showing the correlation between all samples

As can be seen on Figure 2, some samples with higher concentrations of total cholesterol are far from the trending line. Usually those samples would be considered outliers, although it was decided to keep all the samples. Based on the fact that more samples are going to be added to this ANN, making higher the chances to the weight of those samples to be recalculated.

The results for all stages of the ANN are shown in Figure 3. It can be noticed that in all steps as in, training, validation and test the outputs were close enough to the target values to proceed with this research and enlarge the training data set to cover all the cholesterol levels and rise the correlation coefficient to above 95% of correlation between outputs and targets.

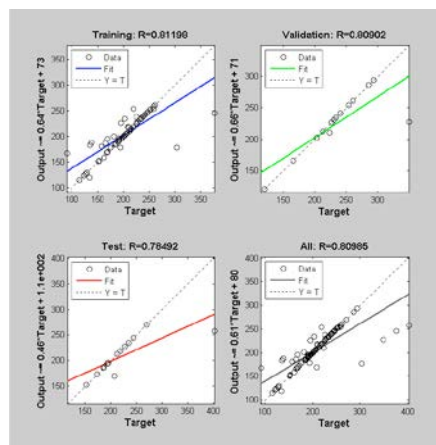


Figure 3. ANN regression, displaying all stages of the network usage

4. Conclusion

The first study of one of the components from the glycaemic and lipid profiles, showed that the artificial neural network was capable to solve the nonlinearity concerning the infrared spectra and the biochemical levels. From this stage of the research, it is the intention to build a full neural network with the entire components from the profiles with a larger data set (500 samples). A larger group of samples will provide more information for every level of each parameter (e.g. total cholesterol, HDL, LDL, triglycerides and glucose).

References

- Barbosa, Luiz Cláudio de Almeida. *Espectroscopia no infravermelho na caracterização de compostos orgânicos*. Viçosa: Universidade Federal de Viçosa, 2007.
- Irudayaraj, J. (2002). Rapid determination of vitamin C by NIR, MIR and FT-Raman techniques. *J. Pharm. Pharmacol.*, London, v.54, n.9, p.1247-1255, 2002.
- Jessen, T. E., Hoskuldsson, A. T., Bjerrum, P. J., Verder, H., Sorensen, L., Bratholm, P. S., . . . Jensen, M. A. (2014). Simultaneous determination of glucose, triglycerides, urea, cholesterol, albumin and total protein in human plasma by Fourier transform infrared spectroscopy: direct clinical biochemistry without reagents. *Clin Biochem*, 47(13-14), 1306-1312. doi: 10.1016/j.clinbiochem.2014.05.064
- Karmowski, K. A. S., Jacek Majda, Piotr Rubisz, Stanisław Han e Andrzej Karmowski. (2014). Lipid index changes in the blood serum of patients with hyperplastic and early neoplastic lesions in the ovaries. *Journal of Ovarian Research*, 7(1), 1-4.
- LIMA, ISAÍAS. PINHEIRO, A. M. CARLOS, SANTOS, A. O. FLÁVIA. *Inteliência artificial*. 1. ed. Rio de Janeiro: Elsevier, 2014.
- Low-Yinga, S., R. Anthony Shawa, Michael Leroux, Henry H. Mantscha. (2002). Quantitation of glucose and urea in whole blood by mid-infrared spectroscopy of dry films. *Vibrational Spectroscopy*, 28, 111-116.
- Mohd Ali, J., Hussain, M. A., Tade, M. O., & Zhang, J. (2015). Artificial Intelligence techniques applied as estimator in chemical process systems – A literature survey. *Expert Systems with Applications*. doi: 10.1016/j.eswa.2015.03.023
- Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., & Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology*, 529, 302-315.
- Song, S. Y., Lee, Y. K., & Kim, I. J. (2016). Sugar and acid content of Citrus prediction modeling using FT-IR fingerprinting in combination with multivariate statistical analysis. *Food Chem*, 190, 1027-1032. doi: 10.1016/j.foodchem.2015.06.068
- Shahin, M. A. (2014). State-of-the-art review of some artificial intelligence applications in pile foundations. *Geoscience Frontiers*. doi: 10.1016/j.gsf.2014.10.002
- Yadav, J., Rani, A., Singh, V., & Murari, B. M. (2015). Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy. *Biomedical Signal Processing and Control*, 18, 214-227.

Apreensibilidade e Qualidade da Informação: Bases de uma Avaliação Textual Automática na Área da Saúde

Asdrubal Falavigna, Carine G. Webber, Fernando A. Abel, Marco A. Koff, Maurício K. Santos, Natália Lisboa

Universidade de Caxias do Sul
Rua Francisco Getúlio Vargas, 1130 - CEP 95070-560
Caxias do Sul (RS) - Brasil

cgwebber@ucs.br

Abstract. *The internet is usually employed to research health-related information. However, available texts lack proper evaluation. In this context, a systematic review of literature on the online evaluation of texts on spinal pathologies has been performed. Selected articles evaluate the quality and the readability of the texts by means of metrics widely accepted in the scientific community, which serve as a basic source to develop a hybrid software for automatic text analysis, based on Artificial Intelligence techniques.*

Resumo. *A Internet é comumente utilizada para pesquisa de informações na área da saúde. Contudo, os textos disponíveis carecem de devidas avaliações. Neste escopo, realizou-se uma revisão sistemática da literatura referente à formas de avaliação de textos sobre patologias da coluna vertebral. Selecionou-se artigos que avaliam a qualidade e a apreensibilidade (ou readability) dos textos por meio de métricas difundidas na comunidade científica para construção de um software para automação da análise textual, em uma abordagem híbrida que empregue técnicas de Inteligência Artificial.*

1. Introdução

Variadas fontes de informação são observadas com a expansão da Internet, nem sempre abordando um tema de forma clara e correta. Em certas áreas, sobretudo naquelas relacionadas à Saúde, é muito importante que existam procedimentos para avaliar informações textuais com relação a critérios como a apreensibilidade (ou *readability*) e a qualidade da informação.

A apreensibilidade da informação textual foi definida por Klare (1963) como sendo a facilidade do entendimento ou compreensão de um texto a partir do estilo da escrita utilizada. Desde então, diversos métodos foram criados a fim de avaliar a apreensibilidade de um texto. Tais métodos vêm sendo, em grande parte, desenvolvidos e aplicados para textos em língua inglesa. A avaliação da apreensibilidade corresponde a uma análise quantitativa da informação (contagem de frases, palavras e sílabas de um texto conforme regras gramaticais), que pode produzir resultados numéricos. É escasso o estudo da aplicabilidade de métodos quantitativos para a língua portuguesa.

A qualidade pode também ser verificada por meio de técnicas de Inteligência Artificial para análise textual (aprendizagem automática e processamento de linguagem natural), extraíndo-se padrões, agrupando e classificando textos [Luger, 2009]. As análises textuais qualitativas em domínios restritos produzem resultados coerentes e úteis [Weiss, 2010; Feldman e Sanger, 2006].

O acesso facilitado de pessoas e pacientes a fontes de informação sobre saúde ampliou a necessidade de que tais informações disponíveis sejam revisadas e analisadas, principalmente em contextos onde o paciente deve participar da decisão do seu tratamento. Nesta problemática, já foi observado em uma amostra populacional que as informações disponíveis não foram compreensíveis a usuários não especialistas, ocasionando indefinições e insegurança no andamento dos respectivos tratamentos. O objetivo deste artigo é promover uma revisão sistemática sobre os métodos e modelos de avaliação de conteúdos da internet sobre doenças da coluna vertebral, servindo como base para o desenvolvimento de um software de análise textual.

2. Materiais e Métodos

2.1 Fase 1: Revisão Sistemática como base científica para desenvolvimento de inovação

Foi realizada uma revisão sistemática da literatura com o objetivo de avaliar a qualidade e apreensibilidade das informações disponíveis na internet sobre doenças da coluna vertebral. A metodologia utilizada foi de pesquisa no banco de dados do PubMed utilizando a seguinte lista de palavras-chave: “education”, “readability”, “evaluation”, “back pain”, “spinal diseases” e “internet”.

Os critérios de seleção dos artigos foram: (1) estudos que continham avaliação da informação em sites; (2) avaliação de sites com doenças da coluna vertebral; (3) publicações em inglês; (4) publicação no período compreendido entre Janeiro de 2000 e Setembro de 2015. Os critérios de exclusão foram: (1) avaliação de um único site; (2) sites com informação contida em vídeos e (3) sites não disponíveis ao paciente.

2.2 Fase 2: Desenvolvimento do Software para Avaliação Textual Automática na Área da Saúde

Para solucionar o problema da avaliação textual na área de saúde na Língua Portuguesa propõe-se o desenvolvimento de um software que empregue técnicas quantitativas e qualitativas, integrando técnicas de Inteligência Artificial.

3. Resultados

3.1 Fase 1: Revisão Sistemática como base científica para desenvolvimento de inovação

O término do levantamento bibliográfico foi na data de 05/10/2015. Foram selecionados 167 artigos, dos quais 104 não estavam relacionados a patologias, sintomas ou cirurgias de coluna. Os resumos foram analisados em 63 artigos, sendo a avaliação do texto completo realizada em 24 artigos. Os 23 artigos foram incluídos no estudo.

As principais ferramentas utilizadas para avaliação da qualidade da informação e apreensibilidade dos sites foram o Flesch-Kincaid Grade Level, o DISCERN tool e o Flesch Reading Ease (Tabela 1).

Tabela 1. Ferramentas Utilizadas na Análise da Informação

Modalidade de Avaliação	Ferramenta de Análise	Número de Artigos Presentes
Avaliação da Qualidade	DISCERN tool/instrument + 3 revisores	5
	JAMA Benchmark Criteria	3
Avaliação da Apreensibilidade	Flesch Reading Ease	4
	Flesch-Kincaid Grade Level	7
	SMOG Grading	2
	Coleman-Liau Index	1
	FORCAST Formula	1
	New Dale-Chall Readability Formula	1
	Gunning Fog Index	3
	New Fog Count	1
	Fry Graphical Analysis	1
	Raygor Readability Estimate	1

Os doze diferentes programas de análise de textos utilizados na avaliação da informação dos sites são em Língua Inglesa. Posteriormente foi verificada a possibilidade de utilizar os programas de análise de texto da Língua Inglesa para aplicação ou adaptação em Língua Portuguesa e de processamento por um computador (Tabela 2).

Tabela 2. Estudo das ferramentas encontradas

Ferramenta de Análise	Aplicação para o português?	Computável?	Considerações
DISCERN Instrument	Sim	Não	Mecanismo realizado manualmente por avaliadores, utilizando-se 16 perguntas sobre o tema.
JAMA Benchmark Criteria	Sim	Não	Sem considerações.
Flesch Reading Ease	Sim	Sim	Métrica é calculada usando o número médio de sílabas por palavra e comprimento médio da frase. Já existe uma adaptação para a língua espanhola.
Flesch Kincaid Grade Level	Sim	Sim	Semelhante ao Flesch Reading Ease. Utilizado para medir o nível de escolaridade de pessoas nos EUA.
SMOG Grading	Não	Sim	Requer um mínimo de 30 frases para aplicação, menos que isso resulta em resultados estatisticamente inválidos. Existem avaliações de textos farmacêuticos de Portugal, no entanto os textos tiveram de ser traduzidos em inglês.
Coleman-Liau Index	Sim	Sim	Métrica calcula o nível de ensino baseado apenas nas médias dos comprimentos das sentenças e a média do número de caracteres por palavra.
FORCAST formula	Sim	Sim	Métrica ideal para manuais técnicos e formulários (textos sem narrativas).A fórmula não utiliza o tamanho das frases, portanto pode ser utilizada em textos sem frases.
New Dale-Chall Readability Formula	Não	Sim	Diferente de outras métricas, essa fórmula utiliza uma lista de 3000 “hard words” para cálculo do grau de escolaridade.
Gunning Fog Index	Sim	Sim	Semelhante ao Flesch Reading Ease.
New Fog Count	Sim	Sim	Métrica ideal para documentos técnicos e manuais. Desenvolvida pela Marinha dos EUA.
Fry Graphical Analysis	Talvez	Sim	Método inicialmente desenvolvido para o Inglês.
Raygor Readability Estimate	Não	Sim	Métrica de mensuração do grau de escolaridade de textos em Inglês. Necessita grande adaptação para o Português.

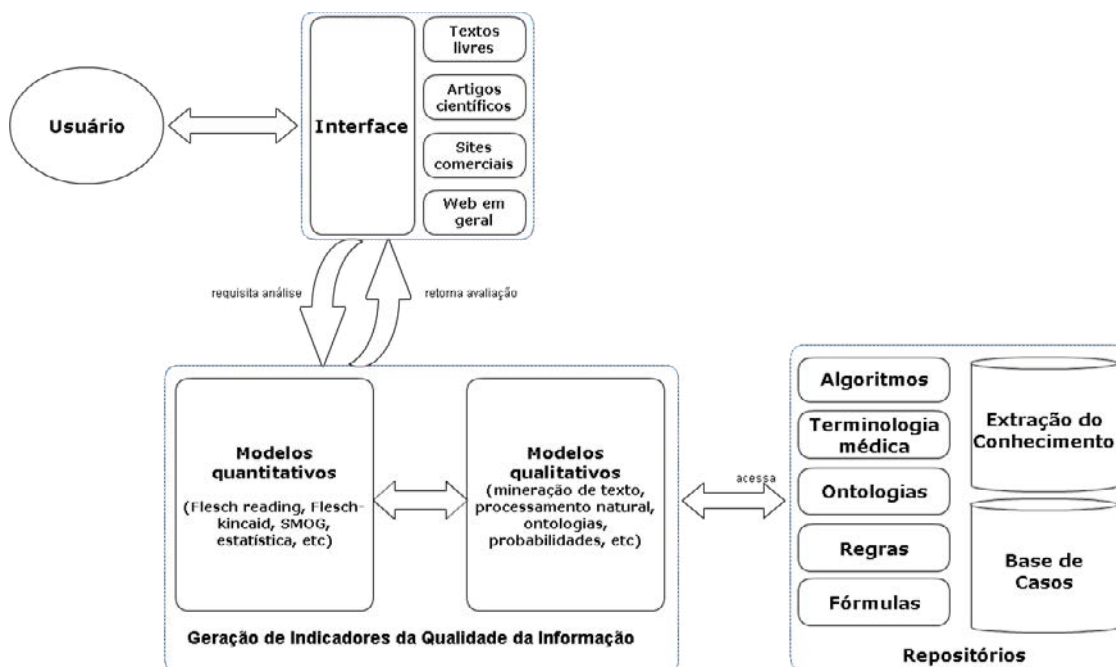
3.2 Fase 2: Desenvolvimento do Software para Avaliação Textual Automática na Área da Saúde

A concepção e modelagem do software encontra-se em desenvolvimento. O software apresentará uma interface visual e textual, com possibilidade de navegação na web para seleção e avaliação de textos (figura 1).

A análise quantitativa será realizada utilizando as ferramentas elencadas na tabela 1. O processamento qualitativo será gerado através de algoritmos de aprendizagem automática e processamento da linguagem natural (máquinas de suporte vetorial, aprendizagem simbólica e análise de latência semântica). O software deve ainda contemplar uma camada de persistência para armazenar dados coletados, calculados, resultados de análises, bem como termos da área da saúde, vocabulários, stop words e bases de casos. O resultado das análises será apresentado na forma de

interfaces gráficas com recursos de visualização que permitam inferir sobre a qualidade da informação e apreensibilidade identificadas automaticamente.

Figura 1 - Arquitetura de software de análise de apreensibilidade da informação



4. Contextualização e Perspectivas Futuras

A Internet, apesar de ser uma fonte útil e relevante na busca por conhecimento, deve oferecer mecanismos objetivos de verificação, usando métricas de apreensibilidade e qualidade. A fim de construir uma fundamentação sólida para a avaliação textual optou-se por uma abordagem híbrida (quantitativa e qualitativa). Em termos quantitativos pretende-se avaliar características de apreensibilidade. Utilizando técnicas oriundas da Inteligência Artificial busca-se implementar mecanismos que permitam avaliar a qualidade da informação. Tais técnicas se complementam, trazendo ao usuário dados que lhe permitam julgar um conteúdo textual a partir de métricas bem definidas. As próximas etapas deste trabalho envolvem a implementação do software, testes e análise dos resultados por meio da comparação da avaliação realizada por especialistas humanos em uma amostra de documentos extraídos da Internet. Espera-se desta forma contribuir com um mecanismo de avaliação correto e preciso dos sites de conteúdos de doenças da coluna vertebral, permitindo que informações claras e verdadeiras sejam oferecidas aos usuários.

5. Referências

- Feldman, R.; Sanger, J. The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press, 2007.
- Klare, G. R.: The measurement of readability. Iowa: Iowa State University Press,(1963)
- Luger, George.F. Artificial Intelligence. Person Education, 2009,774 p.
- Weiss, Sholom M., Indurkha, Nitin, Zhang, Tong. Fundamentals of Predictive Text Mining, Springer-Verlag London, 2010.

Identificação de Situações de Risco para Pacientes em Reabilitação Cardíaca Explorando uma Arquitetura de Software na Internet das Coisas

Douglas Scheunemann¹, Adenauer Yamin¹, João Lopes², Cláudio Geyer²

¹Universidade Católica de Pelotas (UCPEL)
Pelotas – RS – Brasil

²Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

scheunemann.d.a@gmail.com, adenauer.yamin@ucpel.edu.br
{jlblopes, geyer}@inf.ufrgs.br

Abstract. *The Internet of Things (IoT) is influencing how computer systems are developed, enabling new application scenarios and a more proactive user interaction, expanding features of mobility and availability. In this work the infrastructure provided by IoT is applied in an architecture to identify risk situations for patients in cardiac rehabilitation. The proposed architecture, called EXEHDA-HP, supports hybrid processing of context data combining techniques based on specification and learning to identify situations.*

Resumo. *A Internet das Coisas (IoT) vem influenciando a maneira como os sistemas computacionais são desenvolvidos, possibilitando novos cenários de aplicação e uma interação mais proativa com os usuários, expandindo características de mobilidade e disponibilidade. Neste trabalho a infraestrutura oferecida pela IoT é explorada por uma arquitetura na identificação de situações de risco para pacientes em reabilitação cardíaca. A arquitetura proposta, denominada EXEHDA-HP, suporta o processamento híbrido de contexto combinando o emprego de técnicas baseadas em especificação e aprendizado para a identificação de situações.*

1. Introdução

Avanços em tecnologias de sensores, atuadores e protocolos de rede sem fio tem permitido o desenvolvimento de soluções de computação embarcada para diferentes finalidades. Dispositivos como utensílios domésticos, carros, equipamentos médicos e *smartphones* podem ser conectados e interagir a qualquer momento, em qualquer lugar com qualquer pessoa ou outro dispositivo através de uma rede unificada, denominada de Internet das Coisas (IoT) [Razzaque et al. 2015].

A principal contribuição deste trabalho é uma arquitetura de software, integrada ao middleware EXEHDA (*Execution Environment for Highly Distributed Applications*) [Lopes et al. 2014a], chamada EXEHDA-HP (EXEHDA - Hybrid Processing), a qual provê ciência de situação para aplicações em IoT, permitindo o processamento híbrido de dados de contexto, combinando técnicas baseadas em especificação e aprendizado para detecção de situações. Esta combinação provida pela abordagem híbrida, contribui para qualificar o mecanismo de detecção de situações [Ye et al. 2012].

Através do estudo de caso em reabilitação cardíaca foi possível identificar o potencial de uso da arquitetura proposta em aplicações médicas que necessitem de funções para ciência de situação no cenário distribuído típico da Internet das Coisas.

2. EXEHDA-HP: Concepção da Arquitetura

O middleware EXEHDA possui dois tipos de servidores: (i) Servidor de Borda, responsável por interagir com ambiente através de sensores e atuadores; e (ii) Servidor de Contexto, responsável por prover funcionalidades para ciência de situação. Estes servidores são alocados em células no ambiente gerido pelo EXEHDA, onde cada célula possui um Servidor de Contexto e pode possuir vários Servidores de Borda.

O Servidor de Contexto é formado por cinco módulos: Aquisição, Atuação, Notificação, Comunicação e Processamento [Lopes et al. 2014b]. A arquitetura do EXEHDA-HP, apresentada na Figura 1, está integrada ao Módulo de Processamento.

O bloco **Gerenciador de Contexto** é responsável pela funcionalidade de processamento dos dados contextuais. As etapas de fusão e interpretação utilizam modelos baseados em aprendizado, regras, ou ambos, caracterizando assim um modelo híbrido. Um dos principais desafios do modelo híbrido proposto é permitir que o treinamento e uso de algoritmos de aprendizado possa ser feito de uma forma transparente para as camadas de aplicação do EXEHDA. Para adicionar essa funcionalidade ao EXEHDA, foram criados os componentes do bloco **Mecanismos de Aprendizado**.

Utilizando a **API de Treinamento** o administrador da aplicação poderá disponibilizar os dados de treinamento, escolher a técnica de processamento e especificar as características que devem ser extraídas dos dados para executar a técnica de aprendizado. Após a execução do treinamento, os parâmetros obtidos são armazenados no repositório **Modelos de Aprendizado**. Esses parâmetros são acessados através da **API de Execução** durante a utilização da aplicação.

O bloco **Extração de Características** executa o pré-processamento das variáveis de contexto ou de sinais obtidos diretamente de sensores, extraindo o conjunto de características relevantes para o treinamento e a utilização dos algoritmos de aprendizado. O **Classificador de Situações** permite a inferência de situações considerando as informações de contexto geradas pelo processamento híbrido. São aplicados modelos de raciocínio fuzzy em conjunto com técnicas de aprendizado para classificação.

3. EXEHDA-HP: Avaliação da Arquitetura

Para realizar a avaliação das funcionalidades da arquitetura foi prototipada uma aplicação para monitoramento de pacientes em reabilitação cardíaca após um acidente vascular. Estudos na área de cardiologia indicam que exercícios físicos podem reduzir o tempo de recuperação destes pacientes. Porém, respostas desproporcionais na frequência cardíaca podem indicar situações de risco para o paciente. Desta forma, o reconhecimento da atividade física e sua correlação com a frequência cardíaca pode permitir uma recuperação mais segura [Negrão and Barreto 2010]. O risco inerente a situação do paciente foi classificado no domínio linguístico: baixo, moderado e alto.

No trabalho de Kwapisz et al. [Kwapisz et al. 2011] foram capturados sinais de acelerômetros de *smartphones* de voluntários durante a execução das seguintes atividades

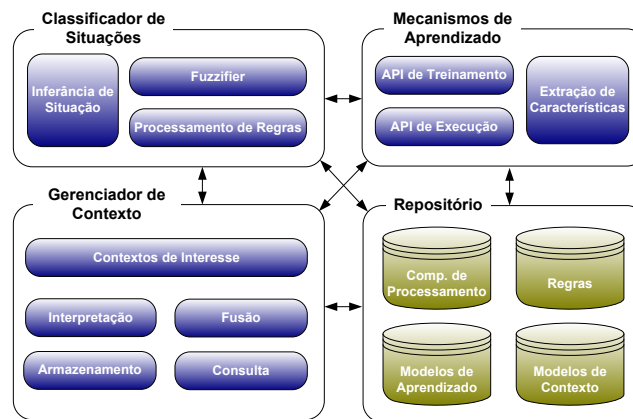


Figura 1. Arquitetura do EXEHDA-HP

físicas: caminhando, correndo, subindo escada, descendo escada, sentado e de pé. Os sinais foram disponibilizados em um banco de dados, o qual foi utilizado neste trabalho para verificar o funcionamento dos mecanismos de aprendizado do EXEHDA-HP na classificação de atividades.

Os resultados de classificação de atividades para diferentes algoritmos de aprendizado podem ser vistos na Tabela 1. Dentre os métodos, o que apresentou a maior exatidão geral foi Árvore de Decisão J48. Além disso, esse método também necessitou um menor tempo de processamento para seu treinamento.

Tabela 1. % de atividades corretamente classificadas.

	Árvore de Decisão J48	Redes Bayesianas	Redes Neurais	Regressão
Caminhando	87.33%	82.94%	89.23%	92.37%
Correndo	95.80%	94.19%	98.89%	97.41%
Subindo Escadas	59.21%	39.31%	52.96%	24.51%
Descendo Escadas	54.66%	30.64%	40.17%	10.35%
Sentado	98.66%	94.98%	98.33%	95.32%
De Pé	97.56%	84.96%	93.90%	91.06%
Total	84.88%	77.50%	84.39%	79.05%

A frequência cardíaca apropriada para cada atividade física deve ser previamente informada pelo terapeuta. O domínio linguístico para a frequência cardíaca é dado por: muito baixa, baixa, normal, alta e muito alta. No bloco **Fuzzifier** os valores numéricos obtidos do sensor de frequência cardíaca são transformados para o domínio fuzzy aplicando uma função de pertinência trapezoidal. Na Figura 2, é mostrado o algoritmo de raciocínio fuzzy aplicado para determinar a situação que representa o risco atual do paciente.

4. Trabalhos Relacionados

O estudo de trabalhos relacionados foi feito considerando as especificações do EXEHDA-HP: arquitetura utilizada em diferentes domínios de aplicação; arquitetura distribuída; modelo híbrido de processamento de contexto; suporte para treinamento de algoritmos de aprendizado e identificação de situações.

Yuan e Herbert [Yuan and Herbert 2014] propõem uma arquitetura para suporte ao monitoramento reconhecimento de atividades e hábitos. Neste trabalho é utilizado apenas um tipo de algoritmo de aprendizado, o qual é baseado em casos. No EXEHDA-

HP quatro algoritmos de aprendizado foram testados, além disso, é oferecido suporte para inclusão de novos algoritmos no repositório de componentes de processamento.

```

GET activity
GET activity_normal_params
IF heart_rate IS very_high OR heart_rate IS very_low THEN
    risk IS high
IF heart_rate IS high OR heart_rate IS low THEN
    risk IS moderate
IF heart_rate IS normal THEN
    risk IS low

```

Figura 2. Algoritmo para inferência de situação

Haghighi et al. [Haghighi et al. 2014] apresentam uma arquitetura para uma aplicação específica, sem suporte a processamento distribuído. Já o EXEHDA-HP foi especificado para ser independente do domínio de aplicação e sua arquitetura foi construída para poder operar de forma distribuída.

5. Considerações Finais

A principal contribuição do EXEHDA-HP é uma arquitetura para prover ciência de situação, com suporte ao processamento híbrido de dados de contexto, aplicando técnicas baseadas em especificação e aprendizado para detectar situações.

Como trabalho futuro pretende-se otimizar as rotinas de treinamento dos algoritmos de aprendizado, incluindo duas novas funções: seleção automática de características e seleção automática de técnica de aprendizado.

Referências Bibliográficas

- Haghighi, D. P., Averi, P., Indrawan-Santiago, M., and Minh, H. T. (2014). Situation-aware mobile health monitoring. *ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)*, pages 248—256.
- Kwapisz, J. R., Weiss, G. M., and Moore, S. a. (2011). Activity recognition using cell phone accelerometers. *ACM SIGKDD Explorations Newsletter*, 12:74.
- Lopes, J., Souza, R., Geyer, C., Costa, C., Barbosa, J., Pernas, A., and Yamin, A. (2014a). A middleware architecture for dynamic adaptation in ubiquitous computing. *Journal of Universal Computer Science*, 20(9):1327–1351.
- Lopes, J., Souza, R., Pernas, A., Yamin, A., and Geyer, C. (2014b). A distributed architecture for supporting context-aware applications in ubicomp. In *Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on*, pages 584–590.
- Negrão, C. E. and Barreto, A. C. P. (2010). *Cardiologia do Exercício: do Atleta ao Cardiopata*. Manole, Barueri, SP - Brazil, 3 edition.
- Razzaque, M. A., Milojevic-Jevric, M., Palade, A., and Clarke, S. (2015). Middleware for Internet of Things: a Survey. *Internet of Things Journal, IEEE*, PP(99):1.
- Ye, J., Dobson, S., and McKeever, S. (2012). Review: Situation identification techniques in pervasive computing: A review. *Pervasive Mob. Comput.*, 8(1):36–66.
- Yuan, B. and Herbert, J. (2014). Context-aware hybrid reasoning framework for pervasive healthcare. *Personal and Ubiquitous Computing*, 18(4):865–881.

Aplicativo Multiplataforma para Planejar Cirurgias de Estrabismo

Igor T. C. Furtado¹, João D. S. Almeida¹, Jorge A. Meireles-Teixeira¹, Geraldo B. Junior¹

¹Núcleo de Computação Aplicada (NCA) – Universidade Federal do Maranhão (UFMA)
Caixa Postal 65.085–580 — São Luís — MA — Brasil

{ygortalysson, jorgemeireles1, ge.braz}@gmail.com, joao.dallyson@ufma.br

Abstract. *Strabismus is a pathology that affects about 4% of the population, causing aesthetic problems and irreversible sensorial changes. Most strabismus cases required surgical treatment for their elimination. Variables involved in the surgical plan are very complex, involving both participation of theoretical scientific knowledge and surgeon experience, which most often uses median values based on formulas and tables. These values do not always apply to all cases. Thus, this work presents a mobile application device for manage and indicate the strabismus surgical plan.*

Resumo. *O estrabismo é uma patologia que afeta cerca de 4% da população, provocando problemas estéticos e alterações sensoriais irreversíveis. A maioria dos casos de estrabismo exigem tratamento cirúrgico para a sua eliminação. As variáveis envolvidas no plano cirúrgico são complexas, envolvendo tanto a participação do conhecimento científico teórico como a experiência do cirurgião, que na maioria das vezes utiliza valores medianos baseados em fórmulas e tabelas. Estes valores nem sempre se aplicam a todos os casos. Assim, o presente trabalho apresenta um aplicativo para auxiliar os oftalmologistas a gerenciar e planejar cirurgias de estrabismos.*

1. Introdução

O estrabismo é uma patologia que consiste na ausência do normal paralelismo entre os eixos visuais [Souza-Dias and Almeida 1998]. Esta patologia resulta de uma alteração óculo-motora que impede a fixação bifoveal. A intervenção cirúrgica está indicada para a maioria dos casos com desvios acima de 15Δ (quinze dioptrias prismáticas). Existem basicamente dois tipos de procedimentos cirúrgicos. No primeiro tipo, podem ser feitos recuos milimétricos da posição original do músculo ocular, ressecção de uma parte dele ou mesmo suturas. Já no segundo, realizam-se ressecções milimétricas do comprimento do músculo ocular, reposicionamento anterior da inserção e pregueamento do tendão muscular. Estes procedimentos são realizados em um ou mais músculos oculares, dependendo do tipo de estrabismo que o paciente apresenta.

É evidente que os avanços da tecnologia móvel melhoram o acesso e auxiliam os profissionais da saúde no diagnóstico e tratamento das patologias. Já existem diversas aplicações oftalmológicas disponíveis e que transformam os *smartphones* em dispositivos médicos sofisticados [Zvornicanin et al. 2014], porém, essa não é uma realidade dentro da subespecialidade estrabismo, o que dificulta o diagnóstico precoce e o planejamento de cirurgias de estrabismos.

Diante dessa problemática, apresenta-se um aplicativo móvel multiplataforma para planejar cirurgias de estrabismo. Através de uma interface portátil e de fácil manuseio, o especialista gerencia e elabora seus planos cirúrgicos ou recebe uma sugestão automática de plano com base em técnicas de aprendizado de máquina, utilizando vetores de suporte para regressão (SVR). Para tanto, o aplicativo incorpora a solução proposta por Almeida et al. (2015) para planejar cirurgias de estrabismo através de múltiplos regressores SVR.

O aplicativo oferece, também, a oportunidade de o usuário solicitar a sugestão de plano a outros especialistas, disponibilizando uma terceira opinião, já que o aplicativo fornece a segunda opinião. Por meio dessa colaboração, o aplicativo contribuirá para a melhoria da qualidade do plano cirúrgico elaborado pelos especialistas e, conseqüentemente, para a prevenção de erros do plano previsto, evitando a necessidade de realização de novo procedimento cirúrgico.

2. Arquitetura

A arquitetura utilizada no desenvolvimento do aplicativo está dividida em 3 camadas, como apresentadas na Figura 1. O usuário tem acesso somente à camada de aplicação, onde se encontra o módulo gerenciador do plano cirúrgico, responsável pela manipulação dos planos e organização das informações nas telas da aplicação. A camada de serviços é composta pelos serviços utilizados pelos componentes da camada de aplicação, sendo estes: a ferramenta Qt Creator¹, o repositório local de arquivos, o módulo de automação e controle do plano cirúrgico e o gerenciador de sincronização. A camada de infraestrutura contém o servidor de banco de dados centralizado e os sistemas operacionais.

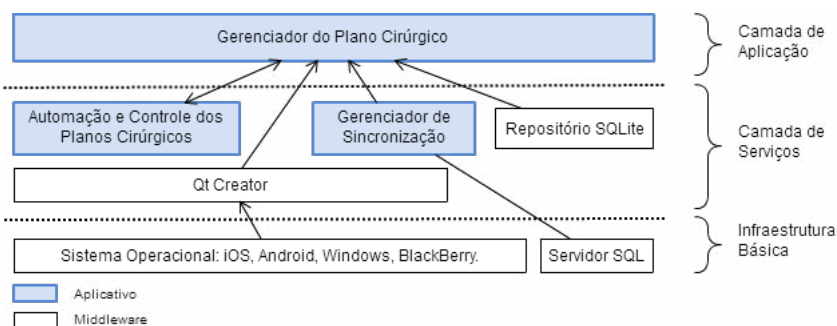


Figura 1. Diagrama da arquitetura da aplicação

Inicialmente, o aplicativo armazena os dados dos planos cirúrgicos gerados em um arquivo local. Em uma versão futura, esses dados serão armazenados e sincronizados em um único repositório central de dados hospedado em nuvem, possibilitando a comunicação e compartilhamento dos planos entre os usuários.

O plano cirúrgico pode ser registrado de duas maneiras: semiautomática, usando a base de planos já realizados; ou manual, baseado na sugestão do próprio especialista. Em ambas, o especialista dependerá dos dados do diagnóstico de cada paciente. Esses dados, provenientes do especialista e do método aqui proposto em Almeida et al. (2015), serão agrupados e submetidos para o algoritmo de aprendizado de máquina supervisionado,

¹<http://www.qt.io/>

cujos regressores estimarão o valor em milímetros de retrocesso (recuo) ou ressecção que se deve operar em cada músculo ocular.

3. Desenvolvimento do Aplicativo

Na tela inicial da aplicação, foi definido um botão para controlar o acesso do especialista ao aplicativo (Figura-2a), sendo que, de acordo com os requisitos, o usuário só poderá criar e gerenciar os planos cirúrgicos se estiver devidamente cadastrado (Figura-2b). O cadastro é composto por dados básicos do especialista (Figura-2c), somente para controle das informações. Qualquer plano gerado no programa estará relacionado ao respectivo usuário.



Figura 2. Telas iniciais da aplicação: (a)Tela inicial, (b)Acesso à aplicação, (c)Cadastro do especialista.

Após realizar o acesso, o especialista tem a opção de criar um novo plano cirúrgico ou manipular os já armazenados na base local, mesmo para aqueles gerados automaticamente pelo método computacional. Caso queira criar o seu próprio plano, precisará dispor dos dados do paciente diagnosticado com estrabismo (Figura-3) e, em seguida, preencher obrigatoriamente todos os campos antes de clicar em salvar.



Figura 3. Telas de criação de um novo plano cirúrgico elaborado pelo especialista: (a)Dados do diagnóstico do paciente, (b) Dados das versões, (c) Dados do procedimento a ser aplicado.

Um mesmo paciente pode ter tantos planos cirúrgicos quantos forem necessários para que o especialista tenha certeza de qual método cirúrgico aplicar. Em seu auxílio,

poderá criar um plano automaticamente, usando o mesmo diagnóstico para comparar se o seu procedimento é parecido com o sugerido pelo aplicativo.

A geração automática dos planos, ainda em desenvolvimento, utiliza técnicas de aprendizado de máquina supervisionada. Por conta disso, fez-se necessário obter uma base de dados de prontuários de pacientes que já foram submetidos a procedimentos cirúrgicos. De cada paciente, foram extraídas 23 características: fixação binocular, equivalente esférico direito e esquerdo, desvios das cinco posições diagnósticas, tipo de desvio, acuidade visual e as 12 medidas de versões dos olhos esquerdo e direito. Estas características foram escolhidas conforme relevância da informação para o plano cirúrgico elaborado pelo especialista [Almeida et al. 2015].

Após passar por um processo de normalização para uma faixa de valores comuns entre -1 e 1, estima-se o plano cirúrgico para cada diagnóstico de paciente utilizando os regressores SVR treinados com os mesmos parâmetros utilizados por Almeida et al. (2015). A estimação do plano é realizada de forma instantânea.

4. Considerações Finais

Este trabalho apresentou uma ferramenta para gerenciar planos de cirurgia de estrabismo, servindo tanto para auxiliar o especialista como para treinar profissionais da oftalmologia e médicos residentes, uma vez que estes últimos poderão utilizar os planos indicados pelo aplicativo ou o plano sugerido em colaboração por outro especialista como um tira-provas.

Certamente, apesar de estar ainda em desenvolvimento, podem-se antever diversas melhorias em versões futuras do aplicativo. Por exemplo, a integração com banco de dados em nuvem e, conseqüentemente, a atualização das informações em tempo real, permitindo, assim, a comunicação entre os especialistas e demais usuários do sistema. Citem-se ainda a inclusão e a integração com módulo de diagnóstico automático do estrabismo em imagens para fornecer as medidas do desvio, tipo de estrabismo e olho fixador.

Por fim, vale destacar que após a realização dos devidos testes qualitativos e de validação do aplicativo, este poderá ser utilizado nos mais diversos centros oftalmológicos. A utilização do aplicativo nestes centros poderá contribuir para uma melhor gestão dos recursos financeiros. Faz-se a ressalva de que o objetivo não é substituir o especialista, mas oferecer-lhe suporte por meio de outras opiniões na elaboração do plano cirúrgico.

5. Agradecimentos

Os autores agradecem à FAPEMA e ao CNPq, pelo apoio financeiro.

Referências

- Almeida, J. D. S., Silva, A. C., Teixeira, J. A. M., Paiva, A. C., and Gattass, M. (2015). Surgical planning for horizontal strabismus using support vector regression. *Computers in biology and medicine*, 63:178–186.
- Souza-Dias, C. and Almeida, H. (1998). Estrabismo. *Conselho Brasileiro de Oftalmologia. São Paulo: Rocca*, pages 12–128.
- Zvornicanin, E., Zvornicanin, J., and Hadziefendic, B. (2014). The use of smart phones in ophthalmology. *Acta Informatica Medica*, 22(3):206.

WIT - Women in Information Technology Fórum Meninas Digitais

Apresentação

O X *Women in Information Technology* (WIT) é uma iniciativa da SBC para discutir os assuntos relacionados a questões de gênero em Tecnologia de Informação (TI) no Brasil – histórias de sucesso, políticas de incentivo e formas de engajamento e atração de jovens, especialmente mulheres, para as carreiras associadas à TI. Organizado na forma de palestras convidadas e painéis, o workshop é centrado em debater problemas relacionados à mulher e ao seu acesso à TI, tanto do ponto de vista de mercado de trabalho quanto de inclusão e alfabetização digital. Os temas abordados se concentram na necessidade de educar, recrutar e treinar mulheres, como uma política estratégica para o desenvolvimento e competitividade nacional e regional.

O **Fórum Meninas Digitais** faz parte das atividades do WIT desde 2011 e é uma das ações do **Programa Meninas Digitais** da SBC. Tal Programa é direcionado às alunas do ensino fundamental, médio e tecnológico, para que conheçam melhor a área de informática e das Tecnologias da Informação e Comunicação, de forma a motivá-las a seguir carreira nessas áreas. Durante o evento os multiplicadores desta proposta aproveitam a oportunidade para discutir projetos e parcerias, de forma a disseminar esta ideia no território nacional.

Celebrando sua 10ª edição, pela primeira vez, o WIT, juntamente com o Programa Meninas Digitais, realiza a chamada para a submissão de trabalhos apresentados na forma de pôsteres. Nos últimos anos, houve um crescimento do número de projetos nesta área, fomentados por órgãos financiadores, instituições de ensino e empresas, caracterizados tanto como pesquisa quanto extensão. Neste cenário, o registro de experiências e ações faz-se necessário. Para esta edição, foram selecionados vinte e cinco trabalhos que apresentam propostas de ações, resultados de pesquisa e extensão, estudos de caso e experiências relevantes ao tema “Mulheres e Meninas na Computação”.

Priscila Solis (UnB)
Juliana Freitag Borin (UNICAMP)
Ana Benso (PUCRS)
Cristiano Maciel (UFMT)
Sílvia Amélia Bim (UTFPR - Curitiba)
Érica Calil Nogueira (UFRJ)
Karen Figueiredo (UFMT)

Comitê de Organização

Coordenação Geral WIT

Priscila Solis (UnB)
Juliana Freitag Borin (UNICAMP)

Coordenação Local WIT e Fórum Meninas Digitais

Ana Cristina Benso da Silva (PUCRS)

Coordenação Geral Fórum Meninas Digitais

Cristiano Maciel (UFMT)
Sílvia Amélia Bim (UTFPR - Curitiba)

Comitê de Organização Fórum Meninas Digitais

Érica Calil Nogueira (UFRJ)
Karen Figueiredo (UFMT)

Coordenação da Chamada de Trabalhos WIT + Meninas Digitais

Sílvia Amélia Bim (UTFPR - Curitiba)

Comitê de Programa WIT + Meninas Digitais

Ana Paula Chaves Steinmacher - UTFPR - Campo Mourão
Andreia Malucelli - PUCPR
Anita Fernandes - UNIVALI
Carlos Silla - UTFPR - Cornélio Procópio
Christiane Heemann - UNIVALI
Cláudia Nalon - UNB
Cristiano Maciel - UFMT
Ecivaldo Matos - UFBA
Fabiola Guerra Nakamura - UFAM
Fabiana Mendes - UNB
Fernanda Cunha - UNIVALI
Idalmis Milián Sardiña - UFRN
Igor Steinmacher - UTFPR - Campo Mourão
José Viterbo - UFF

Juliana Freitag Borin - UNICAMP
Karen Figueiredo - UFMT
Karin Becker - UFRGS
Karina Mochetti - UFF
Luciana Frigo - UFSC
Luciana Salgado - UFF
Luiz Merkle - UTFPR - Curitiba
Maria Augusta Silveira Netto Nunes - UFS
Maria Claudia Emer - UTFPR - Curitiba
Mariangela Gomes Setti - UTFPR - Curitiba
Marilia Amaral - UTFPR - Curitiba
Maristela Holanda - UNB
Milene Silveira - PUCRS
Mirella Moro - UFMG
Nadia Kozievitch - UTFPR - Curitiba
Rosiane de Freitas - UFAM
Silvia Amelia Bim - UTFPR - Curitiba
Tainá Medeiros - IFRN
Tanara Lauschner - UFAM

Painéis / Palestras

Comemorando o WIT - Aventuras e Desventuras de uma Pesquisadora Brasileira

Claudia Bauzer Medeiros (UNICAMP)

A palestra vai comentar alguns fatos interessantes da experiência de ensino e pesquisa da palestrante como aluna, professora e pesquisadora de Computação.

Vamos lá Gurias!

Karin Becker (UFRGS)

Como uma profissional realizada, demorei a perceber o problema e o impacto da redução do gênero feminino nos diferentes cursos relacionados à área da informática. Nesta palestra, descrevo minha jornada pessoal na causa das mulheres na computação. Descrevo os fatores que me levaram a acreditar que a causa era importante, e ações que temos desenvolvido com as alunas da UFRGS para atrair mais meninas em nossos cursos de computação, e mantê-las motivadas e unidas ao longo do curso!

Mulheres na Maratona de Programação

Rosiane de Freitas Rodrigues (UFAM)

Carlos Ferreira (USP)

Rodolfo Jardim de Acevedo (UNICAMP)

Maristela Terto de Holanda (UnB) - moderadora

O painel discutirá as experiências e a importância da participação das mulheres na Maratona de Programação.

Fórum Meninas Digitais - Desafios para atrair e manter mais mulheres na Computação

Lisandro Zambenedetti Granville (UFRGS) - Presidente da SBC

Thais Vasconcelos Batista (UFRN) - Vice-presidente da SBC

Tanara Lauschner (UFAM) - Coordenadora do Projeto Cunhantã Digital

Karin Breitman - EMC Brazil - R&D Division

Cristiano Maciel (UFMT) - Coordenador do Programa Meninas Digitais

Sílvia Amélia Bim (UTFPR-Curitiba) - Coordenadora do Programa Meninas Digitais

Neste painel, discute-se os desafios em diversas esferas para atrair e manter mais mulheres na Computação. Há aspectos culturais que constroem e reforçam estereótipos, comportamentos e julgamentos que afastam as mulheres da área. Existem também os aspectos educacionais, que levam a formulação e proposta de currículos pouco atrativos. Há ainda aspectos políticos e econômicos, que dificultam a realização de ações para reverter o cenário atual. Os painelistas deste Fórum são convidados a apresentarem suas considerações sobre estes diversos aspectos em busca de um mapeamento de ações que possam lidar com os desafios enfrentados pela comunidade brasileira.

Trabalhos aceitos

Ação para Incentivar Meninas do Ensino Médio a Cursar Carreiras Tecnológicas da Universidade Federal do Rio Grande do Norte

Idalmis Milián Sardiña (UFRN)
Cristiano Maciel (UFMT)

Análise da Influência de Grupos de Inclusão de Mulheres em Comunidades de Software Livre

Mariana Prudêncio (UTFPR-Campo Mourão)
Ana Paula Chaves Steinmacher (UTFPR-Campo Mourão)

Ciência da Computação também é coisa de menina!

Karina Mochetti (UFF)
Luciana Salgado (UFF)
Ana Valladares Zerbinato (UFF)
Beatriz Souza (UFF)
Meirylen Rosa Emidio Avelino (UFF)

Computasseia: destacando a participação feminina na História da Computação

Karen Figueiredo (UFMT)
Jean Carlos Santos (UFMT)

Cunhantã Digital: programa de incentivo a participação de mulheres da região amazônica na Computação e áreas afins

Tanara Lauschner (UFAM)
Rosiane de Freitas (UFAM)
Fabiola Guerra Nakamura (UFAM)
Ludymila Lobo de Aguiar Gomes (UFAM)

Curto-circuito na Escola

Kaleandra Almeida (UFSC - Araranguá)
Ana Carolina Marcelino (UFSC - Araranguá)
Antony B. Turra (UFSC - Araranguá)
Luciana Frigo (UFSC - Araranguá)
Eliane Pozzebon (UFSC - Araranguá)

Divulgar para Atrair, Motivar para Manter

Silvia Amelia Bim (UTFPR-Curitiba)
Marília Amaral (UTFPR-Curitiba)
Nadia Kozievitch (UTFPR-Curitiba)
Maria Claudia Emer (UTFPR-Curitiba)
Mariangela Gomes Setti (UTFPR-Curitiba)
Luiz A. Pelisson (UTFPR-Curitiba)
Luiz Merkle (UTFPR-Curitiba)

Empoderamento Feminino com Dispositivos Móveis

Joice Cardoso (UFSC - Araranguá)
Laís Dalle Mulle (UFSC - Araranguá)
Luciana Frigo (UFSC - Araranguá)
Eliane Pozzebon (UFSC - Araranguá)

I Fórum Gurias na Computação: relato de experiências e desdobramentos

Isadora Ferrão (UNIPAMPA - Alegrete)
Amanda Melo (UNIPAMPA - Alegrete)

Incentivando meninas do ensino médio a ingressarem nas áreas tecnológicas com curso de programação e robótica

Tatiana Santos (UFSC - Araranguá)
Maite Manenti (UFSC - Araranguá)
Eliane Pozzebon (UFSC - Araranguá)
Luciana Frigo (UFSC - Araranguá)

Incentivando o Ingresso de Mulheres nos Cursos de Engenharia e Tecnologia

Marcelly Homem Coelho (UFSC - Araranguá)
Morgana Sartor (UFSC - Araranguá)
Luciana Frigo (UFSC - Araranguá)
Eliane Pozzebon (UFSC - Araranguá)

Levantamento e análise dos dados da presença feminina discente no curso de Bacharelado em Sistemas de Informação do IFC-Cam

Ana Elisa Schmidt (IFC - Camboriú)
Mozara Koehler (IFC - Camboriú)
Hiryne Britto (IFC - Camboriú)

Mapeamento de iniciativas brasileiras que fomentam a entrada de mulheres na computação

Maria Augusta Silveira Netto Nunes (UFS)
Carolina Louzada (UFS)
Edilayne Salgueiro (UFS)
Beatriz Andrade (UFS)
Patricia S de Lima (Escola Atheneu)
Raquel Monteiro (UFS)

Mapeamento de iniciativas estrangeiras que fomentam a entrada de mulheres na computação

Maria Augusta Silveira Netto Nunes (UFS)
Carolina Louzada (UFS)
Edilayne Salgueiro (UFS)
Beatriz Andrade (UFS)
Patricia S de Lima (Escola Atheneu)
Raquel Monteiro (UFS)

Meninas Digitais Regional Mato Grosso: Práticas Motivacionais no Ensino Médio para a Equidade de Gênero nas Carreiras e Cursos de Computação e Tecnologias

Karen Figueiredo (UFMT)

Pedro Clarindo da Silva Neto (IFMT - Tangará da Serra)

Cristiano Maciel (UFMT)

Meninas e Jovens na Computação: Cursos de Programação Web EaD

Letícia Heinzmann (IFRS - Bento Gonçalves)

Maria Eduarda Romagna (IFRS - Bento Gonçalves)

Julia Marques Carvalho da Silva (IFRS - Bento Gonçalves)

Meninas, Computação e Música

Thayna Gimenez (UTFPR - Cornélio Procópio)

Caroline Rezende Barroso (UTFPR - Cornélio Procópio)

Jéssika Pereira (UTFPR - Cornélio Procópio)

André Przybysz (UTFPR - Cornélio Procópio)

Adriano Rivolli (UTFPR - Cornélio Procópio)

Carlos Silla (UTFPR - Cornélio Procópio)

Meninas.comp: Um Relato da Experiência de Integração Ensino Médio e a Universidade

Maristela Holanda (UnB)

Aleteia Araujo (UnB)

Maria Emilia Walter (UnB)

Carlos Alberto Jesus de Oliveira (Centro de Ensino Médio Paulo Freire)

Professoras negras na pós-graduação em Ciência da Computação: uma proposta de pesquisa

Ecivaldo Matos (UFBA)

Maria José dos Santos (UFPE)

Juliana Santos (UFBA)

Ana Carolina Ferreira (UFBA)

Emilayne Feitosa Corlett (UFBA)

Projeto SciTechGirls: desenvolvimento de aplicativos e participação em competições de programação científicas e tecnológicas

Rosiane de Freitas (UFAM)

Ludymila Lobo de Aguiar Gomes (UFAM)

Tayana Conte (UFAM)

Reflexões sobre Gênero, Tecnologia e Processo Vocacional

Karen Figueiredo (UFMT)

Cristiano Maciel (UFMT)

Robótica na aprendizagem de física e matemática para alunas do ensino médio

Maite Manenti (UFSC - Araranguá)
Everton Jasinski (UFSC - Araranguá)
Tatiana Santos (UFSC - Araranguá)
Eliane Pozzebon (UFSC - Araranguá)
Luciana Frigo (UFSC - Araranguá)

Trazendo Meninas para a Computação

Helena Ribeiro (UCS)
Iraci C. S. de Carli (UCS)
Maria de Fátima W. P. Lima (UCS)
Simone C. M. Paiva (UCS)
Naura A. Luciano (UCS)

**Um Estudo Inicial Sobre As Alunas Ingressantes no Curso de Licenciatura em
Ciência da Computação da Universidade Federal da Paraíba**

Vanessa Dantas (UEPB - Rio Tinto)
Renata Figueiredo (UEPB - Rio Tinto)
Gabriela Souza (UEPB - Rio Tinto)
Ryslânia Rayana Santos Nascimento (UEPB - Rio Tinto)
Ana Caroline Nascimento (UEPB - Rio Tinto)
Beatriz Martins (UEPB - Rio Tinto)

"Você é o melhor menino desse grupo!"

Ana Cristina Oliveira (UEPB - Campina Grande)
Karolyne Oliveira (UEPB - Campina Grande)
Yngrid Cabral (UEPB - Campina Grande)
Edlane Alves (UEPB - Campina Grande)
Marcela Santos (UEPB - Campina Grande)
Thayanne Sousa (UEPB - Campina Grande)
Fanny Vieira (UEPB - Campina Grande)

Ações para Incentivar Meninas do Ensino Médio a Cursar Carreiras Tecnológicas da Universidade Federal do Rio Grande do Norte

Idalmis Milián Sardiña¹, Cristiano Maciel²

¹ Escola de Ciências e Tecnologia - Universidade Federal de Rio Grande do Norte (UFRN) - Campus Universitário Lagoa Nova - Natal - RN - 59078-970 - Brasil

² Instituto de Computação - Universidade Federal de Mato Grosso (UFMT) Av. Fernando Corrêa da Costa, 2367, Boa Esperança - Cuiabá - MT - 78060-900 - Brasil.

idalmissilian@ect.ufrn.br, cmaciel@ufmt.br

Abstract. *Studies show that there is a deficit in women IT courses, constituting the female least 30% of the total volume of students. On the other hand, research shows that women trained in these areas currently represent a relevant and successful percentage of the population, resulting in excellent professionals. This research continues a previous initiative to integrate the Federal University of Rio Grande do Norte (UFRN) to public high schools, promoting different articulations. Currently UFRN already performs concrete integrated activities with the State School President Roosevelt and exchange with the Havana-Cuba, for the exchange of cultural models and teaching, as well as acquisition of new experiences. This particular proposal, suggests a new action to this project, where the high school public school students know technological careers University, promoting associated activities.*

Resumo. *Estudos realizados apontam que existe um déficit de mulheres em cursos TI, constituindo o sexo feminino menos de 30% do volume total de alunos. Por outro lado, as pesquisas mostram que mulheres formadas nestas áreas, atualmente, representam um percentual relevante e bem sucedido da população brasileira, resultando em excelentes profissionais. Assim, esta pesquisa dá continuidade a uma iniciativa de integrar a Universidade Federal do Rio Grande do Norte (UFRN) às escolas públicas de ensino médio, promovendo diferentes articulações. Neste momento a UFRN já executa atividades integradas concretas com a Escola Estadual Presidente Roosevelt (EEPR) de Parnamirim e um intercâmbio com a Havana-Cuba, na troca de modelos culturais e de ensino, assim como na aquisição de novas experiências. Na proposta em tela sugere-se uma nova ação a este projeto, na qual as alunas de escolas da rede pública de ensino médio possam conhecer as carreiras tecnológicas da Universidade, por meio de atividades associadas.*

1. Introdução

Para muitos estudantes brasileiros, abrir o espaço da universidade significa também abrir as portas para um mundo distante, quase inacessível. Uma iniciativa de articulação entre a educação básica e superior foi proposta em Milián et al. (2012), de forma a aproximar os estudantes da escola pública com a universidade. Tal iniciativa buscou aumentar a integração entre os estudantes em processo de escolarização e os que

chegaram a universidade; socializar informações sobre o universo universitário e profissional; apresentar o espaço físico da universidade; e despertar o interesse dos estudantes para com o papel social da universidade. E, desta forma, ampliar seus horizontes quanto a possibilidades de ingressar nas carreiras.

O processo de inclusão na educação se amplia e se efetiva na medida em que há oportunidade de acesso e permanência de jovens na escola. A ampliação de vagas na escola e a implementação de políticas públicas, já é uma realidade no Brasil e têm de certa forma, garantido maior inserção. Porém, o grande desafio é fazer com que estes jovens não só entrem, mas que permaneçam na escola com possibilidade de inserção no ensino superior. Nesta perspectiva, algumas estratégias organizadas coletivamente pelos grupos de profissionais que atuam nas escolas, podem ser o caminho possível para fomentar, nos adolescentes, o desejo de permanência na escola e o alcance posterior de seus objetivos pessoais. Esta premissa deve promover a construção do Projeto Político Pedagógico escolar em busca do cumprimento de sua função social possibilitar o acesso ao conhecimento socialmente construído. Paralelamente a este objetivo pedagógico, há a necessidade de alimentar “projetos de vida” coletivos e individuais, capazes de impulsionar o sujeito em busca de seus ideais e necessidades. Este é um objetivo transdisciplinar da escola, ou seja, ser capaz de levar o indivíduo a resignificar sua própria vida, redimensionando-a. Inclui-se aí, a necessidade de pensar em criar e planejar situações de articulação e vínculo com este objetivo.

Para corrigir as distorções e os altos índices de insucessos da Educação Básica, foram adotadas recentemente, algumas políticas públicas das quais se destacam: o Pacto Nacional pela Alfabetização na Idade Certa (PNAIC/2012) e o Pacto Nacional pelo Fortalecimento do Ensino Médio (2013). Nessa perspectiva, distintas ações propostas no projeto inicial (padrinhos das escolas, intercâmbios socioculturais com outros países e atividades extracurriculares culturais e de pesquisa), já estão sendo desenvolvidas na UFRN, como expõe Milian et al. (2015).

O presente trabalho representa a inclusão de uma nova ação no projeto, que busca criar ambientes de interlocução com jovens do sexo feminino na interface da educação formal e não formal (Trilla, 2008). O objetivo principal é identificar a vocação destas jovens em relação a áreas de Tecnologia da Informação (TI) com vistas a um possível futuro ingresso em cursos TI da UFRN, como por exemplo, nos bacharelados do Instituto Metr pole Digital e da Escola de Ci ncia e Tecnologia. O projeto ora proposto est  em conson ncia com as metas do Plano Nacional da Educa o (PNE), recentemente aprovada na forma de Lei (Lei no. 13.005 de 25 de Junho de 2014). Este, insere-se junto as metas 3 e 12 que tratam respectivamente da universaliza o do Ensino M dio e a democratiza o do acesso ao Ensino Superior. Acredita-se que trabalho possa contribuir no crescimento do n mero de mulheres interessadas nestas carreiras.

2. Da proposta e sua origem

Neste trabalho uma nova a o foi proposta, como parte do projeto de ensino m dio iniciado anteriormente em Mili n et al. (2012). A a o inclu da visa o aumento da participa o de mulheres estudando em cursos TI da UFRN. Neste sentido, algumas atividades est o sendo definidas e outras j  est o em andamento, como por exemplo, alguns professores de cursos tecnol gicos, em particular da Escola de Ci ncia e Tecnologia e do Instituto Metr pole Digital da UFRN, promovem palestras e atividades de extens o, que atingem especificamente grupos de alunas do segundo ano do ensino

médio da Escola Estadual Presidente Roosevelt (EEPR). A proposta coincide e tem aderência com os objetivos do Programa Meninas Digitais da Sociedade Brasileira de Computação¹, iniciado em 2011. O objetivo deste programa é divulgar a área de Computação para despertar o interesse de estudantes do ensino médio/tecnológico ou dos anos finais do ensino fundamental, para que conheçam melhor a área e, desta forma, motivá-las a seguir carreira em Computação (Maciel; Bim, 2016). As atividades deste programa são realizadas por meio de projetos realizados em diferentes instituições do Brasil e de forma diversificada: oferta de minicursos e oficinas; realização de dinâmicas; palestras com estudantes e profissionais que já atuam na área compartilhando suas experiências etc.

As primeiras instituições que já executam algumas das atividades propostas para esta ação são a UFRN e a EEPR. A UFRN, com sede na cidade do Natal, é o principal centro universitário do estado de Rio Grande do Norte e uma das dez maiores universidades federais do Brasil em número de alunos de graduação (INEP, 2015). A EEPR se destaca por iniciar no ano 2012 uma nova metodologia de ensino, o ensino médio inovador e apresenta uma equipe de professores qualificados com cursos de pós-graduação.

Com o trabalho realizado desde 2012 no projeto, um dos resultados obtidos na EEPR, foi o incremento da motivação dos estudantes a ingressar em cursos superiores. Mais de 50% dos alunos do 3º ano desta escola matricularam em universidades públicas e privadas (Milián et al., 2015). Outro resultado importante foi eles conhecer o funcionamento e as possibilidades oferecidas pelas universidades federais para os estudantes, como bolsas de estudo, intercâmbios com outros países etc. Uma feira de civilizações e um intercâmbio cultural com Cuba (Milián et al., 2014), foram também ações anteriores que levaram os alunos a embarcarem em uma viagem de conhecimento sobre povos e nações, em seus aspectos geo-históricos e culturais (Sacristan, 2002).

Já a nova ação proposta neste trabalho e que começou a ser implementada recentemente pela UFRN, está direcionada a trabalhar somente com grupos de estudantes do sexo feminino de escolas de ensino médio. O objetivo principal é apresentar a estas jovens distintas áreas TI dentro da Universidade, assim como começar a criar algumas habilidades em programação e desenvolvimento de jogos. Para o próximo semestre foram planejados alguns cursos (Web, Banco de dados, HTML, Androide, Redes etc.) que serão dados na escola EEPR por professores da Escola de Ciência e Tecnologia da UFRN. Por outro lado, visitas semanais ao Instituto Metrópole Digital foram programadas e vão permitir as alunas que participam da ação, conhecer os diferentes laboratórios de informática, o Data center, ou seja, a infraestrutura de TI em geral, mostrando assim detalhes da arquitetura dos computadores, das características da rede e o funcionamento da internet.

A interdisciplinaridade, tão discutida nos meios acadêmicos, é inerente às ações de extensão. Com a introdução da ação proposta, a computação será uma nova disciplina inserida na EEPR para grupos específicos. Neste sentido, a participação dos sujeitos nos projetos desta natureza promove, não somente a interdisciplinaridade, mas também a compreensão da abrangência da educação não formal entendida como um processo sociopolítico, tecnológico, cultural e pedagógico para a formação integral da cidadania (Gohn, 2010).

¹ <http://sbc.org.br/institucional-3/meninas-digitais>

Cabe salientar que um fator que distingue os estudantes das instituições particulares em relação aos das públicas é o acesso a informação; fato que é oriundo de questões referentes aos próprios recursos que a escola oferece (corpo docente, material didático, equipamentos, estrutura física da instituição) ou até mesmo das próprias condições sociais do estudante. Sendo assim, por acreditarem que existe uma considerável discrepância, qualitativa e quantitativa no conteúdo assimilado no ensino médio, grande parte desses alunos de escolas públicas sentem-se em desvantagem frente aos alunos de escolas privadas. Alguns pensam que mesmo após o ingresso na universidade será difícil acompanharem o curso, por disporem de menores recursos financeiros.

3. Discussões

O objetivo é ir além, avançando qualitativamente por distintos caminhos, no relacionamento entre a academia e o seu público-alvo: os estudantes de ensino médio. Todavia, na etapa atual, ações específicas com meninas do ensino médio, em especial de Escola Públicas, estão sendo planejadas, em consonância com o Programa Meninas Digitais.

Crê-se que a experiência dos projetos anteriores, de atuação junto ao Ensino Médio, é salutar para entender esse contexto e buscar alternativas viáveis para o projeto. As atividades extensionistas possibilitam criar ambientes privilegiados através da qual os estudantes de graduação, professores e pesquisadores, podem e devem se permitir, sair do seu cotidiano das ações formais de ensino aprendizagem e adentrar nas reflexões sobre uma nova cultura organizacional que exige a aprendizagem de habilidades extracurriculares.

As mudanças para uma visão mais ampliada da educação, para a inovação dos canais existentes, para o estabelecimento de alianças e utilização de recursos de forma a universalizar o acesso à educação e fomentar a equidade (Gohn, 2010), só poderão encontrar ressonâncias a partir das ações dos sujeitos despertados para tal possibilidade por meio de atividades que o ambiente de extensão possibilita (Thiollent, 1985). Neste sentido, este trabalho vislumbra a continuidade das ações propostas em Milián et al. (2015), inserindo o objetivo proposto neste trabalho.

4. Referências

- Brasil. (2014) “Lei do Plano Nacional da Educação (PNE)”. Lei no. 13005 de 25 de Junho de 2014. Presidência da República, Casa Civil.
- _____. Ministério da Educação. Manual do pacto: Pacto pela Alfabetização na Idade Certa: o Brasil do futuro com o começo que ele merece. Brasília, DF, 2012.
- _____. Ministério da Educação. Portaria nº 1140, de 22 de novembro de 2013. Pacto Nacional pelo Fortalecimento do Ensino Médio, Brasília, DF, 2013.
- Gohn, M. G. (2010) “Educação não formal e o educador social – atuação no desenvolvimento de projetos sociais”. São Paulo, Cortez, 2010.
- INEP (2015) “Resumo Técnico da Educação Superior 2013”. Diretoria de Estatísticas Educacionais DEED, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília-DF, 82 p.

- Maciel, C., Bim, S. A. (2016) “Programa Meninas Digitais - ações para divulgar a Computação para meninas do ensino médio”. In: *Computer on the Beach 2016*, Florianópolis, SC. pp. 327-336.
- Milián, I. *et al.* (2015) “Atividades de extensão propostas para aproximar a UFRN às escolas públicas de ensino médio”. XIII Congresso Latinoamericano de Extensão Universitária de Educação Superior, Havana. Cuba
- Milián, I. *et al.* (2014) “Ações para Aproximar a Universidade com Escolas da Rede Pública de Ensino Médio no Brasil”. 9º. Congresso Internacional de Educação Superior - Universidade 2014, Havana. Cuba
- Milián, I. *et al.* (2012) “Integração dos ensinos fundamental, médio e superior no Brasil: Uma experiência da Universidade Federal de Rio Grande do Norte”. 8º. Congresso Internacional de Educação Superior - Universidade 2012, Havana. Cuba
- Sacristan, G. (2002) “Educar e Conviver na Cultura Global”. *Artmed*.
- Thiollent, M. (1985) “Metodologia da pesquisa-ação”. SP: Cortez: Autores Associados (Coleção temas básicos de pesquisa-ação).
- Trilla, J. (2008). “A educação não formal”. In: ARANTES, Valéria Amorin (org.). *Educação formal e não-formal: pontos e contrapontos*. São Paulo: Summus, p. 15-55.

Análise da Influência de Grupos de Inclusão de Mulheres em Comunidades de Software Livre

Mariana Prudêncio¹, Ana Paula Chaves¹

¹Departamento Acadêmico de Computação
Universidade Tecnológica Federal do Paraná – Câmpus Campo Mourão
Caixa Postal 271 – 87301-899 – Campo Mourão – PR – Brasil

mari.mvargas@gmail.com, anachaves@utfpr.edu.br

Abstract. *Research shows a small participation of women in Open Source Software (OSS) Projects. In response to this, a number of initiatives for inclusion and retention has been created in OSS communities in order to change this disproportionality scenario. This ongoing research investigates the influence of these initiatives with regard to engagement and Psychological Sense of Community (PSC) of the women who are OSS communities' members.*

Resumo. *Pesquisas mostram que a participação de mulheres em software livre é baixa. Como resposta, diversas iniciativas para inclusão e retenção tem sido criadas nas comunidades de software livre com a finalidade de mudar este cenário de desproporcionalidade. Essa pesquisa em andamento aborda a influência dessas iniciativas no que diz respeito ao engajamento e sentimento psicológico de comunidade (SPC) das mulheres participantes de comunidades de software livre.*

1. Introdução

A baixa representatividade de mulheres no ramo de tecnologia é uma realidade. No ano de 1983, as mulheres compunham cerca de 43% da força de trabalho em tecnologia da informação, de acordo com dados da *U.S. Bureau of Labor Statistics Current Population Survey* [Coder et al. 2009]. Entretanto, em 2008, mesmo que a força de trabalho tenha mais que duplicado, o percentual de mulheres caiu para cerca de 25% [Coder et al. 2009]. Tratando de software livre, de acordo com [Ghosh et al. 2002], em 2002 menos de 1,5% dos colaboradores eram mulheres. O cenário não mudou: de acordo com o relatório da UNESCO, publicado em 2012, o percentual de mulheres participantes de software livre era de apenas 2% [UNESCO 2012].

A preocupação com a ausência de mulheres em software livre não é apenas inclusiva uma vez que, em uma sociedade baseada em TICs (Tecnologias de Informação e Comunicação), o software é produzido para o uso de ambos os sexos [Nafus et al. 2006]. Portanto, é essencial que se conheça os requisitos para atender os mais diversos tipos de usuários. Além disso, a diversidade de desenvolvedores leva ao desenvolvimento de melhores tecnologias [Powell et al. 2010].

[Lin 2005] pontua uma série de dilemas que impedem mulheres de colaborarem efetivamente com software livre. Um dos dilemas trata da falta de modelos femininos a serem seguidos em uma comunidade. Uma vez que o percentual de mulheres em comunidades de software livre é baixo, é difícil para novatas enxergarem as mulheres presentes na comunidade.

Para minimizar este problema, uma das soluções mais utilizadas é a criação de grupos de apoio e inclusão direcionados para mulheres. [Haché et al. 2011] justifica o fenômeno de criação desses grupos como uma iniciativa para que mulheres se sintam mais confortáveis em comunidades de software livre. Além disso, grupos tendem a destacar as contribuições realizadas pelas mulheres e conseqüentemente dão maior visibilidade às mulheres da comunidade.

Com base na observação desse fenômeno, este trabalho parte da hipótese de que grupos de mulheres em comunidades de software livre exercem influência sobre o Sentimento Psicológico de Comunidade (SPC) das participantes. O SPC se trata do sentimento de fazer parte de uma rede de suporte mútuo, de que se pode depender e que está sempre disponível [Sarason 1977]. Dessa forma, acredita-se que os grupos de mulheres em comunidades de software livre auxiliam na inclusão e retenção de mulheres nessas comunidades, uma vez que age como uma entidade de apoio, aumentando seu vínculo com a comunidade. Entretanto, não foram encontradas referências que tratem da influência desses grupos nas comunidades de software livre.

Para analisar esta hipótese, o objetivo desta pesquisa em andamento é responder à questão “*Grupos de inclusão e retenção de mulheres exercem influência sobre o engajamento e/ou sentimento de pertença de mulheres em comunidades de software livre?*”. Para isso, propõe-se uma pesquisa quali-quantitativa que será respondida por mulheres que participam de comunidades de software livre. A pesquisa divide as comunidades de software livre em dois grupos: **Grupo 1**, formado por comunidades que não possuem uma iniciativa formal para inclusão/retenção de mulheres; e **Grupo 2**, formado por comunidades que possuem uma iniciativa formal para inclusão/retenção de mulheres. O objetivo de delimitar esses dois grupos é comparar o SPC das mulheres pertencentes a cada grupo, visando encontrar evidências empíricas da influência que esses grupos exercem sobre o SPC.

O restante do trabalho está organizado da seguinte maneira: a seção 2 apresenta o método de pesquisa que será utilizado para atingir o objetivo preestabelecido e comprovar a hipótese. A seção 3 apresenta as ações já realizadas bem como as futuras, assim como o que se espera obter de resultados. Por fim, na seção de referências, as referências utilizadas neste trabalho.

2. Método de Pesquisa

Para responder à questão de pesquisa “*Grupos de inclusão e retenção de mulheres exercem influência sobre o engajamento e/ou sentimento de pertença de mulheres em comunidades de software livre?*”, um conjunto de ações foram definidas, como pode ser visto na Figura 1.



Figura 1. Fluxo de execução das ações propostas

A primeira etapa do trabalho consistiu na formulação de um questionário que in-

daga sobre como iniciativas de inclusão auxiliam na entrada e retenção de mulheres em comunidades de software livre. O questionário foi elaborado com base no instrumento de pesquisa definido por [Obst 2004]. O questionário estará disponível em dois idiomas, inglês e português, sendo que o primeiro idioma foi escolhido por ser geralmente utilizado na comunicação em comunidades de software livre, o que aumenta a abrangência da pesquisa e o segundo idioma foi escolhido por se tratar da língua nativa dos pesquisadores.

O questionário está dividido em quatro partes. A primeira visa obter dados do perfil da participante, contendo questões relativas à idade, escolaridade, estado civil, entre outras. A segunda parte procura saber o grau de identificação da participante em relação à seu gênero. A terceira parte procura saber o quão importante a participante considera ser membro de uma comunidade de software livre. Para a segunda e terceira parte, foi utilizada escala Likert [Likert 1932], contendo 15 afirmações para cada parte. Por fim, a quarta parte contém duas perguntas abertas, que servem como um espaço para que a participante se expresse livremente, fornecendo sua opinião de forma mais abrangente a respeito do conteúdo da segunda e da terceira parte. O questionário será disponibilizado online, em um link que será posteriormente enviado às participantes convidadas.

A segunda atividade consiste em encontrar as comunidades de software livre do **Grupo 1**. Para encontrar as comunidades, serão analisados projetos que estejam hospedados no GitHub¹. Tal plataforma foi escolhida por possuir uma quantidade significativa de projetos hospedados. A escolha de comunidades do **Grupo 1** será feita com base nos critérios a seguir:

- **1 - Comunidades consolidadas:** para esse trabalho, considera-se uma comunidade consolidada aquela que possui oitenta contribuidores ou mais e que estejam hospedados na plataforma GitHub há pelo menos dois anos. É esperado que existam comunidades mais antigas do que a data de sua hospedagem no GitHub. Por esta razão, a data servirá apenas como um filtro inicial para a busca. Caso não sejam encontradas comunidades suficientes para a execução do estudo, a busca será expandida.
- **2 - Não possuir iniciativas de inclusão de mulheres:** as comunidades do **Grupo 1** são aquelas que não possuem grupos para inserção de mulheres em software livre, portanto, a existência ou não dessas iniciativas será investigada.

Tendo escolhido as comunidades do **Grupo 1**, é necessário escolher as comunidades do **Grupo 2**. Para encontrar essas comunidades, serão utilizados dois recursos: a plataforma GitHub e o mapa colaborativo “Mapa das Mina”². O “Mapa das Mina” consiste em um mapa colaborativo que possui marcações sobre localizações onde existem projetos de mulheres relacionados a empreendedorismo, tecnologia, saúde entre outras áreas. Alguns dos projetos de tecnologia presente no “Mapa das Mina” são relacionados a comunidades de software livre que possuem projetos hospedados no GitHub. Logo, o mapa se torna uma ferramenta interessante para este trabalho.

As comunidades do **Grupo 2** serão escolhidas utilizando o primeiro critério de escolha do **Grupo 1** (comunidades consolidadas). A partir disso, será feita uma análise mais profunda das ações que visam a inclusão de mulheres em software livre presentes nessas comunidades. Após a seleção de comunidades, é necessário entrar em contato

¹<https://github.com/>

²<https://goo.gl/5RqMFQ>

com as mulheres que participam dessas comunidades. Essas mulheres serão convidadas a responder os questionários previamente elaborados.

Por fim, os resultados obtidos serão analisados, comparando as respostas ao questionário fornecidas pelas participantes do **Grupo 1** e do **Grupo 2**, com a finalidade de atingir o objetivo preestabelecido. Os resultados devem fornecer evidências empíricas da influência dos grupos de inclusão/retenção no engajamento das mulheres na comunidade.

3. Resultados Esperados

Este trabalho é uma pesquisa em andamento que possui como objetivo analisar a influência de grupos de inclusão e retenção de mulheres em projetos de software livre, verificando se as participantes se sentem mais engajadas com a comunidade quando esta oferece iniciativas de empoderamento. No momento da escrita deste artigo, o trabalho está na fase de escolha de comunidades de software livre de interesse. Os resultados serão divulgados em publicações futuras. Para próximas ações, é necessário que o questionário seja enviado para as colaboradoras das comunidades escolhidas. Os questionários serão enviados via email, com texto personalizado para cada colaboradora, relacionando suas contribuições com a comunidade de software livre com os objetivos da pesquisa, a fim de motivá-las a responder. Após o levantamento dos dados, os mesmos serão analisados.

A principal contribuição deste artigo é descrever o método para analisar a influência das ações de inclusão e fomentar a investigação a respeito da eficácia dessas ações. Uma vez que várias ações para inclusão feminina em software livre vêm sendo criadas, espera-se que os resultados desse estudo, quando alcançados, indiquem a influência de tais ações. Espera-se também que os resultados sirvam como base para a ampliação e manutenção das ações de empoderamento de mulheres em software livre e que influencie os demais pesquisadores da área de software livre a estudar iniciativas de inclusão de mulheres nas comunidades.

Referências

- Coder, L., Rosenbloom, J. L., Ash, R. A., and Dupont, B. R. (2009). Economic and business dimensions - increasing gender diversity in the IT work force. *Commun. ACM*, 52(5):25–27.
- Ghosh, R. A., Glott, R., Krieger, B., and Robles, G. (2002). Free/libre and open source software: Survey and study. *RA Ghosh, International Institute of Infonomics, University of Maastricht and Berlecon Research GmbH*.
- Haché, A., Cruels, E., and Vergés, N. (2011). Mujeres programadoras y mujeres hackers. uma aproximación desde lela coders.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Lin, Y. (2005). *Inclusion, Diversity and Gender Equality: Gender Dimensions of the Free/libre Open Source Software Development*. IDEA Groups.
- Nafus, D., Leach, J., and Krieger, B. (2006). Gender: integrated report of findings. *FLOSS POLS Deliverable D, University of Cambridge*.
- Obst, P. (2004). *Community connections: psychological sense of community and identification in geographical and relational settings*. PhD thesis, Queensland University of Technology.

Powell, W. E., Hunsinger, S., and Medlin, D. (2010). Gender differences within the open source community: An exploratory study. *Journal of Information Technology Management*, 21(4).

Sarason, S. (1977). *The psychological sense of community: prospects for a community psychology*. Jossey-Bass behavioral science series. Jossey-Bass.

UNESCO (2012). Unesco is looking for initiatives promoting women's role in free and open source software.

Ciência da Computação também é coisa de menina!

**Karina Mochetti, Luciana Salgado,
Ana Valladares Zerbinato, Beatriz Lopes Souza, Meirylen Rosa Emidio Avelino**

¹Instituto de Computação – Universidade Federal Fluminense (UFF)
Niterói – RJ – Brazil

{kmochetti, luciana}@ic.uff.br

{ana_zerbinato, beatrizls, meirylenerea}@id.uff.br

Resumo. *A participação de mulheres nas áreas de Computação e Tecnologia da Informação é extremamente baixa. Tal fato é preocupante já que diversidade em qualquer ambiente é de vital importância para sua evolução, principalmente no cenário atual em que os usuários passaram de consumidores para produtores de tecnologias e conteúdos digitais. Esse trabalho descreve a atividade inaugural realizada por nós na Universidade Federal Fluminense (UFF) com os alunos recém-inscritos no curso de Ciência da Computação. O objetivo principal do evento era promover uma discussão sobre o que leva as meninas a escolherem ou (não) a área de tecnologia da informação e analisar a interação entre os alunos, principalmente com as meninas, nos primeiros dias de aula.*

Abstract. *Women's participation in Computer and Information Technology field is extremely low. This fact is disturbing since diversity in any environment is vital to a good development, especially in technology where the role of end users is changing and they have gone from consumers to producers. This paper describes our project's inaugural activity at Fluminense Federal University (UFF) with the newly enrolled students in the Computer Science course. Our goal was to promote a discussion about what makes girls to choose or not the information technology field and analyze the interaction among students, especially girls in the first days of school.*

1. Introdução

A diversidade de profissionais em todas as áreas é fundamental para que seja possível representarmos todos os indivíduos da sociedade moderna e, portanto, termos soluções mais adequadas aos problemas por ela enfrentados. Diante desta afirmativa é preocupante o baixo número de mulheres atuando na área de Tecnologias da Informação. Segundo o PNAD, em 2009 somente 18,84% dos profissionais de TI eram mulheres [10].

A dificuldade de inclusão de mulheres em ambientes dominados por homens não é um problema atual, e traz à tona fatores importantes como o papel da mulher na sociedade, cujas regras são impostas e criadas por homens [9]. Mais especificamente na área de TI, temos atualmente grandes congressos e grupos formados com o objetivo de estudar esse problema e tentar revertê-lo. Podemos citar, por exemplo, o congresso Grace Hopper Celebration of Women in Computing do Instituto Anita Borg [4] e a Association for Women in Mathematics [1].

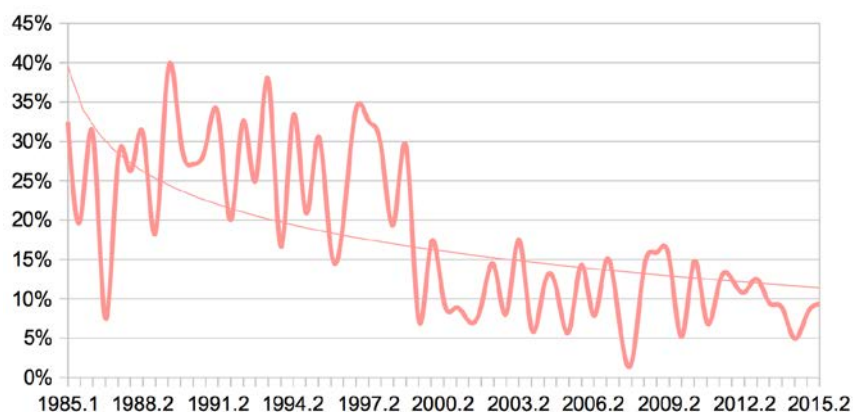


Figura 1. Gráfico de porcentagens de alunas nas turmas do curso de Ciência da Computação da Universidade Federal Fluminense (UFF).

Dentro desse contexto, diversos projetos nacionais foram criados, o principal deles sendo o Programa Meninas Digitais da Sociedade Brasileira da Computação (SBC) [6] sob coordenação dos professores Cristiano Maciel (UFMT) e Silvia Amélia Bim (UTFPR). Esse programa tem como objetivo motivar alunas do ensino médio e fundamental a seguir carreira na área de Tecnologias da Informação. Ele já possui algumas filiais regionais como o Meninas Digitais: Regional Sul [8], o Meninas Digitais: Regional Mato Grosso [7], o Emili@as [3] e o Cunhatã Digital [2], e recentemente uma filial no estado do Rio de Janeiro, o `#include <meninas.uff >` [5].

Este artigo descreve a primeira pesquisa, e seus respectivos resultados, realizada no contexto do Projeto `#include <meninas.uff >`, do Instituto de Computação (IC) da Universidade Federal Fluminense (UFF).

2. Objetivo

Como primeira tarefa do Projeto `#include <meninas.uff >`, organizamos uma atividade com os alunos recém-inscritos no curso de Ciência da Computação da Universidade Federal Fluminense (UFF), no contexto do Evento de Acolhimento dos Calouros, promovido pelo Diretório Acadêmico do curso. Dos 70 alunos inscritos, somente 8 são meninas. Nosso principal objetivo era promover uma atividade para analisar qual seria o comportamento dos alunos com relação às novas alunas e promover um debate sobre o tema.

O curso de Ciência da Computação da Universidade Federal Fluminense (UFF) existe desde 1985. Hoje o curso tem em média 520 alunos matriculados e atualmente são oferecidas 140 vagas anuais para o ingresso duas vezes ao ano. Dos mais de 3000 alunos que já passaram pelo Instituto de Computação da Universidade Federal Fluminense (UFF), somente 16% eram mulheres. Nota-se, contudo, que esse número já pequeno diminuiu muito nos últimos anos, chegando a valores alarmantes como a inscrição de somente uma menina em 2008. A Figura 1 mostra um gráfico da porcentagem de meninas durante os 30 anos de curso.

Com base nessa grave estatística, nosso projeto pretende contribuir para que esse problema seja contornado, motivando alunas do ensino médio/fundamental a seguir carreira na área de Computação e emponderando as alunas de graduação e pós-graduação já inclusas na área.

3. Metodologia

A atividade foi dividida em 3 partes principais, cada uma com o objetivo de analisar e avaliar ações e comportamentos diferentes dos alunos. Ao todo, participaram como observadoras da pesquisa cinco membros do Projeto **#include <meninas.uff>**: duas professoras e três alunas da graduação.

Ainda na fase de recrutamento dos participantes, foi pedido aos organizadores do acolhimento que não avisassem os alunos que a atividade estaria relacionada com a questão da falta de mulheres nas áreas de tecnologia da informação. Primeiramente, os alunos participantes da atividade (37 no total, dos quais 5 meninas) foram posicionados num grande círculo na intenção de que cada um se apresentasse para todos os demais. Ao final desta etapa pedimos a alguns dos participantes que dissessem o nome de algum outro colega, selecionado por nós, pesquisadoras.

A segunda etapa foi uma dinâmica em grupo. Os alunos foram divididos em 5 grupos, com uma menina por grupo. A divisão foi feita dando-se um número para cada menina na frente de todos, enquanto cada garoto ao receber seu número deveria procurar a menina correspondente. Desta forma queríamos que as meninas tivessem a função indireta de liderança, sem que este tenha sido imposto diretamente por nós. O objetivo dessa dinâmica era que cada grupo desvendasse uma dica de uma caça ao tesouro feita baseada na Cifra de César [11]. A intenção dessa etapa era analisar como os grupos se organizariam e se as meninas manteriam (ou não) a posição de liderança.

A terceira e última etapa promoveu um debate aberto com os alunos sobre os motivos para um número cada vez mais baixo de mulheres na área de Tecnologias da Informação. Foi aberto um debate sobre o porquê esse número é tão baixo e as meninas tiveram a oportunidade de dividir histórias pessoais a respeito da sua decisão de cursar uma faculdade dominada por homens.

4. Resultados

Na primeira etapa da atividade ficou clara a posição de destaque que as meninas ocupam e a diferença do tratamento dado pelos colegas. No grande círculo de apresentação, as meninas se mantiveram sempre próximas umas das outras e na hora de se apresentar não só todos os meninos fizeram total silêncio prestando muita atenção, como elas foram as únicas a receber aplausos. No momento em que foi perguntado o nome dos alunos, a maioria não sabia o nome de todos, mas ao perguntarmos o nome das meninas os próprios alunos disseram “Ah! Obviamente todos irão lembrar os nomes das meninas!”.

Esta parte da atividade nos deu evidências de como as meninas ganham destaque, querendo ou não, simplesmente por fazerem parte de uma minoria. Apesar de poder ser benéfico para algumas, isso pode ser um grande obstáculo para outras durante sua carreira, sabendo que sempre estão sendo observadas e analisadas por professores e colegas.

Na segunda etapa, os grupos foram divididos de forma que as meninas ficassem indiretamente numa posição de liderança. Elas davam nome aos grupos, elas tinham o número dos grupos e era responsabilidade dos meninos achá-las para formar seus grupos. No entanto, em nenhum momento nós verbalizamos que as meninas iriam liderar ou coordenar os grupos. Nesta fase observamos como a dica (escrita em um pedaço de papel) rapidamente foi tirada da mão das meninas. Em um grupo, apesar de ter sido en-

tregue para a menina, um menino logo interferiu, segurando o papel ele mesmo sem que a menina sequer tocasse. Na maioria dos grupos não houve uma liderança propriamente dita, mas das cinco meninas, podemos dizer que três mantiveram essa posição durante a atividade.

Na etapa final os alunos foram convidados a compartilhar posições e experiências sobre os motivos de atualmente haver tão poucas mulheres na computação e esse número estar cada vez menor. Essa parte foi a mais rica de toda a atividade com grande participação tanto dos meninos quanto das meninas.

As opiniões foram dadas livremente e sem uma ordem pré-determinada de quemalaria primeiro. Observamos que tanto meninos quanto meninas se interessaram em dar suas opiniões sobre o tema apresentado. Um dos alunos comentou: “Provavelmente essa é a primeira vez que alguns meninos estão ouvindo e conversando sobre esse assunto”. Para as meninas foi claramente muito bom poder dividir um pouco das suas experiências e medos com outros alunos. Elas nos contaram que há uma cobrança da sociedade para que elas não façam cursos dominados por homens, com familiares e amigos as alertando para o problema de ser um curso “para homens” e as incentivando a fazer um curso na área de humanas. Elas também reclamaram da falta de exemplos de mulheres engenheiras e na área de exatas com uma delas deixando claro como a mãe dela, engenheira mecânica, foi uma grande influência e incentivo para que ela entrasse num curso de exatas.

Também foi compartilhado o medo e receio que elas têm principalmente durante os primeiros dias de aula com os trotes e as atividades de acolhimento. Elas disseram que procuram logo outra menina e se sentem mais seguras próximas umas das outras. Esse fato pode ser confirmado já que durante toda a atividade as meninas estavam sempre que possível num grupo próximas umas das outras e nunca divididas entre os meninos.

Após os depoimentos das meninas, alguns meninos comentaram que nunca haviam considerado a posição deles no curso. Um dos alunos chegou a dizer que agora eles ocupavam uma posição de opressor, mesmo que a maioria fosse oprimida no ensino médio por ser mais estudioso ou por qualquer outra característica, agora eles são os opressores das meninas por estarem em maioria. Segundo ele era preciso “prestar atenção no machismo que a gente comete sem perceber”. Um outro aluno disse achar que “as mulheres são interessadas naturalmente por humanas” já que em Portugal, onde ele estudou, há aulas de programação no ensino médio e isso não aumentou o número de meninas na computação. Isso logo foi discordado pelas meninas e por grande parte dos meninos presentes na discussão.

5. Conclusão

A primeira atividade organizada pelo Projeto **#include <meninas.uff>** mostra como é importante um trabalho não só com as mulheres, mas também com os homens. Criar um debate com grande participação masculina pode ajudar na conscientização deles do papel de opressor que eles têm como maioria, o que pode contribuir para a diminuição da hostilidade no meio, atraindo mais meninas e, principalmente, reduzindo o número de desistências por parte das alunas. Adicionalmente, tivemos inúmeras evidências verbais do quanto as meninas são desmotivadas pela sociedade (inclusive pela família) a procurar carreira na área de exatas. Como trabalhos futuros pretendemos estender esta investigação com as alunas já veteranas do curso de Ciência da Computação.

Referências

- [1] Association for Women in Mathematics.
sites.google.com/site/awmmath.
- [2] Cunhatã Digital.
www.facebook.com/cunhantadigital.
- [3] Emili@as.
<https://www.facebook.com/emiliasarmacaoembits>.
- [4] Grace Hopper Celebration of Women in Computing.
ghc.anitaborg.org.
- [5] #include <meninas.uff >.
www.facebook.com/include.meninas.uff.
- [6] Meninas Digitais.
sbcmt.wordpress.com/meninasdigitais/.
- [7] Mulheres Digitais: Regional Mato Grosso.
www.facebook.com/MeninasDigitaisMT.
- [8] Mulheres Digitais: Regional Sul.
www.facebook.com/meninasdigitaisufsc/.
- [9] Simone de Beauvoir. *O Segundo Sexo*. Nova Fronteira, 1949.
- [10] Bárbara Geraldo de Castro. *Afogados em contratos : o impacto da flexibilização do trabalho nas trajetórias dos profissionais de TI*. PhD thesis, UNICAMP, 2013.
- [11] Dennis; Gordon Prichett Luciano. Cryptology: From caesar ciphers to public-key cryptosystems. *The College Mathematics Journal*, 18, 1987.

Computasseia: destacando a participação feminina na História da Computação

Karen da Silva Figueiredo, Jean Carlos Oliveira Santos

Instituto de Computação – Universidade Federal do Mato Grosso (UFMT)
Cuiabá, Mato Grosso, Brasil

karen@ic.ufmt.br, jcoliveira93@gmail.com

Abstract. *This paper presents a card game for teaching History of Computing at different levels of education. The proposed game includes equanimous content about women's contribution to Computing.*

Resumo. *O presente trabalho apresenta um jogo de cartas para o ensino de História da Computação em diferentes níveis de ensino que inclui de forma equânime conteúdos sobre a contribuição feminina para a Computação.*

1. Introdução

É fundamental para estudantes de Computação a compreensão não somente dos aspectos técnicos, mas também dos aspectos culturais da sua área (e.g. a História da Computação) para a evolução dos seus estudos acadêmicos [Berry *et al.* 2011]. Segundo Fonseca Filho (2007), estudar a História da Computação é essencial para conhecer os fundamentos da área, incentivar à educação para o desenvolvimento de qualidade, tornar claros e relacionar os fatos históricos, acompanhar novas tendências e valorizar o potencial humano.

Atualmente, cursos e carreiras em Computação e áreas afins contam com uma baixa participação feminina [Abbate 2012]. Pesquisas, ações afirmativas e outras estratégias vêm sendo empregadas por diversos grupos nacionais e internacionais para atrair mais mulheres para a área. A presença e a obra de várias mulheres na ciência são geralmente pouco reconhecidas e valorizadas, seus nomes e contribuições são frequentemente “apagados” da história da ciência [Oreskes 1996] e com a Computação não é diferente, e.g. as programadoras do ENIAC foram reconhecidas historicamente apenas décadas depois de sua atuação¹. Desta forma, ao ensinar História da Computação é indispensável que a participação feminina seja incluída, não só pelo respeito à veracidade histórica, mas também como estratégia para a atração de mais mulheres para a Computação. A possibilidade de atração de novos talentos para a Computação através da história é possível pelo fator inspirador que o ensino das histórias dos grandes feitos computacionais e personalidades podem promover [Fonseca Filho 2007].

O objetivo deste trabalho é apresentar um jogo sério para o ensino de História da Computação chamado Computasseia. Computasseia é um jogo de cartas de estratégia elaborado como recurso educacional de apoio ao ensino de História da Computação, pensando na dificuldade encontrada algumas vezes por educadores ao trabalharem de

¹ <http://eniacprogrammers.org/>

forma atrativa para os alunos tais temáticas históricas repletas de nomes, fatos e datas importantes. O jogo entra no processo de ensino-aprendizagem como ferramenta lúdica, auxiliando na aprendizagem e retenção de conteúdos pelos alunos. No Computasseia são apresentados de forma equânime conteúdos sobre a contribuição feminina para a Computação, ressaltando para o jogador informações que muitas vezes são colocadas em segundo plano e mostrando que as mulheres também somam parte importante para a História da Computação.

2. Computasseia

O Computasseia [Santos e Figueiredo 2016] é um jogo de cartas de estratégia desenvolvido, prioritariamente, para dar suporte ao ensino da História da Computação em diferentes níveis de ensino (técnico, graduação e pós-graduação). O jogo é atualmente composto por um conjunto de 80 cartas distintas², cada uma representando uma parte significativa da História da Computação. Cada carta do jogo pertence a uma Categoria (Personalidades, Instituições e Eventos) e a uma ou mais Áreas do Conhecimento (Hardware, Software, Gestão de Dados, Teoria Computacional e Matemática e Computação e Sociedade), a fim de melhor organizar os conteúdos e trazer mais um nível de informação ao jogador.

O conjunto de cartas do Computasseia conta com 30 cartas de Personalidades para representar pessoas que contribuíram de forma significativa com teorias, invenções, ideias ou princípios para pelo menos uma das Áreas do Conhecimento. Destas, 15 cartas são sobre personalidades femininas, a citar: Ada Lovelace, Grace Hopper, Hedy Lamarr, Jean Jennings Bartik, Kathleen Booth, Carol Shaw, Susan Kare, Sophie Wilson, Adele Goldberg, Emmy Noether, Marissa Mayer, Ginni Rometty, Margaret Whitman, Shafrira Goldwasser e Anita Borg. Além destas cartas, estão presentes no jogo de forma equânime cartas de Eventos e Instituições que abordam criações femininas e locais em que a atuação feminina foi destaque, *e.g.* Primeiro Algoritmo Processado por Máquina, Primeiro Compilador, Bletchley Park e The Anita Borg Institute.

As cartas do Computasseia são compostas dos seguintes elementos gráficos que podem ser visualizados na Figura 1: **1) Título** - o nome da carta, indicando de forma direta o que a carta representa; **2) Imagem ilustrativa** - representação imagética da carta; **3) Área do Conhecimento** - a área ou áreas associadas à carta; **4) Categoria** - a categoria da carta (Personalidade, Instituição ou Evento); **5) Descrição** - descrição textual que introduz o conceito da carta, complementando a informação do título e expandindo sua contextualização histórica e relevância computacional; e **6) Data** - marco temporal que está associado ao fato histórico que a carta representa.

Todas as cartas e demais informações sobre o jogo Computasseia estão disponíveis no site do jogo³.

A dinâmica do jogo Computasseia consiste na elaboração conjunta pelos jogadores de uma Linha do Tempo⁴ de marcos da história da Computação, formada pela

² O conjunto de cartas do jogo pode ser estendido, para incluir novas cartas no futuro, seguindo a padronização do jogo.

³ <http://gameinclass.ic.ufmt.br/computassy/>

⁴ Dinâmica inspirada no jogo infanto-juvenil Timeline da Galapagos Jogos.

disposição sequencial em ordem crescente da Data das cartas, constituindo uma cronologia. Cada jogador recebe no início do jogo um conjunto de seis cartas atribuídas de forma aleatória. As demais cartas restantes do jogo são empilhadas em um montante denominado Montante de Compra com a visão da frente voltada para cima. O Computasseia comporta de dois a seis jogadores devido à sua quantidade de cartas atual.



Figura 1. Exemplo de carta de Personalidade feminina do jogo Computasseia e seus elementos indicados

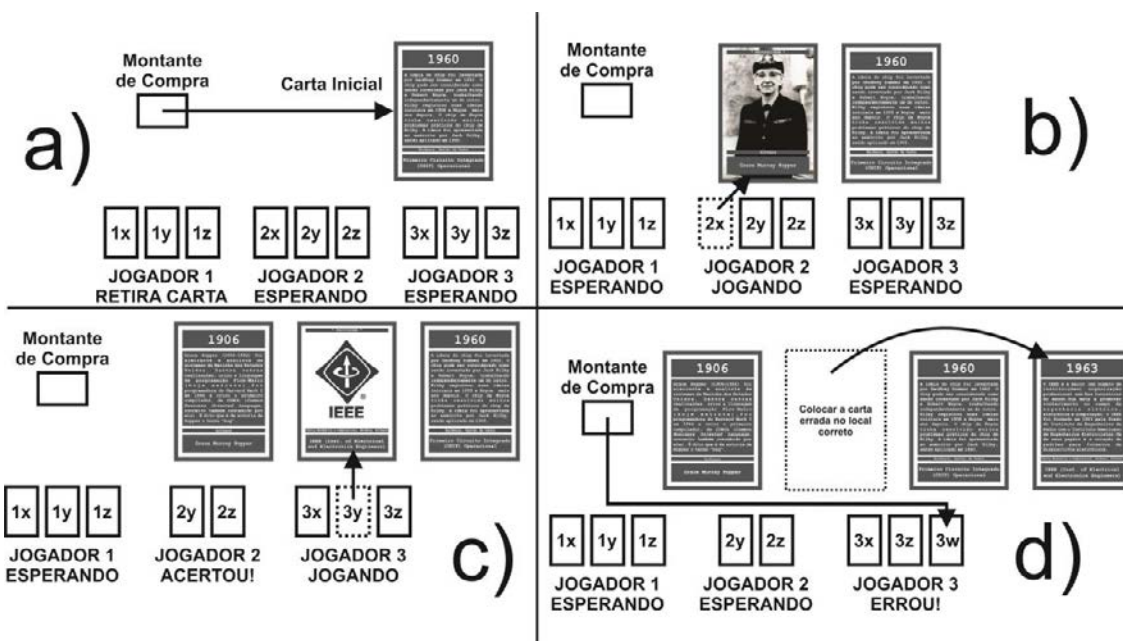


Figura 2. Passos demonstrativos de uma jogada no Computasseia

Ao receber as cartas, o jogador que as recebe deve continuar observando suas cartas apenas a partir da visão frontal das mesmas, sem observar o lado das cartas que

contém a Data e a Descrição. Os jogadores se posicionam em volta do local em que será montada a Linha do Tempo (ver Figura 2a). Um por um, em sentido escolhido pelos jogadores, o jogador da vez deverá escolher uma das cartas em sua mão para utilizar em sua jogada. Assim que fizer a escolha, deverá tentar inseri-la na posição que julga ser a correta na Linha do Tempo (ver Figura 2b). Feito isto, a face da carta jogada é virada, revelando suas informações de Data e Descrição contidas no verso. Logo, é verificado o acerto ou erro da jogada (ver Figura 2c e d).

Caso o jogador tenha errado a posição da carta, essa carta é mantida na Linha do Tempo, porém a mesma deve ser recolocada na sua posição apropriada e o jogador deve retirar uma nova carta no Montante de Compra para as suas cartas da mão (ver Figura 2d). A vez de jogada é passada ao próximo jogador. O jogo acaba quando um dos jogadores consegue descartar de forma apropriada todas as cartas de sua mão para a Linha do Tempo do jogo, sendo o mesmo declarado vencedor da partida, ou quando o Montante de Compra se esgota - neste caso, o vencedor da partida é o jogador (ou jogadores) com o menor número de cartas em mão. A Figura 2 apresenta passos ilustrativos de uma partida de Computasseia.

3. Conclusões

Este trabalho apresentou a versão inicial de um jogo sério de cartas e estratégia para o ensino de História da Computação chamado Computasseia, que aborda fatos importantes de diferentes áreas do conhecimento da Computação, incluindo as contribuições femininas para a História da Computação distribuídas de forma equilibrada no jogo. Os autores do presente trabalho acreditam que o jogo proposto é um recurso acessível aos educadores (podendo ser facilmente impresso no site do jogo e utilizado até de outras maneiras – e.g. estilo “jogo da memória”, imprimindo frente e verso como cartas separadas), podendo funcionar como estratégia motivacional para atrair mais mulheres para a área e para que os jogadores em geral conheçam e valorizem a participação feminina na construção do passado, presente e futuro da Computação.

Referências

- Abbate, J. (2012) “Recoding gender: women's changing participation in computing”, MIT Press, 264p.
- Berry, G., Sheard, J., Quartly, M. (2011) “A Virtual Museum of Computing History: an educational resource bringing the relationship between people and computers to life”, In: 13th Australasian Computing Education Conference-Volume 114, Australian Computer Society, Inc., p. 79-86.
- Fonseca Filho, C. (2007) “História da computação: O Caminho do Pensamento e da Tecnologia”, EDIPUCRS, Porto Alegre, 205 p.
- Oreskes, N. (1996) “Objectivity or heroism? On the invisibility of women in science”, In: *Osiris*, 11, p. 87-113.
- Santos, J. C. O. e Figueiredo, K. S. (2016) “Computasseia: Um Jogo para o Ensino de História da Computação”, In: 24° WEI - Workshop sobre Educação em Computação, XXXVI Congresso da Sociedade Brasileira de Computação, Porto Alegre.

Cunhantã Digital: programa de incentivo à participação de mulheres da região amazônica na Computação e áreas afins

Tanara Lauschner, Rosiane de Freitas, Fabíola Nakamura, Ludymila Lobo

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Manaus – AM – Brasil

{tanara, rosiane, fabiola, llag}@icomp.ufam.edu.br

Resumo. *Este artigo apresenta o Programa Cunhantã Digital, que envolve ações e projetos voltados ao incentivo à participação de mulheres da região amazônica na área de ciência e tecnologia. Tal programa contempla atividades como a realização de palestras e atividades lúdicas em eventos científicos e a capacitação de alunas para participar de competições de programação e desenvolver aplicativos computacionais, dentre outras. Algumas ações já resultaram em destaques obtidos em competições de programação e prêmios obtidos por meio do desenvolvimento de aplicativos, além de terem proporcionado um ambiente acadêmico mais propício à discussão de questões relacionadas ao gênero feminino nas áreas de computação e exatas na comunidade amazonense. Os próximos passos visam atingir também escolas de ensino fundamental e médio, para que as mulheres se sintam motivadas a escolher suas profissões independentemente dos pré-conceitos sociais.*

Abstract. *This article presents the Cunhantã Digital program involving actions and projects aimed at encouraging the participation of women in Science and Technology in the Amazon region. This program includes activities such as lectures at scientific meetings, training students to participate in competitions and develop computer applications. These actions resulted in highlights obtained in programming contests and prizes obtained through application development, and have provided a more conducive academic environment to discuss issues related to the female gender in the science and technology fields. The next steps are also intended to include primary and secondary schools so that women feel motivated to choose your career regardless of social prejudices.*

1. Introdução

Estatisticamente, o percentual de mulheres nas áreas de Computação no Brasil é expressivamente menor que o de homens, representando uma das menores porcentagens de mulheres quando comparado a todos os outros cursos universitários [IPEA 2011]. Diante deste cenário, os estudos mostram diferentes motivos que podem ser apontados como causas para a disparidade entre a proporção de gêneros nestas áreas.

Alguns estudos apontam sexismo e cultura negativa como as principais causas de afastamento de mulheres de áreas tecnológicas [Hanton 2015]. O sexismo seria um fator que causa um ambiente não favorável à adesão e permanência de mulheres no setor. Já a

cultura negativa, refere-se aos costumes sociais que vão desde o tipo de atividade lúdica realizada na infância até a expectativa da sociedade, e acabam ocasionando o desinteresse pela área, já que fatores biológicos não podem ser apontados como causa. Outros estudos [Medeiros e Borges 2014], a partir de análises em conversas de blogs de tecnologia, apontam o estereótipo negativo como sendo um dos principais fatores que repelem as mulheres nesta área.

Outro aspecto que pode ser considerado como desestimulante às alunas novatas na graduação é o grau de inibição em questionar o professor em sala de aula, conforme apresentado por Sales *et. al.* [Sales et al. 2014] em uma amostragem na qual 61% das alunas correspondeu à categoria "muito inibido", contra 29% do total em alunos. Complementando este fator, a pesquisa também demonstrou maior ausência de experiências anteriores relacionadas à lógica e à programação, na qual 37% dos homens já possuíam o conhecimento prévio, em contraste com apenas 17% do total de mulheres. Sendo assim, o grau de dificuldade apontado pelas mulheres foi maior que o dos homens cursando a mesma disciplina de Computação. A falta de incentivo na infância e adolescência das meninas a jogos e brinquedos que estimulam a lógica e/ou experiência em programação também é apontado como um fator que influencia nesta dificuldade inicial. Este estudo, apresentado na *National Science Foundation* (NSF), conclui que justamente na era em que o computador passou a ser pessoal, a frequência de acesso aos computadores passou a ser predominantemente masculina, devido ao marketing realizado na época ser direcionado aos homens, associando o universo tecnológico como algo voltado ao público masculino, fator que gerava desinteresse na maioria das mulheres. Assim, a NSF também aponta que a falta de conhecimento anterior à realização do curso de computação era um fator prejudicial.

Tendo em vista a ocorrência de tais problemas, a literatura apresenta diversas propostas para a solução, como a promoção de movimentos que visam a inserção de mulheres na área da computação realizado por lideranças do país, instituições de ensino e empresas [IPEA 2011], a propagação da ampla discussão sobre o assunto e estudos de medidas sobre como reverter o cenário [Medeiros and Borges 2014]. Para contribuição com tal causa, foi proposto o programa Cunhantã Digital, abordado neste trabalho, baseado em casos de grande impacto positivo proporcionado por ações voltadas para mulheres no desenvolvimento de aplicativos e participações em competições de programação científicas e tecnológicas [de Freitas et al. 2016] [MSR 2016], e na participação de ações nacionais da Sociedade Brasileira de Computação, por meio do *Workshop Women In Technology* (WIT) e o projeto Meninas Digitais [Maciel and Bim 2016].

O restante do artigo está estruturado como segue. Na Seção 2 são apresentados alguns programas e ações pelo Brasil e mundo, relacionados ao mesmo tema. Na Seção 3 são abordados projetos, ações e casos de sucesso do Programa Cunhantã Digital. E, na Seção 4 são feitas as considerações finais, na qual se resume os resultados e repercussão obtidos pelo programa, bem como as próximas ações a serem executadas.

2. Ações e programas no Brasil e pelo mundo

Existe um movimento mundial para promover uma maior participação de mulheres em áreas de ciências e tecnologia ou STEM (*Science, Technology, Engineering and Mathematics*), na qual as principais sociedades científicas possuem setores específicos para ações

e estudo de tais questões, como a ACM-W (*ACMW supporting, celebrating and advocating for Women in Computing*) [ACM-W 2016], IEEE-W (*IEEE Women in Engineering (WIE)*) [IEEE-W 2016] e NCWIT (*National Center for Women Information Technology*) [NCWIT 2016]. A principal conferência científica especialmente sobre o tema é a *Grace Hopper Celebration of Women in Computing* [GHC 2016].

No Brasil, existe o Programa Meninas Digitais da Sociedade Brasileira de Computação [Maciel and Bim 2016], que tem como objetivo divulgar a área de Computação para despertar o interesse de estudantes do ensino médio/tecnológico ou dos anos finais do ensino fundamental, para que conheçam melhor a área e, desta forma, motivá-las a seguir carreira em Computação. Através deste programa, são realizados projetos como Meninas Digitais, MD-MT e elas++ [Maciel and Bim 2016]. A principal conferência científica sobre o tema no Brasil é o *Women In Technology (WIT)* realizado como parte do principal congresso de Computação do Brasil, o CSBC (Congresso da Sociedade Brasileira de Computação), e que está em sua 10ª edição neste ano de 2016. Para a América Latina, há o *Latin American Women in Computing Congress (LAWCC)*, evento satélite do CLEI (*Latin American Computing Conference*), um evento mais recente e que também está crescendo cada vez mais.

3. Sobre o programa Cunhantã Digital e suas ações

O Programa Cunhantã Digital sedimenta no Amazonas este movimento mundial crescente envolvendo ações para um maior envolvimento da mulher em áreas de ciência e tecnológica. Foi criado após várias iniciativas de sucesso envolvendo mulheres, em especial, em competições de programação científica e tecnológicas, bem como no desenvolvimento de aplicativos, como parte do Projeto *SciTechGirls* [de Freitas et al. 2016]. Nas subseções a seguir, será apresentado as ações de divulgação e conscientização realizadas por meio deste programa, tais como workshop e fóruns.

3.1. I Workshop Cunhantã Digital

O Workshop Cunhantã Digital teve sua primeira edição em 21 de agosto de 2015, com duração de 6 horas úteis, cujo objetivo consistiu em estimular a maior participação das mulheres da região amazônica em áreas de ciência e tecnologia, promover a interação entre mulheres profissionais e meninas em formação, bem como estimular o ingresso de mulheres em cursos de graduação e carreiras de tais áreas. Teve a participação de mais de 200 pessoas, a maioria mulheres estudantes ou profissionais, mas também, para grande surpresa, contou com o interesse e participação de um número significativo de homens, representando 22% do total, que buscaram se inteirar mais desta problemática feminina e discutir questões de gênero na academia e mercado de trabalho. Houve palestras ministradas por pesquisadoras convidadas, dinâmicas de grupo, apresentação de aplicativos e casos de sucesso de algumas alunas de destaque e, também, um debate envolvendo mulheres de diversos segmentos, como política, indústria e academia. Uma nova edição está prevista para o segundo semestre deste ano de 2016.

3.2. I Fórum Cunhantã Digital

Com base no sucesso do Workshop, foi realizado em janeiro deste ano de 2016 o I Fórum do Cunhantã Digital, voltado para educadores, de tal modo a se discutir e planejar ações

para definição e planejamento de atividades a serem realizadas nas escolas do ensino fundamental e médio, bem como nos cursos universitários. O evento contou com muitos educadores da rede estadual e municipal de ensino, e das principais universidades da capital Manaus, e de alguns outros municípios do Amazonas, como Itacoatiara, Parintins, Benjamim Constant e Coari. Também contou a Superintendente da Zona Franca de Manaus (SUFRAMA), órgão máximo para gestão das políticas e fomentos do Pólo Industrial do Amazonas, o que demonstra o prestígio alcançado pelo movimento Cunhantã Digital e o seu potencial de crescimento e fortalecimento de suas ações. Além do evento em si, é importante registrar o grande interesse da imprensa local pelo tema.

3.3. Histórias de sucesso e formas de engajamento

Apesar do pouco tempo de criação, já houve alguns casos de sucesso envolvendo alunas participantes de projetos e ações do Programa Cunhantã Digital, que são descritos em seguida.

Na Maratona de Programação 2015, promovida no Brasil pela SBC e mundialmente pela ACM - ICPC [MP 2016] [ICPC-ACM 2016], uma equipe do Instituto de Computação (IComp) da UFAM (Universidade Federal do Amazonas), composta somente por mulheres, se classificou para as finais brasileiras da competição, sendo a de melhor desempenho dentre as 08 (oito) equipes formadas somente por mulheres, em um total de mais de 600 equipes (com enorme predominância de homens). Também, algumas alunas individualmente se destacaram em diversas competições e eventos científicos. Houve a participação com destaque das 3 (três) únicas edições do *International Women's Hackathon* promovida pela Microsoft Research [MSR 2016], entre 2013 e 2014, culminando com o desenvolvimento de aplicativos de temática feminina: *Mommy's BeneFIT*, *Make UPhi*, *What the Hack* e *How To Help*.

4. Considerações finais

O Programa Cunhantã Digital tem sido muito bem recebido pela comunidade amazônica, já sendo observado um impacto muito positivo e que envolve educadores, profissionais e estudantes em torno da temática voltada para o aumento da participação feminina em cursos e profissões das áreas de ciência e tecnológica, com especial ênfase às áreas de Computação. Um canal de interação entre os educadores e profissionais promotores do programa e os jovens estudantes interessados no movimento (a maioria mulheres, mas, também, muitos homens) tem sido mantido via rede social Facebook [Lauschner et al. 2016]. Como atividades em andamento, tem sido realizadas competições de programação e sendo promovidas atividades envolvendo o desenvolvimento de aplicativos [de Freitas et al. 2016]. Como próximos passos, já está prevista a realização do II Workshop e do II Fórum do programa Cunhantã Digital, bem como a realização de um ciclo de palestras a serem ministradas por alunas universitárias e profissionais da área em escolas do ensino médio e fundamental. Também houve um convite para o programa ser apresentado na 7a. Feira Norte do Estudante, que costuma receber mais de dez mil estudantes do ensino fundamental e médio, na qual é oportunizado aos alunos conhecerem o curso que desejam escolher, e onde com isto se espera motivar também as meninas estudantes a conhecerem mais os cursos e carreiras de Computação e áreas afins, colaborando para o aumento do ingresso de meninas em cursos universitários destas áreas.

Referências

- ACM-W (2016). Acm-w supporting, celebrating and advocating for women in computing. Association for Computing Machinery (ACM). Disponível em <https://women.acm.org/>.
- de Freitas, R., Conte, T., Lobo, L., and Salles, J. (2016). Projeto scitechgirls. Instituto de Computação, Universidade Federal do Amazonas. Disponível em <http://scitechgirls.icomp.ufam.edu.br/>.
- GHC (2016). Grace hopper celebration of women in computing. Anita Borg Institute. Disponível em <http://ghc.anitaborg.org/>.
- Hanton, P. B. (2015). The lack of women in technology: The role culture and sexism play. Master's thesis, Master's Capstone Theses.
- ICPC-ACM (2016). International collegiate programming contest. Association for Computing Machinery (ACM). Disponível em <https://icpc.baylor.edu/>.
- IEEE-W (2016). Ieee women in engineering (wie). Institute of Electrical and Electronics Engineers, (IEEE). Disponível em http://www.ieee.org/membership_services/membership/women/index.html.
- IPEA (2011). Retrato das desigualdades de gênero e raça. Instituto de Pesquisa Econômica Aplicada. 4a. edição.
- Lauschner, T., Nakamura, F., de Freitas, R., Conte, T., Lobo, L., and Cruz, M. (2016). Programa cunhantã digital. Instituto de Computação, Universidade Federal do Amazonas. Disponível em <https://www.facebook.com/cunhantadigital/?fref=ts>.
- Maciel, C. and Bim, S. A. (2016). Programa meninas digitais - ações para divulgar a computação para meninas do ensino médio. In *Anais do Computer on the Beach*. Sociedade Brasileira de Computação. Disponível em <http://www.computeronthebeach.com.br/arquivos-2016/Anais completos - Computer on the Beach 2016.pdf>.
- Medeiros, C. and Borges, J. (2014). Abram-se Às mulheres todas as portas!": Conversas em blogs de mulheres em carreira de ti. Em *Blogs De Mulheres Em Carreira De TI. Gestão Contemporânea*.
- MP (2016). Maratona de programação. The ICPC-ACM South America - Brazil. Sociedade Brasileira de Computação. Disponível em <http://maratona.ime.usp.br/>.
- MSR (2016). Microsoft research. Disponível em <http://research.microsoft.com/>.
- NCWIT (2016). National center for women and information technology. Disponível em <https://www.ncwit.org/>.
- Sales, A., Calado, B., Silva, D., Mattos, G., and Moreira, J. (2014). Dificuldades para o ingresso e permanência na ciência e engenharia da computação: Um olhar feminino. 18º Redor - Universidade Federal Rural de Pernambuco - Recife.

Curto-Circuito na Escola

Kaleandra C. Almeida, Ana Carolina Marcelino, Antony B. Turra, Luciana B. Frigo, Eliane Pozzebon

LabTeC– Universidade Federal de Santa Catarina(UFSC)
Araranguá – SC – Brazil

meninasdigitaisufsc@gmail.com, luciana.frigo@ufsc.br

***Abstract.** This paper describes a report on experiences about electric circuit workshops for high school girls. The workshop has basic concepts about electricity where experimentation is the key point for building dynamic activities. The girls are divided into small groups and they use didactic kits with colorful pieces that can be assembled and connected in many ways, making the experience easy and secure. This workshop is part of a project that aims to present various professions related to technology, engineering and computing that can offer great job opportunities for women who are interested in any of these areas.*

***Resumo.** Este artigo apresenta a descrição de um relato de experiência de uma oficina de circuitos elétricos para alunas do ensino médio de escola pública. A oficina abordou conceitos básicos de eletricidade, onde a experimentação é ponto chave para a dinâmica das atividades. As alunas foram divididas em pequenos grupos e foram utilizados kits didáticos com o conceito de blocos de encaixar, o que torna a atividade mais fácil e segura. Esta oficina faz parte do escopo de um projeto que tem por finalidade apresentar diversas profissões relacionadas a tecnologia, engenharia e computação que podem oferecer ótimas oportunidades de trabalho para as mulheres que tiverem interesse por alguma destas áreas.*

1. Introdução

Aulas práticas devem ser o ambiente propício para o questionamento e para a transformação dos fatos em evidências. Este artigo apresenta um relato de experiência sobre a utilização do kit educacional de circuitos elétricos Snap Circuits em uma oficina realizada com alunas do ensino médio de escola pública apresentando conceitos básicos de eletricidade, com ênfase na associação de resistores e capacitores. A oficina faz parte de uma série de atividades extracurriculares, com alunas do ensino médio, com o intuito de apresentar profissões que envolvam tecnologia e computação. Acredita-se que muitas meninas desconhecem as carreiras nas áreas de tecnologia o que acaba excluindo estas áreas no momento da tomada de decisão para uma escolha profissional.

O ensino de circuitos elétricos na prática é descrito no projeto de (HERTZ & HAYER, 2013) como uma atividade que pode auxiliar as minorias a como um meio para aumentar sentimentos de empoderamento, autoestima e encantamento pela busca do conhecimento.

Outro artigo que trata de tema semelhante é de Rebelo e Ramos (2009) que apresenta um relato de investigação sobre o processo de aprendizagem de circuitos elétricos, no Ensino Médio, contando, principalmente, com o auxílio de simulações e de maquetes.

A necessidade de mão de obra qualificada e de profissionais com bons conhecimentos de ciência e tecnologia é cada vez maior. A carência de profissionais é bastante elevada e passa a ser ainda mais preocupante quando se olham os números referentes à participação feminina nestas áreas. Segundo o relatório do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), dos 1.683 engenheiros da computação formados em 2010, somente 161 eram mulheres. Dados de 2013 enfatizam que as mulheres continuam dando preferência de escolha nas áreas do conhecimento tradicionalmente “femininas”, como educação (pedagogia), saúde (enfermagem, psicologia, fisioterapia) e bem-estar social (serviço social) destaque pra área de arquitetura que apresenta significativa participação feminina, enquanto a preferência masculina recai principalmente em cursos de engenharia e tecnologia (MOREIRA, MATTOS & SILVA, 2014).

Em virtude do contexto sócio-cultural que este quadro se apresenta, existe hoje um movimento mundial com o objetivo de resgatar estas meninas que por um motivo ou outro perdem o interesse em atuar em tecnologia, computação e engenharias, mesmo sendo uma área tão promissora e carente de profissionais qualificados.

O *Projeto Meninas Digitais - UFSC* propõe atividades atraentes e interativas que mostram que o ensino de computação e engenharia pode ser lúdico e motivador para alunas do ensino médio. Também atua em conjunto para avançar na formação escolar, já que ajuda a desenvolver a criatividade, concentração, raciocínio lógico, objetividade e procura desmistificar essa área, para que mais mulheres atuem nas áreas de STEM(Ciência, Tecnologia, Engenharia e Matemática).

2. Metodologia

Este relato foi escrito a partir da oficina realizada na Escola Estadual Básica Dite Freitas na cidade de Tubarão no estado de Santa Catarina. A oficina foi realizada em um único dia, com a duração de três horas. Um conteúdo básico de eletricidade foi apresentado para alunas no ensino médio. As atividades aconteceram no turno vespertino, todas as alunas faziam parte do ensino integral. Participaram da oficina oito alunas que foram selecionadas pela própria escola.

Com o objetivo de verificar o conhecimento prévio das alunas foi realizado um questionário inicial. Após o questionário foi feita uma explicação sobre os conceitos envolvidos e na sequência um primeiro contato com o circuito elétrico que havia sido levado para a sala de aula. Durante a oficina as alunas puderam realizar uma série de experimentos didáticos que são propostos nas apostilas disponibilizadas pelo fabricante dos kits. Os experimentos são classificados em básicos, médios e difíceis. Os kits educacionais são desenvolvidos de forma a tornar o ensino atraente e motivador, são classificados como brinquedos.

Snap Circuits (SNAPCIRCUITS, 2016) é um kit educacional que faz uso de componentes eletrônicos reais onde as conexões são feitas por encaixe, conforme ilustra a Figura 1. Seguindo o conceito de aprender fazendo é uma forma importante de

assimilação de conteúdos e desenvolvimento da aprendizagem. Eletricidade é um tema pouco aprofundado no ensino médio, o seu estudo é mais recorrente em cursos superiores de graduação, o Snap Circuits preenche uma lacuna, onde se podem simular circuitos elétricos reais, de forma fácil e segura. O Snap Circuits permite que os alunos exercitem conceitos de física e matemática de uma maneira menos abstrata e enfadonha.



Figura 1. Circuito Elétrico com Snap

Professores e monitores estão disponíveis no tempo da oficina para auxiliarem as alunas na montagem dos circuitos. Os monitores são alunas de graduação, entre bolsistas e voluntárias, da universidade, participam quatro alunas, sendo assim duas voluntárias, duas bolsistas e uma professora.

3. Resultados e Considerações

Os resultados da oficina são resultados qualitativos baseados nas respostas das alunas a um questionário. Na oficina realizada foi possível perceber o quanto nossas escolas carecem de atividades de experimentação, todas as alunas afirmaram que a experiência trará benefícios no seu aprendizado, pois o uso dos kits didáticos tornou o aprendizado mais interessante e podem refletir de alguma forma no desempenho escolar delas. Os principais comentários foram: “interessante”, “interessante e criativo”, “muito legal”, “divertida, dinâmica e interessante”, “muito legal, amei”, etc. Das 8 alunas que participaram da oficina, 7 delas demonstraram querer conhecer mais sobre os circuitos elétricos.

O uso das ferramentas para a oficina despertou o interesse das alunas, provando que antes da oficina, como apresenta os dados do questionário inicial, nenhuma tinha uma opinião formada sobre os circuitos elétricos. Assim, as oficinas despertaram o interesse de aprendizagem, sendo que a facilidade de auxílio das ferramentas propostas na atividade contribuiu para prender a atenção das estudantes, visto que nenhuma das alunas teve dificuldade para realizar os experimentos. No final foi feito um

questionamento solicitando sugestões para melhoria da oficina. A maior exigência foi o acréscimo da carga horária das atividades propostas, para que elas possam experimentar mais

As participantes do curso, independente de já terem feito suas escolhas profissionais e estarem ou não voltadas às áreas da tecnologia computacional, participaram ativamente da oficina e dos desafios apresentados. Com os experimentos realizados, as estudantes puderam observar o comportamento de alguns circuitos elétricos, como para o funcionamento de rádios e o acionamento de lâmpadas. Além do aprofundamento da física, há também o desenvolvimento do raciocínio lógico, tanto para problemas do cotidiano, quanto para resolução de problemas específicos.

Espera-se que iniciativas para mudar o abismo de gênero nestas áreas se multipliquem cada vez mais, tentando com isso, desmistificar o paradigma que estas áreas são apenas para meninos. No entanto, para superar esta situação são necessários esforços conjuntos de toda a sociedade, que precisa perceber que a diversidade de gênero, de comportamentos e de culturas trará muitos benefícios.

Agradecimentos

Agradecemos o apoio do CNPq/VALE S.A. Nº 05/2012 – Forma-Engenharia, MCTI/CNPq/SPM-PR/Petrobras nº 18/2013 e ao Edital Probolsa 2014 da Universidade Federal de Santa Catarina.

Agradecemos também a Escola Jovem Dite Freitas – Tubarão/SC e ao Instituto Alcoa.

Referências

MOREIRA, J.; MATTOS, G.; SILVA, L. (2014) Um Panorama da Presença Feminina na Ciência da Computação. Disponível em: <<http://www.ufpb.br/evento/liti/ocs/index.php/18redor/18redor/paper/view/1935/853>>. Acesso em 12 out. 2015.

REBELLO, A. P. e RAMOS M. G. (2009) Simulação computacional e maquetes na aprendizagem de circuitos elétricos: um olhar sobre a sala de aula. *Experiências em Ensino de Ciências – V4(1)*, pp.23-33, Porto Alegre.

HERTZ, G., Guimarin, A. & HAYES, G. (2013) Toy Hacking- Preliminary Results in Creative Electronic Workshops for Informal Science Education. *FabLearn 2013: Digital Fabrication in Education Conference*. October 27-28, 2013.

SNAPCIRCUITS. (2016) Make Learning a Snap. Disponível em: <<http://www.snapcircuits.net>> Acesso em 31 maio 2016

Divulgar para Atrair, Motivar para Manter

Sílvia Amélia Bim, Marília A. Amaral, Nádia P. Kozievitch, Maria Claudia F. P. Emer, Mariângela O. G. Setti, Luiz A. Pellison, Luiz E. Merkle

Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná (UTFPR) – Curitiba, PR - Brasil

{sabim, mariliaa, nadia, mcemer, mariangelasetti, pelisson, merkle}@utfpr.edu.br

Abstract. *The increasing masculinization of Computing is a global concern. The scenario requires diverse actions among all the people involved. This article presents some of the initiatives carried out in the context of an extension project which aims to diversify the dimensions of Computing, in order to both diversify the participation and attract more women to the area. We also present actions carried out to motivate current students of Information Systems and Computer Engineering degrees of a Brazilian public university, in order to contribute to lower the drop-out rates of women in these degrees.*

Resumo. *A crescente masculinização da área de Computação é uma preocupação mundial. O cenário exige ações diversificadas com todas as pessoas envolvidas. Neste artigo apresentamos algumas das iniciativas realizadas no contexto de um projeto de extensão que visa diversificar as dimensões da Computação com o objetivo de também diversificar a participação, atraindo mais mulheres para a área. Além disso, também apresentamos algumas ações desenvolvidas para motivar os atuais estudantes dos cursos de Sistemas de Informação e Engenharia da Computação de uma universidade pública brasileira, de forma a contribuir para um menor índice de evasão de mulheres nestes cursos.*

1. Introdução

Desde os anos 80 observa-se um decréscimo da presença de mulheres dentre os/as concluintes dos cursos da área de Computação no Brasil [Oliveira et. al. 2014]. Para tentar reverter este cenário, diversas ações estão sendo realizadas no país. Uma iniciativa é o Programa Meninas Digitais da Sociedade Brasileira de Computação (SBC) [Maciel e Bim, 2016] que incentiva e apoia a criação de projetos que divulguem a Computação para estudantes mulheres nos ensinos médio e fundamental. Anualmente, este Programa também realiza o Fórum Meninas Digitais, incluso na programação do WIT - *Women in Information Technology*, evento base do Congresso da Sociedade Brasileira de Computação.

Embora a SBC tenha iniciativas que buscam promover a maior participação de mulheres na Computação ainda há muito a ser feito. Em 38 anos de existência, a SBC só teve uma presidente mulher - Claudia Bauzer de Medeiros (2003 a 2007 - duas gestões) e uma vice presidente mulher, Thais Vasconcelos Batista (2016 a 2018 - atual

gestão). Na gestão atual, de oito diretorias, três (37,5%) são dirigidas por mulheres. Em relação às 36 comissões especiais da SBC, em 2015, 10 (27%) são coordenadas por mulheres.

Entretanto, além de programas de incentivo e atração de novos talentos para a Computação, neste caso mulheres, também são urgentes ações que viabilizem a permanência de mulheres profissionais na área e incentivem as mulheres a ocuparem posições de liderança nos contextos onde atuam.

Este artigo apresenta o Projeto Emíli@s - Armação em Bits¹, suas ações e resultados obtidos desde a sua criação. Tal projeto realiza ações para divulgar a computação para estudantes do ensino médio de escolas públicas de Curitiba. Além disto, promove a integração de alunas de pós-graduação com as alunas de graduação, realiza atividades para incrementar a permanência de mulheres nos cursos de Sistemas de Informação e Engenharia de Computação e, conseqüentemente, diminuir os índices de evasão. O Emíli@s também organiza encontros para docentes e estudantes, de ambos os sexos, para discutir questões de gênero, pois acredita-se que é necessário ampliar e aprofundar as reflexões sobre o tema Mulheres na Computação, a fim de que seja possível, por meio da colaboração de todas e todos, uma diversificação da participação na área.

O artigo é estruturado como segue. Na seção seguinte, o histórico do projeto é apresentado. As Seções 3 e 4 abordam ações de divulgação e ações de motivação, respectivamente. As considerações finais são feitas na Seção 5 que encerra o artigo.

2. O projeto Emíli@s - Armação em Bits

O Projeto Emíli@s - Armação em Bits é uma iniciativa do Departamento Acadêmico de Informática (DAINF) da UTFPR - Universidade Tecnológica Federal do Paraná, câmpus Curitiba. A primeira ação do projeto foi realizada no dia 08 de março de 2013. Neste primeiro ano o projeto foi cadastrado na universidade como projeto de pesquisa. A partir de 2014 o projeto foi cadastrado como projeto de extensão, sendo que no ano de 2014 teve o apoio financeiro do edital MCTI/CNPq/SPMPR/Petrobras nº18/2013 Meninas e Jovens Fazendo Ciências Exatas, Engenharias e Computação. Em 2016 o projeto conta com a colaboração de um casal de docentes da UTFPR-Campo Mourão, que com a colaboração de estudantes realiza ações no campus onde atuam.

O nome do projeto é inspirado na personagem de Monteiro Lobato², Emília³. O objetivo é reconhecer a voz das mulheres na Computação, as atuais e futuras. Entende-se que é necessário incentivá-las a armar, propor e resolver problemas da sociedade, usando a Computação, no sentido clássico de estruturar, aparelhar, abastar. Por isto, armação em bits.

Ao longo de sua história, o projeto teve a participação direta de oito docentes, cinco mulheres e três homens, na equipe que coordena as ações. Em 2014, foram alocadas bolsas, com os recursos do edital citado, a uma estudante do curso de Sistemas de Informação da UTFPR, e a quatro estudantes do ensino médio, e à coordenadora

¹ <http://emilias.dainf.ct.utfpr.edu.br/>

² https://pt.wikipedia.org/wiki/Monteiro_Lobato

³ [https://pt.wikipedia.org/wiki/Emília_\(personagem\)](https://pt.wikipedia.org/wiki/Emília_(personagem))

pedagógica da instituição parceira, o Colégio Estadual Dr. Xavier da Silva. Em todos os anos vários e várias estudantes voluntários(as) colaboraram com as ações, sendo em sua maioria integrantes de dois Programas de Educação Tutorial (PET) existentes no DAINF: PETCoCE⁴ - PET Computando Culturas em Equidade e PETECO⁵ - PET de Engenharia de Computação. Outras ações foram realizadas em parceria com o Compute você Mesm@: Conexões de saberes e fazeres para inclusão digital (projeto de extensão do DAINF), outras ainda, contaram com a participação de estudantes do Programa de Pós-graduação em Computação Aplicada (PPGCA/ DAINF).

3. Ações de divulgação

Ao longo de três anos e meio o projeto Emíli@s realizou diversas ações de divulgação da área de Computação. Buscaram-se, com estas ações, mostrar diversas dimensões da Computação a fim de divulgar as inúmeras possibilidades de atuação na área. Enfatizaram-se as demandas e as vantagens de profissionais com perfis diversificados, de modo a melhor compreender diferentes modos de computar, de diferentes pontos de vista. Como resultado destas ações, espera-se que os esteriótipo profissional em Computação possa ser desmistificado e criticado.

Uma das ações foi a realização de uma palestra para docentes da escola parceira - Colégio Estadual Dr. Xavier da Silva. A coordenadora do projeto Emíli@s apresentou ao corpo docente, durante a semana de formação, o cenário atual da Computação e a necessidade de atrair mais mulheres para a área.

Quatro oficinas práticas foram ofertadas para as estudantes do ensino médio da escola parceira. Estudantes de outras escolas também participaram:

- Oficina prática de Interação Humano-Computador [Amaral et al, 2015];
- Oficina prática de Banco de Dados [Rodriguez et al, 2016];
- Oficina para desenvolvimento de aplicativos Android usando App Inventor;
- Oficina de Webdesign Básico.

Mulheres profissionais, com diferentes perfis (formação e atuação), foram convidadas a compartilharem suas experiências fazendo palestras. Tais palestras aconteceram no próprio ambiente universitário. Uma delas também foi reapresentada na escola parceira.

Por fim, foram organizados eventos em diferentes épocas do ano. “Debatendo Gênero” foi realizado em parceria com outros projetos da universidade. A programação consistiu em duas mesas redondas com a participação de estudantes, docentes e profissionais buscando apresentar diferentes pontos de vista e experiências sobre o tema. Em 2014 e 2015, nos meses de outubro, foram realizadas atividades para celebrar o Ada Lovelace Day [Bim, 2016]. Nos anos de 2013, 2014 e 2016 o projeto organizou atividades para o Dia Internacional da Mulher. Em 2016, por exemplo, houve a exibição do filme Code Girl⁶, palestras com profissionais da área de Engenharia da Computação,

⁴ <http://www.dainf.ct.utfpr.edu.br/petcoce/>

⁵ <http://dainf.ct.utfpr.edu.br/peteco/>

⁶ <http://www.codegirlmovie.com/>

depoimento da única estudante brasileira que participou da Grace Hopper Conference em 2015 e uma palestra sobre feminismo.

4. Ações de motivação

Para ter o apoio das estudantes dos cursos de Sistemas de Informação, Engenharia da Computação e da pós-graduação em Computação Aplicada, nas ações do projeto Emíli@s identificou-se a necessidade de conhecer a experiência destas estudantes nos cursos. Para isto foram realizados edições de dois eventos.

O Pipocando Ideias foi uma ação realizada para promover o encontro das estudantes dos cursos graduação e pós-graduação, assim como discutir ações para receber calouras destes cursos e ações voltadas a estudantes do ensino médio.

O Bate Papo de Emíli@s teve o objetivo de proporcionar uma oportunidade de encontro entre estudantes dos cursos e algumas docentes, de modo que pudessem compartilhar suas experiências no contexto acadêmico.

5. Considerações finais

É possível identificar resultados de diferentes naturezas a partir das ações do Projeto Emíli@s - Armação em Bits. Um destes resultados foi a realização de duas pesquisas sobre a presença de mulheres no curso de Sistemas de Informação, realizadas por duas duplas de estudantes mulheres, no contexto da disciplina de Metodologia de Pesquisa, deste mesmo curso. A partir destas pesquisas, foi criada a Diretoria Feminina do Centro Acadêmico de BSI que organizou diversas atividades de integração entre as estudantes mulheres do curso.

Algumas reflexões sobre gênero, que já existiam no DAINF antes da criação do Projeto Emíli@s, contribuíram claramente na reformulação curricular do curso de Sistemas de Informação [Setti et al, 2014].

As avaliações feitas após as atividades são positivas. A parceria com a escola se renova anualmente e sempre há interesse das estudantes em participar das ações propostas. A discussão sobre gênero está presente no cotidiano do Departamento e percebe-se um crescente engajamento dos(as) atuais estudantes nos questionamentos feitos durante os eventos e fora deles. Os resultados parciais alcançados até o momento são positivos.

Referências

- Amaral, M.A., Bim, S.A., Boscaroli, C., Maciel, C. (2015) “Introducing Computer Science to Brazilian Girls in Elementary School through HCI Concepts” In: HCI International 2015 – Session: Women in DUXU, Los Angeles, CA, Lecture Notes in Computer Science - Design, User Experience, and Usability: Users and Interactions. Proceedings, Part II, 2015. v. 9187. p. 141-152.
- Bim, S. A. Ada Lovelace - falando sobre gênero nas aulas de Algoritmos. In: Computer on the Beach, 2016, Florianópolis. Anais [do] Computer on the Beach, 2016. p. 300-309.

- Maciel, C., Bim, S. A. Programa Meninas Digitais - ações para divulgar a Computação para meninas do ensino médio. In: Computer on the Beach, 2016, Florianópolis. Anais [do] Computer on the Beach, 2016. p. 327-336.
- Rodriguez, J. J. ; Kozievitch, N. P. ; Bim, S. A. ; Setti, M. G. ; Emer, M. C. F. P. ; Amaral, M. A. . Uma Proposta para Apresentar a Computação - Banco de Dados no Ensino Médio para o Público Feminino. In: Escola Regional de Banco de Dados - SBC, 2016, Londrina - PR. XII Escola Regional de Banco de Dados - Tema Data Science, 2016. v. 1. p. 155-158.
- Oliveira, A.C., Moro, M.M., Prates, R.O. (2014) “Perfil Feminino em Computação: Análise Inicial” In: XXII Workshop sobre Educação em Computação, Brasília, DF, p. 1465-1474.
- Setti, M.G., Emer, M.C.F.P., Amaral, M.A., Merkle, L.E. e Gonçalves, M.M. (2014) “Proposta de Flexibilização Curricular do Curso de Sistemas de Informação ofertado pela UTFPR-Curitiba”, In: XXII Workshop sobre Educação em Computação, Brasília, DF, p. 1366-1375.

Empoderamento Feminino com Dispositivos Móveis

Joice P. Cardoso, Laís Dalle Mulle, Luciana B. Frigo, Eliane Pozzebon

LabTeC – Universidade Federal de Santa Catarina (UFSC)
Araranguá – SC – Brasil

meninasdigitaisufsc@gmail.com, luciana.frigo@ufsc.br

Abstract. *This article describes the workshop of mobile applications given to high school students by Meninas Digitais UFSC in 2014. During the workshop the participants were able to develop applications using the AppInventor that combines visual programming blocks to facilitate user learning. Some students participated in a competition sponsored by a regional magazine for high school. Girls used the workshop's knowledge and developed a complete application when it was only required to make a theoretical proposal.*

Resumo. *Este artigo descreve a oficina de Aplicativos Móveis ministrada às alunas participantes do projeto Meninas Digitais UFSC no ano de 2014. No projeto as participantes aprenderam a desenvolver aplicativos utilizando a ferramenta AppInventor, desenvolvida pelo MIT, que une programação visual e programação em blocos a fim de facilitar o aprendizado do usuário. Ao final as alunas participaram de um concurso promovido por uma revista para escolas de ensino médio e utilizaram os conhecimentos adquiridos na oficina para desenvolver um aplicativo quando era somente necessária a elaboração teórica de uma proposta.*

Palavras-chave: *Gênero e tecnologia, programação no ensino médio, desenvolvimento, aplicativos móveis.*

1. Introdução

Historicamente a participação feminina nas áreas de ciência e tecnologia é pouco representativa, devido a tal fato, ve-sê a necessidade de iniciativas que investigem as causas e atuem no sentido de mudar essa realidade como apresentado em Frigo et al. (2015).

Há trabalhos como o do projeto “Android Smart Girls” Ramos et al. (2015) o qual visa a aproximação de garotas do ensino médio as áreas de tecnologia e computação. Para despertar o interesse das meninas utilizou-se a criação de aplicativos para smartphones, valendo-se da ferramenta App Inventor, e ao final traçou-se um perfil das participantes.

Em Duda et al. (2014), alguns discentes do Campus Irati do Instituto Federal do Paraná (IFPR), desenvolveram dois aplicativos, sendo o Aplicativo 1 - Resolução de regra de três com grandezas diretamente proporcionais e o Aplicativo 2 – Comparação de preços de combustíveis. Ao final da elaboração de cada aplicativo, os alunos apresentaram um tutorial do desenvolvimento do mesmo. Dessa forma, foi possível conferir se os conceitos matemáticos pesquisados foram utilizados de forma correta e verificar se foram incorporadas as ferramentas necessárias para o funcionamento do aplicativo. A existência desses trabalhos correlatos substancia a relevância e importância do tema

exposto no presente artigo.

O projeto de extensão Meninas Digitais UFSC por meio de oficinas e palestras, ministradas por alunas da graduação, leva até alunas do ensino médio de escolas públicas, atividades que proporcionam uma amostra do universo da engenharia e da computação, de forma lúdica e criativa.

Dentre as atividades oferecidas pelo projeto, a oficina de desenvolvimento de aplicativos móveis destaca-se por representar uma área do conhecimento difundida globalmente e que faz parte do dia a dia das pessoas. A oficina permite unir diferentes áreas como programação e interdisciplinaridade envolvendo aspectos de inovação e empreendedorismo para as alunas. Este artigo descreve a oficina de desenvolvimento de aplicativos móveis ofertada para a Escola Dite Freitas, na cidade de Tubarão/SC.

2. Ferramenta

A necessidade de um mecanismo inteligível adequado para a oficina se deve ao fato das meninas não terem prévio conhecimento no campo de programação e desenvolvimento de aplicativos. Para um bom andamento da oficina é necessário proporcionar uma atividade mais divertida desse universo visto como sério e por vezes considerado incompreensível. Para tanto, se buscou uma ferramenta gratuita e de fácil acesso, sendo que o software que mais se adequou às necessidades apresentadas foi o App Inventor, uma aplicação desenvolvida pelo MIT (Massachusetts Institute of Technology) que transforma a complexa programação em linguagem de alto nível em uma linguagem visual estruturada em blocos para a construção de aplicativos Android.

Um dos pontos positivos da aplicação é sua execução ser feita em navegadores web, tornando-a uma ferramenta prática, dessa forma pode ser utilizada em qualquer computador, desde que o mesmo esteja conectado à rede, uma vez que a maior parte dos computadores da escola possuem o sistema operacional Linux.

O desenvolvimento de aplicativos no App Inventor divide-se em duas seções, uma de design e outra de programação. O design é o ambiente onde o usuário desenha o aplicativo, adicionando botões, animações, fotos, entre outros. Na parte da programação, faz-se uso da linguagem de blocos lógicos orientada a eventos, como citada anteriormente, a qual se baseia em jogos de encaixe bastante conhecidos.

Há uma gama de aplicativos desenvolvidos com o uso do App Inventor, encontrando-se tanto aplicativos lúdicos quanto de cunho científico, fatos que podem ser verificados mediante o trabalho de Gomes e Melo (2013) o qual apresenta uma metodologia para o ensino da lógica de programação por meio da criação de jogos com pouca complexidade para dispositivos móveis Android e de Rocha et al. (2015) no qual foi desenvolvido um aplicativo para leitura de códigos QR provenientes de aplicações botânicas.

3. Metodologia

A fim de saber o grau de conhecimento das alunas sobre o desenvolvimento de aplicativos para dispositivos móveis, antes da oficina aplicou-se um questionário. Nesta oficina participaram 27 alunas dos três anos do ensino médio. Quando questionadas sobre se costumavam baixar aplicativos em seus celulares 20 garotas (74,1%) responderam que o faziam habitualmente. Das 27 que participaram da pesquisa 24 mostraram interesse em aprender a fazer aplicativos, evidenciando o fato de que as

meninas tem curiosidade em como é o mundo do desenvolvimento. Com o intuito de perceber o nível de compreensão sobre a elaboração de aplicativos 40,7% delas admitiram nunca ter pensado sobre o assunto, em contrapartida, 44,4% das meninas sabiam como os aplicativos são implementados. Por fim, nenhuma das estudantes havia feito algum aplicativo móvel, o que resultaria em um conhecimento novo para todas as alunas. Foram realizadas duas oficinas sobre o desenvolvimento de aplicativos móveis, cada uma com duração média de 3h.

Para introduzir as meninas no universo do desenvolvimento de aplicativos móveis buscou-se primeiramente passar conceitos básicos para então dar início à produção dos aplicativos. Foram usadas apostilas e slides como material de apoio para o processo de aprendizagem.

A plataforma foi apresentada às alunas que fizeram o cadastro para terem acesso a plataforma de desenvolvimento e poderem iniciar as atividades. Foi feita uma introdução sobre o uso da ferramenta com base na criação de um botão, sendo este, o meio mais didático encontrado para apresentar os conceitos de configuração. Este botão foi configurado de forma que durante a parte de montagem dos blocos fosse possível gerar um link. Em seguida, foi feita a programação de blocos, onde a funcionalidade do mesmo foi desenvolvida, ou seja, quando clicado iria para determinado caminho.

Um dos questionamentos da enquete inicial se tratava do tipo de aplicativo que elas gostariam desenvolver. Com base na análise dos resultados foi escolhido desenvolver um editor de fotos. Um tutorial foi ministrado explicando o passo a passo do editor de imagem, onde a foto poderia ser tirada pelo próprio aplicativo ou selecionada na biblioteca de imagens do celular. A principal edição na foto era a inclusão de texto.

A primeira aplicação, mesmo sendo bastante simples, mostrou que o desenvolvimento de um aplicativo era possível de ser realizado sem um conhecimento aprofundado em programação. As alunas puderam salvar suas fotos e as editar utilizando seu próprio programa, o que foi motivo de orgulho e motivação para elas.



Figure 1. Oficina de Desenvolvimento de Aplicativos Móveis

4. Resultados

Ao fim das oficinas as meninas se mostraram animadas com os aplicativos desenvolvidos. A direção da escola relatou que houve grande repercussão entre os alunos da escola.

Ainda como resultado da oficina, uma das alunas do projeto, T. C., junto a mais 11 colegas participaram do concurso “Desafio do Conhecimento”, realizado pela Revista Its, onde mais de 100 escolas do estado de Santa Catarina estavam competindo. Durante o concurso os alunos deveriam realizar 8 tarefas variadas com o tema “todos por um planeta melhor”, objetivando agregar habilidades aos estudantes e a busca pelo conhecimento. O aplicativo foi submetido a tarefa chamada “Ambientes inteligentes, pessoas mais conectadas”, onde deveriam desenvolver o protótipo de um aplicativo mobile que proporcionasse ambientes inteligentes para um planeta melhor. A aluna não só fez o protótipo, montou a parte visual, realizou testes, programou, planejou e desenvolveu o aplicativo. O aplicativo foi nomeado Question Tags e desenvolvido na plataforma App Inventor, que tinha por propósito ampliar o vocabulário do usuário, mostrando diferentes palavras, as quais o utilizador deveria associar o significado. A execução desta tarefa permitiu ampliar a pontuação das escola perante as demais escolas, resultando na terceira colocação geral na competição.

5. Conclusão

A partir do que foi exposto, vê-se que o desenvolvimento de aplicativos pode agregar um novo conhecimento e abrir novos horizontes para as meninas, já que muitas, como mostrado no questionário, não imaginavam como se dava o desenvolvendo de aplicativos. A maior consequência positiva da oficina foi o aproveitamento dos novos conhecimentos em atividades fora do projeto.

A ONU Mulheres e o Pacto Global criaram os Princípios de Empoderamento das Mulheres que são um conjunto de considerações que ajudam a comunidade empresarial a incorporar em seus negócios valores e práticas que visem à equidade de gênero e ao empoderamento de mulheres (ONU, 2016). Um dos sete princípios é o de *Promover educação, capacitação e desenvolvimento profissional para as mulheres*, neste contexto acredita-se que o projeto Meninas Digitais – UFSC com a oficina de desenvolvimento de aplicativos móveis promoveu educação e contribuiu para a equidade de gênero e empoderamento de mulheres.

O projeto de extensão Meninas Digitais UFSC continua ministrando oficinas para que as alunas do ensino médio possam vislumbrar o universo da computação e da tecnologia, para que assim tenham conhecimento de possíveis áreas profissionais.

Agradecimentos

Agradecemos o apoio do CNPq/VALE S.A. N° 05/2012 – Forma-Engenharia, MCTI/CNPq/SPM-PR/Petrobras n° 18/2013 e ao Edital Probolsa 2014 da Universidade Federal de Santa Catarina.

Agradecemos também a Escola Jovem Dite Freitas – Tubarão/SC e ao Instituto Alcoa.

Referências

- Duda, Rodrigo; Zontini, Diego D.; Silva, Sani C. R.; Grossi, Luciane (2012). “Elaboração de Aplicativos Para Android Com Uso do APP Inventor: Uma Experiência no Instituto Federal do Paraná – Campus Irati”, In: IV Simpósio Nacional de Ensino de Ciência e Tecnologia.
- Frijo, L. B.; Cardoso, P. B. ; Cardoso, J. P. ; Coelho, M. H.; Pozzebon, Eliane (2015). “Práctica del incentivo a la inserción de mujeres en carreras tecnológicas y de Ingeniería Robótica Educacional”, In: Novática, v. 231, p. 48-52.
- Gomes, Tancicleide C. S.; Melo, Jeane C. B. (2013). “Lógica de Programação: Iniciação Lúdica com App Inventor” In: III Escola Regional de Informática de Pernambuco 6-8 de Novembro, Garanhuns, PE, Brasil.
- ONU. Princípios de Empoderamento das Mulheres. (2016) Disponível em: <http://www.onumulheres.org.br/referencias/principios-de-empoderamento-das-mulheres/> Acesso em: junho 2016.
- Ramos, Nadja; Freitas, Carmen; Avila, Sandra; Costa, Paula D. P.; Testoni, Vanessa; Borin, Juliana F. (2015). “Ensino de Programação para Alunas de Ensino Médio: Relato de uma Experiência”, In: Anais do XXI Workshop de Informática na Escola.
- Rocha, Luís A. G.; Cruz, Fabiana M.; Leão, Alcides L. (2015). “Aplicativo Para Educação Ambiental” In: Periódico Eletrônico Fórum Ambiental da Alta Paulista, volume 11, número 4.

I Fórum Gurias na Computação: relato de experiências e desdobramentos

Isadora Garcia Ferrão, Amanda Meincke Melo

Universidade Federal do Pampa (Unipampa) – Campus Alegrete
Av. Tiarajú, 810 – Ibirapuitã – 97.546-550 – Alegrete – RS – Brasil

isadora-gf@hotmail.com, amanda.melo@unipampa.edu.br

Abstract. *The incipient presence of women in the field of Information Technology has been a recurring subject of debate in Brazil and abroad. In this context, the “1st Forum Gurias in Computing: experiences that inspire” was proposed and it was conducted in the scope of “7th Academic Week at Technological Campus of Alegrete” to encourage the exchange of experiences among Computing professionals and students, as well as disseminate the area to high school students. We aim at sharing our experience in promoting this forum and its results, which includes the proposal of an activity to promote women's participation in Computing.*

Resumo. *A presença incipiente de mulheres na área de Tecnologia de Informação tem sido alvo recorrente de debates no Brasil e no exterior. Nesse contexto, com o tema “Experiências que inspiram”, foi proposto o I Fórum Gurias na Computação, integrado à “VII Semana Acadêmica do Campus Tecnológico de Alegrete”, que teve como proposta incentivar a troca de experiências entre profissionais da área de Computação e estudantes, assim como divulgar a área a estudantes do Ensino Médio. Busca-se, com este trabalho, compartilhar os resultados dessa experiência e seus desdobramentos, que inclui a proposta de uma ação de extensão para promover a participação feminina na Computação.*

1. Introdução

A última Pesquisa Nacional por Amostra de Domicílios (PNDA), do Instituto Brasileiro de Geografia e Estatística (IBGE), explicita que dentre os 580 mil profissionais de Tecnologia de Informação (TI) no Brasil, 20% são mulheres. Além disso, a pesquisa evidencia que a evasão feminina de cursos relacionados à área de TI chega a 79% no primeiro ano [UOL 2013]. Essa presença incipiente de mulheres na área tem sido alvo recorrente de debates no Brasil e no exterior e uma série de iniciativas, inclusive no âmbito da Sociedade Brasileira de Computação (SBC), tem destacado contribuições de mulheres à Computação e incentivado a participação feminina na área. Dentre elas, desde 2011, existe o programa Meninas Digitais, que visa a despertar o interesse de estudantes do Ensino Médio/Tecnológico ou dos anos finais do Ensino Fundamental para que tenham maior conhecimento a respeito da área de Computação para, assim, incentivá-las a seguir na carreira [SBC 2015].

Nesse contexto, no Campus Alegrete da Universidade Federal do Pampa (Unipampa), no dia 09 de Junho de 2015, realizou-se o “I Fórum Gurias na Computação: experiências que inspiram”, como ação vinculada ao programa de extensão “GEInfoEdu – Grupo de Estudos em Informática na Educação” [Ferrão e

Melo, 2015]. Integrado à VII Semana Acadêmica do Campus Tecnológico de Alegrete, o fórum teve como proposta incentivar a troca de experiências entre profissionais da área de Computação e estudantes, assim como divulgar a área a estudantes do Ensino Médio, contando com a participação de egressas dos cursos de Computação do próprio Campus.

Com este trabalho, pretende-se compartilhar os resultados desta experiência e seus desdobramentos. A seção 2 apresenta a metodologia para a organização, a realização e a divulgação de resultados do fórum. A seção 3 apresenta alguns de seus resultados. Finalmente, realizam-se as considerações finais.

2. Metodologia

O “I Fórum Gurias na Computação: experiências que inspiram” foi organizado por uma docente e uma acadêmica da área da Computação do Campus Alegrete da Unipampa. Através de um grupo na rede social *Facebook*, estas envolveram acadêmicas dos bacharelados em Ciência da Computação e Engenharia de Software na definição do tema, da programação geral do evento e do *banner* de divulgação do fórum.

Uma vez definida a programação e confirmadas as colaborações, a discente envolvida na organização do fórum propôs três versões para o *banner* de divulgação do evento e as colocou em votação. A Figura 1 ilustra a versão escolhida.



I Fórum Gurias na Computação
Experiências que inspiram

Data: 09/06/15 - Terça feira

Local: Sala 101 - Unipampa -Campus Alegrete

19:00-19:30 Abertura - Professora Sílvia Amélia Bim (UTFPR)

19:30-20:30 - Roda de conversa com egressas do Curso da área da Computação. Amanda Estrela (Bacharel em Engenharia de Software), Tatiana Cartagena (Bacharel em Ciência da Computação) e Greicy Sayuri (Bacharel em Ciência da Computação)

20:30-21:00 Intervalo

21:00-22:00 - Palestra - A Presença Feminina nas Startups, Tatiana Cartagena de Oliveira (Emdime)

Logos: Unipampa, CEC, GelnfoEdu

Organização: Amanda Meincke Melo e Isadora Ferrão.

tecpib.com/

Figura 1. Banner do Primeiro Fórum Gurias na Computação: experiências que inspiram.

O fórum foi amplamente divulgado em grupos da rede social *Facebook*, no portal institucional da Universidade [Unipampa 2015] e para o PROJOVEM-Adolescente – programa da Assistência Social do município de Alegrete/RS.

Para a abertura do evento, com apoio do sistema de webconferência institucional, contou-se com a participação da professora Sílvia Amélia Bim da Universidade Tecnológica Federal do Paraná (UTFPR), a frente do projeto “Emílias – Armação em Bits”. Acompanharam esse momento, também *online*, as egressas Amanda Caricatti Estrela – Bacharel em Ciência da Computação e Greicy Sayuri Akamine – Bacharel em Engenharia de Software que, em seguida, colaboraram com relatos de

experiência. Presencialmente, o evento contou com a presença da egressa Tatiana Cartagena de Oliveira do Empreendedorismo Digital Feminino (Emdime) – Bacharel em Ciência da Computação – que realizou relato de sua experiência e palestrou sobre a presença feminina nas *startups*. Participaram do fórum acadêmicos dos cursos de Ciência da Computação e de Engenharia de Software, incluindo gurus e gurias. A Figura 2 apresenta momentos do fórum.

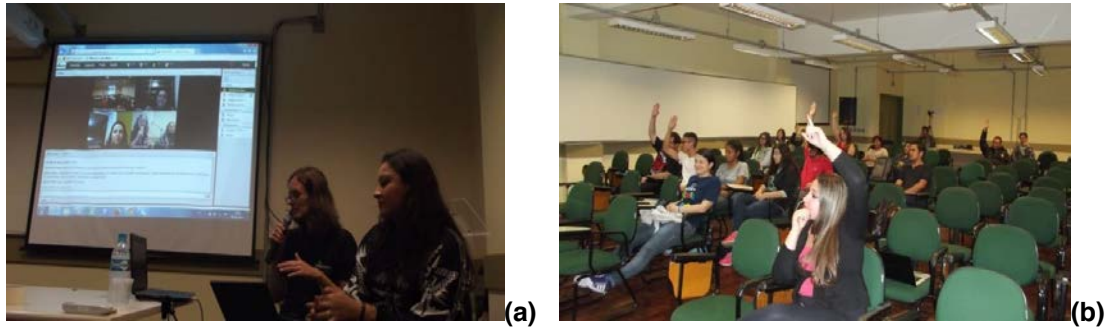


Figura 2. (a) webconferencia institucional, (b) troca de experiências.

Resultados do fórum foram divulgados no portal da Universidade [Unipampa 2015] e no evento VII Salão Internacional de Ensino, Pesquisa e Extensão (VII SIEPE) – organizado anualmente pela Universidade [Ferrão e Melo, 2015].

3. Resultados

O “I Fórum Gurias na Computação: experiências que inspiram”, além de valorizar experiências de egressas da Unipampa, possibilitou aos participantes conhecerem mais sobre a inserção da mulher no mercado de trabalho de TI. Durante o fórum, pode-se evidenciar a necessidade de estratégias para divulgar a área às gurias da comunidade, incentivar a permanência de mulheres nos cursos de Computação, dialogar sobre estereótipos e preconceitos, além de realizar proposições para um próximo fórum.

A confecção de um pôster para apresentação do fórum durante o VII SIEPE levou à elaboração de um logotipo para a temática “Gurias na Computação”, desenvolvido por um acadêmico do curso de Ciência da Computação, que adotou como referência um site que disponibiliza vetores livres e gratuitos [Freepik 2015]. Esse logotipo foi aplicado no pôster e em uma camiseta confeccionada para a apresentação do trabalho. A Figura 3 apresenta o logotipo escolhido em votação na rede social *Facebook* por acadêmicas da área da Computação.



Figura 3. Logotipo Gurias na Computação.

Impulsionou-se com o fórum a proposta de uma ação integrada ao programa de extensão “Programa C – Comunidade, Computação, Cultura, Comunicação, Ciência, Cidadania, Criatividade, Colaboração” com o objetivo de promover a participação

feminina na área da Computação. Esta ação é atualmente denominada “Gurias na Computação”.

Considerando a ausência de estudantes do Ensino Médio no fórum, a proposta de realização de relatos de experiências das docentes da área da Computação do Campus e a necessidade de serem pensadas estratégias para promover a participação feminina na área da Computação com a participação de discentes, organizou-se a segunda edição do fórum. Esta foi realizada no dia 12 de Maio 2016, integrada à VIII Semana Acadêmica do Campus Tecnológico de Alegrete, com o tema “Experiências docentes e protagonismo discente em perspectiva” [Unipampa 2016].

4. Considerações Finais

A realização do “I Fórum Gurias na Computação: experiências que inspiram” oportunizou, no Campus Alegrete da Unipampa, uma primeira aproximação à discussão sobre gênero na área da Computação. Resultou em uma ação de extensão para promover a participação feminina na área, além de oferecer subsídios à organização da segunda edição do fórum.

Referências

- Ferrão, I. G., Melo, A. M. (2015), Primeiro Fórum Gurias na Computação: experiências que inspiram, SIEPE 2015.
- Gallina, M. (2015) “Porque há menos mulheres no setor de tecnologia”, <http://epocanegocios.globo.com/Informacao/Dilemas/noticia/2015/08/por-que-ha-menos-mulheres-no-setor-de-tecnologia.html>
- SBC, “Meninas Digitais”, <http://sbc.org.br/institucional-3/meninas-digitais>, Abril
- Unipampa (2015) “Campus Alegrete realiza Fórum Gurias na Computação”, http://porteiros.unipampa.edu.br/alegrete/index.php?option=com_content&view=article&id=890%3Agurias&catid=1%3Anoticias-unipampa&Itemid=62, Abril
- Unipampa (2016) “II Fórum Gurias na Computação acontece na Unipampa”, http://porteiros.unipampa.edu.br/alegrete/index.php?option=com_content&view=article&id=921:ii-forum-gurias-na-computacao-acontece-no-unipampa, Maio
- UOL HOST (2015) “A difícil missão de ser mulher no mercado de ti”, <http://www.uolhost.uol.com.br/academia/noticias/tecnologia/2015/08/25/a-dificil-missao-de-ser-mulher-no-mercado-de-ti.html#rmcl>, Abril

Incentivando meninas do ensino médio a ingressarem nas áreas tecnológicas com curso de programação e robótica

Tatiana Nilson dos Santos¹, Maitê Thomazi Manenti², Eliane Pozzebon¹, Luciana Bolan Frigo¹

¹Programa de Pós-Graduação em Tecnologias da Informação e Comunicação –
Universidade Federal de Santa Catarina (UFSC)
Araranguá – SC – Brasil

²Laboratório de Tecnologias Computacionais – Universidade Federal de Santa Catarina
(UFSC) – Araranguá – SC – Brasil

tatiana.santos@posgrad.ufsc.br, mayte_manenti@hotmail.com,
{elianepozzebon, luciana.frigo}@ufsc.br

Abstract. *This article presents the results of a course conducted by the Federal University of Santa Catarina in partnership with the High School Macário Borba, located in Sombrio-SC. The course lasted two (2) months and was restricted to high school girls. The course aimed to encourage them to follow the technological areas, due to the low percentage of women found in these areas. The course presented the contents as computational logic and assembly and programming of robotic kit (LEGO). The results were positive, but noted the lack of incentive for students of basic education, lack of knowledge by the same in nearby universities courses*

Resumo. *Este artigo apresenta os resultados de um curso realizado pela Universidade Federal de Santa Catarina em parceria com a Escola de Ensino Médio Macário Borba, localizada em Sombrio-SC. O curso teve duração de 2 (dois) meses e foi restrito a meninas do ensino médio. O curso teve como objetivo incentivá-las a seguirem nas áreas tecnológicas, devido à baixa percentagem de mulheres encontradas nessas áreas. O curso apresentou os conteúdos como lógica computacional e montagem e programação do kit robótico (LEGO). Os resultados foram positivos, porém notou-se a falta de incentivo para os estudantes da educação básica, falta de conhecimento, pelo mesmo, nos cursos de universidades próximas.*

1. Introdução

No último vestibular do Instituto Tecnológico de Aeronáutica (ITA) em 2016, 23,1% dos candidatos inscritos eram mulheres e 10% dos candidatos convocados foram mulheres. Atualmente, 8% dos estudantes em curso são mulheres [ITA, 2016]. Na Universidade Federal de Santa Catarina (UFSC) – Campus Araranguá, também considerando os estudantes classificados no último vestibular (de 2015) para os três cursos considerados tecnológicos, sendo eles: Engenharia de Energia, Engenharia da Computação e Tecnologias da Informação e Comunicação, a porcentagem de estudantes do sexo feminino também apresentou queda [COPERVE, 2015].

O curso de Engenharia da Computação apresentou apenas 6% de público feminino no total de aprovados neste vestibular, sendo que em 2013 elas representavam 15%. Entre os aprovados no curso de Engenharia de Energia as mulheres representavam 38%

do total, onde no vestibular de 2013, elas representavam 43%. Analisando o curso de Tecnologias da Informação e Comunicação, o primeiro curso tecnológico da UFSC (Araranguá), a porcentagem de mulheres aprovadas no vestibular em 2013 representava 35% sendo que dois anos depois, elas representavam apenas 21% dos aprovados [COPERVE, 2015].

Levando em consideração, o baixo número de mulheres nas áreas, tanto das engenharias quanto nas áreas de tecnologias no geral, a UFSC (Araranguá), realizou um minicurso de programação e robótica exclusivamente para meninas do ensino médio, em parceria com a Escola de Ensino Médio Macário Borba, localizada na cidade de Sombrio-SC com o objetivo de estimular as jovens a seguir carreira nessas áreas.

2. Introdução à programação

A maioria dos cursos de tecnologia consta em sua grade curricular a disciplina de programação e se não for dada a devida importância a mesma, ela pode causar problemas no processo de ensino do estudante, ocasionando em alguns casos a desistência do curso. Aspectos como, i) motivação, ii) didática, iii) metodologia e iv) raciocínio lógico, deverão ser levados em consideração caso não se deseje um alto índice de reprovação nos cursos tecnológicos [LIMA JUNIOR; VIEIRA e VIEIRA, 2015].

Uma forma de auxiliar esses estudantes é apresentá-los à estes conteúdos antes mesmos deles ingressarem na universidade. Se os estudantes conhecerem a programação e os algoritmos, souberem o que é, para que serve, como programar, ao ingressarem na universidade este conhecimento será complementado com o conteúdo das disciplinas [LIMA JUNIOR; VIEIRA e VIEIRA, 2015].

Logo, estimular a formação nestas áreas no Brasil passa antes de tudo pela desmistificação da profissão diante dos estudantes de ensino médio. Para este público, que já encontra dificuldades em disciplinas como física, química e matemática, a escolha da engenharia e tecnologia, nas suas diversas habilitações, muitas vezes nem é cogitada, por acreditarem ser de grande dificuldade [WATANABE et. al. 2014].

Desta forma, também surge a responsabilidade da comunidade acadêmica, vinculada à estes cursos, propondo ações que tenham como objetivo diminuir a tensão e o pré-conceito que existe com as áreas da ciência, junto ao ensino médio e também o esclarecimento do real papel destes profissionais em nossa sociedade.

3. Proposta das Atividades

Este projeto surgiu da necessidade de envolver instituições públicas da região com a comunidade acadêmica, com vistas a criar oportunidades para as meninas das escolas públicas, motivando encaminhando as alunas a continuação dos estudos nas áreas de tecnologia.

Ocorreram 6 (seis) aulas entre os meses de maio a julho do ano de 2014, sempre aos sábados, totalizando uma carga horária de 4 (quatro) horas/semanais. Foram escolhidas 30 (trinta) meninas, na faixa etária de 15 (quinze) a 17 (dezesete) anos, por meio de um questionário aplicado juntamente com duas professoras da escola que também participaram do curso. As respostas do mesmo foram analisadas e as meninas que se enquadravam mais no perfil tecnológico foram escolhidas.

Na primeira aula foi passada às estudantes uma introdução à lógica de programação, conceito de variável e montaram um pseudocódigo. A partir da segunda

aula foi apresentado às estudantes a Lógica Computacional e os tipos de operadores (lógico, relacional e aritmético) com realização de exercícios práticos.

Na terceira aula, as estudantes desenvolveram pseudocódigos, no ambiente de programação *VisualG* sobre os conteúdos de estrutura de seleção múltipla e vetores. Na quarta aula, as meninas adaptaram os pseudocódigos desenvolvidos até aquele momento para a linguagem de programação C.

Já na quinta aula (Figura 1), as alunas montaram protótipos de robô (Kit LEGO, disponibilizados pelo Laboratório de Tecnologias Computacionais (LaBTeC) da UFSC - Araranguá). Na última aula, a tarefa era programar o robô a partir dos conteúdos aprendidos, para que o mesmo realizasse uma atividade qualquer.

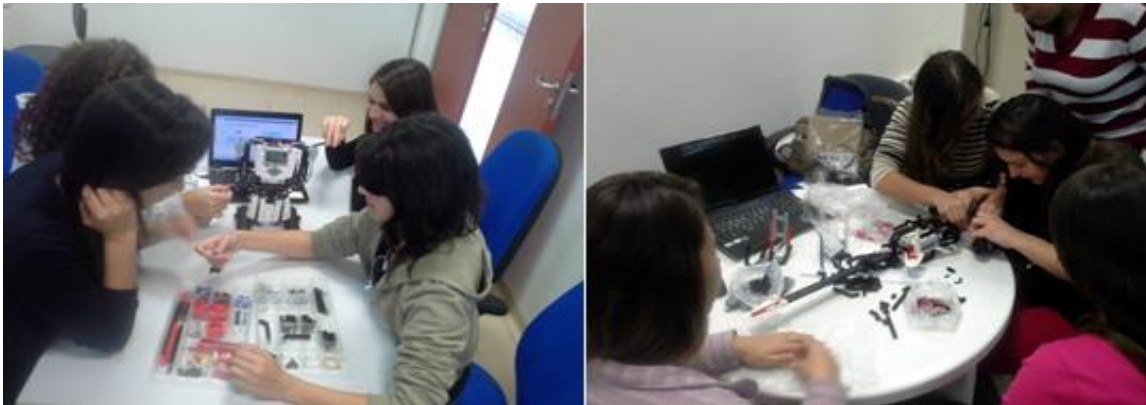


Figura 1: Montagem dos robôs

As monitoras foram estudantes dos cursos de Engenharia da Computação e Tecnologias da Informação e Comunicação (UFSC – Araranguá), sendo que a escolha pelo gênero feminino foi proposital uma vez que se pretende desmistificar o assunto sobre mulheres na área tecnológica, além de atrair novas profissionais.

4. Resultados e Discussões

Como resultado esperava-se que as participantes tivessem conhecimento mínimo de programação. Esperava-se também que estas tivessem mais motivação para o estudo das disciplinas do ensino médio, bem como um conhecimento maior sobre a universidade e seus cursos, principalmente nos cursos das áreas tecnológicas.

Pode-se perceber então, que o curso realmente apresentou os resultados positivos esperados, como pode ser percebido por meio de observação indireta e pelo questionário aplicado às alunas, sobre o desempenho do curso e da influência do mesmo sobre a vida escolar/acadêmica das mesmas e o interesse pela área que se fez despertar, ou não, ao longo do curso.

Questionadas sobre seus futuros (Figura 2), 38% das meninas afirmaram que se interessaram pela área de programação a ponto de se imaginar trabalhando nela futuramente, contra apenas 10% que afirmam não ter vontade de tornar-se programadoras. Por outro lado, 52% das alunas afirmam que esta poderá ser uma opção para elas no futuro.

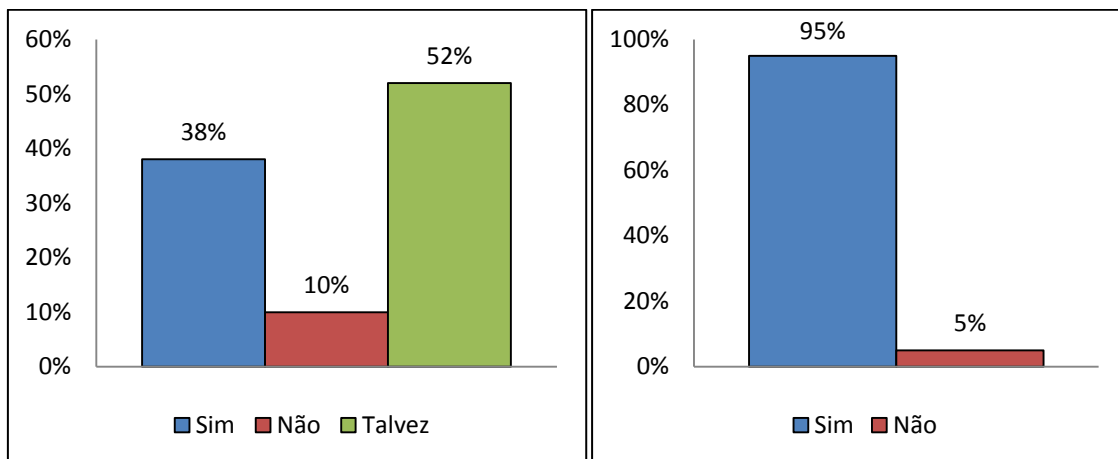


Figura 2: Você se imagina nesta área

Figura 3: Interesse após o curso

Sobre o interesse das alunas por programação (Figura 3), 95% afirmaram que o mesmo aumentou após a realização do curso e, apenas 5% afirmaram que não se interessavam e ele não foi despertado após do curso.

5. Considerações

A ideia do projeto é reforçar a importância da participação direta entre a universidade e as estudantes para o esclarecimento das dúvidas sobre as carreiras profissionais e o incentivo para que as mesmas sigam nas áreas tecnológicas, com o intuito de mudar a realidade sobre a escassez de mulheres nestas áreas.

Vários outros grupos, pelo mundo inclusive, estão se mobilizando em prol deste objetivo, tem-se como exemplo o grupo *Girls Who Code* (Meninas Que Programam) [GIRLS WHO CODE, 2016]. No Brasil, por exemplo, uma universidade que investe nessa ideia é Universidade de São Paulo (USP), em São Carlos que apresentou o programa Codifique, um curso de programação básica voltada para estudantes do 9º ano do ensino fundamental. Os principais tópicos abordados assemelham-se aos deste projeto, pois ao final os participantes desenvolvem projetos, tais como jogos e calculadoras, aplicando na prática o que haviam aprendido [WATANABE et. al. 2014].

Referências

- COPERVE. (2015) “Relatório Oficial do Vestibular: UFSC 2015”, http://vestibular2015.paginas.ufsc.br/files/2015/06/Relatorio_Oficial_V2015.pdf, Maio.
- GIRLS WHO CODE. (2016) “Who we are”, <http://girlswhocode.com/>, Março.
- ITA, Instituto Tecnológico de Aeronáutica (2016) “Dados Estatísticos Relativos ao Exame Vestibular de 2016”, <http://www.vestibular.ita.br>, Março.
- Lima Junior, J. A. T.; Vieira, C. E. C.; e Vieira, P. P. (2015) “Dificuldades no processo de aprendizagem de Algoritmos: uma análise dos resultados na disciplina de AL1 do Curso de Sistemas de Informação da FAETERJ - Campus Paracambi”. Cadernos UniFOA, Volta Redonda, n. 27, p. 5-15.
- Watanabe, F. Y.; Francisco, c. A.; França, C. A.; e Ogashawara, O. (2014) “A questão do gênero e as iniciativas de incentivo à formação de mais engenheiras na UFSCar”. In: XLII Congresso Brasileiro de Educação em Engenharia, Juiz de Fora - MG.

Incentivando o Ingresso de Mulheres nos Cursos de Engenharia e Tecnologia

Marcelly Homem Coelho¹, Morgana Sartor¹, Luciana Bolan Frigo¹, Eliane Pozzebon¹

meninasdigitaisufsc@gmail.com, luciana.frigo@ufsc.br

Abstract. *This article describes a workshop with Computer Engineering undergraduate young women and high school girls where the former give classes to the latter. The purpose of the workshop is to provide real experience with usually abstract concepts of electronics and technological subjects. The high school girls' challenge was to make a prototype automation doll house (domotics) using their knowledge. The project's main goal is for high school girls to experience and explore a field traditionally dominated by men, with the intention of enabling them to reflect and expand their career possibilities to include opportunities within the technological area as well as help in building new skills.*

Resumo. *O presente artigo descreve uma experiência de interação entre estudantes de graduação do curso de Engenharia de Computação com um grupo de alunas do Ensino Médio, por meio da realização de uma oficina. O objetivo da oficina é de proporcionar uma experimentação real acerca dos conceitos abstratos de eletricidade, eletrônica e tecnologia. O desafio das alunas foi de conhecer a domótica, realizando um protótipo de automação de uma casa de bonecas. O principal objetivo do projeto é que as meninas possam experimentar e explorar um campo tradicionalmente dominado e vivido pelos homens, a fim de que elas possam refletir e expandir suas possibilidades de carreira incluindo a área de exatas no rol de oportunidades, além de auxiliar na construção de novas habilidades e competências.*

1. Introdução

A desproporcionalidade da representação das mulheres em áreas tecnológicas como na eletrônica, por exemplo, possui causas variadas, sejam estas de ordem sociocultural, econômica ou cognitiva. O desenvolvimento de tecnologia ainda é considerada por grande parte da população como sendo intrínseca ao universo masculino. E desta forma, é fortemente difundido na sociedade que a mulher não possui talentos para as ciências exatas, mesmo que historicamente elas tenham conquistado papéis importantes na área da tecnologia, como por exemplo, o fato do primeiro algoritmo ter sido escrito por Ada Lovelace, mostrando que as mulheres possuem competências para atuarem nesta profissão, conforme Coelho et al. (2015).

Existem vários estudos que tem como objetivo entender os motivos que levam à diminuição do número de garotas em áreas de tecnologia e computação. De acordo com Teague (2002), uma das razões para essa redução é o fato das garotas terem uma percepção equivocada sobre as carreiras da computação fazendo com que a grande maioria delas opte por cursos mais tradicionais.

Para incentivar o aprendizado dos estudantes é de suma importância buscar alternativas para proporcionar um processo de ensino-aprendizagem atraente e um ambiente motivador. Um dos motivos de utilizar tecnologias é esclarecido no livro *Connected Code – Why Children Need to Learn Programming* de Kafai e Burke (2014), o qual afirma que os computadores se tornaram acessíveis dentro e fora do ambiente escolar, despertando assim um interesse em aprender estas tecnologias.

Moratori (2003) ressalta que utilizar recursos tecnológicos torna possível potencializar o desenvolvimento de competências diversas, uma vez que possibilita a reestruturação do modo como alunos e professores se relacionam.

Introduzir tecnologias como forma de entretenimento e fazer com que se torne também um momento de troca de conhecimento e aperfeiçoamento dos assuntos abordados em sala de aula é uma ótima maneira de instigar alunas do Ensino Médio a correlacionarem e solidificarem seus aprendizados (Coelho et al. 2015).

Neste contexto, surge a proposta do projeto de extensão *Meninas Digitais - UFSC* que realiza oficinas de diferentes temas que utilizam práticas pedagógicas modernas buscando apresentar profissões voltadas à computação e engenharias e de promover a reflexão sobre os papéis desempenhados por homens e mulheres nestas áreas.

2. Condução das oficinas

Com o intuito de mitigar o impacto negativo do senso comum que relaciona tecnologia a profissões masculinas que o Projeto *Meninas Digitais – UFSC* oferece oficinas e palestras de diversas áreas tecnológicas para incentivar meninas do Ensino Médio a obterem interesse na área da computação. As oficinas são ministradas por alunas do curso de Engenharia de Computação que fazem parte do projeto.

Neste artigo será descrita a oficina de iniciação à eletrônica, realizada em 2015, que faz uso do kit didático de eletrônica *LittleBits* (Fitzpatrick e Aneja, 2015), (*LittleBits*, 2016) onde as peças e componentes do circuito não estão conectadas a um protoboard e se encaixam umas as outras por contato e permanecem conectadas por atração magnética.

A oficina segue a seguinte ordem: (i) aplicação de um questionário inicial; (ii) distribuir apostilas com informações importantes sobre cada grupo de peças, sua utilização e aplicações; (iii) apresentar conceitos básicos de circuitos; (iv) mostrar as principais funcionalidades do kit; (v) propor uma situação problema que no caso foi de realizar a automação de uma casa de boneca, conforme ilustra a figura 1 e (vi) aplicação do questionário final.



Figura 1: Desenvolvimento da automação residencial.

A oficina aconteceu no laboratório de informática com um grupo de oito alunas, selecionadas previamente pela escola, o pequeno número de alunas, se justifica pelo fato da oficina ter ocorrido na final do ano letivo e muitas alunas que estavam estudando para o vestibular, não quiseram participar desta oficina.

Foi feita uma atividade demonstrando o uso do kit e na sequência as alunas deveriam solucionar, com os materiais disponíveis, o problema apresentado. O desafio envolveu conceitos de circuitos eletrônicos com aplicações em automação residencial, estimulando e fortalecendo o raciocínio lógico, além de provocar um estímulo de reflexão sobre as possibilidades de perfil do profissional atuante nesta área. A Figura 2 mostra o resultado da atividade de automação residencial realizada pelas alunas: (i) automação do portão; (ii) uso de sensores para acendimento automático das lâmpadas, (iii) efeitos sonoros etc. Os circuitos são desenvolvidos em grupos e qualquer uma das garotas poderia se registrar no site do LittleBits e publicar o projeto elaborado em sala, compartilhando o conhecimento adquirido com a comunidade que faz uso do kit ao redor de todo o mundo.



Figura 2: Casa de boneca automatizada pelas estudantes.

3. Resultados da Oficina e Conclusão

Na oficina aqui relatada, participaram oito alunas do ensino médio, que responderam aos questionários inicial e final. A primeira pergunta tinha o objetivo de constatar se elas já haviam tido algum contato com a montagem de um circuito elétrico. A grande maioria das alunas (87,5%) não tinha qualquer contato anterior com a construção de circuitos elétricos.

Quando questionadas sobre uso do kit LittleBits, por unanimidade, as participantes avaliaram que esse recurso tecnológico proporcionou à oficina a característica de ser mais interessante e dinâmica, uma vez que o minicurso possibilitou experiências não encontradas no ensino médio tradicional.

Em comparação aos resultados do questionário inicial, observou-se que as novas respostas, relacionadas ao interesse das estudantes em aprender mais sobre circuitos, foram gratificantes e motivadoras, pois todas demonstraram interesse.

No que diz respeito às alunas de Engenharia de Computação, a realização de oficinas é uma maneira de reforçar conhecimentos e realizar experiências práticas, pois algumas experimentações não são oportunizadas ao longo do curso.

Este artigo tem como finalidade relatar sucintamente a experiência de uma oficina prática de circuitos eletrônicos, que teve sucesso na sua realização em proporcionar um modelo alternativo e que pode ser replicado em outras instituições de ensino.

Agradecimentos

Agradecemos o apoio do CNPq/VALE S.A. Nº 05/2012 – Forma-Engenharia, MCTI/CNPq/SPM-PR/Petrobras nº 18/2013 e ao Edital Probolsa 2014 da Universidade Federal de Santa Catarina.

Agradecemos também a Escola Jovem Dite Freitas – Tubarão/SC e ao Instituto Alcoa.

Referências

COELHO, M. H. ; FRIGO, L. B. ; CARDOSO, J. P. ; SOUZA, R. F. ; POZZEBON, E. (2015) O Desafio de Mudar o Papel das Mulheres na Indústria de Games. In: Computer on the Beach, 2015, Florianópolis. Anais do Computer on the Beach. p. 249-258.

FITZPATRICK, A.; ANEJA, A. (2015) LittleBits Are Like Electronic Legos For Kids and Adults. TIME. Disponível em: < <http://time.com/4022048/littlebits/>>.

KAFAI, Y. B.; BURKE, Q. (2014) “Connected Code: Why Children need to learn programming”, Cambridge: The Mit Press.

LITTLEBITS (2016) Disponível em: < <http://littlebits.cc>>

MORATONI, P. B. (2003) “Por Que Utilizar Jogos Educativos no Processo de Ensino Aprendizagem?” Trabalho de Conclusão de Curso. UFRJ. Rio de Janeiro. Disponível em:

<http://www.nce.ufrj.br/ginape/publicacoes/trabalhos/t_2003/t_2003_patrick_barbosa_moratori.pdf>

SANTO, P., E. (2008) “Os estudos de gênero na ciência da informação”. Em Questão, Porto Alegre, v. 14, n. 2, p. 317 - 332, jul./dez.

TEAGUE, J, (2002) “Women in computing”. SIGCSE Bulletin, v. 34, n. 2, p. 147.

Levantamento e análise dos dados da presença feminina discente no curso de Bacharelado em Sistemas de Informação do Instituto Federal Catarinense, campus Camboriú

Ana Elisa Schmidt, Hiryna Philipps de Britto, Mozara Dias Koehler¹

¹Instituto Federal Catarinense- Campus Camboriú- Joaquim Garcia S/N-Camboriú- SC
anaelisa@ifc-camboriu.edu.br, hiryna@ifc-camboriu.edu.br, mozara@ifc-camboriu.edu.br

***Abstract.** This article presents a survey and data analysis about the student female presence in the Information Systems Bachelor course at Instituto Federal Catarinense, campus Camboriú. Comparative data with the male students' presence is presented and further discussed.*

***Resumo.** Este artigo apresenta o levantamento e análise dos dados da presença feminina discente no curso de Bacharelado em Sistemas de Informação do IFC-Camboriú desde a criação do curso. Apresenta-se comparativos com dados dos discentes do sexo masculino visando evidenciar as diferenças encontradas.*

1. Mulheres na informática

Segundo Felitti, em 1974 as mulheres representavam 70% da primeira turma do bacharelado em ciência da computação do Instituto de Matemática e Estatística (IME). Muito antes disso a equipe que programava o ENIAC era formada por mulheres. Mas a partir dos anos 2000 a realidade na área de computação começou a mudar. Um exemplo disso é citado pela influenciadora digital Camila Achutti [Achutti, 2016] que diz ter sido única mulher formada no IME no curso de ciência da computação [Felitti, 2015].

O Google preocupou-se com essa diferença entre número de homens e mulheres em 2014, avaliando que somente 30% dos funcionários da empresa em todo mundo eram mulheres. Segundo análise do Google, o baixo número de mulheres que se formam em ciência da computação, cerca de 18%, seria um dos motivos para a presença inferior das mulheres no mercado de TI. Nas grandes empresas de tecnologia a porcentagem de mulheres empregadas nas mesmas é bastante parecida com as apresentadas pelo Google [Cruz, 2014].

Para alguns autores, o estímulo e a interação com objetos tecnológicos nos homens vem desde a infância, enquanto nas mulheres isso não acontece. As escolas mistas deveriam reduzir os estereótipos, mas acontece o inverso. As inovações tecnológicas seduzem muito mais os homens, fato este que inibe o avanço das mulheres no mercado de trabalho tecnológico além das diferenças salariais [Rapkiewicz, 1998].

Visando colaborar na discussão sobre o esclarecimento e incentivo à participação feminina dentro dos cursos superiores da área de ciência da computação, este artigo apresenta um levantamento e análise dos dados da presença feminina discente no curso de Bacharelado em Sistemas de Informação do IFC-CAM desde sua criação em 2010 até os dias de hoje.

2. O ingresso feminino discente no curso de Bacharelado em Sistemas de Informação do IFC- CAM

O curso de Bacharelado em Sistemas de Informação do Instituto Federal Catarinense - Campus Camboriú - BSI/IFC-CAM - teve seu início no ano de 2010. O ingresso neste curso é anual, tendo atualmente como forma de ingresso a seleção pelo SISU e, posteriormente, caso haja vagas remanescentes, a análise curricular. As disciplinas são ofertadas semestralmente no período matutino.

O curso tem seu foco voltado para a formação do profissional que irá desenvolver ou gerenciar sistemas aplicativos e de gestão de informações. Entre os objetivos do curso estão: preparar profissionais para identificar as necessidades das organizações, criar e gerir soluções tecnológicas e desenvolver sistemas informatizados para organizar os fluxos de informação das empresas [IFC-CAM, 2016].

A seguir será apresentado o resultado do levantamento da presença feminina no curso BSI/IFC-CAM, através dos gráficos e suas respectivas discussões.

A figura 1 mostra o gráfico comparativo entre o número de homens e mulheres ingressantes no curso desde o primeiro ano, em 2010. O percentual médio de ingresso masculino no período em questão é de 80,3%, enquanto que o feminino é de 19,7%. Observa-se assim a superioridade no ingresso do público masculino, desde a criação do curso, o que nos remete a pergunta para discussão futura: quais os fatores que influenciam a predominância do ingresso masculino no curso do BSI/IFC-CAM.

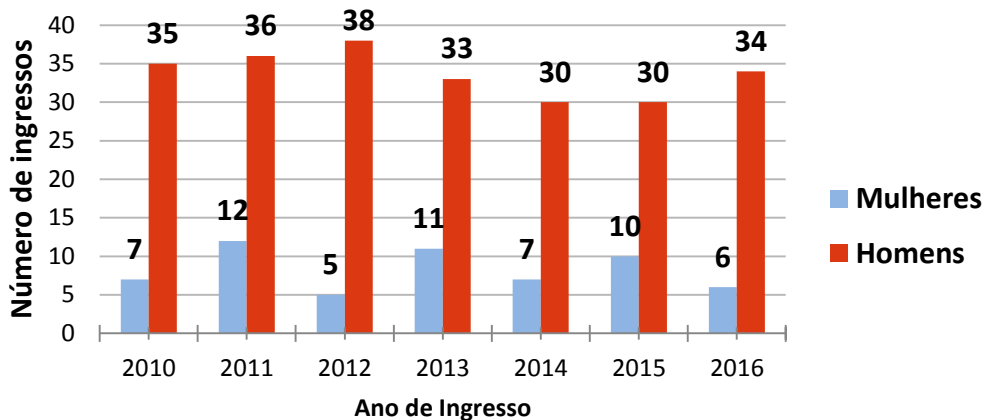


Figura 1- Relação de mulheres e homens ingressantes

A figura 2 apresenta a trajetória do ingresso das mulheres no BSI/IFC-CAM desde 2010. Observa-se que no ano de 2011 o curso teve o maior número de mulheres ingressantes, mas no ano seguinte também foi o menor ingresso. A média anual de mulheres ingressantes é de 07 mulheres/ano, o que ressalta o alto índice de ingresso feminino no ano de 2011, com 05 representantes a mais que a média.

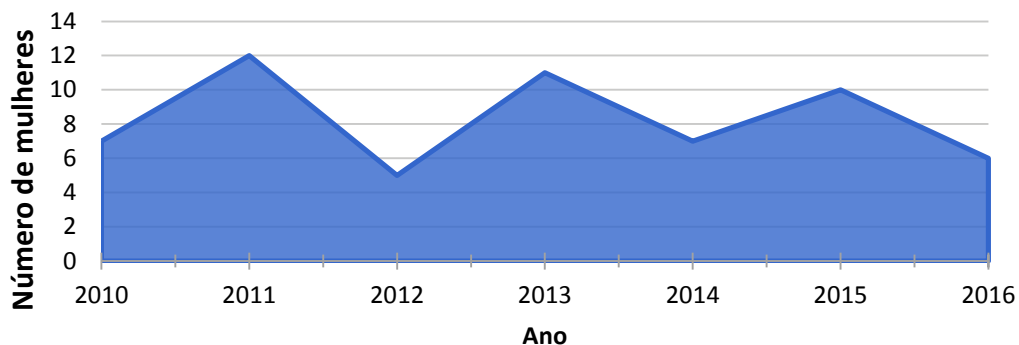


Figura 2- Trajetória do ingresso feminino

A figura 3 demonstra que o número de discentes matriculados a cada ano do sexo masculino cresceu de forma linear enquanto o dos matriculados do sexo feminino obteve um crescimento menos acentuado, permanecendo praticamente estável desde 2014.

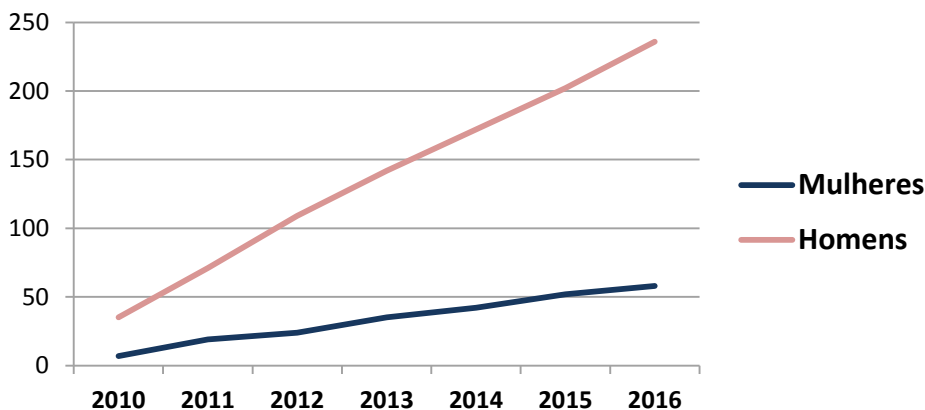


Figura 3 – Número de discentes matriculados por gênero

3. Formadas x formados

A figura 4 apresenta o número de formados por gênero. O gráfico apresenta uma quantidade quase seis vezes maior de formados do sexo masculino em relação aos discentes do sexo feminino.

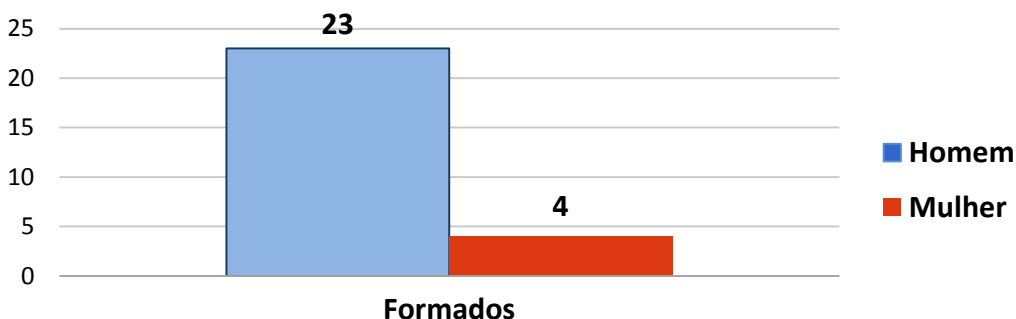
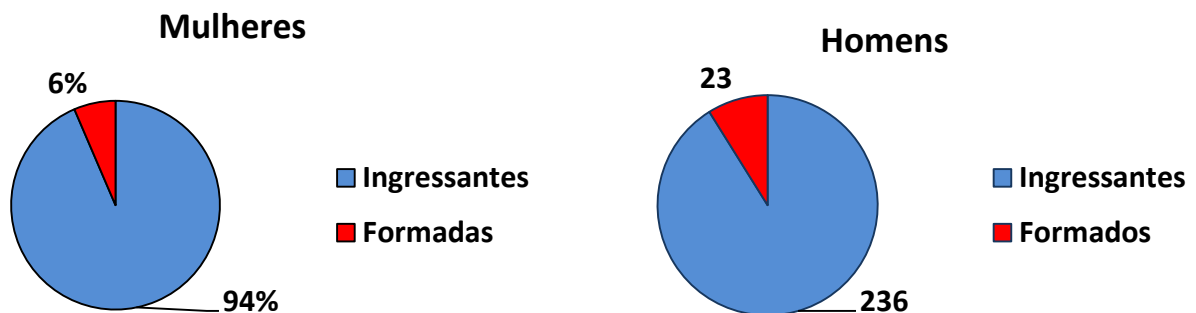


Figura 4 - Comparação entre formados por gênero

A seguir as figuras 5 e 6 apresentam o percentual de formados com relação ao ingresso por gênero. Estes gráficos demonstram que o número de mulheres formadas é somente 2,4% menor que o número de homens formados.



Figuras 5 e 6 - Porcentagem de mulheres formadas e porcentagem de homens formados

4. Considerações finais

Analisando todos os gráficos pode-se observar o pequeno ingresso feminino em relação ao masculino, mas a relação ingressantes/formados é equivalente proporcionalmente entre os gêneros.

Esta constatação específica aos anos analisados e ao curso do BSI/IFC-CAM, nos leva a proposição de um novo trabalho onde possa-se aprofundar as pesquisas para verificar quais os fatores que leva a um baixo ingresso feminino, já que dentre a amostra analisada, podemos evidenciar que as mulheres não têm dificuldades maiores que os homens ao fazer o curso do BSI/IFC-CAM pois a porcentagem das formadas em relação ao ingresso é a mesma que os homens.

Referências

- Achutti, C. (2016) “Mulheres na computação”, <https://mulheresnacomputacao.com/>, Maio.
- Cruz, L. (2014) “Mulheres buscam mais espaço e diversidade no mercado de tecnologia”, <http://link.estadao.com.br/noticias/geral,mulheres-buscam-mais-espaco-e-diversidade-no-mercado-de-tecnologia,10000030331>, Maio.
- Felitti, G. (2015) “Por que há menos mulheres no setor de tecnologia?”, <http://epocanegocios.globo.com/Informacao/Dilemas/noticia/2015/08/por-que-ha-menos-mulheres-no-setor-de-tecnologia.html>, Maio.
- IFC- CAM, Instituto Federal Catarinense- Campus Camboriú (2015) “Bacharelado em Sistemas de Informação”, <http://www.camboriu.ifc.edu.br/cursos-superiores/bacharelado-em-sistemas-de-informacao/>, Maio.
- Rapkiewicz, C. E. (1998). Informática: domínio masculino? <http://periodicos.sbu.unicamp.br/ojs/index.php/cadpagu/article/view/4172931>, páginas 169–200. Cadernos Pagu, Maio.

Mapeamento de iniciativas brasileiras que fomentam a entrada de mulheres na Computação

Maria Augusta S. N. Nunes^{1,2,4}, Carolina S. Louzada^{1,4}, Edilayne M. Salgueiro^{1,4},
Beatriz T. Andrade^{1,4}, Patrícia S. de Lima^{3,4}, Raquel M. C. T. Figueiredo^{2,4}

¹ DCOMP- Universidade Federal de Sergipe/UFS - São Cristóvão- SE

² PROCC -Universidade Federal de Sergipe/UFS -São Cristóvão- SE

³ Escola Estadual Atheneu Sergipense –Aracaju – SE

⁴ Projeto meninasnacomputacao.ufs.br –Aracaju – SE

gutanunes@gmail.com, carolina.louzada@hotmail.com, edilayne@ufs.br,
beatriz@ufs.br, paslima@globo.com, raquelmctf@gmail.com

Abstract. *This paper provides an overview of Brazilian initiatives that promote women into the computing area. The methodology used was based on Systematic Mapping. After analyzing the data we concluded that despite the low representation of women in the computing area, there is currently a significant number of Brazilian initiatives that encourage the migration of women in computing.*

Resumo. *Esse artigo tem como objetivo obter um panorama das iniciativas brasileiras que fomentam a entrada de mulheres da área da Computação. A metodologia utilizada foi baseada no processo do Mapeamento Sistemático. Após a análise dos dados concluiu-se que apesar da baixa representatividade das mulheres, há atualmente uma significativa quantidade de iniciativas brasileiras que fomentam a entrada destas na Computação.*

1. Introdução

De acordo com [Lima 2013], as mudanças culturais pelas quais a sociedade continuamente passa mudam a forma como as mulheres se veem e são vistas com relação à escolha das carreiras a seguir. Entretanto, apesar de suas capacidades, as mulheres ainda enfrentam obstáculos para serem aceitas e reconhecidas em áreas onde historicamente os homens predominam. Na Computação, as dificuldades e falta de incentivo refletem no baixo número de mulheres ingressantes nos cursos da área. O baixo percentual de mulheres nessa área não é só preocupante com relação às tentativas de minimizar as desigualdades dos gêneros, mas também preocupa o mercado de Tecnologia da Informação (TI) atual. Isso porque, de acordo com a IDC, no ano de 2013 houve uma carência de 39,9 mil profissionais de tecnologia e até 2015 esse número

deveria crescer para 117 mil [INFOEXAME 2013].

Diante desse contexto, nas universidades e no mercado de trabalho existe uma preocupação em criar estratégias de incentivo que visem diminuir essa baixa representatividade das mulheres na Computação [Louzada e outros 2014]. Tais estratégias, de iniciativa pública e privada, utilizam-se de uma significativa variedade de meios para divulgação como blogs, cursos, *workshops*, *mentoring*, palestras, oficinas, projetos, entre outros.

O objetivo deste artigo é obter um panorama das iniciativas brasileiras que fomentam a entrada de mulheres da área da Computação. O artigo está organizado da seguinte forma: na Seção 2 descreve-se a metodologia; na Seção 3 analisa-se os dados e na Seção 4 conclui-se o artigo.

2. Método de Pesquisa

A busca por iniciativas que estimulem o ingresso de mulheres na área da Computação é um trabalho amplo e não há um método de pesquisa específica para verificar e garantir que todas as iniciativas existentes sejam efetivamente inseridas nessa pesquisa. Assim, os autores deste artigo entenderam que era necessária um método próprio, mas semelhante ao método do Mapeamento Sistemático (MS), proposto por [Petersen e outros 2008] e [Kitchenham 2007].

As etapas executadas foram divididas como apresentadas nas seções a seguir, onde foram definidas (i) as questões de pesquisa e busca das iniciativas; (ii) os critérios de inclusão e exclusão; (iii) a classificação das iniciativas; (iv) e finalmente, a análise destas iniciativas. O mapeamento foi realizado em Agosto de 2014.

2.1 Questões de pesquisa e busca das iniciativas

Na primeira etapa do método, questões foram definidas de forma a limitar os dados procurados. As questões de pesquisa definidas foram: (i) Quais e quantas são as iniciativas brasileiras com base nas informações encontradas na web? (ii) Qual o ano de criação das iniciativas brasileiras?

Na segunda etapa, utilizaram-se *strings* de busca para pesquisar essas iniciativas brasileiras por meio de URLs. As *strings* foram formuladas em inglês e português objetivando mapear as iniciativas brasileiras também redigidas na língua inglesa, visando a internacionalização de tal iniciativa. Como não há uma base de dados científica para os tipos de dados procurados, tal como SCOPUS, IEEE, ACM, SPRINGER, como recomenda o método de Mapeamento Sistemático, optou-se, então, por realizar uma adaptação metodológica e, por conseguinte, recorreu-se à Internet usando como motor de busca o *Google*, que retorna os resultados mais acessados e populares de acordo com as *strings* utilizadas. O motor de busca, retornou então, uma quantidade significativa de resultados, mais de 600 milhões. De forma a limitá-los, estipulou-se pelos autores (por conveniência) que os dados seriam procurados nas dez primeiras páginas encontradas. Foram revisadas cem iniciativas com os mesmos termos abaixo, em português e cem em inglês. Com isso, foram selecionados um total de

duzentos sites. Deve-se ressaltar que as *strings* aplicadas diretamente no motor de busca Google foram aplicadas aos títulos das páginas e utilizaram a sintaxe a seguir: ((*TITLE: woman OR TITLE: women*) AND (*TITLE: computer*)); ((*TITLE: woman OR TITLE: women*) AND (*TITLE: computing*)); ((*TITLE: woman OR TITLE: women*) AND (*TITLE: information technology*)); ((*mulher OR mulheres*) AND (*computador*))OR ((*mulher OR mulheres*) AND (*Computação*))OR ((*mulher OR mulheres*) AND (*tecnologia da informação*)).

2.2 Critérios de inclusão e exclusão de dados

Tendo a relação das URLs encontradas na pesquisa, utilizaram-se critérios de inclusão e exclusão para encontrar os sites com dados relevantes para a pesquisa. Os sites incluídos foram aqueles que se propõem especificamente ao incentivo à mulheres no campo da Computação ou diminuem sua evasão. Já os excluídos, classificados neste artigo como não relevantes, não apresentam propostas de incentivo às mulheres na Computação ou somente mencionam as iniciativas e/ou projetos, como no caso de sites jornalísticos.

2.2 Classificação de dados

A partir desses critérios, foi aplicado o critério de classificação dos dados, assim, após análise de inclusão e exclusão, o número de sites foi reduzido. Dos duzentos sites encontrados, no total dezesseis foram de iniciativas brasileiras relevantes (redigido em português ou inglês).

As iniciativas brasileiras encontradas foram:

Women in Information Technology (WIT); Fórum Meninas Digitais; Android Smart Girls; meninas.comp; Projeto da UNIPLAC; Mulheres na Tecnologia; Roda da hacker; T.I.mosia; Emílias armação em bits; Feminino livre; Girl Geek Dinners Brazil; Inspirada na Computação; Meninas na engenharia; Projeto da UFMT- Mulheres na Computação; Projetomeninasmais; Blog mulheres na Computação.

As referências às URLs das iniciativas brasileiras estão disponíveis em: (<http://200.17.141.213/~gutanunes/wit2016/appendiceBrasil.pdf>).

3. Análise dos dados

Muitas das iniciativas pesquisadas utilizam mais de uma fonte de fomento ou patrocinador. Neste contexto, destacam-se as organizações criadas especificamente com o propósito de incentivar mulheres e universidades que desejavam inspirar e aumentar o número de mulheres em seus cursos de Computação. A análise desses dados é apresentada nas subseções seguintes.

3.1 Catalogação, quantidades e tipos de iniciativa

De acordo com o método de pesquisa, o primeiro passo foi catalogar título e as URLs

das respectivas iniciativas. Após essa etapa separaram-se os dados segundo critérios de inclusão e exclusão de iniciativas brasileiras. Após a seleção dos sites que atenderam esses critérios foi possível classificar os dados necessários. Esses dados classificados, ao serem analisados, ajudaram a responder às questões de pesquisa definidas na primeira etapa. O Gráfico 1 apresenta essas classificações obtidas dentre as diversas iniciativas.

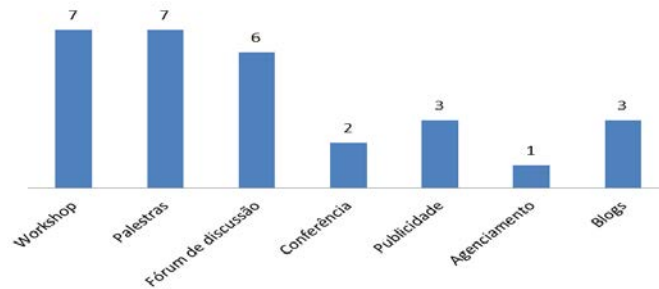


Gráfico 1. Relação entre quantidade e tipos de iniciativas brasileiras

Analisando o Gráfico 1, verifica-se que no Brasil os *workshops* e palestras são as iniciativas mais comuns com o intuito de fomentar as mulheres, obtendo sete sites de iniciativas cada um. O agenciamento aparece com um site, sendo o tipo de iniciativa menos utilizada. O caráter mais dinâmico e prático dos *workshops* e palestras faz que estas sejam estratégias úteis e eficientes para incentivar a entrada de mulheres na área de TI ou mantê-las, no caso de já estarem na área de TI. Os *blogs* possuem caráter pessoal e são, geralmente, administrados por uma pessoa que deseja informar a respeito de tudo que encontra e sabe sobre o tema em questão. Essa característica dos *blogs* pode explicar o menor número de sites com esse intuito, já que para incentivar as mulheres é preciso mais que veiculação de informações pela Internet.

3.2 Ano de criação

De acordo com a pesquisa verificou-se que existia baixa representatividade em termos quantitativos de sites que apresentassem iniciativas brasileiras específicas para as mulheres na Computação, como pode-se observar no Gráfico 2.

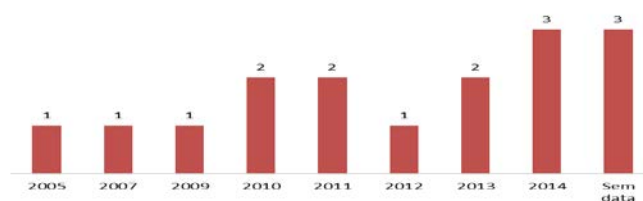


Gráfico 2. Ano de criação das iniciativas brasileiras

De 2005 a 2009 verificou-se que apenas uma iniciativa por ano foi criada. Entre 2010 e 2013 a média de iniciativas criadas aumentou. No ano 2014 observou-se a

criação de três iniciativas, o maior número encontrado desde o ano de 2005. Do total pesquisado, três iniciativas ou projetos não possuíam data de criação expressa no site, o que impediu sua extração.

4. Considerações Finais

Com o uso do método de pesquisa baseado no Mapeamento Sistemático, esse estudo teve como objetivo verificar iniciativas que fomentem a entrada de mulheres na Computação no Brasil. Apesar de ser uma pesquisa limitada devido à grande quantidade de resultados retornados e a não capacidade de lidar com eles na totalidade, tanto pela falta de informações nos sites, como pelo critério que usou-se para definir o limite de páginas efetivamente incluídas, obteve-se, ainda, um mapeamento preliminar das iniciativas brasileiras considerado interessante. Esse mapeamento forneceu uma visão geral das iniciativas existentes para a comunidade acadêmica, de forma que, a partir desse estudo, com o uso das *URLs* referenciadas, seja possível a extensão de estudos futuros que tratem de outros aspectos mais profundos não tratados nesse artigo. O baixo número de iniciativas brasileiras encontradas, embora seja representativo na data que a pesquisa foi realizada, pode não apresentar a situação atual das iniciativas por limitação do próprio método utilizado no mapeamento e pela data de aplicação. Como trabalho futuro sugere-se o uso de outro método mais adaptado a esse tipo de levantamento estatístico, que reflita precisamente as condições atuais das iniciativas que fomentem a entrada de mulheres na Computação no Brasil.

5. Agradecimentos

Ao MCTI/CNPq/SPM pelo fomento ao projeto N° 420160/2013-2. Ao CNPq pelas bolsas ITI e DT concedidas. À CAPES. À FAPITEC.

Referências

- Kitchenham, B. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Technical Report. Keele University at Staffordshire and University of Durham.UK.
- Infoexame (2013). Falta de profissionais de TI se agravará, diz IDC. Caderno: Notícias-Carreira. 19/03/2013. Disponível em: < <http://exame.abril.com.br/> >
- Lima, M. P. (2013).As mulheres na Ciência da Computação. In:Revista Estudos Feministas. Flórianópolis.
- Louzada, C. S.; Gomes, W. F.; Nunes, M. A. S. N.; Salgueiro, E. M.; Andrade, B. T.; Lima, P. S. (2014). Um mapeamento das publicações sobre o ingresso das mulheres na Computação. In: CLEI 2014 - VI Congresso da Mulher Latino-americana na Computação, Montevidéu.
- Petersen, K. *et al.*(2008). *Systematic mapping studies in software engineering*, ACM. New York.

Mapeamento de iniciativas estrangeiras em língua inglesa que fomentam a entrada de mulheres na Computação

**Maria Augusta S. N. Nunes¹, Carolina S. Louzada¹, Edilayne M. Salgueiro¹,
Beatriz T. Andrade¹, Patrícia S. de Lima², Raquel M. C. T. Figueiredo¹**

¹ DCOMP/PROCC- Universidade Federal de Sergipe/UFS— São Cristóvão- SE

² Escola Estadual Atheneu Sergipense –Aracaju— SE

gutanunes@gmail.com, carolina.louzada@hotmail.com, edilayne@ufs.br,
beatriz@ufs.br, paslima@globo.com, raquelmctf@gmail.com

***Abstract.** This paper provides an overview of foreign initiatives based on english language that promote women into the computing area around the world. The methodology used was based on Systematic Mapping. As a conclusion we found that the foreign initiatives of women in the computing area based on english language is quite representative worldwide that could encourage the migration of women in computing.*

***Resumo.** Esse artigo tem como objetivo obter um panorama das iniciativas estrangeiras baseadas em língua inglesa que fomentam a entrada de mulheres da área da Computação no mundo. A metodologia utilizada foi inspirada no processo de Mapeamento Sistemático. Concluiu-se que há, atualmente, uma significativa quantidade de iniciativas estrangeiras baseadas em língua inglesa distribuídas mundialmente que fomentam a entrada de mulheres na área da Computação.*

1. Introdução

Na Computação, as dificuldades e a falta de incentivo refletem no baixo número de mulheres ingressantes em cursos dessa área. O baixo percentual de mulheres nessa área não é só preocupante com relação às tentativas de minimizar as desigualdades dos gêneros, mas também preocupa o mercado de Tecnologia da Informação (TI) atual. Diante desse contexto, há uma crescente preocupação em criar estratégias de incentivo que visam diminuir essa baixa representatividade das mulheres na Computação [Louzada e outros 2014]. Existem algumas estratégias e ações, tais como a da Universidade de Stanford que disponibiliza um curso introdutório de Computação que foi adaptado para atender um leigo, além dos tradicionais graduandos da área; nesse curso foram apresentadas personalidades importantes na área de Computação incluindo também mulheres de sucesso. Outras diversas universidades tem criado ações, tais como minicursos ou oficinas para alunos de ensino médio, onde são apresentados os conceitos de programação e lógica de programação [Gomes e outros 2014; BlackGirlCode 2015]; ações, também, apresentadas em forma de concursos que estimulam o empreendedorismo desde o ensino médio, para fomento ao gênero [Technovation 2015]. Durante essas ações tem se percebido que as meninas, geralmente, não conhecem o perfil do profissional de TI e, conseqüentemente, não vislumbram as oportunidades de trabalho na área, esse é um dos fatores que pode determinar a falta de interesse delas nos cursos da área de Computação. Observa-se, também, que a formação de grupos de

suporte às mulheres nas universidades auxilia na retenção dessas mulheres nos cursos de Computação [Beck 2007], demonstrando que as medidas adotadas nas universidades podem ser um importante estímulo para manter o interesse das mulheres no mercado de TI. O objetivo deste artigo, então é buscar e quantificar iniciativas estrangeiras¹ baseadas em língua inglesa de fomento ao gênero para a área da Computação. Nesse artigo utilizou-se o método de Mapeamento Sistemático de URLs relacionados à inclusão de mulheres na área de Computação e afins no contexto estrangeiro usando a língua inglesa como base. Esse trabalho está organizado da seguinte forma: na Seção 2 descreve-se a metodologia; na Seção 3 realiza-se uma análise dos dados encontrados e na Seção 4 descreve-se as considerações finais.

2. Metodologia

A busca por iniciativas que estimulem o ingresso de mulheres na área da Computação é um trabalho amplo. Não há uma metodologia específica para verificar e garantir que todas as iniciativas estrangeiras baseadas em língua inglesa existentes estejam todas incluídas na pesquisa. Foi utilizada uma adaptação do método do Mapeamento Sistemático (MS), proposto por *Petersen e outros* (2008) e *Kitchenham* (2007). De acordo com ambos os autores, o MS é uma metodologia que fornece uma visão geral do que se quer mapear por meio da quantificação de evidências. A partir dessa quantificação é possível verificar e interpretar os dados encontrados e, dessa forma, encontrar tendências que são essenciais para a formação de um panorama do tema abordado. As etapas executadas foram divididas da seguinte forma: (i) Definição das questões de pesquisa; (ii) Busca de iniciativas, projetos e organizações; (iii) Seleção de iniciativas, projetos e organizações relevantes; (iv) Extração e classificação de dados; (v) Análise dos dados. Os dados foram coletados em Agosto de 2014.

2.1 Questões de pesquisa e busca de projetos e iniciativas

Na primeira etapa do método, questões foram definidas de forma a limitar os dados procurados. As questões de pesquisa proposta foram: 1. Quais e quantas são as iniciativas estrangeiras baseadas em língua inglesa com base nas informações encontradas na web? 2. Quando essas iniciativas foram criadas (ano de fundação)? 3. Em quais países essas iniciativas são aplicadas? Na segunda etapa, utilizaram-se *strings* de busca para pesquisar essas iniciativas estrangeiras baseadas em língua inglesa por meio de URLs. Como não há uma base de dados científica para os tipos de dados procurados, tal como SCOPUS, IEEE, ACM, SPRINGER, como recomenda a metodologia de mapeamento sistemático, foi realizada uma adaptação metodológica e, então, recorreu-se a uma base de dados da Internet. Foi utilizado o motor de busca *Google*, que retorna os resultados mais acessados e populares de acordo com as *strings* pesquisadas. O motor de busca, retornou então, uma quantidade significativa de resultados, mais de 600 milhões. De forma a limitá-los, estipulou-se que os dados seriam procurados nas dez primeiras páginas encontradas, visando criar um panorama inicial para o mapeamento. Foram revisadas 100 iniciativas, na língua inglesa, com os mesmos termos: *((TITLE: woman OR TITLE: women) AND (TITLE: computer)); ((TITLE: woman OR*

¹ A palavra « estrangeira » refere-se a « fora do Brasil » . Nesse artigo não usou-se o termo « internacional » pois objetivou-se mapear as iniciativas desenvolvidas fora do Brasil, fossem elas internacionais ou não. Exemplificando, conforme a visão dos autores, uma iniciativa brasileira, pode ser considerada internacional, quando ela atinge um público estrangeiro, uma forma de atingir esse público é a página web da iniciativa estar desenvolvida, também, em uma língua estrangeira, como o inglês, por exemplo.

TITLE: women) AND (TITLE: computing)); ((TITLE: woman OR TITLE: women) AND (TITLE:information technology)).

2.2 Critérios de inclusão e exclusão e extração de dados

Tendo a relação dos URLs encontrados na pesquisa, utilizaram-se critérios de inclusão e exclusão para encontrar os *sites* com dados relevantes para a pesquisa. Os *sites* incluídos foram aqueles, na língua inglesa, criados pelas iniciativas que apresentam propostas estrangeiras que incentivem especificamente mulheres no campo da Computação ou diminuam sua evasão. Já os excluídos, classificados neste artigo como não relevantes, não estão na língua inglesa, não são brasileiros e não apresentam propostas de incentivo às mulheres na Computação ou somente mencionam as iniciativas e projetos, como no caso de *sites* jornalísticos. A partir desses critérios o número de *sites* foi reduzido. Dos cem *sites* encontrados no total, cinquenta e nove foram de iniciativas estrangeiras baseadas em língua inglesa, sendo elas: *Women in Technology (Austrália)*, *IEEE Women in Engineering*, *NCWIT*, *Women in Technology(Reino Unido)*, *Associação Profissional WITI*, *WIT Education Foundation*, *Technovation Challenge*, *Blog ACM-W COUNCIL WOMEN IN COMPUTING NEWS*, *Railsgirls*, *Girls who code, she + +*, *pyladies*, *Women techmakers*, *ACM-W*, *The Center for Women In Technology(CWIT)*, *Debian*, *Genderchangers*, *Women into Science and Engineering*, *IWT*, *Fabfems*, *National Girls Collaborative Project*, *The Computer Research Association's Committee on the Status of Women in Computing Research*, *Women in Engineering and Computing Career Day*, *Women in Computer Science*, *Arabwic*, *Maryland Center for Women in Computing*, *Association for Women in Computing*, *Valentina project*, *Jinnah University For Women*, *Anitaborg*, *College of Computing and Information Women in Technology*, *Young Women in Computing*, *Ada initiative*, *woman@CC*, *Black girls code*, *Women in Computer Science at North Carolina State University*, *NSF ADVANCE Program*, *CloudNOW*, *BinaryGirl*, *EngineerGirl*, *Institute for Women in Trades, Technology & Science's*, *Women in games international*, *Girlstart*, *DIYgirls*, *Her ideas in motion*, *Girls in Tech*, *Duchess-france*, *Girls learning code*, *Linuxchix*, *Projeto Ubuntu Women*, *Women in Computer Science in Boston University*, *Webgrrls*, *Devfest*, *Women2*, *Million women mentors*, *women in stem*, *Women who code*, *Digigirlz* (Links disponíveis em: (<http://200.17.141.213/~gutanunes/wit2016/apendiceEstrangeiro.pdf>) Após a seleção dos *sites* que atenderam aos critérios foi possível extrair os dados necessários. Esses dados analisados ajudaram a responder às questões de pesquisa definidas. Os dados extraídos foram: o título da iniciativa, países de alcance ou ação e ano de fundação.

3. Análise dos dados

Muitas das iniciativas pesquisadas utilizam mais de um modo/fonte de fomento. Neste contexto, destacam-se as organizações criadas especificamente com o propósito de incentivar mulheres e universidades que desejavam inspirar e aumentar o número de mulheres em seus cursos de Computação. A análise desses dados é apresentada nas subseções seguintes.

3.1 Catalogação, quantidades e tipos de iniciativas

De acordo com a metodologia, o primeiro passo foi catalogar o título e as URLs das respectivas iniciativas estrangeiras baseadas em língua inglesa. Por meio da leitura das propostas contidas nos *sites* das iniciativas foi possível verificar os tipos de iniciativas pesquisadas. Ao final da leitura dessas iniciativas/propostas foram verificados os seguintes tipos de ações: *workshops*, palestras, fóruns, tutoria, cursos, conferências, consultoria, publicidade, agenciamento, competição, jogos de ensino, desenvolvimento de *softwares*, programas pedagógicos e blogs. O Gráfico 1 apresenta essas classificações

obtidas dentre as diversas iniciativas estrangeiras baseadas em língua inglesa.

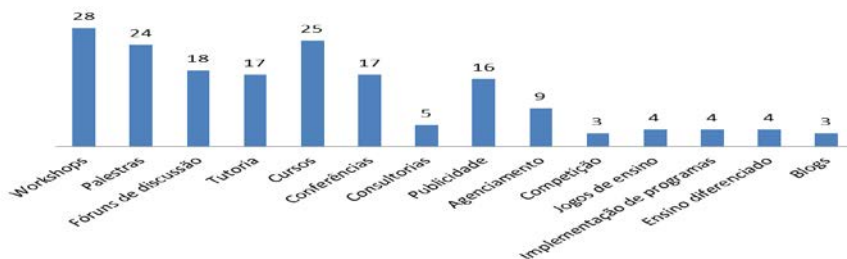


Gráfico 1. Relação entre quantidade e tipos de iniciativas estrangeiras

3.2. Ano de fundação

Com relação às iniciativas estrangeiras baseadas em língua inglesa verificou-se um grande contraste nos anos de fundação, como é visto no Gráfico 2. A iniciativa mais antiga encontrada na pesquisa foi fundada no ano de 1978. A criação de uma ou duas iniciativas por ano manteve-se até o ano de 2003, em que houve o lançamento de duas iniciativas estrangeiras baseadas em língua inglesa. A partir de 2004 a média de iniciativas aumentou e no ano de 2012 verificou-se o maior número de iniciativas criadas, com a criação de sete iniciativas. Em 2013 novamente esse número caiu e voltou a crescer em 2014. Devido à falta de informações nos *sites* não foi possível extrair a data de criação de quatro iniciativas. Contudo, essa falta de informação não prejudica o entendimento quanto ao crescimento de iniciativas de incentivo às mulheres na Computação, sejam públicas ou privadas.

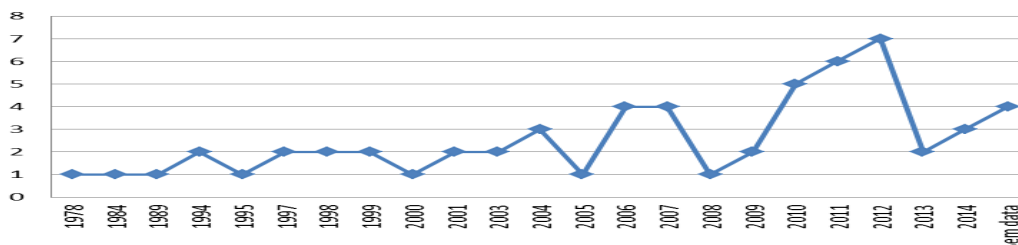


Gráfico 2. Ano de fundação das iniciativas

Em Louzada *et. al* (2014) foi encontrado no ano de 2002, o maior número de publicações científicas sobre o incentivo a entrada de mulheres na Computação. Comparando com o Gráfico 2, no mapeamento de iniciativas, verifica-se que dez anos depois, em 2012, foram criadas o maior índice numérico de estratégias de incentivo. A primeira publicação relacionada surgiu em 1975, no entanto observa-se que a primeira iniciativa foi fundada somente em 1978.

3.3. Países de alcance

A partir da leitura do Gráfico 3 verificou-se que 56% das iniciativas têm como foco os EUA, e em segundo lugar, as iniciativas de alcance internacional, com 32%. As iniciativas de alcance internacional são aquelas que são criadas em determinado país e, também, atuam em outros países. O Canadá possui o menor percentual com relação ao número de iniciativas, ficando com 1%. Também, com um baixo número de iniciativas, ocupando 2% do total, tem-se a Austrália, França, Emirados Árabes e Paquistão. Essa relação entre países de alcance é importante para entender-se a forte conexão internacional entre diversos países que possuem a necessidade de incentivar as mulheres

na Computação.

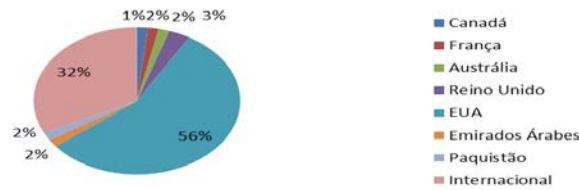


Gráfico 3. Relação percentual dos países em que as iniciativas atuam

4. Considerações Finais

Por meio do Mapeamento Sistemático, esse artigo teve como objetivo verificar as iniciativas estrangeiras baseadas em língua inglesa que fomentem a entrada de mulheres na Computação. Foram encontradas cinquenta e nove iniciativas estrangeiras baseadas em língua inglesa. A criação de iniciativas de alcance internacional mostra a preocupação latente em conseguir bons profissionais para o mercado de TI e diminuir a diferença percentual entre homens e mulheres na área. Os EUA são foco de 56% do total das iniciativas pesquisadas e isso pode mostrar um maior percentual na entrada de mulheres nos cursos de Computação com relação a outros países, mesmo considerando a língua base como o inglês. Muitos países que, também, possuem o inglês como primeira língua apresentam uma baixa representatividade, nesse mapeamento, em termos de iniciativa, tais como: Reino Unido com 3%, Austrália com 2% e parte do Canadá (anglófona) com 1%. Independente do país de origem das iniciativas, 36 % destas iniciativas, em língua inglesa, tiveram um alcance internacional, ou seja, as iniciativas criadas em determinado país não se concentraram somente nesse mesmo país, mas buscaram atingir diversos outros países. Sugere-se para trabalhos futuros que se investigue outras línguas para que se tenha um panorama mundial concreto considerando a distribuição territorial das iniciativas nos dias atuais. É interessante também pesquisar sobre a relação dos países e o efeito causado pelas iniciativas na efetiva entrada de mulheres nos cursos de Computação de suas Universidades.

Referências

- Beck, J. (2007). *Forming a Women's Computer Science Support Group*. In: ACM SIGCSE'07. EUA.
- Gomes, W. F.; Louzada, C. S.; Nunes, M. A. S. N.; Salgueiro, E. M.; Andrade, B. T. (2014) Incentivando meninas do ensino médio à área de Ciência da Computação usando o *Scratch* como ferramenta. In: CBIE 2014 – 20^o Workshop de Informática na Escola (WIE).
- Kitchenham, B. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*, Technical Report. Keele University.UK.
- Louzada, C. S.; Gomes, W. F.; Nunes, M. A. S. N.; Salgueiro, E. M.; Andrade, B. T.; Lima, P. S. (2014). Um mapeamento das publicações sobre o ingresso das mulheres na Computação. In: CLEI 2014 - VI Congresso da Mulher Latino-americana na Computação, Montevidéu.
- Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M.(2008). *Systematic mapping studies in software engineering*, ACM. New York.
- BlackGirlCode(2016). Disponível em: <http://www.blackgirlsgcode.com/>. Acesso 19/04/2016.
- Technovation(2016). Disponível em: <http://www.technovationchallenge.org/>Acesso 19/04/2016.

Meninas Digitais Regional Mato Grosso: Práticas Motivacionais no Ensino Médio para a Equidade de Gêneros nas Carreiras e Cursos de Computação e Tecnologias

Karen da Silva Figueiredo¹, Pedro Clarindo da Silva Neto², Cristiano Maciel¹

¹Instituto de Computação – Universidade Federal de Mato Grosso (UFMT)
Av. Fernando Corrêa da Costa, 2367, Boa Esperança – Cuiabá– MT – Brasil

²Instituto Federal de Mato Grosso – Campus Avançado de Tangará da Serra (IFMT)
Rua 28, 980N, Jd Horizonte – 78300-000 - Tangará da Serra – MT – Brasil.
karen@ic.ufmt.br, pedro.neto@tga.ifmt.edu.br, cmaciel@ufmt.br

***Abstract.** This paper presents the actions realized between 2015 and 2016 by the project Digital Girls of Regional Mato Grosso which aims to realize motivational and informational practices with High School students in the state of Mato Grosso. The purpose of the project is to achieve gender equality in careers and courses of Computing in the region by encouraging and promoting women's participation.*

***Resumo.** Este trabalho apresenta as ações do projeto Meninas Digitais da Regional Mato Grosso que tem como propósito a realização de práticas de caráter motivacional e informativo com alunas de Ensino Médio no estado de Mato Grosso, visando a equidade de gênero nas carreiras e cursos das áreas de Computação e Tecnologias da região pelo incentivo e promoção da participação feminina. Este artigo tem como objetivo relatar os trabalhos realizados pela regional Mato Grosso durante o período de 2015-2016.*

1. Introdução

De modo geral, as pesquisas costumam revelar que as mulheres são minoria na Computação. Entretanto, dados recentes revelam um cenário ainda pior: as mulheres brasileiras perderam representatividade nos cursos relacionados à Computação. Em 1991, as mulheres somavam 34,9% dos alunos matriculados em cursos de Computação e afins, em 2013 passaram a representar 15,53% dos ingressantes, e destas apenas 13,6% concluem o curso, segundo o Censo da Educação Superior [INEP 2015].

Buscando uma solução para o cenário descrito, diversas iniciativas estão emergindo em prol do incentivo e promoção da participação feminina nas áreas de Computação e afins. Uma destas iniciativas é o Programa Meninas Digitais da Sociedade Brasileira de Computação (SBC) que tem como objetivo divulgar a área de Computação para despertar o interesse de estudantes (principalmente do gênero feminino) do ensino médio/tecnológico ou dos anos finais do ensino fundamental [Maciel e Bim 2016]. Deste programa, surgiu uma série de projetos em instituições no Brasil, sendo uma delas o Meninas Digitais Regional Mato Grosso (MD-MT).

O MD-MT propõe a realização de ações estratégicas que possam motivar alunas do Ensino Médio/Tecnológico a seguirem carreiras nas áreas de tecnologia. Este projeto teve início em maio de 2015 e possui um cronograma de longo prazo, prevendo a execução de atividades em diferentes municípios do estado de Mato Grosso. O projeto possui caráter extensionista, mas suas ações buscam estreitar os laços entre ensino, pesquisa e extensão.

O MD-MT tem como objetivo geral a realização de práticas de caráter motivacional e informativo com alunas de Ensino Médio no estado de Mato Grosso visando à equidade de gêneros nas carreiras e cursos das áreas de Computação e Tecnologias da região pelo incentivo e promoção da participação feminina.

A fim de alcançar o objetivo geral do projeto, o mesmo conta com os seguintes objetivos específicos:

- 1) Realização de pesquisas sobre os índices estaduais de participação feminina em Computação e Tecnologias nas IES e no mercado do estado;
- 2) Produção de materiais informativos sobre os cursos de Computação e Tecnologias do estado e sobre as diversas carreiras possíveis nas áreas de Computação e Tecnologias;
- 3) Planejamento, elaboração e execução de atividades motivacionais e informativas com alunas de Ensino Médio sobre conteúdos, disciplinas e carreiras das áreas de Computação e Tecnologias;
- 4) Realização de atividades integradoras com as alunas de Ensino Médio e alunas de graduação, alunas de pós-graduação, professores e outros profissionais das áreas de Computação e Tecnologias;
- 5) Avaliação dos impactos das atividades realizadas pelo projeto no cenário estadual.

Para alcançar cada um dos objetivos específicos, o projeto Meninas Digitais Regional Mato Grosso possui uma série de ações estratégicas programadas. Na próxima seção, são descritos de forma breve, devido ao limite deste artigo, os resultados preliminares das ações executadas no primeiro ano do projeto (2015-2016).

2. Práticas Realizadas

No período de 2015-2016, o Meninas Digitais Regional Mato Grosso já realizou diversas práticas informativas e motivacionais com alunas do Ensino Médio, Fundamental, Superior e a comunidade em geral, atendendo mais de 580 pessoas de 12 municípios do estado. As ações do projeto são concentradas principalmente nos municípios de Cuiabá (capital do estado) e Tangará da Serra (pólo do projeto no interior).

A seguir, são listadas algumas das principais práticas realizadas pelo projeto neste período (veja mais detalhes no site¹), enumeradas por objetivo supracitado:

¹ <http://meninasdigitaismatogrosso.org/>

- 1.1)** Levantamento e análise dos dados da participação feminina no curso de Bacharelado de Sistemas de Informação no estado [Mendes e Figueiredo 2016];
- 2.1)** Produção e distribuição de material gráfico impresso (flyer) sobre os cursos de Computação e afins do estado e sobre as diferenças entre eles;
- 2.2)** Produção e divulgação de material gráfico e informativo online sobre diversas temáticas relacionadas à mulheres na Computação e tecnologias através do site do projeto²;
- 3.1)** Realização de oficinas de MIT App inventor, Desenvolvimento de Jogos (Construct2), Programação Web (HTML e CSS), História da Computação, Gêneros e Tecnologia, Edição de Som Digital com alunas de Ensino Médio;
- 3.2)** Realização de oficina de programação básica com a Hora do Código com alunos de Ensino Médio e Fundamental;
- 3.3)** Realização de palestras sobre equidade de gêneros nas carreiras e cursos de Computação e Tecnologias em empresas, IES e eventos do estado³;
- 3.4)** Participação com alunas de Ensino Superior na Maratona de Programação da SBC 2015;
- 3.5)** Participação com alunas de Ensino Médio e Superior no Technovation Challenge 2016;
- 4.1)** Realização de eventos: 1º Encontro App Inventor, I Fórum Meninas Digitais Regional MT: Tangará da Serra⁴, CineDebate do filme CODEGIRL, Apresentação do Technovation Challenge Brasil 2016 em comemoração ao aniversário da Ada Lovelace, 1º Fórum TPM – Teorias, Programação e Métodos, Mesa Redonda Meninas Digitais: Iniciativas inclusivas da mulher no mundo tecnológico;
- 4.2)** Participação com alunas de Ensino Técnico de Informática em desfile cívico municipal;
- 5.1)** Participação em eventos nacionais para apresentação e discussão dos resultados do projeto: 12ª Semana Nacional de Ciência, Tecnologia e Inovação⁵, Computer on the Beach 2016, 9º WIT Women in Information Technology e Fórum Meninas Digitais 2015, I Diálogo Elas nas Exatas.

A Figura 1 apresenta alguns registros imagéticos de algumas das ações previamente citadas.

² <http://meninasdigitaismatogrosso.org/>

³ http://www.bemnoticias.com.br/cidades/id-338848/forum_regional_apresenta_projeto_meninas_digitais

⁴ <http://www.radiopioneira.com.br/index.php/new/31474/1-forum-do-projeto-meninas-digitais-e-realizado-no-ifmt-tangara>

⁵ <https://sbcmmt.wordpress.com/2015/10/08/ri-da-ufmt-cuiaba-leva-a-semana-da-sbc-para-a-12a-semana-nacional-de-ciencia-tecnologia-e-inovacao/#more-195>



Figura 1. Fotos de oficinas do projeto: A) oficina Gêneros e Tecnologias com alunas de Ensino Médio; B) oficina MIT App Inventor com alunas de Ensino Médio; C) oficina Hora do Código com alunas de Ensino Fundamental e Médio

Observou-se o sentimento de apropriação das participantes das ações, no sentido de entenderem que podem ser estudantes e profissionais da área da computação, de que mulheres contribuíram e contribuem de forma significativa para a ciência e que também pode ser divertido trabalhar nessa área, desmistificando estereótipos comuns. Foi possível observar também a assiduidade na frequência nos encontros periódicos no período de execução apresentado.

3. Conclusões

Como parte da sociedade, as universidades têm papel fundamental na formação dos profissionais e devem criar estratégias de conscientização da população com o intuito de aumentar a visibilidade dos problemas referentes às relações de gêneros na educação e no mercado. Acredita-se que iniciativas como o projeto Meninas Digitais Regional Mato Grosso são fundamentais para motivar e informar as alunas com relação aos cursos e carreiras de Computação e afins, enquanto estas estão em um momento de amadurecimento e escolhas no Ensino Médio, principalmente em uma região que aparece em penúltimo lugar na quantidade de cursos da área [SBC 2013], fazendo-se necessário um incentivo ainda maior. Deste modo, o Meninas Digitais Regional Mato Grosso pretende continuar aplicando seu plano de ações, visando alcançar os 5 objetivos traçados e buscando ainda ampliar a sua atuação em números no estado.

Referências

- INEP (2015) “Resumo Técnico da Educação Superior 2013”, Diretoria de Estatísticas Educacionais DEED, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, Brasília-DF, 82 p.
- Maciel, C., Bim, S. A. (2016) “Programa Meninas Digitais - ações para divulgar a Computação para meninas do ensino médio”, In: Computer on the Beach 2016, Florianópolis, SC, p. 327-336.
- Mendes, L. B., Figueiredo, K. S. (2016) “Analisando a Participação Feminina no Curso de Sistemas de Informação da Universidade Federal de Mato Grosso”, In: Computer on the Beach 2016, Florianópolis, SC, p. 109-117.
- SBC (2013) “Educação Superior em Computação, Estatísticas – 2013”, Relatório Técnico, Disponível em: <<http://www.sbc.org.br/documentos>>.

Meninas e Jovens na Computação

Maria Eduarda Romagna¹, Letícia Heinzmann², Júlia Marques Carvalho da Silva³

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - Campus
Bento Gonçalves¹²³

maria.e.romagna@gmail.com¹, leticiaheinzmann@gmail.com²,
julia.silva@bento.ifrs.edu.br³

Abstract. *Over the years the Information Technology underwent an expansion that is being accompanied by the female audience and the young, although not in the same proportion as others. The project “Meninas e Jovens na Computação (which means, in literal translation, “Women and Teenagers in Information Technology”) carries out actions aimed at contributing to include this population in the area of IT, through events to integrate and bring the community into the activities developed. In addition to encouraging the permanence in the course, also contributing to people discover their interest by IT. The main of these activities was the development of web programming distance education courses and the research “Mulheres na Computação” (“Women in IT”).*

Resumo. *Ao longo dos anos a Informática passou por uma expansão exponencial que foi acompanhada pelo público feminino e pelos jovens, mas notavelmente em menor proporção do que pelos outros públicos. O projeto Meninas e Jovens na Computação realiza ações que visam contribuir para aumentar a participação deste público na área de TI por meio de eventos para integrar e abranger a comunidade nas atividades desenvolvidas, além de incentivar a permanência nos cursos, também contribuindo para as pessoas descobrirem seu interesse pela área de TI. A principal dessas ações foi o desenvolvimento de cursos de programação web EAD, com vagas reservadas a mulheres e jovens e a pesquisa “Mulheres na Computação”.*

1. Introdução

A Informática é uma área em grande expansão, segundo o IBGE (2006). Conforme levantamento realizado pela IDC Brasil, a área já apresenta 40 mil vagas disponíveis e a previsão é que o número suba para 117 mil no final de 2015. Diante disso, a desigualdade de gênero se faz ainda mais evidente. Conforme a pesquisa realizada pelo PNAD - IBGE em 2009, 79% das mulheres que ingressam em cursos acadêmicos de TI não concluem o curso, e apenas 20% dos profissionais de Tecnologia da Informação no Brasil são mulheres. Ainda, para ampliar o número de mulheres e jovens interessados pela informática, o Ministério da Ciência, Tecnologia e Inovação lançou, em 2013, o Programa Ensino Médio Inovador, onde buscavam incentivar principalmente meninas cursando o ensino médio ou a graduação a inserir-se na área, além de permanecer na mesma. Com base nesses dados, o projeto Meninas e Jovens na Computação tenta entender o porquê das desistências e busca atrair mais mulheres e jovens para a área por

meio de ações de empoderamento, como o curso de capacitação de programação web EaD e a pesquisa “Mulheres na Computação”.

2. Os cursos

Para incentivar a aprendizagem da comunidade foi oferecido um curso presencial de informática básica, abrangendo temas como a utilização do computador e seus softwares básicos. Também foram oferecidos cursos de programação web online abertos a toda a comunidade. A plataforma utilizada nos cursos foi o *Moodle*, ferramenta que oferece um ambiente propício para a realização dos cursos. Cada aula tinha a duração de uma semana, tendo cada semana um assunto. Por exemplo, na primeira semana de PHP viu-se Introdução ao PHP, na segunda estudou-se Operadores e Desvios Condicionais e assim nas seguintes semanas. Dentre os cursos oferecidos estiveram Lógica de Programação, HTML, CSS, JavaScript e PHP.

O curso de capacitação foi uma das grandes ações realizadas pelo projeto, vindo com a proposta de atingir um número maior de pessoas e auxiliando a disseminar o conhecimento de programação web a um público de nível superior. O mesmo encontra-se finalizado e obteve, num total de soma entre todos os cursos, 77 inscritas mulheres e 139 homens.

Para dar suporte aos alunos matriculados, os bolsistas do projeto Meninas e Jovens na Computação foram escalados como tutores/professores. Dividiram-se para que cada tutor receba-se, distribuídos de acordo com a sua localização geográfica, um pequeno número de alunos. O tutor era responsável por corrigir os exercícios enviados, monitorar presença do aluno no curso e responder eventuais questionamentos.

O curso de HTML foi o primeiro a ser concluído. Pela grande procura foi ofertado em três versões. Somando seus participantes houve 45 mulheres e 77 homens inscritos sendo que desses apenas 8 mulheres concluíram em contraste aos 18 homens. Em CSS, oferecido duas vezes, o curso obteve 45 mulheres e 77 homens inscritos, assim como em HTML, mas o número de concluintes é ainda menor e mais alarmante: 6 mulheres para 17 homens. Para JavaScript a procura foi mais baixa, totalizando 41 mulheres e 64 homens participando e apenas 1 mulher concluinte para 6 homens. No curso de PHP foram, entre mulheres e homens, respectivamente, 46 e 76 inscritos e 5 e 10 concluintes. Por último, Lógica de Programação obteve, por sua vez, 53 mulheres e 79 homens inscritos. Foi o único dos cursos a ter um número maior de mulheres concluintes: 13 para apenas 9 homens. A alternativa escolhida para combater esse quadro foi destinar 50% das vagas dos cursos exclusivamente para mulheres e jovens.

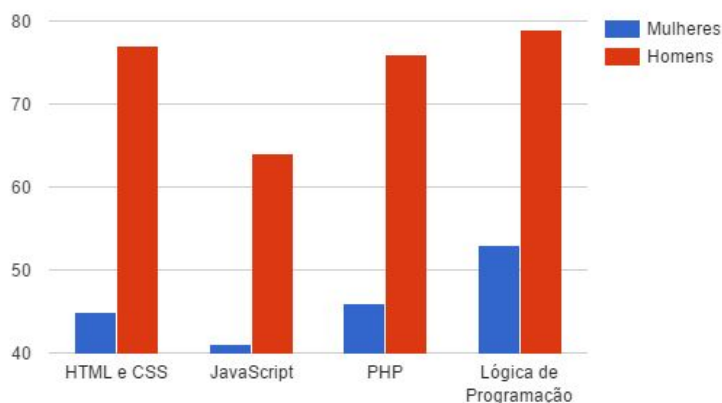


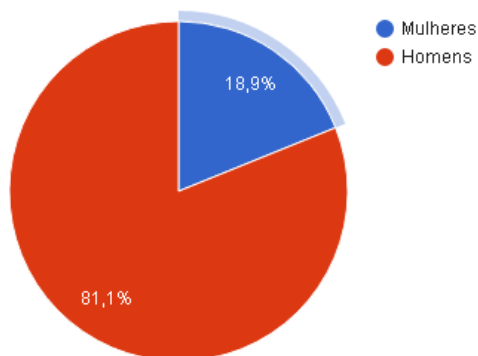
Figura 1. Gráfico comparativo de homens e mulheres nos cursos de programação web EAD.

3. Pesquisa “Mulheres na Computação”

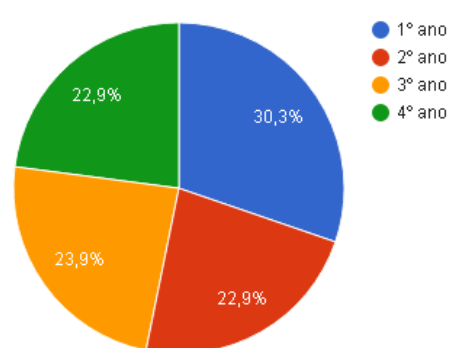
A pesquisa foi realizada a fim de descobrir e avaliar a presença da mulher dentro da área de TI atualmente. A pesquisa foi aplicada a alunos dos cursos de Informática para Internet da instituição. O questionário apresentava questões que buscavam entender os motivos para a escolha do curso, interesses em seguir na carreira, entre outras.

Por meio desta pesquisa é possível observar a desigualdade no número de mulheres na área quando comparado ao número de homens. Dentre todas as turmas entrevistadas pôde-se concluir que em média havia somente 20% de presença feminina nos discentes de informática. Dentre as alunas mulheres do 1º, 2º, 3º e 4º ano, respectivamente, há 30,3%, 22,9%, 23,9% e 22,9% em comparação aos alunos homens; 28,4% das mulheres questionadas alegaram já ter sofrido algum preconceito de gênero na escola ou em seu ambiente de trabalho; 94,6% das pessoas acredita que a maior parte da turma seja formada por homens; 86,5% concorda que há, sim, desequilíbrio entre homens e mulheres na área; e 81,1% dos entrevistados são homens.

Integrantes do curso



Mulheres no curso



Figuras 2 e 3. Gráfico comparativo de integrantes do cursos Técnico em Informática para Internet do IFRS-BG; gráfico que mostra a distribuição das integrantes do sexo feminino.

4. Considerações Finais

Conforme a pesquisa realizada em 2009, há desigualdade de gênero e baixo interesse na área de TI, em especial, de mulheres e jovens. O projeto “Mulheres e jovens na computação” visa, através de múltiplas ações, a inserção deste público. Para isto foram desenvolvidas atividades para a comunidade interna e externa da instituição.

O curso de programador web colabora para muitas pessoas terem sua primeira experiência na área de programação além de instigar a aprendizagem na mesma. As ações descritas tem o propósito de que possam auxiliar as pessoas envolvidas, no descobrimento de seu interesse na área de TI.

Referências

- IBGE (2009). Pesquisa Nacional de Amostra Por Domicílio (PNAD), Síntese de Indicadores Sociais. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/sinteseindicossociais2008/default.shtm>. Acessado em 01 Novembro 2015
- IDC Brasil (2015). IDC Brasil aponta que mercado de TIC deve crescer 5% e movimentar US\$ 165,6 bilhões no país em 2015. Disponível em: <http://br.idclatin.com/releases/news.aspx?id=1779>. Acessado em 02 Novembro 2015.
- IBGE (2006). O Setor de Tecnologia da Informação e Comunicação no Brasil, 2003-2006. Disponível em: <http://www.ibge.gov.br/home/estatistica/economia/stic/>. Acessado em 30 Outubro 2015.
- Pereira, L. F. D.; Sampaio, F. F.; Oliveira, C. E. T.; Lapolli, F.; Motta, C. L. R. (2010). Ateliê de Objetos de Aprendizagem: Uma Abordagem para o Ensino de Computação em Cursos Técnicos. In: *Revista Brasileira de Informática na Educação*, Volume 18, Número 3.
- MEC. Ensino Médio Inovador. Disponível em: <http://portal.mec.gov.br/component/content/article?id=13439:ensino-medio-inovador>. Acessado em 01 Novembro 2015.

Meninas, Computação e Música

Thayna Gimenez¹, Caroline R. Barroso¹, Jéssika M. Pereira¹,
André L. Przybysz¹, Adriano Rivolli¹, Carlos N. Silla Jr.¹

¹Laboratório de Computação e Tecnologia Musical
Universidade Tecnológica Federal do Paraná (UTFPR)
Cornélio Procópio – PR – Brasil

{gimenezthayna, carol.rezendebarroso, jessikamachado17}@gmail.com

{andreelp, rivolli, carlosjunior}@utfpr.edu.br

Abstract. *This work presents an overview of the project Girls, Music and Computer Science. The goal of this project is to promote Computer Science-related courses to pre-university female students. The project is aimed at public school girls and teaches them Java programming along music theory and practice.*

Resumo. *Este trabalho descreve o projeto Meninas, Computação e Música, que busca incentivar meninas para ingressarem em cursos da área de Computação. O projeto propõe o ensino de programação em linguagem Java aliado ao ensino de teoria e práticas musicais para meninas do ensino médio da escola pública.*

1. Introdução

O famoso ENIAC, primeiro computador do mundo, foi programado por seis jovens mulheres, em 1946. A história ao decorrer dos anos fez questão de ocultar esse fato de tamanha importância para as mulheres e que mostrava seu posicionamento no âmbito computacional [Schwartz et al. 2006]. Hoje, porém; é comum observar que os homens são a grande maioria nas áreas referentes à computação. A mitologia grega diz que Prometeu foi o grande responsável pelo nascimento da tecnologia, mas e se este fato histórico tivesse sido marcado por uma mulher, uma deusa grega? A presença feminina e o olhar para as mulheres da área de computação seriam diferentes? A grande maioria de indivíduos na área de computação ainda seria masculina? [Cabral and Bazzo 2005]

O astrofísico americano Neil DeGrasse Tyson, homem negro que enfrentou preconceitos raciais, disse em uma discussão que antes mesmo de debatermos nossas diferenças genéticas, devemos nos tratar igualmente, gerando oportunidades iguais a todos e somente a partir disso poderemos discutir sobre nossas genéticas [Lima 2014]. Assim como o preconceito racial vivido por Tyson, o mesmo depoimento é válido em relação ao preconceito que as mulheres sofrem na área de computação, por se tratarem da minoria.

Essa presença masculina em massa pode se tornar um problema às mulheres. Os casos de preconceito relatados por mulheres da área de exatas constatarem que são obrigadas a conviver com a impressão de que a mulher não é capaz de exercer um cargo como programadora ou coisa do tipo. A sociedade traçou um estereótipo de que mulheres se enquadram melhor nos cursos de humanas, tais como enfermagem ou pedagogia e que cursos de computação são bastante complexos para o entendimento das mulheres; assim, as jovens que estão finalizando o ensino médio já se imaginam em cursos relacionados a

área de humanas e não a área de exatas [Cabral and Bazzo 2005]. Não há um incentivo que mostra a estas jovens que elas são capazes e que também podem ingressar nessa área.

Com o intuito de dar este incentivo, o projeto Meninas, Computação e Música, mostra a alunas do ensino médio, ensinando conceitos de programação aliados ao ensino de música, que elas também são capazes de entrar em cursos relacionados a computação e que esse não é um privilégio único e exclusivo do público masculino.

2. Meninas, Computação e Música

Visando diminuir os problemas de preconceitos relatados por mulheres e derrubar o paradigma de que cursos de exatas são restritos ao público masculino, o projeto Meninas, Computação e Música busca trazer meninas/mulheres para os cursos de Computação da universidade. Para isso, foram selecionadas quatro meninas do ensino público, com idade entre 16 e 17 anos.

As alunas aprenderam programação em Java aliado ao aprendizado de música, o que tornava o processo mais lúdico e um problema visível do mundo real para ser tratado por elas. O laboratório recebia as meninas duas vezes por semana, sendo quatro horas diárias. Nesses dias, professores das áreas de música e computação, ensinavam conceitos de cada área. No primeiro dia da semana, as meninas tinham aulas sobre música, onde aprendiam o modo de manuseio de cada instrumento, como tocar notas básicas e notas mais complexas, como acordes. No segundo dia, recebiam aulas relacionadas a programação Java, onde aprendiam conceitos que envolviam declaração e tipos de variáveis, estruturas de repetição, estruturas condicionais, como capturar os sons provenientes dos instrumentos e também como trabalhar com o ambiente de desenvolvimento dos jogos, o Greenfoot [Kolling 2009].

Ao decorrer do projeto, também foram ministradas aulas relacionadas a edição de imagens, para que as meninas pudessem criar suas próprias telas para os jogos. As meninas também contavam com o apoio de alunas dos cursos de Análise e Desenvolvimento de Sistemas e Engenharia de Computação para o desenvolvimento dos jogos e também para que as mesmas tivessem um contato mais próximo com alunas da área.

3. Materiais e Métodos

O projeto foi desenvolvido utilizando quatro instrumentos musicais reais, sendo eles: bateria, contrabaixo, guitarra e teclado. Todos os instrumentos possuem entradas para conexão MIDI.

O cabo conector MIDI é responsável por realizar a comunicação entre computador e instrumento musical. Ele transforma o som do instrumento em linguagem de máquina, para que o computador seja capaz de interpretar esta linguagem. Assim, é possível manipular os sons recebidos e os que serão transferidos ao instrumento [Guerin 2005].

Para manipulação dos dados recebidos através do cabo conector MIDI, foi utilizada a biblioteca open source JFugue. A biblioteca recebe através do cabo uma sequência de instruções e os manipula gerando eventos musicais [Koelle 2008]. A partir das sequências recebidas do instrumento, é possível comparar com as instruções exigidas no jogo e conferir o desempenho do usuário [Koelle 2008].

Os jogos foram desenvolvidos na plataforma interativa Greenfoot. Os efeitos de mudanças realizadas no código podem ser observados em tempo real no projeto; aumentando a compreensão do usuário no desenvolvimento [Kolling 2009].

4. Resultados

Com as aulas dadas pelos professores e com a orientação das graduandas, cada aluna desenvolveu um jogo para um determinado instrumento musical.

4.1. Jogo para a Bateria

Neste jogo, o participante deve tocar na bateria real o som que a bateria virtual emitir. Em caso de acerto, ganhará pontos ao término do jogo. O mesmo trabalha a audição do jogador e sua capacidade de saber diferenciar qual parte da bateria está sendo tocada.

4.2. Jogo para o Contrabaixo

Este jogo é voltado para iniciantes que já possuem um conhecimento básico a respeito de música e do instrumento contrabaixo. O usuário deve tocar as mesmas notas que estão dispostas na tela e em seguida são exibidas, dizendo se estão corretas ou incorretas.

4.3. Jogo para a Guitarra

A ideia deste jogo é ter uma aplicação em que o usuário possa aprender novas escalas e assim ficar mais hábil no uso do instrumento. O jogo possui diferentes níveis e ao final é mostrada a porcentagem de erros e acertos do usuário.

4.4. Jogo para o Teclado

O jogo baseia-se em tocar as notas que aparecem dispostas na partitura. É composto por notas simples e por tríades (acordes), com o objetivo de ensinar partituras e acordes no teclado. O jogo também é voltado para treinamento.

5. Feedback das Alunas

As meninas que participaram do projeto se sentiram muito interessadas e motivadas em participar após descobrirem que seria desenvolvido jogos relacionados a determinados instrumentos musicais e que elas mesmas iriam desenvolver estes jogos, aprendendo sobre programação e música.

Foram realizados questionários com as alunas no início e fim do projeto. Para elas, os cursos relacionados a Computação eram vistos como cursos complexos por envolver programação. Ao término do projeto, elas puderam notar que programar não era tão difícil e que bastava dedicação para aprender os conceitos que foram passados. Como as alunas desenvolveram todo o jogo, desde as telas, botões, a programação, a lógica e os conceitos musicais utilizados, as mesmas ficaram muito surpresas com o resultado final do projeto e acharam interessante conseguir interagir com um objeto do mundo real (instrumento) através do programa desenvolvido.

O projeto também auxiliou na escolha do curso para o vestibular, onde algumas alunas consideraram a possibilidade de cursar algo relacionado a Computação.

De forma espontânea, o projeto foi divulgado pelas meninas nas escolas para seus amigos e colegas e acabou despertando interesse em diversos outros alunos, meninas e meninos; surgindo possíveis novas alunas para uma próxima turma e gerando reclamação por parte dos meninos que também gostariam de participar do projeto.

6. Considerações finais

Com o projeto Meninas, Computação e Música, foi possível incentivar as alunas e mostrar que mulheres podem e tem direito igual em ingressar nos cursos de exatas, principalmente na área de computação e que não podem ser motivos de preconceitos.

Além do mais, foi possível mostrar as alunas que o estereótipo gerado pela sociedade em relação às mulheres na área de computação pode ser desmistificado, mostrando que com esforço e empenho, qualquer pessoa é capaz de ingressar e permanecer nesta área, independente de sexo. O ensino da música paralelo ao ensino de programação foi essencial para que as alunas pudessem notar essa quebra de paradigma; pois além de ser uma prática lúdica, também representava um problema do mundo real e que elas são capazes de transformar este problema em linhas de códigos.

O projeto auxiliou as alunas na tomada de decisão em relação ao vestibular, onde cogitaram a possibilidade de ingressar em cursos de computação por terem uma mínima noção sobre programação; uma vez que antes do projeto esta possibilidade não tinha sido considerada.

Agradecimentos

Agradecemos o apoio financeiro recebido do Ministério da Ciência, Tecnologia e Inovação (MCTI), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Secretaria de Políticas para as Mulheres da Presidência da República (SPM-PR), da Petrobras, da Fundação Araucária e da Universidade Tecnológica Federal do Paraná (UTFPR).

Referências

- Cabral, C. G. and Bazzo, W. A. (2005). As mulheres nas escolas de engenharia brasileiras: História, educação e futuro. *Revista de Ensino de Engenharia*, pages 3–9.
- Guerin, R. (2005). *MIDI Power: The Comprehensive Guide*. Cengage Learning PTR, 2nd edition.
- Koelle, D. (2008). The complete guide to jfugue: Programming music in java. Disponível em: <http://www.jfugue.org/4/book.html>.
- Kolling, M. (2009). *Introduction to Programming with Greenfoot: Object-Oriented Programming in Java with Games and Simulations*. Pearson, 1st edition.
- Lima, F. A. (2014). Mulheres na tecnociência: Depoimentos e vivências de mulheres nos cursos de computação da universidade tecnológica federal do paran . Master's thesis, Programa de P s-Gradua o em Tecnologia e Sociedade, Universidade Tecnol gica Federal do Paran , Curitiba, Brazil.
- Schwartz, J., Casagrande, L., Leszczynski, S. A., and Carvalho, M. G. (2006). Mulheres na inform tica: quais foram as pioneiras? *Cadernos Pagu*, 27:255–278.

Meninas.comp: Um Relato da Experiência de Integração entre o Ensino Médio e a Universidade de Brasília

Maristela Holanda¹, Aleteia Patrícia F. de Araújo¹, Maria Emília T. Walter¹, Carlos Alberto Jesus de Oliveira²

¹Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Campus Darcy Ribeiro - CEP 70910-900 – Brasília – DF – Brasil

²Centro de Ensino Médio Paulo Freire
Asa Norte SGAN 610 - CEP 70860-100 – Brasília – DF – Brasil

***Abstract.** This article reports some integration actions among girls at high school and the University of Brasília (UnB), in computer science. These actions were developed as part of the project Meninas.comp, coordinated by female Professors of the Department of Computer Science at UnB. The project Meninas.comp has the main objective of publicizing the computer science area to the girls of the public high schools of Distrito Federal, at Brazil.*

***Resumo.** Este artigo relata atividades de integração entre meninas de escola de Ensino Médio e a Universidade de Brasília (UnB), em computação. As atividades foram desenvolvidas como parte das ações do projeto Meninas.comp, coordenadas por docentes mulheres do Departamento de Ciência da Computação da UnB. O projeto Meninas.comp tem como principal objetivo a divulgação da área de computação para meninas do ensino médio de escolas públicas do Distrito Federal, no Brasil.*

1. Introdução

Nos últimos anos, vem sendo realizadas discussões sobre as razões da incipiente participação feminina na área de Computação, além de estratégias para atrair as meninas para essa área [Cohoon, 2002], [Gurer, 2002], [Maia, 2016], [Khan, 2016]. Computação não parece ser a primeira escolha de formação universitária dessas jovens, que no Ensino Médio devem decidir sobre as suas possibilidades de atuação profissional. Esse fenômeno também ocorre no Distrito Federal, no Brasil. Nesse contexto, no ano de 2010, iniciou-se na Universidade de Brasília (UnB) o projeto “Meninas.comp”, com o objetivo de apresentar a área de Computação para alunas do ensino médio de escolas públicas do Distrito Federal, de modo que elas pensassem nos cursos de Computação como possibilidades entre as suas escolhas profissionais.

O projeto “Meninas.comp” propõe diversas ações, que promovem a discussão das atividades profissionais das mulheres na área de Computação, e vem sendo coordenada por docentes mulheres do Departamento de Ciência da Computação da Universidade de Brasília (UnB). Dentre as atividades do projeto, uma tem se destacado nos últimos dois

anos, que é a solução de desafios na área de computação, de maneira integrada entre alunas da UnB e do Ensino Médio, público. Este artigo tem como objetivo relatar essas experiências.

Este artigo está dividido nas seguintes seções. Na Seção 2, é descrita a metodologia dos desafios, resolvidos de forma integrada entre o Ensino Médio e a Universidade. Na Seção 3, são apresentadas as atividades realizadas em 2015. Na Seção 4, são ressaltados aspectos importantes das atividades do projeto. Por fim, na Seção 5, este artigo é concluído.

2. Metodologia

Esta seção apresenta a metodologia utilizada no projeto Meninas.comp, no que tange ao desenvolvimento de desafios, trabalhados entre alunas do Ensino Médio e da Universidade. A metodologia é composta pelas fases, descritas detalhadamente em seguida (Figura 1): definir a equipe, planejar os desafios, desenvolver os desafios e apresentar os resultados.

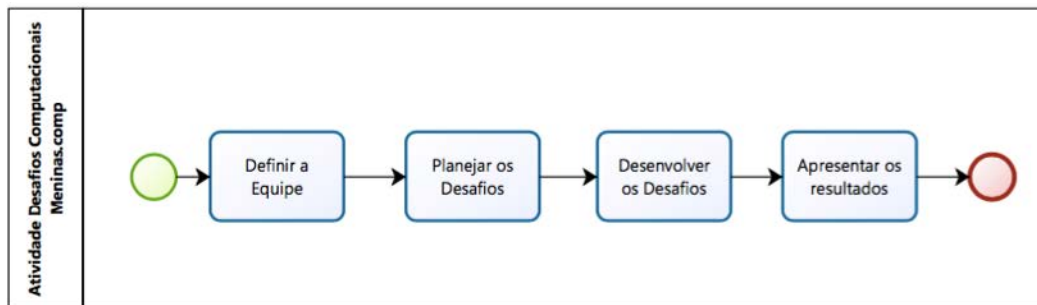


Figura 1. Metodologia da Atividade Desafios Computacionais.

2.1 Definir a Equipe

No início do ano, é definida a equipe composta por alunas de Ensino Médio e da graduação da UnB, que têm interesse em participar do projeto. Para escolher as alunas do Ensino Médio, é fundamental a participação do Professor da Escola de Ensino Médio na qual o projeto será executado. Esse Professor tem como missão divulgar o projeto e convidar as alunas.

Em seguida, após a definição da equipe da Escola, o projeto Meninas.comp é apresentado em uma reunião na própria Escola, que conta com a participação das docentes da Universidade de Brasília, do Professor da Escola e das alunas do Ensino Médio e da Universidade. Esse primeiro contato entre as alunas e os Professores visam integrar todas as pessoas envolvidas no projeto. Um resultado importante dessa reunião é a definição do horário e local das reuniões semanais, que também ocorrem na Escola.

2.2 Planejar os Desafios

Em seguida, são realizadas reuniões de discussões de possíveis problemas (desafios). Nessa etapa, é importante que as alunas do Ensino Médio tragam ideias a serem implementadas na vigência do projeto, por 1 ano. Assim, os desafios nascem das próprias alunas do Ensino Médio, devendo ser as suas soluções implantados em conjunto com as alunas de graduação. Nesse processo, os professores do Ensino Médio e da Universidade atuam como moderadores. Essa etapa pode necessitar de mais de uma reunião, pois além

dos desafios, deve-se definir as necessidades de equipamentos e de conhecimentos, que permitam desenvolver as atividades que solucionarão os desafios.

Um aspecto importante é que o projeto tenha suporte financeiro que permita adquirir os equipamentos. No caso do Meninas.comp, ao longo dos 6 anos de execução do projeto, alguns equipamentos utilizados durante as suas atividades foram comprados, dentre eles: Kit lego educacional, robôs sparks, placas de Arduino, diferentes sensores (temperatura, umidade, presença, dentre outros) e *notebooks*.

2.3 Desenvolver os Desafios

Como as atividades dos desafios são realizadas principalmente na Escola, o papel do Professor da Escola e o apoio do Diretor são fundamentais, pois é necessário um espaço físico onde as atividades possam ser desenvolvidas, todas as semanas.

Nas reuniões semanais, a primeira atividade é uma aula que discute os conceitos básicos necessários para as soluções dos desafios. Essas aulas podem ser ministradas pelos Professores ou por alunas da Universidade. Neste ano de 2016, o projeto tem atividades de programação e conhecimento básico de Arduino.

Em seguida, a implementação dos desafios é desenvolvida. As alunas do Ensino Médio são divididas em grupos, onde cada grupo deve implementar um desafio diferente. Apesar dos desafios serem diferentes, estimula-se uma forte integração entre todas as meninas, de forma que cada aluna possa participar de mais de um desafio.

Durante as soluções dos desafios, acontecem visitas nos laboratórios da UnB, nas quais são discutidos temas relacionados a esses desafios. Esses encontros são importantes, pois as meninas do Ensino Médio podem ter contato direto com alunas e alunos da Universidade que trabalham com temas de pesquisa semelhantes.

2.4 Apresentar os Resultados

Nessa etapa, os projetos são apresentados em feiras de ciência, locais, regionais e nacionais (como na Semana Nacional de Ciência e Tecnologia, promovido anualmente pelo MCTI). As próprias alunas do ensino médio apresentam os desafios e as soluções, implementados durante o ano, para outras alunas e alunos do Ensino Médio.

3. Resultados das Atividades do Ano de 2015

3.1 Definir a equipe

Em 2015, a equipe foi composta por 1 Professor e 6 alunas do Centro de Ensino Médio Paulo Freire (CEM Paulo Freire) do Governo do Distrito Federal, sendo 2 alunas do primeiro ano, 2 alunas do segundo ano, e 2 alunas do terceiro ano, todas bolsistas em programas de Iniciação Científica da UnB. A equipe da UnB foi composta por 3 docentes e 1 aluna do curso de Engenharia de Computação.

3.2 Planejar as Atividades

Após 3 reuniões iniciais de planejamento, foram definidos cinco desafios: construir estação meteorológica com medição de temperatura e umidade; construir carro de controle remoto; realizar corrida com robôs spark; montar cubo mágico com o kit educacional do

Lego; e construir uma casa inteligente composta de sensores controlados por placas Arduino.

As alunas forma divididas em equipes para desenvolver as atividades dos desafios, tal que cada desafio tinha uma líder responsável pelo acompanhamento das atividades. Os Professores supervisionaram continuamente a realização das atividades.

3.3 Desenvolver os Desafios

As atividades foram realizadas principalmente no CEM Paulo Freire, às sextas-feiras à tarde, quando aconteciam as reuniões do projeto. Algumas visitas na UnB foram realizadas. A visita mais relevante para as atividades de 2015 foi feita no Laboratório de Robótica DROID, onde as alunas do ensino médio conheceram alunas do ensino superior que também desenvolviam pesquisas com a placa Arduino e com robôs spark.

3.4 Apresentar os Resultados

Os desafios desenvolvidos podem ser visto na Figura 2. Na Figura 2a, é mostrada a estação meteorológica com placa Arduino e sensores de temperatura e umidade. Na Figura 2b, tem-se a casa inteligente com sensor de presença e sensor de iluminação, dentre outros controlados com Arduino. Na Figura 2c, são apresentados os projetos com o robô lego: montagem do cubo mágico, robôs sparks que andam na mesa sem cair, e o carrinho com controle remoto desenvolvido pelas alunas.



a) Estação Meteorológica.

b) Casa Inteligente.

c) Projeto com robôs.

Figura 2. Projetos Implementados em 2015.

4. Discussão

Após seis anos de realização do projeto Meninas.comp, por docentes mulheres do Departamento de Ciência da Computação da UnB, com diferentes ações em Escolas de Ensino Médio público do Distrito Federal, os seguintes pontos devem ser destacados:

- Ter um professor responsável na escola de ensino médio é fundamental para o sucesso do projeto, pois ele conhece a rotina da escola, os locais disponíveis para reunião, os horários das aulas, e tem acesso direto às alunas e ao Diretor;
- Ter alunas da Universidade envolvidas no projeto, pois a linguagem usada é muito importante para a comunicação. Assim, uma jovem universitária fala “de igual para igual” com as alunas de ensino médio, facilitando a comunicação;

- Trazer as alunas do ensino médio para visitas nos laboratórios da Universidade também foi uma excelente experiência, pois isso diminui a distância ente as Escolas públicas de Ensino Médio e a Universidade, além de motivar as alunas a acreditarem que é possível ingressar em uma Universidade Pública Federal.

5. Conclusões

A área de Computação, nos últimos anos, tem tido a participação de um número pequeno de profissionais mulheres, indicando que as meninas não têm tido interesse em se formar e seguir carreira na área. Neste contexto, o projeto “Meninas.comp” tem como objetivo divulgar a área de Computação para as meninas do ensino médio e público, e vem sendo realizado já há seis anos.

Esse artigo apresentou as ações e resultados do projeto na atividade de desafios computacionais, realizados em 2015. Notamos que duas das alunas integrantes do projeto, que cursavam o terceiro ano em 2015, agora são alunas da Universidade de Brasília, dos cursos de Física e Arquitetura.

Para finalizar, é fundamental o financiamento para o sucesso do projeto. As atividades previstas, decorrentes dos desafios, só puderam ser desenvolvidas em razão da compra dos equipamentos, propiciada por editais de 2013 e 2014 do CNPq, o que também permitiu consolidar a integração entre alunas e Professores do Ensino Médio e da Universidade.

Agradecimentos

Agradecemos a Universidade de Brasília que apoiou o projeto desde a sua criação em 2010, e ao CNPq, que apoiou o projeto em 2013 por meio do Edital Forma Engenharia, e em 2014 com o Edital Mulheres nas Exatas. Por fim, gostaríamos de agradecer especialmente ao Centro de Ensino Médio Paulo Freire do Governo do Distrito Federal que tem apoiado o projeto desde 2013.

Referências

- Cphoon, J. McGrath. (2002) Recruiting and retaining women in undergraduate computing majors (2002). ACM SIGCSE Bulletin - Women and Computing. Vol. 34 (2), pp. 48-52, ACM. USA.
- Gurer, Denise and Camp, Tracy (2002). An ACM-W literature review on women in computing. ACM SIGCSE Bulletin - Women and Computing. Volume 34 Issue 2, June 2002 pp. 121-127.
- Maia, Marcel Maggion. (2016). Limites de gênero e presença feminina nos cursos superiores brasileiros do campo da computação. Cadernos Pagu. n.46, pp. 223-244. ISSN 1809-4449. <http://dx.doi.org/10.1590/18094449201600460223>.
- Khan, Nazish Zaman and Luxton, Andrew (2016). Is computing for social good the solution to closing the gender gap in computer science?. Australasian Computer Science Week Multiconference. DOI: 10.1145/2843043.2843069.

Professoras negras na pós-graduação em Ciência da Computação: uma proposta de pesquisa

Ecivaldo S. Matos¹, Maria José dos Santos², Juliana M. S. Oliveira¹, Ana Carolina C. Ferreira¹, Emilayne F. Corlett¹

¹Departamento de Ciência da Computação – Universidade Federal da Bahia (UFBA)
Caixa Postal 15.064 – 91.501-970 – Salvador – BA – Brazil

²Núcleo de Estudos Afro-Brasileiros
Universidade Federal de Pernambuco (UFPE) – Recife, PE – Brazil

ecivaldo@ufba.br

Abstract. *Demographic data have shown increased participation of black people in higher education. This increase is not equitable to all knowledge areas neither for all genders. However, with regard to graduate courses, there is no conclusive studies on the participation of black women on the faculty of graduate programs in Computer Science. In this sense, this article presents a work-in-progress that aims to identify and analyze, in a pilot study, the presence and role of black women as teachers in graduate programs in Computer Science from Brazil.*

Resumo. *Dados demográficos tem mostrado aumento na participação de negros no ensino superior. Sabe-se que esse aumento não é equitativo para todas as áreas de conhecimento nem para todos os gêneros. Todavia, no tocante à pós-graduação, não há estudos conclusivos sobre a participação de mulheres negras no corpo docente de programas de pós-graduação em Ciência da Computação. Nesse sentido, este artigo apresenta uma proposta de pesquisa em andamento que pretende identificar e analisar, em um estudo piloto, a presença e a atuação de mulheres negras como professoras em programas de pós-graduação em Ciência da Computação do Brasil.*

1. Introdução

Dados apresentados pela Pesquisa Nacional por Amostra de Domicílio (PNAD) revelam forte disparidade entre a população branca e a negra em vários indicadores educacionais. Hoje, temos 47,7% que se declaram brancos, 43,1% pardos, 7,6% pretos, 1,1% amarelos e 0,4% indígenas. Somando-se os quantitativos de pretos e pardos, teríamos 50,7% de negros (IPEA, 2014).

Ainda que tenhamos indicativos de avanços decorrentes da implementação de políticas afirmativas de acesso à educação superior, é fato que a baixa proporção de negros e negras a concluir o ensino médio continua a ser um entrave à equalização do acesso entre negras(os) e brancas(os) (IPEA, 2014). E o que se dizer daqueles que seguem a carreira de pesquisador e professor do magistério superior?

Segundo Lima (2002), "as universidades continuam a reproduzir o androcentrismo e a representar seus espaços como masculinos, principalmente em se

tratando da pesquisa, da direção e da administração institucional [...]" (p. 53). Ainda assim, no imaginário da população temos a pessoa que desenvolve ciência como uma construção simbólica do homem branco e aburguesado. Mas segundo Lima (2002), a real imagem da ciência no Brasil tem mudado nos últimos anos, ainda que o imaginário social demore mais a mudar. Em 2002 já éramos formados por cerca de 44% de mulheres, sendo maioria nas faixas etárias mais jovens.

Todavia, ao analisar esses dados, percebe-se outro aspecto ainda latente, a dicotomização de áreas pelo gênero. Segundo Lima (*ibid*), "a maioria masculina está concentrada nas ciências agrárias, exatas, sociais, engenharia e computação, enquanto a maioria feminina está nas ciências biológicas, da saúde, humanas, letras e artes." (p. 54, grifo nosso). No campo da Ciência da Computação, a experiência cotidiana tem mostrado que apesar de universo de estudantes ter-se modificado ao longo dos anos, a presença está longe de ser equitativa, seja no tocante aos gêneros, seja no tocante às raças/etnias.

Segundo Medeiros (2014), primeira mulher a presidir a Sociedade Brasileira de Computação (SBC), o crescente desinteresse de mulheres pela profissão tem desencadeado diversos programas estratégicos nos países da América do Norte e Europa para atrair mais mulheres aos cursos de Computação. No Brasil, desde 2007, anualmente ocorre o WIT (*Women in Information Technology* - Mulheres em Tecnologia da Informação), um *workshop* para discutir assuntos relacionados a questões de gênero e tecnologia. Além dessa iniciativa, a SBC articula nacionalmente o Programa Meninas Digitais. Iniciativas como essa mostram a preocupação da SBC em promover ações para diminuir as diferenças dentro da área. Mas isso ainda é muito pouco, pois a busca pela equidade passa por questões de gênero, mas também por questões étnico-raciais.

Entendemos que muitos são os desafios para ser docente negra(o) em um departamento e em um programa de pós-graduação e, além disso, exteriorizar preocupações para além dos limites epistemológicos disciplinares da ciência exata, especialmente nas ciências historicamente mais prestigiadas, como as Engenharias, de onde a Ciência da Computação possivelmente herdou o seu pseudo-prestígio.

Dada a pluralidade de potenciais questões de pesquisa, neste estudo nossa questão de pesquisa norteadora será:

- Qual é o quadro atual da presença e da atuação de mulheres negras no corpo docente da pós-graduação em Ciência da Computação no Brasil?

Nesse sentido, está em curso um *estudo piloto* para identificar e analisar a presença e atuação de mulheres negras como professoras nos programas de pós-graduação em Ciência da Computação.

2. Fundamentação teórica

Desde os primórdios da colonização, o acesso à educação se estruturou de forma diferenciada em relação aos diversos grupos sociais. Nos estudos de História da Educação, predomina a ideia de que a população negra escravizada ou livre não teve acesso ao sistema educacional. Estudos recentes, no entanto, como o de Fonseca (2009)

tem demonstrado que, em determinados locais, populações negras e pobres também estiveram na escola elementar do século XIX.

Nesse contexto, podemos identificar o surgimento de mobilizações da população negra e a criação de organizações que começam a pautar em suas lutas questões relacionadas à educação, como enfatizam Gomes e Silva (2000).

O atual contexto histórico evidencia que movimentos negros em toda parte do Brasil incorporam as lutas em prol da presença negra nas instituições de ensino, de modo equitativo, e no combate à discriminação. A história das mulheres negras brasileiras, nesse contexto, está atrelada a um passado repleto de obstáculos na trajetória de formação deste país, enfrentados pela população negra, com consequências que perduram até os dias atuais.

São muitos os desafios para que as mulheres negras possam superar tais obstáculos, porém no decorrer de sua história, elas aprenderam a utilizar suas experiências de vidas e transformá-las em armas, com as quais foram enfrentando os diversos obstáculos e, em sua maioria, superados, especialmente no processo de construção da cidadania de seu povo.

Inúmeras pesquisas realizadas nos últimos anos mostram que a mulher negra apresenta menor nível de escolaridade, recebe remuneração menor que o homem, e poucas conseguem romper as barreiras do preconceito e da discriminação racial, ascendendo socialmente (Santos, 2014). Segundo Santos, as mulheres negras são as que até hoje carregam o mito da mulher sexuada, visão concebida em uma sociedade patriarcal onde sempre se determinou o poder dos homens sobre a mulher.

Essas mulheres estão em um campo de disputa e necessitam comprovar sua competência profissional, lidando com o preconceito e a discriminação racial que lhes impõem maiores desafios para a conquista do ideal sonhado e desejado. A questão de gênero é um complicador, mas quando somada à de raça, se traduz nas maiores dificuldades. A discriminação racial na vida das mulheres negras é constante. Apesar disso, muitas constituíram estratégias próprias para superar as dificuldades decorrentes dessa problemática. (SANTOS, 2014, p. 58)

As lutas da população negra, especialmente das mulheres negras, para conquistarem melhores cargos no mercado de trabalho exigem uma força maior do que outros setores da sociedade. Acreditamos na possibilidade de a população negra se identificar e conviver em uma realidade menos repressora, especialmente para as mulheres negras.

3. Metodologia

Dada a natureza do nosso objeto de investigação, essencialmente descritivo-analítico, e dos nossos objetivos de pesquisa, utilizaremos duas técnicas fundamentais para coleta de dados (em duas etapas): (i) *survey* e (ii) entrevista.

Na primeira etapa, aplicar-se-á *survey* elaborado no formato eletrônico e enviado aos coordenadores dos programas de pós-graduação em Ciência da Computação, após pré-testes e contato prévio com esses coordenadores para apresentação da pesquisa. Para

análise, os dados serão tabulados e sumarizados por meio de descrições, tabelas e gráficos.

Na segunda etapa, será realizada entrevista apenas com alguns professores pré-selecionados e contatados de acordo com um conjunto de critérios a serem definidos. O objetivo será aprofundar a percepção sobre a atuação de alguns desses professores negros/as na carreira acadêmica em Ciência da Computação, na busca pela identificação de elementos não levantados no *survey* e/ou esclarecimento de pontos que mereçam mais detalhamento. Dado o tempo previsto para a pesquisa, consideraremos a priori o máximo de três entrevistados. Para análise consideraremos a perspectiva da *Grounded Theory* para composição dos códigos e categorias de análise.

Cuidados éticos

Para participar da pesquisa, os sujeitos deverão ler e aceitar o(s) Termo(s) de Consentimento Livre e Esclarecido (TCLE). Esse documento possuirá explicações acerca da natureza da pesquisa, local de realização da pesquisa, informações sobre os pesquisadores e financiadores, objetivos, contatos, sigilo da identidade e esclarecimentos sobre a voluntariedade da participação.

4. Agradecimentos

Agradecemos à *Fundação Carlos Chagas* pelo provimento de recursos para financiamento desta investigação.

Referências

- Fonseca, M.V. (2009) População negra e educação: o perfil racial das escolas mineiras no século XIX. Belo Horizonte: Mazza Edições.
- Gomes, L.A.O.; Silva, B.P.G. (2000) Movimento Negro e Educação. *Revista Brasileira de Educação*, Rio de Janeiro, v. 15, p.134 - 156, set./out./nov./dez.
- IPEA (2014) (Instituto de Pesquisa Econômica e Aplicada). Políticas Sociais: acompanhamento e análise. Anexo Estatístico. Brasília, n. 22. 40 p. Disponível em: <http://www.ipea.gov.br/portal/images/stories/PDFs/politicas_sociais/140930_bps22_anexo.pdf>. Acesso em: 29 mar. 2016.
- Lima, N.R.L.B. (2002) As mulheres nas ciências: o desafio de uma passagem... a passagem do privado para o público. In: Costa, A.A.A.; Sardenberg, C.M.B. (Org.). *Feminismo, Ciência e Tecnologia*. Salvador: REDOR/NEIM-FFCH/UFBA. p. 51-65.
- Medeiros, C.B. (2014) Grandes desafios e mulheres na Computação. In: Bigonha, R.S. et alii. (Org.) *Memórias da Sociedade Brasileira de Computação*. Porto Alegre: SBC. p. 125-137.
- Santos, M.J. (2014) *Trajetória Educacional de Mulheres Quilombolas no Quilombo das Onze Negras dos Cabo de Santo Agostinho - PE*. Recife: Ed.Universitária da UFPE.

Projeto SciTechGirls: desenvolvimento de aplicativos e participação em competições de programação científicas e tecnológicas

Rosiane de Freitas, Ludymila Lobo, Tayana Conte

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM)
Manaus – AM – Brasil

{rosiane, llag, tayana}@icomp.ufam.edu.br

Resumo. *Este artigo apresenta o projeto SciTechGirls, do Instituto de Computação da Universidade Federal do Amazonas (IComp/UFAM), com ações voltadas para o envolvimento de alunas dos cursos de graduação em Computação, no desenvolvimento de aplicativos computacionais e na participação em competições de programação científicas e tecnológicas, como forma de incentivá-las a aprofundar suas habilidades de programação e, assim, também desmistificar o falso mantra dito por muitos da área de que meninas não programam ou não gostam de programar. Neste trabalho são apresentados casos de sucesso no envolvimento de alunas universitárias em competições de programação, bem como apresentados alguns dos aplicativos de temática feminina desenvolvidos.*

Abstract. *This article presents the SciTechGirls project, from the Institute of Computing of the Federal University of Amazonas (IComp/UFAM), with actions aimed at the involvement of women students of undergraduate computing courses, in the development of computer applications and in the participation in scientific and technological programming contests, as a way to encourage them to further develop programming skills and thus also demystify the false mantra spoken by many that girls can not code or do not like to code. This work presents success stories in involving women university students in programming contests and present some of the developed applications feminine themes.*

1. Introdução

O projeto *SciTechGirls* [de Freitas et al. 2016] faz parte do programa Cunhantã Digital [Lauschner et al. 2016], que por sua vez se vincula ao programa nacional "Meninas Digitais" da Sociedade Brasileira de Computação - SBC [Maciel and Bim 2016]. Sua origem se deu através de professoras e professores do Instituto de Computação da Universidade Federal do Amazonas (IComp/UFAM), que divulgavam e promoviam competições de programação científicas e tecnológicas. Em contrapartida, observou-se que além do número de mulheres ingressantes nos cursos de Computação ser proporcionalmente muito menor do que o de homens [IPEA 2011], era muito frequente a desistência de mulheres nos primeiros períodos dos cursos, principalmente por se sentirem desmotivadas e sem identificação com as atividades propostas.

Foi neste cenário que, em 2013, surgiu uma oportunidade apresentada pela *Microsoft Research* [MSR 2016], de professores e alunas do IComp/UFAM participarem de uma competição internacional pioneira de programação tecnológica [MSR 2014], voltada para alunas universitárias desenvolverem aplicativos computacionais: a "*International Women's Hackathon*". Paralelamente a isto aconteciam as Olimpíadas Brasileiras de Computação e a Maratona de Programação [MP 2016], parte do *International Contest Programming Contest* da *Association of Computing Machinery* (The ICPC-ACM) [ICPC-ACM 2016], ambas promovidas pela SBC no Brasil, e que estimulam alunos e alunas na resolução de problemas por computador envolvendo a aplicação de conteúdos teóricos, técnicas de projeto de algoritmos e programação avançada. Houve, então, um esforço para envolver e viabilizar a participação das alunas em tais competições, sendo oficializado um projeto em torno do tema.

Após a implantação do projeto, um número maior de alunas passaram a participar de competições como a maratonas de programação, olimpíada de robótica e *hackathons* (competições tecnológicas de programação), incluindo todas as três edições do *International Women's Hackathon*. Posteriormente, em 2015, houve a criação do programa Cunchantã Digital e a ampliação deste movimento em prol da maior participação de mulheres da região amazônica em cursos e profissões de Computação e demais áreas de ciências e tecnologia.

Neste contexto, a existência de histórias de sucesso é importante para manter outras meninas interessadas e motivadas. Pôde-se observar que as conquistas obtidas por algumas alunas fizeram com que outras se interessassem e passassem a acreditar em seus próprios potenciais e habilidades, daí o interesse em se criar um projeto para dar uma maior ênfase no incentivo à participação de meninas estudantes em competições de programação.

O restante deste artigo está estruturado como segue. Na Seção 2 são abordadas algumas competições das quais as alunas integrantes do Projeto *SciTechGirls* participaram. Na Seção 3, aborda-se os principais aplicativos desenvolvidos. E, na Seção 4 são feitas as considerações finais, onde se resume os resultados e repercussão obtidos pelo projeto *SciTechGirls*, bem como as próximas ações a serem executadas.

2. Participação em competições de programação científicas e tecnológicas

Processos competitivos, desde que adequadamente conduzidos, proporcionam ao estudante um aprimoramento das habilidades e dos conhecimentos, além de ser motivacional, fazendo o aluno sentir-se capaz de realizações. Estes fatores são importantes na inclusão de grupos minoritários, como é o caso das mulheres nas áreas tecnológicas e científicas, sendo internacionalmente cunhado o termo STEM (*Science, Technology, Engineering and Mathematics*), para expressar os cursos e carreiras profissionais relacionadas à ciência e tecnologia (que incluem às Engenharias, Computação, áreas de exatas e carreiras tecnológicas em geral).

No caso do projeto *SciTechGirls*, definiu-se como ponto de abordagem principal, se trabalhar a habilidade na resolução de problemas por computador através da participação em competições de programação científicas e tecnológicas, bem como através do desenvolvimento de aplicativos móveis de temática feminina.

Dentre as competições que as alunas participaram, podem ser destacadas as seguintes:

- *hackathons* (competições tecnológicas de programação) promovidas pela *Microsoft Research* (2013 a 2014), uma delas parte da competição *Imagine Cup*, envolvendo equipes de alunas para o desenvolvimento de aplicativos móveis.
- Maratonas de programação (The ICPC-ACM), regionais e final brasileira, sendo que no ano de 2015, uma equipe foi classificada para a final brasileira, sendo a melhor equipe formada somente por mulheres dentre as 8 existentes.
- Olimpíada Brasileira de Informática - OBI (modalidade universitária), 2013-2015, 1a fase e finais brasileira.
- Competições de robótica.
- *hackathons* regionais (competições tecnológicas) envolvendo o desenvolvimento de aplicativos.

Com participações individuais, outras competições foram: *Technovation Challenge* - melhor projeto eleito; 2o lugar na *Startup Weekend* em Providence (RI) USA; prêmio Empreendedorismo Sustentável: projeto eleito pelo Santander como um dos 32 melhores projetos brasileiros; 2º lugar na 1a *Hackathon* do Instituto Nokia de Tecnologia.

3. Aplicativos com temática feminina desenvolvidos

A participação na competição de programação tecnológica promovida pela *Microsoft Research* (EUA) e várias entidades internacionais de representação feminina, a *International Women's Hackathon*, possibilitou o desenvolvimento de alguns aplicativos, em sua maioria de temática do universo feminino. Ao todo, foram desenvolvidos quatro aplicativos: *MakeUPhi*; *Mommy's Benefit*; *What The Hack*; e, *How to Help*. Tais aplicativos são brevemente descritos a seguir.

3.1. *Make UPhi* app

O aplicativo para dispositivos móveis *MakeUPhi* permite, com base na foto tirada da câmera de um aparelho celular, se conseguir dicas de maquiagem geradas através da análise das proporções do rosto de uma mulher, com base no princípio matemático da razão áurea (*golden ratio*). A razão áurea ou divina proporção é uma propriedade matemática representada por uma constante irracional denotada pela letra grega *Phi*, encontrada entre os números da sequência de Fibonacci. Usando a proporção áurea, foi criada uma máscara que representa o rosto mais simétrico possível, ou seja, mais agradável aos olhos humano. Através deste cálculo, o aplicativo fornece dicas de beleza e maquiagem, de acordo com o rosto da usuária de tal forma a deixá-lo mais próximo das proporções áureas (mas, também dá a liberdade para que se brinque com este conceito e se obtenha proporções ainda mais assimétricas e também interessantes).

3.2. *Mommy's BeneFIT* app

O aplicativo *Mommy's BeneFIT* visa auxiliar mulheres grávidas na prática de atividades físicas que contribuem na manutenção da saúde antes, durante e após o período de gestação, através de uma base de informações e rede de contatos, com dicas de lugares para praticar o exercício escolhido, página para troca de conhecimentos entre as grávidas, possibilidade de realizar *check-in* dos lugares em que praticou esportes, e "*gamification*" através de selos (*badges*) que a usuária conquista à medida que realiza *check-in* nos lugares para prática de atividades, e uma agenda para organizar dias e horários dos exercícios realizados.

3.3. *What the Hack app*

What the Hack é um aplicativo móvel "gamificado", voltado para as meninas que desejam obter mais informações sobre carreiras tecnológicas e científicas, e o papel das mulheres nestas áreas. É uma ferramenta para se obter informações sobre eventos tecnológicos, facilitando o compartilhamento de conhecimento e incentivo à participação em tais eventos. O aplicativo possibilita a interação e troca de experiências entre mulheres jovens e profissionais, auxiliando na construção de uma base de conhecimento, geração de cultura e ampliação da rede de contatos (*networking*) das usuárias. O ponto-chave do aplicativo é o processo de "*gamification*" para tornar seu uso mais divertido, excitante e desafiador. É, portanto, um aplicativo baseado em jogo (*serious game*) com a proposta de informar e envolver os usuários em eventos de tecnologia.

3.4. *Outros - How to Help app*

Este foi o único aplicativo desenvolvido sem a temática feminina, mas, assim como os outros, foi desenvolvido por uma equipe de cinco meninas para uma *hackathon* feminina. O aplicativo trata-se também de um *serious game* (aplicativo com propósitos sérios, como o de instruir ou auxiliar grandes grupos em situações complexas), composto por três funcionalidades: guia explicando o que fazer em casos de catástrofes naturais (como enchentes, desmoronamento de terras, furacões, terremotos); jogo de perguntas e respostas (*quiz*) para testar os conhecimentos sobre as melhores medidas a serem adotadas nestes casos; e, botão de emergência que permite o envio automático de mensagens SMS com a localização do usuário para os contatos de emergência cadastrados previamente.

4. Considerações finais

As ações do projeto *SciTechGirls*, através da participação em competições de programação e desenvolvimento de aplicativos, possibilitaram um maior envolvimento das alunas dos cursos de graduação do IComp/UFAM em atividades das disciplinas que envolviam programação, bem como em atividades extras, além das competições, envolvendo participação em eventos e palestras sobre o tema, possibilitando uma ampliação dos seus conhecimentos e redes de contatos e interessados: em suma, incentivou o empoderamento feminino nos cursos de Computação e afins da UFAM e região. Como preparação, as equipes se reuniam, recebiam treinamento e trocavam conhecimentos constantemente. Em todos os casos, se destacaram nas competições e foram alvo de matérias no site internacional da *Microsoft Research* [MSR 2014]. Os aplicativos e as meninas participantes também ganharam destaque nas mídias sociais e jornais regionais e nacionais, e notou-se que o fato de possuir apenas mulheres envolvidas no projeto era o maior motivo de entusiasmo da mídia [Camelo 2016].

Como repercussão do projeto *SciTechGirls*, foram recebidos convites para participação em eventos, reportagens, entrevistas e apresentação trabalhos, como os seguintes:

- Aceite e participação no *First MobiCom Mobile App Competition* do MOBICOM 2013 (com o aplicativo passando para a 3a etapa).
- Aceite e apresentação no Workshop sobre Ensino de Interação Humano-Computador (WEIHC) 2013 [Ferreira et al. 2013].
- Apresentação no Workshop Cunhantã Digital, evento realizado como parte do Simpósio Brasileiro de Qualidade de Software (SBQS) 2015.

- Aceite e apresentação no Workshop de Ferramentas e Aplicações (WFA) do Web-Media 2015.

Como perspectivas futuras, pretende-se estender o escopo do projeto *SciTechGirls*, em conjunto com outros projetos e ações do programa Cunhantã Digital, para meninas do ensino fundamental e médio, e assim ajudá-las no momento de escolha da área profissional abstraindo qualquer pré-conceito da sociedade relacionado ao gênero. Além disso, almeja-se que tais ações sejam uma referência na região Norte do Brasil, para que outras mulheres se encorajam e se estimulem a seguir carreiras na áreas de ciência e tecnologia.

Agradecimentos

Nossos agradecimentos à Juliana Salles e Rane Johnson (Microsoft Research).

Referências

- Camelo, L. (2016). Alunas da ufam desenvolvem aplicativos de celular voltados ao universo feminino. *Jornal A Crítica*. Disponível em <http://acritica.uol.com.br/vida/Mulheres-tecnologia-aplicativos-UFAM-evento-Manaus-Ama>
- de Freitas, R., Conte, T., Lobo, L., and Salles, J. (2016). Projeto scitech-girls. Instituto de Computação, Universidade Federal do Amazonas. Disponível em <http://scitechgirls.icomp.ufam.edu.br/>.
- Ferreira, B., Costa, I., Dantas, A., Gomes, L., Aoki, M., Conte, T., and Rodrigues, R. (2013). Teoria à prática: Aplicando técnicas de ihc em um aplicativo móvel voltado para grávidas. In *WEIHC13*.
- ICPC-ACM (2016). International collegiate programming contest. Association for Computing Machinery (ACM). Disponível em <https://icpc.baylor.edu/>.
- IPEA (2011). Retrato das desigualdades de gênero e raça. Instituto de Pesquisa Econômica Aplicada. 4a. edição.
- Lauschner, T., Nakamura, F., de Freitas, R., Conte, T., Lobo, L., and Cruz, M. (2016). Programa cunhantã digital. Instituto de Computação, Universidade Federal do Amazonas. Disponível em <https://www.facebook.com/cunhantadigital/?fref=ts>.
- Maciel, C. and Bim, S. A. (2016). Programa meninas digitais - ações para divulgar a computação para meninas do ensino médio. In *Anais do Computer on the Beach*. Sociedade Brasileira de Computação. Disponível em <http://www.computeronthebeach.com.br/arquivos-2016/Anais completos - Computer on the Beach 2016.pdf>.
- MP (2016). Maratona de programação. The ICPC-ACM South America - Brazil. Sociedade Brasileira de Computação. Disponível em <http://maratona.ime.usp.br/>.
- MSR (2014). Inspiring female hackers in brazil. Microsoft Research Blog. Disponível em https://blogs.msdn.microsoft.com/msr_er/2014/04/18/inspiring-female-hackers-in-brazil/.
- MSR (2016). Microsoft research. Disponível em <http://research.microsoft.com/>.

Reflexões sobre Gêneros, Tecnologia e Processo Vocacional

Karen da Silva Figueiredo, Cristiano Maciel

Programa de Pós-Graduação em Educação, Instituto de Educação, Universidade Federal do Mato Grosso (UFMT), Cuiabá, Mato Grosso, Brasil

karen@ic.ufmt.br, cmaciel@ufmt.br

Abstract. *This paper aims to analyze theoretical issues related to gender, technology and vocational process in order to discuss how gender could influence the decision-making of the female high school students' vocational process in choosing CTE courses.*

Resumo. *Este trabalho analisa aportes teóricos relacionados às relações de gênero, tecnologia e processo vocacional a fim de discutir como o gênero poderia influenciar nas tomadas de decisão do processo vocacional de jovens alunas de Ensino Médio para ingresso em cursos de superiores de CET.*

1. Introdução

A escolha de uma profissão é uma das primeiras grandes responsabilidades exigidas aos jovens, um momento de decisão que geralmente é engatilhado no período do Ensino Médio (EM) [McWhirter *et al.* 2000]. Considerando que a escolha profissional impacta não somente a dimensão da carreira, mas também outras dimensões pessoais como família, lazer, saúde etc., é de fundamental importância que esta escolha seja feita pelos jovens com consciência dos seus interesses e aptidões e total liberdade para priorizarem os seus valores pessoais. Segundo Saavedra (2009), os gêneros destacam-se como barreiras que mais limitam a liberdade de escolha dos e das jovens, influenciando a forma como concretizam as suas decisões ao longo do ciclo de vida. Taveira (2000) distingue os gêneros como variável capaz de influenciar na exploração e indecisão vocacional, interferindo na concretização de projetos vocacionais.

Dados quantitativos [Portal Brasil 2015; Saavedra 2009] apontam que a presença de mulheres em cursos de Ensino Superior das áreas de Computação, Engenharias e Tecnologias (CET) é muito baixa quando comparada a outras áreas. Entretanto, pesquisas mostram que o número de jovens de ambos os sexos que se interessam por estas áreas de conhecimento é praticamente idêntico até os treze anos de idade [Saavedra *et al.* 2010]. A partir dos treze anos, é possível verificar o decréscimo no número de jovens do sexo feminino que escolhem estes domínios tipicamente masculinos, em todos os níveis de ensino subsequentes [Xu 2008].

Para Jacobs (1996), as teorias educacionais e pesquisas estão focadas em disparidades de classes e quando as relações de gênero são discutidas, recebem pouca atenção. Pocinho *et al.* (2010) acrescentam o fato de que: no que diz respeito a estudos sobre indecisão de carreira, é ainda escasso o registro de diferenças de gêneros. Assim, o objetivo deste trabalho é analisar aportes teóricos relacionados às relações de gênero, tecnologia e processo vocacional. Tal atividade faz parte de uma pesquisa maior que intenta averiguar como o gênero poderia influenciar nas tomadas de decisão do processo

vocacional de alunas de EM, para ingresso em cursos de Ensino Superior em CET.

2. Gêneros e Tecnologia

Historicamente, os campos da ciência e tecnologia foram percebidos como espaços de atuação masculina, sendo também associados às características comumente atribuídas aos homens, tais como: neutralidade, objetividade, imparcialidade e universalidade [Muzi e Da Luz 2012]. Graña (2004) aponta a associação do homem com o domínio da ciência, que por sua vez domina a natureza, considerada passiva e associada ao feminino. A existência dessa associação que reforça modelos patriarcais faz parecer que os homens estiveram e sempre estarão no domínio de todas as ordens do saber. Por conseguinte, a aproximação de mulheres nestes campos nem sempre foi estimulada. As mulheres começaram a avançar no domínio da ciência e tecnologia a partir de questionamentos sobre o determinismo biológico que as mantinha majoritariamente em atividades consideradas como mais adequadas ao seu sexo [Muzi e Da Luz 2012].

Foi a conhecida segunda onda do feminismo, nas décadas de 60 e 70, que detectou o apagamento e a minimização das mulheres na história da ciência e tecnologia [Brunet e Natansohn 2012]. Segundo Wacjman (2006) podem ser analisadas três posições teóricas em relação às feministas dessa época e a tecnologia: o feminismo liberal, o feminismo radical e o feminismo socialista.

Descrevendo resumidamente conforme Brunet e Natansohn (2012), o feminismo liberal considera que todo o problema da separação histórica das mulheres e a tecnologia se reduz a uma questão de acesso e oportunidades. A tecnologia seria neutra e os problemas poderiam ser resolvidos com políticas de igualdade de oportunidades. O feminismo radical considera que as tecnologias reprodutivas são instrumentos de eugenia, controle racial e patriarcal, colocando o problema da opressão das mulheres nas suas capacidades reprodutivas. Já o feminismo socialista reinsere a crítica de Marx às tecnologias e à divisão de trabalho, incorporando discussões de gêneros às discussões de classe e trabalho doméstico. A tecnologia seria considerada uma extensão da dominação patriarcal e capitalista e as mulheres seriam suas vítimas.

Atualmente, na terceira onda, conta-se com as discussões do ciberfeminismo, surgido na década de 90 com trabalhos conhecidos de Plant, Haraway e Turkle [Brunet e Natansohn 2012]. O ciberfeminismo traz uma visão otimista da tecnologia, em um discurso mais atrativo para as mulheres jovens que cresceram imersas na cultura digital. O ciberfeminismo deixa a “tecnofobia” para trás e enxerga a tecnologia como aliada para empoderamento feminino, mudanças políticas e organização e articulação de ideias e saberes. Com estes princípios, emergem uma série de políticas e iniciativas, como as já citadas na seção 2, unindo a visão ciberfeminista do empoderamento tecnológico feminino com um pouco do viés de criação de oportunidades do feminismo liberal.

Hoje, as mulheres já são maioria como usuárias da internet, maioria nas redes sociais e maioria entre os jogadores de jogos eletrônicos. O desafio maior está em transformar estas mulheres de simples consumidoras de tecnologia em produtoras de tecnologia, participando ativamente do projeto e desenvolvimento dos produtos.

3. Processo Vocacional

Uma das principais características da adolescência é o desenvolvimento de vários

aspectos da identidade e a formação de planos de carreira desempenha um importante papel no processo de formação dessa identidade, a identidade vocacional [Super, 1963]. A identidade vocacional deveria refletir um padrão de interesses e objetivos do jovem. Entretanto, a teoria de aprendizagem de Bandura (1986) e a Teoria Social Cognitiva do Desenvolvimento de Carreira de Lent *et al.* (2002) chamaram atenção para a influência de fatores cognitivos que antecedem o desenvolvimento de interesses e objetivos e para a avaliação da autoeficácia.

As pessoas tendem a acreditar que possuem capacidade de desempenhar determinados comportamentos e este fato determina o quão bem elas vão se engajar nestes comportamentos. O contrário também é verdade. Para os teóricos sociais cognitivos de carreira, autoeficácia é um fator crucial que media os interesses, objetivos e desenvolvimento de ações em um domínio em particular [Gushue *et al.* 2006].

Apesar da escolha de uma profissão já não ser necessariamente uma decisão para a vida toda, é um ponto colocado para os jovens como uma grande responsabilidade, culminando usualmente no período do EM sobre qual caminho seguir: procurar determinado emprego, fazer um curso técnico ou de nível superior em determinada área do conhecimento. Como o processo vocacional e a tomada de decisão é importante nos anos do EM, os jovens examinam a carreira sob uma perspectiva de avaliação da sua autoeficácia [Taylor e Betz 1983].

Segundo Saavedra (2009), a autoeficácia pode afetar os interesses vocacionais das jovens. Para a autora, as adolescentes evitam estas áreas por uma série de fatores: dificuldade de realização pessoal com a matemática, pouca exposição a modelos de mulheres em domínios não tracionais, maiores níveis de ansiedade nas tarefas associadas ao masculino e menor encorajamento de prosseguirem em uma carreira profissional. Pode concluir-se que, de uma forma geral, “meninas e mulheres possuem expectativas de autoeficácia a carreiras mais baixas e menos generalizadas do que os homens” [Saavedra 2009], explicando a subexploração das suas capacidades em áreas não popularmente consideradas femininas e a limitação de suas escolhas profissionais.

4. Conclusões

Diante do exposto, os autores do presente trabalho acreditam que como parte da sociedade, cabe às universidades fomentar pesquisas sobre as questões relativas aos gêneros na educação e no mercado, fornecendo informações e ajudando a contribuir com soluções para os problemas levantados. Paridade de acesso das jovens nos cursos de CET não é uma questão somente de equidade de gêneros. Muzi e Da Luz (2012) destacam que a baixa participação de mulheres em cursos superiores de CET não é prejudicial só para as mulheres, mas sobretudo, para essas áreas de conhecimento e principalmente para a sociedade que não pode contar com a contribuição de metade da população na busca do desenvolvimento científico, tecnológico, econômico e principalmente social do país.

Desta forma, investir em pesquisas que investiguem a fundo como o gênero interfere na escolha vocacional de cursos de CET das jovens durante o EM pode contribuir para a sociedade em diversos níveis de desenvolvimento. Como passos futuros, os autores pretendem realizar uma pesquisa qualitativa exploratória acompanhando o percurso de formação do processo vocacional de alunas de EM.

Referências

- Bandura, A. (2002) “Social foundations of thought and action”, In: The health psychology reader, p. 94-106.
- Brunet, K. S. e Natansohn, G. (2012) “Labdebug: práticas de cultura digital livre para mulheres”, In: Gênero e Tecnologias. EdUFMT, p. 101-116.
- Graña, F. (2004) “Ciencia y tecnologia desde uma perspectiva de género”, Montevideu: Facultad de Humanidades y Ciencias de la Educación.
- Gushue, G. V. et al. (2006) “The relationship of career decision-making self-efficacy, vocational identity, and career exploration behavior in African American high school students”, In: Journal of Career Development, v. 33, n. 1, p. 19-28.
- Jacobs, J. A. (1996) “Gender inequality and higher education”, In: Annual Review of Sociology, p. 153-185.
- Lent, R. W. et al. (2002) “Social cognitive career theory”, In: Career choice and development, v. 4, p. 255-311.
- McWhirter, E. H. et al. (2000) “The effects of high school career education on social-cognitive variables”, In: Journal of Counseling Psychology, v. 47, n. 3, p. 330.
- Muzi, J. L. C. e Da Luz, N. S. (2012) “Relações de gênero na UFTPR: participação das mulheres na produção científica e tecnológica”, In: Gênero e Tecnologias. Tecnologias do Gênero, EdUFMT, p. 69-84.
- Pocinho, M. D. (2010) “Influência do gênero, da família e dos serviços de psicologia e orientação na tomada de decisão de carreira”, In: Revista Brasileira de Orientação Profissional, v. 11, n. 2, p. 201-212.
- Portal Brasil. (2015) “Mulheres são maioria no ingresso e na conclusão de cursos superiores. Educação”, Disponível em: <<http://www.brasil.gov.br/educacao/2015/03/mulheres-sao-maioria-no-ingresso-e-na-conclusao-de-cursos-superiores>>.
- Saavedra, L. (2009) “Assimetrias de Gênero nas Escolhas Vocacionais” In: Guião de educação - Gênero e cidadania, Lisboa, CIG.
- Saavedra, L. et al. (2010) “A subrepresentatividade das mulheres em áreas tipicamente masculinas: Factores explicativos e pistas para a intervenção”, In: Revista Brasileira de Orientação Profissional, v. 11, n. 1, p. 49-59.
- Super, D. E. (1963) “Vocational development in adolescents and early adulthood: Tasks and behaviors”, In: Career development: Self concept theory, Research Monograph.
- Taveira, M do C. (2000) “Exploração e desenvolvimento vocacional de jovens”, Centro de Estudos em Educação e Psicologia, Universidade do Minho, 2000.
- Taylor, K. M. e Betz, N. E. (1983) “Applications of self-efficacy theory to the understanding and treatment of career indecision”, In: Journal of Vocational Behavior, v. 22, n. 1, p. 63-81.
- Wajcman, J. (2006) “El tecnofeminismo”, Universitat de València, Catedra, p. 200.
- Xu, Y. J. (2008) “Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions”, In: Research in Higher Education, v. 49, n. 7, p. 607-624.

Robótica na aprendizagem de física e matemática para alunas do ensino médio

Maitê Thomazi Manenti¹, Éverton Fabian Jasinski², Tatiana Nilson dos Santos³,
Eliane Pozzebon³, Luciana Bolan Frigo³

¹Laboratório de Tecnologias Computacionais – Universidade Federal de Santa Catarina (UFSC) – Araranguá – SC – Brasil

²Laboratório de Automação e Robótica Móvel – Universidade Federal de Santa Catarina (UFSC) – Araranguá – SC – Brasil

³Programa de Pós-Graduação em Tecnologias da Informação e Comunicação – Universidade Federal de Santa Catarina (UFSC) – Araranguá – SC - Brasil

mayte_manenti@hotmail.com, everton.fabian@ufsc.br,
tatiana.santos@posgrad.ufsc.br, {elianepozzebon,
luciana.frigo}@ufsc.br

Abstract. *This article presents the results of a project where robots were applied LEGO kit as a teaching tool in teaching physics and mathematics to support learning in the average level disciplines of students of High School Macário Borba located in Sombrio-SC. This course lasted 3 (three) months and the students solved practical exercises on kinematics, algebraic equations and energy conservation using robots projects that facilitate understanding. As a result realized the great motivation of students to learn in a more didactic and motivating way, considered complicated content.*

Resumo. *Este artigo apresenta os resultados de um projeto onde foram aplicados robôs do kit LEGO como ferramenta didática no ensino de física e matemática para auxiliar a aprendizagem nas disciplinas do nível médio das estudantes da Escola de Ensino Médio Macário Borba localizada em Sombrio-SC. Este curso teve duração de 3 (três) meses e as estudantes resolveram exercícios práticos sobre cinemática, equações algébricas e conservação de energia utilizando projetos de robôs que facilitassem o entendimento. Como resultado percebeu-se a grande motivação das estudantes ao aprenderem de uma maneira mais didática e motivadora, conteúdos considerados complicados.*

1. Introdução

Diversas empresas têm demonstrado grande interesse no aumento de contratação de mulheres no seu quadro de pessoal. Conforme afirma Frigo, Yevseyeva e Pozzebon (2013), a multinacional Volkswagen, tem como objetivo chegar a uma proporção de 30% de mulheres graduadas em Engenharia no seu processo de recrutamento.

Mas o número de mulheres nestes cursos ainda é baixo e esta falta de incentivo começa no ensino médio, onde as estudantes sentem-se desmotivadas em estudar os

conteúdos das disciplinas que serão base dos cursos de engenharia (como física e matemática, por exemplo), por estas tratarem de temas que são de difícil assimilação. Em virtude da abordagem técnica, o baixo número de ferramentas que auxiliem os docentes a reforçar determinados tópicos destas disciplinas é um fator impactante no processo de ensino/aprendizagem [POZZEBON e FRIGO, 2013].

Desta forma, a utilização da robótica nestas disciplinas, consideradas mais complicadas, visa instigar a curiosidade e a criatividade das estudantes para solucionarem diferentes tipos de problemas, além de proporcionar um ensino mais motivador e didático [POZZEBON e FRIGO, 2013].

O objetivo deste projeto foi estimular a formação de engenheiras no Brasil, combatendo a evasão das graduandas nos primeiros anos do curso, despertar o interesse das estudantes de ensino médio pela profissão de engenheira de computação, além de promover uma união entre as instituições públicas da região com a comunidade acadêmica.

2. Robótica Educacional

A robótica educacional é uma área multidisciplinar, pois engloba o conteúdo de diferentes disciplinas com o objetivo de construir robôs para desempenhar as mais variadas tarefas [NEHMZOW, 2000].

Por ser uma área que desperta bastante curiosidade, a robótica pode ser usada como ferramenta didática para auxiliar os docentes em diferentes conteúdos dependendo da disciplina a ser ensinada. Para as estudantes, a robótica é uma excelente ferramenta para exercitar a criatividade, estudar e praticar conceitos relacionados as disciplinas [NEHMZOW, 2000].

A vantagem de se construir robôs é o total conhecimento sobre todas as partes que o compõem, proporcionando ao docente a oportunidade de lançar desafios aos alunos, para que estes construam robôs, desde os modelos mais simples até os considerados mais complexos. Como por exemplo, o docente da disciplina de física pode utilizar os robôs para demonstrar na prática os conteúdos teóricos sobre os efeitos de cinemática, dinâmica, inércia, movimento retilíneos, etc [SILVA, et. al. 2008].

A complexidade inerente à robótica irá depender do grau de exploração do docente, uma vez que esta proporcionará uma visão de aprendizagem mais significativa para os alunos. Além disso, ela cria condições de discussão sobre assuntos diversificados promovendo uma interação maior entre docente e discente [POZZEBON e FRIGO, 2013].

3. Desenvolvimento das Atividades

O curso iniciou suas atividades em dezembro de 2014 e estendeu-se até março de 2015, foram realizadas oficinas de física e matemática com as estudantes do primeiro ano do ensino médio da Escola de Ensino Macário Borba localizada em Sombrio-SC, com cerca de 15 (quinze) meninas, para tornar as aulas divertidas e principalmente mais visuais.

O curso era constituído de uma aula, com duração média de 2 (duas) horas, realizada em um dia de aula, ou seja, de segunda a sexta-feira. Foram utilizados os kits da Lego® Mindstorm®, disponibilizados pelo Laboratório de Tecnologias Computacionais - LabTeC. Esses kits consistem de várias peças que foram utilizadas para criar a estrutura de um pequeno robô. Também eram encaixados junto a essa

estrutura motores e sensores para permitir uma maior interação entre o robô e o ambiente de atuação do mesmo.

Uma das finalidades do curso era a de expor aplicações da física e da matemática com o auxílio da robótica. Como a ementa da oficina foi compatível com o que as estudantes trabalharam em sala de aula então tivemos 3 (três) assuntos principais: i) Conceitos Básicos de Cinemática: móvel, referencial, movimento e repouso, deslocamento, velocidade média, aceleração média, funções horárias; ii) Equações algébricas: incógnitas, igualdade; iii) Conservação de Energia: energia potencial gravitacional, energia cinética, energia mecânica.

Foi elaborada uma apostila onde foram apresentados três exercícios em que as estudantes deveriam anotar os valores obtidos em cada experimento e calcular utilizando os conceitos e fórmulas, observando que o objetivo era que elas conseguissem aplicar e entender a conceituação ensinada pelo docente, com isso, comprovar a veracidade das fórmulas.

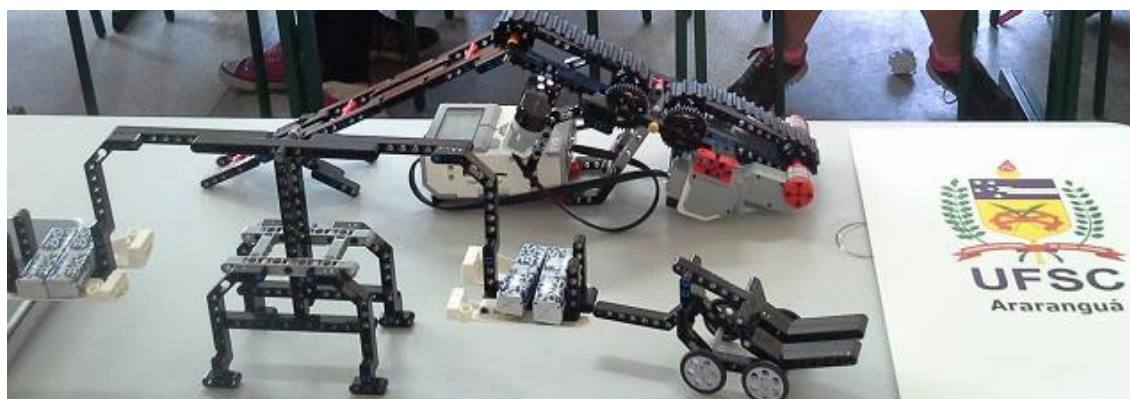


Figura 1: Balança para o experimento das equações algébricas

Para essa experiência, foi criada com os kits de robótica uma balança (Figura 1) para o experimento das equações algébricas, fazendo-se uso do chocolate para o equilíbrio da balança.

Finalizando com o entendimento do conteúdo através da associação de que uma balança pode ser uma igualdade, uma montanha-russa para a questão da conservação de energia, utilizando um motor para fazer com que a trilha de movimento e consequentemente o trem feito com o kit ganha velocidade, então a energia potencial se transforma em energia cinética e permite aos carros completarem o percurso, ou parte dele, através desta força adquirida.

A energia cinética é novamente transformada em energia potencial enquanto o trem se move novamente para o próximo pico, neste caso, há um único pico, e um carrinho para demonstrar o movimento acelerado, mudando as potências do motor, e retardado, simulando uma corrida.

Participaram do curso, exercendo a função de monitoras, estudantes do gênero feminino dos cursos de Engenharia da Computação e Tecnologias da Informação e Comunicação (UFSC – Araranguá). Esta escolha se deu, propositalmente, para incentivar ainda mais as estudantes do ensino médio, mostrando a participação das mulheres nos cursos das áreas tecnológicas.

4. Resultados e Discussões

Os resultados até o momento apresentaram-se positivos; avaliados por meio de observação, notou-se o interesse das estudantes pelo curso, onde aprenderam na prática conceitos de física e matemática, até então considerados complexos auxiliados pelo artefato tecnológico (robótica).

As alunas que participaram deste projeto comentaram em depoimentos sobre a importância de realizar atividades relacionadas com o conteúdo das disciplinas de maneira mais atrativas e interativas. Quanto aos resultados, 100% das alunas interagiram rapidamente com a tecnologia, não houve resistência quanto ao aprendizado com um novo instrumento. Em torno de 35% fizeram questionamentos pertinentes ao conteúdo e cerca de 93% relataram que gostaram do aprendizado diversificado.

Tem-se como expectativa, com a continuidade do projeto, o aumento no número de ingressantes do gênero feminino no curso de Engenharia de Computação e também em outras engenharias, pois este é o objetivo do curso, incentivá-las a ingressarem nas áreas tecnológicas.

Como resultado extra, alguns docentes de escolas que participam do Programa de Mestrado em Física da UFSC (Araranguá) se interessaram pela proposta do projeto e estão desenvolvendo suas pesquisas em novos problemas-desafios de física utilizando os robôs Lego EV3.

5. Considerações

A aplicação dos robôs do kit LEGO como ferramenta didática no ensino de física e matemática para auxiliar a aprendizagem nas disciplinas do nível médio poderá ser utilizada como um exemplo para outras escolas e disciplinas. Ao longo do projeto foram desenvolvidas as seguintes atividades, i) Definição dos problemas de física e matemática que foram abordados com os kits Lego; ii) Preparação do material didático que foi utilizado nas atividades; iii) Treinamento das estudantes bolsistas para o manuseio e o desenvolvimento de soluções com os kits Lego; iv) Atividades didáticas com os problemas de física e matemática; v) Preparação de relatórios e artigos científicos para posterior submissão para congressos e revistas especializadas e vi) Participação da 13ª SEPEX - Semana de Ensino, Pesquisa e Extensão da Universidade Federal de Santa Catarina, objetivando a divulgação do projeto.

Referências

- Frigo, L. B.; Yevseyeva, O.; e Pozzebon, E. (2013) “El análisis de la diferencia de género en la educación. Estudio de caso en Araranguá-Brasil”. In: Mundo Nuevo, Venezuela, n. 13, ano, V, p. 131-145.
- Nehmzow, U. (2000) “Mobile Robotics: A Practical Introduction”. Springer, 243p.
- Pozzebon, E.; e Frigo, L. B. (2013) “Robótica no processo de ensino e aprendizagem”. In: International Conference on Interactive Computer Aided Blended Learning, p. 104-107.
- Silva, A. F.; Agaé, A.; Gonçalves, L. M.; Guerreiro, A. M. G.; Pitta, R.; e Aranibar, D. B. (2008) “Utilização da Teoria de Vygotsky em Robótica Educativa”. UFRN: Natal, 2008. Departamento de Engenharia da Computação e Automação, Universidade Federal do Rio Grande do Norte, Natal, 2008.

Trazendo Meninas para a Computação*

Helena G. Ribeiro¹, Iraci C. S. de Carli¹, Maria de Fátima W. P. Lima¹,

Simone C. M. Paiva¹, Naura A. Luciano¹

¹Centro de Ciências Exatas e da Tecnologia– Universidade de Caxias do Sul (UCS)
Caixa Postal 1352 – 95001-970 – Caxias do Sul – RS – Brasil

{hgrib, icsilvei, mfwplima, scmendes, nalucian}@ucs.br

Abstract. *This article describes the activities performed by a group of high school girls, who participated in the project Bringing Girls for Computer so that they could experience and learn activities executed by professionals in this area. These activities were organized in workshops and visits in the university, involving logical thinking, programming, web page development and robotics, that are roles and functions needed in project development environment. The activities were oriented by teachers of Computer Science on the University of Caxias do Sul.*

Resumo. *Este artigo descreve as atividades realizadas por um grupo de meninas do ensino médio que participaram do projeto Trazendo Meninas para a Computação para que elas pudessem experimentar e conhecer atividades desenvolvidas por profissionais dessa área. Essas atividades foram organizadas em oficinas e visitas que envolveram raciocínio lógico, programação, desenvolvimento de páginas web e robótica, papéis e funções necessários em ambiente de desenvolvimento de projetos. As atividades foram orientadas por professoras da área de computação da Universidade de Caxias do Sul.*

1. Introdução

A área da Tecnologia da Informação na Universidade de Caxias do Sul (UCS) – e em todas as demais Instituições de Educação Superior, historicamente, recebem, em média, quatro ou cinco vezes menos meninas e jovens do que meninos. Como a tecnologia de computação torna-se cada vez mais difundida, esta subrepresentação da mulher na área se traduz em uma perda de oportunidade para os indivíduos, uma perda de talentos para o trabalho, e uma perda de criatividade na construção do futuro da nossa sociedade. Segundo agências como *National Center for Women and Information Technology*¹ e *Girl Scout Research Institute*², as meninas acreditam que os meninos são naturalmente melhores em ciência da computação, que cientistas da computação gostam de trabalhar sozinhos, e que ciência da computação não ajuda as pessoas, mas 90% das meninas querem ajudar as pessoas. Isso termina por desestimular as meninas a seguirem nessa área.

No Brasil, segundo (Medeiros, 2005), as mulheres representavam mais de 60%

* Projeto CNPq 420546/2013-8, edital 18/2013 MCTI/CNPq/SPM-PR/Petrobras - Meninas e Jovens Fazendo Ciências Exatas, Engenharias e Computação

1 <http://www.ncwit.org/>

2 <http://www.girlscouts.org/>

dos graduados no país, mas desses apenas 5% escolhiam a área de Ciência da Computação, apesar de ainda serem maioria em outras áreas científicas como Matemática, Biologia e Medicina. O Censo da Educação Superior de 2011 do INEP apontava que nos cursos de graduação em ciência da computação, 85,4% das matrículas eram masculinas (Anúnciação, 2014). Algumas iniciativas como o Programa “Meninas Digitais” da SBC (SBC, 2016) tem promovido ações para tentar mudar esse quadro.

Atualmente, nos cursos da área da Tecnologia da Informação da UCS, que agregam as graduações em Ciência da Computação, Sistemas de Informação, Tecnologias Digitais Engenharia da Computação e Tecnologia em Análise e Desenvolvimento de Sistemas, tem-se uma proporção significativamente maior de alunos do sexo masculino em relação ao sexo feminino (em 2013, 13% das matrículas eram do sexo feminino). Para compreender os motivos desta diferença, e aproveitando um edital do CNPq (“Meninas e Jovens Fazendo Ciências Exatas, Engenharias e Computação”) que buscava projetos para atrair o público feminino para áreas como a Computação, uma equipe de cinco professoras da área de computação da UCS decidiu criar estratégias que visassem despertar o interesse feminino para a área. Elaborou-se assim o projeto Trazendo Meninas para a Computação, com o objetivo maior de cativar meninas para as profissões relacionadas às áreas de Computação e Informática.

A seção 2 apresenta as atividades e oficinas propostas, e como foram realizadas. A seção 3 relata as contribuições do projeto, e a seção 4 apresenta algumas conclusões e resultados.

2. Atividades e metodologia

Para divulgar a área de computação e desmistificar o fato de a mesma ser identificada como uma área masculina, um grupo de professoras da UCS elaborou um conjunto de atividades que trazem práticas da área de computação em diferentes níveis e setores. Tais atividades compreenderam oficinas presenciais, visitas e encontros, tendo mulheres da área como coordenadoras ou como participantes ativas.

Em parceria com uma escola pública de ensino médio³, foi constituído um grupo de 15 meninas dos segundos e terceiros anos para participar de atividades na UCS durante o ano de 2014. Organizou-se um calendário prevendo dezesseis encontros durante o ano. Os quinze primeiros tiveram a duração de 3 horas cada um, e o último foi de 6 horas marcando o encerramento do projeto.

2.1. Oficinas

Foram elaboradas seis oficinas (Tabela 1), realizadas presencialmente na UCS, em encontros quinzenais, cada uma delas ministrada por uma das professoras da área de Computação.

Um professor da escola acompanhava as meninas durante as oficinas e também em atividades complementares ao trabalho presencial. Duas das quinze meninas foram escolhidas como monitoras (bolsistas do projeto) e auxiliaram o professor no acompanhamento das atividades, tanto nas atividades presenciais como na atividades complementares. Uma aluna do curso de Sistemas de Informação foi selecionada como bolsista para acompanhar a execução das oficinas e encontros, e dar apoio às atividades

³ Escola Estadual de Ensino Médio Décio Martins Costa – Picada Café - RS (<http://escoladecio.com.br/>)

do projeto.

Tabela 1. Oficinas realizadas durante o projeto Trazendo Meninas para a Computação

	Oficinas
1	“O que vai rolar?” (Apresentação do projeto. Uso do MovieMaker para as meninas apresentarem suas expectativas em relação ao projeto) - 6 h
2	“Qual é o desafio?” (Oficina de Raciocínio Lógico: uso de material concreto - Tangram construído em cartolina; blocos lógicos para problemas matemáticos, arames, palitos e ábaco - para resolver desafios, descrevendo os passos para sua resolução) - 6h
3	“Começando a programar!” (Oficina de Scratch: exploração de ambiente para prática de programação, com uso de recursos diversos para construção de pequenas aplicações) - 12h
4	“Criando páginas na Internet” (Oficina de HTML: levantamento de requisitos e desenvolvimento para a Internet) - 12h
5	“Robótica” (Oficina de Robótica: montagem de protótipos com kits Lego e noções de programação dos protótipos) - 6h
6	“Como fazer programas?” (Oficina de projeto e desenvolvimento de protótipos, realizadas por equipes com participantes exercendo diferentes funções na equipe de projeto: motivação para as diferentes carreiras da área de computação) - 3h

As oficinas foram planejadas para: propiciar atividades interessantes e motivadoras para as meninas compreenderem a sua capacidade e talento, apresentar a área de Computação como sendo uma opção de carreira, propiciar contato com linguagens de programação, realizar atividades de raciocínio lógico através da utilização de materiais concretos, conscientizar a respeito da importância da mulher como profissional de Tecnologia da Informação, e vivenciar as diferentes carreiras ou oportunidades de trabalho na área de TI.

2.2. Encontro com mulheres que escolheram a Computação

Como atividades finais foram realizadas um “Open Day” e um “Dia na UCS”, que proporcionaram um contato mais direto com mulheres que atuam na área de computação.

O “Open Day” foi uma visita técnica a uma empresa de desenvolvimento de software para observar e ver de perto como funciona cada cargo dentro de uma empresa da área de Tecnologia da Informação, e nesse contexto conhecer as atividades desempenhadas por mulheres nessa empresa. As meninas tiveram oportunidade de conversar com as profissionais e de apresentar seus portfólios.

O “Dia na UCS” proporcionou, entre outras atividades elaboradas para conhecer um pouco da estrutura da universidade, uma visita ao setor de Gestão de Tecnologia de Informação (GTI) da UCS, com destaque para as tarefas sob a responsabilidade de mulheres. Após essa visita, houve um encontro das meninas com quatro alunas dos cursos de graduação da área de computação da UCS, para troca de ideias e experiências sobre motivação para ser profissional da área, como escolheram o

curso, incentivo para seguir na área, entre outros assuntos.

3. Contribuições do projeto

As alunas do ensino médio que participaram das oficinas, como a maior parte dos jovens hoje em dia, são usuárias exaustivas da tecnologia e da informática. Elas acreditavam que era muito difícil programar e tinham preferência pela área da saúde. As oficinas permitiram que elas conhecessem e experimentassem um pouco mais detalhadamente ao “como funciona”, “como se faz”, daquilo que utilizam, na ideia de que essa introdução mais lúdica ao mundo da computação pudesse motivá-las a estudar e conhecer mais ainda, e até mesmo trazer outros colegas para a área. No final do projeto muitas expressaram que mudaram sua visão inicial sobre a complexidade da computação, salientando que é preciso dedicação e que há oportunidades atraentes para mulheres nas diferentes carreiras da área. Porém, apenas duas meninas manifestaram interesse em seguir na área.

Uma contribuição do projeto foi na formação de recursos humanos especializados para a educação básica e superior. Na educação básica, as meninas que participaram do projeto e o professor que as acompanhou divulgaram as atividades realizadas na escola, e elas tiveram repercussão bastante positiva na comunidade na qual se insere a escola. Na UCS, as atividades realizadas nas oficinas foram divulgadas junto a outros projetos que incentivam alunos e alunas do ensino médio a ingressarem nas áreas das tecnologias e das engenharias, e abriram-se perspectivas de novas atividades nesses projetos.

4. Conclusões e resultados

Espera-se assim que iniciativas como esse projeto possam atrair mais estudantes do ensino médio (principalmente meninas) para os cursos de graduação na área de Tecnologia da Informação. Essa interação de quase um ano (abril a novembro de 2014) entre professoras da UCS, alunas da escola, professor-bolsista e alunas-bolsistas aproximou a escola e suas alunas da área de TI. Assim, possibilitou às professoras que ministraram as oficinas desenvolver estratégias de aprendizagem incluindo motivação, planejamento de práticas na área com foco no público feminino, externo à área (alunas de ensino médio geral, sem base de formação na tecnologia), mas na expectativa de trazer esse público para os cursos de graduação.

Após a finalização do projeto, não foi possível realizar novamente um conjunto de atividades combinadas na área de computação com um grupo dedicado, mas mantem-se algumas oficinas de programação e robótica como parte de um projeto para atração do público feminino para as Engenharias da UCS (“Encorajando Meninas do Ensino Médio para as Engenharias”).

Um dos pontos fortes deste trabalho foi a metodologia utilizada (sociointeracionista sistêmica), propiciando as meninas a refletir sobre estratégias com o professor da oficina na busca das possíveis soluções aos desafios propostos, com o objetivo de resolver problemas e trocar ideias em conjunto. O principal impacto do projeto foi a certeza de que é muito importante investir em oportunidades de disseminação, divulgação e atração das estudantes do ensino médio para a área de TI, pois a maioria delas não se sente atraída pela área, não conhece as atividades envolvidas e as diferentes carreiras que a área oportuniza, além de suas características.

5. Referências bibliográficas

Medeiros, C.B. (2005) “From Subject of Change to Agent of Change — Women and IT in Brazil”. Em: Women and ICT. June 12–14, 2005, Baltimore, MD – ACM.

Anuniação, S. (2014) “Lugar de Menina é na Computação”. Em: Jornal da UNICAMP, Campinas, 28 de abril de 2014 a 11 de maio de 2014 – Nº 595. (<http://www.unicamp.br/unicamp/ju/595/lugar-de-menina-e-na-computacao>)

SBC - Sociedade Brasileira de Computação (2016). Meninas Digitais. Acesso em 30/05/2016. Disponível em <<http://sbc.org.br/institucional-3/meninas-digitais>>

Um Estudo Inicial Sobre As Alunas Ingressantes no Curso de Licenciatura em Ciência da Computação da Universidade Federal da Paraíba

Renata V. de Figueiredo¹, Vanessa Farias Dantas¹, Gabriela Coutinho¹, Ryslânia R. S. Nascimento¹, Ana Caroline C. Nascimento¹, Beatriz Pamela M. Nunes¹

¹Universidade Federal da Paraíba (UFPB) – Campus IV
Departamento de Ciências Exatas
Rio Tinto – Paraíba - Brasil

{renata, vanessa, gabriela, ryslania.rayana, ana.caroline, beatriz.pamela}@dcx.ufpb.br

Abstract. *The decline in female participation in the courses of Information Technology area has attracted the attention of companies and universities in several countries at a time when much is discussed about gender equality and opportunities. In addition to promoting initiatives to encourage the entry of girls in these courses, it is necessary to conduct studies to know the profile of the freshmen and identify patterns that can direct actions against evasion. This paper analyzes data from computer science students in a graduation course, since its creation until nowadays.*

Resumo. *A queda da participação feminina nos cursos da área de Tecnologia da Informação tem chamado a atenção de empresas e universidades em vários países, em um momento em que muito se discute a igualdade de gênero e de oportunidades. Além de promover iniciativas para estimular o ingresso de garotas nesses cursos, faz-se necessário realizar estudos para conhecer o perfil das ingressantes e identificar padrões que possam direcionar ações contra a evasão. O presente trabalho analisa dados referentes a todas as alunas do curso de Licenciatura em Ciência da Computação, desde sua criação até os dias atuais.*

1. Introdução

Os indicadores de ingresso e permanência de mulheres nos cursos de Computação e Ciências Exatas têm se mostrado bastante reduzidos e em queda. Segundo Medeiros (2005), o número de ingressantes mulheres na graduação em cursos de Computação caiu de entre 30% e 40% na década de 80 para 5% e 10% nos anos 2000, ainda que o número de cursos de ensino superior em Computação tenha aumentado 10 vezes, e o número de alunos 50 vezes.

Os dados do CENSO 2014 compilados pelo INEP e publicados pela SBC (2016) reforçam essa ideia. Ainda que a quantidade de mulheres matriculadas em cursos de Tecnologia da Informação (TI) tenha crescido de 28.281 em 2001 para 49.042 em 2014, considerando-se a quantidade total de matrículas nesses cursos, esses números representam uma queda de 24,10% para 15,27%. O mesmo estudo mostra também que o

percentual de mulheres que terminam cursos na área de TI caiu de 31,78% para apenas 16,28%.

Embora Maciel e Bim (2016) relatem várias iniciativas que estão em andamento para divulgar cursos da área de tecnologia para alunas do Ensino Médio em vários Estados brasileiros, ainda existem poucos estudos sobre o perfil e o desempenho das alunas ingressantes dos cursos de TI. Para entender as dificuldades encontradas e as principais causas de evasão, faz-se necessário conhecer a realidade das alunas que se matriculam e traçar estratégias para que estas permaneçam nos cursos.

Alguns trabalhos já investigam a presença feminina na área da computação, como Oliveira *et. al* (2014), que mostra indicativos de pesquisas relacionados aos perfis das mulheres que estudam e trabalham na área de tecnologia da informação, através resultados de questionários respondidos por mulheres de todo o Brasil. A partir desta amostra de dados nacionais, observou-se a necessidade de pesquisar dados por região, para que a realidade do Estado seja analisada de forma mais detalhada.

O presente trabalho propõe um estudo quantitativo inicial baseado em dados acadêmicos de todas as meninas já matriculadas no curso de Licenciatura em Ciência da Computação da Universidade Federal da Paraíba – UFPB, Campus IV, ao longo de dezoito semestres consecutivos. Uma breve caracterização do curso em questão encontra-se na Seção 2, enquanto a metodologia usada para a realização do estudo aparece descrita na Seção 3. A Seção 4 apresenta os resultados observados, bem como breves análises sobre cada indicador, e a Seção 5 contém algumas considerações e indicações de trabalhos futuros.

2. O Curso de Licenciatura em Ciência da Computação

O curso foi criado no segundo semestre de 2006 no campus IV da UFPB, situado em Rio Tinto, um município com aproximadamente 24.000 habitantes situado a 62 km da capital do Estado.

Composto por nove semestres, totalizando 3015 horas-aulas, o curso propõe disciplinas que integram conteúdos de formação básica nas áreas de Ciência da Computação, Matemática, Pedagogia e Ciências Humanas.

Desde o início, o intuito do curso tem sido formar profissionais para atuar no ensino da Computação em diferentes níveis (educação básica, técnica e suas modalidades). Os egressos deste curso devem ser capazes de especificar, projetar, desenvolver, avaliar e implantar tecnologias para prover soluções pedagógicas em ambientes escolares e não escolares. Além disso, o egresso poderá atuar também na formação de usuários da infraestrutura de software e hardware das organizações e no planejamento didático e tecnológico de soluções de ensino-aprendizagem em ambientes acadêmicos e corporativos.

Embora os cursos da área de licenciatura culturalmente tenham um perfil mais feminino, o mesmo não ocorre no curso de Licenciatura em Ciência da Computação, em que a participação feminina tem se mostrado minoritária. Tal particularidade despertou o interesse para a investigação do perfil das alunas ingressantes no curso e de sua atuação durante a graduação.

3. Metodologia da Pesquisa

A pesquisa quantitativa realizada visa compreender o ingresso, permanência e evasão das alunas no curso de Licenciatura em Ciência da Computação da UFPB. A partir de consultas realizadas no Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), foram identificados os nomes, semestres de ingresso, formas de ingresso e situações atuais das 151 alunas ingressantes no curso desde sua fundação (semestre 2006.2) até o semestre corrente (2015.1).

As consultas foram feitas semestre a semestre, e os dados acadêmicos das discentes foram compilados de forma a obter respostas para as seguintes questões:

- Quantas alunas entram semestralmente no curso?
- Quais as formas de ingresso dessas alunas?
- O que acontece com essas alunas ao longo dos semestres?

4. Resultados e Análise de Indicadores

Após a realização das consultas no SIGAA, os dados obtidos foram agrupados e tabulados considerando os critérios necessários para responder cada uma das questões de pesquisa. Os resultados detalhados aparecem nas seções a seguir.

4.1 Quantidade de Alunas Ingressantes

Para entender a proporção de alunas em relação ao total de alunos ingressantes em cada semestre, foram analisados os dados de cada um dos dezoito semestres de existência do curso. É importante destacar que até 2014 a UFPB contava com uma política de matrícula automática de alunos no primeiro semestre de cada curso, não permitindo que fossem feitos trancamentos nem cancelamentos de matrícula. Os resultados obtidos encontram-se na Tabela 1.

Tabela 1 – Quantidade Semestral de Ingressantes por Sexo.

	2006.2	2007.1	2007.2	2008.1	2008.2	2009.1	2009.2	2010.1	2010.2	2011.1	2011.2	2012.1	2012.2	2013.1	2013.2	2014.1	2014.2	2015.1	Total
Masculino	31	38	29	44	37	44	44	51	43	50	44	44	41	40	8	44	36	39	707
Feminino	9	3	11	6	14	6	7	14	11	9	4	8	11	16	6	6	1	9	151
Feminino (%)	23	7	28	12	27	12	14	22	20	15	8	15	21	29	43	12	3	19	17,6
Total	40	41	40	50	51	50	51	65	54	59	48	52	52	56	14	50	37	48	858

Considerando todos os ingressos no curso até o momento, as alunas representam apenas 17,6% do total, e menos de 30% das vagas a cada semestre são preenchidas por garotas. O semestre 2013.2 teve uma situação atípica, pois seu início após uma longa greve ocasionou uma grande sobra de vagas e uma quantidade de ingressantes bem abaixo da média dos demais semestres. Apenas nesse semestre, a quantidade de alunas ingressantes foi quase igual à quantidade de alunos ingressantes. Nos demais, entretanto, a disparidade nos números é notável.

4.2 Formas de Ingresso das Alunas

Existem diversas formas de ingresso nos cursos da UFPB: vestibular (adotado até 2013), ENEM/SISU (adotado a partir de 2011), transferência institucional, reopção de curso, ou ingresso de graduados. Para entender como as alunas têm entrado no curso, foram compilados os dados apresentados na Figura 1.

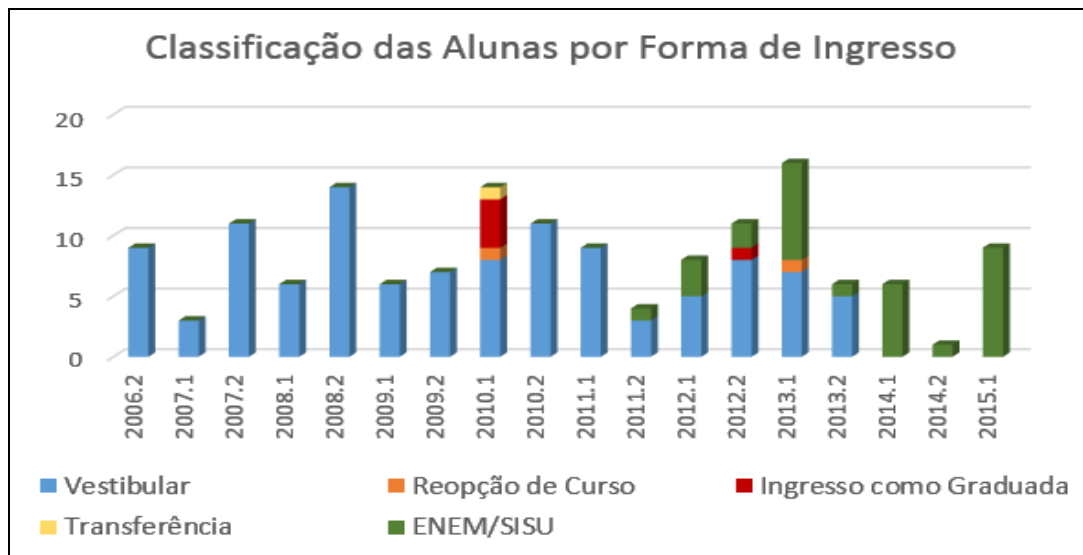


Figura 1 – Classificação das alunas por forma de ingresso.

Os dados coletados indicam que as formas de ingresso mais comuns são o vestibular (74,17%) e o ENEM/SISU (20,52%). Nos dezoito semestres analisados, houve apenas um ingresso por transferência (em 2010.1), dois ingressos por reopção de curso (em 2010.1 e 2013.1), e cinco ingressos de graduadas (em 2010.1 e 2012.2).

4.3 Situação Atual das Ingressantes

Além de conhecer os dados sobre a entrada das alunas do curso, é importante identificar se existe evasão (trancamentos ou cancelamentos de curso) e quantas das alunas ingressantes efetivamente se formaram. Para isso, foi feita uma análise da situação atual de cada aluna ingressante por semestre, e os dados obtidos encontram-se na Tabela 2.

Tabela 2 – Situação Atual das Alunas Ingressantes Por Semestre.

	2006.2	2007.1	2007.2	2008.1	2008.2	2009.1	2009.2	2010.1	2010.2	2011.1	2011.2	2012.1	2012.2	2013.1	2013.2	2014.1	2014.2	2015.1	Total
Total de Ingressantes	9	3	11	6	14	6	7	14	11	9	4	8	11	16	6	6	1	9	151
Concluíram o curso (%)	33,3	33,3	36,4	50	21,4	16,7	0	7,1	0	0	0	0	0	0	0	0	0	0	10,6
Trancaram ou Cancelaram (%)	66,7	66,7	63,6	50	71,4	83,3	42,9	78,6	45,5	66,7	50	50	18,2	18,8	16,7	33,3	0	44,4	50,3
Estão Ativas (%)	0	0	0	0	7,1	0	57,1	14,3	54,5	33,3	50	50	81,8	81,3	83,3	66,7	100	55,6	39,1

Considerando os nove anos de existência do curso, ter apenas dezesseis alunas concluintes indica um índice de sucesso muito baixo. Outro dado preocupante observado foi que 76 das 151 alunas que já ingressaram no curso trancaram ou cancelaram suas matrículas, o que representa mais de 50% de evasão. Esses números apontam para a necessidade de investigar as razões que levam tantas alunas a desistir, e a importância de iniciativas que possam oferecer encorajamento durante a graduação.

5. Considerações Finais e Trabalhos Futuros

Os resultados apresentados no presente estudo quantitativo sobre as alunas ingressantes no curso de Licenciatura em Ciência da Computação representam um esforço inicial no sentido de conhecer o perfil das alunas e identificar problemas de evasão. Espera-se que estudos qualitativos posteriores possam indicar as razões que estão por trás de indicadores tão preocupantes, e que seja realizada em paralelo uma análise sobre os dados dos alunos, para que sejam comparados. Só assim será possível identificar quais resultados estão relacionados a gênero e quais dizem respeito a possíveis problemas do curso.

Com o objetivo de investigar essas e outras questões da participação feminina nos cursos da área de TI, e também de estimular as alunas ingressantes a permanecer no curso, foi criado há um ano na citada universidade o projeto IT Girls – Garotas na Tecnologia da Informação. A realização de algumas ações, como rodas de conversa e palestras, e o incentivo à participação em eventos relacionados ao tema já mostra resultados, como um maior entrosamento entre as discentes e uma confiança mais firme na escolha profissional que fizeram. Espera-se que o projeto tenha ainda vários frutos no sentido de não apenas atrair mais garotas para os cursos da área de TI, mas também de apoiar a permanência delas na academia.

Referências

- INEP - Censo da Educação Superior 2014. Disponível em <http://sistemascensosuperior.inep.gov.br/censosuperior_2014/>. Acessado em 01 de junho de 2016.
- Maciel, C., Bim, S. A. “Programa Meninas Digitais: ações para divulgar a Computação para meninas do ensino médio”, Anais do Computer On The Beach, Florianópolis, 2016.
- Medeiros, C. “From subject of change to agent of change: women and IT in Brazil”, In: Proceedings of the international symposium on Women and ICT: creating global transformation. ACM, p. 15, 2005.
- Oliveira, A., Moro, M., Prates, R. “Perfil Feminino em Computação: Análise Inicial”. Anais do XXII Workshop sobre Educação e em Computação, Brasília, 2014.
- SBC - Educação Superior em Computação – Estatísticas – 2014 Disponível em <http://www.sbc.org.br/documentos-da-sbc/summary/133-estatisticas/1007-estatisticas-da-educacao-superior-2014>. Acessado em 01 de junho de 2016.

Você é o melhor menino desse grupo!

Ana Cristina Oliveira¹, Karolyne Oliveira^{1,2}, Edlane Alves¹, Marcela Tassyany¹,
Yngrid Cabral¹, Thayanne Sousa¹, Fanny Batista¹

¹Grupo de Pesquisa em Redes Convergentes (GPRC)
Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB)
Campina Grande, Paraíba, Brasil

²Dataprev – Empresa de Tecnologia da Informação da Previdência Social
João Pessoa, Paraíba, Brasil

ana.oliveira@ifpb.edu.br, karolyne.alves@dataprev.org.br,
{edlaneoliveiragusmao2,marcelatassyany,yngrid.tecinfo,thayannevls}@gmail.com

Abstract. *This article presents an account of the experiences of a group of women who are joined the technology area seeking to occupy positions in this area. The objective of this paper is to present their personal experiences and which factors influenced and motivated the entry of these women in the area of Information Technology to foster the development of actions that attract and encourage more women to this field.*

Resumo. *Este artigo expõe um relato de experiências de um grupo de mulheres que se uniu buscando ocupar posições de trabalho no universo da tecnologia. O principal propósito é apresentar suas experiências pessoais e quais fatores influenciaram e motivaram o ingresso dessas mulheres na área de Tecnologia da Informação, visando contribuir com a elaboração de ações que possam atrair e estimular uma maior atuação feminina nesse meio.*

1. Introdução

“Você é o melhor menino desse grupo!” Esse foi um elogio recebido por uma das autoras desse artigo do coordenador de um projeto em que apenas ela era mulher. O coordenador fez o comentário na intenção de que seria o melhor elogio que ela poderia receber. Ela deveria se sentir muito feliz por esse elogio? Afinal de contas, ela não era um menino.

Ela estranhou a frase e lembrou do quanto o próprio grupo às vezes desprezava as suas opiniões e das vezes em que se deparou com a vontade de alertar aos colegas de trabalho de que não era um menino. Às vezes no começo da frase, às vezes nem prestando atenção e muitas vezes chegando à mesma conclusão que ela havia sugerido semanas atrás após algum tempo de estudo dos meninos do grupo. Finalmente, depois de um bom tempo, ela foi reconhecida profissionalmente. Ela foi reconhecida ou ele foi reconhecido?

A identidade do profissional de Tecnologia da Informação (TI) é subliminarmente entendida como masculina. O menino da informática para consertar o computador, o programador, o cara que trabalha em campo passando fibra ótica nos postes. As mulheres são igualmente capazes. Só que elas mesmas podem não saber disso. Como fazer para começarem a acreditar nelas mesmas?

Ter mulher no ambiente de trabalho traz à mente a lembrança de entrar em uma sala e sentir um cheiro bom, ver sempre um hidratante por perto, umas canecas coloridas e enfeites na mesa. Vem à mente competência? Quem é mulher na área de TI sabe da obviedade de que não existe diferença em termos de potenciais e capacidades intelectuais apenas por questão de gênero. Existem bons e maus profissionais e para isso não importa o gênero. Uma mulher pode ser femininamente inteligente e competente na área de informática. Um homem pode não se sair tão bem assim.

Duas recentes polêmicas no Brasil desencadearam discursões sobre o empoderamento das mulheres em vários seguimentos. O primeiro por causa de uma matéria de circulação nacional descrevendo o exemplo de mulher que teria o vice-presidente da república Michel Temer, que era bela, recatada e do lar (Linhares, 2016). Esse modelo foi durante muito tempo sendo modificado no país de modo que as mulheres pudessem ocupar diferentes cargos de destaque. Um outro fato polêmico foi a questão de, ao se tornar presidente, Michel Temer não ter nomeado nenhuma mulher para cargos de primeiro escalão. Esses episódios, longe de serem uma exceção, mostram que setores tradicionalmente dominados por homens estão voltando a serem ligados a eles. A computação não foge deste panorama, bem como em outras áreas “difíceis” (Medeiros, 2011). Estudos mostram que as mulheres vêm cada vez mais se distanciando da área de TI (Henn, 2014).

Neste artigo serão tratados relatos de sucesso de mulheres que iniciaram na área de TI ainda no ensino médio, que continuaram seus estudos e que descobriram o quanto essa área é para mulheres, assim como é para homens. O objetivo deste artigo é de compartilhar o que as motivou e encantou na tecnologia, objetivando propor e colaborar com ações que possam atrair e estimular uma maior participação de mulheres.

2. Desigualdade de Gênero

Fatores externos e questões impostas culturalmente tendem a ser internalizadas pelos indivíduos ao longo de sua vida e formar a visão de mundo deles. No que diz respeito ao gênero não é diferente. Desde a infância é comum o estabelecimento de estereótipos e tabus. O conceito de que existem profissões exclusivamente masculinas e femininas é um exemplo disso. A sociedade impõe categorias de gênero: “*Na grande maioria da sociedade, essas supostas diferenças são usadas para justificar diferenças de poder e divisões de trabalho em casa e no mercado de trabalho*”. (Abbate, 2012).

O modelo conservador ainda pode ser visto como sendo o “ideal”. O fato de a mulher contemporânea poder ocupar cargos de liderança ainda assusta. A teoria do materialismo cultural trata da complexidade e dos problemas em se discutir um determinado conceito sem antes analisar o contexto histórico. A discriminação se dá a partir da visão cultural, social e das práticas vivenciadas pelo indivíduo ao longo de sua formação e do passado histórico, o qual ele não vivenciou, mas teve contato por meio das práticas dos seus pais, avós e entre outros (Williams, 2011).

Há um crescente número de iniciativas de empresas, ONG’S e governos que procuram estimular a participação feminina na área tecnológica, como *Technovation Challenge, He for She, Digi Girlz, Women who code, Women Techmakers, WIE, WIT*, entre outras. Essas iniciativas sofrem críticas de pessoas que não compreendem como ações desse tipo podem ajudar a diminuir a desigualdade de gênero. Contudo, elas

contribuem para a desconstrução do estereótipo de que mulheres deveriam ocupar posições “femininas”. Essas ações servem para motivar empresas a realizarem parcerias e trazer mulheres para ocupar cargos de gestão, por exemplo. É importante estimular a presença de mulheres na área tecnológica e outras áreas “difíceis” por meio de políticas públicas, com leis e regulamentações específicas, visando expandir oportunidades econômicas para mulheres (Banco Mundial, 2012).

O apoio dos homens para redução da desigualdade de gênero também vem sendo promovido. Ações como o *He for She* (Eles por elas) buscam combater a violência e a discriminação, mas também reforçar a ideia de igualdade entre homens e mulheres (He for She, 2016). Thyanne tem o exemplo de um irmão que a ajudou a descobrir potenciais na área tecnológica. *“Agradeço ao meu irmão que me convenceu a fazer o técnico, pois agora tenho certeza de que quero continuar na área e seguir para graduação em Ciência da Computação. Eu sei que muitas mulheres não tiveram a mesma sorte que eu em ter apoio para estudar o que tem vontade sem sofrer preconceitos, principalmente nas áreas de exatas e tecnologia. Mas afirmo com certeza de que é apenas necessário você acreditar em si mesma e estar disposta mesmo com as dificuldades a continuar. Ninguém poderá impedir você de trilhar o caminho que quer e existe muito suporte online para quem queira começar, como eu estou agora”*.

3. Quem São Elas?

As alunas que serão apresentadas nesta seção possuem desempenho escolar bem acima da média. Elas estudam com rapazes e são reconhecidas por suas competências por alunos, alunas, professores e professoras. Um ponto importante a ser ressaltado é que todas as autoras deste artigo iniciaram nesta área por meio de cursos técnicos. Nesses cursos, elas se identificaram com a área e optaram por fazer um curso superior que estivesse dentro da meta que traçaram para o futuro profissional.

A crescente procura por parte das empresas, a remuneração cada vez mais igualitária entre ambos os sexos, bem como a mudança cultural da visão da mulher em todas as áreas inclusive em TI, torna-se um estímulo para que as mulheres continuem no curso e sigam uma futura profissão nesta área (Huallem, 2013).

Thyanne Luiza tem 16 anos e atualmente está concluindo o ensino médio e também o técnico integrado em informática em um instituto federal. Tudo começou quando ela tinha 13 anos e começou a ter interesse sobre alguns assuntos de programação. Como tinha um irmão que cursava Ciência da Computação, acabou pedindo ajuda e sendo influenciada por ele a entrar nesse mundo.

“Meu irmão teve um papel importante, orientou-me e me ensinou programação. Logo depois, teve a ideia de eu tentar ingressar no IFPB para aprofundar meus conhecimentos e conhecer mais sobre a área da computação”. Thyanne ingressou no IFPB em 2014, teve suas primeiras experiências com olimpíadas de programação, OPI (Olimpíada Paraibana de Informática) e OBI (Olimpíada Brasileira de Informática), assim como participações em projetos de pesquisa e atualmente está participando de um time que está desenvolvendo uma ferramenta que trabalha com QR-Code e Android para autenticação de produtos.

O curso técnico foi de grande importância para motivar Thyanne a se dedicar à área de TI. *“Apesar de em muitos lugares encontrarmos preconceito contra mulheres no campo da computação, no IFPB o que recebi foi apoio e suporte para expandir meu conhecimento e trabalhar no que eu gostava, tanto quanto o reflexo que tínhamos das nossas professoras mulheres na computação”*.

O interesse de Fanny pela área de informática começou com a popularização das redes sociais. Ela desejou saber como funcionava esse “universo”. Até que um dia soube de uma escola que oferecia o curso técnico de informática juntamente com o ensino médio. Era necessário participar de um processo seletivo. Ela se esforçou para conseguir a vaga. Atualmente se encontra no último ano do curso. *“Percebo que todo esforço que tive para aprender programação e afins, valeu a pena”*.

As alunas Edlane (técnica em Telecomunicações), Marcela (técnica em Manutenção e Suporte em Informática) e Yngrid (técnica em Tecnologia da Informação) são graduandas do 5º período do curso superior de Tecnologia em Telemática do IFPB. Um tecnólogo em Telemática alia conhecimento em três principais áreas: Telecomunicações, Redes de Computadores e Informática, estando capacitado a trabalhar no projeto e gerência de infraestruturas de TI e no desenvolvimento de sistemas. As três formam uma equipe feminina que desde o início do curso se uniu com o propósito de fazer a diferença em um universo predominantemente masculino, que é o universo da tecnologia. *“Nossa área de atuação predominante é redes de computadores, com foco em computação em nuvem, virtualização e segurança, mas também já trabalhamos em telecomunicações com comunicações ópticas sem fio”*.

Ana Cristina e Karolyne são irmãs. Começaram juntas ainda muito jovens a estudar sobre Internet e operação de microcomputadores. Fizeram curso técnico integrado em informática, o curso superior em Tecnologia em Telemática na instituição que hoje se chama IFPB, mestrado e doutorado em Ciência da Computação. Ana Cristina tornou-se professora do IFPB e Karolyne é analista de Tecnologia da Informação na Dataprev.

Karolyne relata: *“eu comecei a me interessar desde a adolescência, quando chegaram os primeiros computadores na minha cidade. Nesta época, houve curso de capacitação no parque tecnológico (PaqTc). Era de fato um curso voltado para homens adultos e apenas nós (eu e minha irmã) éramos do sexo feminino e ainda mais adolescentes. Desde então, gostamos muito da área e nos aprofundamos. Mesmo com bastante dedicação de nossa parte, percebemos a predileção de cargos para pessoas do sexo masculino, mas não esmorecemos e fomos à luta. Formei-me como única mulher da turma. Dentro da área de informática, atuei em vários segmentos. Comecei como programadora, fui professora de programação Web. Neste período, era a única programadora da empresa em que trabalhava. Passei em um concurso público para empresa pública do setor de TI, onde majoritariamente os cargos mais elevados são sempre ocupados por homens...”*

Ana Cristina atualmente coordena de projetos de pesquisa, extensão e de inovação, já atuou como coordenadora dos cursos técnicos em Informática e é coordenadora do curso superior de Tecnologia em Telemática no IFPB campus Campina Grande. Ela é membro dos grupos de mulheres na em TI da ACM e do IEEE (WIT - Women in Information Technology - e WIE – Women in Engineering), desenvolve ações

para incentivar mulheres a despertarem interesse para área, como apoio a projetos de extensão para visitar escolas públicas e fazer demonstrações da prática de TI. “Trabalhar no IFPB para mim é uma oportunidade de poder retribuir a formação que recebi e os professores e as professoras inspiradores que eu tive”.

4. Conclusões

Ainda é comum as mulheres enfrentarem preconceitos em universidades, empresas e até mesmo dentro da família. Desta forma, é de fundamental importância o desenvolvimento de projetos que visem a conscientização, quanto ao que se rotula ser o lugar da mulher, principalmente para as meninas que ainda se encontram no ensino médio e receiam a falta de espaço nessa área ou que ainda não a conhecem.

O grande problema é que muitas meninas e mulheres tem medo de arriscar, justamente pelo preconceito e falta o incentivo e campanhas que possam influenciar essas mulheres a deixar de lado esse tabu que se criou. Nas instituições e universidades deveriam existir mais campanhas e incentivos que possam conscientizar e abrir espaço para diálogos e discussões sobre o problema. Futuramente, esperamos que o número de mulheres na área de Tecnologia da Informação no Brasil possa crescer e quebrar essas barreiras de preconceito que existem nas escolas, universidades e ambientes de trabalho, dando oportunidade para todas as mulheres brasileiras.

Referências

- Abbate, J. (2012) *Recoding Gender: Women's Changing Participation in Computing*. The MIT Press.
- Banco Mundial. (2012) Relatório sobre o desenvolvimento mundial: igualdade de gênero e desenvolvimento. <http://siteresources.worldbank.org/INTWDR2012/Resources/7778105-1299699968583/7786210-1315936231894/Overview-Portuguese.pdf>
- HeForShe. (2016) Eles por elas. www.heforshe.org. Acessado em: 02/05/2016.
- Henn, S. (2014) When women stopped coding. <http://www.npr.org/sections/money/2014/10/21/357629765/when-women-stopped-coding>. Acessado em 02/05/2016.
- Huallem, D. Mulheres na TI: um bem escasso e precioso. (2013) http://www.catho.com.br/cursos/mulheres_na_ti_um_bem_escasso_e_precioso. Acessado em: 02/05/2016.
- Linhares, J. (2016) Marcela Temer: bela, recatada e “do lar”. *Revista Veja*. Publicado em: 18/04/2016. <http://veja.abril.com.br/noticia/brasil/bela-recatada-e-do-lar>. Acessado em: 02/05/2016.
- Medeiros, C. R. de O., Borges, J. F. (2011) Abram-se às Mulheres todas as Portas!": Conversas em Blogs de Mulheres em Carreira de TI. In *XXXV Encontro da ANPAD*, Rio de Janeiro.
- Schwartz, J., Casagrande, L. S., Leszczynski, S. A. C., e Carvalho, M. G. (2006) “Mulheres na informática: quais foram as pioneiras?”, *Cadernos Pagu*, (27), 255-278. <https://dx.doi.org/10.1590/S0104-83332006000200010>.
- Williams, R. (2011) *Cultura e sociedade: de Coleridge a Orwell*. Petrópolis, RJ: Vozes.

WPerformance 2016 – XV Workshop em Desempenho de Sistemas Computacionais e de Comunicação

Apresentação

O estudo e a análise de desempenho nas mais diversas áreas da computação é uma prioridade nos dias de hoje. O Workshop em Desempenho de Sistemas Computacionais e de Comunicação, mais conhecido como WPerformance, é um fórum para discussão e disseminação de ideias sobre métodos, técnicas e ferramentas para avaliação de desempenho de sistemas computacionais e de comunicação. São de particular interesse do evento os trabalhos que apresentam novas metodologias ou que aplicam de forma inovadora os métodos existentes para modelar, projetar, avaliar e otimizar o desempenho de sistemas computacionais em geral.

O WPerformance 2016 é a 15a. edição do evento colocada no CSBC (Congresso da Sociedade Brasileira de Computação). Nesta edição, foram submetidos 25 trabalhos para trilha principal, sendo 9 aceitos para publicação como trabalhos completos o que significa uma taxa de aceitação de 36%. Os trabalhos aceitos englobam um amplo espectro de tópicos, como avaliação de desempenho de bancos de dados, avaliação de desempenho de redes de computadores e avaliações de transações BitCoin, o que caracteriza a boa aceitação de diversas comunidades.

Esperamos que o WPerformance 2016 seja bastante proveitoso para todos!

Daniel de Oliveira (Universidade Federal Fluminense - UFF)

Wagner Meira Jr. (Universidade Federal de Minas Gerais - UFMG)

Comitê de Organização

Coordenação Geral

Daniel de Oliveira (UFF)
Wagner Meira Jr. (UFMG)

Coordenação Local

Afonso Salles (PUC-RS)

Coordenação do Comitê de Programa

Daniel de Oliveira (UFF)
Wagner Meira Jr. (UFMG)

Comitê de Programa

Alex Borges Vieira (UFJF)
Ana Paula Couto da Silva (UFMG)
Antônio Abelém (UFPA)
Antonio Rocha (IC/UFF)
Antonio Tadeu Gomes (LNCC)
Artur Ziviani (LNCC)
Carlos Alberto Vieira Campos (UNIRIO)
Daniel Batista (IME-USP)
Daniel de Oliveira (IC/UFF) – *co-chair* 2016
Daniel Figueiredo (UFRJ)
Daniel Menasche (UFRJ)
Edmundo de Souza e Silva (COPPE/UFRJ)
Edmundo Madeira (UNICAMP)
Eduardo Ogasawara (CEFET/RJ)
Eduardo Xavier (UNICAMP)
Esteban Clua (IC/UFF)
Fabio Andre Machado Porto (LNCC)
Genaina Rodrigues (UnB)
Gustavo Figueiredo (UFBA)
Islene Garcia (UNICAMP)
João Netto (UFRGS)
José Ferreira de Rezende (UFRJ)
José Augusto Suruagy Monteiro (UFPE)
Kary Ocaña (LNCC)
Lasaro Camargos (UFU)

Laura Assis (CEFET/RJ)
Leonardo Murta (IC/UFF)
Lucia Drummond (IC/UFF)
Luidi Simonetti (COPPE/UFRJ)
Luiz Manoel Rocha Gadelha Júnior (LNCC)
Magnos Martinello (UFES)
Marcelo Carvalho (UnB)
Marta Mattoso (COPPE/UFRJ)
Miguel Elias Mitre Campista (UFRJ)
Morganna Diniz (UNIRIO)
Paulo Aguiar (UFRJ)
Paulo Fernandes (PUCRS)
Pedro Velloso (UFRJ)
Rafael Melo (UFBA)
Rafael Pasquini (UFU)
Ricardo Czekster (UNISC)
Rodolfo Villaca (Ifes)
Romulo Silva de Oliveira (UFSC)
Rosa Leão (UFRJ)
Sidney Lucena (UNIRIO)
Thais Webber (PUCRS)
Thiago Noronha (UFMG)
Tiberius Bonates (UFC)
Vanessa Braganholo (IC/UFF)
Wagner Meira Jr. (UFMG) – *co-chair* 2016
Yuri Frota (IC/UFF)

Painéis / Palestras

Métricas de Desempenho e Qualidade em Sistemas de Informação Web

Altigran Soares da Silva (UFAM)

Resumo: Como qualquer outro tipo de sistema de informação, sistemas de informação Web devem ser capazes de responder rapidamente as demandas de seus usuários. No entanto, diversos outros critérios de desempenho de ser levados em consideração e são determinantes na percepção da qualidade que os usuários tem destes sistemas. Em particular, sistemas de busca e recomendação de produtos em sites de comércio eletrônico devem atender a dois requisitos importantes: qualidade da resposta fornecida aos usuários e efetividade na tarefa de conversão de vendas. Nesta apresentação, discutirei um pouco da minha experiência em métricas de avaliação para este tipo de sistema, tanto do ponto de vista acadêmico, como pesquisador das áreas de gerência de dados e recuperação de informação para a Web, como do ponto de vista prático, como fundador de uma startup na área de tecnologia de informação para o varejo digital.

Biografia: Altigran Soares da Silva é professor associado do Instituto de Computação da Universidade Federal do Amazonas (IComp/UFAM) onde atua como pesquisador, professor e orientador na graduação, mestrado e doutorado. Concluiu seu doutorado em Ciência da Computação pela UFMG em 2002. Seus interesses de pesquisa envolvem Gerência de Dados, Recuperação de Informação e Mineração de Dados com ênfase no ambiente da World-Wide Web e Mídias Sociais. Sobre estes temas, tem coordenado e participado de dezenas de projetos de pesquisa que resultaram em mais de 100 publicações científicas em periódicos e anais de conferência de boa qualidade nestas áreas. Em 2007 foi o Coordenador do Comitê de Programa do Simpósio Brasileiro de Bancos de Dados (SBBBD) e em 2010 atuou com co-chair da trilha "Bridging Structured and Unstructured Data" da International World Wide Web Conference. Participou também como membro de comitês técnico de programa em cerca de 40 conferências e workshops realizados no Brasil e no Exterior. Em 2012, foi indicado como palestrante convidado para o SBBBD. Exerceu entre 2007 e 2009 a Pró-reitoria de Pesquisa e Pós-Graduação da UFAM. No triênio 2011-2013 foi o Coordenador Adjunto da área de Computação na CAPES e entre 2005 e 2015 foi membro da diretoria da Sociedade Brasileira de Computação (SBC), sendo atualmente membro do conselho da Sociedade. É co-fundador de empreendimentos de tecnologia, entre eles a Akwan Information Technologies, adquirida pela Google Inc. em 2005, e a Neemu.com, empresa de tecnologia para aplicações em varejo on-line que líder no e-commerce brasileiro e que foi adquirida pela Linx Sistemas em 2015. Em 2013 uma tese de doutorado sob sua orientação recebeu o Primeiro Lugar no Concurso de Teses e Dissertação da Sociedade Brasileira de Computação (SBC) e Menção Honrosa no Prêmio CAPES de Teses. Recebeu também em 2013 o prêmio de Sócio Destaque da SBC por sua atuação junto às Comissões Especiais da sociedade, contribuindo para o aperfeiçoamento do Qualis CAPES de Conferências na área de Ciência da Computação. Este ano foi ganhador de um dos "Google Research Awards in Latin America" como orientador.

Design using Processor in Memory and Multicore

Luigi Carro (UFRGS)

Resumo: In multi-core architectures DDR memories are being used to provide data for the increasing number of cores. However, in order to provide a higher bandwidth, the DDR memories require a large number of channels and memory controllers, in order to hide the DRAM latency through the parallelism of the devices. Considering the increasing bandwidth requirements, 3D stacked memories like the Hybrid Memory Cube (HMC) are becoming popular, for their high bandwidth capabilities. Moreover, the extra logic layer can provide processing inside the memory, without expensive data transfers. In this talk we will discuss how these new memory devices can impact algorithm design, as well as the hardware design tools for verification and simulation.

Biografia: Luigi Carro was born in Porto Alegre, Brazil, in 1962. He received the Electrical Engineering and the MSc degrees from Universidade Federal do Rio Grande do Sul (UFRGS), Brazil, in 1985 and 1989, respectively. From 1989 to 1991 he worked at ST-Microelectronics, Agrate, Italy, in the R&D group. In 1996 he received the Dr. degree in the area of Computer Science from Universidade Federal do Rio Grande do Sul (UFRGS), Brazil. He is presently a full professor at the Applied Informatics Department at the Informatics Institute of UFRGS, in charge of Computer Architecture and Organization courses at the undergraduate levels. He is also a member of the Graduation Program in Computer Science at UFRGS, where he is co-responsible for courses on Embedded Systems, Digital signal Processing, and VLSI Design. His primary research interests include embedded systems design, validation, automation and test, fault tolerance for future technologies and rapid system prototyping. He has advised more than 20 graduate students, and has published more than 150 technical papers on those topics. He has authored the book *Digital systems Design and Prototyping* (2001-in Portuguese) and is the co-author of *Fault-Tolerance Techniques for SRAM-based FPGAs* (2006-Springer), *Dynamic Reconfigurable Architectures and Transparent optimization Techniques* (2010-Springer) and *Adaptive Systems* (Springer 2012). In 2007 he received the prize FAPERGS - Researcher of the year in Computer Science. His most updated resume is located in <http://lattes.cnpq.br/8544491643812450>. For the latest news, please check www.inf.ufrgs.br/~carro.

Trabalhos aceitos

A preemptive fair scheduler policy for disco MapReduce framework

Augusto Souza (UNICAMP)

Islene Garcia (UNICAMP)

LoadTime is money! Avaliação da performance Web de e-commerces no Brasil

Cristiano Politowski (UFESM)

Gabriel Freytag (UFESM)

Vinícius Maran (UFESM)

Lisandra Fontoura (UFESM)

Usando a elasticidade de recursos em nuvem para aumentar o desempenho de aplicações pipeline

Rodrigo Righi (UNISINOS)

Mateus Aubin (UNISINOS)

Cristiano Costa (UNISINOS)

Guilherme Galante (UNIOESTE)

Análise de desempenho de handover em rede heterogênea usando MIH

Marco Teixeira (UTFPR)

Alexandre Veloso Matos (UDESC)

Mario Augusto (UDESC)

Análise de desempenho de brokers MQTT em sistema de baixo custo

Andrei Torres (UFC)

Atslands Rocha (UFC)

Neuman Souza (UFC)

Dinâmica das transações do Bitcoin: uma abordagem quantitativa

Saulo M. R. Ricci (UFRJ)

Alex Borges Vieira (UFJF)

Helder Luiz Caldas (IFNMG)

Daniel Menasché (UFRJ)

Eduardo Ferreira (UFRJ)

Energy consumption evaluation of NoSQL DBMSs

Carlos Araújo (UFPE)

Eduardo Tavares (UFPE)

Meuse Junior (UFPE)

sAnalytical modeling approach of routing deflection for intra-domain networks

Rodolfo Gomes (UFES)

Cristina Dominicini (IFES-ES)

Alextian Liberato (IFES-ES)

Moises Ribeiro (UFES)

Magnos Martinello (UFES)

FLECHA: a Forecasting eLEction meCHANism for semantic collectors sensor nodes

Maurício Moreira Neto (UFC)

Leonardo Moreira (UFC)

Danielo G. Gomes (UFC)

A Preemptive Fair Scheduler Policy for Disco MapReduce Framework

Augusto Souza¹, Islene Garcia¹

¹ Instituto de Computação
Universidade Estadual de Campinas – Campinas, SP – Brasil

augusto.souza@students.ic.unicamp.br, islene@ic.unicamp.br

Abstract. *Disco is an open source MapReduce framework and an alternative to Hadoop. Preemption of tasks is an important feature which helps organizations relying on the MapReduce paradigm to handle their heterogeneous workload usually constituted of research (long duration and with low priority) and production (short duration and with high priority) applications. The missing preemption in Disco affects the production jobs when these two kinds of jobs need to be executed in parallel: the high priority response is delayed because there aren't resources to compute it. In this paper we describe the implementation of the Preemptive Fair Scheduler Policy which improved largely our experimental production job execution time with a small impact on the research job.*

1. Introduction

MapReduce as proposed by Google in 2004 had the objective of giving developers a computational model in which common distributed systems problems could be addressed by the framework and not by each application [Dean and Ghemawat 2004]. This model is based on a simple abstraction in which programmers need to implement two main functions called Map and Reduce. The first one divides the entry into small groups to be processed in parallel by the cluster's machines, the outputs of the mappers - processes responsible for executing the Map function - are the entry for the reducers. The Reduce function is responsible for combining the data from multiple mappers into a final output. For all of this to work effectively, it is important to have a reliable and distributed file system - e.g. GFS [Ghemawat et al. 2003], HDFS [Shvachko et al. 2010], and DDFS [Disco Distributed Filesystem Website].

Based on Google's MapReduce and File System, some open source projects started to appear and address the same issue in a more accessible way, since the Google's version is only available for their internal use. The main one is Apache Hadoop [Apache Hadoop Website], but there are others like the one we study in this paper, developed by Nokia in 2008 and called Disco [Mundkur et al. 2011]. They both address similar problems but with different architectures and sizes. Hadoop is much larger in terms of code base, investment and complexity than Disco. The last one by being simpler also brings to the table a more flexible solution. To help us compare the complexity in terms of project size we gathered some statistics from their repositories with the help of a tool called *cloc* [Cloc Project at Github.com]. For this, we used the most recent git tags for these projects in the time of this writing: "release-2.7.1" for Hadoop and "0.5.4" for Disco. With much less lines of code than Hadoop (see Table 1), the development of

Table 1. Comparison of projects sizes of Disco and Hadoop.

	Hadoop	Disco
Two Most Used Languages	Java and XML	Erlang and Python
Files in Both These Languages	6,474	190
Lines of Code in Both These Languages	1,688,407	19,134

new features in Disco is faster, easier to implement and test, and also has more chances of becoming an improvement integrated in a future project release. Disco is written in Erlang, a language proposed by Ericsson and focused on distributed systems and concurrency, so it can use all the power the language and its virtual machine to support a distributed computation [Armstrong 2003]. Also, Disco MapReduce applications are majorly written in Python, a language with a large history in scientific computing and data processing, also known for being productive and quick to prototype in. In Section 2, we briefly describe how the configuration of Disco can be performed - the cluster administrator is able to do every configuration step through an intuitive web interface - this is another area in which Disco shines when compared to Hadoop and its dozens of XML files of configuration.

The large adoption of the MapReduce model by the industry caused a necessity not yet addressed by Disco, but already addressed by Hadoop: preemption of tasks. Often, clusters are used by different types of applications with different kinds of priorities, and for some uses it is important to be able to suspend the execution of parts of the applications with less priorities as soon as another with more priority arrives. We divide the jobs of this heterogenous workload into two main categories: research (long duration and with low priority) and production (short duration and with high priority) [Cheng et al. 2011]. Nowadays, Disco is only able to schedule applications respecting FIFO (First in First Out) and Fair policies. Both have serious limitations: FIFO needs to wait for a job to complete in order to schedule a new one, while Fair takes time to give to the production (high priority) application its resources. The limitation related to Fair scheduler policy is our concern in this paper. With no preemption of tasks, the production job waits for the tasks of any running research job to finish before being able to share the cluster resources and starts its execution. With the preemption mechanism we propose, the applications submitted to the Preemptive Fair Scheduler should get its resources as soon as possible.

Besides the submission of tasks, another critical period for scheduling tasks in Disco environment occurs between transition phases of the jobs. These are the phases between the end of the map and the begin of the shuffle and the end of the shuffle and the begin of the reduce. It is during shuffle that reducers copy the mappers output to start their computation. The Preemptive Fair Scheduler waits the transition phases to finish in order to get more accurate information about the running jobs and allocate the resources in accordance to its policy.

The rest of the paper is organized as follows. We present Disco architecture and its schedulers in detail in Section 2. In Section 3, we explain the proposed Preemptive Fair Scheduler Policy and advocate for its necessity. In Section 4, we provide some results obtained by our implementation. In Section 5, we discuss related work. We analyze future work and conclude in Section 6.

2. Disco's Architecture

The architecture we study is the one of version 0.5.4 of Disco. In this version, just like the Hadoop pre-YARN (Yet Another Resource Negotiator) [Vavilapalli et al. 2013], Disco works with a simple master-slave architecture. It is able to achieve its objective of running on a large cluster of networked commodity machines with this simple model. The master is responsible for two key activities:

- initialization of every other component of Disco, like the logging mechanism, web interface, and workers which run in the slaves nodes;
- scheduling, monitoring and allocating resources for tasks from new job submissions of the clients.

The Disco's master needs few configurations to start its slaves: just the number of workers each node will run and their hostnames. The usual is to give to the hosts the same number of workers as they have of CPU cores, since workers are used to run job tasks. After this setup, the master uses the hostnames to launch a single slave Erlang node in these hosts. The slave nodes are responsible for launching and monitoring the execution of tasks specified by the submitted jobs. Also, the slave nodes are the units of storage used by Disco Distributed File System (DDFS) [Disco Distributed Filesystem Website]. Administrators can add and remove slave nodes in real time.

The decision of using the Erlang Virtual Machine for Disco's core shines when we take a look at how nodes interact with each other. The connections between master and slaves are monitored and maintained by Erlang's monitor node. Since the language and its virtual machine have been designed especially for distributed systems, there are useful features provided out of the box for handling things like communication, fault tolerance, reliability and scalability. This way when a node fails, the master is notified and can reschedule the task in another host. Also, all the clustering setup is not responsibility of Disco, it can rely on the Erlang Virtual Machine for communication between cluster machines.

A summary of Disco's master-slave architecture and the communication between these nodes is illustrated by Figure 1. One of the objectives in this architecture is to let Disco MapReduce applications to be written in a range of programming language. In order to achieve this design goal, there is a protocol for the communication of the slave nodes and its workers, nowadays besides the default client in Python there are others in OCaml [OCaml Worker at Github.com], Golang [Go Worker at Github.com], LFE [LFE Worker at Github.com], and Haskell [Haskell Worker at Github.com], giving the programmers the ability to use the programming language they feel more comfortable with. The protocol uses the standard error file descriptor (stderr) for messages from the worker to Disco slave and the standard input file descriptor (stdin) for messages from the Disco slave to the worker. The message flow is always initiated by the worker and there are messages to notify about a bunch of events, such as the startup, task information request, errors while retrieving inputs, and others.

Another interesting piece of the Disco's architecture is its Pipeline. Like Hadoop developed YARN to support a more vast range of applications than just MapReduce, Disco developed the Pipeline for similar purpose. The idea is to define jobs as a linear sequence of stages constituted of a computation and a grouping function. The output

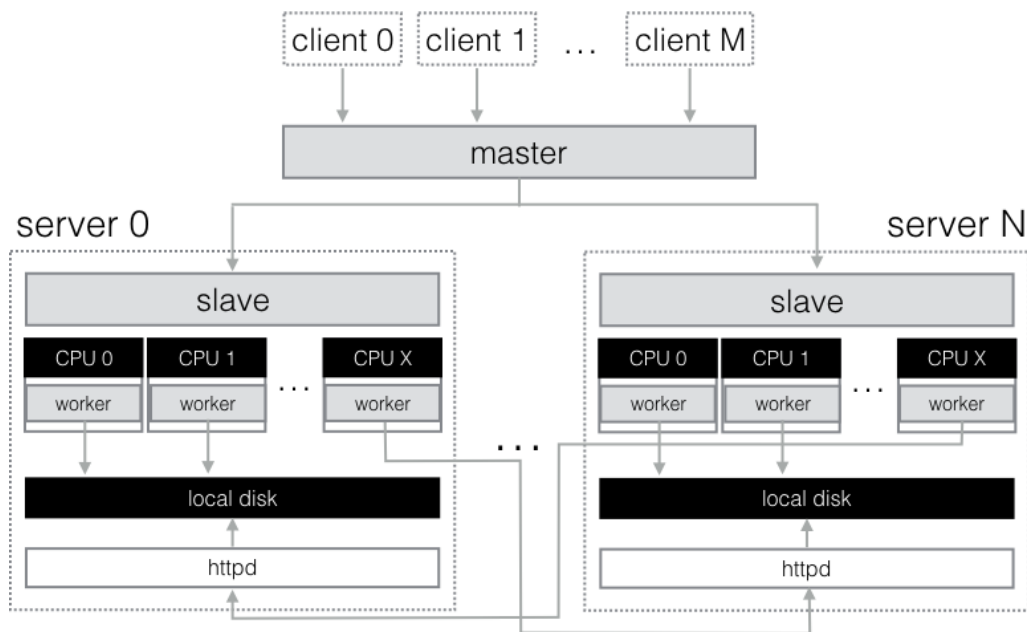


Figure 1. Disco's architecture overview. Gray scale squares are the main components (master, slave and workers) [Mundkur et al. 2011].

of a stage is the input of the next one and there are a bunch of pre-defined grouping possibilities to help the programmers of Disco applications.

Disco has two possible scheduling policies by default: FIFO and Fair. The first one works as a simple queue in which the running job gets all cluster resources it needs and executes up to its end while the others wait. In FIFO policy, when there are available resources the first job in the queue is always selected by the scheduler, if it has no more tasks ready to be executed the next job in queue is chosen. The Fair scheduler tries to divide the cluster in equal shares giving the running jobs the same amount of cluster resources in average to run their tasks. In order to achieve this behavior the Fair scheduler policy cycles through the running jobs when asked to select a task from one of them to give an available resource to. Also, it calculates a deficit that is increased by the amount of time a job had to wait or had fewer resources than the other jobs. The default fair policy always try not to have unused resources while there is work to be done.

Figure 2 shows a case in which four jobs have been submitted to a cluster. The left side of the figure, shows a scheduler configured with FIFO policy, while the right side shows the case for Fair policy. The scheduler makes three of these jobs wait in a queue for the FIFO policy scenario. In the Fair policy case, the scheduler does not need a queue, instead it uses a pool to store the information related to the running jobs and without prioritization it tries to schedule every job with the same amount of resources.

Both these policies have problems when trying to work with a workload composed of research and production jobs. FIFO does not divide the cluster resources within the running jobs, so the production job starts only after a long time, because the research job takes every resource of the cluster for a long time. Fair policy is not ready to preempt running tasks of the research job to give the resources needed by the production job faster. Also, Fair policy isn't ready to avoid the research job from taking every resource of the

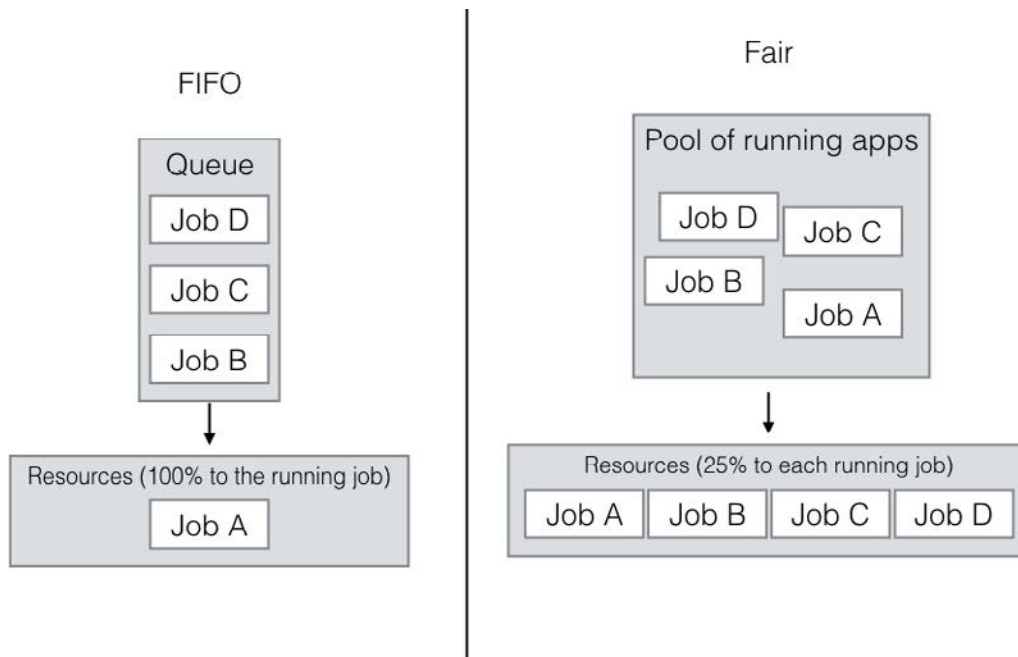


Figure 2. Cluster resources distributions comparison for both FIFO and Fair scheduler policies of Disco.

job during transition phases. These are special periods in a job lifetime for Disco and if the scheduler policy does not monitor that the production job is in this period and is still calculating the amount of remaining work, it can make a mistake by giving the resources for the research job. It greedily gets and locks more resources than it should. The Preemptive Fair Scheduler Policy we propose in this work addresses these limitations.

3. Preemptive Fair Scheduler Policy

Disco's scheduler has a pluggable policy architecture to rely on when some events occur, e.g. a new job submission, the end of a job's execution or an new available slot for task execution. Respecting this pluggable architecture, there are already FIFO and Fair policies as options for the cluster administrator as we described in the previous section. We decided to implement a new policy to add preemption of tasks to the scheduler. The Preemptive Fair Scheduler Policy acts when one of the following events occur:

New job submission:

In order to be fair with the new job as soon as possible, the policy calculates the fair share for it and kill tasks evenly from the running jobs making room for it;

New slot is available:

There is room in the cluster for a new task to be scheduled in, but the Preemptive Fair Scheduler policy must first check if there is a running job in a transition phase. If there is one, then it is better to wait by not scheduling any job in the available slot, because the amount of remaining work is still unknown by Disco at this moment. If there is no job in a transition phase, then the policy checks if a recently submitted job needs to get its first resources, if so it gives them as soon as the job is able to be executed; if not, it makes a list of the jobs sorted by their

running tasks quantity, then filter only those with the number of running tasks smaller than the fair share, the most unfavored job is the first one in this list and is the one that is chosen to use the available slot;

Execution of job finished:

The policy removes this job from every internal data structure it has to maintain to help with the scheduling decisions;

This policy focus on being fair by giving the new job resources for its execution as soon as possible using preemption. It also makes sure the fair number of resources will be available for the desiring job even if some waiting needs to be done during transition phases.

4. Experiments

Our experiments have been done with the sponsorship of Digital Ocean [Digital Ocean Website], we configured a 5 nodes cluster (connected as a private network) based in their New York data center. Each node has 2 cores, 2 GB of RAM memory, 40 GB of SSD Disk, and 3 TB transfer. They are running Disco on top of Ubuntu 14.04.3 LTS, Erlang/OTP R16B03, and Python 2.7.6. We compared two scheduler policies of Disco, for this we used a setup with the original version 0.5.4 got from the Github repository [Disco at Github.com] configure with the original Fair Scheduler Policy and a setup of our own adapted to run the Preemptive Fair Scheduler Policy [Disco Fork at Github.com].

In order to simulate the heterogenous workload of research and production jobs we use the canonical Word Count MapReduce example, in which large files of plain text are the inputs and the application calculates the number of times each word appears in these texts. Our research job has 25 mappers and a sleep interval of 1 millisecond per processed word during map phase. The objective of this sleep is to turn its tasks into long time consuming activities, while our production job is a Word Count with the same number of mappers but with no sleep time. The sleep time is the simplest yet effective way for us to experiment with long running computations. The time a task is running is what will affect the scheduler decisions, the kind of the computation being performed does not matter. The second job is submitted 20 seconds after the first job starts. Charts in Figure 3 compare the total execution time for the production job after 10 runs of this experiment in both policies.

The preemption mechanism and the special treatment we give to the transition phases of the job produced an interesting result: our production job now runs 8.4 times faster in average. But this improvement comes with a cost: an impact in the research job execution time as shown in Figure 4. In average, with the Preemptive Fair Scheduler policy the research job is 4% slower. It had the execution time increased in 9 seconds in our experiments.

Figure 5 shows the cluster resources division between research and production jobs during their lifetime for the original Fair Scheduler Policy. It is possible to observe two large waiting times to execute the production job. The first one is at the beginning of the production job, just after it is submitted: it only starts to run after 1 minute waiting

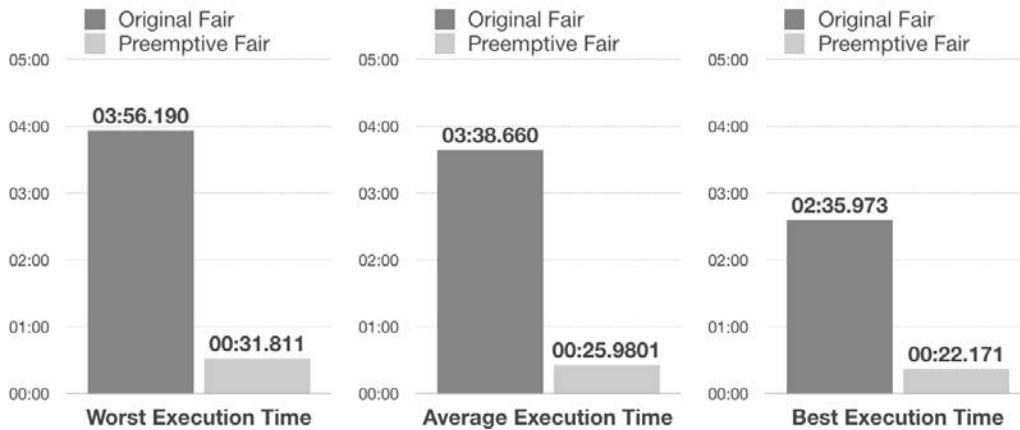


Figure 3. Production job execution times.

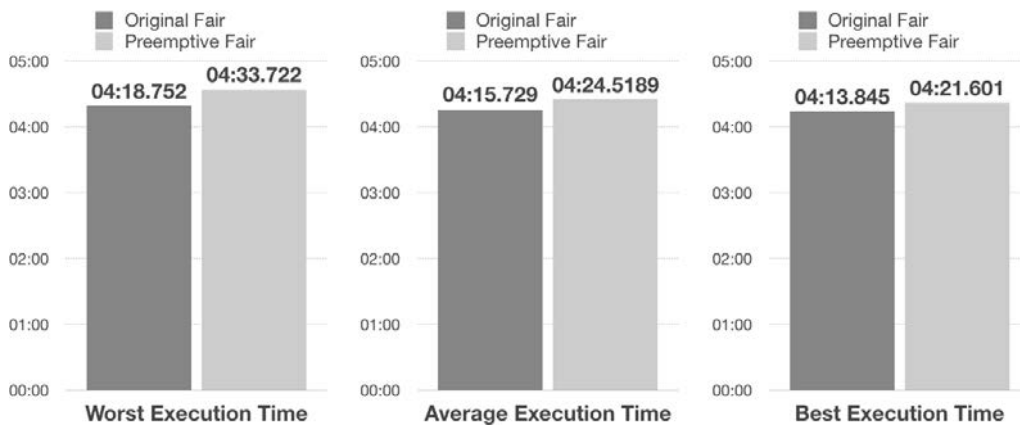


Figure 4. Research job execution times.

(01:20s in the chart, remembering that it was submitted 20 seconds after the research job). The second waiting happens between the end of the map phase and the start of shuffle phase: the production job has no tasks running between 01:41s and 02:42s.

The Preemptive Scheduler policy tries to address this two waiting periods as seen in Figure 6. So the production job gets its resources right away and there is almost no waiting after its submission. The amount of work lost because of the preemption is proportional to the fair share of the resources and the interval before the job submission in our experiments, so 100 seconds of computation is lost because we have 5 slots for each job and 20 seconds of interval. But, in a general case, the lost work could vary, since each task could be in a different stage of the computation.

Also, while running the scheduler with the Preemptive Fair Policy there is a period in which the cluster is underutilized. This happens during the shuffle and reduce phases of the production job, because the policy decides it is better to wait than to allocate resources for the greed research job. In Figure 6 check that between 00:20s and 00:30s in the chart there is a period in which the sum of running tasks for both jobs is less than the available resources (10 nodes). This corresponds to the period in which the amount of remaining work is still unknown by Disco because there is a job in a transition phase, as we described in section 3. Like Figure 5 shows, the research job with its long running tasks would take

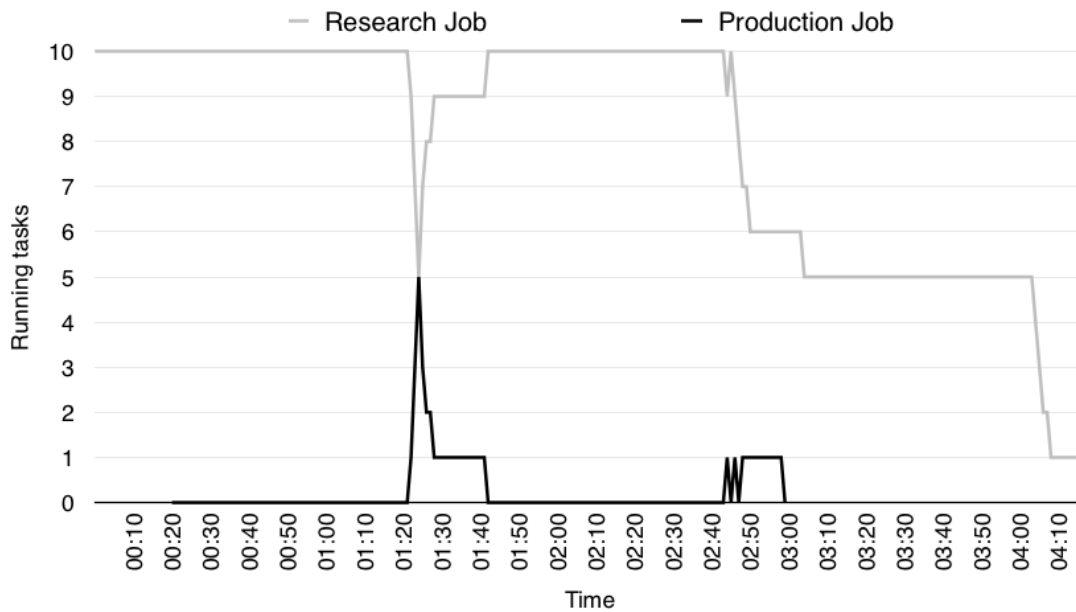


Figure 5. Cluster utilization for Disco configured with the original Fair scheduler policy.

every resource of the cluster and make the production job wait for a long time.

5. Related work

The closest work and one of the main inspirations for this is the Hadoop YARN preemption mechanism [Vavilapalli et al. 2013]. It works with the Hadoop's Capacity Scheduler, a feature enhanced version of the Fair Scheduler based on being Fair with the jobs between the same queue while each queue gets a percentage of the cluster's resources. Also, it preserves the work of the tasks to be preempted with a checkpointing service (which is something we intend to look as a future work). Finally, it tries to preempt tasks that made only little progress not any task like ours. These changes are still open issues in the Hadoop's source code changes control system (JIRA) [MAPREDUCE-5269 , MAPREDUCE-4584], so even they could be tested for the time the paper was published they are not available for the majority of Hadoop's users.

There is competing work regarding YARN preemption that was not accepted by Hadoop's community [Cho et al. 2013]. It is based in an interesting approach of eviction policies for both jobs and tasks, separating the logic necessary for the scheduler to chose the job that will be penalized with preemption and then the task from this job that will be preempted. This two level preemption policy is simple and powerful. Our work nowadays does not look in policies with this separation, our policy only choses the job to preempt a task from and do not chose a task wisely (we only chose randomly from the set of running tasks). The separation of policies could be a future work.

The idea of waiting a little before taking a scheduling decision have been explored for Hadoop [Zaharia et al. 2010]. This is similar to what we have done during job's transition phases. Differently than our scheduling decision that waits to have more information about the remaining work, it prefers to wait for some tasks to complete and use their vacant slots to schedule other task that will be benefitted by the data-locality

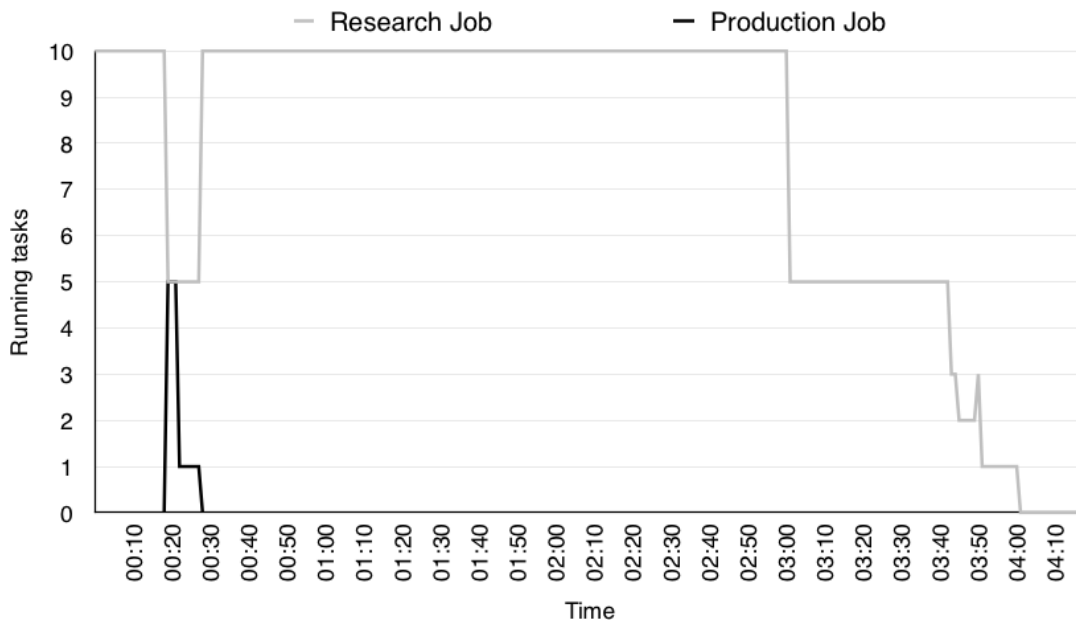


Figure 6. Cluster utilization for Disco configured with the Preemptive Fair Scheduler policy.

of the inputs it needs to use. Even Disco being location aware of its Data, we decided not to look at data-locality before preempting tasks at this moment. Another interesting information that improves Hadoop's jobs is an estimation of the end time of the jobs before deciding if a backup task can improve its execution time [Zaharia et al. 2008]. This is interesting for Hadoop, but Disco doesn't work with the concept of stragglers and backup tasks. Neither these works support preemption of tasks.

The problem we have with schedulers that greedily allocate resources for long running jobs also occurs in Hadoop schedulers [Wang et al. 2013]. Preemption of tasks in a work-preserving manner for the long running reduce tasks can help solve this problem after a proven not so good scheduling decision is made. Our work shows that this issue can happen in Disco too, and introduces both a preemption mechanism and a delay scheduling during transitory phases to improve production job's execution time. But, we underutilize the cluster during the reduce phase of the production job to avoid the greedily behavior of the research job.

Finally, the heterogeneous workload scheduling problem in which research and production jobs compete for cluster's resources and commonly the scheduling policies can't be fair by their defaults implementations exists for Hadoop too. The Global Preemption (GP) technique [Cheng et al. 2011] acts mainly in the tasks to be preempted decision of the scheduler. In Hadoop's default implementation the most recent tasks that overflow the running jobs share are preempted to give room for the tasks of the just submitted job. GP proposes not to look at tasks from each job separately, but to check every running task in the cluster and kill the newly launched ones, even if fairness can be lost during this process for a small amount of time. Our work prefers to stick with the default implementation since we doesn't look at tasks running time before preempting yet.

Every related work described in this section have been implemented for Hadoop, our work intends to take some of the good mechanisms Hadoop have to a less explored system: Disco. We also think Disco is a great project for MapReduce related research, since it is open source, simple, robust, mature, well architected and uses the Erlang Virtual Machine to address most of the distributed systems problems related to network communication, giving the researchers the freedom to focus on new features instead of configuration.

6. Conclusion

This paper describes a work performed in a robust, elegant and not so explored MapReduce system called Disco. Using the power of Erlang in a well architected pluggable scheduler policy environment, we have been able to address the heterogenous workload problem in this system by creating the Preemptive Fair Scheduler Policy. In order to implement this policy only a small quantity of lines of code were changed and added to Disco's source code. This way it can continue to be a lean alternative to Hadoop.

Usually, MapReduce jobs have different priorities and running times and can be divided into two main categories: research (long duration and with low priority) and production (short duration and with high priority). The default schedulers provided by Disco doesn't support preemption of tasks and are not able to address this kind of workload respecting the job's priority and staying fair. Using previous researches from the well known Hadoop framework as basis, we could adapt best practices related to preemption and to delayed scheduling decisions to a new Fair scheduler policy, achieving great results in our experimental environment. The application with more priority runs 8.4 times faster in average with the Preemptive Fair Scheduler Policy when compared to the default Fair Scheduler Policy. But this optimization comes with a small price: underutilization of the cluster during the production job final phases and the low priority research job has a 4% increase in its execution time.

For future work, we want to preserve the preempted tasks state with a checkpointing mechanism. We intend to avoid the cluster underutilization during the shuffle and reduce phases of the production job: in our experiments, the research job could have taken more resources than those the scheduler gave during this period. Another great improvement would be to only preempt when jobs with more priority are submitted, to achieve this we think in giving the clients the power to send jobs to Disco with a parameter indicating if preemption of tasks must be enabled for this submission. Finally, we want to take these ideas to Disco's community and work more closely to them in the most important features.

Acknowledgements

Special thanks to Digital Ocean [Digital Ocean Website] for the sponsorship for our experiments.

References

- Apache Hadoop Website. <http://hadoop.apache.org/>. Accessed: 2016-04-07.
- Armstrong, J. (2003). Making reliable distributed systems in the presence of software errors.

- Cheng, L., Zhang, Q., and Boutaba, R. (2011). Mitigating the negative impact of preemption on heterogeneous mapreduce workloads. In *Proceedings of the 7th International Conference on Network and Services Management, CNSM '11*, pages 189–197, Laxenburg, Austria, Austria. International Federation for Information Processing.
- Cho, B., Rahman, M., Chajed, T., Gupta, I., Abad, C., Roberts, N., and Lin, P. (2013). Natjam: Design and evaluation of eviction policies for supporting priorities and deadlines in mapreduce clusters. In *Proceedings of the 4th Annual Symposium on Cloud Computing, SOCC '13*, pages 6:1–6:17, New York, NY, USA. ACM.
- Cloc Project at Github.com. <https://github.com/AlDanial/cloc>. Accessed: 2016-04-07.
- Dean, J. and Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. In *Proceedings of the 6th Conference on Symposium on Operating Systems Design & Implementation - Volume 6, OSDI'04*, pages 10–10, Berkeley, CA, USA. USENIX Association.
- Digital Ocean Website. <https://www.digitalocean.com>. Accessed: 2016-04-07.
- Disco at Github.com. <https://github.com/discoproject/disco>. Accessed: 2016-04-07.
- Disco Distributed Filesystem Website. <http://disco.readthedocs.org/en/latest/howto/ddfs.html>. Accessed: 2016-04-07.
- Disco Fork at Github.com. <https://github.com/discoproject/disco>. Accessed: 2016-04-07.
- Ghemawat, S., Gobiuff, H., and Leung, S.-T. (2003). The Google File System. In *Proceedings of the nineteenth ACM symposium on Operating systems principles, SOSP '03*, pages 29–43, New York, NY, USA. ACM.
- Go Worker at Github.com. <https://github.com/discoproject/goworker>. Accessed: 2016-04-07.
- Haskell Worker at Github.com. https://github.com/zuzia/haskell_worker. Accessed: 2016-04-07.
- LFE Worker at Github.com. <https://github.com/oubiwann/lfe-disco>. Accessed: 2016-04-07.
- MAPREDUCE-4584. Umbrella: Preemption and restart of MapReduce tasks - JIRA. <https://issues.apache.org/jira/browse/MAPREDUCE-4584>. Accessed: 2016-04-07.
- MAPREDUCE-5269. Preemption of Reducer (and Shuffle) via checkpointing - JIRA. <https://issues.apache.org/jira/browse/MAPREDUCE-5269>. Accessed: 2016-04-07.
- Mundkur, P., Tuulos, V., and Flatow, J. (2011). Disco: A computing platform for large-scale data analytics. In *Proceedings of the 10th ACM SIGPLAN Workshop on Erlang, Erlang '11*, pages 84–89, New York, NY, USA. ACM.

- OCaml Worker at Github.com. <https://github.com/discoproject/odisco>. Accessed: 2016-04-07.
- Shvachko, K., Kuang, H., Radia, S., and Chansler, R. (2010). The Hadoop Distributed File System. In *Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on*, pages 1–10.
- Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., Saha, B., Curino, C., O’Malley, O., Radia, S., Reed, B., and Baldeschwieler, E. (2013). Apache Hadoop YARN: Yet Another Resource Negotiator. In *Proceedings of the 4th Annual Symposium on Cloud Computing, SOCC ’13*, pages 5:1–5:16, New York, NY, USA. ACM.
- Wang, Y., Tan, J., Yu, W., Zhang, L., Meng, X., and Li, X. (2013). Preemptive reduced task scheduling for fair and fast job completion. In *Proceedings of the 10th International Conference on Autonomic Computing (ICAC 13)*, pages 279–289, San Jose, CA. USENIX.
- Zaharia, M., Borthakur, D., Sen Sarma, J., Elmeleegy, K., Shenker, S., and Stoica, I. (2010). Delay scheduling: A simple technique for achieving locality and fairness in cluster scheduling. In *Proceedings of the 5th European Conference on Computer Systems, EuroSys ’10*, pages 265–278, New York, NY, USA. ACM.
- Zaharia, M., Konwinski, A., Joseph, A. D., Katz, R., and Stoica, I. (2008). Improving mapreduce performance in heterogeneous environments. In *Proceedings of the 8th USENIX Conference on Operating Systems Design and Implementation, OSDI’08*, pages 29–42, Berkeley, CA, USA. USENIX Association.

LoadTime is Money! Avaliação da Performance Web de E-Commerces no Brasil

Cristiano Politowski¹, Gabriel Freytag¹, Vinícius Maran², Lisandra Fontoura¹ *

¹ Programa de Pós-Graduação em Informática (PPGI)
Universidade Federal de Santa Maria (UFSM) — Santa Maria — RS — Brasil

²Coordenadoria Acadêmica
Universidade Federal de Santa Maria (UFSM) — Cachoeira do Sul — RS — Brasil

{cpolitowski,gfreytag,viniciusm,lisandra}@inf.ufsm.br

Abstract. *Performance of web applications is an underexplored topic by developers and academic work in Brazil. However, the access speed to pages is directly related to the user experience and, therefore, reflects directly the revenue from E-Commerces. This research presents an evaluation of performance of the 29 most accessed virtual stores in Brazil according to the Alexa ranking. There is a wide difference between the best and the worst placed. Moreover, the numbers show that the pages of Brazilian e-commerces, with few exceptions, are worse than the amounts considered appropriate by experts and / or global averages.*

Resumo. *Performance de aplicações web é um tópico pouco explorado por desenvolvedores e em trabalhos acadêmicos no Brasil. No entanto, a velocidade de acesso às páginas está diretamente relacionada com a experiência do usuário e, assim sendo, reflete diretamente na receita dos e-commerces. Nesta pesquisa é apresentada uma avaliação da performance web das 29 lojas virtuais mais acessadas no Brasil segundo o ranking da Alexa. Por meio do estudo foi possível ter um panorama da performance geral das páginas web. Há uma larga diferença entre o melhor e o pior colocado. Além disso, os números mostram que as páginas dos e-commerces brasileiros, com poucas exceções, estão piores que os valores considerados adequados por especialistas e/ou médias mundiais.*

1. Introdução

Comércios eletrônicos, lojas virtuais ou *e-commerces* fazem parte do cotidiano de consumidores de todo o mundo. Grandes empresas como a *Amazon* e *AliExpress* veem sua receita aumentar a cada ano por meio da venda de grandes quantidades de produtos à preços baixos. Pequenas e médias empresas também tem espaço nesse mercado em constante crescimento. Segundo *eMarketer* [eMarketer 2016], em 2015, o total arrecadado

*O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico — Brasil.

por empresas desse tipo foi de aproximadamente U\$ 1.600.000.000.000 (um trilhão e seiscentos bilhões de dólares americanos) com previsão de aumento de mais de 110% até 2019. Os *e-commerces* no Brasil, apesar de apresentarem números mais modestos, tiveram uma receita em 2015 de cerca de U\$ 19.600.000.000 (dezenove bilhões e seiscentos milhões de dólares americanos) com uma previsão de aumento de cerca de 50% até 2019, ou seja, mais de trinta bilhões¹. Para efeito de comparação, este valor corresponde, hoje, ao Produto Interno Bruto (PIB) do Paraguai².

A competitividade do mercado obrigou donos de lojas virtuais à investir em diferenciais para seus negócios. Como consequência, a complexidade dos *e-commerces*, ou seja, o número de requisições e o tamanho total das páginas, aumentou. Segundo os dados de *HTTPArchive* [HTTPArchive 2016], no mês de março, o tamanho médio de uma página na Internet teve crescimento de 1311 KB em 2013 para 1703 KB em 2014, 2008 KB em 2015 e este ano chegou à 2281 KB, um aumento de mais de 74%.

Contrário ao crescimento das páginas, usuários não aceitam mais atrasos na compra de produtos via *e-commerces* [IPNews 2015]. Estudos realizados pela indústria e livros técnicos de profissionais especializados como Steve Souders, Ilya Grigorik e Patrick Meenan [Souders 2016, Grigorik 2016, Meenan 2016] comprovam esse fato. Segundo um estudo realizado pelo *Aberdeen Group* em 2008 [Aberdeen Group 2008], um segundo de atraso no tempo de resposta de aplicações *web* pode impactar na satisfação do usuário em 16%, nas páginas visualizadas em 11% e nas conversões (ato de compra do usuário) em 7%. De acordo com a empresa *Akamai* [Akamai 2002], a instabilidade e imprevisibilidade da performance de uma aplicação pode impactar negativamente nos clientes, ou seja, na escolha do *e-commerce* na qual realizarão suas compras.

Segundo Everts [Everts 2015], em 2006 usuários de *e-commerces* esperavam que uma página carregasse em 4 segundos ou menos. Atualmente, no entanto, cerca de metade dos usuários esperam um tempo de carregamento de 2 segundos ou menos [Forrester Consulting 2009] e 18% esperam que a página carregue instantaneamente. O mesmo autor ainda cita alguns casos onde a performance influenciou diretamente na receita das empresas:

- Cada 1 segundo retirado do tempo de carregamento da página equivaleu a um aumento de 2% na taxa de conversão de clientes em *Walmart.com*;
- *Staples.com* retirou 1 segundo do tempo de carregamento da página, aumentando a taxa de conversão em 10%;
- *AutoAnything.com* cortou o tempo de carregamento da página pela metade, aumentando a taxa de conversão em 9%;
- A plataforma de captação de fundos da *Mozilla* retirou 2.2 segundos do tempo de carregamento da página, aumentando o número de downloads em 15,4%;

¹Dados retirados de [Statista](#) sendo a pesquisa realizada entre 2014 e 2015 por [eMarketer](#).

²Dados retirados dos sites: [FMI](#), [World Bank](#) e [Nações Unidas](#).

- *Intuit.com* cortou o tempo de carregamento pela metade, aumentando a conversão em 14%.

Dada esta situação, como podemos otimizar um *website*? Souders [Souders 2007] apresenta a *Performance Golden Rule*, mostrando que é muito mais viável concentrar os esforços de otimização nos elementos “do lado do cliente” (*client-side* ou *front-end*) pois correspondem a 80–90% dos componentes da aplicação:

“Only 10–20% of the end user response time is spent downloading the HTML document. The other 80–90% is spent downloading all the components in the page.” — Steve Souders.

Com a previsão de crescimento da receita dos *e-commerces* no Brasil, com o crescente aumento do tamanho e do número de requisições das páginas *web* e pelo fato do *front-end* corresponder à grande parte de uma página *web*, duas perguntas centrais devem ser respondidas para se ter um panorama da performance dos *e-commerces* no Brasil:

1. Qual o tamanho médio das páginas das lojas virtuais brasileiras? Como estão em comparação com a média mundial? Qual a média de *assets* por página?
2. Quais páginas mais se preocupam com o usuário, ou seja, qual o *ranking* das métricas relacionadas a experiência do usuário (*Time To Start Render*, *Speed Index*, etc)?

Portanto, o objetivo deste trabalho é capturar e analisar dados de performance *web* das lojas virtuais mais acessadas no Brasil. Um panorama dos principais *e-commerces* é importante e se faz necessário por ser um ponto de partida para uma estratégia de otimização. Além disso, novas ideias e *insights* surgem por meio da análise de performance e podem auxiliar pesquisadores de outras áreas como *User experience (UX)* e *Search Engine Optimization (SEO)*.

O artigo está estruturado da seguinte forma: na seção 2 são apresentados os conceitos e termos necessários para compreender o trabalho. Na seção 3 é mostrado o método utilizado na pesquisa. Na seção 4 os dados obtidos por meio da pesquisa são analisados. Na seção 5 são apresentados os trabalhos relacionados. Por fim, na seção 6 são descritas as considerações finais.

2. Referencial Teórico

A visualização de uma página *web* é composta de uma requisição para cada *asset* (arquivo). Nos primórdios da internet, o arquivo HTML continha quase a totalidade dos bytes transferidos. Hoje, o cenário mudou, e o tamanho do arquivo HTML representa uma porção muito pequena do total de bytes transferidos.

Durante muitos anos a performance *web* foi — e continua sendo, porém, em menor escala — associada à otimização *backend*. No entanto, de acordo com Souders

[Souders 2007], se formos capazes de cortar ao meio o tempo de resposta do *backend*, o tempo de resposta do usuário final seria reduzido apenas 5–10%. Porém, se nós reduzíssemos a performance do *front end* ao meio, o tempo de resposta total seria reduzido de 40–45%. Isso mostra a importância da otimização da performance do *front end*.

2.1. Métricas de Performance em Sistemas Web

Quando se fala em performance *web*, o primeiro detalhe a considerar são quais dados monitorar, ou seja, por meio de quais métricas a performance será analisada. Para tanto, existem várias métricas com diferentes propósitos que, inclusive, vão evoluindo ou sendo desconsideradas conforme mais estudos são realizados. De acordo com Viscomi [Rick Viscomi 2015], entre as métricas mais usadas encontram-se:

- TTFB (*Time To First Byte* — tempo até o primeiro byte): é o tempo do início da requisição até o navegador receber o primeiro byte da resposta;
- TTSR (*Time to Start Render* — tempo até o início da exibição na tela): é o tempo do início da requisição até o navegador exibir o primeiro pixel do conteúdo na tela;
- *Speed Index* (índice de velocidade): é uma representação do progresso visual de uma tela;
- *Load Time* (tempo de carregamento): é o tempo de carregamento total da página;
- *Requests* (requisições): o número de requisições HTTP antes do evento;
- *Connections* (conexões): é a quantidade de conexões realizada com servidores;
- *Bytes In* (bytes de entrada): é o número total de bytes recebidos no teste;
- *DOM Elements* (número de elementos contidos no DOM): é o número de elementos no documento.

2.2. Performance Percebida

Métricas por si só não garantem que um *website* pareça rápido para o usuário. Hoje, engenheiros de performance estão dando uma maior ênfase na performance percebida do que em números por si só. De acordo com a escala citada por Nielsen [Nielsen 2009], ações que levam até 0.1 segundos para serem concretizadas passam a impressão de tempo real ao usuário. Entre 0.1 e 1 segundo, o usuário, caso esteja em plena atenção, percebe um atraso mínimo mas mantém o controle da aplicação navegando livremente. Após 1 segundo, o usuário começa a ficar impaciente, o que vai aumentando conforme o tempo passa. 10 segundos marca o tempo de quebra do fluxo da navegação, normalmente fazendo o usuário abandonar o página. Grigorik [Grigorik 2013] vai mais além e define uma “regra de ouro” (*rule of thumb*) para performance *web*: “*Renderize suas páginas, ou pelo menos, forneça um feedback visual em menos de 250 milissegundos para manter o usuário engajado.*” (tradução livre). Ou seja, a principal preocupação agora é dar uma resposta o mais cedo possível para o usuário, idealmente em até 250 milissegundos.

Como as métricas de rede não conseguem traduzir com precisão a performance percebida pelos usuários durante a navegação em *websites*, novas formas de medição foram criadas. Por exemplo, *Time to Start Render* é a métrica que mede o tempo decorrido

da requisição até o momento que o primeiro pixel é mostrado na tela, ou seja, é a primeira indicação percebida pelo usuário de que a página está sendo renderizada, o que não significa que a página está totalmente carregada. *Above-The-Fold Time* é a métrica que mede o tempo entre o clique no *link* de um *website* e o carregamento da parte visível da página, ou seja, a parte inicial do *website* que não necessita de *scroll* para ser visualizada.

Speed Index é uma métrica criada pela ferramenta WebPagetest. De acordo com [Rick Viscomi 2015], o índice de velocidade de uma página é derivado do progresso visual de cada tela, até ser totalmente carregada, sendo que uma página que exibe mais ao usuário e mais rapidamente tem um índice de velocidade inferior, ou melhor do que uma página que é mais lenta para exibir o conteúdo.

3. Metodologia da Pesquisa e Avaliação

Inicialmente definiu-se quais seriam os *e-commerces* analisados. Como base, utilizou-se o *ranking* dos 500 *websites* mais acessados no Brasil, segundo o site Alexa³ no mês de março. Desta lista, filtrou-se apenas os *websites* de comércio eletrônico, resultando em uma lista de 30 *e-commerces* com o melhor posicionamento nesse *ranking* (Tabela 1).

Tabela 1. Lista dos e-commerces analisados.

Loja	URL	Posição	Loja	URL	Posição
Americanas	www.americanas.com.br	23	Peixe Urbano	www.peixeurbano.com.br	248
Submarino	www.submarino.com.br	40	Livraria Cultura	www.livrariacultura.com.br	256
Walmart	www.walmart.com.br	54	Posthaus	www.posthaus.com.br	274
Extra	www.extra.com.br	55	Mobly	www.mobly.com.br	292
Casas Bahia	www.casasbahia.com.br	56	Kalunga	www.kalunga.com.br	319
Ponto Frio	www.pontofrio.com.br	63	Passarela	www.passarela.com.br	325
Saraiva	www.saraiva.com.br	83	Sou Barato	www.soubarato.com.br	339
Magazine Luiza	www.magazineluiza.com.br	87	Renner	www.lojasrenner.com.br	350
Dafiti	www.dafiti.com.br	89	Privalia	www.privalia.com.br	358
Ultrafarma	www.ultrafarma.com.br	92	oBoticario	www.boticario.com.br	386
Shoptime	www.shoptime.com.br	132	Marisa	www.marisa.com.br	421
Kabum	www.kabum.com.br	143	Polishop	www.polishop.com.br	423
Ricardo Eletro	www.ricardoeleto.com.br	173	Sephora	www.sephora.com.br	427
Kanui	www.kanui.com.br	217	Tricae	www.tricae.com.br	447
Natura	www.natura.com.br	232	OQVestir	www.oqvestir.com.br	482

Foram definidos quatro cenários de testes diferentes para cada amostra na lista. Todos os testes fizeram uso de um dispositivo *desktop* localizado no Brasil em uma instância da *Amazon EC2* com o navegador *Google Chrome*. Nos cenários 1 e 3 não houve a utilização de cache do navegador, simulando o primeiro acesso do usuário. Nos cenários 1 e 2 foi utilizado uma rede cabeada com velocidade de 5 megabits de download e 1 megabit de upload por segundo com RTT (*Round Trip Time* — demora para uma informação chegar ao destinatário e o remetente receber sua confirmação) de 28 milissegundos enquanto que nos cenários 3 e 4 a velocidade da rede simula um acesso utilizando uma rede

³<http://www.alexa.com>

Tabela 2. Lista da configuração dos testes realizados para cada amostra.

Cenário de teste	Dispositivo	Local do Teste	Navegador	Cache	Tipo de Conexão
1	Desktop	Brazil — EC2	Chrome	Não	Cable (5/1 Mbps 28ms RTT)
2	Desktop	Brazil — EC2	Chrome	Sim	Cable (5/1 Mbps 28ms RTT)
3	Desktop	Brazil — EC2	Chrome	Não	Mobile 3G (1.6 Mbps/768 Kbps 300ms RTT)
4	Desktop	Brazil — EC2	Chrome	Sim	Mobile 3G (1.6 Mbps/768 Kbps 300ms RTT)

3G com 1.6 megabits de download e 768 megabits de upload por segundo respectivamente e RTT de 300 milissegundos. A velocidade média média a cabo no Brasil é de 3,6 megabits⁴ e 3G é de 2 megabits⁵ (3G em São Paulo).

Os testes foram realizados no dia 08 de março de 2016. A página principal de cada *e-commerce* listado na Tabela 1 foi testado uma vez (sem descartes — *outliers*) nos quatro cenários descritos na Tabela 2, sempre com a mesma configuração de rede para cada *website* em cada cenário. Em ambos os testes nenhum usuário estava logado nas páginas web, visto que os logins normalmente expiram em um curto espaço de tempo e quando um usuário real acessa um *website* (principalmente no primeiro acesso) não está logado. Foram armazenados os *dashrefs* referentes ao teste de cada *e-commerce* nos quatro cenários para posterior análise do desempenho.

A ferramenta escolhida para os testes de performance foi a WebPagetest⁶, uma ferramenta *online* para testes de desempenho de páginas *web*. WebPageTest permite gerar resultados e cenários de diversas formas, inclusive vídeos com o processo de renderização das páginas e *filmstrips* (linha do tempo do processo de renderização da página com *prints* da tela acima da borda a cada 0,5 segundos). A ferramenta ainda permite a execução de dois testes de um mesmo *website* em uma única execução por meio da seleção da opção *First View and Repeat View* nas configurações avançadas. Dessa forma, os testes 1 e 2 assim como os testes 3 e 4 descritos na Tabela 2 foram realizados em duas execuções e, portanto, as métricas de cada teste estão presentes nos mesmos arquivos. WebPagetest também permite salvar um resumo das métricas dos testes, as métricas detalhadas de cada requisição de cada teste e também um arquivo com todas as informações das transações HTTP realizadas pela página *web*. Em cada teste foram armazenados os seguintes arquivos: o *Waterfall* (visão do carregamento sequencial do *website* em forma de requisições fluindo constantemente para baixo), o *Waterfall* das conexões realizadas com cada servidor de conteúdo, a captura de tela, o *filmstrip* e o vídeo de todo o carregamento dos *websites*.

⁴Velocidade média a cabo no Brasil: <http://goo.gl/B1CLWT>

⁵Velocidade média 3G no Brasil: <http://goo.gl/8Kk4Mk>

⁶<http://www.webpagetest.org>

4. Resultados e Discussão

Nas Tabela 3 e 4 são apresentadas as métricas e resultados considerados mais importantes, retirados dos testes 1 e 3⁷. Na tabela, os dados estão separados por cores, sendo azul e vermelho os valores abaixo e acima da média da amostra respectivamente. Ainda, na tabela constam valores mínimos (MIN) e máximos (MAX) de cada métrica, ambos em destaque com fundo vermelho e verde. Os testes foram realizados em 29 dos *e-commerces* listados, isto porque os testes não puderam ser efetuados no domínio <http://www.ricardoeletro.com.br/> pois o servidor da loja rejeitou o acesso efetuado pelo servidor do *WebPageTest*.

Tabela 3. Resultados do teste 1 (Cable (5/1 Mbps 28ms RTT), no cache).

Loja	TTFB	TTSR	SpeedIndex	LoadTime	Request	Connections	Bytes In	DOM Elem.
Americanas	156	2185	13104	14212	211	66	4430441	3271
Boticario	417	6784	17625	21510	496	193	4627767	6549
Casas Bahia	151	3194	3971	8509	221	110	2091314	3392
Dafiti	129	1895	4275	13612	297	154	2168881	1986
Extra	255	3789	9496	13005	272	128	2443574	1132
Kabum	364	6604	9426	11533	240	80	2565895	5809
Kalunga	586	7586	10303	10295	146	33	3782604	1609
Kanui	94	3100	5611	7822	208	92	3743222	1578
Livraria Cultura	276	3186	5570	10105	154	69	2561703	2216
Renner	1300	3888	8149	10006	287	140	2322744	1253
Magazine Luiza	554	6198	8974	16910	166	81	3452265	2994
Marisa	222	2986	6266	17945	357	179	4183569	2238
Mobly	256	2090	4156	2518	233	112	2031785	1807
Natura	255	4091	6286	21881	347	48	4818540	4373
OQVestir	354	7298	17459	19957	249	83	2760183	2937
Passarela	415	2294	4304	20734	367	107	3463009	1290
Peixe Urbano	2498	4789	10277	13844	158	98	2762172	1544
Polishop	381	8287	15681	25811	565	238	11167147	1675
Ponto Frio	163	7196	13059	11714	204	105	2316903	1821
Posthaus	211	3185	5629	9585	201	89	1520350	1422
Privalia	1065	3887	19512	12077	163	51	4115997	2304
Saraiva	147	5086	10074	13453	304	123	4539035	1232
Sephora	248	3880	9715	16223	430	144	3185218	1359
Shoptime	292	3100	18742	13823	352	64	12744964	1386
Sou Barato	139	2584	6801	10645	167	46	3535307	4253
Submarino	143	3485	17322	19111	271	87	6181333	2133
Tricae	286	2291	7328	7877	124	32	3434053	1343
Ultrafarma	1102	4081	13430	16866	112	25	3563155	4123
Walmart	571	2486	5937	11663	166	67	1642101	2213
MÉDIA	444,96	4250,68	10090,89	13985,11	260,79	99,18	3946897,50	2465,32
MIN	94	1895	3971	2518	112	25	1520350	1132
MAX	2498	8287	19512	25811	565	238	12744964	6549

Entre os dados do teste 1, os melhores resultados estão bem distribuídos entre as amostras pois cada loja se sobressai em algum aspecto. No entanto, vale ressaltar que

⁷Todos os demais resultados podem ser visualizados na página do projeto em <http://inscale.github.io/perf-e-commerces-br/>.

Tabela 4. Resultados do teste 3 (Mobile 3G (1.6 Mbps/768 Kbps 300ms RTT), no cache).

Loja	TTFB	TTSR	SpeedIndex	LoadTime	Request	Connections	Bytes In	DOM Elem.
Americanas	981	7794	29805	33181	212	70	4459254	1986
Boticario	1202	11692	24187	45149	492	200	4598730	1544
Casas Bahia	1085	10194	12963	19459	223	113	2091289	1578
Dafiti	931	3096	8984	36756	306	163	2202368	1290
Extra	957	9289	15589	22148	288	136	2515380	6548
Kabum	949	4484	14073	20267	231	74	2545747	2216
Kalunga	1344	22088	27584	33507	146	33	3778838	3270
Kanui	1820	4577	8413	23258	208	92	3743231	5809
Livraria Cultura	1132	5201	8310	20380	150	70	2381045	1609
Renner	2252	6487	15204	20127	292	143	2325531	2995
Magazine Luiza	1402	13584	16217	31372	153	78	3322118	4379
Marisa	947	5391	6083	29411	323	171	3707094	1810
Mobly	969	4191	9127	7294	272	133	2045635	1253
Natura	1073	9488	11449	36269	347	50	4821433	2240
OQVestir	2295	14601	28540	35281	255	86	2729064	3392
Passarela	1344	4292	5294	34175	367	110	3455837	2937
Peixe Urbano	2845	5284	13703	23402	201	119	2821609	1132
Polishop	1176	5888	27784	50257	585	239	12355696	4122
Ponto Frio	980	12887	28284	22811	205	107	2239099	1673
Posthaus	1011	7184	9552	19773	226	91	1542579	1335
Privalia	3650	10187	37076	29110	165	60	4126116	2133
Saraiva	930	6191	15354	25718	288	122	4533687	1359
Sephora	951	5091	11248	28931	429	148	3174608	1422
Shoptime	966	8184	33679	27978	353	67	12753517	2305
Sou Barato	969	4986	11296	25643	171	49	3539113	2217
Submarino	950	7827	31996	53441	314	103	6379051	1375
Tricae	1988	7684	25358	27759	124	32	3434141	1232
Ultrafarma	1651	8884	18279	24439	112	25	3563296	1819
Walmart	2035	5082	8341	18072	151	71	1525921	4253
MÉDIA	1383,93	8097,36	18051,11	28832,00	265,64	103,00	3970896,64	2392,14
MIN	930	3096	5294	7294	112	25	1542579	1132
MAX	3650	22088	37076	53441	585	239	12753517	6548

Casas Bahia apresentou o melhor índice de velocidade, seguido por *Mobly* e *Dafiti*. No entanto, a loja *Polishop* obteve os piores resultados em quatro métricas: *TTSR*, *Load-Time*, *Requests* e *Connections*. Se considerarmos os valores abaixo das médias, *Tricae* com 7, *Shoptime* com 6 e *Polishop* com 5 são as piores colocadas. No teste 3 os valores estão ainda mais distribuídos. O melhor índice de velocidade aqui fica com a *Loja Passarela*, seguida de *Marisa* e *Livraria Cultura*. *Boticario* e *Polishop* com 6 juntamente com *OQVestir* e *Privalia* com 5 são as lojas com maior número de valores abaixo das médias.

Também procuramos correlações entre os valores, ou seja, se o aumento/diminuição dos valores de uma métrica possuíam relação direta com outra métrica (Tabela 5). Entre as correlações mais relevantes (consideramos valores entre -0,99 a -0,50 e 0,560 a 0,99) estão *SpeedIndex* com *TTSR*, visto que ambos se referem à renderização de conteúdo na tela e de *SpeedIndex* com *Bytes In*, pois a velocidade de exibição do conteúdo depende do download do conteúdo e consequentemente tem relação

com o tamanho dos conteúdos. Há também correlações de *LoadTime* com *Requests* e de *LoadTime* com *Bytes In*, pois o tempo de carregamento depende da quantidade de requisições realizadas e do tamanho destas requisições. Além disso, há correlação de *Requests* com *Connections* e de *Requests* com *Bytes In*, visto que o número de conexões varia de acordo com o número de requisições, assim como o tamanho total varia de acordo com o número de requisições.

Tabela 5. Tabela de correlações entre as métricas.

Correlações	TTFB	TTSR	SpeedIndex	LoadTime	Request	Connections	Bytes In	DOM Elem.
TTFB	1,00	0,12	0,26	-0,09	-0,34	-0,22	-0,15	0,11
TTSR	0,12	1,00	0,55	0,23	-0,20	-0,29	0,01	0,16
SpeedIndex	0,26	0,55	1,00	0,47	0,05	-0,14	0,52	-0,05
LoadTime	-0,09	0,23	0,47	1,00	0,55	0,30	0,55	0,00
Request	-0,34	-0,20	0,05	0,55	1,00	0,80	0,55	0,01
Connections	-0,22	-0,29	-0,14	0,30	0,80	1,00	0,21	0,04
Bytes In	-0,15	0,01	0,52	0,55	0,55	0,21	1,00	0,09
DOM Elem.	0,11	0,16	-0,05	0,00	0,01	0,04	0,09	1,00

Traçando um panorama geral dos resultados do teste 1, temos que, a página de uma loja virtual brasileira possui, em média: 260 requisições em sua página inicial, 103 conexões são realizadas com o servidor da aplicação ou CDNs, o número total de elementos carregados pelo navegador cliente é de cerca de 2392, totalizando 3,94 Megabytes de dados, sendo que o primeiro byte chega ao navegador em 444 milissegundos e a renderização começa depois de mais de 4 segundos de espera. Já no teste 3, com rede móvel 3G, os valores permanecem praticamente os mesmos, com exceção do *time to first byte* de 1383 milissegundos (aumento de três vezes) e o *time to start render* de mais de 8 segundos (duas vezes mais lento).

Se considerarmos os piores casos, a situação fica preocupante. Por exemplo, a página inicial da loja Kalunga leva mais de 22 segundos para mostrar algo na tela em uma conexão 3G. Isso significa que o usuário ficará todo esse tempo vendo uma tela em branco. Provavelmente neste intervalo já terá desistido da compra ou até mesmo pensado que o *website* está *offline*. Este caso pode ser melhor compreendido visualizando o *filmstrip* da Figura 1. Cada retângulo branco é meio segundo em que o usuário está esperando pelo conteúdo.

5. Trabalhos Relacionados

A empresa RADWARE realiza anualmente trabalhos de pesquisa sobre a performance de *websites* de várias áreas como lojas virtuais, sites de esportes, hotéis, etc. Por exemplo, em 2015 [RADWARE 2015] foram analisados os cem primeiros *e-commerces* listados no *ranking* do site Alexa.com por meio de algumas métricas como TTI e LoadTime e também outros detalhes como a composição das páginas em relação aos seus *assets*. Seus achados mostram que somente 12% dos sites analisados mostravam algum conteúdo abaixo



Figura 1. Filmstrip da página inicial da loja Kalunga testada em uma rede 3G.

da barreira dos 3 segundos; o tamanho e complexidade das páginas estão diretamente ligados com tempo de carregamento (*loadtime*); imagens não estão sendo devidamente otimizadas.

Recentemente a mesma empresa divulgou outro estudo [RADWARE 2016] mas dessa vez englobando também outras áreas. Seus resultados mostraram que lojas virtuais tiveram a melhor média de TTI (3.1 segundos, algo bem superior ao estudo prévio); o tamanho médio das páginas é 1,4 MBs, com 97 requisições e um tempo de carregamento total de 7,9 segundos; as páginas oscilaram entre 276 KB a 9 MB e 34% dos sites analisados não usavam uma compressão de imagens adequada;

No estudo apresentado neste trabalho, no entanto, o foco foi uma análise detalhada das lojas virtuais mais acessadas do Brasil, abrangendo um gama maior de métricas e uma análise da renderização das páginas considerando conexões comuns em *desktops* e redes móveis. Além disso, o trabalho também trata de detalhes relacionados a otimização *front end*.

6. Considerações Finais

Este trabalho apresentou uma avaliação da performance *web* dos 29 *e-commerces* mais acessados no Brasil segundo o ranking da Alexa. Analisando os melhores resultados de cada métrica dos *e-commerces* testados é possível perceber que poucos estão próximos das métricas ideais.

Por exemplo, de acordo com as especificações da ferramenta de medição de desempenho *web PageSpeed Insights* [GoogleDevelopers 2016], o TTFB ideal é igual ou menor que 200 ms. No Teste 1, apenas oito amostras (cerca de 27%) ficaram abaixo desse valor.

Outras métricas, no entanto, tiveram resultados bem inferiores dos desejados. Para

a métrica TTSR, o valor ideal é abaixo de 1,2 segundos [WebSiteOptimization 2016], no entanto, nenhum *e-commerce* se aproximou desse valor, sendo o mais próximo 3,09 segundos da loja Dafiti no Teste 1. Outro exemplo é a métrica *SpeedIndex*, que segundo Paul Irish [Irish 2016], o valor ideal é abaixo de 1000, mas o melhor resultado alcançado foi de 3971 pela loja Americanas.

Comparando o tamanho médio das páginas brasileiras (3854 kilobytes) com o tamanho médio mundial no mesmo período (2281 kilobytes) [HTTPArchive 2016], é possível observar que as páginas de comércio eletrônico brasileiras estão cerca de 68% maiores. Esta diferença é duas vezes maior quando comparada com a média mundial no setor de *e-commerces* [RADWARE 2015], cujo valor é de cerca de 1905 kilobytes por página. Além disso, a quantidade de elementos no DOM se mostraram altas — principalmente o *website* da loja *Shoptime* com 12 megabytes. A média mundial de elementos presentes no DOM, no mesmo período, era de 905, e o resultado médio do Teste 1 foi de 2465, quase três vezes maior.

Otimização da performance de aplicações *web* ainda é um tópico pouco discutido na indústria e ainda menos no meio acadêmico. Os dados apresentados aqui reforçam a necessidade de maior apreço pelo assunto. Mais pesquisas, técnicas e ferramentas precisam ser feitas para propiciar melhores experiências para usuários. Como trabalho futuro há previsão de extensão dos testes também para dispositivos móveis, abrindo possibilidade de outras análises como responsividade da página e comparação entre páginas *mobile* e páginas únicas. Além disso, a criação e execução de uma estratégia de otimização juntamente com a comparação do antes e depois são os próximos passos da pesquisa. Por fim, ainda é possível realizar um trabalho complementar, com viés mais qualitativo, visando capturar a percepção dos usuários finais.

Referências

- [Aberdeen Group 2008] Aberdeen Group (2008). The Performance of Web Applications: Customers Are Won or Lost in One Second. Technical Report November.
- [Akamai 2002] Akamai (2002). Why Performance Matters. Technical report.
- [eMarketer 2016] eMarketer (2016). emarketer. Disponível online em <http://www.emarketer.com>.
- [Everts 2015] Everts, T. (2015). *Time is Money The Business Value of Web Performance*.
- [Forrester Consulting 2009] Forrester Consulting (2009). *eCommerce Web Site Performance Today*. Technical report.
- [GoogleDevelopers 2016] GoogleDevelopers (2016). Melhorar o tempo de resposta do servidor. Disponível online em <https://goo.gl/1ZXawT>.
- [Grigorik 2013] Grigorik, I. (2013). High performance networking in google chrome. Disponível online em <https://goo.gl/37RctB>.

- [Grigorik 2016] Grigorik, I. (2016). Ilya grigorik. Disponível online em <https://www.igvita.com/>.
- [HTTPArchive 2016] HTTPArchive (2016). Http archive. Disponível online em <http://goo.gl/82e14w>.
- [IPNews 2015] IPNews (2015). Paciência dos brasileiros ao utilizarem um aplicativo dura três segundos. Disponível online em <http://goo.gl/sB6Vc0>.
- [Irish 2016] Irish, P. (2016). Fast enough. Disponível online em <https://goo.gl/4FBPZv>.
- [Meenan 2016] Meenan, P. (2016). Patrick meenan. Disponível online em <http://blog.patrickmeenan.com/>.
- [Nielsen 2009] Nielsen, J. (2009). Powers of 10: Time scales in user experience. Disponível online em <https://goo.gl/zJoUc9>.
- [RADWARE 2015] RADWARE (2015). STATE OF THE UNION — SUMMER 2015 ECOMMERCE PAGE SPEED & WEB PERFORMANCE Content and Speed: The Magic Formula. Technical report.
- [RADWARE 2016] RADWARE (2016). MULTI-INDUSTRY WEB PERFORMANCE 2016 STATE OF THE UNION (desktop edition) Time is money when it comes to microwaves. Technical report.
- [Rick Viscomi 2015] Rick Viscomi, Andy Davies, M. D. (2015). Using webpagetest: Web performance testing for novices and power users. Technical report.
- [Souders 2007] Souders, S. (2007). *High Performance Web Sites: Essential Knowledge for Frontend Engineers*.
- [Souders 2016] Souders, S. (2016). Steve souders. Disponível online em <https://stevesouders.com/>.
- [WebSiteOptimization 2016] WebSiteOptimization (2016). Optimize start render time. Disponível online em <http://goo.gl/s4RoFm>.

Usando a Elasticidade de Recursos em Nuvem para Aumentar o Desempenho de Aplicações Pipeline

Rodrigo da Rosa Righi¹, Mateus Aubin¹, Cristiano A. da Costa¹, Guilherme Galante²

¹Programa Pós-Graduação em Computação Aplicada – Unisinos – Brasil
{rrrighi, maubin, cac}@unisinos.br

²Centro de Ciências Exatas e Tecnológicas – Unioeste – Brasil
guilherme.galante@unioeste.br

Abstract. *Although offering clear benefits for Web and business-critical demands, the use of cloud elasticity still imposes challenges when trying to reap its benefits over complex applications such as those modeled as pipelines. This often happens because replication, the standard technique for resource reorganization, by default, doesn't address both task-level parallelism and communication between working VMs. Taking into account this background, here we propose a model named Elastipipe to provide automatic elasticity over pipeline-structured applications. Elastipipe's contribution consists in a framework designed to provide the notion of flexible superscalar pipelines, in which scaling operations and load balancing take place over different elasticity units. An elasticity unit refers to a set of one or more stages of a pipeline that will be grouped together under the same elasticity rules. Using a real workload from a Brazilian company, the Elastipipe prototype presented performance gains of up to 60% when confronted with a non-elastic approach.*

Resumo. *Embora ofereça benefícios claros para aplicações Web, o uso da elasticidade em nuvem ainda impõe desafios ao tentar usá-la em aplicações complexas como aquelas modeladas em pipeline. Isso acontece porque a técnica de replicação, que é padrão na reorganização de recursos, normalmente não trata o paralelismo em nível de tarefas e comunicação entre VMs replicadas. Neste contexto propomos um modelo chamado Elastipipe para fornecer elasticidade automática sobre aplicações baseadas em linhas de montagem, ou pipelines. A contribuição de Elastipipe consiste num arcabouço projetado para oferecer o conceito de pipeline superescalar flexível, no qual operações de elasticidade e balanceamento de carga são realizadas sobre diferentes unidades de elasticidade. Uma unidade de elasticidade se refere a um conjunto de um ou mais estágios, os quais são agrupados sob as mesmas regras de elasticidade. Usando a carga de trabalho de uma empresa brasileira, o protótipo apresentou ganhos de desempenho de até 60% quando comparado com uma execução não elástica.*

1. Introdução

Vários sistemas de computação vêm sendo portados para execução em nuvem graças a capacidade de auto-escala, ou elasticidade, desse ambiente de computação distribuída [Herbst et al. 2015]. O serviço de elasticidade é melhor representado em sistemas de comércio e transações Web críticos que devem satisfazer a um determinado acordo de nível de serviço (SLA), como por exemplo, limites máximos de

tempo de resposta para diferentes tipos de requisição de entrada [Ali-Eldin et al. 2012, Islam et al. 2012, Jamshidi et al. 2014]. Além de atuar em questões de desempenho e economia de energia elétrica, a elasticidade horizontal e/ou vertical é especialmente pertinente para ambientes dinâmicos, onde a intervenção humana é difícil ou ainda, impossível [Farokhi et al. 2015]. A técnica padrão está baseada no uso de um balanceador de carga que recebe requisições e realiza operações de aumento ou redução da infraestrutura usando réplicas de máquinas virtuais (VMs). Essa interação apresenta uma arquitetura fracamente acoplada, onde réplicas não estabelecem comunicação entre si, somente cooperando através de mensagens de requisição e resposta com o balanceador de carga.

Aplicações Web como e-commerce e EFT (*Electronic Funds Transfer*) tiram proveito de uma organização de elasticidade fracamente acoplada para oferecer uma melhor experiência para o usuário [Righi et al. 2015]. Na mesma linha, essa ideia vem sendo cada vez mais explorada no contexto de sistemas de computação de alto desempenho (PAD), onde é desejável fornecer uma reorganização automática de recursos em aplicações Mestre-Escravo, Sacola de Tarefas (*Bag of Tasks*) e MapReduce [Raveendran et al. 2011]. Nesse escopo, a ideia é explorar o paralelismo de dados, uma vez que cada VM, ou réplica, executa o mesmo código sobre requisições de entrada diferentes. Entretanto, esse tipo de projeto de elasticidade não endereça aplicações complexas com comportamento de carga e comunicação entre processos imprevisíveis. Essas interações, por exemplo, podem incluir dependência de dados, causando comunicação entre processos onde origem e destino somente podem ser conhecidos em tempo de execução.

Com o intuito de mitigar esse problema, o estado-da-arte propõe como principais alternativas o uso de interfaces de programação (APIs) para extrair o paralelismo, forçando os usuários a mudarem ou inserirem diretivas no próprio código fonte para colher os benefícios dessa capacidade da computação em nuvem [Dustdar et al. 2012, Imai et al. 2012, Raveendran et al. 2011, Zhang et al. 2010]. Embora viável, essa solução pode ser classificada como custosa em relação ao tempo de implementação (*time-consuming*) e não portátil, endereçando um dueto particular composto de aplicação e infraestrutura. Nesse contexto, planeja-se investigar como eficientemente integrar aplicações complexas e a capacidade de elasticidade em nuvem; mais precisamente, pretende-se focar em aplicações estruturadas em pipelines. Um pipeline é um conjunto de elementos processadores, ou estágios, conectados em série onde a saída de um elemento é a entrada do próximo. Somado ao emprego nas indústrias de CPUs e GPUs, pipelines estão presentes em workflows de PAD, bem como em procedimentos de empresas que exploram o paralelismo em nível de tarefas [Jahn et al. 2013].

Considerando o contexto exposto previamente, esse artigo apresenta o Elastipipe: um modelo de elasticidade que fornece reorganização dinâmica de recursos para aplicações estruturadas em pipeline a nível de SaaS (Software como Serviço). O usuário simplesmente adiciona tarefas para serem computadas no pipeline, o qual é gerenciado pela nuvem com operações de expansão ou consolidação de VMs. A reorganização de recursos é relevante para evitar execuções nos estados com baixa ou alta carga de CPU, enquanto mantém igual ou melhora o tempo de execução de um conjunto de tarefas quando comparado com pipelines não elásticos. A contribuição científica de Elastipipe consiste em um arcabouço para explorar o conceito de Pipeline Superescalar Elástico (PSE) no qual o administrador da aplicação pode projetar unidades de elasticidade para um con-

junto arbitrário de estágios. Por exemplo, suponha que os estágios e_1 e e_2 sejam classificados como CPU-bound, assim neste caso é possível mapeá-los para a mesma unidade de elasticidade: regras e operações de elasticidade atuam sobre VMs otimizadas para uso de CPU, as quais incluem ambos os estágios citados. Baseado no modelo proposto também foi desenvolvido um protótipo que foi avaliado com uma carga de trabalho real de uma empresa brasileira, tanto em termos de definição de estágios do pipeline quanto das tarefas que serão executadas. Os resultados são encorajadores, com ganhos de desempenho de até 60%, fazendo com que a empresa planeje migrar o seu sistema para uma nuvem na qual seja possível implementar sua aplicação conforme o modelo Elastipipe.

O restante desse artigo irá introduzir os trabalhos relacionados na Seção 2. A Seção 3 descreve o modelo Elastipipe. A Seção 4 apresenta a metodologia de avaliação, enquanto que a Seção 5 mostra uma discussão sobre os resultados. Por fim, a Seção 6 enfatiza as principais conclusões e contribuições.

2. Trabalhos Relacionados

A exploração da elasticidade em nuvem é cada vez mais perceptível em nuvens públicas e privadas (incluindo Amazon AWS, Windows Azure e RightScale para o primeiro grupo, e Eucalyptus, OpenStack e OpenNebula, para o segundo). Muitas iniciativas de pesquisa focam em utilizar esses ambientes, explorando a elasticidade através da técnica horizontal (expansão ou consolidação de VMs), onde a replicação serve para criar *templates* que servirão como unidades básicas para as operações de elasticidade [Dustdar et al. 2012, Imai et al. 2012, Loff and Garcia 2014, Marshall et al. 2010, Martin et al. 2011, Rajan et al. 2011, Raveendran et al. 2011, Righi et al. 2015, Tran et al. 2011, Zhang et al. 2010]. Para essa seção, foram selecionados quatro artigos relacionados ao tratamento de cargas de trabalho complexas [Imai et al. 2012, Marshall et al. 2010, Tran et al. 2011, Zhang et al. 2010].

Marshall, Keahey e Freeman [Marshall et al. 2010] propuseram o sistema Elastic Site, o qual eficientemente adapta os serviços oferecidos dentro de um domínio, como escalonadores em lote, armazenamento ou Web Services, para tirarem proveito de um provisionamento elástico de recursos. A ideia é promover uma extensão de clusters físicos para dentro da nuvem, onde nós trabalhadores adicionais são dinamicamente adquiridos ou desligados da nuvem baseado em mudanças na fila de trabalhos do cluster. Tran, Skhiri e Zimányi [Tran et al. 2011] desenvolveram o modelo Elastic Queue Service (EQS), o qual pode ser visto como uma arquitetura de fila de mensagens (message queue) elástica e um algoritmo elástico que pode ser adaptado para muitos sistemas de fila de mensagens com o intuito de explorar esta característica. De acordo com os autores, sistemas de filas de mensagens existentes estão somente habilitados para fornecer escalabilidade para um número reduzido de clientes e não estão aptos para tratar da dinamicidade quanto ao provisionamento de recursos. A arquitetura de EQS é realizada em camadas organizadas em componentes distribuídos, cada qual com as suas próprias regras de elasticidade. Os autores também propuseram um algoritmo de balanceamento de carga em nível de usuário, o qual é capaz de rebalancear o trabalho depois das ações de elasticidade.

Imai, Chestna e Varela [Imai et al. 2012] apresentaram o Cloud Operating System (COS), que atua como um middleware que suporta cargas de trabalho de forma elástica e escalável através da estratégia de reconfiguração baseada em migração de máquinas vir-

tuais. Enquanto outros arcabouços (por exemplo, MapReduce ou Google App Engine) forçam com que os desenvolvedores desenvolvam suas aplicações segundo uma interface de programação (API) específica, COS fornece escalabilidade através de um arcabouço de propósito geral baseado em uma linguagem de programação orientada a atores. Essa estratégia é pertinente para melhorar o tempo de migração, uma vez que somente o estado da aplicação é de fato migrado. Zhang et al. [Zhang et al. 2010] propuseram um modelo de aplicação elástico que habilita recursos da nuvem para aumentar a capacidade de computação em dispositivos móveis. A principal ideia foi a partição de uma aplicação em múltiplos componentes chamadas Weblets, os quais podem ser executados em um dispositivo móvel ou migrados para a nuvem. Atuando com conceitos similares a uma arquitetura orientada a serviço, Weblets permitem migração em nível de usuário e, assumindo que eles operam sobre HTTP, podem facilitar a interoperabilidade com provedores de nuvem. Assim, uma aplicação elástica pode aumentar as capacidades de dispositivos móveis incluindo capacidade de CPU, armazenamento e largura de banda através de recursos provenientes da nuvem.

Numa análise das quatro iniciativas citadas nessa seção, pode-se enfatizar que há uma lacuna de trabalho na exploração de um modelo de elasticidade para endereçar aplicações estruturadas em pipelines. Esse modelo de aplicação apresenta diferentes estágios de processamento, cada qual com suas próprias funcionalidades, consequentemente impondo diferentes *templates* de VMs para as operações de elasticidade. Portanto, uma solução de elasticidade deve trabalhar a reorganização de recursos em nível de estágios do pipeline para que se possa aumentar a eficiência total do sistema ao mesmo tempo que evita ignorar o fato de que cada estágio é diferente e pode exigir recursos computacionais distintos entre si.

3. Modelo Elastipipe

Essa seção apresenta o modelo Elastipipe, o qual foi desenvolvido em nível de SaaS de uma nuvem para gerenciar a elasticidade de aplicações estruturas em pipeline. Com o intuito de obter melhor desempenho, a ideia é explorar a reorganização de recursos para tornar mais rápida a execução de um conjunto de tarefas através de um mecanismo com baixo esforço por parte dos usuários, os quais não precisam mudar suas aplicações. Consequentemente, no ponto de vista do sistema, essa melhoria na perspectiva de tempo também contribui para que se obtenha um melhor índice de vazão nas tarefas. Em adição as operações de aumento de recursos que estão diretamente ligadas com o desempenho da aplicação, Elastipipe também suporta operações de consolidação de VMs, focando num melhor uso dos recursos e reduzindo o consumo energético. A Figura 1 ilustra a arquitetura de Elastipipe. O modelo pode ser visto como um arcabouço que facilita a migração de aplicações em pipeline para serem executadas na nuvem, oferecendo redução nas perspectivas de custo e riscos [Rajan et al. 2011]. Na Figura 1, os módulos Gerente de Elasticidade e Monitoramento de Recursos são comumente suportados por provedores de nuvens através de interface em linha de comando ou gráfica; enquanto que os módulos Entrada de Tarefas, Controle e Execução são de responsabilidade do modelo Elastipipe.

3.1. Arquitetura

Esta subseção descreve a arquitetura do modelo Elastipipe, detalhando seus módulos e interações. O elemento de entrada para o sistema é uma tarefa representada por T_z , as

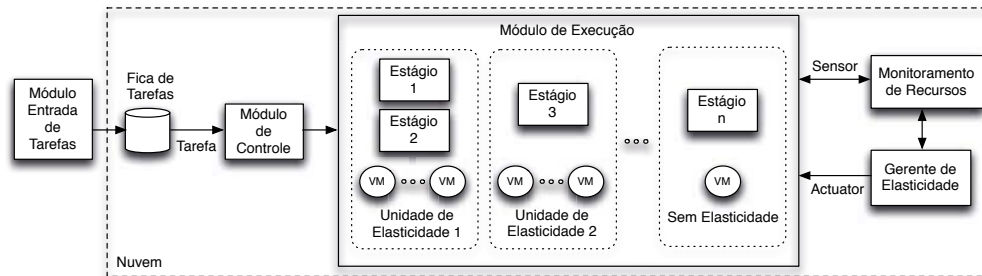


Figura 1. Arquitetura de Elastipipe. Aqui, detalha-se um exemplo no módulo de Execução: estágios podem ser agrupados sob uma mesma unidade de elasticidade, ou ainda, eles podem não apresentar a capacidade de elasticidade.

quais são numeradas da seguinte forma: $1 \leq z \leq l$. Cada tarefa possui um determinado tipo, o qual indica o pipeline com os estágios apropriados para processá-la. Cada tarefa pode apresentar um conjunto de dependências (para cada tarefa t_a , tem-se um conjunto de dependência D_a composto por tarefas arbitrárias que estão no intervalo $[1, a - 1]$). Assim, a tarefa t_a irá somente entrar no pipeline quando todas as demais em D_a estiverem finalizadas. Um pipeline, representado por P_i , numerados de $1 \leq i \leq m$, é uma unidade lógica que é composto por um conjunto de estágios (representado por s_j , numerados de $1 \leq j \leq n$), podendo compartilhar estágios com outros pipelines. Uma vez que o modelo é de propósito geral, não define-se a funcionalidade de cada estágio; esse procedimento é conduzido em tempo de implementação de acordo com as necessidades da aplicação.

3.1.1. Módulo de Entrada de Tarefas e Controle

O Módulo de Entrada de Tarefas é responsável por inserir uma tarefa na fila (aqui habilitada por um sistema de fila de mensagens), de modo que o Módulo de Controle esteja habilitado para, na sequência, ler dados dessa área compartilhada. Em particular, a fila atua como um elemento para controlar o efeito de Jitter, suavizando picos de demanda e assim habilitando a entrega regular para o Módulo de Controle. O Módulo de Controle é responsável pelas seguintes funcionalidades: (i) capturar uma tarefa da fila de tarefas; (ii) validar uma tarefa; (iii) controlar o paralelismo; (iv) tratar erros e; (v) disparar e monitorar tarefas. Após capturar uma tarefa que deve ser computada, a funcionalidade ii garante a integridade da tarefa e sua aderência ao Módulo de Execução. A ideia é reduzir a quantidade de erros na aplicação devido a mensagens desconhecidas ou mal formadas, evitando problemas que possam interferir no desempenho de outras tarefas que já estejam em execução. O controle de paralelismo é associado a definição de qualidade de serviço (QoS) para gerenciar a superescalaridade de cada pipeline. Nesse sentido é possível definir um número mínimo e máximo para tarefas que serão executadas em paralelo, o qual impacta no número de VMs usado para tratar cada unidade de elasticidade (que por sua vez pode ser composta por um único estágio ou uma coleção deles, mais detalhes na Subseção 3.2). O tratamento de erros endereça exceções, assim como a repetição da execução de tarefas que falharam devido a problemas de comunicação ou infraestrutura. A consolidação de uma VM que tenha pelo menos uma tarefa em execução é um exemplo da necessidade de relançamento de uma ou mais tarefas.

Para viabilizar a quinta funcionalidade, foi projetado um despachador de tarefas que identifica qual pipeline é o destino de uma determinada tarefa, além de controlar o fluxo

de tarefas no pipeline e a dependência entre tarefas. Para cobrir tais casos, a estratégia foi a mesma aplicada em pipelines de microprocessadores, onde conflitos de dados ou controle podem prejudicar a eficiência do pipeline com pausas nas execuções. Quando situações como essa são detectadas, define-se que o estado da tarefa é em espera até que todas as dependências estejam completamente satisfeitas. A Figura 2 ilustra um exemplo de despacho onde cinco tarefas e três pipelines estão envolvidos. O fluxo de execução de cada tarefa é o seguinte: o Módulo de Controle decide o pipeline correto e despacha a tarefa para o primeiro estágio, recebendo a saída da computação na sequência e, então, despachando novamente a tarefa (e seu conjunto de dados de saída) para o próximo estágio e assim por diante. Em outras palavras, baseado em [Rajan et al. 2011], foi projetada uma organização mestre-escravo para gerenciar a execução das tarefas, onde o Módulo de Controle representa o mestre e os estágios de processamento, os escravos. Por fim, é possível observar na Figura 2 que um estágio em particular pode aparecer em mais do que um pipeline. Isso é esperado uma vez que pipelines, na modelagem do Elastipepe, são fluxos lógicos de trabalho e a execução é tratada individualmente através dos estágios.

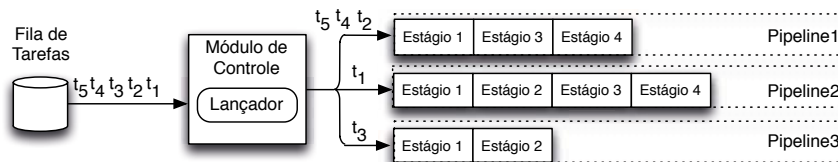


Figura 2. Execução de tarefas em diferentes pipelines.

3.1.2. Módulo de Execução

Este módulo abriga os componentes responsáveis pela execução de cada estágio do pipeline. O projeto do módulo de execução foi feito tendo em mente o conceito de microserviços [Namiot and Sneps-Snepe 2014]: um microserviço representa um pequeno e independente serviço o qual se comunica através de protocolos simples e bem definidos. O uso de microserviços tem crescido como blocos básicos para a programação em nuvem dada as características que proporciona como melhor escalabilidade, *compatibility* e facilidade de mapeamento para recursos [Newman 2015]. Newman afirma que é comumente observável que transações CPU ou IO-bound são endereçadas para uma mesma unidade de elasticidade a qual pode penalizar um dos tipos mencionados ou acarretar o desperdício de recursos [Newman 2015]. Através da adoção da abordagem de microserviços, administradores podem escalar cada unidade de maneira que melhor se encaixe às suas necessidades, assim obtendo reduções de tempo de execução e custos.

Embora isolados em processos diferentes, microserviços podem se comunicar e, ainda, depender um do outro para que um objetivo comum seja alcançado. Tal afirmação facilita a construção de aplicações multi-plataforma, a emergência de novas aplicações através da combinação de microserviços, e facilita a criação de pipelines definidos em software. Através da adoção de microserviços para a implementação dos estágios do pipeline, é possível melhorar o reconhecimento de possíveis gargalos e promover um diagnóstico mais preciso em nível de aplicação e infraestrutura no momento de detecção de tarefas em falha. Por fim, microserviços são pertinentes como uma unidade de elasticidade e para implementar aplicações MPMD (*Multiple programs, multiple data streams*) como as que são organizadas em pipelines. Nessa organização, é possível explorar o pa-

ralismo em nível funcional (ou de código) com múltiplos processadores autônomos que operam em diferentes nós, representados aqui por estágios do pipeline.

3.2. Modelo de Elasticidade

Elastipipe tira vantagem do monitoramento de recursos e gerenciamento de elasticidade já disponíveis em provedores de nuvem. Portanto, essa subseção objetiva classificar a estratégia de elasticidade de Elastipipe, enquanto também apresenta os requisitos que o provedor de nuvem deve preencher. A abordagem de elasticidade adotada é horizontal, automática e reativa. Levando em conta a palavra-chave reativa, Elastipipe considera: (i) dois limiares (*thresholds*), um mínimo e outro máximo, para guiar as ações de elasticidade quando violados; (ii) a técnica de replicação sem um único *template* de VM, mas sim uma coleção deles que varia de acordo com o número de unidades de elasticidade; (iii) um balanceador de carga por unidade de elasticidade. Essa combinação foi selecionada levando em consideração as seguintes premissas: facilidade de configuração, eficiência das ações de elasticidade e disponibilidade de implementação na maioria dos ambientes de computação em nuvem vigentes.

A abordagem de elasticidade horizontal é conhecida pelas operações de expansão ou consolidação de VMs. A abordagem vertical (também conhecida por redimensionamento de recursos) foi preterida porque é limitada às capacidades de um único nó computacional (CPU, memória, disco e rede), não sendo uma solução escalável para suportar aplicações do tipo pipeline. Do ponto de vista do usuário, a reorganização dos recursos é feita de forma automática e de acordo com a demanda de computação, sem intervenções em tempo de execução. Os únicos parâmetros necessários para o modelo são os limites usados para guiar a elasticidade e as ações pré-configuradas pelo administrador da aplicação. Basicamente, “se” uma determinada regra for satisfeita, “então” uma ação de elasticidade é disparada para otimizar os objetivos estabelecidos pelo administrador, os quais podem ser desempenho, energia, custo ou, ainda, uma combinação deles.

Limites, regras e ações são usados para controlar unidades de elasticidade. A Figura 3 ilustra um exemplo de quatro mapeamentos, três deles com unidades de elasticidade. Uma unidade de elasticidade pode ser vista como um agrupamento de um ou mais estágios sobre o mesmo *template* de VM; portanto, todas as operações de elasticidade irão sempre instanciar ou consolidar este elemento funcional. Cada unidade de elasticidade possui as suas próprias regras e ações, em adição a um balanceador de carga que decide a réplica alvo que está menos carregada e um estágio em particular (quando existe mais de um estágio na mesma VM) para executar determinada tarefa. Balanceadores de carga atualizam as suas conexões com as réplicas a cada alteração na infraestrutura. Usando unidades de elasticidade como blocos básicos, Elastipipe propõe o conceito de Pipeline Superescalar Elástico, como segue:

- Definição 1. Pipeline Superescalar Elástico: O usuário ou administrador da aplicação está apto para configurar tantas unidades de elasticidade quanto requeridas para executar a sua aplicação organizada em pipeline, provendo o isolamento da elasticidade e a possibilidade de tirar proveito da decomposição funcional para agrupar estágios de acordo com a natureza de recursos exigida em cada estágio (CPU, memória, disco ou rede).

É importante enfatizar que a noção de pipeline é lógica, ou seja, uma unidade de elasticidade pode servir vários pipelines. Por exemplo, é possível acomodar os estágios

s_1 , s_2 e s_3 na mesma unidade de elasticidade, mas s_1 e s_2 pertencerem ao pipeline P_1 enquanto s_2 e s_3 pertencem ao pipeline P_2 . Por fim, considerando que aplicações complexas são compostas por vários estágios, cada qual com os seus requisitos de recursos, os conceitos de unidade de elasticidade e Pipeline Superescalar Elástico são úteis para balancear a duração de cada estágio. Esse equilíbrio é essencial para evitar gargalos no pipeline, uma vez que o estágio mais lento será responsável pelo acúmulo de tarefas, prejudicando a eficiência da técnica de linha de montagem [Jahn et al. 2013]. Como um efeito desejável, o uso de unidades de elasticidade também contribui para aumentar a confiabilidade do sistema: o uso de réplicas favorece a tolerância a falhas. Assim, falhas sobre um único componente não irão tornar a aplicação indisponível [Zhang et al. 2010].

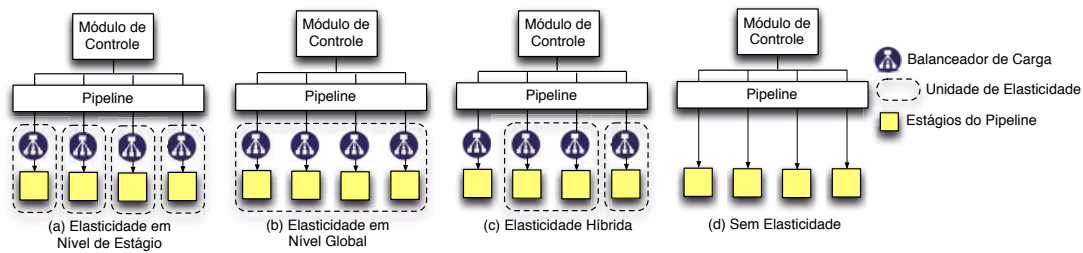


Figura 3. Unidade de elasticidade e Balanceador de Carga: Permitindo diferentes estratégias de elasticidade de acordo com o comportamento de cada estágio.

4. Metodologia de Avaliação

Essa seção descreve o protótipo Elastipipe e parâmetros envolvidos na sua avaliação.

4.1. Protótipo

Foi desenvolvido um protótipo no qual as tarefas de entrada são representadas por arquivos texto. Em adição, foi selecionado o ambiente Amazon Web Services (AWS) como provedor de nuvem e um conjunto de softwares da Microsoft para o desenvolvimento dos módulos do Elastipipe: compilador C#, arcabouço de desenvolvimento .NET e banco de dados SQL Server. Quanto à AWS, foram usados os seguintes componentes: (i) EC2 para gerenciamento de máquinas virtuais; (ii) CloudWatch para monitoramento; (iii) AutoScaling para gerenciamento e configuração da elasticidade; (iv) Elastic Load Balancing (ELB) para habilitar balanceamento de carga dentro das unidades de elasticidade; (v) SQS como serviço de mensagens distribuído e; (vi) S3 para armazenamento. O protótipo opera com o seguinte caminho de execução: (a) um componente primeiro detecta a presença de arquivos em S3 e gera uma mensagem de notificação que é inserida na fila de mensagens; (b) outro componente, responsável por capturar as mensagens dessa fila, detecta a notificação e inspeciona o arquivo para detectar qual o pipeline que a tarefa pertence; (c) levando em conta essa informação, o Módulo de Controle despacha a tarefa para o serviço apropriado para executar cada estágio do pipeline. Uma vez que os serviços operam de forma isolada e sem estado, eles podem ser seguramente instanciados quantas vezes for necessário sem comprometer a integridade da aplicação.

Foram implementados dois pipelines: P_1 e P_2 . P_1 possui 5 estágios como segue: (i) validação de tarefas (caráter CPU-bound); (ii) conversão de leiaute (caráter CPU-bound); (iii) análise de consistência (caráter memory-bound); (iv) persistência em banco de dados (disk-bound); (v) estágio de distribuição, o qual é responsável por enviar um relatório para outras fontes de dados e aplicações relacionadas (network-bound). P_2 representa somente dois estágios: (i) validação de esquema, o qual é responsável pela inspeção

Tabela 1. Descrição dos cenários propostos.

Cenário	Elasticidade	Descrição da VM	Regras de Elasticidade	
1	sem elasticidade	m4.large (fixa para todos estágios)	—	
2	elasticidade global	m4.large (única unidade de elasticidade para todos estágios)	$\uparrow CPU > 75\%$	$\downarrow CPU < 35\%$
3	elasticidade funcional	c4.large (unidade otimizada para CPU)	$\uparrow CPU > 75\%$	$\downarrow CPU < 35\%$
		r3.large (unidade otimizada para Memória)	$\uparrow RAM > 75\%$	$\downarrow RAM < 35\%$
		i2.xlarge (unidade otimizada para Disco/IO)	$\uparrow IOPS > 3k$	$\downarrow IOPS < 1k$
		m4.large (unidade otimizada para Rede)	$\uparrow MBps > 40$	$\downarrow MBps < 10$

de uma tarefa e a sua validação de acordo com leiautes pré-definidos presentes em XML Schema (CPU- e memory-bound); (ii) persistência em banco de dados (disk-bound). Considerando a definição dos estágios foram definidas 5 unidades de elasticidade usando o pacote AutoScaling da AWS: (a) Otimizado para CPU, para os estágios de validação, conversão e esquema; (b) otimizado para memória, para o estágio de consistência; (c) otimizado para rede, para o estágio de distribuição; (d) otimizado para disco, para o estágio de persistência em banco de dados; (e) tratamento global onde todos os estágios são inseridos juntos em uma única unidade de elasticidade.

4.2. Definindo a Carga de Trabalho do Pipeline e Métricas de Avaliação

Para os testes, foi utilizada uma carga de trabalho capturada a partir de amostras provenientes de uma empresa parceira localizada no parque tecnológico da Unisinos (Tecnosinos). A carga se refere a traços capturados durante uma semana, com um total de 11.459 tarefas em forma de arquivos texto. Cada arquivo varia entre 3.500 e 480.000 linhas de dados, gerando um total de até 110 megabytes de dados. Em adição, foram criados três cenários para avaliar diferentes estratégias de uso da elasticidade, como pode ser visto na Tabela 1. Essa tabela também apresenta as regras de elasticidade para cada unidade de elasticidade. Três métricas são consideradas para a avaliação do protótipo Elastipipe: tempo, recurso e custo. A primeira se refere a desempenho, apresentando o tempo total de execução para todo o conjunto de dados. Através de observações periódicas feitas pela ferramenta CloudWatch, foi realizada a estimativa da métrica de recursos levando em consideração uma configuração da nuvem com i VMs e $obs(i)$ como o número de observações em que essa configuração aparece. Por fim, usa-se a noção de custo de sistemas paralelos para que seja obtida a métrica custo: multiplicação de tempo por recursos. As Equações 1 e 2 computam os valores das métricas recurso e custo. Na Equação 1, v representa o máximo de VMs que podem ser instanciados nas unidades de elasticidade.

$$recursos = \sum_{i=1}^v (i \times obs(i)) \quad (1)$$

$$custo = recursos \times tempo \quad (2)$$

5. Resultados

A Tabela 2 apresenta os resultados. O tempo de execução total nessa tabela confirma os benefícios de se usar a elasticidade em nuvem para tratar demandas de aplicações pipeline. A estratégia de mapeamento de serviços (estágios do pipeline) de acordo com o comportamento do seu recurso alvo (CPU-bound, network-bound, memory-bound ou disk-bound), em adição a regras particulares para endereçá-los, foi responsável pelas melhores taxas de desempenho a favor do cenário 3. Considerando o desempenho do cenário 1 como um fator de normalização, foram obtidas melhorias de 62% e 72% quando habilita-se a

Tabela 2. Resultados. Cenário 1: sem elasticidade; Cenário 2: elasticidade através de uma única unidade de elasticidade; Cenário 3: elasticidade com múltiplas unidades de elasticidade

Cenário	Duração Média por Tarefa (s)	Vazão por Minuto	por	Duração Total	Número Médio de VMs	Observações	Uso de Recursos	Custo
1	5,52	108,70		17:34:12	1,0	1.054	$1.054 * 1 = 1.054$	18.518
2	2,62	594,74		06:45:36	5,2	405	$40 * 1 + 40 * 2 + 40 * 3 + 40 * 4 + 40 * 5 + 40 * 6 + 40 * 7 + 120 * 8 = 2.080$	14.056
3	2,13	1.007,45		04:55:12	8,5	295	$29 * 4 + 29 * 6 + 145 * 9 + 87 * 10 = 2.465$	12.131

elasticidade. Comparando somente os cenários 2 e 3, o último atingiu 27% de ganho de desempenho enquanto usa 39% mais VMs para atingi-lo.

A Figura 4 (a) ilustra o comportamento de alocação de recursos ao longo da execução em cada cenário. Pode-se observar um gargalo de desempenho no cenário 2: para a maior parte do tempo de execução é utilizado um número saturado de VMs, estabilizando perto do final da execução. Isto explica o efeito escada nesse cenário: uma nova VM é alocada, mas logo se torna sobrecarregada, implicando em uma nova ação de elasticidade. Esse gargalo é minimizado no cenário 3 através da distribuição de unidades de elasticidade desde o início, percebe-se portanto um equilíbrio no uso de recursos mais cedo (veja a Figura 4 (b)).

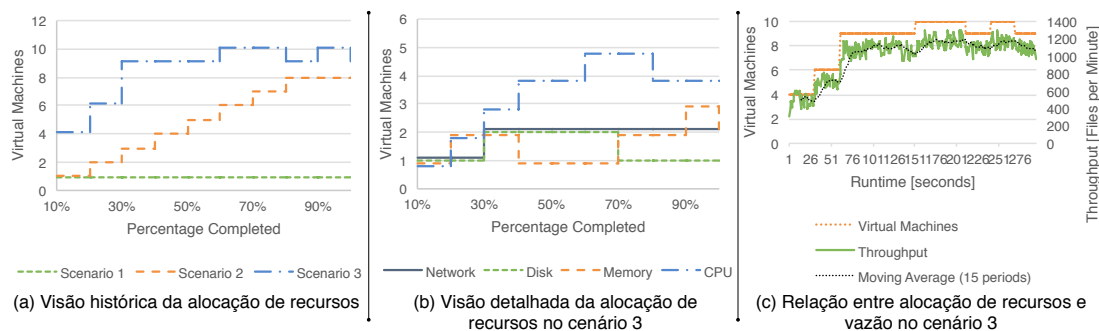


Figura 4. Observando a alocação de recursos e vazão ao longo da execução.

O cenário 1 utiliza uma única VM para acomodar os quatro conjuntos de estágios. Nesse contexto, utiliza-se a família de máquinas *m4*, a qual é classificada pela AWS como servidores de propósito geral que balanceiam requisitos de processamento, disco, memória e rede. Esse mapeamento é responsável pelo menor consumo de recursos obtidos nos testes. Entretanto, ele apresenta o pior custo uma vez que o tempo de execução foi muito maior do que aquele observado nos demais cenários. O cenário 2 também começa com uma única VM, mas a elasticidade foi responsável por aumentar o consumo de recursos enquanto melhora também o tempo de execução. Em média, tem-se 26 tarefas sendo executadas concorrentemente nos pipelines quando habilita-se elasticidade sobre uma única unidade de elasticidade que abriga todos os estágios.

O cenário 3 trabalha com agrupamento de estágios em diferentes unidades de elasticidade de acordo com as características dos recursos. Nessa linha, o fato de usar mais nós neste cenário foi responsável pelo maior uso de recursos. Em particular, Baliga et al. [Baliga et al. 2011] afirmam que o número de VMs em um nó não é o maior fator de

influência para o consumo de energia; ao invés disso, eles informam que o fator mais relevante é se o nó está ligado (energizado) ou não. Sobrepondo o cenário 2, no cenário 3 foi medida uma quantidade de 36 tarefas em paralelo sendo executadas nos pipelines, quase dobrando a vazão de tarefas por minuto (de 594,74 no cenário 2 para 1.007,45 no cenário 3). A Figura 4 (c) ilustra a vazão de tarefas por minuto e sua relação com a alocação de recursos. Por fim, observa-se que o uso da decomposição funcional dos estágios do pipeline para projetar as unidades e elasticidade foi o fator que mais influenciou o desempenho, também facilitando a manutenção e a detecção de problemas.

6. Conclusão

O conceito de linhas de montagem, originado na indústria automotiva e popularizado em projetos de microprocessadores, está presente em cargas de trabalho complexas oriundas de uma variedade de áreas do conhecimento. Neste contexto, este artigo propôs um modelo chamado Elastipipe para abordar essas aplicações apresentando, ao mesmo tempo, o conceito de Pipeline Superescalar Elástico para apoiar a reorganização dinâmica de recursos. Essa reorganização é atingida através da definição de unidades de elasticidade por usuários ou administradores da aplicação, podendo contemplar um ou mais estágios do pipeline. Cada operação de expansão ou consolidação de VMs considera essa unidade como grão de recurso para aumentar ou reduzir a quantidade de recursos que melhor mapeia para a execução da aplicação em um determinado momento.

Os resultados com um protótipo revelaram a viabilidade do uso da elasticidade em nuvem, onde cenários com elasticidade apresentaram melhores taxas de custo quando comparados a uma abordagem não-elástica. Além disso, a decomposição funcional dos estágios em unidades de elasticidade de acordo com suas características (CPU-, memory-, network- ou disk-bound) foi decisiva para a obtenção dos melhores resultados de desempenho. Esse cenário de implantação alcançou ganhos de 27% quando comparado com uma situação em que todos os estágios estão concentrados em uma única unidade de elasticidade. Investigações futuras incluem o uso de cargas de trabalho sintéticas para analisar o comportamento do modelo com diferentes características de carga: ascendente, descendente e flutuante (onda).

Agradecimentos

Os autores gostariam de agradecer aos seguintes órgãos: CNPq, CAPES e FAPERGS.

Referências

- Ali-Eldin, A., Tordsson, J., and Elmroth, E. (2012). An adaptive hybrid elasticity controller for cloud infrastructures. In *Network Operations and Management Symposium (NOMS), 2012 IEEE*, pages 204–212.
- Baliga, J., Ayre, R., Hinton, K., and Tucker, R. (2011). Green cloud computing: Balancing energy in processing, storage, and transport. *Proceedings of the IEEE*, 99(1):149–167.
- Dustdar, S., Guo, Y., Han, R., Satzger, B., and Truong, H.-L. (2012). Programming directives for elastic computing. *Internet Computing, IEEE*, 16(6):72–77.
- Farokhi, S., Jamshidi, P., Brandic, I., and Elmroth, E. (2015). Self-adaptation challenges for cloud-based applications : A control theoretic perspective. In *10th International Workshop on Feedback Computing (Feedback Computing 2015)*. ACM.

- Herbst, N. R., Kounev, S., Weber, A., and Groenda, H. (2015). Bungee: An elasticity benchmark for self-adaptive iaas cloud environments. In *Proc. Int. Symp. Software Engineering for Adaptive and Self-Managing Systems, SEAMS '15*, pages 46–56, IEEE.
- Imai, S., Chestna, T., and Varela, C. A. (2012). Elastic Scalable Cloud Computing Using Application-Level Migration. In *2012 IEEE Fifth International Conference on Utility and Cloud Computing*, pages 91–98, Honolulu. IEEE.
- Islam, S., Keung, J., Lee, K., and Liu, A. (2012). Empirical prediction models for adaptive resource provisioning in the cloud. *Future Gener. Comput. Syst.*, 28(1):155–162.
- Jahn, J., Pagani, S., Kobbe, S., Chen, J.-J., and Henkel, J. (2013). Optimizations for configuring and mapping software pipelines in many core systems. In *Proc. Annual Design Automation Conference, DAC '13*, pages 130:1–130:8, ACM.
- Jamshidi, P., Ahmad, A., and Pahl, C. (2014). Autonomic resource provisioning for cloud-based software. In *Proc. of the 9th Int. Symp. on Software Engineering for Adaptive and Self-Managing Systems, SEAMS 2014*, pages 95–104, ACM.
- Loff, J. and Garcia, J. (2014). Vadara: Predictive Elasticity for Cloud Applications. In *2014 IEEE 6th International Conference on Cloud Computing Technology and Science*, pages 541–546, Singapore. IEEE.
- Marshall, P., Keahey, K., and Freeman, T. (2010). Elastic Site: Using Clouds to Elastically Extend Site Resources. In *2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, pages 43–52, Melbourne. IEEE.
- Martin, P., Brown, A., Powley, W., and Vazquez-Poletti, J. L. (2011). Autonomic management of elastic services in the cloud. In *2011 IEEE Symposium on Computers and Communications (ISCC)*, pages 135–140, Kerkyra. IEEE.
- Namiot, D. and Sneps-Sneppe, M. (2014). On Micro-services Architecture. *International Journal of Open Information Technologies*, 2(9):24–27.
- Newman, S. (2015). *Building Microservices*. O'Reilly Media, Inc., Sebastopol.
- Rajan, D., Canino, A., Izaguirre, J. A., and Thain, D. (2011). Converting a high performance application to an elastic cloud application. In *Proc. Int. Conf. on Cloud Computing Technology and Science, CLOUDCOM '11*, pages 383–390, IEEE.
- Raveendran, A., Bicer, T., and Agrawal, G. (2011). A framework for elastic execution of existing mpi programs. In *Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on*, pages 940–947.
- Righi, R., Rodrigues, V., Andre daCosta, C., Galante, G., Bona, L., and Ferreto, T. (2015). Autoelastic: Automatic resource elasticity for high performance applications in the cloud. *Cloud Computing, IEEE Transactions on*, PP(99):1–1.
- Tran, N.-L., Skhiri, S., and Zimányi, E. (2011). EQS: An Elastic and Scalable Message Queue for the Cloud. In *2011 IEEE Third International Conference on Cloud Computing Technology and Science*, pages 391–398, Athens. IEEE.
- Zhang, X., Jeong, S., Kunjithapatham, A., and Gibbs, S. (2010). Towards an Elastic Application Model for Augmenting Computing Capabilities of Mobile Platforms. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, volume 48, pages 161–174. Springer, Heidelberg.

Análise de desempenho de handover em rede heterogênea usando MIH

Marco Antonio Simões Teixeira¹, Alexandre Veloso de Matos², Mário Ezequiel Augusto²

¹Engenharia Elétrica e Informática Industrial
Universidade Federal Tecnológica do Paraná (UTFPR) – Curitiba – PR – Brasil

²Departamento de Sistemas de Informação
Universidade do Estado de Santa Catarina (UDESC) – São Bento do Sul – SC – Brasil

mantonio.t91@gmail.com, {alexandre.matos,mario.augusto}@udesc.br

Abstract. *Nowadays mobile devices have more than one network interface, in a way that they can access the Internet using many wireless network technologies. However, in this heterogeneous environment, handovers cause data and connection losses. IEEE proposed IEEE 802.21 protocol (MIH), an interlayer architecture to assist heterogeneous handovers and to minimize packet loss. One of its advantages is that it can preview the need of a handover, warning higher layers about the event. This paper presents results of simulations of a wireless heterogeneous environment where we analysed the benefits of the Link_Going_Down event of IEEE 802.21. The protocol showed significant results related to throughput and handover delay.*

Resumo. *Atualmente os dispositivos móveis possuem várias interfaces de redes, podendo acessar a Internet usando várias tecnologias de rede sem fio. Porém, neste ambiente heterogêneo, as trocas de rede ocasionam perda de dados e de conexão. O IEEE propôs o protocolo IEEE 802.21 (MIH), uma arquitetura inter-camadas para auxiliar na troca de redes de tecnologias diferentes com a intenção de minimizar a perda de pacotes. Uma das vantagens é que ela pode prever a necessidade de uma troca de rede e avisar as camadas superiores sobre o evento. Este artigo apresenta os resultados da simulação de um ambiente heterogêneo sem fio onde nós analisamos os benefícios do evento Link_Going_Down do IEEE 802.21. O protocolo apresentou resultados significativos relacionados à vazão e ao tempo de handover.*

1. Introdução

As tecnologias de comunicação e transmissão de dados estão evoluindo bastante. Na última milha, é crescente o uso do meio sem fio para transmitir e receber dados, através de tecnologias como Wi-Fi e rede celular (3G e 4G). Em paralelo com o avanço das tecnologias sem fio, tem-se um avanço também nos dispositivos capazes de usufruir desses serviços. *Tablets*, *smartphones*, e outros dispositivos cada vez menores e mais potentes, possuem a capacidade para se conectar a redes sem fio, e a maioria possui a capacidade de se conectar em mais de uma tecnologia, como Wi-Fi e a telefonia móvel.

Os aplicativos e serviços também evoluíram e passaram a depender mais do acesso à Internet. Outras aplicações surgiram, como o caso da aplicação VoIP (Voz sobre IP), a qual nós utilizamos como caso de uso nesse trabalho. O serviço de VoIP tem como objetivo estabelecer uma ligação de voz entre dois dispositivos, utilizando a Internet e o protocolo IP. Qualquer degradação do sinal de rádio pode provocar perda de pacotes, o que pode acarretar em atrasos na fala, ou até mesma a perda da ligação. Quando o dispositivo do usuário realiza uma troca de rede, a tendência é perder mais pacotes, principalmente se a troca ocorrer entre tecnologias diferentes ou redes diferentes. Este procedimento de troca de rede é denominado de *handover*.

Para tornar o *handover* suave (*seamless*) e diminuir o tempo de troca e a perda de pacotes, o IEEE criou um protocolo denominado IEEE 802.21 (2009) ou Media Independent Handover (MIH). Esse protocolo possui regras e normas para serem adotadas ao se praticar a troca entre redes, seja entre mesma tecnologia (*handover* horizontal) como, por exemplo, entre Wi-Fi e Wi-Fi, seja entre tecnologias diferentes (*handover* vertical) como Wi-Fi e rede celular. O MIH é um protocolo intercamadas, o qual encaminha comandos, eventos e informações para camadas inferiores e superiores.

Como este protocolo não é implementado nos equipamentos de rede dos principais fabricantes, não é possível conhecer o desempenho real do MIH através de ambientes reais, ou seja, não se sabe o quanto este protocolo pode melhorar as trocas de rede. Diversos pesquisadores têm realizado simulações do MIH para analisar seus benefícios.

Esse trabalho estudou, através de simulação, o quanto o IEEE 802.21 pode ajudar no problema de perda de dados ao se trocar de rede entre tecnologias distintas (*handover* heterogêneo ou vertical), mais especificamente, entre redes Wi-Fi e UMTS (rede celular 3G), embora também tenha sido analisado um *handover* homogêneo (ou horizontal), entre mesma tecnologia. Nós analisamos o uso do evento “Link Going Down” do MIH. Este evento avisa a aplicação que a conexão de rede atual será perdida.

Para realizar a simulação, nós utilizamos um simulador de redes comumente usado em pesquisas científicas, o simulador NS-2 (2015). Esse simulador possibilita virtualizar casos reais de transmissão e envio de dados. Para simular o MIH, nós acrescentamos o módulo NIST (2014) ao NS-2.

Os resultados demonstraram quanto o IEEE 802.21 pode otimizar as trocas de rede. Houve uma redução significativa na perda de dados e, conseqüentemente, aumento da vazão. A quantidade de pacotes perdidos durante a troca diminuiu em 98,4%.

Além desta introdução, a seção 2 apresenta o protocolo IEEE 802.21 e seus serviços. Na seção 3 são apresentados alguns trabalhos relacionados ao gerenciamento de *handovers* e MIH. A seção 4 apresenta a metodologia utilizada neste trabalho, juntamente com o cenário heterogêneo simulado. A seção 5 apresenta os resultados de desempenho do MIH. As conclusões e trabalhos futuros encontram-se na seção 6.

2. Media Independent Handover (MIH)

O protocolo IEEE 802.21 (2009) tem como função facilitar o *handover* heterogêneo, principalmente entre tecnologias móveis, com o intuito de minimizar a perda de pacotes e aumentar o desempenho de redes. Ele deve ser implementado no nó móvel e também

na estação base ou controlador. O protocolo cria uma camada intermediária entre a camada de redes e a camada de enlace, denominada MIHF – Media Independent Handover Function, à qual implementa três serviços: o *Media Independent Command Service* (MICS), o *Media Independent Information Service* (MIIS) e o *Media Independent Event Services* (MIES), como pode ser visto na Figura 1.

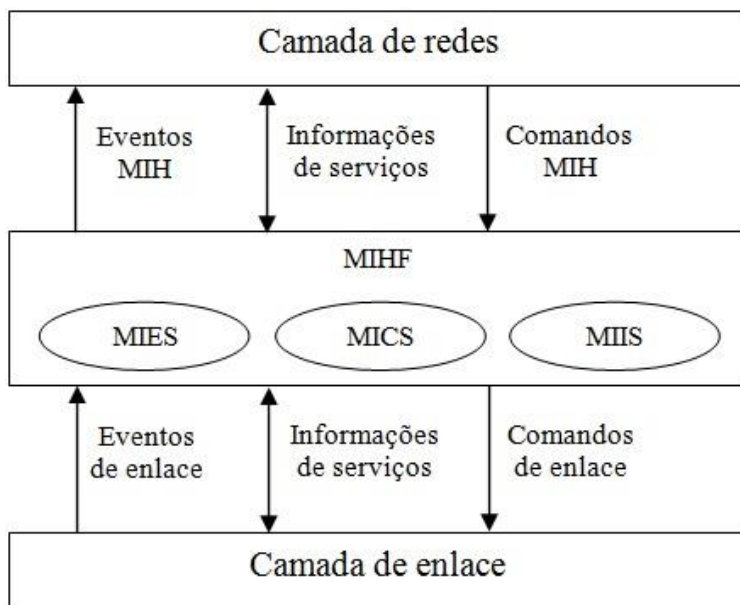


Figura 1 Arquitetura em camadas do IEEE 802.21 (MIH)

Na camada de rede e acima são implementados os MIH Users, programas que se inscrevem nos eventos do MIH, enviam comandos e recebem informações do serviço de informações (MIIS). O serviço de eventos (MIES) detecta eventos como, por exemplo, queda na conexão, nova rede detectada, degradação do sinal, entre outros, e alerta os MIH Users inscritos naquele evento. O serviço de informações (MIIS) fornece informações sobre as redes disponíveis para os MIH Users. O serviço de comandos (MICS) possibilita o envio de comandos entre as camadas do MIH como, por exemplo, inscrição em eventos e configuração de limites (*thresholds*).

Um dos eventos gerados pelo MIH é o *Link Going Down*, o qual avisa que as condições do enlace estão se degradando e a perda de conexão é iminente. Desta forma, um MIH User (gerenciador de conexões, por exemplo) fica sabendo antecipadamente que a conexão será perdida. Assim, prepara a troca de rede antes que ela ocorra, diminuindo o tempo de *handover* e a quantidade de pacotes perdidos. Os efeitos deste evento são analisados neste trabalho.

3. Trabalhos relacionados

Em um trabalho anterior (Augusto, 2012), nós realizamos simulações de *handovers* homogêneos, ou seja, entre redes utilizando a mesma tecnologia de acesso. Foram medidos o número de *handovers*, a quantidade de bytes recebidos, e quanto o usuário pagou pelo acesso. O ambiente era formado de várias antenas Wi-Fi e não foi utilizado o protocolo IEEE 802.21 devido à homogeneidade do ambiente.

Relacionado ao trabalho atual, vários autores também simularam o IEEE 802.21 em um ambiente heterogêneo. Iqbal(2011) faz uso do IEEE 802.21 para a troca entre tecnologias de rede, utilizando como instrumento de testes o simulador de redes NS-2 com o pacote NIST (2014). Nos testes, os autores comprovam a eficiência do protocolo com *handover* heterogêneo, onde a vazão aumentou em 49,8%. Porém, o ambiente analisado continha redes Wi-Fi e WiMAX. Outro trabalho parecido envolvendo cenário Wi-Fi e WiMAX é encontrado em Lim(2009).

Santos (2011) analisou o *handover* heterogêneo entre redes WiMAX e UMTS usando IP móvel e MIH através do simulador NS-2. Os autores analisaram a vazão, o atraso e o *jitter* médio, e compararam os resultados com os requisitos de Qualidade de Serviço especificados pelo 3GPP (3rd Generation Partnership Project).

Em Machan (2008), os autores simulam um cenário com antenas Wi-Fi e UMTS e comparam a utilização do MIH versus a não utilização. Os autores se concentram apenas na troca da rede Wi-Fi para a UMTS, medindo perda de pacotes e tempo perdido com a troca de rede. Os autores utilizaram o simulador NS-2 para a análise. Nosso trabalho analisou a troca em ambos os sentidos, além do *handover* homogêneo Wi-Fi para Wi-Fi.

Alguns trabalhos apresentam esquemas para otimização de *handover* heterogêneo utilizando MIH, como proposto por Augusto (2013), Wu (2009), Cicconetti (2011) e também fazem a análise de desempenho da utilização do protocolo.

Algumas implementações do IEEE 802.21 estão disponíveis. Além da implementação do NIST (2014) para o simulador NS-2, Corujo (2011) apresenta o ODTONE, uma implementação do protocolo que tem código-fonte aberto e é independente de sistema operacional. Outra implementação é a OpenMIH de Lopez (2009).

Para diferenciar as trocas de rede entre tipos diferentes, Mapp (2009) classifica os *handovers* entre imperativos e alternativos. *Handovers* imperativos são obrigatórios para evitar queda severa de desempenho ou queda na conexão, e do tipo alternativo ocorrem por opção (preço, incentivos, contexto, serviços oferecidos ou preferência do usuário). Mapp também classifica as trocas entre *hard handover*, quando a conexão atual é finalizada antes do dispositivo se conectar a outra rede, e *soft handover*, quando a conexão atual é finalizada após o dispositivo se conectar a outra rede. Nós utilizamos esta classificação para melhor comparar as trocas realizadas nas simulações.

4. Metodologia e cenário analisado

Nós realizamos a simulação utilizando-se o simulador de redes NS-2 (2015), juntamente com a implementação do IEEE 802.21 (MIH) e da tecnologia UMTS desenvolvida pelo NIST (2014). As tecnologias de redes escolhidas para realizar a simulação foram Wi-Fi e UMTS, devido à sua atual predominância de acesso a dados no Brasil (Opensignal, 2015).

O cenário é apresentado na Figura 2. As duas antenas Wi-Fi (Wi-Fi1 e Wi-Fi2) estão instaladas ao longo de uma avenida nas posições 100 metros e 180 metros, respectivamente, com raio de alcance de 50 metros. Desta forma, tem-se uma sobreposição de 20 metros entre ambas, viabilizando-se assim a análise de um *hard*

handover homogêneo imperativo. O nó P2r é o roteador controlador das antenas Wi-Fi. O sinal da antena UMTS (Base Station) cobre todo o cenário. Os nós RNC, SGSN e GGSN são, respectivamente, *Radio Network Controller*, *Serving GPRS Support Node*, e *Gateway GPRS Support Node* do domínio UMTS. As duas tecnologias da forma como estão posicionadas no cenário, viabilizam um *soft handover* heterogêneo alternativo (UMTS → WiFi1) e um *soft handover* heterogêneo imperativo (WiFi2 → UMTS), além de um *hard handover* homogêneo imperativo (WiFi1 → WiFi2). As velocidades dos enlaces estão anotadas na figura.

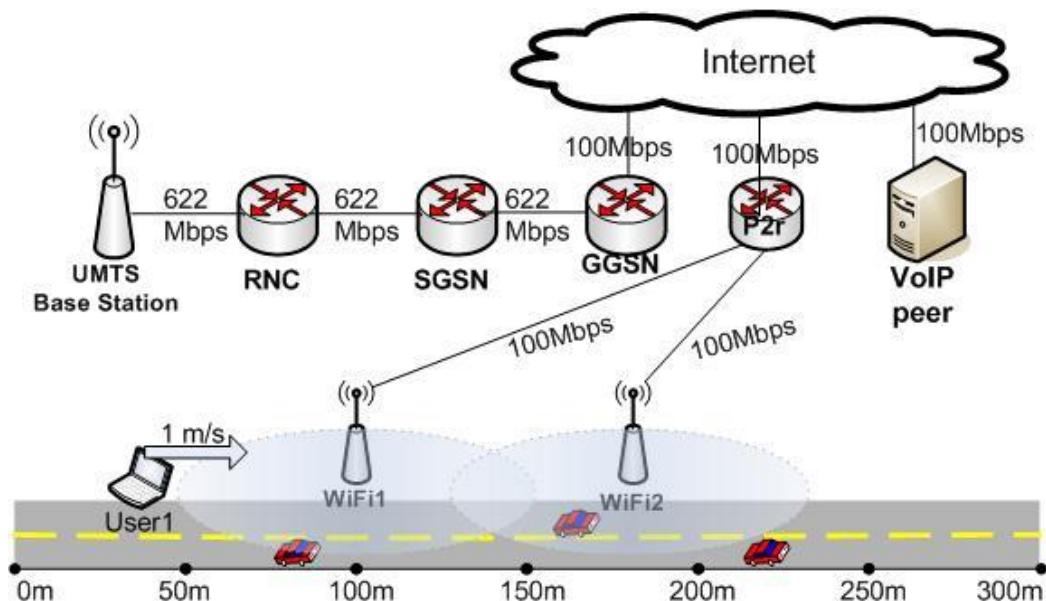


Figura 2 Cenário heterogêneo analisado

Um usuário móvel (User1) percorre a avenida a uma velocidade de 1 m/s, simulando uma pessoa caminhando, saindo da posição 40 metros aos 10 segundos de simulação até a posição 250 metros, passando pelas três antenas citadas anteriormente. O tempo total de simulação foi de 240 segundos. Neste percurso, o nó móvel primeiramente se conecta à antena UMTS, e em seguida conecta-se à antena WiFi1, depois conecta-se à antena WiFi2, e quando perde o sinal desta última, volta a trafegar os dados pela antena UMTS.

Nós simulamos um tráfego VoIP entre o nó VoIP peer e o usuário móvel (User1). Os pacotes gerados foram do tipo UDP a uma taxa constante de bits de 64 kbps (codec G.711). O tráfego teve início juntamente com a movimentação do nó móvel aos 10 segundos, e finalizou após o nó móvel parar de se mover, aos 240 segundos. No momento que este tráfego se inicia, o dispositivo já se encontra conectado à rede UMTS. Durante toda a movimentação do nó móvel, o tráfego estava fluindo. Para possibilitar a continuidade do tráfego após a troca de rede, nós utilizamos o protocolo Mobile IPv6 (Perkins, 2011).

5. Resultados

Nós realizamos simulações comparando a geração ou não do evento *Link Going Down* (LGD), comentado na Seção 2. Quando o evento LGD não estava ativado, a quantidade total de bytes recebidos pelo usuário móvel foi de 1663 KB (1703520 bytes) em um

período de quase 4 minutos (230 segundos) trafegando o fluxo VoIP. Na simulação em que o evento LGD estava ativado, a quantidade de bytes recebidos aumentou para 1801 KB (1844960 bytes) no mesmo período, ou seja, um acréscimo de 8,3%.

O acréscimo na quantidade total de bytes em si pode não ser tão significativo. Porém, considerando a aplicação em questão (VoIP), com suas exigências de Qualidade de Serviço, em dois dos três *handovers* na simulação em que nós não utilizamos o evento LGD, a conexão VoIP estaria comprometida. Os gráficos de vazão são apresentados na Figura 3.

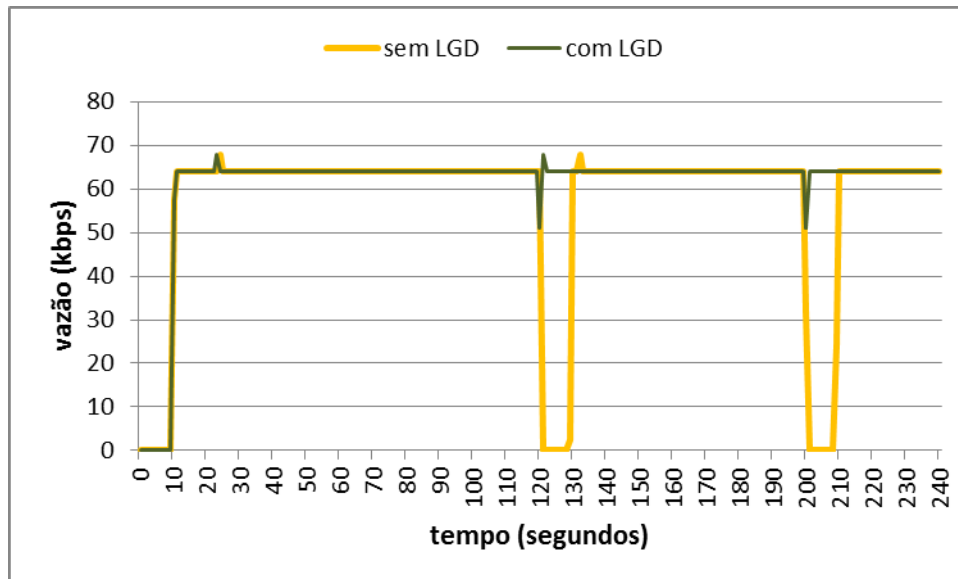


Figura 3 Gráfico dos resultados de vazão

O eixo X na Figura 3 refere-se ao tempo de simulação em segundos e o eixo Y refere-se à vazão do tráfego VoIP em kbps. Os três *handovers* ocorreram aproximadamente aos 22, 120 e 200 segundos. Estas três trocas são detalhadas a seguir.

5.1. *Soft handover* heterogêneo alternativo (UMTS → WiFi1 aos 22 segundos)

No gráfico da Figura 3, pode-se notar o fluxo VoIP iniciando aos 10 segundos de simulação. Neste momento o nó móvel já se encontrava conectado à antena UMTS. Neste mesmo momento, o nó móvel inicia sua caminhada a uma velocidade de 1 m/s. Aproximadamente aos 22 segundos de simulação, ocorre a troca entre a antena UMTS para a antena WiFi1. Enquanto os dados ainda trafegavam pela rede UMTS, o dispositivo detectou o sinal da antena WiFi1 e a conexão foi estabelecida, portanto troca do tipo *soft handover* heterogêneo. Como o dispositivo realizou a troca por uma questão de preferência, pois a implementação do simulador NS-2 prioriza as antenas Wi-Fi devido à maior velocidade de transmissão, o *handover* é classificado como sendo do tipo alternativo.

Em ambos os casos, sem LGD e com LGD, o tempo de *handover* foi zero pois o nó móvel conectou à rede Wi-Fi enquanto ainda trafegava dados normalmente pela interface UMTS. Conseqüentemente, a perda de pacotes também foi zero.

5.2. *Hard handover* homogêneo imperativo (WiFi1 → WiFi2 aos 120 segundos)

Aproximadamente aos 120 segundos, a interface Wi-Fi perde sinal da antena WiFi1 pois já estava a 50 metros de distância, obrigando a interface a conectar-se com outra antena. Consultando as mensagens de log do simulador, é possível acompanhar cada passo do processo de *handover*. O processo é detalhado a seguir.

No caso da simulação sem LGD, aos 120 segundos a interface Wi-Fi perde sinal com a antena WiFi1. Aos 128 segundos o prefixo Mobile IP da rede WiFi1 expira e o dispositivo inicia processo de *scanning*, ou seja, procura outras antenas disponíveis. Aos 129 segundos a interface Wi-Fi recebe resposta ao escaneamento da antena WiFi2 e a conexão é estabelecida. Aos 131 segundos, o agente MIPv6 recebe *acknowledgement* para redirecionamento do fluxo VoIP pela antena WiFi2. O tempo total sem recebimento de dados foi de 11,1 segundos. Neste intervalo de tempo sem receber dados, a quantidade de pacotes perdidos em função da troca de rede foi de 445 pacotes.

No caso da simulação com LGD, antes da interface perder o sinal da WiFi1, aos 117 segundos o protocolo MIH gerou o evento *Link Going Down*. Este evento se repete a cada 0,02 segundos e, a cada repetição, é gerada uma probabilidade de perda de conexão que inicia em 0%. Quando esta probabilidade alcança 81%, aos 119 segundos, a interface inicia processo de *scanning*. Logo em seguida já recebe resposta da antena WiFi2. Com isso a interface desconecta da WiFi1 e conecta-se à WiFi2. Aos 120 segundos, o agente MIPv6 recebe *acknowledgement* para redirecionamento do fluxo VoIP pela antena WiFi2. O tempo de *handover* foi de 0,8 segundos. Devido ao menor tempo para realizar a troca, a quantidade de pacotes perdidos diminuiu para apenas 7 pacotes.

5.3 *Soft handover* heterogêneo imperativo (WiFi2 → UMTS aos 200 segundos)

Aproximadamente aos 200 segundos, a interface Wi-Fi perde sinal da antena WiFi2 pois já estava a 50 metros de distância. No caso da simulação sem LGD, aos 199 segundos a interface Wi-Fi perde sinal com a antena WiFi2 e alguns pacotes começam a ser perdidos. Aos 208 segundos o prefixo da rede WiFi2 expira, o dispositivo inicia processo de *scanning* e o agente MIPv6 recebe *acknowledgement* para redirecionamento do fluxo VoIP pela antena UMTS. O processo de *scanning* continua até o final da simulação, sem obter resposta de antenas Wi-Fi.

No caso da simulação com LGD, aos 197 segundos o protocolo MIH gerou o evento *Link Going Down*, com probabilidade de perda de conexão inicialmente de 0%. Aos 199 segundos, a probabilidade alcança 81% e o agente de *handover* do MIH decide redirecionar o fluxo VoIP para a interface UMTS. No mesmo segundo, o agente MIH gera comando para iniciar *scanning* e o agente MIPv6 recebe *acknowledgement* para redirecionamento do fluxo VoIP pela antena UMTS. Logo em seguida, o evento *Link Going Down* indica probabilidade de 92% de perda de conexão. Com isso, a conexão Wi-Fi com a antena WiFi2 é derrubada. O processo de *scanning* continua até o final da simulação e o fluxo VoIP continua pela interface UMTS.

Em ambos os casos, sem e com LGD, o tempo de *handover* foi zero pois a interface UMTS já estava conectada. Porém, na simulação sem LGD o fluxo VoIP ficou 8,9 segundos sem recebimento de dados pois o *handover* demorou a ser percebido pela

camada de rede, apenas quando houve a expiração do prefixo é que o fluxo foi redirecionado de interface. Na simulação com LGD, como houve a notificação do evento *Link Going Down*, o dispositivo se adiantou à perda da conexão e redirecionou o fluxo para a interface UMTS antecipadamente, evitando consideravelmente a perda de pacotes. Neste *handover*, na simulação sem LGD, um total de 453 pacotes foram perdidos. Com a geração do evento *Link Going Down*, esta quantidade caiu para apenas 7 pacotes.

5.4 Repetição das simulações

O simulador NS-2 utiliza geração de números aleatórios, cuja “semente” (seed) foi inicialmente fixada no valor 2 para ambas as simulações (com e sem LGD) para permitir uma comparação mais confiável entre ambos. Após a análise, comparação e geração dos gráficos de vazão, nós realizamos novamente as simulações, porém, utilizando seed aleatória. As simulações sem LGD resultaram em um total de bytes recebidos de 1699520, 1728320 e 1713600 bytes. Ou seja, o valor inicial 1703520 bytes ficou dentro do intervalo destes outros valores. As simulações com LGD resultaram na mesma quantidade de bytes recebidos que a inicial (1844960 bytes), provavelmente porque a troca é feita antes de haver depreciação significativa do sinal.

Os gráficos de vazão gerados nestas últimas simulações são parecidos com os gráficos gerados inicialmente, com os dois vales aos 120 e 200 segundos, em que a segunda e terceira troca foram feitas.

6. Conclusões e trabalhos futuros

Atualmente, a utilização de dispositivos móveis com várias interfaces de rede e a oferta de redes sem fio de diferentes tecnologias demanda o aperfeiçoamento das trocas de rede para possibilitar que o usuário aproveite este cenário para permanecer sempre conectado da melhor forma possível (*Always Best Connected*). O protocolo IEEE 802.21 (MIH - *Media Independent Handover*) tem como objetivo possibilitar as trocas de rede entre domínios diferentes e, principalmente, entre tecnologias de acesso diferentes.

O evento *Link Going Down* do IEEE 802.21 ocorre quando a interface de rede percebe que a força do sinal está se degradando, e avisa as camadas superiores (módulo de gerenciamento de mobilidade e aplicativos) que a conexão será perdida. Com isso, os aplicativos podem se antecipar ao *handover*, preparando a troca de rede antes de haver perdas significativas de dados.

Este trabalho simulou a utilização do evento *Link Going Down* em um cenário heterogêneo envolvendo redes Wi-Fi e UMTS (rede celular 3G), na qual um usuário móvel percorre uma trajetória passando por algumas antenas e realizando *handovers*. As simulações mostraram como o protocolo IEEE 802.21 pode auxiliar nas trocas de rede e também o quanto o evento *Link Going Down* pode otimizar estas trocas. Houve uma redução significativa na perda de dados e, conseqüentemente, aumento da vazão, evitando desta forma que um fluxo VoIP fosse interrompido. A grande vantagem deste evento é que os aplicativos e/ou um módulo gerenciador de *handover* pode antecipar a conexão com outra antena, diminuindo o tempo de troca.

Neste trabalho, o dispositivo procurou por redes disponíveis próximas ao usuário, quando necessário, através de *scanning*. Os autores sugerem, como trabalhos futuros, a utilização do serviço de informações (MIIS) do IEEE 802.21 e a implementação de alguma política de decisão de acesso em um cenário com vários provedores. Também, utilizando as implementações disponíveis do MIH, citadas na seção Trabalhos Relacionados, é possível montar um *testbed* com várias antenas e o MIH sendo executado no nó móvel e na rede.

7. Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina – FAPESC pelo apoio financeiro por meio do Termo de Outorga N° 2014TR3656, Edital Chamada Pública FAPESC N° 04/2014 – Apoio à infraestrutura para grupos de pesquisa da UDESC.

Referências

- Augusto, M. e Moreira, E. S. (2012), “Access on Demand on Wi-Fi Networks: the Impact of Dynamic Switching of Providers for the Mobile User”, Workshop em Desempenho de Sistemas Computacionais e de Comunicação (Wperformance 2012), Anais do XXXII Congresso da Sociedade Brasileira de Computação, Julho.
- Augusto, M. E., Vanni, R. M. P., Guardia, H. C., Aiash, M., Mapp, G. e Moreira, E. S. (2013), “MYHand: a Novel Architecture for Improving Handovers in NGNs”, The Ninth Advanced International Conference on Telecommunications - AICT, Roma. Proceedings of the Advanced International Conference on Telecommunications, v. 1. p. 211-218, Junho.
- Ciconetti, C., Galeassi, F., e Mambrini, R. (2011), “A software architecture for network-assisted handover in IEEE 802.21”, Journal of Communications, vol. 6, no. 1, pp. 44–55.
- Corujo, D., Guimaraes, C., Santos, B. e Aguiar, R. L. (2011), “Using an open-source IEEE 802.21 implementation for network-based localized mobility management”, IEEE Communications Magazine, vol. 49, no. 9, pp. 114-123, doi: 10.1109/MCOM.2011.6011742, Setembro.
- IEEE 802.21 Institute of Electrical and Electronics Engineers (2009), “IEEE Standard for Local and Metropolitan Area Networks – Part 21: Media Independent Handover Services, IEEE Std 802.21-2008”, DOI: 10.1109/IEEESTD.2009.4769367, Janeiro.
- Iqbal, F. e Javed, M. Y. (2011), “Performance Analysis of MIH based Multihoming Approach of Vertical Handover in Heterogeneous Wireless Networks”. International Journal of Future Generation Communication & Network, Vol. 4 Issue 4, p103, Dezembro.
- Lim, S., Kim, D., Suh, Y. e Won, J. (2009), “Implementation and performance study of IEEE 802.21 in integrated IEEE 802.11/802.16e networks”, Computer Communications, pp. 134-143.
- Lopez, Y. e Robert, E. (2009), “OpenMIH, an Open-Source Media-Independent Handover Implementation and Its Application to Proactive pre-Authentication”,

- Mobile Networks and Management: First International Conference, MONAMI 2009, pp. 14-25, DOI 10.1007/978-3-642-11817-3_2, Outubro.
- Machan, P., Swewin, S., e Wozniak, J. (2008), “Performance of mobility support mechanisms in a heterogeneous UMTS and IEEE 802.11 network offered under the IEEE 802.21 standard”, International Conference on Information Technology.
- Mapp, G., Shaikh, F., Aiash, M., Vanni, R.M. P., Augusto, M. e Moreira, E. S. (2009), “Exploring efficient imperative handover mechanisms for heterogenous wireless”, Proceedings of the 12th International Conference on Network-Based Information Systems NBIIS, p. 286-291, Agosto.
- NIST – National Institute of Standards and Technology (2014), “Seamless and Secure Mobility Tool Suite (NIST Mobility)”, Disponível em http://www.nist.gov/ctl/wireless-networks/ssm_tools.cfm, Outubro.
- NS-2 Network Simulator (2015), Disponível em <http://nslam.isi.edu/nslam>, Outubro.
- OpenSignal (2015), OpenSignal Site, Disponível em: <http://opensignal.com>, Dezembro.
- Perkins, C. (2011), “Mobility Support in IPv6 – RFC6275”, ISSN 2070-1721.
- Santos, W. P., Reis, S. O., Nogueira, R. S. e Figueiredo, F. L. P. D. (2011), “Modelo de handover vertical suave entre redes WiMAX e UMTS”, XVI Workshop de Gerência e Operação de Redes e Serviços (WGRS), Campo Grande, Mato Grosso do Sul.
- Wu, J., Yang, S., e Hwang, B. (2009), “A terminal-controlled vertical handover decision scheme in IEEE 802.21-enabled heterogeneous wireless networks”, International Journal of Communication Systems, vol. 22, no. 7, pp. 819–834, Julho.

Análise de Desempenho de Brokers MQTT em Sistema de Baixo Custo

Andrei B. B. Torres, Atslands R. Rocha, José Neuman de Souza

¹Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat)
Universidade Federal do Ceará (UFC) – Fortaleza – CE – Brazil

andreibosco@great.ufc.br, atslands@ufc.br, neuman@ieee.org

Abstract. *This paper presents a performance analysis (CPU usage, memory consumption and message throughput) of MQTT brokers in a low-cost hardware, the Raspberry Pi 2 Model B. The objectives of the analysis are to ascertain which MQTT broker implementation is best suited to the limitations of the hardware and to verify if the Raspberry Pi 2 is actually able to function as a gateway in a sensor and actuator network for the Internet of Things (IoT). The results showed that the Raspberry Pi 2 can handle large numbers of connections and that the implementation in Erlang (eMQTT) obtained the best results in data throughput, while the implementation in C (Mosquitto) obtained the lowest CPU load and memory consumption.*

Resumo. *Este artigo apresenta uma análise de desempenho (uso de CPU, consumo de memória e envio de mensagens) de brokers MQTT em um hardware de baixo custo, o Raspberry Pi 2 Modelo B. Os objetivos da análise são averiguar qual implementação de broker MQTT é a mais adequada às limitações do hardware e se o Raspberry Pi 2 é realmente capaz de funcionar como gateway de uma rede de sensores e atuadores para a Internet das Coisas (IoT). Os resultados mostraram que o Raspberry Pi 2 consegue lidar com grandes números de conexões, sendo a implementação em Erlang (eMQTT) a que obteve melhor resultado em vazão de dados, enquanto a implementação em C (Mosquitto) apresentou a menor carga de processamento e consumo de memória.*

1. Introdução

A Internet das Coisas (IoT - *Internet of Things*) é um conceito de tornar a Internet e a comunicação entre objetos ou ‘coisas’ pervasiva, onde eles são capazes de interagir e de cooperar entre si para alcançar um objetivo [Singh et al. 2014]. Por meio de sensores e atuadores embutidos em tais “coisas inteligentes” seremos capazes de receber dados e de controlá-los remotamente. Algumas áreas com grande potencial para aplicação da IoT são: domótica, saúde, cidades inteligentes, escritórios, comércio, uso militar, dentre outros. A IoT possui um potencial comercial entre 3,5 a 11,1 trilhões de dólares, podendo alcançar entre 25 a 50 bilhões de dispositivos até 2025 [Manyika et al. 2015].

Para tornar a IoT viável e permitir a conexão de centenas de milhares de coisas, é necessário que tais coisas sejam de baixo custo, o que implica baixa capacidade de processamento, armazenamento e comunicação. Deve-se considerar que, ao contrário de *smartphones*, em que o usuário possui em média uma unidade, as coisas inteligentes irão permear o ambiente. Dependendo da limitação de recursos, em alguns cenários torna-se

necessário um dispositivo que funcione como *gateway* de acesso à Internet, repassando os dados das coisas para o usuário final ou até mesmo para outras coisas. As plataformas de hardware de baixo custo são fortes candidatas para preencher este espaço de *gateway* na IoT, e o Raspberry Pi¹ é um dos mais famosos dentre elas. O projeto Raspberry Pi é um projeto educacional britânico cujo foco era criar um hardware simples e acessível para crianças, além de motivar o aprendizado de eletrônica e computação. Outras plataformas de baixo custo são: Beaglebone Black², Banana Pi³, MinnowBoard MAX⁴, dentre outras. O Raspberry Pi 2 foi escolhido para este trabalho, dentre as plataformas citadas, devido a facilidade de compra do equipamento no Brasil, menor preço em comparação às plataformas similares⁵ e a existência de uma ampla gama de acessórios e componentes.

Os protocolos para a comunicação entre os componentes da IoT devem lidar com fatores como baixa largura de banda, alta latência e instabilidade da comunicação. Alguns protocolos foram criados exatamente para lidar com tais fatores: CoAP (*Constrained Application Protocol* [Shelby et al. 2014]), MQTT (*Message Queuing Telemetry Transport* [MQTT Version 3.1.1 2014]), WAMP (*Web Application Messaging Protocol* [WAMP Draft 2 2015]), dentre outros. Neste artigo, o protocolo MQTT foi adotado para a realização dos experimentos, devido à ampla gama de *brokers* implementados em diferentes linguagens e à crescente adoção no mercado. A Seção 2 aborda o protocolo MQTT mais detalhadamente e o motivo de ter sido escolhido para este experimento.

Os objetivos deste artigo são averiguar a viabilidade de uso do Raspberry Pi 2 modelo B como *gateway* de uma rede de sensores e atuadores para IoT, tendo que lidar com um grande número de clientes e de mensagens sendo enviadas em modo QoS 0⁶, além de determinar qual *broker* MQTT é a implementação mais adequada a este hardware.

O restante deste trabalho está dividido da seguinte forma: a Seção 2 apresenta o protocolo MQTT e os *brokers* estudados. A Seção 3 apresenta trabalhos relacionados à análise de desempenho de protocolos e de dispositivos voltados à comunicação machine-to-machine (M2M) e/ou à IoT. A Seção 4 aborda a análise de desempenho em si, apresentando os métodos, os materiais e a metodologia utilizada. A Seção 5 apresenta e analisa os resultados encontrados, e a Seção 6 apresenta uma análise geral com base nos resultados discutidos e possíveis trabalhos futuros.

2. Protocolos de Comunicação para IoT

A IoT ainda está em rápida evolução e está em aberto qual protocolo de comunicação se tornará o padrão do mercado. Dentre os diversos protocolos existentes (MQTT, CoAP, XMPP, SNMP, WAMP) que podem ser usados para IoT, este artigo focou no protocolo MQTT, o qual tem sido largamente adotado por diversas empresas, seja com foco no uso na IoT ou apenas como um protocolo de comunicação. Um exemplo fora do contexto de IoT é a adoção do MQTT pelo Facebook como protocolo de comunicação de

¹<https://www.raspberrypi.org/> - Acesso em 06/04/2016.

²<http://beagleboard.org/black> - Acesso em 06/04/2016.

³<http://www.bananapi.org/> - Acesso em 06/04/2016.

⁴<http://www.minnowboard.org/meet-minnowboard-max/> - Acesso em 06/04/2016.

⁵De acordo com pesquisa de preço realizada online pelos autores em 15/05/2016.

⁶MQTT possui 3 níveis de garantia de entrega de mensagem (*QoS - Quality of Service*), que vai de 0 (nenhuma garantia de entrega) a 2 (garantia de entrega sem duplicidade)

seu sistema de *instant messaging* [Zhang 2011], e dentro do contexto de IoT, a Amazon adotou MQTT, HTTP e Websockets como protocolos padrões em sua plataforma AWS IoT [Amazon Web Services 2016]. Segundo [Skerrett 2015], diretor de marketing da fundação Eclipse, “me parece que o MQTT se tornou o padrão a ser suportado por qualquer provedor sério de soluções para IoT”⁷. Portanto, a análise de implementações do protocolo MQTT é bastante relevante no cenário atual.

2.1. MQTT

Criado em 1999 pela IBM, MQTT (*MQ Telemetry Transport*), é um protocolo aberto de mensagens projetado para comunicação M2M, na qual deve lidar com alta latência, instabilidade na comunicação e baixa largura de banda. O protocolo MQTT foi padronizado pelo OASIS em 2013 e atualmente está na versão 3.1.1 [MQTT Version 3.1.1 2014], sendo livre de *royalties* desde 2010.

O protocolo MQTT adota o protocolo TCP e o padrão de mensagens *publisher/subscriber* (publicador/assinante), onde todos os dados são enviados para um intermediário, chamado *broker*, que se encarrega de enviar as mensagens aos destinatários corretos. Esta estrutura permite desacoplar o produtor do cliente, assim, apenas o endereço do *broker* precisa ser conhecido, possibilitando a comunicação de um para um (*one-to-one* - ver Figura 1), um para muitos (*one-to-many* - ver Figura 1) ou muitos para muitos (*many-to-many* - ver Figura 1). Já no quesito de segurança, até a versão 3.1.1, MQTT não implementa qualquer tipo de criptografia (SSL pode ser utilizado independentemente), existindo apenas um método de autenticação com nome de usuário e senha.

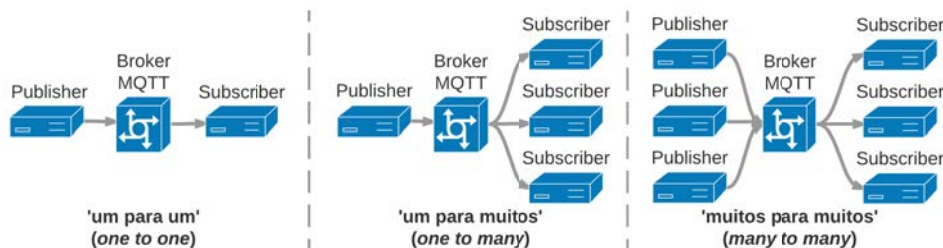


Figura 1. Tipos de distribuição de mensagem suportados pelo protocolo MQTT.
Fonte: os autores.

3. Trabalhos relacionados

Existem outros trabalhos que focam na análise de desempenho, seja de protocolos e serviços ou de hardware e dispositivos, porém nenhum dos artigos encontrados através de um processo de revisão sistemática foca especificamente no uso de hardware de baixo custo com MQTT. A metodologia da revisão sistemática está detalhada na Tabela 1.

O trabalho mais similar ao aqui apresentado foi realizado por [Scalagent 2015], que realiza uma análise comparativa entre diversos *brokers* MQTT, porém com um escopo diferente. O cenário é montado como uma comunicação de muitos-para-muitos (chegando a um máximo de 100 *subscribers* e 100.000 *publishers*), com foco na escalabilidade dos *brokers*, latência no tempo de envio das mensagens e a vazão de entrega, enquanto o presente artigo objetiva principalmente o desempenho do hardware em lidar

⁷Tradução livre dos autores.

Tabela 1. Metodologia da Revisão Sistemática

String de busca	((“mqtt”OR “raspberry pi”) AND (“performance”OR “analysis”OR “simulation”)) AND IoT
Fonte de dados	ACM Digital Library, IEEE Digital Library, Science@Direct, Scopus
Critérios	Inclusão: análise de desempenho Exclusão: survey, não lidar diretamente com MQTT ou Raspberry
Período	2012 - atual
Resultado	113 artigos, 4 relevantes, nenhum com foco em hardware de baixo custo e MQTT

com um elevado número de *subscribers* (máximo de 10.000). O hardware utilizado foi um Intel Core 2 Duo 3.00 GHz com 4 GB de memória, muito superior ao Raspberry Pi 2 utilizado neste artigo, mas também muito mais caro.

3.1. Benchmark de hardware

O artigo de [Kruger and Hancke 2014] realiza uma análise comparativa de diversos dispositivos *off-the-shelf* (“disponíveis na prateleira” - soluções completas, autocontidas): Raspberry Pi 1, BeagleBone, BeagleBone Black. Os objetivos do artigo foram determinar se o uso de soluções de armazenamento de maior desempenho causam alguma influência notável e avaliar os dispositivos como *gateways* CoAP, onde o foco da análise foi o tempo de latência das requisições. Os resultados apontaram que utilizar armazenamento de maior desempenho e mais caro não causa impacto significativo, e o fator mais importante na escolha da solução é o processador.

O mesmo autor possui um artigo mais recente, [Kruger et al. 2015], onde estuda o uso do Raspberry Pi como *gateway* de uma rede de sensores sem fio, utilizando, novamente, dispositivos *off-the-shelf* para monitoramento do consumo de água. O autor conclui que a utilização de dispositivos *off-the-shelf* é viável, reduzindo consideravelmente o tempo do ciclo de desenvolvimento de produtos.

3.2. Benchmark de software

Outros artigos relevantes, mas com foco no software, são os apresentados por [Thangavel et al. 2014], [Kovatsch et al. 2014] e [Collina et al. 2014]. O primeiro realiza uma análise comparativa entre MQTT e CoAP através de um *middleware* desenvolvido pelos autores, utilizando atraso e consumo de banda como métricas. Os resultados experimentais apontam que o desempenho depende das condições da rede, em que mensagens MQTT obtiveram menor atraso em uma rede com pouca perda de pacotes, e maior atraso em rede com elevada perda de pacotes. Também foi identificado que CoAP gera menos tráfego adicional do que o MQTT para garantir a entrega de mensagens.

O segundo artigo apresenta um *benchmark* do *framework* Californium, com foco na criação de um servidor CoAP na nuvem que seja escalável e capaz de lidar com um número massivo de clientes (acima de 100k). O *framework* CoAP Californium obtém uma vazão de dados de 33 a 64 vezes superior à do HTTP.

E o terceiro artigo realiza uma análise de desempenho quantitativa dos protocolos MQTT e CoAP, e analisa o comportamento dos protocolos em diversas situações de rede

(perda de pacote, atraso, dentre outros fatores). No estudo, o MQTT apresentou a melhor vazão de dados, enquanto o CoAP apresentou a menor latência em caso de baixo tráfego e baixa probabilidade de perda de pacote.

4. Experimentos

Nesta Seção são descritos os materiais utilizados e o design experimental dos testes. O objetivo dos experimentos foi avaliar o funcionamento e desempenho do Raspberry Pi 2 e dos *brokers* MQTT em um cenário de alta carga, simulando seu uso como um *gateway* acessível pela internet, respondendo a milhares de requisições de leituras realizadas simultaneamente. Um exemplo de cenário prático é o de um sensor público, como um sensor de temperatura ou de poluição, acessível via web ou por um aplicativo móvel.

4.1. Ambiente

4.1.1. Hardware

Um Raspberry Pi 2 modelo B (especificações na Tabela 2) e uma máquina *desktop* foram utilizados nos experimentos. No *desktop* foi executado o aplicativo Apache JMeter que realizou a carga simulada de *subscribers* (esse processo é explicado na Subseção 4.2).

Tabela 2. Especificações do Raspberry Pi 2 modelo B

Processador	ARM Cortex A7 900MHz Quad-core	Disco	Micro SDHC 16GB UHS Class 1
Memória	1 GB	S.O.	Raspbian Wheezy 2015-05
Ethernet	10/100 Mbps	Alimentação	Micro USB 5V/1.8A

4.1.2. Software: *Brokers* MQTT

Uma ampla gama de servidores MQTT estão disponíveis, tanto em código aberto quanto em software proprietário⁸. Para a execução dos experimentos relatados neste artigo, foram selecionados apenas *brokers* de código aberto e de diversas linguagens de programação (C, Java, Javascript, Erlang) para aferir qual obteria um melhor desempenho (ver Tabela 3).

Tabela 3. Brokers MQTT analisados

Broker	Versão	Linguagem	Implementação	Extras
Mosquitto ⁹	1.4.8	C	MQTT 3.1.1	–
Apache ActiveMQ Apollo ¹⁰	1.7.1	Java	MQTT 3.1	Suporte a STOMP, AMQP, MQTT, OpenWire
Mosca ¹¹	1.1.2	Javascript (node.js)	MQTT 3.1.1	Sem suporte a QoS 2
eMQTT ¹²	0.17.0-beta	Erlang	MQTT 3.1.1	Foco em clusterização
Ponte ¹³	0.0.16	Javascript (node.js)	Não informado	Realiza ponte entre HTTP (REST), MQTT e CoAP.

⁸Listagem mantida pela comunidade disponível em <https://github.com/mqtt/mqtt.github.io/wiki/servers> - Acesso em 06/04/2016.

⁹<http://mosquitto.org/> - Acesso em 13/05/2016.

4.2. Design Experimental

Para testar o desempenho do hardware foi realizada uma carga de trabalho progressiva, onde o número de clientes (*subscribers*) é o fator de controle, de acordo com a seguinte sequência de passos (ver Figura 2):

1. Iniciar captura de dados;
2. Aguardar 60 segundos;
3. Iniciar 200 conexões de *subscribers* a cada 30 segundos;
4. ao atingir a carga máxima de 10 mil *subscribers*, mantê-la durante 180 segundos;
5. Encerrar 25 conexões a cada 1 segundo;
6. Aguardar 5 minutos após encerrar a última conexão;
7. Encerrar captura de dados.

O controle do processo de conexões, conforme apresentado na Figura 2, foi realizado por meio do software Apache JMeter¹⁴, em conjunto com os plugins *Stepping Thread Group*¹⁵, que permite definir uma carga progressiva em degraus, e *mqtt-jmeter*¹⁶, que habilita o JMeter a utilizar o protocolo MQTT.

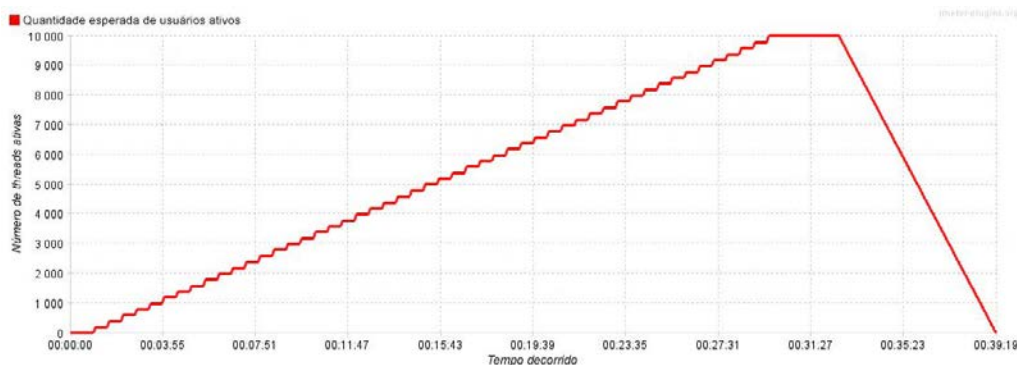


Figura 2. Gráfico da carga de trabalho progressiva adotada

A coleta de dados de desempenho foi realizada por meio do software *collectl*, que permite a coleta de diversas métricas, onde as seguintes foram elencadas:

- CPU: porcentagem de uso do processo (e seus filhos);
- Memória: quantidade de uso do processo (e seus filhos).

Para a coleta de dados de rede, o software *tshark*¹⁷ foi utilizado, gerando um arquivo que pode ser analisado posteriormente utilizando o software *wireshark*¹⁸. Já o papel de *publisher* foi realizado por um *script* bash, executado no próprio Raspberry Pi 2, enviando a mensagem 'hello' em modo QoS 0 (envio sem confirmação de recebimento), sem retenção, a cada 1 segundo. Toda a estrutura de comunicação entre os componentes do experimento está explicitada na Figura 3.

¹⁰<https://activemq.apache.org/apollo/> - Acesso em 13/05/2016.

¹¹<http://www.mosca.io/> - Acesso em 13/05/2016.

¹²<http://emqtt.io/> - Acesso em 13/05/2016.

¹³<https://github.com/eclipse/ponte> - Acesso em 13/05/2016.

¹⁴<http://jmeter.apache.org/> - Acesso em 06/04/2016.

¹⁵<http://jmeter-plugins.org/wiki/SteppingThreadGroup/> - Acesso em 06/04/2016.

¹⁶<https://github.com/tuanhiep/mqtt-jmeter> - Acesso em 06/04/2016.

¹⁷<https://www.wireshark.org/docs/man-pages/tshark.html> - Acesso em 06/04/2016.

¹⁸<https://www.wireshark.org/> - Acesso em 06/04/2016.

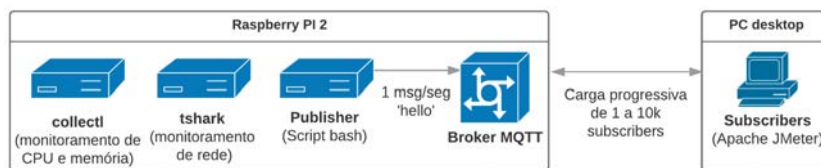


Figura 3. Diagrama de comunicação e monitoramento

Todos os experimentos foram executados 10 vezes após a reinicialização do sistema (*warm boot*), com comandos executados via SSH e na seguinte sequência:

1. Iniciar o *broker* MQTT;
2. Iniciar *script* de publicação de mensagens (*publisher*).
3. Iniciar coleta de dados de rede (*tshark*);
4. Iniciar coleta de dados de desempenho (*collectl*).

Dentre os *brokers* elencados para o experimento, alguns necessitaram de ajustes nas configurações para a realização dos experimentos¹⁹:

- *Brokers* Mosca e Ponte: aumentar o limite de arquivos abertos;
- *Broker* eMQTT: aumentar o limite máximo de processos;
- *Broker* Apollo: aumentar limite máximo de conexões e ajuste de uso de memória.

A Tabela 4 apresenta resumidamente o design experimental dos testes, incluindo as métricas, fatores e carga de trabalho.

Tabela 4. Critérios para análise de desempenho

Critérios	Descrição
Sujeitos	Raspberry Pi 2 e <i>brokers</i> MQTT
Delineamento	Análise comparativa de <i>brokers</i> utilizando uma carga progressiva de clientes
Métricas	Uso de CPU, Consumo de memória, Mensagens enviadas
Parâmetros	CPU e memória
Fatores	Número de conexões (<i>subscribers</i>)
Técnica de avaliação	Medição
Iterações	10 iterações por <i>broker</i>
Carga de trabalho	Carga progressiva em escada com 200 novas conexões a cada 30s. Ao atingir a carga máxima de 10 mil conexões ela é mantida por 180s, e então é iniciado o processo de desconexão de 25 conexões por segundo.
Análise dos dados	Interpretação dos resultados das médias de uso de CPU, memórias e total de mensagens enviadas.

5. Resultados

Com exceção do Apollo, todos os *brokers* conseguiram completar com sucesso as 10 iterações de experimento. Dentre os 5 *brokers* testados, a expectativa era de que aqueles

¹⁹Código disponível online em <https://gist.github.com/andreibosco/b9e70757ed6315d4ea57>

que foram programados em *Javascript via node.js* (uma linguagem interpretada de alto nível) teriam pior desempenho (maior carga de processamento e uso de memória) do que aqueles programados em C, Java ou Erlang (linguagens compiladas). Porém esta hipótese foi refutada, em parte, pelos experimentos.

A Figura 4 apresenta um resumo dos resultados obtidos através de uma média das 10 iterações. Nela é possível notar a similaridade de desempenho dos *brokers* Mosca, Ponte e Mosquitto, e que o Apollo apresenta um resultado excêntrico. Nas próximas subseções cada tópico dos resultados será abordado de maneira mais aprofundada.

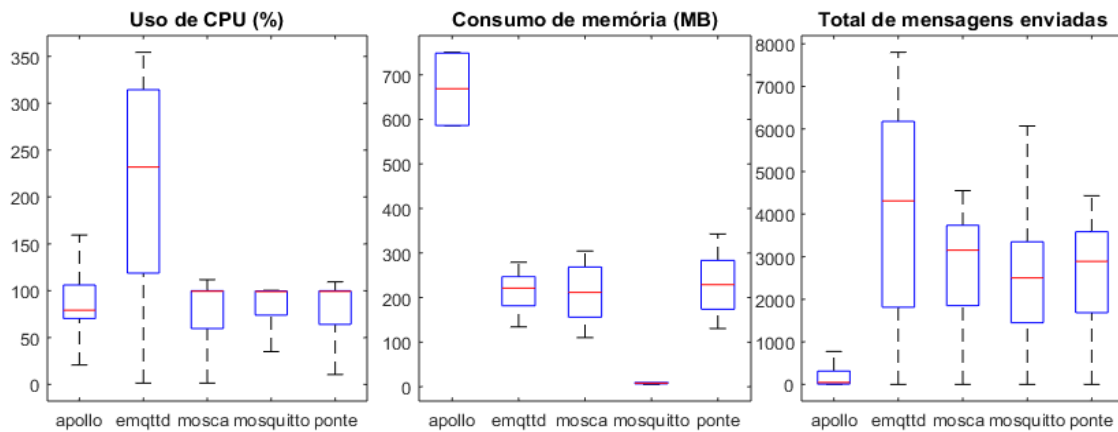


Figura 4. Resumo dos resultados obtidos

5.1. CPU

Conforme apresentado na Figura 5, os *brokers* Mosquitto (C), Mosca (node.js) e Ponte (node.js) apresentaram a menor carga de processamento. Ponte é um *broker* de múltiplos protocolos de comunicação (CoAP, MQTT e HTTP REST), e sua implementação de MQTT é baseado no Mosca, portanto, a similaridade de desempenho já era esperada. O fato mais marcante é o Mosca e Ponte terem obtido desempenho equivalente ao Mosquitto, que é utilizado como *broker* MQTT de referência.

Representando a linguagem Erlang temos o *broker* eMQTT, cujo foco é em clusterização e escalabilidade²⁰. Talvez devido a esse foco, sua arquitetura deve ter sido criada para utilização em máquinas mais robustas do que o Raspberry Pi 2. O *broker* eMQTT conseguiu concluir os experimentos com sucesso, porém obteve a maior carga de processamento.

Já para o *broker* programado em Java, os resultados foram problemáticos. O Apollo não chegou a concluir todos os experimentos, mesmo após diversas tentativas de ajustes nas configurações do *broker*. Na Subseção 5.2, fica claro o momento e o motivo de travamento do *broker*. No *log* de comunicação entre o *broker* e o JMeter (representando os *subscribers*), foi reportada uma série de erros *Uncaught exception: java.lang.NullPointerException*. Portanto, apesar de algumas iterações de experimento terem concluído com sucesso, seu funcionamento foi errático, e na Subseção 5.3 fica claro o problema no envio de mensagens.

²⁰A linguagem Erlang foca em multiparalelismo e aplicações distribuídas. Mais informações: http://www.erlang.org/download/armstrong_thesis_2003.pdf - Acesso em 06/04/2016.

5.2. Memória

A Figura 6 apresenta o resultado do consumo de memória, e as implementações em node.js novamente obtiveram desempenho similar, tendo o *broker* Ponte uma leve desvantagem, talvez pelo projeto não ser desenvolvido tão ativamente²¹. O *broker* Mosca obteve o segundo menor consumo de memória, chegando a um pico de ~275MB.

No caso do Apollo, ele obteve o maior consumo de memória, chegando a travar em algumas iterações reportando memória insuficiente para execução (*There is insufficient memory for the Java Runtime Environment to continue*). Por padrão, o Apollo inicia com um parâmetro que indica 1GB de memória como o máximo alocável para a máquina virtual Java (JVM - *Java Virtual Machine*), que é exatamente a quantidade existente no Raspberry Pi 2, sendo assim um valor inadequado para a plataforma. Tentou-se repetir os experimentos com um valor menor (700MB), porém o comportamento instável persistiu.

Contrastando com sua carga de processamento, o eMQTT obteve o terceiro menor consumo de memória, apresentando um desempenho próximo ao apresentado pelos *brokers* Ponte e Mosca. Um fato curioso é o aumento do consumo de memória no término do período da carga máxima e início do processo de desconexão dos *subscribers*.

Já o resultado obtido pelo Mosquitto foi surpreendente, chegando a um pico de apenas 9.5MB. Era esperado que a implementação em C obtivesse melhor resultado, porém não era esperado um valor tão díspar ao segundo colocado, eMQTT. Interessante notar na Figura 7 como o consumo de memória do Mosquitto representa claramente a carga progressiva representada na Figura 2.

5.3. Rede

Na Figura 8 são apresentados os resultados de vazão das mensagens (número de pacotes entregues por minuto). Os *brokers* Mosca e Ponte obtiveram não apenas bom desempenho na carga de uso do processador, baixo consumo de memória, mas também bons índices de entrega de mensagens. Mosca e Ponte estão tecnicamente empatados, novamente lembrando que ambos possuem uma base de código similar.

Apesar do Mosquitto ter apresentado um excelente resultado no quesito consumo de memória, ele aparenta ter alcançado um limite de conexões por volta de 15 minutos, onde houve um claro declínio no envio de mensagens. Em comparação, ele obteve apenas 68% do índice de entrega de mensagens do Mosca, o melhor colocado.

O eMQTT apresentou a maior carga de processamento dentre os *brokers* analisados ao alcançar o primeiro lugar em quantidade de mensagens entregues, conseguindo quase o dobro da vazão em comparação ao Mosca e Ponte. Em último lugar temos o Apollo. A quantidade ínfima de mensagens entregues, em comparação com os outros *brokers*, confirma o comportamento errático apontado pelos dados do desempenho de memória e processamento. Devido à geração do gráfico através de mediana e suavização das curvas, o Apollo sequer chega a ser representado.

Quase todos os *brokers*, com exceção do eMQTT, aparentam ter alcançado um limite após aproximadamente 15 minutos de experimento, onde a taxa de envio de mensagens estagnou, chegando até a decair no caso do Mosquitto. Esse tempo equivale a cerca de 4 mil conexões, e é necessário investigar o que causou esse limite.

²¹Projeto Ponte teve 8 *commits* desde o início de 2016, enquanto o projeto Mosca teve 38.

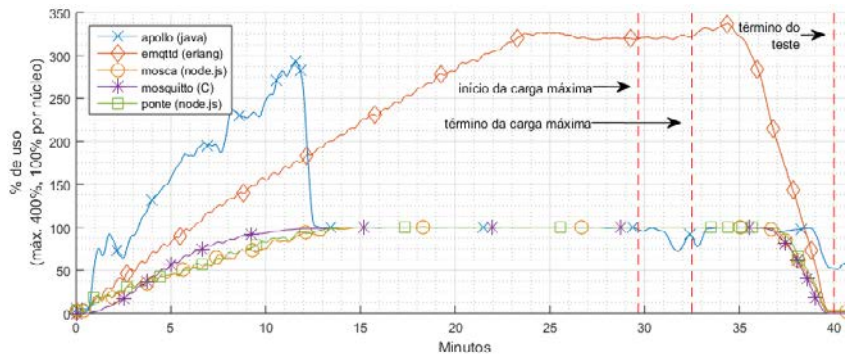


Figura 5. Carga de uso de processador

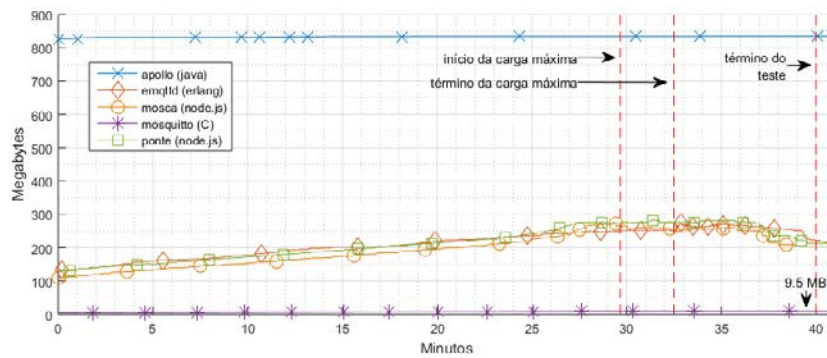


Figura 6. Consumo de memória

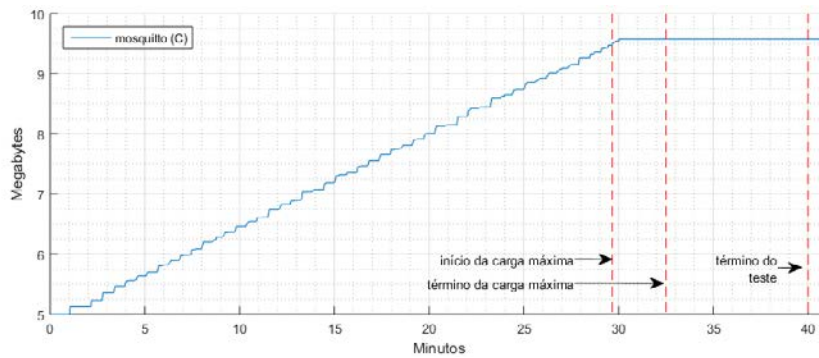


Figura 7. Consumo de memória do *broker* Mosquitto

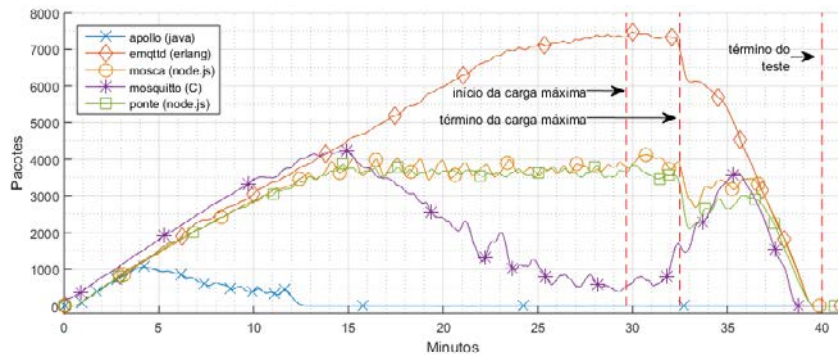


Figura 8. Pacotes por minuto

6. Conclusão

Foi realizada uma avaliação do uso de CPU e consumo de memória de *brokers* MQTT em um hardware de baixo custo, Raspberry Pi 2 Modelo B. Os objetivos foram averiguar a viabilidade de uso de tal hardware como *gateway* para IoT e de determinar qual *broker* MQTT possui a implementação mais adequada a este hardware. Essa avaliação foi realizada por meio de um experimento de comunicação “um para muitos”, com 1 *publisher* se comunicando com 10 mil *subscribers* através de um *broker* instalado no Raspberry Pi 2.

A hipótese inicial de que a implementação em C, Mosquitto, teria o melhor desempenho (menor uso de CPU e menor consumo de memória) foi confirmada pelos resultados obtidos nos experimentos, apresentando consumo de memória em uma ordem de grandeza abaixo do segundo melhor colocado, a implementação em node.js, Mosca. No entanto, melhor desempenho não implicou em melhor vazão de dados, com o Mosquitto atingindo um limite em aproximadamente 4 mil conexões e passando a se comportar erraticamente. Os *brokers* Mosca e Ponte também apresentaram um limite de aproximadamente 4 mil conexões, porém mantiveram uma vazão de dados constante. O *broker* Apollo (Java) não chegou a concluir todos os experimentos por falha de falta de memória, sendo o pior em todos os quesitos analisados. Já a implementação em Erlang, o eMQTT, demonstrou um *trade-off*, com o maior uso de CPU e o terceiro menor consumo de memória (muito próximo ao apresentado pelo Mosca), porém com a maior vazão de mensagens.

Quanto ao hardware, o Raspberry Pi 2 provou ser um hardware capaz de lidar com altos números de requisições, sendo a baixa quantidade de memória o ponto frágil do hardware. Quanto ao software, o *broker* Mosquitto é o mais recomendado caso o número de conexões seja inferior a 4 mil e a necessidade de baixo consumo de memória e baixa carga de processamento sejam primordiais. Porém, caso a necessidade seja alta vazão e grande número de conexões, e onde carga de processamento, e, conseqüentemente, consumo energético, não seja um problema, o eMQTT é a melhor opção.

Como trabalhos futuros, pretende-se realizar experimentos de comunicação do tipo “muitos para muitos”, analisar os motivos que causaram a barreira de quatro mil conexões encontrada nos experimentos, e a comparação do MQTT com outros protocolos, como CoAP, HTTP/2 e WAMP.

Referências

- Amazon Web Services (2016). AWS IoT. <https://aws.amazon.com/pt/iot/>. Acesso em 17/03/2016.
- Collina, M., Bartolucci, M., Vanelli-Coralli, A., and Corazza, G. E. (2014). Internet of Things application layer protocol analysis over error and delay prone links. *2014 7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, pages 398–404.
- Kovatsch, M., Lanter, M., and Shelby, Z. (2014). Californium: Scalable cloud services for the Internet of Things with CoAP. In *2014 International Conference on the Internet of Things (IOT)*, pages 1–6. IEEE.
- Kruger, C. P., Abu-Mahfouz, A. M., and Hancke, G. P. (2015). Rapid prototyping of a wireless sensor network gateway for the internet of things using off-the-shelf compo-

- nents. In *2015 IEEE International Conference on Industrial Technology (ICIT)*, pages 1926–1931. IEEE.
- Kruger, C. P. and Hancke, G. P. (2014). Benchmarking Internet of things devices. In *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, pages 611–616. IEEE.
- Manyika, J., Chui, M., Bisson, P., Woetzel, J., Dobbs, R., Bughin, J., and Aharon, D. (2015). *The Internet of Things: Mapping the value beyond the hype*. *McKinsey Global Institute*, page 144.
- MQTT Version 3.1.1 (2014). Edited by Andrew Banks and Rahul Gupta. OASIS Standard. <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html>. Latest version: <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html>. Acesso em 06/04/2016.
- Scalagent (2015). Benchmark of MQTT servers. Technical Report January.
- Shelby, Z., Hartke, K., and Bormann, C. (2014). The constrained application protocol (coap). RFC 7252, RFC Editor. <http://www.rfc-editor.org/rfc/rfc7252.txt>. Acesso em 06/04/2016.
- Singh, D., Tripathi, G., and Jara, A. J. (2014). A survey of Internet-of-Things: Future vision, architecture, challenges and services. *2014 IEEE World Forum on Internet of Things, WF-IoT 2014*, pages 287–292.
- Skerrett, I. (2015). Case Study MQTT: Why Open Source and Open Standards Drive Adoption. <https://ianskerrett.wordpress.com/2015/03/04/case-study-mqtt-why-open-source-and-open-standards-drives-adoption/>. Acessado em 16/11/2015.
- Thangavel, D., Ma, X., Valera, A., Tan, H.-X., and Tan, C. K.-Y. (2014). Performance evaluation of MQTT and CoAP via a common middleware. In *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–6. IEEE.
- WAMP Draft 2 (2015). The Web Application Messaging Protocol. Technical report. <https://tools.ietf.org/html/draft-oberstet-hybi-tavendo-wamp-02>. Acesso em 06/04/2016.
- Zhang, L. (2011). Building Facebook Messenger. <https://www.facebook.com/notes/facebook-engineering/building-facebook-messenger/10150259350998920>. Acesso em 16/11/2015.

Dinâmica das transações do Bitcoin: uma abordagem quantitativa

Saulo Ricci^{1,2}, Alex Borges³, Helder Luiz⁴, Daniel S. Menasché¹,
Eduardo Ferreira¹

¹Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ – Brasil

²Skry, Palo Alto CA, USA

³Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG – Brasil

⁴Instituto Federal do Norte de Minas Gerais (IFNMG), Teófilo Otoni, MG – Brasil

Resumo. Apesar do crescente interesse em cripto moedas, tanto indústria quanto a academia sentem falta de análises quantitativas desses sistemas. Neste trabalho, caracterizamos o Bitcoin, um dos mais populares sistemas de cripto moedas, sob a ótica quantitativa. Visamos métricas importantes relacionadas a transações, como a probabilidade uma transação ser confirmada e o tempo decorrido para essa confirmação. Nossos resultados mostram que há um número não negligenciável de transações que não são confirmadas 24 horas após serem postadas na rede. Nesse caso, nós observamos uma alta correlação entre a taxa de uma transação, seu volume e a suspeita de que a transação não será confirmada. O trabalho mostra que transações do Bitcoin geralmente são confirmadas em períodos curtos, mas ainda assim, muito acima de tempos usuais de sistemas de cartão de crédito. Finalmente, nós provemos uma análise de nível de atividade simples, comparando um período importante na existência do sistema do Bitcoin –o fechamento do Silk Road– e um período subsequente. Em resumo, concluímos que a rede Bitcoin se apresenta resiliente e não é dependente de hubs altamente centralizados.

Abstract. Despite the growing interest in cryptocurrencies, both industry and academia lack on quantitative analysis of such systems. In this work, we have characterized the cryptocurrency system Bitcoin under a quantitative perspective. We have addressed important transaction metrics, such as the transaction confirmation probability and the time to confirmation, and found a non-negligible number of transactions that are not confirmed 24 hours after being issued. We also observed a high correlation between the transaction fee, value and probability of confirmation. Bitcoin transactions usually are confirmed in short periods, but still much larger than conventional credit card systems. Finally, we have provided a simple analysis of activity level on a specific period of Bitcoin life –the ending of Silk Road– and contrasted it against subsequent periods. In summary, our preliminary results indicate that the Bitcoin network is resilient and do not depend on main hubs.

1. Introdução

Nos últimos anos o interesse por moedas virtuais tem aumentado e atraído a atenção de diversos setores da sociedade. Especificamente, o Bitcoin atualmente se destaca como

solução *de facto* para comércio digital que envolve cripto moedas. Estima-se que milhares de dólares circulem diariamente na rede. Ainda assim, alguns elementos fundamentais do Bitcoin ainda não são plenamente compreendidos. Portanto, uma caracterização das propriedades das suas transações, levando em conta aspectos temporais, torna-se necessária.

Neste trabalho são destilados alguns aspectos fundamentais relacionados às transações que circulam na rede Bitcoin. Para tal, contamos com dados públicos, disponíveis na rede, bem como dados colhidos ao longo de um ano por centenas de monitores instrumentados em diversos pontos do mundo. Nossos monitores obtêm informações que não estão convenientemente acessíveis ao público geral, por exemplo, sobre transações que não foram confirmadas bem como o instante em que cada transação foi publicada na rede Bitcoin. As informações colhidas pelos monitores são enviadas para um banco de dados administrado pela startup *Skry*. A partir delas, podemos inferir propriedades sobre o tempo de espera experimentado pelos clientes, bem como sobre a chance de uma dada transação ser eventualmente confirmada.

Dentre as perguntas que visamos responder neste trabalho, temos:

1. é possível inferir se uma dada transação será efetivamente confirmada ou não, no momento em que ela é primeiramente observada na rede BitCoin?
2. qual a distribuição do tempo de confirmação das transações que são confirmadas?
3. o quão robusta é a rede Bitcoin com relação a remoção de nós da rede e/ou ataques que visem quebra de anonimidade?

Com a finalidade de responder estas perguntas, assumimos que uma transação que não é confirmada em 24 horas é considerada como inválida. Analisando os *datasets* coletados, nossas contribuições, com relação às perguntas anteriores são:

1. verificamos que várias das transações inválidas possuem propriedades que permitem prever, com alto grau de acurácia, que de fato tais transações não serão confirmadas;
2. caracterizamos a distribuição do tempo de confirmação das transações válidas, e parametrizamos distribuições clássicas a fim de que se possa utilizar nossas caracterizações de forma simples e conveniente em trabalhos futuros;
3. identificamos que a rede Bitcoin é robusta com relação a remoção de nós previamente considerados como centrais; mas que certas transações no Bitcoin não seguem os padrões recomendados pelo protocolo oficial, e apontamos potenciais formas de se explorar assimetrias na rede a fim de quebrar anonimidade de transações.

O trabalho segue com a definição de conceitos-chave na seção 2. Os *datasets* coletados para análise são descritos na seção 3. Na seção 4 são apresentados caracterizações de alguns aspectos orgânicos das transações, enquanto na seção 5 são abordados os aspectos que refletem o estado e dinâmica como um todo da rede do Bitcoin. É mostrado um caso específico sobre a robustez da rede na seção 6. Os trabalhos relacionados são levantados na seção 7 enquanto a seção 8 fecha o trabalho com a conclusão e oportunidades futuras de pesquisa nesta linha.

2. Fundamentos

Nesta seção são abordados os principais conceitos para o entendimento do Bitcoin [Narayanan et al. 2016]. Em particular, focamos nos conceitos relevantes para a compreensão do restante do artigo.

Endereço: Para participar da rede e realizar transações, um usuário precisa de uma chave pública (referida como endereço público, ou simplesmente endereço) e uma chave privada.

Transação: Uma transação contém a informação da movimentação de moedas entre endereços. Ela está associada a duas listas de endereços, uma de entrada e uma de saída. Os endereços de entrada (resp., saída) transferem (resp., recebem) moedas. Uma transação típica pode conter, por exemplo, dois endereços de entrada correspondentes a duas carteiras de um mesmo usuário físico e duas saídas - uma correspondente à carteira do destinatário e a outra ao troco.

Grau da transação: O grau de entrada (resp., saída) de uma transação é o número de endereços na lista de entrada (resp., saída). O grau da transação é a soma dos graus de entrada e saída.

Bloco: Consiste de um conjunto de transações. A tarefa dos mineradores é coletar transações pendentes de confirmação, formar blocos a partir destas e validar os blocos formados. O instante em que cada bloco é validado é estampado no próprio bloco.

Mineração: O ato de emitir novos blocos é chamado de mineração. Em dado momento existem diversos mineradores competindo entre si para serem os primeiros a validar o próximo bloco e ganharem as respectivas recompensas em satoshis¹. O tempo para confirmação de transações e validação de blocos será discutido nas seções 4 e 5 deste trabalho.

Bloco confirmado: Um bloco validado por um minerador e efetivado no *blockchain*.

Transação confirmada: Transação que faz parte de um bloco confirmado.

Prova de Trabalho (Proof of Work): O primeiro minerador a validar um bloco recebe uma recompensa. Validar o bloco envolve resolver um desafio criptográfico, e desta forma pode-se dizer que a recompensa é fruto da prova de trabalho (*proof of work*). O fato de o estado atual da rede ser fruto da solução de inúmeros desafios criptográficos é uma das medidas usadas para minimizar ataques, tornando tais ataques muito custosos.

Cadeia de Blocos: Uma sequência (lista encadeada) de blocos de transações confirmadas, considerada como registro de todas as transações realizadas pelo sistema.

Cadeia de Blocos Principal (blockchain): A maior sequência de blocos de transações confirmadas existente na rede. A cadeia de blocos principal é muito longa (acumulando mais de 121 milhões de transações em abril de 2016), e envolveu uma quantidade de trabalho grande para ser criada tendo em vista que a inclusão de cada bloco requer a solução de um desafio criptográfico. Assim, existe um consenso sobre a natureza da maior parte dos blocos da cadeia principal do Bitcoin. Eventualmente, pode existir ambiguidade apenas sobre os últimos blocos incluídos na cadeia. Os blocos que não venham a ser considerados como parte da cadeia principal constituem cadeias órfãs.

Prioridade de Transações: Existe um limite na quantidade de transações que podem ser incluídas por bloco. Para determinar quais transações incluir em cada bloco, os mineradores em geral utilizam uma fórmula padrão de prioridades que envolve a idade da transação, seu tamanho e incentivos na forma de taxas oferecidos por cada transação. Conforme veremos na Seção 5, prioridades afetam de forma significativa o tempo de confirmação

¹O satoshi é a menor unidade da moeda BitCoin. Uma unidade de satoshi equivale a 10^{-8} Bitcoin.

experimentado pelas transações.

Taxa da transação (fee): Gorjeta não obrigatória oferecida pelos proponentes da transação para recompensar os mineradores que venham a confirmar tal transação. Valores maiores de gorjeta aumentam a prioridade na validação da transação, mas valores muito altos podem gerar suspeitas de operações ilícitas (*high-fee spam*).

3. Dados Utilizados

Nesta seção, descrevemos os três *datasets* utilizados neste trabalho.

O primeiro conjunto de dados, usado na seção 4, foi coletado no período de 23-29/05/2015 referente às transações confirmadas e não confirmadas. Os dados são originários de um *streaming* do sistema de nós monitores desenvolvidos pela *Skry* e que estão espalhados por todos os continentes. Cada transação, portanto, é anotada com o *timestamp* do momento em que ela foi primeiro vista por alguns dos nós da rede de monitores. Cada transação, assim que coletada, é armazenada durante 24 horas² e verifica-se se de fato esta foi confirmada usando uma API desenvolvida também pela *Skry*. Ao final desse processo, foram gravadas 697,392 transações, cada uma possuindo informações como o tamanho em KB, o *fee* associado, o volume transacionado, grau de entrada e de saída e o tempo de confirmação.

O segundo *dataset*, usado na seção 5, é uma composição de registros do banco de dados proveniente da empresa *Skry* que compreende transações confirmadas entre 22/09/2015 e 11/01/2016. Para cada transação, temos o instante em que esta foi vista pela primeira vez por algum dos monitores. Este *dataset* contém 12,725,212 transações confirmadas e que foram incluídas no *blockchain*.

O terceiro conjunto de dados, usado na seção 6, é proveniente de informações retiradas do *blockchain*. Foram avaliados os meses de outubro, novembro e dezembro de 2013, fevereiro de 2014, além de todos os meses de novembro do sistema Bitcoin, desde seu início de funcionamento. Após a extração dos dados, realizamos a modelagem das informações deste *dataset* em um banco de dados relacional, tal qual apresentado por [Spagnuolo 2013]

4. Prevendo Confirmações

No Bitcoin, assim que uma transação é gerada, suas informações são imediatamente espalhadas ao longo da rede a fim de serem, eventualmente, efetivadas dentro de um bloco a ser confirmado. Nesse trabalho, definimos como tempo de confirmação de uma transação como a diferença entre os instantes de tempo em que o bloco em que ela está inserida é confirmado por algum minerador e o instante em que a transação foi primeiro vista por um de nossos monitores.

É de valia um modelo que faça previsão para confirmação de transações conforme elas são geradas. De fato, como é de conhecimento comum, o tempo de confirmação no Bitcoin é bem maior que o de transações em cartões de crédito. Assim, a previsão se faz necessária e interessante para instituições que analisam riscos de créditos e que desejam ter relacionamento com a rede do Bitcoin.

²No trabalho uma transação é considerada como não confirmada se não houve sua efetivação no *blockchain* em até 24 horas

Para construção do modelo de previsão, é necessária a caracterização prévia de aspectos que podem ser agrupados dentre aqueles orgânicos e inerentes à própria transação bem como aqueles referentes aos estado e dinâmica de todo o sistema de nós e mine-radores. Assim, nesta seção concentramos nos fatores orgânicos das transações como a quantidade de endereços creditados pela transação, razão entre o *fee* e o volume total pago de satoshis e o próprio tamanho das transações.

A CDF na Figura 1 mostra o cenário geral para as transações recém geradas e que foram confirmadas durante a semana de maio de 2015. Em torno de 90% delas foram confirmadas em até 30 minutos a partir do momento que foram vistas pela primeira vez pelos monitores.³ Portanto para entidades que desejam utilizar o *Bitcoin* como meio de pagamento é desejável que o tempo de confirmação de transações seja o menor possível e portanto pode haver um impacto significativo no nível de negócio, uma vez que as transações podem ser confirmadas em no máximo 30 minutos.

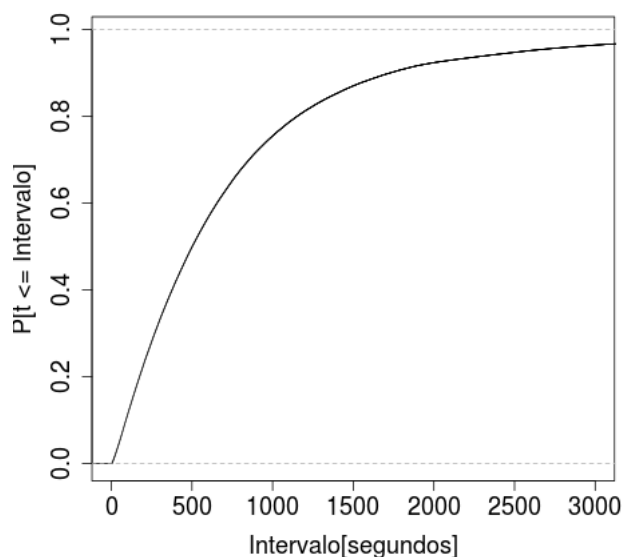


Figura 1. CDF do tempo de confirmação para as transações confirmadas

Por outro lado, no grupo de transações não confirmadas houve casos em que uma transação possui mais de 100 endereços de saída, ou cujo o *fee* foi superior a mais de 20 vezes o valor da própria transação, ou cujo tamanho foi maior do que 20MB. Este cenário é inexistente no grupo de transações confirmadas e pode explicar o fator de *delay* no tempo de confirmação por parte dos mineradores.

Os mineradores optam por não adicionarem transações relativamete grandes, ou que possuem o seu *fee* destoante do valor de referência definido pelo protocolo do Bitcoin (i.e. 0.0001 BTC/KB [Bitcoinwiki 2016]). Atacantes que geram tais transações sugerem induzir interesse ou até mesmo burlar o minerador para que suas transações sejam confirmadas. Estas transações foram reportadas em fóruns tais como em [Bitcointalk 2016] e são conhecidas como *high-fee spam*.

³A precisão do tempo de confirmação está atrelada à cobertura dos monitores que estão espalhado por todos os continentes. Assim temos uma visão da rede.

5. Caracterização do Tempo de Confirmação

Nesta seção, são caracterizados alguns aspectos relacionados com a dinâmica e estado da rede no Bitcoin tais como o tempo entre geração de blocos, tempo entre chegadas de transações e quantidade de blocos ativos no sistema. Além disso propomos um modelo que relaciona tais aspectos com o tempo de confirmação das transações.

A linha de tempo na Figura 2 mostra a dinâmica relacionada com o processo de chegada de transações para serem mineradas, bem como a indicação dos intervalos entre chegada de transações dentro de um bloco. T e B denotam intervalos entre divulgação de transações e entre confirmação de blocos, respectivamente. $T_{j,i}$ denota o intervalo entre a divulgação da i -ésima transação pertencente ao bloco j e a divulgação da transação imediatamente anterior a ela. B_j denota o intervalo de tempo entre a confirmação do bloco B_j e a confirmação do bloco B_{j-1} .

Seja $D_{j,i}$ o tempo de confirmação da i -ésima transação do bloco j . $D_{j,i}$ é igual ao instante de tempo em que o bloco no qual a transação está contida foi confirmado, menos o instante de tempo no qual tal transação foi primeiro divulgada.

O tempo de atividade do bloco j é denotado por S_j . Ele é igual ao intervalo de tempo entre a primeira transação do bloco j ter sido divulgada e o instante em que o bloco j foi confirmado.

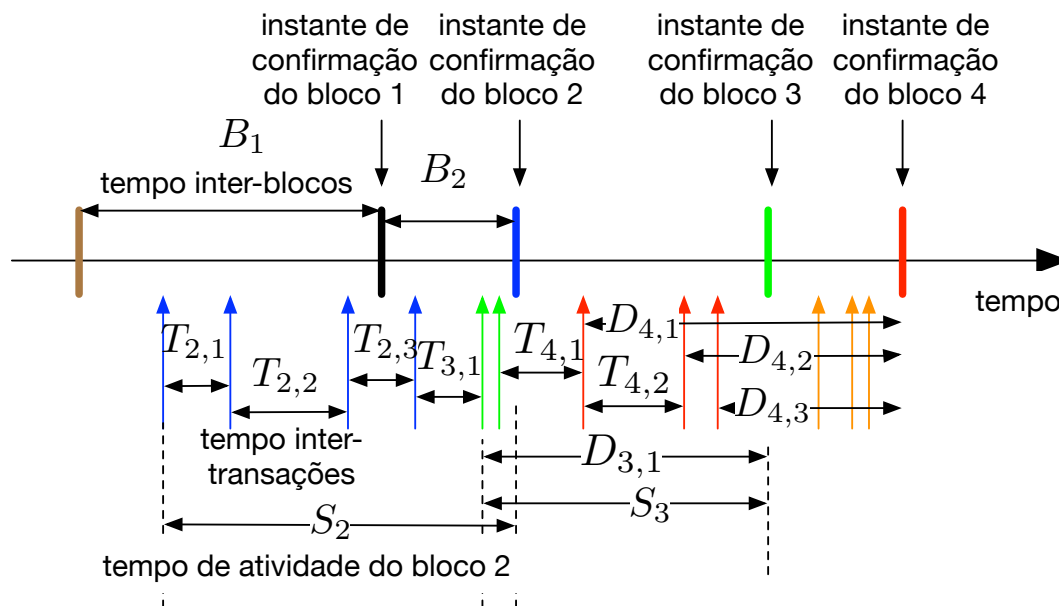
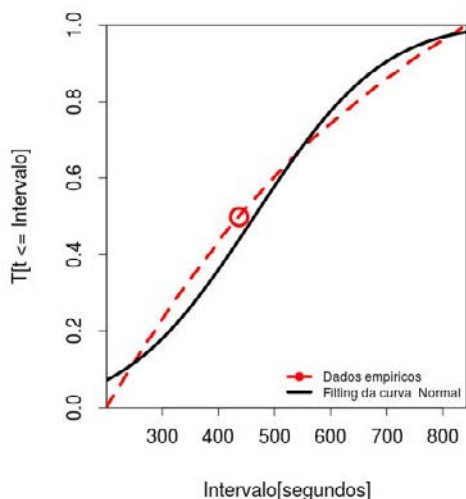


Figura 2. Caracterização do tempo de confirmação.

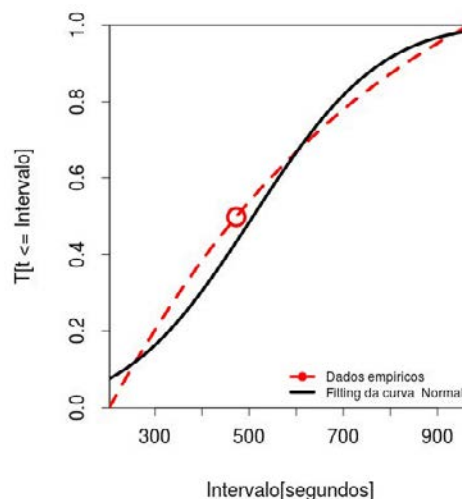
A fim de minimizar o efeito de ruídos gerados pelas ferramentas de coleta de dados executadas por nossos monitores, realizamos uma filtragem dos dados da série temporal de tempo entre transações, e consideramos apenas valores que estão entre o primeiro e terceiro quartis da distribuição. Após feita esta filtragem, os tempos entre transações consecutivas no *trace* analisado variou entre 0 e 1 segundo.

A CDF da distribuição dos tempos entre blocos, bem como a curva que melhor se adequa à distribuição empírica, estão representadas na Figura 3(a). Seja B a variável

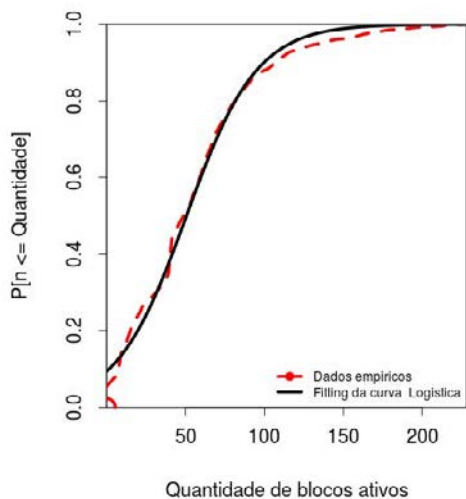
aleatória que caracteriza o tempo entre confirmação de blocos. A melhor distribuição que descreve o intervalo entre confirmação de blocos é a normal com parâmetros $E(B) = 464.30$ segundos e $V(B) = 179.57$. Desta forma temos um coeficiente de variação de 0.029. Em média os blocos gastam quase 8 minutos para serem minerados. Note que esse valor é um pouco menor que o alvo de 10 minutos visado no projeto da rede Bitcoin (vide <https://bitcoinwisdom.com/bitcoin/difficulty> para mais detalhes).



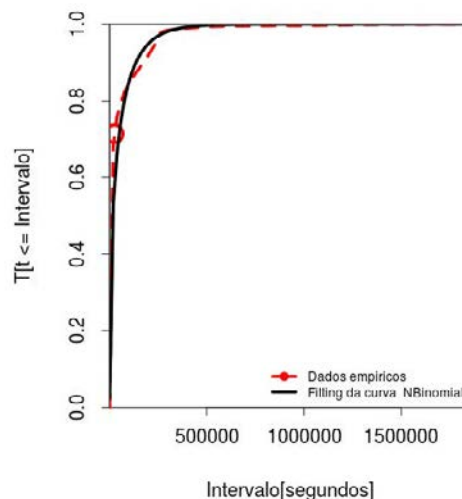
(a) Fitting da CDF do tempo entre blocos, B



(b) Fitting da CDF do tempo de confirmação das transações, D



(c) Fitting da CDF da quantidade de blocos ativos, M



(d) Fitting da CDF do tempo de atividade dos blocos, S

Figura 3. Fittings de métricas para a dinâmica de confirmação e tempos relacionados com transações e blocos

Já a CDF representada na Figura 3(b) reflete o intervalo de tempo que uma determinada transação espera por sua confirmação. Seja D a variável aleatória que caracteriza o tempo de confirmação de uma transação. Percebemos que, assim como para a distribuição dos tempos entre blocos, a melhor distribuição que descreve os dados empíricos foi a normal, com parâmetros $E(D) = 508.51$ segundos e $V(D) = 212.11$,

cujo coeficiente de variação calculado é 0.029. Portanto, em média as transações esperam um pouco mais de 8 minutos para serem confirmadas. Note que $E(D) > E(B)$, o que reflete o fato de que as transações tipicamente precisam esperar, na média, por pelo menos uma validação de bloco antes de serem consideradas como candidatas a validação.

Seja $E(B_r)$ o valor esperado da vida residual do tempo entre blocos. Quando uma nova transação chega ao sistema, ela espera $E(B_r)$ até enxergar o próximo bloco sendo confirmado,

$$E(B_r) = E(B^2)/(2E(B)) \quad (1)$$

O tempo médio de confirmação de uma transação é dado pela vida residual do tempo até a confirmação do próximo bloco, somada à espera pela confirmação de blocos adicionais até que esta seja efetivamente confirmada. Seja α o parâmetro que quantifica o número médio de blocos adicionais que se aguarda até uma confirmação. Então,

$$E(D) = \alpha E(B) + E(B_r) \quad (2)$$

Podemos estimar α experimentalmente. Conforme descrito nesta seção, temos que $E(B_r) = (179.57 + (464.30)^2)/(2(464.30)) = 232.34$, $E(D) = 508.51$ e $E(D) - E(B_r) = 276.16$. Ou seja, $\alpha = 0.6$, o que significa que na média as transações precisam esperar a vida residual de um bloco mais aproximadamente 0.6 do tempo entre blocos para serem servidas. Em outras palavras,

$$E(D) = 0.6(E(B) + E(B_r)) + 0.4E(B_r) \quad (3)$$

Na média 40% das transações já são atendidas no primeiro bloco que tem oportunidade de encontrar ao chegarem no sistema, e 60% tem de esperar por mais um bloco para serem confirmadas.

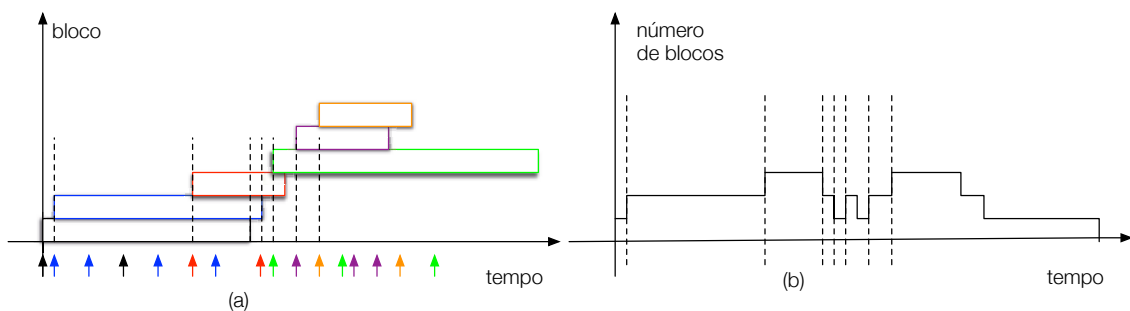


Figura 4. Blocos ativos: (a) cada barra inicia quando ocorre a primeira transação associada ao bloco, e termina quando o bloco é confirmado; (b) número de blocos ativos em função do tempo. A área embaixo desta curva dividida pelo tempo produz o número médio de blocos ativos.

Para obter maior entendimento sobre o tempo experimentado pelas transações, estudamos também o número de blocos ativos no sistema. Seja $E(M)$ o número esperado de blocos encontrados no sistema. $E(M)$ é igual à área embaixo da curva mostrada na figura 4 dividida pelo tempo de observação.

A figura 3(c) representa a CDF empírica da quantidade de blocos ativos, bem como a curva de *fitting* que melhor se adequou aos dados: a distribuição logística com

$E(M) = 50.60$ blocos ativos e $V(M) = 22.30$. Em princípio, tal média de blocos ativos é alta se comparada com o número médio de confirmações de blocos que uma transação típica observa antes de ser confirmada (em torno de 1.6 confirmações, conforme discutido acima). Um dos fatores que explicam tal discrepância é o esquema de prioridades implementado na rede (vide Seção 2). Algumas transações precisam aguardar muito mais que outras para serem confirmadas, o que favorece um aumento do número de blocos ativos.

Em relação ao tempo de atividade de cada bloco, a figura 3(d) mostra a sua CDF empírica, bem como o melhor *fitting* encontrado, dado pela distribuição binomial com $n = 47158$ e probabilidade de sucesso $p = 0.37$. Denotemos por \hat{S} o estimador do tempo de atividade de cada bloco. Temos a média estimada $E(\hat{S}) = np = 17448.46$ segundos e variância $V(\hat{S}) = np(1 - p) = 104.85^2$, com coeficiente de variação em torno de 0.006. Em média os blocos estão ativos em torno de 5 horas. Esse número, bem maior que o tempo médio entre gerações de novos blocos que é de aproximadamente 8 minutos, mais uma vez reflete o fato de que algumas transações ficam pendentes no sistema por muito tempo antes de serem confirmadas, acarretando em tempos de atividades de blocos na ordem de horas (figura 3(b)).

De acordo com o resultado de Little, temos que $E(M) = \lambda_B E(S)$, onde λ_B denota a taxa média de chegada de blocos, $\lambda_B = 12968/9605710 = 0.00135$ blocos/segundo (observamos 12,968 blocos de 22/09/2015 19:36:06 até 11/01/2016 23:51:16). Logo, $E(S) = 50.6/0.00135 \approx 10.4$ horas. Note que $E(S)$ é aproximadamente igual ao dobro do valor de $E(\hat{S})$ estimado pelo *fitting* da distribuição binomial discutido no último parágrafo. A discrepância pode ser consequência do fato de o sistema não encontrar-se em estado estacionário ou em equilíbrio. Trabalho futuro consiste em melhor compreender a causa de tal disparidade.

O tempo médio de confirmação de uma transação envolve vários fatores, incluindo as taxas associadas, incentivos etc. Neste trabalho, apresentamos um modelo preliminar para relacionar o tempo médio de confirmação de uma transação e o tempo entre blocos. Trabalho futuro consiste em analisar os diferentes fatores que afetam o tempo de confirmação, levando em conta prioridades.

6. Robustez da Rede

Nesta seção, avaliamos o quanto que a rede Bitcoin é sensível a remoção de nós tidos como centrais, e estudamos alguns aspectos estruturais da rede.

6.1. Remoção de Nós

Em outubro e novembro de 2013 o FBI fechou o site denominado *Silk Road*. Este site da *DeepWeb* ficou amplamente conhecido por realizar vendas de produtos e serviços ilegais através da rede Bitcoin. A fim de capturar possíveis alterações no sistema Bitcoin decorrentes do fechamento, foram analisados os dados referentes à atividade da rede, durante e após o fechamento do *Silk Road*. É possível notar que a quantidade de transações após o fechamento do site praticamente se mantém, conforme as estatísticas sobre a atividade na rede reportadas na tabela 1. Estas estatísticas reforçam a hipótese de que há grande diversidade nas transações do Bitcoin, de forma que aquelas relacionadas com mercado negro correspondem apenas a uma pequena parcela de todas as transações na rede.

Mês	Transações	Mês	Transações	Mês	Transações
Outubro	1,645,153	Novembro	1,959,041	Dezembro	1,941,525

Tabela 1. Número de transações antes e após o fechamento do *Silk Road* (2013).

Além da aparente manutenção da atividade da rede Bitcoin, nota-se que não há sazonalidade desta atividade, como pode ser conferido na figura 5, referente aos dias entre os meses de outubro e dezembro de 2013.

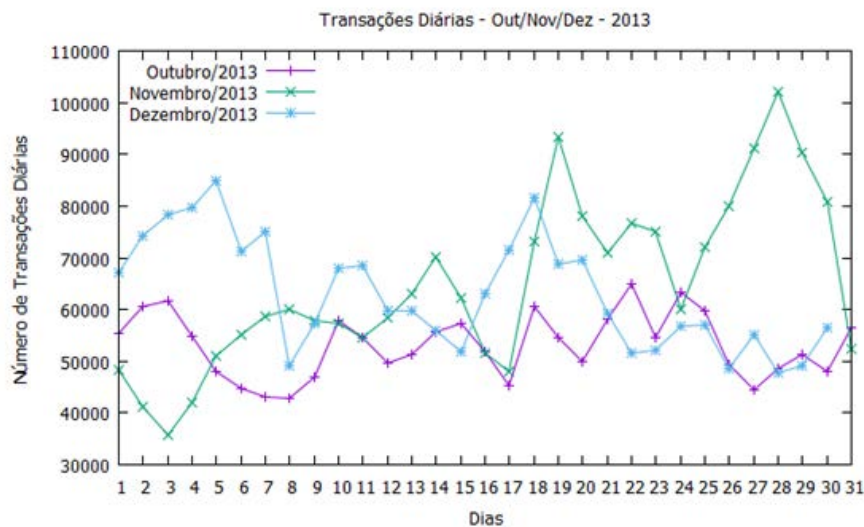


Figura 5. Diárias: outubro, novembro e dezembro de 2013

6.2. Transações de Grau Um e Transações de Grau Zero

Cada transação na rede Bitcoin está associada a um grau de entrada (número de transações que alimentam a transação em questão) e a um grau de saída (número de transações que recebem recursos da transação em questão, incluindo a transação referente ao troco). Uma análise das transações envolvendo graus altos é realizada em [Luiz 2016]. A seguir, em contrapartida, estudamos transações de baixo grau – ou seja, de grau 0 e grau 1. Em novembro de 2013 e fevereiro de 2014, contamos com 999,427 e 1,048,575 transações, respectivamente, e verificamos que existiu uma grande quantidade de transações com grau um e uma pequena parcela de transações com grau zero. Em janeiro e fevereiro, o número de transações de grau zero foi de 1,480 e 1,862, respectivamente. O número de transações de grau um foi de 123,875 e 82,092, respectivamente.

Apesar das transações de grau zero não serem recomendadas pelo protocolo padrão do Bitcoin, elas podem ser usadas para diferentes fins tais como a divulgação de chaves públicas. Entretanto, vislumbra-se que tais transações também possam ser usadas em ataques do tipo negação de serviço (“denial of service”), e por isso elas são desencorajadas. É interessante observar que o caráter descentralizado do protocolo Bitcoin admite que mesmo operações desencorajadas (e proibidas de acordo com o cliente padrão), acabem ocorrendo no *blockchain*.

Muitas das transações no Bitcoin envolvem o pagamento de uma certa quantia e um troco para o pagador. Por este motivo, elas possuem grau de saída maior ou igual

a dois. Entretanto, algumas transações possuem grau de saída igual a um. A maioria das transações de grau de saída igual a um configuram transações geradas pelo protocolo como recompensa por mineração. Estas transações, que possuem grau de entrada zero e grau de saída um, são denominadas de *coinbase*, e cada bloco possui ao menos uma transação deste tipo. Dentre os trabalhos futuros, pretendemos caracterizar todas as transações que possuem grau de saída igual a um (independente do seu grau de entrada), e buscar entender o seu papel na rede de transações.

7. Trabalhos Relacionados

O trabalho que originou o Bitcoin foi publicado na Internet em 2008 para que todos tivessem acesso a idealização do funcionamento do sistema [Nakamoto 2008]. O autor, com pseudônimo Satoshi Nakamoto, nunca foi identificado. Uma das principais contribuições do trabalho seminal consiste no uso do *proof of work* para evitar o problema do gasto duplo.

[Androulaki et al. 2013] e [Reid and Harrigan 2013] chamam atenção para o fato de que as medidas que o sistema Bitcoin adota para tratar o anonimato de seus usuários já não são suficientes. Os trabalhos utilizam técnicas de agrupamento de usuários em função do histórico de transações envolvidas para tentar desvendar características dos mesmos. A quebra de anonimato pode ocorrer mesmo que o usuário crie inúmeros endereços na tentativa de dificultar sua identificação.

Neste trabalho estudamos estatísticas sobre previsão de confirmação de transações, tempo de confirmação e graus das mesmas. Dentre os trabalhos de caracterização da rede Bitcoin, destacamos [Ron and Shamir 2013]. Neste trabalho, os autores analisam diversas propriedades estatísticas associadas às transações: como os usuários gastam seus bitcoins, o saldo de bitcoins dos usuários, como os usuários movem bitcoins entre seus diversos endereços para manter sua privacidade etc.

Outro trabalho de caracterização do Bitcoin é [Meiklejohn et al. 2013]. Neste trabalho, os autores identificam vários participantes do sistema e os separam em grupos. Os usuários identificados no trabalho são usuários com forte atuação no sistema: mineradores, vendedores, casas de câmbio e casas de jogos, além de outros usuários influentes. O trabalho apresenta duas heurísticas de como agrupar esses usuários e defini-los, mostrando mais uma vez que o anonimato no sistema não é tão forte como supostamente foi apresentado. O trabalho também caracteriza variáveis referentes à rede do sistema como: média de transações, menores transações, transações recebidas e um balanço de quanto que cada usuário transaciona com os usuários categorizados pelo trabalho.

Embora a literatura sobre o sistema Bitcoin seja vasta, não é de nosso conhecimento nenhum trabalho anterior que tenha proposto um modelo de filas para estimar o tempo de espera experimentado pelas transações.

8. Conclusão

Neste trabalho, nós apresentamos uma caracterização acerca das transações do Bitcoin, o sistema de cripto moeda mais popular até o presente. Nós conduzimos nosso estudo

a partir dos dados públicos, disponíveis na rede a todos os clientes Bitcoin, bem como dados colhidos ao longo de um ano por centenas de monitores instrumentados executando na rede Bitcoin em diversos pontos do mundo.

Além de métricas importantes relacionadas com transações, como a probabilidade uma transação ser realmente confirmada e o tempo decorrido para essa confirmação, nós provemos uma análise temporal simples, comparando um período referente ao fechamento do Silk Road com períodos subsequentes.

Nossos resultados mostram que há um número não negligenciável de transações que não são confirmadas. Nesse caso, nós observamos uma alta correlação entre o *fee*, seu volume e a suspeita de que a transação não será confirmada. Mais ainda, transações do Bitcoin geralmente são confirmadas em períodos curtos, mas ainda assim, muito acima de tempos usuais de sistemas convencionais de cartão de crédito. Em mais de 90% dos casos, uma transação foi confirmada em até 30 minutos. Em resumo, nós acreditamos que a rede Bitcoin é resiliente e não é dependente de *hubs* altamente centralizados, e que o presente trabalho abre novas avenidas em direção ao melhor entendimento de características básicas do sistema, ao propor um modelo baseado em teoria de filas para inferir o tempo médio de confirmação das transações.

Referências

- Androulaki, E., Karame, G. O., Roeschlin, M., Scherer, T., and Capkun, S. (2013). Evaluating user privacy in bitcoin. In *Financial Cryptography and Data Security*, pages 34–51. Springer.
- Bitcointalk (2016). Bitcoin talk forum. <https://bitcointalk.org/index.php?topic=1382996.0>.
- Bitcoinwiki (2016). Transaction fees. https://en.bitcoin.it/wiki/Transaction_fees.
- Luiz, H. (2016). Estudo da dinamicidade do sistema bitcoin. Master's thesis, UFJF, Juiz de Fora.
- Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., McCoy, D., Voelker, G. M., and Savage, S. (2013). A fistful of bitcoins: characterizing payments among men with no names.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Narayanan, A., Bonneau, J., Felten, E., Miller, A., and Goldfeder, S. (2016). Bitcoin and cryptocurrency technologies. <https://class.coursera.org/bitcointech-001/wiki/readings>.
- Reid, F. and Harrigan, M. (2013). *An analysis of anonymity in the Bitcoin system*. Springer.
- Ron, D. and Shamir, A. (2013). Quantitative analysis of the full bitcoin transaction graph.
- Spagnuolo, M. (2013). Bitiodine: extracting intelligence from the bitcoin network. *Politecnico di Milano*.

Energy Consumption Evaluation of NoSQL DBMSs

Carlos Gomes¹, Eduardo Tavares¹, Meuse Nogueira de O. Junior²

¹Centro de Informática, Universidade Federal de Pernambuco, (CIn,UFPE)

²Instituto Federal de Pernambuco (IFPE)

{cga, eagt}@cin.ufpe.br, meusejunior@recife.ifpe.edu.br

Abstract. *Over the years, NoSQL Database Management Systems (DBMS) have been adopted as an alternative to the constraints of relational/SQL DBMSs. In order to demonstrate their feasibility, works have evaluated NoSQL DBMSs regarding some performance metrics, but energy consumption has not been assessed. Indeed, energy consumption is an issue that should not be neglected due to the rise of energy costs and environmental sustainability. This paper presents a performance and energy consumption evaluation of NoSQL DBMSs, more specifically, Cassandra (column), MongoDB (document-oriented), Redis (key-value). Experiments are based on YCSB benchmark, and results demonstrate energy consumption can vary significantly among the assessed DBMSs for different commands (e.g., read) and workloads.*

1. Introduction

For building and maintain a data center, high investments are involved, and energy consumption plays a remarkable role. According to reports [NRDC 2015], data centers consumed 91 billions KiloWatts in United States during 2013 at a cost of \$ 9 billions USD, and the consumption is estimated to be 140 billions KiloWatts by 2020 (at a cost of \$ 13.7 billions USD). Additionally, governments around the world have tried to regulate energy demand for data centers, and, thus, designs based on energy saving have been crucial for modern data centers [NRDC 2015].

Data storage is an important subsystem in data centers, and it has significantly evolved and increased the capacity due to advent of new paradigms, such as cloud computing. Consequently, Big Data concept has emerged, which requires new tools to deal with a large volume of complex data [Demchenko et al. 2014, Ji et al. 2012].

No relational database management systems (DBMS), termed as NoSQL, have been proposed as a solution to deal with the constraints (e.g., performance) of conventional data management tools, such as relational/SQL DBMSs. Nowadays, there are over 150 NoSQL DBMSs [Kuznetsov and Poskonin 2014], which adopt distinct data models, such as key-value. The abundance and diversity of NoSQL DBMSs have motivated researches to evaluate and compare these DBMSs regarding latency and throughput. However, to the best of our current knowledge, energy consumption is usually neglected, despite its huge importance for modern data centers.

As previously stated, energy consumption is an important non-functional requirement, and the selection of an appropriate NoSQL DBMS may allow cost reduction and less environmental impact. Besides, system reliability is impacted by energy consumption due to the influence on operational temperature [Zhang et al. 2014]. Reducing

consumption in data center provides benefits in terms of heat dissipation and, thus, system reliability.

This paper presents an energy consumption evaluation of representative NoSQL DBMSs with distinct data models, more specifically, Redis (key-value), MongoDB (document-oriented) and Cassandra (column). Experiments are based on Yahoo! Cloud Serving Benchmark (YCSB), in which insert, read, and delete commands are evaluated for each adopted DBMS. Besides, we also provide performance metric for these operations and their correlation to energy consumption. A measurement framework, namely, EMeter, is also presented, which contemplates hardware components and a software tool to estimate energy consumption and execution time.

The remainder of this paper is organized as follows. Section 2 summarizes related works and Section 3 presents an overview of NoSQL DBMS. Section 4 describes the methodology and the measuring framework, and Section 5 presents the results. Finally, Section 6 presents closing remarks and future works.

2. Related Works

According to Intel Labs report [Minas and Ellison 2009], in a server, CPU is the main energy consumer (followed by RAM memory), which indicates processes that require intense CPU usage and memory access, like those utilized by DBMSs, lead to high energy consumption. However, few works have evaluated the performance and energy consumption of DBMSs. Commonly, researchers assess throughput and latency assuming distributed applications and high-capacity hardware - memory and processing power - for evaluating performance under high workloads [Seriatos et al. 2016, Floratou et al. 2012, Neves and Bernardino 2015]. It is difficult to find works related to energy consumption for DBMSs, besides, researches assessing energy consumption in NoSQL DBMS are scarce.

In [Abramova and Bernardino 2013], the authors assessed the execution time of two NoSQL DBMSs, Cassandra and MongoDB, taking into account distinct workloads generated by YCSB benchmark. Results demonstrate Cassandra provided an improved scalability in comparison to MongoDB. [Abubakar et al. 2014] evaluated MongoDB, Elasticsearch, OrientDB and Redis DBMSs considering distinct workloads for read, insert and update, but energy consumption is not taken into account. [Seriatos et al. 2016] carried out a performance evaluation comprising HBase, MongoDB and Cassandra in a cloud adopting YCSB benchmark, considering different scenarios in which different parameters of each DBMS was assessed.

[Neves and Bernardino 2015] evaluated the performance of Voldemort, a NoSQL DBMS, using YCSB benchmark considering an environment with 1, 3 and 6 nodes. The work concluded Voldemort DBMS does not scale considerably, and additional nodes do not have a significant impact on performance. [Cai et al. 2013] evaluated Hbase DBMS using also YCSB benchmark, and the work shows issues related to scalability and concurrent client requests. [Li et al. 2014] assessed energy consumption in NoSQL DBMSs regarding the idle state on cluster configurations. The analysis was performed collecting the idle time of each node, which may not reflect the dynamics and behavioral variations of accessing a DBMS.

Different from previous works, this papers provides an energy consumption eva-

luation and comparison of representative NoSQL DBMSs, also providing a correlation between performance and energy consumption.

3. NoSQL DBMS

Databases comprise data repositories and a common model for manipulating data. To access these repositories, DBMSs are adopted, which are set of mechanisms for providing data storage and manipulation [Abramova and Bernardino 2013]. Over the decades, several DBMSs have been proposed.

The term NoSQL was adopted for the first time in 1998 [Kuznetsov and Poskonin 2014] for a small DBMS that did not adopt the relational model. Nowadays, the term also encompasses DBMSs, which do not fully comply to ACID (Atomicity, Consistency, Isolation, Durability) properties and/or the relational model [Kuznetsov and Poskonin 2014]. Besides, the popularity of NoSQL DBMSs is related to the advent of Big Data, as huge data volume needs to be quickly manipulated, and traditional data models or ACID transactions may considerably affect system performance.

In this work, we evaluate 3 popular NoSQL DBMSs (Redis, MongoDB, Cassandra) with representative data models (key-value, document-oriented, column) (Figure 1), and they are described as follows.

3.0.1. Column

Column model (Figure 1 (a)) is the closest to the relational model, in which data are structured using the following column types [Abramova and Bernardino 2013]:

- **column** represents a set of data structured as key-value;
- **super column** provides a group of columns;
- **column family** represents a data set composed of super columns, which resemble the relational model.

Cassandra [Planet 2015] is a popular column DBMS, which was initially developed by Facebook using Dynamo DBMS (Amazon) and BigTable (Google) as a basis. Additionally, Cassandra provides a set of functionalities to facilitate a deployment onto distributed architectures, and it adopts a query language with a syntax similar to SQL.

3.0.2. Document-oriented

Document-oriented model (Figure 1 (b)) represents data using a document format, usually, based on well-known standards, such as XML and JSON. Each document may have several fields, which may include arrays and even other documents. A document has a unique key, which is the usual mechanism for accessing documents. Besides, documents can be grouped [Abramova and Bernardino 2013].

MongoDB[Mongo 2015] is a representative document-oriented DBMS, which adopts JSON for data storage . To avoid performance issues, MongoDB extensively adopts main memory. Documents are manipulated using conventional operations, such

as read, insertion, exclusion, update. To search for documents, some mechanisms are available based on projection (adopted by YCSB benchmark) and iteration. Besides, during an update, a document can be totally or partially modified.

3.0.3. Key-Value

This data model stores all data using a structure based on key-value (Figure 1 (c)). Each key is unique and data access is performed relating keys to hash values in order to access the desired value. The value can be a primitive type or complex data structure, such as a tree.

In this work, we adopt Redis [Kuznetsov and Poskonin 2014], which is a open-source DBMS that supports key-value model. Besides, Redis supports atomicity and isolation, and data is kept on main memory. To allow persistent storage, Redis provides different mechanisms, in which the default mechanism periodically stores snapshots obtained from the main memory.

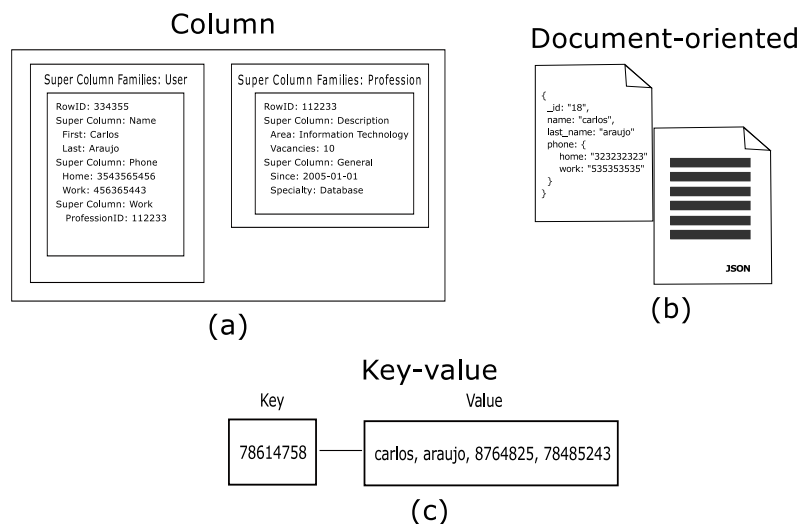


Figura 1. NoSQL Data Models

4. Methodology

The methodology is based on design of experiments [Douglas C. Montgomery 2013], in which a l^k factorial design with r replications is adopted. We have considered 2 factors ($k = 2$) with 3 levels ($l = 3$): (i) DBMS - Cassandra, MongoDB, Redis; and (ii) command - insert, read, update. Besides, 3 different workloads are considered (1,000 operations, 10,000 operations, 100,000 operations) generated by YCSB benchmark [Cooper et al. 2010], and the metrics of interest are energy consumption and execution time for each workload (e.g., 1,000 operations). The workload could be an explicit factor for the proposed evaluation, but it would be the major source of variation in result analysis, and the overall measurement noise (i.e., random errors) would hinder a finer comparison.

YCSB is an open-source benchmark suite for evaluating computer applications, being often adopted to compare the performance of NoSQL DBMSs. Indeed, YCSB benchmark has been the standard for evaluating the performance of general DBMS. Additionally, 75 replications ($r = 75$) are taken into account to obtain mean values (with an approximate normal distribution) and to reduce the impact of measurement noises. A replication is a workload execution (e.g., a single execution of 1,000 operations). In this work, results are analyzed using ANOVA [Douglas C. Montgomery 2013].

The adopted DBMS server is a Core 2 Duo CPU T5450 1.66GHz with of 2GB RAM memory, running Debian 7.8 (Linux) with EXT4 as system file. All operating system (O.S.) services were kept to a minimum to not impact data collection, and no application process (except DBMS software) is executing. Regarding DBMS softwares, Cassandra 2.0.15, MongoDB 2.0.6 and Redis 2.4.14 were adopted with the default configuration.

As follows, the conceived measurement framework is described.

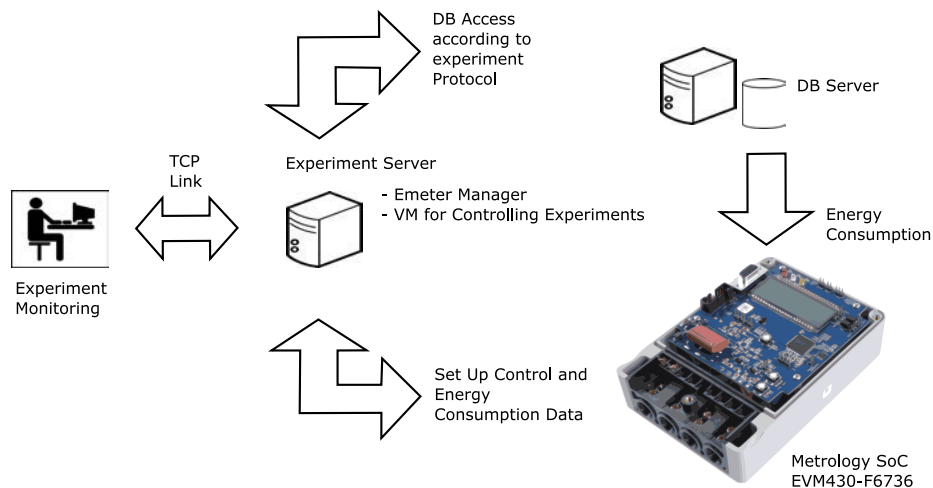


Figura 2. Measurement Framework

4.1. Measurement framework

We conceived a measurement framework (Figure 2), namely, Emeter, to allow the collection of data related to energy consumption and execution time in the DBMS server. The framework contemplates hardware components and a software tool to store and visualize the collected metrics.

The framework adopts a experiment server, in which Emeter software executes, and it communicates with a EVM430-F6736 hardware [Texas-Instruments 2015]. The latter is a specialized device that collects instantaneous current and voltage for estimating instantaneous electric power. Energy consumption is calculated by using a numerical integration using the instantaneous powers and the time interval for executing a workload. EVM430-F6736's firmware was modified, such that Emeter's software can communicate and collect the required data via a serial interface. For managing and controlling, Emeter has a GUI (Graphical Unit Interface), in which real-time consumption and execution

information can be visualized. The collected data is stored on a database or a csv file. Additionally, EMeter supports remote control by TCP/IP communication, in the sense that all experiments can be externally controlled.

The experiment server also executes YCSB benchmark, which triggers the workload on the DBMS server. The time interval for a workload is adopted for estimating execution time and energy consumption, and each execution is kept independent in order to avoid interference between replications (e.g., cache).

5. Experimental results

We have adopted a design of experiment, which is explained in Section 4. YCSB benchmark is taken into account considering workloads with 1,000, 10,000 and 100,000 operations. For each DBMS, the default configuration is adopted, and the metrics of interest are execution time and energy consumption. Besides, each inserted record in the database has 1KB (standard value defined by YCSB), and 10 threads is adopted for mimicking a real environment with concurrent accesses. As follows, the results are presented for each workload using ANOVA analysis [Douglas C. Montgomery 2013]. Next, correlation is discussed for both metrics followed by general remarks.

5.1. Workload results

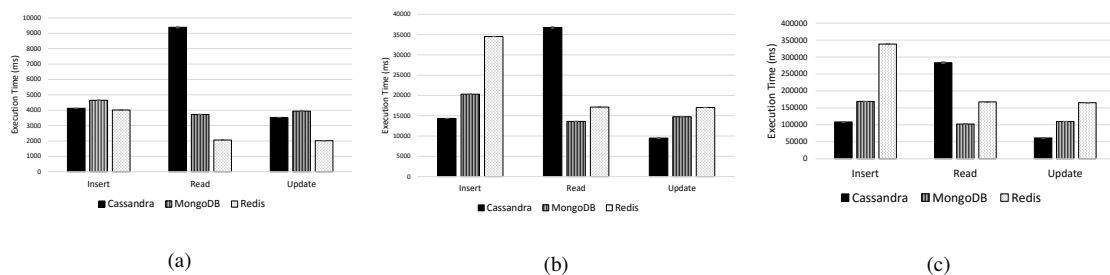


Figura 3. Mean Execution Time Values

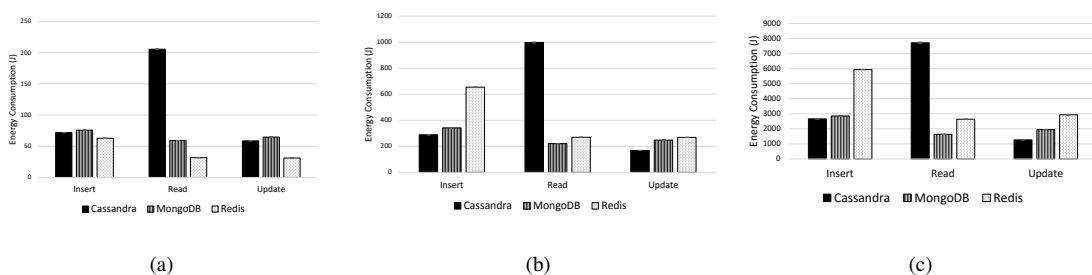


Figura 4. Mean Energy Consumption Values

Table 1 and Table 2 provides the results for ANOVA analysis (significance level $\alpha = 0.05$) for execution time and energy consumption. Figure 3 and Figure 4 depict the results considering the mean values.

Table 1 depicts all factors and their interactions (*source*) significantly impact the execution time (*F statistic -F stat.-* and *p-value*). Depending on the workload (*work.*),

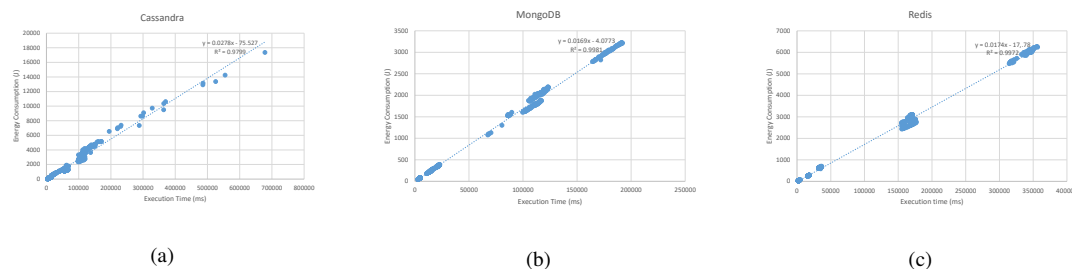


Figura 5. Correlation Energy Consumption X Execution Time

the factor may have a different impact on the metric (*var.%*), as some DBMSs vary their behaviour for the amount of operations executed. In general, the interaction between DBMS and command (*DBMS * Comm*) is the major source of variation, but, in 100,000 workload, the interaction influence changes due to the behaviour of Redis and Cassandra. Similar results are obtained for energy consumption. However, for 100,000 workload, a minor difference occurs due to a sudden increase of energy consumption by Cassandra (explained in the following paragraphs). Other sources of variation are due to experimental errors (i.e., noise in the measurements), and they are reduced with the increase of workload. Such a situation is related to the time interval required for a workload execution, which decreases, for instance, the influence of other O.S. services. The following explanations are based on the results obtained with Tukey’s procedure [Douglas C. Montgomery 2013] (a post-ANOVA test).

Tabela 1. ANOVA: Execution Time

work. 1,000				work. 10,000				work. 100,000			
source	var.%	F stat.	p-value	source	var.%	F stat.	p-value	source	var.%	F stat.	p-value
DBMS	23.15	230.41	<0.001	DBMS	7.97	84.31	<0.001	DBMS	30.80	453.88	<0.001
Comm	9	89.55	<0.001	Comm	16.95	179.28	<0.001	Comm	20.32	299.42	<0.001
DBMS*Comm	30.78	153.16	<0.001	DBMS*Comm	42.74	226.04	<0.001	DBMS*Comm	26.28	193.62	<0.001
Error	37.07			Error	32.33			Error	22.60		

Tabela 2. ANOVA: Energy Consumption

work. 1,000				work. 10,000				work. 100,000			
source	var.%	F stat.	p-value	source	var.%	F stat.	p-value	source	var.%	F stat.	p-value
DBMS	23.24	243.66	<0.001	DBMS	7.54	99.73	<0.001	DBMS	13.28	147.94	<0.001
Comm	9.78	102.55	<0.001	Comm	15.71	207.92	<0.001	Comm	16.62	185.14	<0.001
DBMS*Comm	31.36	164.45	<0.001	DBMS*Comm	51.59	341.34	<0.001	DBMS*Comm	40.19	223.83	<0.001
Error	35.62			Error	25.16			Error	29.90		

Figure 3 (a) depict the mean execution times for running 1,000 operations. Concerning insert command, the difference between values are not statistically significant, and, thus, we do not have enough statistical evidence to reject the equality between the execution times. For read commands, the difference is significant, and Cassandra considerably takes more time than other DBMSs. For instance, it is, respectively, 60.33% and 78.04% slower than MongoDB and Redis (which provided the lowest execution time).

Regarding update, Redis also provides the best value for this workload followed by Cassandra. In the context of energy consumption (Figure 4 (a)), the values indicate a correlation with execution time, since similar arguments can be provided. The values for insert command are also close, and Tukey's test indicates the differences between DBMSs are not expressive. Concerning read command, Cassandra provides the highest energy consumption, which is 71.17% larger than MongoDB and 84.57% than Redis for this command. Redis has the least energy consumption for updates, but MongoDB and Cassandra do not have mean values that significantly differ.

Considering 10,000 operations (Figure 3(b)), Redis appears to be less scalable than other DBMSs for insert command. Redis was 58.51% slower than MongoDB and 41.18% than Cassandra. this situation explains the increase of command (*comm.*) as a source of variation in Table 1. Comparing to 1,000 workload, Redis' execution time increased 88.38%, whereas for Cassandra and MongoDB the executime increased 71.22% and 77.13% , respectively. Cassandra provides the best execution time for insert command, but its behaviour repeats for reading requests, which execution time is much larger than Redis and MongoDB. In this context, MongoDB provides the best performance. For update, Cassandra provides good execution time, in which it is almost 2 times faster than MongoDB and Redis. Figure 4 (b) depicts energy consumption values for this workload, which also points to the correlation between execution time and energy consumption. The values indicate workload variation impacts execution time and energy consumption in different proportions. As an example, the execution time of Redis DBMS increased 88.38% for insert command, but energy consumption increased only 90.41%. For Cassandra and MongoDB, the increment were 75.10% and 77.82%, respectively. In relation to read operation, Cassandra, Redis, MongoDB consumed more 79.42%, 88.25% and 73.19%, respectively. Energy consumption for update correspondingly increased 65.04%, 73.91%, 88.41% for Cassandra, MongoDB and Redis.

Figure 3 (c) shows the mean values for execution times concerning 100,000 operations workload. The behaviour resembles 10,000 workload, in which Redis provides the slowest performance, except for update command. The mean time for executing 100,000 insert commands is the poorest performance in all experiments (surpassing the execution time of read commands for Cassandra). The values is 67.96% higher than Cassandra, and 50.06% larger than MongoDB. Also, Cassandras did not scale well for read commands. The values is 63.87% higher than MongoDB and 40.87% than Redis. Concerning update, Cassandra provides the best performance, followed by MongoDB and Redis. Figure 4 (c) provides an interesting behaviour. Although it also provides evidence for correlation between execution time and energy consumption, energy consumption in Cassandra seems higher than other DBMSs. Although Redis' execution time for insert commands are greater than Cassandra's update commands, Cassandra consumed more energy consumption. Such a behaviour indicates the change of variation (*var.%*) of each factor for both metrics . Particularly, the contribution of command factor raises as well as the noise in the measurement for energy consumption.

5.2. Correlation

Figure 5 depicts the correlation between energy consumption and execution time for each DBMS contemplating all workloads and commands (insert, read and update).

For all DBMS, a strong (linear) correlation is obtained, which is corroborated by the square of correlation coefficient (R^2). The values are 0.9799, 0.9981, and 0.9972 for Cassandra, MongoDB and Redis respectively. Figure 5 (a) depicts the values for Cassandra, in which equation $y = 0.0278x - 75.527$ represents the correlation. Equation $y = 0.0169x - 4.0773$ represents the correlation for MongoDB (Figure 5 (b)), and $y = 0.0174x - 17.378$ for Redis (Figure 5 (c)).

The slope (i.e., the first derivative) provides an interesting information regarding the impact of executing commands on mean energy consumption. For the adopted DBMS, MongoDB is more energy efficient (0.0169 energy consumption/execution time), and Redis is very close (0.0174), despite the issue on 100,000 workload. Cassandra consumes more energy in average, which is represented by the slope 0.0278, which is strongly affected by read commands.

5.3. Remarks

It is important to emphasize that all experiments have been carried adopting the default configuration for each DBMS, and configuration tuning is out of scope for this work. Nevertheless, tuning may provide different results from the values presented in this work. Besides, we have adopted only a single computer to reduce possible interference of other issues associated with a cluster environment on the results. Nevertheless, in a cluster environment, the results may also differ.

Redis had a remarkable performance for 1,000 workload, in the sense that it provided the best performance for this workload concerning read and update commands. Execution time and energy consumption is equivalent to other DBMSs concerning insert command. We believe this workload was not sufficient to activate Redis' snapshot task for persistently saving the new records during a workload execution. However, the snapshot was activated for 10,000 and 100,000 workloads [Abubakar et al. 2014], which affected the performance of insert command and generated the worst execution time in 100,000 workload.

Cassandra performed very well for insert and update commands, and those commands were improved with the increasing of workload. Indeed, for 10,000 and 100,000 workloads, Cassandra provided best results. However, such a DBMS provided worst results for read command in all workloads, and this is the command that consumed more energy. Cassandra adopts several tasks to access persistent data [Cassandra 2015], which include decompress data from a storage device (e.g., hard disk). All samples are independent, and, thus, no data are present in Cassandra's cache during the workload for read command. In [Abramova and Bernardino 2013], the authors observed a similar behaviour for Cassandra.

For all workloads, MongoDB kept a stable performance regarding no abrupt increasing in execution time or energy consumption for the adopted commands. However, it did not provide a dominant performance comparing to other DBMSs. A minor increase in the time for executing insert commands may be related to the journaling mechanism adopted for providing fault-tolerance.

It is important to emphasize that no DBMS provided a dominant behaviour (execution time and energy consumption) in all workloads.

6. Conclusion

Energy consumption is a very important concern, which has gained considerable attention by researchers and practitioners to develop energy efficient platforms and systems. Due to advent of big data, storage systems have also their considerable contribution on energy usage, but such a concern has not been explored on NoSQL DBMSs (although works have evaluated NoSQL DBMSs in the context of throughput and latency).

This paper presented performance and energy consumption evaluation of representative NoSQL DBMSs. Experiments adopted distinct workloads for read, create and update commands using Yahoo! Cloud Serving Benchmark. Although no single DBMS dominates all workloads, MongoDB provided a stable behaviour for the adopted system and default configuration. Cassandra and Redis provided prominent execution time and energy consumption for some scenarios, but they did not perform well for read and insert commands, respectively. As future works, we are planning to evaluate NoSQL DBMSs on a cluster environment.

Referências

- Abramova, V. and Bernardino, J. (2013). Nosql databases: Mongodb vs cassandra. In *Proceedings of the International C* Conference on Computer Science and Software Engineering*, pages 14–22.
- Abubakar, Y., Adeyi, T. S., and Auta, I. G. (2014). Article: Performance evaluation of nosql systems using ycsb in a resource austere environment. *International Journal of Applied Information Systems*, 7(8):23–27.
- Cai, L., Huang, S., Chen, L., and Zheng, Y. (2013). Performance testing of hbase based on the potential cycle. In *Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference on*, pages 359–363.
- Cassandra, W. (2015). Architecture internals. <https://wiki.apache.org/cassandra/ArchitectureInternals>. Accessed: 2016-03-24.
- Cooper, B. F., Silberstein, A., Tam, E., Ramakrishnan, R., and Sears, R. (2010). Benchmarking cloud serving systems with ycsb. In *Proceedings of the 1st ACM Symposium on Cloud Computing*, SoCC '10, pages 143–154.
- Demchenko, Y., De Laat, C., and Membrey, P. (2014). Defining architecture components of the big data ecosystem. In *Collaboration Technologies and Systems (CTS), 2014 International Conference on*, pages 104–112.
- Douglas C. Montgomery, G. C. R. (2013). *Applied Statistics and Probability for Engineers*. Wiley, 6th edition.
- Floratou, A., Teletia, N., DeWitt, D. J., Patel, J. M., and Zhang, D. (2012). Can the elephants handle the nosql onslaught? *Proceedings of the VLDB Endowment*, 5(12):1712–1723.
- Ji, C., Li, Y., Qiu, W., Jin, Y., Xu, Y., Awada, U., Li, K., and Qu, W. (2012). Big data processing: Big challenges and opportunities. *Journal of Interconnection Networks*, 13.

- Kuznetsov, S. D. and Poskonin, A. V. (2014). Nosql data management systems. *Programming and Computer Software*, pages 323–332.
- Li, T., Yu, G., Liu, X., and Song, J. (2014). Analyzing the waiting energy consumption of nosql databases. In *Dependable, Autonomic and Secure Computing (DASC), 2014 IEEE 12th International Conference on*, pages 277–282. IEEE.
- Minas, L. and Ellison, B. (2009). *Energy efficiency for information technology: How to reduce power consumption in servers and data centers*. Intel Press.
- Mongo, M. (2015). Bring your giant ideas to life with mongodb. <https://www.mongodb.com/what-is-mongodb>. Accessed: 2016-03-22.
- Neves, R. and Bernardino, J. (2015). Performance and scalability of voldemort nosql. In *Information Systems and Technologies (CISTI), 2015 10th Iberian Conference on*, pages 1–6.
- NRDC (2015). America’s data centers consuming and wasting growing amounts of energy. <http://www.nrdc.org/energy/data-center-efficiency-assessment.asp>. Accessed: 2016-03-01.
- Planet, C. (2015). What is apache cassandra? <http://www.planetcassandra.org/what-is-apache-cassandra/>. Accessed: 2015-12-10.
- Seriatos, G., Kousiouris, G., Menychtas, A., Kyriazis, D., and Varvarigou, T. (2016). *Comparison of database and workload types performance in Cloud environments*, pages 138–150.
- Texas-Instruments (2015). Evm430-f6736 - msp430f6736 evm for metering. <http://www.ti.com/tool/EVM430-F6736>. Accessed: 2016-03-22.
- Zhang, H., Shao, S., Xu, H., Zou, H., and Tian, C. (2014). Free cooling of data centers: A review. *Renewable and Sustainable Energy Reviews*, pages 171–182.

Analytical Modeling Approach of Routing Deflection for Intra-domain Networks

Rodolfo R. Gomes¹, Cristina K. Dominicini¹, Alextian B. Liberato¹
Moisés R.N. Ribeiro¹, Magnos Martinello¹

¹Software Defined Networks Research Group (NERDS)
Federal University of Espírito Santo, Vitoria/ES, Brazil

rodolfo@ifes.edu.br, cristina.dominicini@ifes.edu.br

alextian@ifes.edu.br, moises@ele.ufes.br, magnos@inf.ufes.br

Abstract. *This paper presents an analytical modeling to **KAR (Key-for-Any-Route)**, an intra-domain resilient routing system in which edge-nodes set a route ID to select any existing route as an alternative to safely forward packets to their destination. KAR-enabled switches explore the existing routes by using special properties of Residue Number System as encoding and forwarding techniques. The encoding technique allows adding resilient forwarding paths to drive deflected packets (due to link failure) to their original destination so that loops are not formed. Three deflection methods are discussed along with their analytical models checked against numerical simulations. Results show that KAR efficiently allows deflected packets to automatically reach their destination.*

1. Introduction

Routing is an important feature in computer networking that selects a path between two nodes to enable their communication. The route can be determined within the network by each node or be entirely computed by ingress edge nodes, also known as source-routing (SR). Although non-SR protocols tend to be scalable, SR networks facilitate traffic engineering because they offer choice of routes for the sources to select the path in a network-wide view with undesirable characteristics [Yang and Wetherall 2006]. Thus, SR technique is the basis of many proposals to improve the reliability and performance of networks, as a promising approach to improve flexibility of the network layer in future Internet architectures.

Source routing has an important issue regarding resilience: its reaction speed to failures of a link or node that belongs to a path. As soon as a core node becomes aware of a link failure, it sends a failure notification to the edge nodes. The source node can, then, select another path that does not include the faulty link after the new topology converge. Even if an alternative path has been pre-selected as a protection (to avoid the delay imposed by convergence), the source node must wait to receive the notification message and the core node must wait the new path signaling. Until that, packets that had already left the source node are dropped.

In order to avoid this packet dropping, another failure reaction can happen within the network. Thus, a node can locally switch to the alternative path as soon as the node detects a failure on one of its directly connected links affecting that path. However, this approach requires that every node/switch should be able to compute and/or store the backup paths, so that there is a dependency between each switch forwarding table and the topology of the entire network. This is normally done by pre-signaling the protection paths to switches, as [Swallow et al. 2005] does, but it tends to cause an initial delay before packets start to flow. Another way is embedding extra paths in packet header, increasing its

length, as it is done in [Nguyen et al. 2011]. Moreover, this process often requires extra packet processing, such as changes on packet header, at each switch while the packet travels its route. Lastly, deflection routing is another approach for failure reaction. It may be done by a core node randomly selecting an outgoing port and relying on the adjacent node to deliver the packets to their destination. In this case, there is no protection signaling and no state requirement to core nodes. However, the packets would follow a random walk and create transient loops [Yang and Wetherall 2006].

We propose **KAR (Key-for-Any-Route)**, that is a new intra-domain routing system with a novel fast failure reaction mechanism which combines the benefits of source-controlled routing with driven deflections to provide network routing resiliency. A path in a KAR network is represented by a number, the Route ID, inserted into the packet header. The next-hop decision relies on the remainder of the division between the Route ID and the switch ID along the path(s). The core nodes are simple stateless forwarding nodes. As for the fast failure reaction mechanism, these core nodes randomly deflects packets from the faulty link instead of dropping them. Those packets then follow by a diverse set of switches by using driven deflection forwarding paths embedded in the route ID due to the encoding technique exploit properties from Residue Number System (RNS) [Garner 1959]. Thus, KAR approach addresses link failures keeping the network connectivity and allowing in-flight packets along the failed path to reach their destination with a small route stretch.

An analytic model is proposed to evaluate the effect of three routing deflection techniques on the number of hops taken to reach destination. The model is validated in a discrete event simulator.

2. KAR Resilient Routing System Design

As a source-routed networking, KAR has two main components: the edge nodes, which are responsible for routing, and the core nodes. In this work, the latter forward packets based on their unique identifier (Switch ID) and the Route ID, a packet header field. They do not need any route and protection signaling.

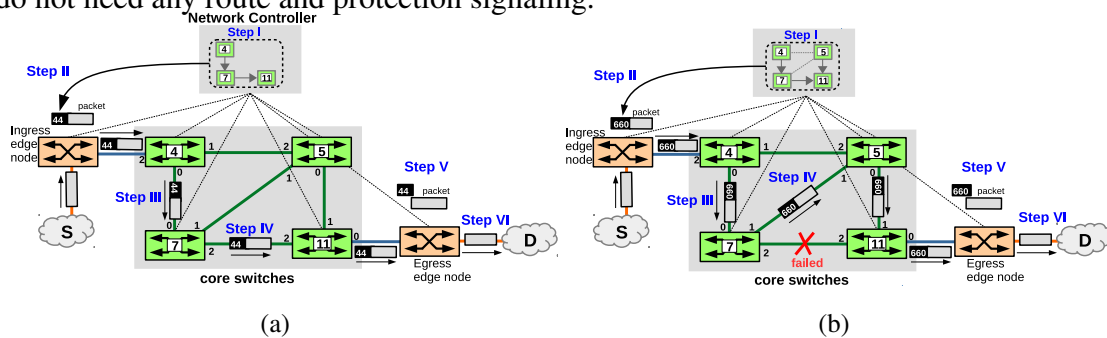


Figure 1. KAR design: (a) routing system based on shortest path, (b) fast failure reaction with driven deflection forwarding paths.

Fig. 1 illustrates an overview of the concept of *KAR* design with 6-node network. When the conventional node *S* in Fig. 1(a) wishes to communicate with another conventional node *D*, the ingress edge node selects an end-to-end path by using a routing algorithm (e.g. shortest path). Suppose it chooses the set of switches $S = \{4, 7, 11\}$ as the primary path. These IDs together with their respective output switch interfaces $P = \{0, 2, 0\}$ allows computing the *Route ID*, e.g. $R = 44$ (**Step I**) that should be assigned to the header of the incoming flow packets by the node itself (**Step II**). *KAR route ID* computation exploits RNS properties, and will be latter explained (section 2.2).

By computing the remainder of the division (denoted as $\langle a \rangle_b \equiv a \text{ modulo } b$) between the *route ID* ($R = 44$) and the *Switch ID*, the core switches know the output port to send packets to. Consequently, switch_ID 4 (SW4) forwards packets with route ID 44 to port $\langle 44 \rangle_4 = 0$ (**Step III**). In its turn, SW7 forwards them to the port $\langle 44 \rangle_7 = 2$ (**Step IV**). Finally, SW11 forwards them to port $\langle 44 \rangle_{11} = 0$ (**Step V**). They then reach the egress edge node, that removes the *route ID* from the packet header (**Step VI**) and delivers them to node D .

A link failure makes the adjacent core switches notify the edge nodes, which recalculate the route ID ignoring the faulty link in their network view. In the meanwhile, all packets sent by the source before the route ID modification would be lost. In order to avoid packet loss (*Hitless* property), KAR uses a deflection routing approach. As for this fast failure reaction, it chooses randomly one of its healthy ports to forward the packets to. Although deflection routing may form transient loops [Yang and Wetherall 2006], KAR provides a guarantee of loop-free routing even in the event of a link failure based on its *Driven Deflection* property. To this end, it is necessary to compose protection paths that are responsible for driving deflected packets to the destination by adding new nodes in the computation of the route ID. Fig. 1(b) illustrates this concept, by including proactively SW5, in (Step I), in the route ID as a protection path that delivers the deflected packets to SW11 when a link failure happens (resulting in $R = 660$). Consider, also, that the selected deflection technique chooses randomly between the available ports when a link fails. Thus, when link SW7-SW11 fails, SW7 forwards packets randomly to port 0 (SW4) or 1 (SW5). Those packets sent to SW4 bounce back to SW7 and suffer another deflection. However, all the packets that reach SW5 by deflection in SW7 (Step IV) will be forwarded to SW11 ($\langle 660 \rangle_5 = 0$) and, in this way, arrive at the destination.

2.1. Deflection Techniques

We propose the following three deflection routing techniques in order to make deflected packets reach the destination by the *Driven Deflections* property or even by chance.

Hot-Potato (HP): once a packet is deflected, its remaining route is completely random.

Any Valid Port (AVP): only when the output port computation does not represent a valid port ID (it does not exist or it is not available), the core node selects at random an active port and send packet to it.

Not the Input Port (NIP): the AVP method is changed by excluding the input port from the set of next-hop candidates even when the computation tells to send the packet back to it. This rule is broken only if the only available port is the ingress one. Besides generating less random paths, it avoids routing loops between two nodes.

There is a good reason to propose AVP or NIP: after deflection, a packet may arrive at a node that leads to the destination path. From there, it will follow the computed path once again. Note that in Fig. 1(a), without any *Driven Deflection* forwarding paths, a packet arriving at SW5 has 50% probability to go to SW11. In contrast, the addition of SW5 in the route ID and the use of NIP deflection technique cause all the packets to be driven through this forwarding path (SW5→SW11).

A final remark in terms of modeling is that an edge node can receive a packet not addressed to it. In this case, the node can choose between two approaches: it directly returns the packet to the network without any change or it queries the controller for a new route ID (path between the edge node and the destination) before returning the packet to the network. In our tests, we considered this second approach.

2.2. Encoding the Forwarding Paths

The generation of Route ID from Switch IDs and output port indexes is based on the Residue Number System, which is described in this section.

Let S be a set of modulo $S = \{s_1, s_2, \dots, s_N\}$ of the N switch IDs on the desired path, in which all elements of the set are pairwise coprimes numbers. Let P be a set of outgoing ports $P = \{p_1, p_2, \dots, p_N\}$, where p_i is the outgoing port index for the packet at the switch s_i .

Let M be

$$M = \prod_{i=1}^S s_i \tag{1}$$

A number $R \in \mathbb{N} | 0 \leq R < M$ can be represented by a residue set given a modulo set:

$$R \xrightarrow{RNS} \{p_1, p_2, \dots, p_N\}_S \tag{2}$$

, where

$$p_i = R \text{ modulo } s_i \tag{3}$$

The Chinese Remainder Theorem [Ding et al. 1996] states that is possible to reconstruct R through its residues in a RNS. This can be done as follows:

$$R = \langle \sum_{i=1}^S p_i \cdot M_i \cdot L_i \rangle_M \tag{4}$$

where

$$\langle a \rangle_b \equiv a \text{ modulo } b \tag{5}$$

$$M_i = \frac{M}{s_i} \tag{6}$$

$$L_i = \langle M_i^{-1} \rangle_{s_i} \tag{7}$$

Eq. (7) means that L_i is the modular multiplicative inverse of M_i . In other words, L_i is an integer number such that

$$\langle L_i \cdot M_i \rangle_{s_i} = 1 \tag{8}$$

Returning to the example of this section, the computation of route ID from S to D is obtained as follows:

Single forwarding path	Path protected by driven deflections
$switches = \{s_1, s_2, s_3\} = \{4, 7, 11\}$	$switches = \{4, 7, 11, 5\}$
$ports = \{p_1, p_2, p_3\} = \{0, 2, 0\}$	$ports = \{0, 2, 0, 0\}$
$M = 4 \cdot 7 \cdot 11 = 308$	$M = 4 \cdot 7 \cdot 11 \cdot 5 = 1540$
$M_1 = 77, M_2 = 44, M_3 = 28$	$M_1 = 385, M_2 = 220, M_3 = 140, M_4 = 308$
$L_1 = \langle M_1^{-1} \rangle_{s_1} = \langle 77^{-1} \rangle_4 = 1$	$L_1 = \langle 385^{-1} \rangle_4 = 1$
$L_2 = \langle 44^{-1} \rangle_7 = 4$	$L_2 = \langle 220^{-1} \rangle_7 = 5$
$L_3 = \langle 28^{-1} \rangle_{11} = 2$	$L_3 = \langle 140^{-1} \rangle_{11} = 7$
$R = \langle L_1 \cdot M_1 \cdot p_1 + L_2 \cdot M_2 \cdot p_2 + L_3 \cdot M_3 \cdot p_3 \rangle_M$	$L_4 = \langle 308^{-1} \rangle_5 = 2$
$R = \langle 0 + 352 + 0 \rangle_{308} = 44$	$R = \langle 0 + 2200 + 0 + 0 \rangle_{1540} = 660$

It can be noticed in Eq. (4) that the Route ID does not store or keep the information about the switch sequence the packet would travel along. Data from each switch (switch

ID and port index) belong to its own addend of the summation and does not influence the other summation addends. As the finite summation is commutative, the switch order is irrelevant to derive the route ID. This property allows embedding, in the route ID, extra switches that are disjoint of the desired route. This is the fundamental concept of the Driven Deflection Forwarding Paths, and it is useful to protect a desired route when a packet is deflected due to a faulty link since it is possible to guide deflected packets to destination via path segments.

3. Analytic Modeling of Deflection Routing

The model presented in this paper is designed to understand the random effect of the deflection techniques HP, AVP and NIP. The basis of the analytic model is Markov chain with absorbing states, as it can be used to trace the random walk of a test packet through a network under deflection routing [Forghieri et al. 1995].

Let N be number of nodes of a given network and Π be the $N \times N$ transition matrix whose elements π_{ij} represent the probability of a *packet* move to node i at its $(k+1)$ th hop, being at node j at its k th hop. This is a static information for a given (static) topology. The vector $\mathbf{p}(k)$ represents the probability distribution of the test packet at k th hop. That is, its $p_i(k)$ elements represent the probability of the packet be at i th node at its k th hop. Therefore, it yields

$$\mathbf{p}(k+1) = \Pi \cdot \mathbf{p}(k) \quad (9)$$

and

$$p_i(0) = \begin{cases} 1 & , \text{if node } i \text{ is source node} \\ 0 & , \text{otherwise} \end{cases} \quad (10)$$

If node d is the destination node, it retains the received packets. As packets cannot go anywhere else, their next hop is their current node.

$$\pi_{id} = \begin{cases} 1 & , \text{if node } i \text{ is the destination node } d \\ 0 & , \text{otherwise} \end{cases} \quad (11)$$

Finally, let n_j be the number of active ports of j th switch.

Each iteration described by Eq. (9) represents a hop and the iterations continue until the destination node has 100% of packages. In this way, it is possible to find the cumulative distribution function of the number of hops needed to reach the destination node from source node, by recording the state vector at each step.

The rest of this section describes how to fill the transition matrix and the state vector for each deflection technique.

Under **HP deflection**, the probability of a packet going to node i from node j is $\pi_{ij} = 1/n_j$, if j has a direct active link with i , or zero otherwise. For example, for the network of Figure 1, the resulting matrix Π is

$$\Pi_{HP} = \begin{bmatrix} 0 & 1/3 & 1/3 & 0 & 1 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 & 1 \end{bmatrix} \quad (12)$$

The first column represents the next hop probability for packets at first node, SW4. As it has three ports, a packet has 33.33% chance to be sent to SW5, to SW7 or back to

S. The last column represents the destination D, and that is why it has 100% of packets going to itself.

The source node for the analysis of HP, unlike the further discussed techniques, cannot be the real one but the node attached to the faulty link. Otherwise, the test packet would not follow the right path until the faulty link, as the probabilities of the traversed path previous to the faulty link were 100% until failure point: it was not marked as a diverted packet yet. So, the computed $\mathbf{p}(k)$ actually represents the state after $k+h$ hops, where h is the number of hops a packet does to reach the failing link on the path from the ingress node.

As for the **AVP**, the state transition matrix Π depends also on the route ID and the switch IDs. When $\langle routeID \rangle_{switchID}$ matches an active port index, this should become the only possible outgoing port. As a result, the element in the row representing the output port is 1.0; others are 0. If $\langle routeID \rangle_{switchID}$ does not match an available port, the switch j randomly forwards it to any of its active ports. So, $\pi_{ij} = 1/n_j$ if j has a direct active link with i , and zero otherwise.

Adopting the route ID 44 computed in section 2.2 for the route S-SW4-SW7-SW11-D, and performing the modulo operation $\langle routeID \rangle_{switchID}$, we find out the output port for each node.

Switches SW4 and SW11 have valid computed ports. Switch 5 has no port numbered as 4 ($\langle 44 \rangle_5 = 4$). In such case, the packet destination is random. Hence, the second column of Π is equal to the HP matrix transition. As D is the destination node, its column keeps unchanged compared to Π_{HP} . When analyzing the failure in link SW7-SW11, SW7 has only two available ports: back to SW4 or deflect to SW5. This can be seen in Eq. (13) in the third column of the following matrix.

$$\Pi_{AVP} = \begin{bmatrix} 0 & 1/3 & 1/2 & 0 & 1 & 0 \\ 1 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (13)$$

For NIP, the state vector $\mathbf{p}(k)$ now has to account for the fact that a given switch bearing the *test packet* at k th hop must not receive it back at $k+2$ th hop. This requires the $\mathbf{p}(k)$ to encompass N^2 states, instead of just N , in order to trace the previous node the test packet has visited. Consequently, the transition matrix Π grows now to $N^2 \times N^2$. The first element of the state vector, $p_1(k)$, represents the probability of the test packet being in the first node coming from itself at k th hop. The second element, in its turn, represents a packet in node 1 coming from node 2, and so on until the N th element. The value of $p_{N+1}(k)$ ($p_5(k)$ in our example) is the probability, at k th hop, of the packet being in second node coming from the first one. Figure 2 illustrates a hypothetical $p(k)$ value for our network example. At a k th hop, node 1 (SW4) has 20% of packets, and they were all at node 4 (D) in previous hop. In its turn, node 2 (SW5) has 40% of packets at this state: 30% came from node 1 (SW4) and 10% from node 3 (SW7). Hence, the matrix Π must be created avoiding the chance of the traveling packet going back to its previous state when the hot-potato is performed only.

The state transition probability matrix is now expanded to 36x36 in order to sup-

port this previous-hop information. The matrix is not shown due to space limitation.

$$p(k) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0.2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.4 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.4 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

} Probability to be at switch 4 at k th-hop: 20%
 } switch 5 40%
 } edge switch of D 40%

Figure 2. Example of state vector content at hypothetical k th hop for NIP

4. Evaluation

The KAR investigation considers one general topology¹ (Fig. 3(a)) with a 15-node network to represent an experimental scenario to illustrate the main insights of KAR routing deflection approach.

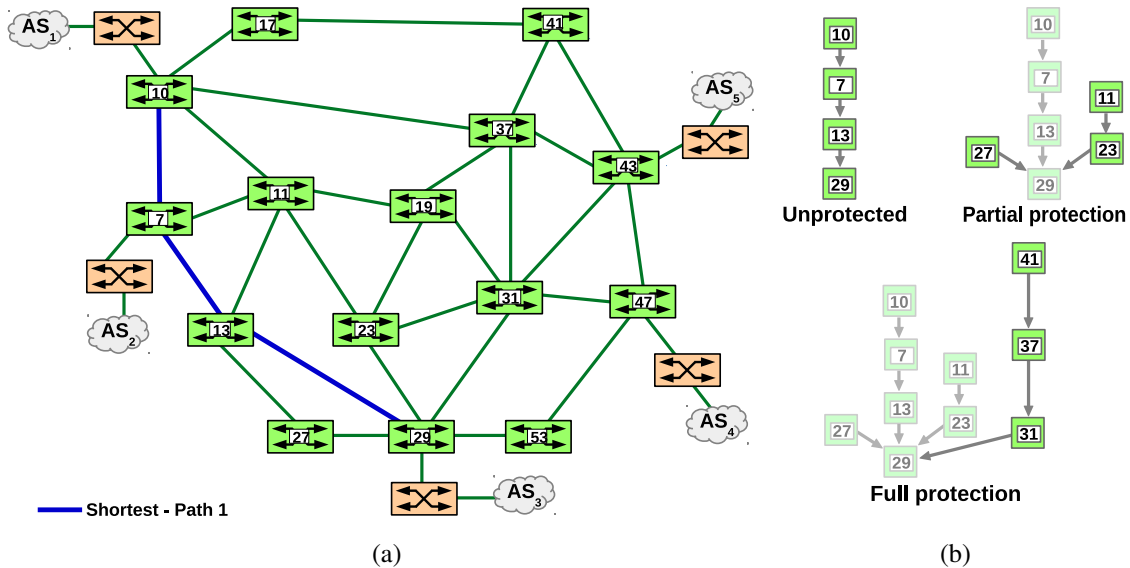


Figure 3. KAR Analysis: (a) topologies with 15-node network, (b) protection mechanism for resilient routing.

Our evaluation methodology is structured in two parts. The first part is related to the validation of our deflection routing model. The second analysis considers the number of hops taken by packets to reach their destination as the performance metric. This study investigates the impact of driven deflection forwarding paths as the network protection mechanism to ensure loop-free (safety) condition.

¹A common range for network topologies <http://www.topology-zoo.org/>

4.1. Deflection Routing Model Validation

In order to validate the analytic model, the network scenario of Fig. 3(a) has been simulated using OMNeT++ V3.0, a discrete event simulator. Simple custom modules were created to simulate KAR nodes using the proposed deflection techniques (HP, AVP and NIP). After each single link failure, 1,000,000 packets are sent by the same route without protection. The number of hops taken by each packet is evaluated.

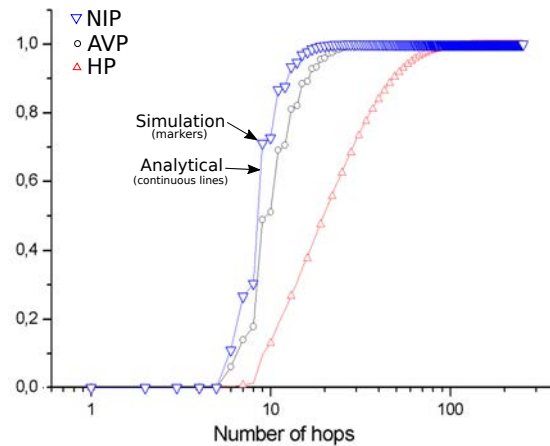


Figure 4. Cumulative Distribution Function (CDF) of the number of hops (Model versus Simulation).

The diameter of the topology Fig. 3(a) was selected (from AS_2 to AS_4 via 7, 11, 19, 31, 47) as the route under analysis. Fig. 4 shows a comparison between analytical and simulation approaches. It presents Cumulative Distribution Functions for both of them, considering the number of hops to reach the destination as a performance metric.

More specifically, Fig. 4 is related to the failure at link (SW11-SW19) where the continuous lines represent the analytical values and the markers (triangles and circles) identify the results obtained by simulation. As it can be seen, the analytical model approach matches precisely the simulation results for the three deflections techniques. The differences between them were lower than 10^{-3} for all cases. It is worth to mention that all other failures at the selected path (diameter) had accuracy at the same order of magnitude.

4.2. Analysis of KAR Resilient Routing System

Let us assume that the shortest path of Fig. 3(a) (Path 1) is selected to allow a communication between AS_1 and AS_3 . If a failure happens at the link (SW10-SW7), then the NIP deflection technique chooses one link uniformly among 3 links to forward packets to switches SW17, SW37 and SW11. Although it implies that packets are not lost, they are likely to be disordered at destination or even to enter in a loop. Thus, a partial protection for KAR resilient routing can be built adding switches SW11 and SW23 to the route ID (Fig. 3(b)), so that packets deflected to SW11 with $1/3$ of probability will be all driven to their destination. In the case of failures at links (SW7-SW13) or (SW13-SW29), they are protected by this tree branch (SW11-SW13). Adding SW27 allows to fully protect the failure near the destination SW29. Therefore, in case of failure at link (SW13-SW29), half of packets is forwarded to SW11 with one more hop than the other half sent to SW27.

Besides, there is still $2/3$ of packets that will be sent to switches SW17 or SW37 in a failure of link (SW10-SW7) close to the source. In order to support the highest resilience for such a route (from AS_1 to AS_3), a full protection route can be ensured by including a branch coming from SW41, e.g. SW41-SW37-SW31, see Fig. 3(b).

In order to evaluate the performance of KAR driven deflections and understand the contribution of the different protection mechanisms and deflection techniques proposed, we considered each individual possible link failure in the shortest path 1 from AS_1 to AS_3 and generated the CDF curves that outline the probability that packets will arrive in the destination within a value less than or equal to a given number of hops. Fig. 5, 6 and 7 consider the selection of shortest **Path 1** (10; 7; 13; 29).

Fig. 5 presents **Path 1** results for a failure in SW10-SW7 link considering Unprotected (Fig. 5(a)), Partial Protection (Fig. 5(b)) and Full Protection (Fig. 5(c)) mechanisms. Regarding protection mechanisms using AVP or NIP deflection techniques, Full Protection curve is much steeper when compared to Unprotected and Partial Protection mechanisms. For instance, considering 99th percentile, Full Protection takes 7 hops for packet delivery while Unprotected and Partial Protection mechanisms take 19 hops. To understand NIP behaviour in this case, consider that when the failure in SW10-SW7 happens there are three possible next hops (SW11, SW17, SW37) with equal probability ($1/3$). However, for Partial Protection, only SW11 hop is included in the defined alternative paths, while for Full Protection all three possible hops lead to protected paths.

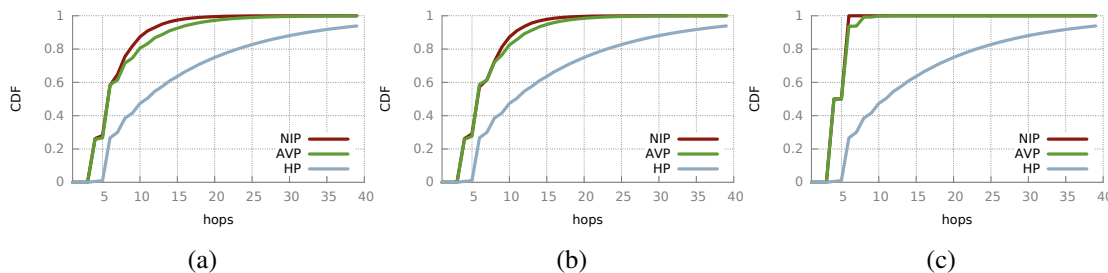


Figure 5. CDF of number of hops to reach AS_3 from AS_1 using shortest Path 1 - varying deflections techniques (failure at link SW10-SW7): (a) Unprotected, (b) Partial Protection, (c) Full Protection.

Fig. 6 presents **Path 1** results for a failure in SW7-SW13 link considering Unprotected (Fig. 6(a)), Partial Protection (Fig. 6(b)) and Full Protection (Fig. 6(c)) mechanisms. This figure depicts no difference in the application of the different protection mechanisms, even in the case of Unprotected mechanism. Regarding deflection techniques, it is possible to note that NIP took about half the number of nodes to reach probability 1 when compared to AVP. Partial and Full Protection have similar behavior, because when the failure in SW7-SW13 happens there are two possible next hops (SW11, AS_2) with equal probability ($1/2$) and, in both cases, SW11 hop is included in the defined alternative paths while AS_2 returns the packet to SW7.

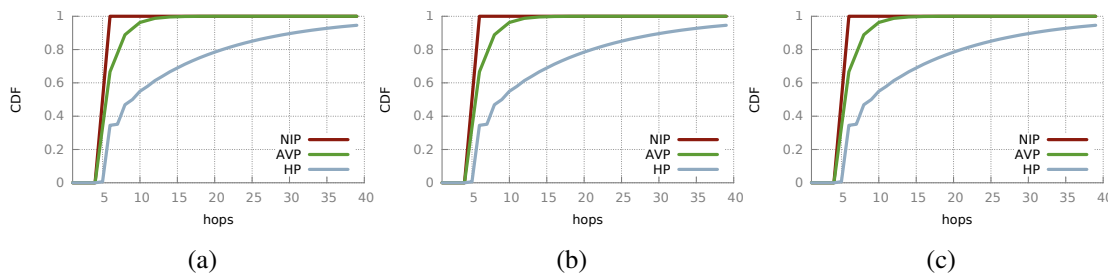


Figure 6. CDF of number of hops to reach AS_3 from AS_1 using shortest Path 1 - varying deflections techniques (failure at link SW7-SW13): (a) Unprotected, (b) Partial Protection, (c) Full Protection.

Fig. 7 presents **Path 1** results for a failure in SW13-SW29 link considering Unprotected (Fig. 7(a)), Partial Protection (Fig. 7(b)) and Full Protection (Fig. 7(c)) mechanisms. This figure depicts almost no difference in the application of the Full and Partial Protection mechanisms, but shows a much slower convergence for the Unprotected case. This performance can be explained because the point of failure is fully surrounded by nodes that are part of the protection path. Regarding deflection techniques, it is possible to note that NIP took about half the number of nodes to reach probability 1 when compared to AVP for Full and Partial Protection, but this difference between NIP and AVP is smaller when using Unprotected Protection. This difference is justified by the fact that AVP allows the packet to return to the previous hop (SW13) while all NIP options guide the packet in the direction of the destination.

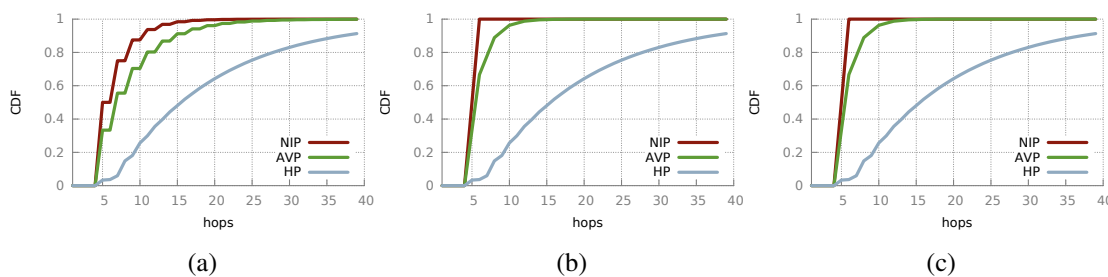


Figure 7. CDF of number of hops to reach node AS_1 from AS_3 using shortest Path 1 - varying deflections techniques (failure at link SW13-SW29): (a) Unprotected, (b) Partial Protection, (c) Full Protection.

For all the results presented in this section, it is important to observe that HP deflection technique increases substantially the number of hops, when compared to AVP and NIP (more than 40 hops). Therefore, the results from the application of HP technique serves as baseline to outline the gains from NIP and AVP techniques in each scenario.

5. Related Work

There has been much work on failure reaction within a network. The most closely related works include MPLS Fast Reroute [Swallow et al. 2005], SafeGuard [Li et al. 2009], and OpenFlow 1.3 Fast Failover [Sharma et al. 2012]. The common part among these proposals is the precomputation of alternative paths to each destination for intra-domain routing, so a router can locally switch to the alternative path without waiting for a topology convergence process. However, these approaches require network states stored at the switches tables (statefull) and lack the flexibility of source-controlled routing. In the case of MPLS Fast Reroute, it still requires the support of a signaling protocol such as Label Distribution Protocol (LDP) for MPLS-enabled switches.

Although source-controlled routing is not in mainstream use of the Internet today, perhaps because source routes do not fit the Internet model in which ISPs set routing policy based on destination addresses, this approach has inspired many innovative proposals for future Internet architectures [Yang and Wetherall 2006, Motiwala et al. 2008, Nguyen et al. 2011, Ramos et al. 2013, Martinello et al. 2014]. Among the main reasons to revisit this approach are [Lee et al. 2015]: i) The data plane becomes simpler because core nodes perform very simple forwarding operations; ii) Traffic engineering is more flexible, allowing application-optimized path selection at the source; iii) Routing stability is improved (e.g. no transient loops) since the path computation is centralized at the source.

Slick Packets [Nguyen et al. 2011] and SlickFlow [Ramos et al. 2013] were proposed to achieve fast data plane failure reaction by embedding alternative routes within the packet headers at the source. The idea is to represent the paths as a sequence of segments that will be used by each switch (or router) to perform the forwarding operation. Also, [Yang and Wetherall 2006] and [Motiwala et al. 2008], both use path label bits set by the source to pseudo-randomly select a next hop at each router or AS. In [Yang and Wetherall 2006], pseudo-random forwarding can lead to forwarding loops. In [Motiwala et al. 2008] routers follow certain rules that ensure loop-free, but reduce path diversity. In contrast to previous works, KAR network core is stateless and its forwarding strategy does not depend on the network topology. The second important difference is related to the driven deflection forwarding paths as the resilient routing mechanism for network protection. Rather than define the complete protection path, only small parts of a path can be included or even a unique node can be added to the route ID. This gives more flexibility, keeps the network core fast and simple and does not increase the route ID size.

RNS-based forwarding is used by [Martinello et al. 2014]. However, KAR routing system has focused on resilient routing that is not taken into account in KeyFlow proposal. In contrast, KAR advances the state of art by dealing with failed links by using routing deflections. In addition, one of the main contributions of KAR is our analytical model approach for evaluating deflection routing. This model was validated against simulation results and allowed us to obtain probabilistic insights on different methods of deflection.

6. Conclusion and Future Work

This paper proposed the analytical modeling of Key-for-Any-Route (KAR), that is a novel fast failure reaction scheme to avoid packet loss and improve resiliency for intra-domain routing systems. Our proposal combines a source routing technique based on the *Residue Number System* (RNS) with *Driven Deflections* property in order to enable efficient routing even in the event of a link failure. This network routing strategy uses stateless core switches, which provide high forwarding performance with the use of simple, low-cost switches. It is a resilient routing scheme because the protection paths enable the packet delivery after a failure through loop-free alternative paths without any reconfiguration on the network nodes.

Analytical models were proposed for three deflection techniques (HP, AVP, NIP) and checked against numerical simulations for different protection mechanisms (Unprotected, Partial Protection, Full Protection). Results show that KAR efficiently allows deflected packets to automatically reach their destination and that NIP and AVP techniques presented substantial performance improvements when compared with a lower bound classical HP technique. NIP deflection with fully protected path presents itself as the best combination in terms of failure reaction.

Finally, mainly when using NIP deflection, KAR poses as a fast failure reaction scheme with small stretch, which is the ratio of the number of hops to achieve the destination after a failure to the original shortest path. In the analyzed scenarios, the addition of protection paths reduces the stretch by at least half.

As future work, we plan to explore the use of multiple paths in the case of redundant links and to investigate the application of KAR in the service chaining of virtualized network functions. Furthermore, we intend to extend the conclusions of the analytical model to enable the prediction of traffic engineering information that could be used to

define more efficient routes.

Acknowledgments

This work has received funding from CNPq, CAPES and FAPES. In addition, it is part of the FUTEBOL project, which has received funding from the European Union's Horizon 2020 for research, technological development, and demonstration under grant agreement no. 688941 (FUTEBOL), as well from the Brazilian Ministry of Science, Technology and Innovation (MCTI) through RNP and CTIC.

References

- Ding, C., Pei, D., and Salomaa, A. (1996). *Chinese Remainder Theorem: Applications in Computing, Coding, Cryptography*. World Scientific Publishing Co., Inc., River Edge, NJ, USA.
- Forghieri, F., Bononi, A., and Prucnal, P. (1995). Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks. *IEEE Transactions on Communications*.
- Garner, H. L. (1959). The residue number system. *Transactions on Electronic Computers*, pages 140 – 147.
- Lee, T., Pappas, C., Basescu, C., Han, J., Hoefler, T., and Perrig, A. (2015). Source-based path selection: The data plane perspective. In *The 10th International Conference on Future Internet, CFI '15*, pages 41–45, New York, NY, USA. ACM.
- Li, A., Yang, X., and Wetherall, D. (2009). Safeguard: safe forwarding during route changes. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies, CoNEXT '09*, pages 301–312, New York, NY, USA. ACM.
- Martinello, M., Ribeiro, M., de Oliveira, R., and de Angelis Vitoi, R. (2014). Keyflow: a prototype for evolving sdn toward core network fabrics. *Network, IEEE*, 28(2):12–19.
- Motiwala, M., Elmore, M., Feamster, N., and Vempala, S. (2008). Path splicing. *SIGCOMM Comput. Commun. Rev.*, 38(4):27–38.
- Nguyen, G. T., Agarwal, R., Liu, J., Caesar, M., Godfrey, P. B., and Shenker, S. (2011). Slick packets. *SIGMETRICS Perform. Eval. Rev.*, 39(1):205–216.
- Ramos, R. M., Martinello, M., and Rothenberg, C. E. (2013). Slickflow: Resilient source routing in data center networks unlocked by openflow. *IEEE Conference on Local Computer Networks*, pages 01–08.
- Sharma, S., Staessens, D., Colle, D., Pickavet, M., and Demeester, P. (2012). A Demonstration of Fast Failure Recovery in Software Defined Networking. *TRIDENTCOM 2012*, pages 411–414.
- Swallow, G., Pan, P., and Atlas, A. (2005). RSVP-TE fast reroute. RFC 4090, <http://www.ietf.org/rfc/rfc4090.txt>.
- Yang, X. and Wetherall, D. (2006). Source selectable path diversity via routing deflections. *SIGCOMM Comput. Commun. Rev.*, 36(4):159–170.

FLECHA: a Forecasting eLEction meCHANism for semantic collectors sensor nodes

Mauricio M. Neto¹, Leonardo O. Moreira², Danielo G. Gomes¹

¹Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat/UFC)
Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI/UFC)

² Instituto Universidade Virtual (UFC Virtual)

Universidade Federal do Ceará (UFC)
Fortaleza – CE, Brasil

{mauriciomoreira, dgomes}@great.ufc.br, leoomoreira@ufc.br

Abstract. *Due to the resource constraints of the sensor nodes, energy provisioning in wireless sensor networks remains a challenging task, particularly in outdoor scenarios. Among the literature proposals to mitigate this problem, we highlight semantic clustering as a recent energy-efficient technique for prolonging the network lifetime. In semantic clustering, each cluster has a semantic leader (collector) which is periodically elected according to an energy-related criterion. However, since collectors' energy depletion is faster than the others members of their cluster, suitable election mechanisms are required to avoid the energy hole problem. Here we propose FLECHA, a mechanism based on the ARIMA model to predict semantic collectors elections with leader-node alternation. Our hypothesis is that by anticipating the best candidates to semantic leaders, we can improve the energy-saving at the node-level, and hence allow the network lifetime to be further extended.*

1. Introduction

Technological advances in microelectronics and wireless communications have enabled a fast development of the so-called Low-Power and Lossy Networks (LLN) [Remy 2015], including Wireless Sensor Networks (WSN), which are comprised of several sensing embedded nodes. Nowadays, these small-size nodes play a key role as a sensing and communication infrastructure for the smart-x applications [Somov and Giaffreda 2015, Mόνton and Picone 2015], such as smart environment, smart metering, smart agriculture, and green smart cities [Gomes and Forster 2015]. Although there are interesting energy-harvesting solutions in outdoor and low power applications [Frohlich et al. 2015], most of the sensor nodes are battery-operated [Hermeto et al. 2013], making energy efficiency a critical goal.

,To achieve a higher energy efficiency related to data transmission, a commonly used technique in wireless sensor networks is clustering. In such approach, the whole set of network nodes is divided into partitions controlled by an elected leader node, namely Cluster Head (CH) [Deshpande and Bhagat Patil 2013]. To form the clusters, physical clustering commonly uses physical metrics such as Euclidean distance and residual energy [Jannu and Jana 2014]. The proximity of the nodes is one of the main parameters of clustering in a physical clustering. Semantic clustering, instead, uses the similarity of

data collected by nodes, i.e., when certain nodes detect the same kind of data, the semantic cluster is created [Hermeto et al. 2013, Rocha et al. 2012]. The semantic clustering also uses the concept of leader (semantic collector), analogous to the CH in a physical cluster.

Concerning the differences between physical and semantic clustering, we highlight that a semantic cluster is independent of physical metrics, such as distance, so that it can be created with separate physical cluster members. It is important to realize that a physical leader can also become a Semantic Collector (SC). However, in case we have two semantic clusters, their collectors cannot be the same node. Given that the semantic clusters are created using data similarity, it is not possible for a same semantic node to join two or more groups. In both physical and semantic clustering, each leader is responsible for receiving data from adjacent semantic nodes (semantic neighbors), summarizing and sending them to the sink node via multihop communication [Avril et al. 2014]. Although energy efficiency in the semantic clustering is better than in the physical clustering [Rocha et al. 2012], the nodes elected as leader also tend to lose a large amount of energy, which leads them to die prematurely. To avoid the energy hole problem, algorithms based on leader-node alternation are required [Jannu and Jana 2014].

Here we present **FLECHA**, a **F**orecasting **e**LECTION **m**eCHANism for semantic collectors in wireless sensor networks. FLECHA uses ARIMA (Auto-regressive Integrated Moving Average) models as the decision mechanism to improve the election of new semantic leaders. ARIMA is considered a general class of models for forecasting a time series. Besides, it has performed well on short-term prediction windows [Moreira et al. 2014, Santos et al. 2013], which can be suitable to avoid temporal disruptions in WSN and other short-term communication failures.

2. Related Work

This section is divided in two subsections that surrounds prior works related to FLECHA, which involve clustering and prediction in WSN.

2.1. Clustering and Prediction in Wireless Sensor Network

Genetic algorithms can be used to select cluster heads in a centralized cluster and can also prolong the network lifetime compared to other clustering protocols such as LEACH [Heinzelman et al. 2002] and LEACH-C [Pal et al. 2015]. The work proposed by Pal et al. (2015) resulted in an improvement of the load balance in relation to a traditional clustering. However, the centralized network approach generates an overhead, causing a greater increase in energy expenditure in the nodes of the network. Our approach is decentralized, resulting in an autonomy of sensor nodes and thus eliminating the network overhead.

The residual energy and distance among the nodes can be used as parameters to elect cluster leaders in a heterogeneous WSN, as DTRE-SEP [Hassan et al. 2015]. The work proposed by Hassan et al. (2015) presented an improvement of the network lifetime in unstable regions compared to clustering algorithms as LEACH. However, such protocols do not have a good performance in networks with random distributions of nodes [Hassan et al. 2015]. FLECHA has autonomic characteristics as self-organization, self-configuration and self-adaptation, having a good performance in random distribution of nodes.

Algorithms based on static and dynamic hierarchical levels of clusters, together with filters to predicts targets positions of the nodes, are used to improve WSNs lifetime, such as proposed by PRATIQUE-a [Souza et al. 2015]. The forecast methods could be utilized to obtain the set of nodes that will detect the next event, reducing the overhead of messages in the network [Souza et al. 2015]. However, the PRACTIQUE-a does not use metrics like residual energy as a parameter to choose cluster leaders, which could generate an improvement in the WSNs lifetime.

2.2. Wireless Sensor Network and Semantic Clustering

The semantic clustering is used for event monitoring in a WSN and its cluster is formed by the similarity of the data captured by the nodes, which elect a semantic leader (collector) [Hermeto et al. 2013, Rocha et al. 2012]. The semantic clustering has better energy efficiency than other classical clustering methods [Rocha et al. 2012].

It is possible to improve the lifetime of the semantic clustering using fuzzy logic to elect the semantic collector [Hermeto et al. 2013]. Using metrics such as residual energy and Euclidean distance it is feasible to have an alternation of leaders to decrease the overhead in the collector (leader) [Hermeto et al. 2013]. FLECHA uses a predictive method to anticipate the election of a new semantic collector through this proactive approach, making it is possible to improve the lifetime of the WSN.

It is possible improve the energy efficiency of a WSN with semantic clustering by introducing a mechanism of decentralized sensor nodes [Rocha et al. 2016]. Fully-decentralized mechanisms tend to have better energy performance than protocols partially decentralized by the reducing the amount of messages and the number of hops in the network [Rocha et al. 2016].

Table 1. Summary of related work.

Reference	Cluster	Self-*	Leader election technique	Leader election parameters
FLECHA (proposal)	Semantic	Organization, configuration and adaptation	ARIMA (prediction)	Residual energy, Euclidean distance
DSENSE [Rocha et al. 2016]	Semantic	Organization, configuration and adaptation	N/A	Fuzzification Message
Hermeto et al. (2013)	Semantic	Organization, configuration and adaptation	Fuzzy Logic	Residual energy
SEMANTK [Rocha et al. 2012]	Semantic	Organization, configuration and adaptation	N/A	Amount of Neighbors
PRATIQUE-a [Souza et al. 2015]	Physical	N/A	N/A	N/A
Pal et al. (2015)	Physical	N/A	Genetic Algorithms	Residual energy
DTRE-SEP [Hassan et al. 2015]	Physical	N/A	N/A	Residual energy, Euclidean distance

Table 1 shows that most related works use the residual energy as a key parameter to (re)elect leaders. SEMANTK and DSENSE use semantic clustering to improve the network lifetime, but the decision to choose the collector is reactive. It is possible to further decrease the energy cost of the WSN using a predictive selection of semantic collectors. FLECHA employs ARIMA to generate predictive curves having residual energy and Euclidean distance as input parameters and, through the trends of these curves, to elect the new semantic leaders of the cluster, avoiding the excessive energy cost.

3. FLECHA: a Forecasting eLEction meCHANism for semantic collectors sensor nodes

FLECHA is a mechanism to elect leaders (collectors) in a WSN based on semantic clustering which metrics are the residual energy and Euclidean distance among nodes. FLECHA extends SEMANTK [Rocha et al. 2012] by using the same oriented events mechanism. Figure 1 shows the FLECHA UML diagram where we can see three modules or phases (separated by colors): (i) data collection and analysis of sensor nodes (yellow); (ii) the detection of semantic neighbors in the physical cluster (orange) and (iii) the predictive election of semantic collectors mechanism (blue). All the FLECHA processes occur locally in each node. In the first phase, each sensor node will collect metrics (e.g., temperature, humidity) based on the application of interest. Moving on the next phase, a new event is expected (e.g., “frequency values higher than X”). In case of the event does not occur, FLECHA comes back to the initial data collection phase. However, if a new event is detected, the detection of semantic neighbor phase on the physical cluster is triggered. This phase is responsible for establishing the relationship between the semantic neighbors based on data similarity [Rocha et al. 2016].

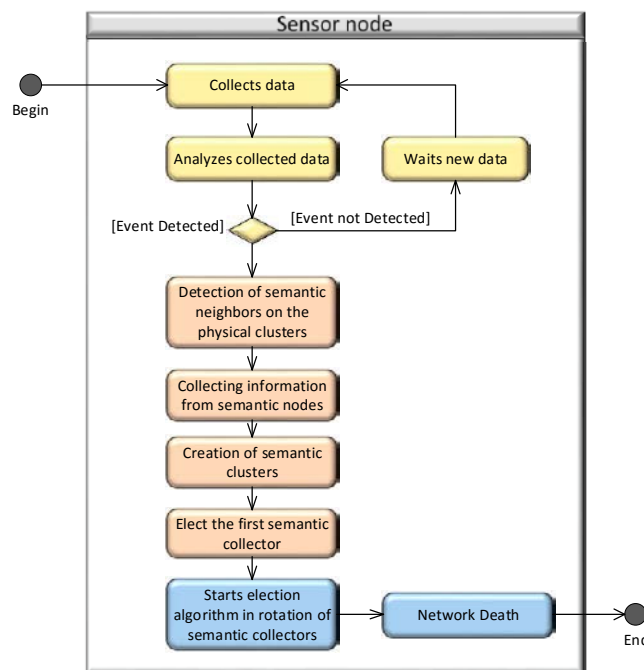


Figure 1. FLECHA UML diagram.

The election of the first semantic collector takes into account the Euclidean distance metric. Despite the main metric considered in FLECHA is the residual energy, it is initially supposed that all nodes of the semantic neighborhood have approximately the same level of residual energy. Consequently, we need to use other metric different from residual energy to elect the first semantic collector. In this work, the node with more semantic neighbors will be the first collector.

The subsequent phase is the predictive election for the next semantic collectors. It is common the cluster leader tends to die before the other nodes of clusters. The semantic collector (leader of the semantic cluster) tends to have a greater energy expenditure than the other sensor nodes of the semantic cluster that only capture the event data.

In our proposal, the election of the next semantic collector occurs after 10% of power depletion of the current collector, which is a reactive approach. To adopt a proactive one, we use ARIMA prediction method to generate the trend curve of residual energy decrease for the collector node. The election of the new semantic collector starts when the trend of the predictive curve is close to 10% of power drop in a current collector (see Algorithm 1). It can be seen by the algorithm 1 that the prediction curve $y[t + n]$ is fed by the residual energy data $x[t]$, where t is the current instant of energy captured and n is the future moment generated by the predictive method.

```

Data: collector election
Result: elect a new semantic collector
initialization (predictiveCurve);
collected data (residualEnergy, euclideanDistance);
 $x[t] \leftarrow residualEnergy$ ;
 $y[t] \leftarrow predictiveCurve$ ;
 $z \leftarrow euclideanDistance$ ;
if  $|y[t + n]| \geq |0.1 \cdot x[t]|$  then
    | collects all the energy and distance data of the semantic cluster nodes;
    | elect the node with larger  $x$ ;
    | if nodes tied == TRUE then
    | | choose the node with more semantic neighbors using  $z$ ;
    | end
else
    | go back to the beginning;
end

```

Algorithm 1: Semantic Collector Election (pseudo-code).

During a new collector election, it is possible that two or more nodes have the same level of residual energy. In this case, the node with more semantic neighbors will be elected as a collector. Finally, when all the nodes of the semantic cluster reach a residual energy value below 10% or the first collector die (First Node Death indicator, see Section 4.2), the WSN will be considered dead.

By using ARIMA, our aim is to elect semantics collectors with better conditions to perform their duties. Thus, we seek to extend the network lifespan by choosing the best semantic collector for the next status of the network. ARIMA was chosen because it performs better with small prediction windows [Santos et al. 2013] [Moreira et al. 2014]. This model can be used to predict certain events like time series showing trends, correla-

tions, and seasonal variations.

4. Material and Methods

The FLECHA mechanism was implemented in C using the Contiki operating system [Dunkels et al. 2004]. The simulations were performed on Cooja [Osterlind et al. 2006], which is used on the communication standard of IEEE 802.15.4 and 250 kbps transmission rate. Cooja also supports various sensor platforms; here we used the MicaZ platform.

4.1. Energy Model

We used the MicaZ energy model as follows [JurdaK et al. 2008]:

$$E_t = P_{send} * P_{length} * TB * I_t * V \tag{1}$$

$$E_r = P_{receive} * P_{length} * TB * I_r * V, \tag{2}$$

where E_t and E_r are, respectively, the energy costs of transmission and packet reception (mJ). P_{send} and $P_{receive}$ represent the number of packets sent and received. P_{length} is the packet size, and TB is the radio response time. I_t and I_r are the respective values of the node radio electric current in the modes of transmission and reception. Lastly, V represents the voltage constant provided by the MicaZ datasheet.

4.2. Evaluation Scenarios

We used the scenarios and values from Hermeto et al. (2013) (Figure 2) but adding more sensor nodes in different situations. We simulated an application for structural health monitoring (SHM) domain as a case study. With a scenario of a five floors building, all physical clusters are deterministic.

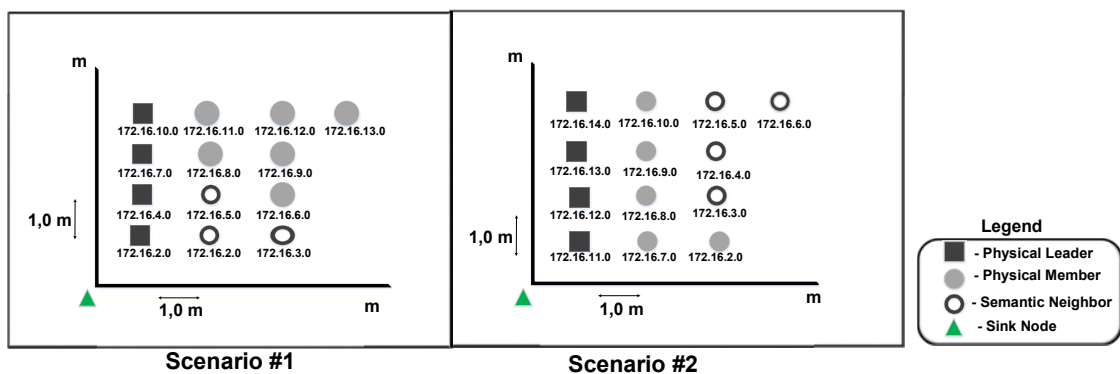


Figure 2. Simulated scenarios (adapted from [Hermeto et al. 2013]).

In these scenarios, the events are triggered simulating damage in the physical structure, which will cause a change of state (shown by modal frequencies, Table 2). When the network starts to operate, all the cluster’s nodes know their respective physical leaders as well the physical leaders know where adjacent leaders are. The sensor nodes transmission time ranged from 500 to 1300 ms and each round lasted 15 min.

At the simulation beginning, all physical nodes are healthy. During the first ten seconds, the physical clusters nodes send messages with the modal frequencies shown in

Table 2. Afterwards, we forced FLECHA to change all nodes frequencies to “Damage 1” for scenario #1 and “Damage 2” to scenario #2, thus causing damage on the 2nd floor (please see Figure 2).

Table 2. Modal Frequencies.

Structural Condition	Mode 1 (Hz)	Mode 2 (Hz)	Mode 3 (Hz)	Mode 4 (Hz)	Mode 5 (Hz)
Damage 1	2.34	7.52	11.62	14.45	17.48
Damage 2	2.44	6.93	11.91	15.14	16.99
Healthy	2.54	7.52	12.01	15.53	17.77

Based on the Euclidean distance, we first identified the nodes closer to the damage. Then, they became semantic neighbors by distribution of the nodes weights. In scenario #1, the IP nodes are 172.16.6.0, 172.16.7.0 and 172.16.8.0. Similarly, in the scenario #2, we used IPs 172.16.9.0, 172.16.11.0, 172.16.13.0 and 172.16.14.0. The event takes place when the frequencies of the nodes are changed, thus generating semantic clustering. When the semantic clustering occurs, the physical leaders know whether there are semantics nodes in their clusters or not. In our simulations, we extended the scenario used in [Hermeto et al. 2013] by doubling the number of nodes.

Since our aim is to reduce the premature loss of semantic collectors, we used the First Node Death indicator [Pal et al. 2015, Dietrich and Dressler 2009]. Our predictive approach was compared to SEMANTK’s election algorithm, which selection criteria for the new semantic leader is the amount of semantic neighbors that exist on each physical cluster, and also to the Hermeto et al.(2013), who used fuzzy logic to elect the leader through the residual energy and distance among nodes.

5. Results and Discussion

Due to their multiple functions, leader nodes have higher energy costs than their neighbors, so that leaders tend to die faster. Thus, we considered the initial energy of all nodes as 5 J. Figure 3 shows the comparison of the semantic collectors lifetime in the scenario #1 between the approaches (a) SEMANTK (b) Hermeto et al. (2013), and (c) FLECHA. Likewise, Figure 4 presents a comparison of lifetime of the semantic collectors in the scenario #2 between the approaches (a) SEMANTK, (b) Hermeto et al. (2013), and (c) FLECHA. Both SEMANTK, Hermeto, and FLECHA algorithms were repeated 20 times for each experiment with 90% confidence intervals. Note that the vertical bars are not shown in figures where they are not visually significant.

Notes that Hermeto’s algorithm took about 350 min to lose the first semantic collector. This result illustrates the gain in survival of about 45% of Hermeto et al. (2013) algorithm with regard to SEMANTK.

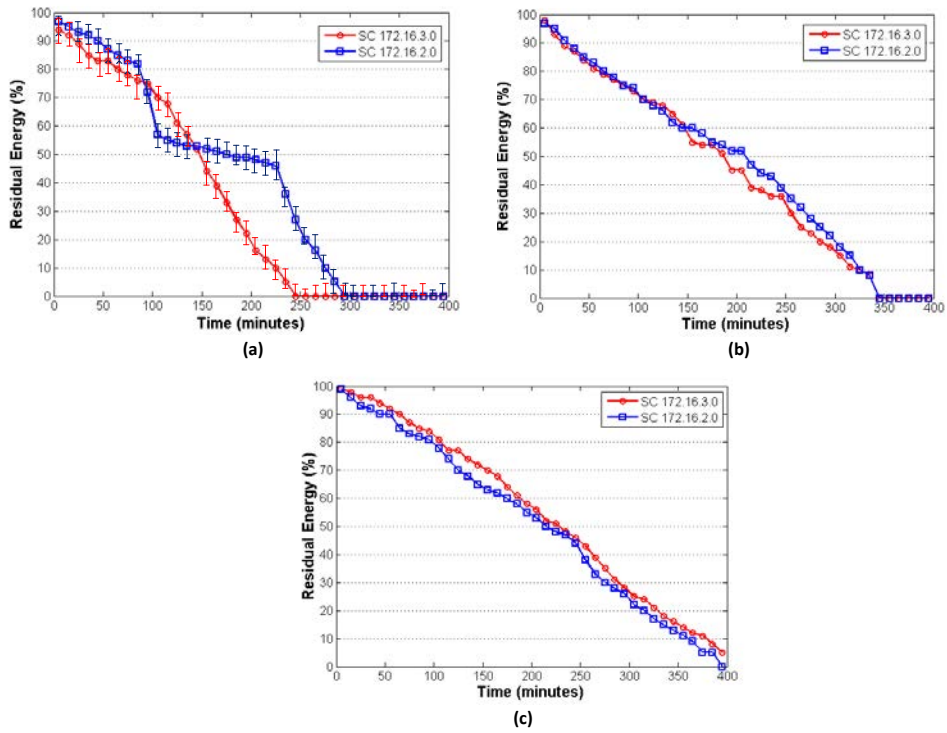


Figure 3. Comparison between approaches in scenario #1.

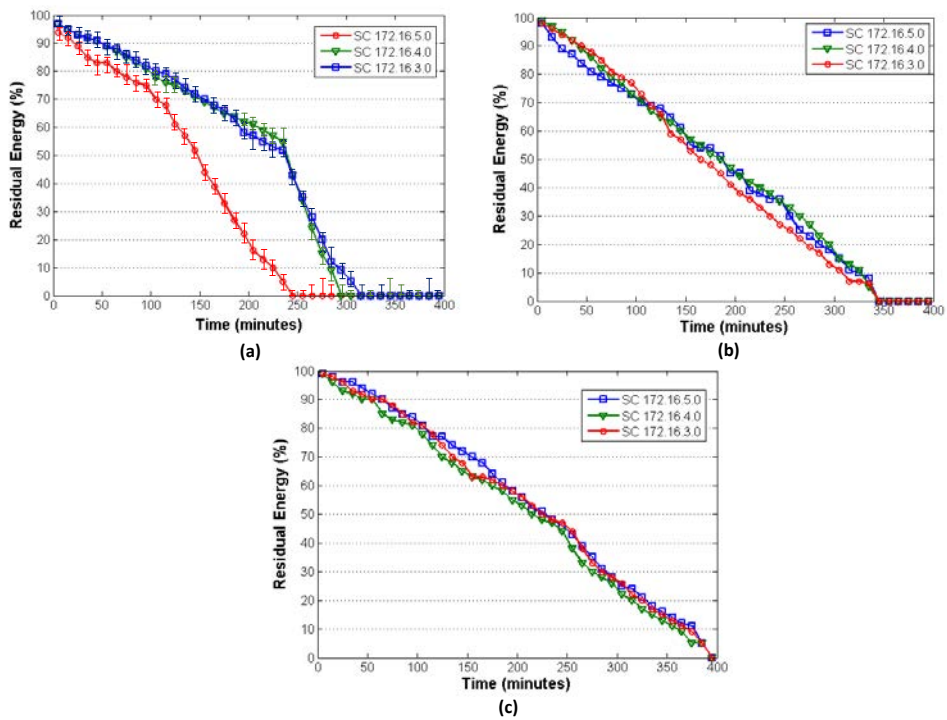


Figure 4. Comparison between approaches in scenario #2.

Since the time when the first semantic node die was around 400 min in the FLECHA approach, it brought a gain of 14% survival compared to the Hermeto approach and of 60% compared to SEMANTK algorithm. It can also be seen from Figures 3 and 4 that the first node was chosen by ARIMA, approximately, during the first 15 minutes of simulation of two scenarios. The last node was elected when it achieves 90% loss of energy. Using FLECHA, when the residual energy nears 10% of the energy loss in the initial state, ARIMA predicts time which will reach this loss. Therefore, the election is made in a predictive way before reaching 10% loss (on average 9%).

Table 3 illustrates the number of nodes chosen in scenarios #1 and #2. We can see in that the SC 172.16.3.0 was elected more times in Hermeto algorithm and FLECHA. We can also note that SC 172.16.5.0 has a better performance in SEMANTK and FLECHA when compared to Hermeto proposal. However, the other nodes elected as collectors (SC 172.16.4.0 and SC 172.16.3.0) present better results through FLECHA and Hermeto algorithm because of the proximity to the sink node.

Table 3. Quantity of nodes elected

Scenario	IP Address	SEMANTK	Hermeto et al. (2013)	FLECHA
#1	172.16.2.0	16	11	12
#1	172.16.3.0	0	12	14
#2	172.16.3.0	0	10	12
#2	172.16.4.0	1	9	9
#2	172.16.5.0	15	8	14

Afterwards, we doubled the number of nodes in scenario #1, which means it had two more candidates to be elected as semantic collectors (SC 172.16.4.0 and SC 172.16.5.0).

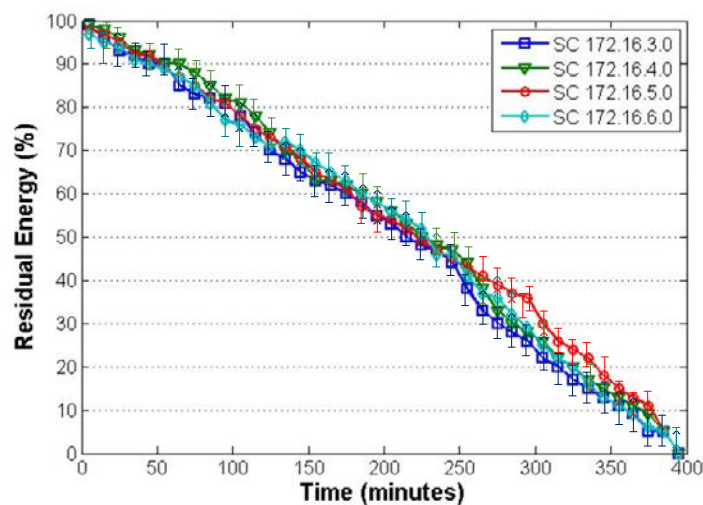


Figure 5. FLECHA execution in scenario #1 with twice nodes.

Figure 5 shows the scenario #1 with twice elected nodes as collectors by FLECHA. When we doubled the nodes number, the network lifetime (400 min) becomes similar to that in Figure 4(c). This is because FLECHA uses ARIMA as predictor to avoid energy waste on the collector node, which causes a new election per collector. As a result, FLECHA allows a gradual network energy depletion and, consequently, increases the network lifetime.

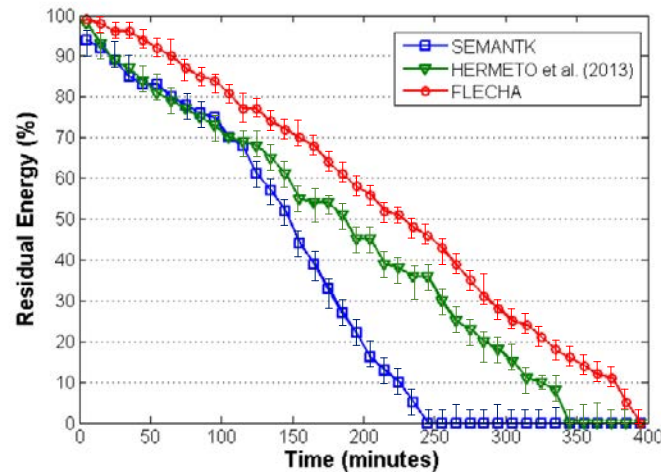


Figure 6. Approaches comparisons in scenario #1

In Figure 6, we see the comparison between SEMANTK, Hermeto et al. (2013) and the FLECHA for the scenario #1. There we note FLECHA brought a gain of 60% in network lifetime when compared to the SEMANTK and 14% compared to Hermeto's algorithm.

In this paper, the leader-node alternation is independent of how far the candidates are from the collectors. This is due to the load balancing of the input variables weights. It should be noted that when an event of interest is detected by a large amount of neighboring clusters, it will lead to an intense exchange of messages among the candidates for collectors until there is a consensus for a new leader election.

6. Conclusion

Here we propose FLECHA, an approach to elect WSN semantic collectors using ARIMA models to forecast the best candidates. FLECHA uses the residual energy of the nodes and the distance between semantic collectors. According to our results, FLECHA improved the network lifetime by 60% compared to SEMANTK [Rocha et al. 2012] and by 14% in relation to Hermeto's proposal [Hermeto et al. 2013]. These results provide insight into the suitability of forecasting algorithms for use in WSNs clustering in such a way that the constrained sensor resources are satisfied.

As future perspectives, we intend to extend and improve FLECHA to make it flexible for mid- and long-term prediction (e.g. using machine learning algorithms) and considering other metrics, such as the most central node, processing time, and memory footprint.

References

- Avril, F., Bernard, T., and Bui, A. (2014). Efficient communication scheduling in clustered wsn. In *Computers and Communication (ISCC), 2014 IEEE Symposium on*, pages 1–6.
- Deshpande, V. and Bhagat Patil, A. (2013). Energy efficient clustering in wireless sensor network using cluster of cluster heads. In *Wireless and Optical Communications Networks (WOCN), 2013 Tenth International Conference on*, pages 1–5.
- Dietrich, I. and Dressler, F. (2009). On the lifetime of wireless sensor networks. *ACM Trans. Sen. Netw.*, 5(1):5:1–5:39.
- Dunkels, A., Gronvall, B., and Voigt, T. (2004). Contiki - a lightweight and flexible operating system for tiny networked sensors. In *Local Computer Networks, 2004. 29th Annual IEEE International Conference on*, pages 455–462.
- Frohlich, A. A., Bezerra, E. A., and Slongo, L. K. (2015). Experimental analysis of solar energy harvesting circuits efficiency for low power applications. *Computers and Electrical Engineering*, 45:143 – 154.
- Gomes, D. G. and Forster, A. (2015). Introduction to the special issue on green engineering: Towards sustainable smart cities. *Computers and Electrical Engineering*, 45:141 – 142.
- Hassan, S., Nisar, M., and Jiang, H. (2015). Dtre-sep: A direct transmission and residual energy based stable election protocol for clustering techniques in hwsn. In *Communication Software and Networks (ICCSN), 2015 IEEE International Conference on*, pages 266–271.
- Heinzelman, W. B., Chandrakasan, A. P., and Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1(4):660–670.
- Hermeto, R. T., Kridi, D. S., Rocha, A. R., and Gomes, D. G. (2013). A distributed algorithm for semantic collectors election in wireless sensors networks. *Journal of Applied Computing Research*, 3(12):1–10.
- Jannu, S. and Jana, P. (2014). Energy efficient grid based clustering and routing algorithms for wireless sensor networks. In *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*, pages 63–68.
- Jurdak, R., Ruzzelli, A., and O’Hare, G. (2008). Adaptive radio modes in sensor networks: How deep to sleep? In *Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON '08. 5th Annual IEEE Communications Society Conference on*, pages 386–394.
- Mónton, M. and Picone, M. (2015). An open-source cloud architecture for big stream iot applications. *Interoperability and Open-Source Solutions for the Internet of Things: International Workshop, FP7 OpenIoT Project, Held in Conjunction with SoftCOM 2014, Split, Croatia, September 18, 2014, Invited Papers*, 9001:73.
- Moreira, L. O., Farias, V. A. E., Sousa, F. R. C., Santos, G. A. C., Maia, J. G. R., and Machado, J. C. (2014). Towards improvements on the quality of service for multi-tenant rdbms in the cloud. In *Data Engineering Workshops (ICDEW), 2014 IEEE 30th International Conference on*, pages 162–169.

- Osterlind, F., Dunkels, A., Eriksson, J., Finne, N., and Voigt, T. (2006). Cross-level sensor network simulation with cooja. In *Local Computer Networks, Proceedings 2006 31st IEEE Conference on*, pages 641–648.
- Pal, V., Yogita, Singh, G., and Yadav, R. (2015). Cluster head selection optimization based on genetic algorithm to prolong lifetime of wireless sensor networks. *Procedia Computer Science*, 57:1417 – 1423. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).
- Remy, L. (2015). Smart gateway for low-power lossy networks. In *Proceedings of the 2015 on MobiSys PhD Forum, PhDForum '15*, pages 13–14, New York, NY, USA. ACM.
- Rocha, A. R., Delicato, F. C., Pirmez, L., Gomes, D. G., and de Souza, J. N. (2016). A fully-decentralized semantic mechanism for autonomous wireless sensor nodes. *Journal of Network and Computer Applications*, 61:142 – 160.
- Rocha, A. R., Pirmez, L., Delicato, F. C., Érico Lemos, Santos, I., Gomes, D. G., and de Souza, J. N. (2012). Wsns clustering based on semantic neighborhood relationships. *Computer Networks*, 56(5):1627 – 1645.
- Santos, G. A. C., Maia, J. G. R., Moreira, L. O., Sousa, F. R. C., and Machado, J. C. (2013). Scale-space filtering for workload analysis and forecast. In *2013 IEEE Sixth International Conference on Cloud Computing*, pages 677–684.
- Somov, A. and Giaffreda, R. (2015). Powering iot devices: Technologies and opportunities. *IEEE Internet of Things Newsletter*.
- Souza, E. L., Pazzi, R. W., and Nakamura, E. F. (2015). A prediction-based clustering algorithm for tracking targets in quantized areas for wireless sensor networks. *Wirel. Netw.*, 21(7):2263–2278.

WPIETFIRTF – III Workshop Pré-IETF/IRTF

Apresentação

O *Internet Engineering Task Force* (IETF) (<http://www.ietf.org>) é uma comunidade internacional de pessoas interessadas em cumprir a missão de produzir padrões que influenciam a maneira como as pessoas projetam, usam e gerenciam os recursos da Internet [RFC3935 (<http://www.ietf.org/rfc/rfc3935.txt>)]. O processo de estabelecer padrões para a Internet é longo, complexo e exige participação intensa, sobretudo de estudantes e pesquisadores. Embora hajam pessoas do mundo inteiro trabalhando ativamente, cada dia mais se verifica a necessidade de novas presenças. Além disso, existem trabalhos e atividades associadas ao *Internet Research Task Force* (IRTF) (<http://www.irtf.org>), que promove pesquisas relevantes à evolução da Internet, através de grupos que trabalham com foco de pesquisas de longo prazo, relacionados a protocolos, aplicações, arquitetura e tecnologia da Internet.

Um grupo, que se reúne através do IETF-LAC, considerou importante a criação de um workshop com o objetivo de expandir a presença Latino Americana no IETF. Além da divulgação em si, considerou-se de extrema importância oferecer um ambiente para que as habilidades de nossos técnicos, estudantes e pesquisadores fossem exibidas, antecipando-se aos três encontros anuais do IETF. Nesse contexto, surgiu a proposta para criar o Workshop Pré IETF/IRTF (WPIETFIRTF), em sua terceira edição em 2016. A primeira edição ocorreu no CSBC 2014, em Brasília, e a segunda edição ocorreu no CSBC 2015, em Recife. Nesta edição, foram submetidos 7 artigos completos e todos foram aceitos.

Leandro Melo de Sales (Universidade Federal de Alagoas)

Luiz Julião Braga Filho (Universidade Presbiteriana Mackenzie)

Christian O'Flaherty (Internet Society)

Comitê de Organização

Coordenação Geral

Leandro Melo de Sales (Universidade Federal de Alagoas - UFAL)
Rafael de Amorim Silva (Universidade Federal de Alagoas - UFAL)
Luiz Julião Braga Filho (Universidade Presbiteriana Mackenzie)
Christian O'Flaherty (Internet Society)

Coordenação Local

Alexandre Agustini (Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS)

Comitê de Programa

Artur Ziviani (Laboratório Nacional de Computação Científica - LNCC)
Carlos Kamienski (Universidade Federal do ABC - UFABC)
Elias Duarte Jr. (Universidade Federal do Paraná - UFPR)
Glaucio Gonçalves (Universidade Federal Rural de Pernambuco - UFRPE)
Heitor Ramos (Universidade Federal de Alagoas - UFAL)
Ivo Calado (Instituto Federal de Alagoas - IFAL)
Javier Baliosian (University of the Republic Uruguay)
Leandro Melo de Sales (Universidade Federal de Alagoas - UFAL)
Lisandro Z. Granville (Universidade Federal do Rio Grande do Sul - UFRGS)
Rafael de Amorim Silva (Universidade Federal de Alagoas - UFAL)
Rafael Sousa (Universidade de Brasília - UnB)
Stenio Fernandes (Universidade Federal de Pernambuco - UFPE)
Thiago de Sales (Universidade Federal de Alagoas - UFAL)

Painéis / Palestras

Estado da Arte do IPv6 no IETF: Padrões e Protocolos

Eduardo Barasal Morales (NIC.br)

A Internet está em evolução constante. A todo momento novas tecnologias e padrões são desenvolvidos e aprovados pela comunidade científica do IETF, impactando assim a vida de todos os usuários da rede. Essa palestra possui o objetivo de apresentar as últimas discussões que aconteceram no primeiro evento do IETF na América Latina com relação ao IPv6, o novo protocolo de internet. Dessa forma, procura-se atualizar os professores e pesquisadores sobre a evolução do protocolo IPv6 e outros padrões relacionados e incentivar/promover a participação no IETF.

Caminhando para a Reunião 100: Um Panorama Sobre o Passado Recente e o Futuro das Ações do IETF/IRTF

Stenio Flávio de Lacerda Fernandes (Universidade Federal de Pernambuco)

A reunião 95 do IETF/IRTF foi um marco importante para a comunidade Latino-Americana de desenvolvimento e pesquisa de protocolos e tecnologias para a Internet, pois pela primeira vez na sua história o encontro ocorreu no continente sul-americano. Um marco claro agora está bem à frente da comunidade: a sua centésima reunião! Nesta palestra, faremos uma reflexão das ações importantes do IETF/IRTF num escopo de 10 reuniões, tratando dos avanços conseguidos em diversos grupos de trabalhos nas últimas 5 reuniões (da 91 a 95), com uma visão das prováveis consolidações nas próximas 5 reuniões (da 96 a 100), bem como dos desafios a serem ainda superados.

Importância de Padrões Abertos para a Internet e os Meios para Contribuir

Alvaro Retana (Cisco)

O IETF / IRTF: história e organização. Os padrões abertos e o IETF. O trabalho atual do IETF e aqueles que devem protagonizar os principais eventos da Internet em futuro próximo. A importância da participação dos diversos segmentos da computação, como voluntários nos diversos grupos de trabalho do IETF e IRTF. O Brasil no cenário das atividades Latino-americanas relacionadas com a Internet.

Painel: Ações Efetivas para Aumentar o Voluntariado da América Latina na IETF/IRTF

Alvaro Retana (Cisco)

Christian O'Flaherty (Internet Society)

Leandro Melo de Sales (Universidade Federal de Alagoas)

Luiz Julião Braga Filho (Universidade Presbiteriana Mackenzie)

Stenio Flávio de Lacerda Fernandes (Universidade Federal de Pernambuco)

Trabalhos aceitos

Análise de Performance do PUSH em Conexões HTTP/2 no Carregamento de Páginas Web

Igor Nogueira de Oliveira (Universidade Federal de Pernambuco)

Wesley Davison (Universidade Federal de Pernambuco)

Djamel Sadok (Universidade Federal de Pernambuco)

Patricia Takako Endo (Universidade Federal de Pernambuco, Universidade de Pernambuco)

Implementação Inicial da RFC 6897 para Auxílio no Tratamento de Fluxos Elefantes

Alan C. Silva (Universidade Federal de Sao Carlos)

Simone Ferlin (Simula Research Laboratory)

Fabio L. Verdi (Universidade Federal de Sao Carlos)

Otimizando a Transmissão TCP para Veículos Aéreos Não Tripulados

Rafael de Amorim Silva (Universidade Federal de Alagoas)

Leandro de Melo Sales (Universidade Federal de Alagoas)

Protocolo de Comunicaciones para Control de la Generación Distribuida de Flujo Multimedia

C. F. Perez-Monte (Universidad Tecnológica Nacional, Universidad Nacional de San Luis)

M. F. Piccoli (Universidad Nacional de San Luis)

M. Perez (Universidad Tecnológica Nacional)

C. Luciano (University of Illinois at Chicago)

S. Rizzi (Argonne National Laboratory)

Problema de Alocação de Controladores SDN: Uma Análise do Custo em Redes de Backbone

Marcelo Santos (Universidade Federal de Pernambuco)

Stênio Fernandes (Universidade Federal de Pernambuco)

Uma Proposta para a Autenticação de Estações de Trabalho em Redes Definidas por Software com Utilização de Certificados Auto-Assinados

Osiel O. Souza (Universidade do Vale do Rio dos Sinos)

Jeferson C. Nobre (Universidade do Vale do Rio dos Sinos)

Um Modelo Formal de Propósito Geral para Simulações de Redes Definidas por Software

Rafael Souza (Universidade Federal de Pernambuco)

Marcelo Santos (Universidade Federal de Pernambuco)

Stenio Fernandes (Universidade Federal de Pernambuco)

Braulio Mello (Universidade Federal da Fronteira Sul)

Análise de Performance do PUSH em Conexões HTTP/2 no Carregamento de Páginas Web

Igor Nogueira de Oliveira¹, Wesley Davison¹, Djamel Sadok¹, Patricia Takako Endo^{1 2}

¹Grupo de Pesquisa em Redes de Computadores e Telecomunicações (GPRT)
Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brazil

²Grupo de Estudos Avançados em Tecnologia da Informação e Comunicação (GREAT)
Universidade de Pernambuco (UPE)
Caruaru – PE – Brasil

{igor.nogueira, davison, jamel}@gppt.ufpe.br, patricia.endo@upe.br

Abstract. *Recently, the HTTP protocol was updated and received several modifications, focused mainly on improvements in the network resource usage. Among these improvements one can cite the addition of push, a feature that allows the server to respond to a request with more than one resource simultaneously. This work presents a performance analysis of push feature on the transport of web pages on HTTP/2 connections. Therefore, experiments were conducted on a prototype using Total Download Time (TDT) as metric, and Web page requests with different amounts and sizes of objects.*

Resumo. *Recentemente, o protocolo HTTP foi atualizado e recebeu diversas modificações, direcionadas principalmente a melhorias na utilização dos recursos da rede. Dentre estas melhorias, pode-se citar a adição do recurso push, que permite que o servidor responda a uma solicitação com mais de um recurso simultaneamente. Este trabalho apresenta uma análise de desempenho do recurso push no transporte de páginas web em conexões HTTP/2. Para tanto, foram realizados experimentos em um cenário real, utilizando TDT (Total Download Time) como métrica de análise, e requisições de páginas web com diferentes quantidades e tamanhos de objetos.*

1. Introdução

Desde sua padronização definida em 1997 pela Internet Engineering Task Force (IETF) na RFC 2068 [Fielding et al. 1997], o protocolo Hypertext Transfer Protocol (HTTP) tem sido utilizado amplamente pela comunidade da Internet. Além de manter sua funcionalidade original de transmissão de arquivos hipertexto, o HTTP também provê o transporte de mídias mais dinâmicas, hoje denominadas hipermídias, tais como áudio e vídeo, não apenas em tempo real, como também de forma cada vez mais interativa. Mesmo com a diversidade de funcionalidades e características dos aplicativos web, o HTTP ainda continua a ser a escolha padrão para transporte de dados.

O HTTP 1.1 não proporciona a melhor utilização da rede possível por utilizar codificação de controle não compactado e em texto plano, sendo este último um fator incremental na complexidade computacional para análise de identificadores de quebra

de linha [Nielsen et al. 1998]. Além disso, segundo os mesmos autores, as restrições tornaram sua documentação desnecessariamente extensa e complexa, dificultando que suas implementações contivessem todas as funcionalidades definidas pelo padrão. Estes e outro pontos já eram identificados como falhos e que demandariam grandes alterações no protocolo.

Recentemente, uma nova proposta do HTTP foi definida [Belshe et al. 2015]. As principais melhorias propostas no Hypertext Transfer Protocol 2.0 (HTTP/2) são direcionadas ao melhor aproveitamento da rede. Para isso foram propostos o uso de codificação binária e uma estratégia específica de compactação para os dados de controle [Peon and Ruellan 2015]. Adicionalmente, a nova versão define multiplexação de requisições e respostas dentro de uma mesma conexão. Com isso, espera-se a diminuição do número de conexões ativas e conseqüentemente menor carga em vários pontos da rede.

Dentre as novas características do HTTP/2, o *push* permite o envio de recursos pelo servidor sem que haja necessidade de uma requisição explícita do cliente, de forma assíncrona. Esta funcionalidade já era, de certa forma, possível na versão 1.1 através da utilização de *pipelining* e técnicas de *polling*. Porém, ela ainda dependia da requisição explícita do cliente antes de efetuar o envio do conteúdo do recurso solicitado. Devido a natureza do modelo solicitação/resposta do HTTP, durante a utilização de *pipelining*, é possível que uma solicitação seja bloqueada por outra em casos de perda de dados ou variações na latência da rede. Este fenômeno é conhecido como Head-Of-Line Blocking (HOL Blocking).

Alguns dos protocolos utilizados como base para a definição do HTTP/2, como o SPDY (SPDY) e o Quick UDP Internet Connections (QUIC), já utilizam o recurso *push* em suas especificações. Contudo, como o HTTP/2 ainda é bastante recente, lançado em maio de 2015, alguns dos recursos definidos nesta versão ainda não são amplamente utilizados por servidores web. Dessa forma, o impacto que estes novos recursos causam ainda não foram extensamente analisados na prática.

Este trabalho apresenta uma análise inicial sobre o recurso *push* no transporte de páginas web em conexões HTTP/2 através de experimentos. Um servidor que implementa o HTTP/2 foi configurado para responder a solicitações de páginas web de diferentes características como quantidade e tamanho de objetos. Com isso foi possível duplicar a carga de rede gerada no carregamento de páginas web de diferentes configurações. A partir deste ambiente foi utilizado um cliente web compatível para acessar e capturar dados de temporização do carregamento de páginas Web em diferentes condições de rede.

Este artigo está estruturado da seguinte forma: a Seção 2 descreve conceitos básicos sobre HTTP/2; a Seção 3 apresenta os trabalhos relacionados a análise de desempenho do push; a Seção 4 descreve os experimentos realizados neste trabalho; a Seção 5 apresenta e discute os resultados obtidos; e por fim, a Seção 6 apresenta as conclusões e trabalhos futuros.

2. HTTP/2

De forma simplificada, o funcionamento do HTTP, independentemente da versão, dá-se da seguinte maneira: um cliente HTTP inicia uma solicitação através de uma conexão TCP para uma porta específica em um servidor HTTP (por padrão, porta 80 para HTTP, e

443 para HTTPS [Internet Assigned Numbers Authority (IANA)]. O servidor, ao receber o pedido, envia de volta um código de *status*, como por exemplo "200 OK", e, opcionalmente, o conteúdo no restante do corpo da resposta. O corpo desta mensagem é normalmente o recurso solicitado, apesar de uma mensagem de erro ou outras informações também poderem ser enviadas no cabeçalho. Os recursos HTTP são identificados e localizados na rede através de uma Uniform Resource Locators (URL), utilizando os esquemas de Uniform Resource Identifier (URI) `http` ou `https` para indicar o uso de conexões inseguras ou seguras, respectivamente.

A Figura 1 descreve um cenário simplificado de operação de uma sessão HTTP.

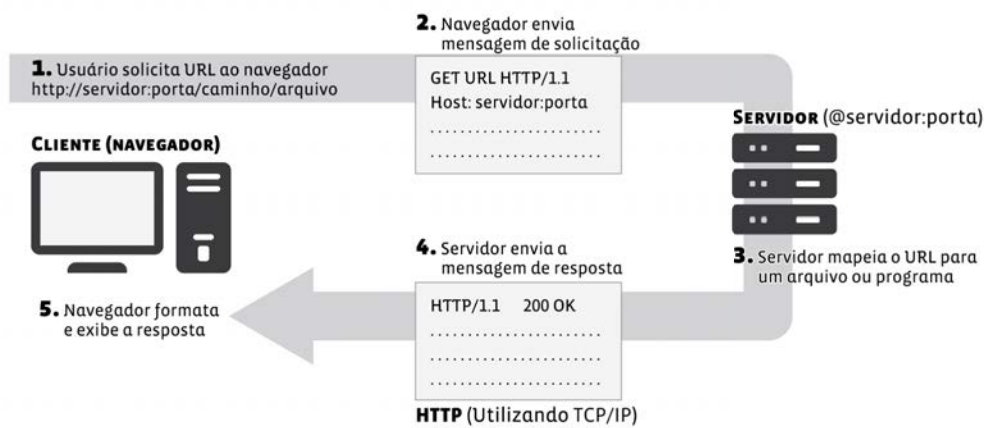


Figura 1. Visão geral de uma sessão HTTP. Adaptado de [Hock-Chuan 2009].

Métodos, como o GET exibido na Figura 1, são utilizados para indicar o tipo de ação em uma determinada solicitação. O HTTP 1.0 definiu 3 métodos, sendo eles: GET, POST e HEAD, utilizados para solicitar conteúdo, enviar conteúdo e solicitar apenas meta informações de um recurso, respectivamente. A versão 1.1 adicionou mais 5 métodos sendo eles: OPTIONS, PUT, DELETE, TRACE e CONNECT. A semântica utilizada para definir estes métodos permite que outros métodos sejam definidos por extensões.

Apesar de ser um protocolo considerado antigo, não houve grandes esforços no desenvolvimento de uma nova versão do HTTP. Apenas em 2007 foi formado pela IETF um novo grupo de trabalho, o HTTPBis. O grupo seria responsável por inicialmente revisar e atualizar todas as considerações definitivas do HTTP 1.1, que resultou nas RFCs 7230 à 7235.

De 2007 até 2012, foram analisados pelo HTTPBis alguns possíveis protocolos que poderiam ser utilizados como base para a versão 2.0. Em julho de 2012, um dos *feedbacks* ao *Call for Expressions of Interest in HTTP/2*¹, fornecido pela equipe de infraestrutura de rede do Facebook², relata quais características consideravam importantes na nova versão para atender as demandas em sua estrutura e quais protocolos já existentes estavam analisando.

¹<http://trac.tools.ietf.org/wg/httpbis/trac/wiki/Http2CfI>

²<http://lists.w3.org/Archives/Public/ietf-http-wg/2012JulSep/0251.html>

Dentre os possíveis protocolos a serem utilizados, como SPDY (formalmente FLIP, Google), HTTP Speed+Mobility (Microsoft) e WAKA (Roy Thomas Fielding, Adobe e Apache Software Foundation), apenas o **SPDY** atendia a maioria das características, além de já ser utilizado internamente na infraestrutura de rede do Facebook e em maior escala pela Google. Neste ponto da história, apenas o SPDY era suportado por dois dos três navegadores Web mais populares, Google Chrome e Mozilla Firefox.

Em novembro de 2012, o primeiro *draft* do HTTP/2 foi publicado, sendo uma cópia direta da especificação do SPDY, até sua padronização oficial pela IETF em maio de 2015 através das Request For Commentss (RFCs) 7540 e 7541. Apesar da existência de diversas implementações e discussões pela comunidade ativa, o processo foi considerado muito rápido e possivelmente tendencioso até sua formalização [Kamp 2014].

As principais limitações do HTTP 1.1, como HOL Blocking, codificação em texto plano e falta de compressão de informações de controle(cabeçalhos), foram consideradas durante o processo de padronização da versão 2.0. Outros fatores técnicos de performance foram identificados para guiar o desenvolvimento do novo protocolo, especificamente, latência observada pelo usuário final, utilização de recursos de rede e recursos dos servidores. Além disso, o principal objetivo era implementar o uso de apenas uma conexão entre o navegador e o servidor Web³.

Dentre os novos recursos do HTTP/2, este trabalho tem como foco o *Server Push*. Através deste recurso é possível que o servidor responda a uma solicitação com mais de um recurso ao mesmo tempo. A implementação deste recurso pode ser descrita sucintamente através dos seguintes passos:

1. O servidor recebe uma solicitação que pode ter **recursos adicionais** como resposta;
2. No mesmo *stream* utilizado para responder à solicitação, o servidor envia um *frame* do tipo PUSH_PROMISE para cada recurso que deseja enviar via *push*. O *frame* PUSH_PROMISE contem um número de *stream* a ser utilizado pelo cliente para identificar futuros *frames* de dados do *push*;
3. O servidor considera como criado os *streams* informados no passo 2 e envia, assim que possível, *frames* de dados dos **recursos adicionais**;
4. O cliente, para cada PUSH_PROMISE recebido, deve optar por:
 - reservar um *stream* com o identificador sugerido pelo PUSH_PROMISE ou;
 - enviar um *frame* do tipo RST_STREAM informando que não irá aceitar o envio de *push*.

3. Trabalhos Relacionados

Estudos anteriores tais como [Padhye and Nielsen 2012] e [Podjarny 2012] analisam performance com uso do SPDY em relação ao HTTP e indicam resultados contraditórios, convergindo apenas no possível baixo desempenho do SPDY em redes móveis.

Em [Erman et al. 2013], os autores efetuam medições ao acessar os 20 sites mais populares do mundo, de acordo com o Alexa, utilizando SPDY e servidores *proxy* HTTP/1 em redes 3G. Eles atestam que o baixo desempenho do SPDY em redes móveis esta

³<https://http2.github.io/>

relacionado à forma como o Transmission Control Protocol (TCP) efetua o controle de congestionamento, em função de perdas e variância de latência da rede (*jitter*). Tendo em vista que o SPDY, assim como o HTTP/2, utiliza apenas uma conexão TCP, o impacto causado pelo controle de congestionamento é mais visível do que no HTTP, que utiliza várias conexões.

Ainda sobre o SPDY, [Wang et al. 2014] apresentou novas métricas e técnicas para caracterizar o desempenho deste protocolo. Através do isolamento do processo de carregamento de rede em relação aos demais processos envolvidos no carregamento de páginas Web, foi identificado que o SPDY possui melhor desempenho em relação ao HTTP. Porém, ao se considerar os demais fatores computacionais, o desempenho do SPDY é consideravelmente afetado. Apesar de propor formas de uso para o recurso *push*, a análise efetuada considera poucos casos, sendo focados números e tamanhos de objetos observados em um conjunto de 200 páginas.

Avançando para o HTTP, o trabalho apresentado em [Saxce et al. 2015] relata análises utilizando HTTP/2 em relação a diferentes fatores de rede como latência, perdas e largura de banda. Os resultados comprovam ganho de performance no uso do recurso *push* em um cenário simples de múltiplos objetos e tamanho fixo. Porém, os autores não estendem as análises em diferentes ambientes.

Por fim, [Han et al. 2015] descreve um *framework* que combina a utilização do *push* em conjunto com *Server Hint* para evitar uso duplicado na transmissão de objetos já armazenados em *cache* pelos clientes. Os comparativos de resultados do uso deste *framework* se baseiam apenas em configurações simples do uso do *push*.

Embora todos os trabalhos relacionado aqui descritos apresentem contribuições significativas, até onde sabemos, não foi encontrada uma análise específica sobre no impacto do uso do recurso *push* em relação aos fatores de tamanho e número de objetos em *streams* concorrentes. Os estudos efetuados consideram cópias de uma pequena fração de páginas Web reais, que podem não representar de forma adequada características da população de Web sites existentes.

4. Experimentos

Os experimentos tem como objetivo comparar o comportamento das **variações no tempo** de carregamento de páginas Web através de conexões HTTP/2 com e sem a utilização do recurso *push*. Os resultados devem ser significativos o bastante para indicar quais fatores devem ser considerados ao se optar por este recurso.

Com o objetivo de recriar ao máximo a utilização de um servidor HTTP/2 em um ambiente real, a escolha da configuração do ambiente do experimento foi focada em soluções mais próximas possível do padrão dos protocolos envolvidos. Na Figura 2, pode-se visualizar as principais entidades envolvidas nos experimentos e como as mesmas estão interligadas.

O ambiente apresentado na 2 descreve um cenário de acesso a um servidor Web HTTP 1.1 utilizando um *proxy* reverso HTTP/2. Decidiu-se utilizar esta configuração de ambiente porque, durante o desenvolvimento deste artigo, o software HTTP/2 utilizado ainda não possuía suporte ao *backend* FastCGI⁴ e, portanto, dificultaria a criação de

⁴<http://www.fastcgi.com/drupal/node/6?q=node/15>

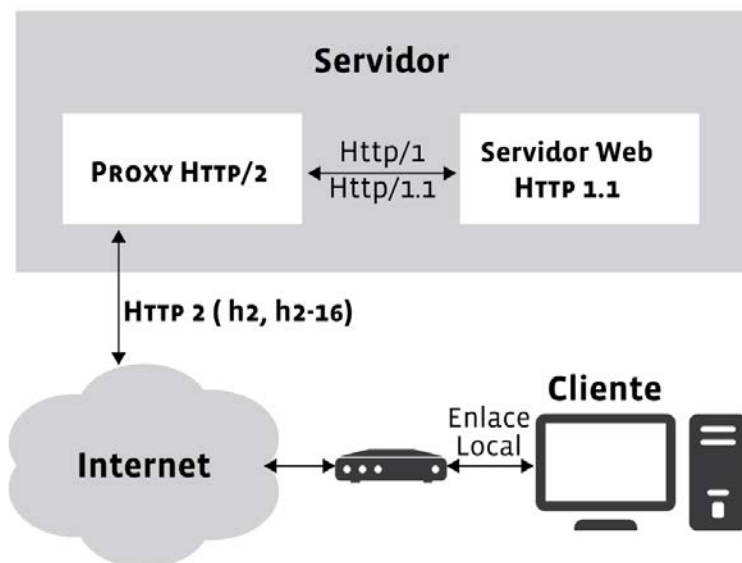


Figura 2. Entidades do ambiente

conteúdo de forma dinâmica.

Todas as requisições do cliente foram geradas por um navegador Web e possuem o mesmo servidor de destino para todos os objetos da página Web. A sequência de passos de uma requisição do experimento realizado são descritos no diagrama de sequência da Figura 3.

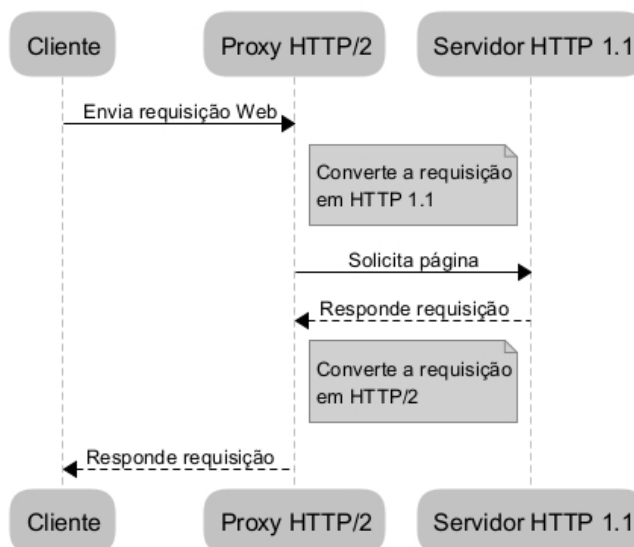


Figura 3. Diagrama de sequência das requisições do cliente

4.1. Infraestrutura de Software

O *backend* de conteúdo do servidor Web HTTP 1.1 utiliza servidor Web Lighttpd versão 1.4.37, configurado na porta TCP 80 e com suporte a *backend* FastCGI habilitado.

Além disso, utiliza *scripts* PHP para gerar conteúdo dinâmico baseado em parâmetros da URL. Desta forma, é possível que um cliente solicite um recurso com tipo e tamanho de dados específico.

Para garantir que solicitações a um mesmo objeto não gerem armazenamento em *cache* pelo navegador Web, além de omitir cabeçalhos de controle de cache, o conteúdo é gerado de forma pseudo-aleatória, através da função `openssl_random_pseudo_bytes`. Esta função não requer acesso a disco para gerar conteúdo, o que garante que os dados gerados sejam rapidamente disponibilizados ao servidor Web.

O **servidor proxy HTTP/2** atua como *gateway* na conexão entre servidor Web e navegador Web do cliente. Este serviço implementa suporte ao protocolo HTTP/2 através do software `Nghttp2`, na versão 1.7.0, configurado na porta TCP 443, utilizando o servidor Web `Lighttpd` como *backend* HTTP. Bibliotecas desenvolvidas neste mesmo software são utilizados em outras ferramentas popularmente utilizadas como `cURL`⁵ e `Wireshark`⁶. O servidor oferece suporte à extensão Application-Layer Protocol Negotiation (ALPN), provido pela camada de segurança Transport Layer Security (TLS). Através desta extensão, é possível informar quais os protocolos ofertados diretamente no final do *handshake* de segurança. Por padrão, o servidor suporta conexões em diferentes protocolos como HTTP/1.1, SPDY/3.1 e versões de desenvolvimento do HTTP/2 como h2-16 e h2-14 além da versão final h2. Para restringir o anúncio de protocolos não relacionados aos experimentos, o parâmetro de configuração “`-nnp-list=h2`” foi utilizado ao iniciar o serviço.

O **cliente** utiliza o navegador Web Google Chrome na versão 48.0.2564.82 (64-bit) que foi escolhido por ofertar suporte nativo ao HTTP/2, além de facilidades para coleta de dados e manipulação interativa através de sua interface de depuração remota. A instalação deste não teve qualquer tipo de configuração adicional e não foram utilizados *plugins* ou modificações internas. A Tabela 1 resume as versões e configurações dos *softwares* do ambiente.

Entidade	Configuração do Software
Servidor Web	Lighttpd versão 1.4.37 TCP 80 <i>backend</i> FastCGI habilitado
Servidor Proxy	Nghttp2 versão 1.7.0 TCP 443 Servidor Web Lighttpd como <i>backend</i>
Cliente	Google Chrome versão 48.0.2564.82 (64-bit)

Tabela 1. Entidades de Software

As entidades em execução no servidor (servidor Web, servidor *proxy* e *scripts* PHP) foram compilados com base no sistema operacional Centos 7.0.1406 com o Kernel Linux 3.10.0123.el7.x86_64. A entidade no cliente, o Google Chrome, é executada no sistema operacional Fedora versão 22, utilizando o Kernel Linux 4.3.4200.fc22.x86_64.

⁵<https://curl.haxx.se/>

⁶<https://www.wireshark.org/>

4.2. Infraestrutura de Hardware

A estrutura de *hardware* do ambiente possui componentes capazes de executar os *softwares* de forma eficiente e sem gerar atrasos inesperados que poderiam influenciar na medição dos experimentos. A Tabela 2 resume as configurações de *hardware* utilizadas.

Componete	Processador	Memória	Rede
Servidor	Dual Xeon L5410	12 GigaBytes	Ethernet 100Mb/s
Cliente	Quad core Intel i7	16 Gigabytes	Ethernet 100Mb/s

Tabela 2. Componentes de Hardware

A interconexão entre os componentes é feita através da Internet, que não possui estrutura fixa de roteamento. Entretanto, durante a execução dos experimentos, foi possível observar um número constante de 13 saltos de rota e latência média entre 80 e 100 milissegundos.

4.3. Métrica, Fatores e Níveis

Este experimento definiu como métrica o **intervalo de tempo para conclusão do download de todos os objetos das páginas Web**, denominado **Total Download Time (TDT)**. Esta métrica considera apenas fatores relacionados a transações de rede, excluindo processos de renderização, validação e análise do conteúdo. A Tabela 3 apresenta todos os fatores e seus respectivos níveis utilizados nos experimentos para a análise da métrica definida.

Fator	Níveis
Tamanho	100KB a 2000KB com incremento de 100KB
Número de Objetos	10 a 200 com incremento de 10

Tabela 3. Fatores e Níveis

Alguns valores de níveis foram baseados em médias mineradas a partir do repositório Http Archive⁷, no intervalo de janeiro de 2013 a dezembro de 2015. O sistema de coleta utilizado pelo Http Archive utiliza arquivos HAR para obter seus dados que não possuem poder descritivo para identificar como os objetos estão apresentados no documento HyperText Markup Language (HTML). Dessa forma, eles serão considerados nos experimentos como sendo o pior cenário possível, onde todos os objetos do tipo *JavaScript* são externos e causam bloqueio no processamento do Document Object Model (DOM).

Para definir o número de objetos do tipo *JavaScript* nos experimentos, utilizou-se como base os dados da Figura 4, que representa a frequência de números de objetos *JavaScript* e, a partir dela, fixou-se o limite de 200 objetos para o experimento.

Já para definir o tamanho dos objetos do tipo *JavaScript*, o nível do fator foi baseado na Figura 5, que representa a frequência de tamanhos de objetos também minerados do repositório Http Archive, coletados de janeiro de 2013 a dezembro de 2015.

⁷<http://httparchive.org>

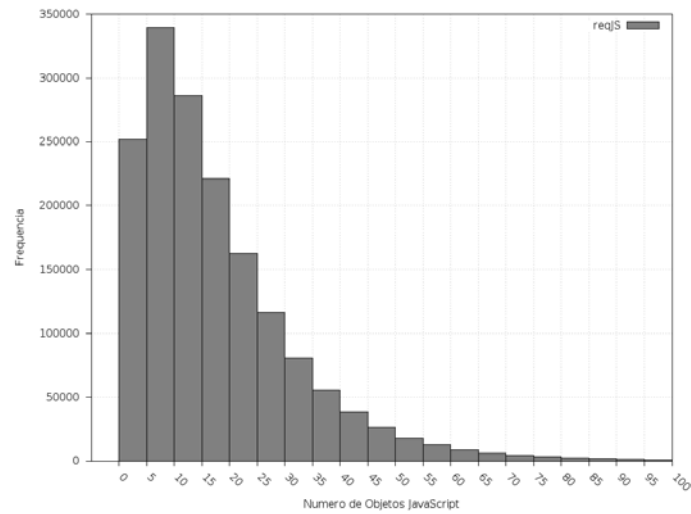


Figura 4. Frequência de número de objetos JavaScript

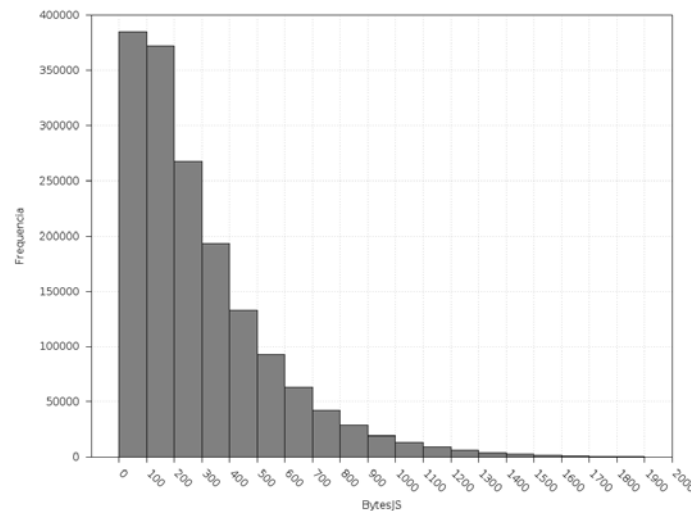


Figura 5. Frequência de tamanho de objetos JavaScript

Por não existir um modelo descritivo padrão para representar páginas Web, outros estudos como [Butkiewicz et al. 2011] e [Sivakumar et al. 2014] utilizam cópias de conteúdo de sites reais em um determinado período do tempo e os reutilizam durante os experimentos. Normalmente, os sites são selecionados com base em classificação de popularidade e tendem a representar o comportamento de uma população maior de sites.

5. Resultados

Os valores obtidos nos experimentos pela variação do tamanho de objetos são apresentados na Figura 6. Os grupos de experimentos apresentam incremento linear em função do tamanho dos objetos. O padrão visual dos intervalos de desvio padrão em relação as médias de cada grupo de experimentos indicou uma possível igualdade estatística. Para confirmação desta hipótese, o teste estatístico ANOVA foi executado⁸. Sendo o nível de significância considerado em 6%, 70% dos experimentos apresentam similaridade.

⁸Dados do teste não apresentados por falta de espaço

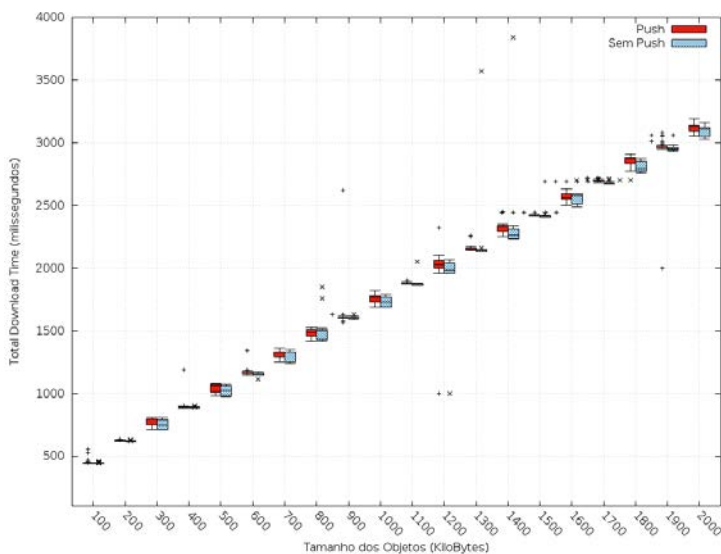


Figura 6. Comparativo do tamanho dos objetos para TDT

A Figura 7 apresenta os dados coletados nos experimentos com variação do fator número de objetos. Os valores apresentam grande dispersão, sendo este efeito mais frequente no uso do recurso *push* no intervalo de 17 a 110 objetos. Apesar desta dispersão, o teste estatístico ANOVA apresentou, para o nível de significância em 5%, que apenas 9 dos 20 experimentos são similares.

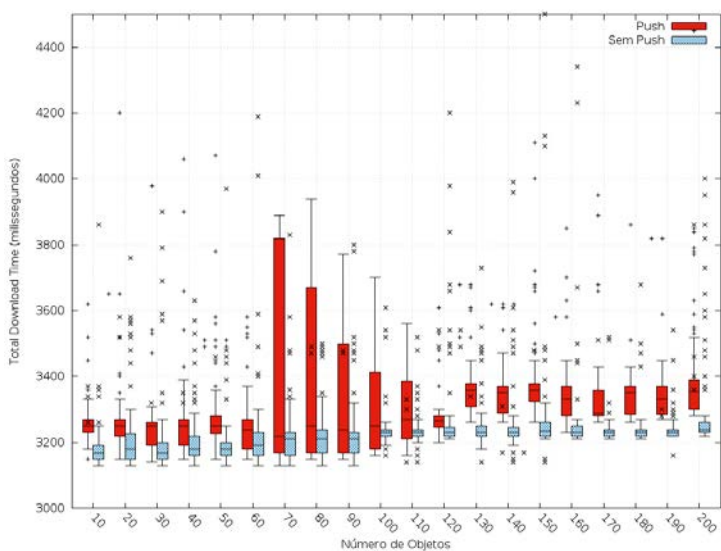


Figura 7. Comparativo do número de objetos em TDT

Os resultados obtidos pela métrica TDT indicam que **a quantidade de streams concorrentes, independentes do tipo, não afetam de forma crítica o tempo no processo de rede**. Apesar de não haver confirmação estatística de igualdade nos *streams* do tipo *push* e não *push*, ambos apresentam comportamento similar nos diferentes cenários. Esta característica é esperada devido o *design* do protocolo HTTP/2, uma vez que o mesmo utiliza multiplexação, e também indica que a compactação de cabeçalhos não

causa *overhead*.

Os níveis de dispersão apresentados com maior evidência durante o uso do recurso *push* nos experimentos com maiores números de objetos podem estar relacionados ao grande número de conexões efetuadas entre o servidor *proxy* e o servidor Web em um curto intervalo de tempo. Este evento ocorre pois para cada objeto a ser enviado através do recurso *push*, uma nova solicitação é gerada ao servidor Web, sendo necessário que o servidor *proxy* aguarde o recebimento do código de *status* antes de efetuar o envio dos respectivos *frames* do tipo PUSH_PROMISE.

6. Conclusões e Trabalhos Futuros

As limitações e resultados observados neste estudo revelam a possibilidade de continuação desta mesma análise considerando outros fatores e novas métricas, como o intervalo de tempo decorrido entre a solicitação das páginas experimentais e o evento de conclusão de carregamento do DOM, denominada PLT (*Page Load Time*).

Além disso, a variação de fatores da rede como latência e perdas devem ser analisadas, especialmente em ambientes reais, a exemplo de redes móveis. Também pode ser considerada a existência de múltiplos servidores como origem de conteúdos de uma mesma página, visto que a utilização de Content Delivery Networks (CDNs) deve continuar a ser fortemente utilizados.

Agradecimentos

Este trabalho foi apoiado pelo RLAM Innovation Center, Ericsson Telecomunicações S.A., Brasil

Referências

- Belshe, M., Peon, R., and Thomson, M. (2015). Hypertext transfer protocol version 2 (http/2). RFC 7540, RFC Editor.
- Butkiewicz, M., Madhyastha, H. V., and Sekar, V. (2011). Understanding website complexity: Measurements, metrics, and implications. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11*, pages 313–328, New York, NY, USA. ACM.
- Erman, J., Gopalakrishnan, V., Jana, R., and Ramakrishnan, K. K. (2013). Towards a spdy'ier mobile web? In *ACM Conference on Emerging Networking Experiments and Technologies, CoNEXT '13*, New York, NY, USA. ACM.
- Fielding, R., Gettys, J., Mogul, J., Frystyk, H., and Berners-Lee, T. (1997). Hypertext transfer protocol – http/1.1. RFC 2068, RFC Editor.
- Han, B., Hao, S., and Qian, F. (2015). Metapush: Cellular-friendly server push for http/2. In *5th Workshop on All Things Cellular: Operations, Applications and Challenges, AllThingsCellular '15*, New York, NY, USA. ACM.
- Hock-Chuan, C. (2009). In introduction to http basics. https://www.ntu.edu.sg/home/ehchua/programming/webprogramming/HTTP_Basics.html. [accessed: 2015-01-20].
- Kamp, P.-H. (2014). Http/2.0 - the ietf is phoning it in. *Queue*, 13(1):10:10–10:12.

- Nielsen, H. F., Spreitzer, M., Janssen, B., and Gettys, J. (1998). Http-ng overview - problem statement, requirements, and solution outline. Internet draft. <https://tools.ietf.org/html/draft-frystyk-httpng-overview-00>.
- Padhye, J. and Nielsen, H. F. (2012). A comparison of spdy and http performance. *Microsoft Res.*
- Peon, R. and Ruellan, H. (2015). Hpack: Header compression for http/2. RFC 7541, RFC Editor.
- Podjarny, G. (2012). Not as spdy as you thought.
- Saxce, H., Oprescu, I., and Chen, Y. (2015). Is http/2 really faster than http/1.1? In *IEEE Conference on Computer Communications Workshops*.
- Sivakumar, A., Puzhavakath Narayanan, S., Gopalakrishnan, V., Lee, S., Rao, S., and Sen, S. (2014). Parcel: Proxy assisted browsing in cellular networks for energy and latency reduction. In *Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies, CoNEXT '14*, pages 325–336, New York, NY, USA. ACM.
- Wang, X. S., Balasubramanian, A., Krishnamurthy, A., and Wetherall, D. (2014). How speedy is spdy? In *USENIX Conference on Networked Systems Design and Implementation, NSDI' 14*.

Implementação Inicial da RFC 6897 para Auxílio no Tratamento de Fluxos Elefantes

Alan C. Silva¹, Simone Ferlin², Fábio L. Verdi¹

¹LERIS – Universidade Federal de São Carlos (UFSCar)
Sorocaba – SP – Brasil

²Simula Research Laboratory, 1364 Fornebu, Norway

{alansilva,verdi}@ufscar.br, ferlin@simula.no

Abstract. *The Multipath TCP (MPTCP) protocol allows applications to better explore the network resources available to multi-connected devices such as mobile phones or multi-homed systems. Here, two main advantages are envisioned: bandwidth aggregation, and the ability to maintain the connection, if one of the network path fails. To extend these capabilities to the application, RFC 6897 defines an API to better control each of MPTCP's subflows, so that these can be added or removed as needed. This paper presents an API implementation as defined in RFC 6897 to give applications the capability to control MPTCP's subflows. To test the API and validate our implementation, we build experiments with elephant flows in datacenter networks.*

Resumo. *O protocolo Multipath TCP (MPTCP) permite que as aplicações possam explorar melhor os recursos de rede disponíveis para dispositivos multi-conectados como os telefones móveis ou sistemas multi-homed. Aqui, duas principais vantagens são previstas: agregação de banda e a habilidade de manter a conexão estabelecida se houver falha em um dos caminhos de rede. Para estender essas capacidades para a aplicação, a RFC 6897 define uma API para permitir um melhor controle de cada subfluxo MPTCP, de modo que esses possam ser adicionados ou removidos conforme necessário. Este artigo apresenta uma implementação da API conforme definida na RFC 6897 para dar as aplicações a capacidade de controlar os fluxos MPTCP. Para testar a API e validar nossa implementação, nós construímos experimentos usando fluxos elefantes em redes de datacenter.*

1. Introdução

O protocolo de controle de transmissão (*TCP*) é um protocolo-chave no paradigma atual da Internet, pois ele permite criar serviços onde os fluxos de bytes que passam são confiáveis e por isso acaba sendo utilizado por diversas aplicações que demandam confiabilidade na transmissão de dados nos mais variados tipos de dispositivos, desde smartphones até servidores em redes de datacenters.

Mesmo pertencendo a uma das partes fundamentais da Internet que conhecemos hoje, o protocolo *TCP* continua em constante processo de evolução. Com o surgimento de uma Internet e serviços mais complexos, uma das evoluções mais recentes que acompanhou esta mudança é o *Multipath TCP (MPTCP)*. [Raiciu et al. 2012, Raiciu et al. 2013].

O *MPTCP* altera uma das premissas básicas da especificação original do protocolo *TCP* que determina que uma conexão *TCP* será sempre identificada através da tupla de 4 elementos que consiste nos endereços *IP* de origem e destino e das portas de origem e destino. Todos os pacotes que são enviados por uma conexão *TCP* sempre serão formados por essa tupla de 4 elementos. O *TCP* como conhecemos não consegue explorar hoje todas os recursos oferecidos pelo hardware ou pelas tecnologias atuais. Por exemplo, um *host multi-homed* ou um *smartphone* que normalmente possuem a capacidade de trabalhar com múltiplas interfaces ou múltiplos caminhos. Nestes casos, uma conexão *TCP* não poderá se beneficiar das múltiplas interfaces e caminhos caso não exista um mecanismo que ofereça suporte para isso.

O *MPTCP* [Raiciu et al. 2013] resolve esta questão permitindo que os pacotes pertencentes a uma determinada conexão sejam transmitidos por diferentes interfaces e endereços através dos subfluxos criados para cada conexão, que, inicialmente se aproveitam da estrutura de uma conexão *TCP* original. Estes subfluxos potencialmente serão enviados individualmente pelas interfaces e caminhos disponíveis, tarefa que convém ao agendador (*scheduler*) e ao mecanismo de controle de congestionamento (*congestion control*) decidirem.

O *MPTCP* ganhou reconhecimento comercial e foi adaptado aos sistemas operacionais recentes da *Apple* [Hesmans et al. 2015b], permitindo a sua utilização em grande escala. Desde 2013, todos os *tablets* e *smartphones* da *Apple* utilizam o *MPTCP* na aplicação de reconhecimento de voz chamada *Siri*, explorando um dos aspectos do protocolo cujo objetivo é manter a conectividade quando há falha em um dos caminhos possíveis na rede. Em julho de 2015, a maior companhia de telefonia sul-coreana, a *Korean Telecom* em parceria com a *Samsung* anunciou a implantação em escala comercial do protocolo *MPTCP* em diversos *smartphones* equipados com *Android*, combinando as interfaces *LTE* com *wifi* disponíveis, aumentando significativamente o ganho na taxa de utilização de banda desses dispositivos [Oh and Lee 2015].

O desenvolvimento do protocolo *MPTCP* foi motivado por alguns cenários que necessitam enviar dados utilizando a mesma conexão através de interfaces e endereços diferentes, como por exemplo, *smartphones* equipados com interface celular e *wifi* [Paasch et al. 2012], ou redes de *datacenters* [Raiciu et al. 2011] cuja malha topológica oferece múltiplos caminhos para os diferentes subfluxos, que seguem possivelmente para um mesmo destino.

O *MPTCP* foi desenvolvido tendo três principais objetivos de compatibilidade em mente [Ford et al. 2011]:

1. O protocolo deve ser utilizado pelas aplicações aproveitando a *API* de *socket* já existente. Atualmente, o *Linux* [Paasch et al. 2013] possui uma implementação do protocolo *MPTCP* que resolve em parte essa situação.
2. O protocolo deve ser compatível com a rede onde está sendo implantado. Para alcançar este objetivo, o protocolo possui diversos mecanismos que possibilitam a sua comunicação com *middleboxes* existentes na rede [Raiciu et al. 2012, Raiciu et al. 2013].
3. O protocolo deve manter a equidade entre os usuários da rede. Para alcançar esse objetivo, vários esquemas de controle de congestionamento foram propostos e implementados.

Este artigo tem como objetivo rever o primeiro item da lista de compatibilidades que é requerido para o *MPTCP* operar de forma compatível na Internet, ou seja, implementar a *API* proposta pela *RFC 6897*. Esta *API* oferece para as aplicações a capacidade de controlar o protocolo *MPTCP* podendo, por exemplo, ligar ou desligar o suporte ao protocolo e ser capaz de adicionar ou remover um ou mais subfluxos. Hoje, a única forma de controlar o número de subfluxos que são fixos e criados no momento de abertura da conexão é via comando *sysctl*.

A criação da *API* permitirá que todas as aplicações que estão rodando no sistema possam ajustar de forma individual a quantidade de subfluxos necessários, servindo como motivação para o trabalho.

Assumindo que a aplicação necessita de um serviço de transmissão de bytes, precisamos de uma interface (*API*) que possa providenciar a adaptação de forma transparente para permitir o controle do protocolo *MPTCP* com a finalidade de se obter melhor utilização. Para validar nossa implementação, usamos a *API* para o tratamento de fluxos elefantes em redes de datacenters.

O artigo está organizado da seguinte forma: na Seção 2 serão apresentados os trabalhos relacionados. A descrição do protocolo *MPTCP* é apresentada na Seção 3. Uma visão mais detalhada do funcionamento da *RFC 6897* será apresentada na Seção 4. Os métodos utilizados para o desenvolvimento e implementação assim como o caso de uso serão apresentados na Seção 5.

2. Trabalhos Relacionados

Os gerenciadores de caminhos (*path managers*) possuem a função de decidir como os subfluxos serão estabelecidos e atualmente há dois disponíveis: *Fullmesh* e *ndiffports*. No caso do *ndiffports*, apenas o cliente é responsável pela criação de subfluxos. O servidor nunca cria subfluxos, pois o cliente pode estar atrás de um *firewall* ou *NAT* que bloqueia a tentativa de conexão [Eardley 2013]. Em *fullmesh* é possível a criação do subfluxo a partir do servidor, caso ele conheça o IP de conexão e trabalhar como passivo, esperando um *ADD* por parte do cliente, porém, o modo comum de operação é o ativo (onde o cliente cria o subfluxo).

O gerenciador de caminhos *fullmesh* escuta os eventos a partir das interfaces de rede subadjacentes e cria um subfluxo para o servidor para cada interface ativa. Estes subfluxos são criados imediatamente após a criação da conexão ou quando a interface é ativada a partir da criação do primeiro subfluxo, o subfluxo principal. Isto permite, por exemplo, que os *smartphones* possam reagir no caso de uma perda de conectividade [Paasch et al. 2012].

O gerenciador de caminhos *ndiffports* cria *N* subfluxos em uma mesma interface de forma imediata após o estabelecimento da conexão. Este gerenciador de caminhos foi desenvolvido tendo o foco em *datacenters* onde são permitidos a utilização de diferentes caminhos que possam ser balanceados com *Equal Cost Multipath (ECMP)* [Raiciu et al. 2011].

Diversos pesquisadores têm explorado como o protocolo *MPTCP* deve gerenciar os fluxos e as interfaces disponíveis. Paasch et Al. [Paasch et al. 2012] avaliam como os dispositivos *wireless* se adaptam a perdas de conectividade. Este artigo propõe três modos

de operação para o protocolo *MPTCP* em *smartphones*: *single path*, *backup* e *full-mptcp*. Bocassi et al. [Bocassi et al. 2013] propõem o gerenciador de caminhos chamado *Binder* que explora o roteamento de origem livre e utiliza o *MPTCP* para agregar caminhos diferentes em redes de malha sem fio. Lim et al. [Lim et al. 2014b] propõem uma extensão para o *MPTCP* que permite adaptar a utilização dos subfluxos baseado na informação extraída a partir da camada de *MAC*. Essa extensão foi validada experimentalmente, porém, a implementação não possui muitos detalhes.

Lim et al. [Lim et al. 2014a] também propõem o *eMPTCP* que retarda o estabelecimento dos subfluxos em *smartphones* que estão utilizando a interface *LTE*. Porém, quando o smartphone troca de interface, por exemplo, (*LTE - Wifi*), o artigo propõe a reinicialização da estimativa do *RTT* do subfluxo utilizado na interface *LTE* para forçar a utilização do mesmo na nova interface selecionada. Essa solução agiliza a utilização do subfluxo *LTE*, mas não é elegante para solucionar o problema do gerenciamento de subfluxos.

Schmidt et al. [Schmidt et al. 2013] propõem a utilização do conceito de *socket intents* que permite as aplicações informarem a conexão sobre o seu conhecimento da comunicação estabelecida, podendo gerar informação que serve como previsão para melhoria da conexão. Estes *intents* incluem informação sobre o tipo de transferência (*query*, *bulk*, *stream*) ou a informação sobre o fluxo (número de bytes, duração).

Hesmans et al. [Hesmans et al. 2015a] definem a criação de um gerenciador de caminhos que utiliza o conceito de *socket intents*, extraindo as informações dos *intents* e usando as mesmas como parâmetros, permitindo que aplicações possam manipular o protocolo *MPTCP*.

Porém, mesmo que a *RFC 6897* [Scharf and Ford 2013] proponha a criação de algumas extensões na interface (*API*) básica de *socket* para permitir que aplicações possam adicionar ou remover subfluxos em uma conexão *MPTCP*, atualmente nenhuma implementação do *MPTCP* faz uso desta extensão. Esta fato serve de motivação para a implementação que é mostrada nesse artigo, dado a necessidade existente demonstrada através do caso de uso escolhido.

Por fim, é importante definirmos o que é fluxo elefante. De acordo com a definição encontrada em [Greenberg et al. 2009], um fluxo é considerado elefante quando ele atinge 100MB. Muito embora existam outras definições na literatura [Curtis et al. 2011, Casado 2013], esta é a mais utilizada. Atualmente, existem algumas soluções que tratam o problema dos fluxos elefantes. Elas se baseiam na ideia de detecção e tratamento no centro da rede usando controladores que fazem o escalonamento através de múltiplos caminhos conforme *Al Fares et. al.* [Al-Fares et al. 2010].

3. Multipath TCP

O *MPTCP* [Raiciu et al. 2013] permite aos hosts trocarem pacotes que pertencem a uma determinada conexão entre várias interfaces ou caminhos na rede. Para que esta tarefa possa ser executada, cada conexão *MPTCP* é composta por várias conexões *TCP* que são denominadas subfluxos [Raiciu et al. 2013].

O *MPTCP* é uma extensão do protocolo *TCP* definida pelo *IETF* em [Ford et al. 2009] dando a possibilidade de uma conexão *TCP*, que por padrão é *single*

path, poder funcionar como uma conexão de múltiplos caminhos, suportando assim uma maior diversidade para os dados de uma mesma conexão.

Uma das principais vantagens do *MPTCP* em relação a outras soluções, como por exemplo, o *SCTP* [Paasch and Bonaventure 2014], é poder usar a mesma estrutura do protocolo *TCP* para trafegar informação, fazendo com que o *MPTCP* seja transparente para aplicações que utilizam o protocolo. Isso é possível pois o *MPTCP* se aproveita da mesma *API* de *socket* do protocolo *TCP* como mostra a Figura 1.

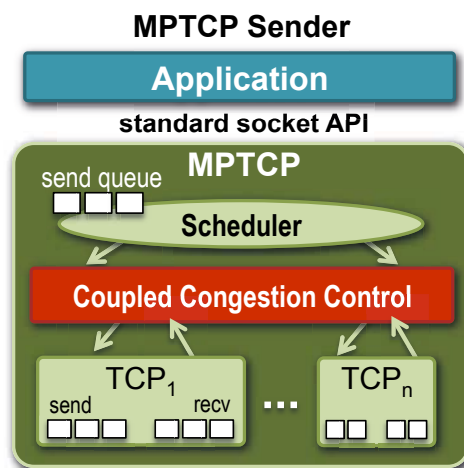


Figura 1. Pilha TCP com suporte ao MPTCP.

O *MPTCP* se aproveita do processo de *three-way handshake* do *TCP* para efetuar a negociação necessária que permite seu uso. Basicamente, ele possui 3 fases que são:

- Estabelecer uma nova conexão *MPTCP*;
- Adicionar subfluxos em uma conexão *MPTCP*;
- Transmitir dados através da conexão *MPTCP*.

Para estabelecer uma nova conexão *MPTCP*, o host de origem envia um pacote *SYN* com a *flag* de *MP_CAPABLE* sinalizada mais uma chave aleatória que será utilizada para o cálculo do *hash* que servirá de identificador para a conexão. Nesse caso, se o *host* de destino suportar o protocolo *MPTCP*, ele devolve a resposta com um *SYN+ACK* contendo o *MP_CAPABLE* sinalizado mais uma chave aleatória escolhida pelo servidor. O terceiro *ACK* do processo de *three-way handshake* também vai incluir a *flag* de *MP_CAPABLE* sinalizada e as chaves que serão utilizadas como identificador, estabelecendo assim a conexão, conforme a Figura 2.

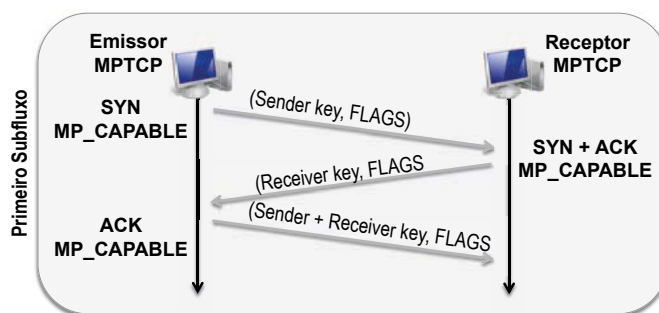


Figura 2. Criação de uma conexão MPTCP (MP_CAPABLE).

Caso seja necessária a criação de um novo subfluxo em uma conexão *MPTCP* existente, é efetuado um novo processo de *three-way handshake*, onde o *host* de origem envia o primeiro *SYN* com o campo *MP_JOIN* que possui o *token* da conexão *MPTCP* existente. Este *token* é o resultado da chave que foi gerada no estabelecimento da conexão.

O *host* de destino recebe o pacote e calcula o *HMAC* a partir do *token*, devolvendo um *SYN+ACK* com o resultado do *HMAC*. O *host* de origem recebe e valida o *HMAC* devolvendo um *ACK* com o *HMAC* e assim estabelecendo a validação do subfluxo, conforme a Figura 3.

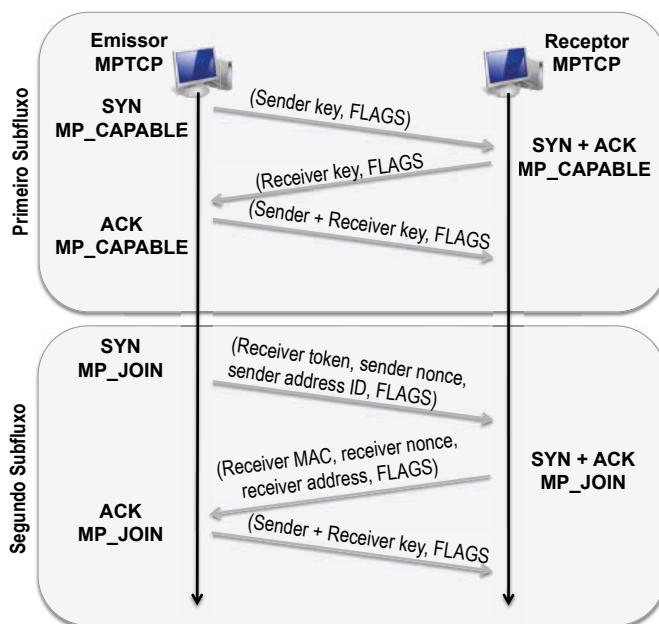


Figura 3. Criação de um novo subfluxo em uma conexão existente (MP_JOIN).

A proposta desse projeto é aproveitar o recurso de subfluxos do *MPTCP* justamente para quebrar os fluxos grandes em subfluxos menores, permitindo assim, o espalhamento desses subfluxos através dos vários caminhos oferecidos e explorados pelo protocolo *MPTCP* por padrão.

4. RFC 6897

Conforme citado anteriormente, o protocolo *MPTCP* adiciona a capacidade de se utilizar caminhos múltiplos em uma conexão *TCP*. As motivações para isso incluem o aumento da taxa de transferência e a resiliência no caso de uma possível falha de rede.

O texto encontrado na *RFC 6897* descreve algumas questões sobre a compatibilidade do protocolo *MPTCP* com aplicações que não suportam o mesmo, e sugere uma interface básica de aplicação que estende a interface *TCP* já existente para permitir esse suporte. A idéia é apresentar quais são os efeitos que o protocolo *MPTCP* exerce sobre as aplicações, como por exemplo, a questão de desempenho comparando-o ao *TCP* e a interoperabilidade entre *MPTCP* e as aplicações.

Assumindo que a aplicação não suporta transporte de dados por caminhos múltiplos sem que haja a necessidade de alteração nas mesmas, se faz necessário especificar uma *API* básica para permitir que estas aplicações sejam *MPTCP-aware*, ou seja, suportem o protocolo.

A motivação para a criação de uma *API* que suporte o protocolo *MPTCP* consiste no fato de permitir que aplicações possam utilizar uma interface já conhecida. Por isso, a implementação é efetuada através da extensão da interface de *sockets* que já é utilizada como padrão na implementação de aplicações que precisam interagir entre si utilizando o protocolo *TCP*. Isso permite a mesma facilidade de implementação, pois quem já utiliza a interface de *sockets TCP* poderá utilizar a *interface MPTCP* da mesma forma.

A *API padrão* já consolidada para aplicações *TCP* é a interface via *socket*. A partir disso, o documento propõe uma maneira abstrata de estender esta interface de forma que possa suportar o *MPTCP* através de operações que obtenham (*get*) e definam (*set*) valores específicos do *MPTCP* através de uma opção de *socket socket option* gerada no nível do protocolo *TCP*. Isso permite que aplicações, linguagens de programação e bibliotecas possam decidir como utilizar essas informações de forma transparente, ou seja, sem nenhuma alteração de código.

Uma *API MPTCP* básica consiste em um conjunto de novos valores que serão associados a um *socket MPTCP*. Estes valores podem ser utilizados para alterar propriedades em uma conexão *MPTCP* ou recuperar uma informação.

Os valores podem ser acessados através das chamadas de sistema já utilizadas na interface de *socket* como *setsockopt()* ou *getsockopt()*, através de símbolos específicos criados especialmente para alterar ou recuperar dados correspondentes ao protocolo *MPTCP* que serão acessados por parâmetros passados para essas chamadas.

As opções que são especificadas na Seção 5.3.1 da *RFC 6897* correspondem a uma versão básica da *API* e devem ser implementadas para uma utilização aceitável por outras aplicações:

- *TCP_MULTIPATH_ENABLE*: Habilita ou desabilita o *MPTCP*;
- *TCP_MULTIPATH_ADD*: Conecta o protocolo *MPTCP* em um conjunto de endereços locais definidos ou adiciona um conjunto de novos endereços locais em uma conexão *MPTCP* existente. Na prática esta opção permite que novos subfluxos sejam adicionados à uma conexão *MPTCP*;
- *TCP_MULTIPATH_REMOVE*: Remove um endereço local de uma conexão

MPTCP. Na prática, esta opção permite que subfluxos existentes sejam removidos de uma conexão *MPTCP*;

- *TCP_MULTIPATH_SUBFLOWS*: Recupera os pares de endereços atualmente utilizados por subfluxos *MPTCP*;
- *TCP_MULTIPATH_CONNID*: Retorna o identificador de conexão local da conexão *MPTCP* atual.

A Tabela 1 mostra o conjunto de funções que serão implementadas na *API* de *socket* para permitir a manipulação do *MPTCP* através das aplicações.

Tabela 1. Operações da API MPTCP

Nome	Get	Set	Tipo de Dados
TCP_MULTIPATH_ENABLE	x	x	Booleano
TCP_MULTIPATH_ADD		x	Lista de Endereços/Portas
TCP_MULTIPATH_REMOVE		x	Lista de Endereços/Portas
TCP_MULTIPATH_SUBFLOWS	x		Lista de Pares de Endereços/Portas
TCP_MULTIPATH_CONNID	x		Inteiro

5. Implementação e Avaliação

A *RFC 6897* define como estender a *API* de *sockets* para agregar funções que consigam interoperabilizar com o protocolo *MPTCP*. Lembrando que a *API* de *socket* é implementada diretamente em nível de *kernel* e, por isso, é necessário que as funções da *API* também sejam implementadas no mesmo nível.

Para isso, adicionamos as chamadas de interface para o protocolo *MPTCP* dentro da interface de *sockets* que são representadas pelas funções *do_tcp_setsockopt*, no caso das funções de alteração de dados, e *do_tcp_getsockopt*, no caso das funções de recuperação de dados. Tais funções são definidas e implementadas no arquivo *tcp.c* localizado em *net/ipv4* dentro da árvore de código do *kernel*.

Abaixo temos um trecho de código, esqueleto em C, da implementação da função que adiciona subfluxos, ilustrando o início do desenvolvimento em modo de *kernel*.

```
static int do_tcp_setsockopt(struct sock *sk, int level,
    int optname, char __user *optval, unsigned int optlen)
{
    ...
    switch (optname) {
        case TCP_MULTIPATH_ADD: {
            //Implementação da função que adiciona subfluxos.
        }
    }
    ...
}
```

A característica padrão de operação do protocolo *MPTCP* é a de estabelecer um número fixo de subfluxos que são criados no momento de abertura da conexão. Atualmente, a única forma de alterar essa operação é por meio de variável de controle do sistema referente ao protocolo e que pode ser definida através do comando *sysctl*. Entretanto, esta alteração passa a valer para

todas as aplicações que estão rodando no sistema, não permitindo, portanto, que cada aplicação, individualmente, ajuste a quantidade de subfluxos necessários.

Outro ponto importante na operação do *MPTCP*, e também relevante para a API que está sendo implementada, é a questão dos *path managers*; já que são eles que definem como o protocolo deve se comportar em relação à criação de subfluxos. A implementação da API feita neste trabalho utiliza o *path manager ndiffports* uma vez que seu modo de operação permite abrir mais de um subfluxo por par de endereços.

As duas operações que devem ser inicialmente implementadas para permitir que uma aplicação execute as funções básicas descritas no protocolo são as de adicionar (*TCP_MULTIPATH_ADD*) e remover subfluxos *TCP_MULTIPATH_REMOVE*.

Estas operações permitem adicionar e remover um subfluxo por vez. Entretanto, neste trabalho, fizemos uma pequena extensão para que uma aplicação possa criar *N* subfluxos em uma única requisição. Com isso, adicionamos à opção *TCP_MULTIPATH_ADD* mais um campo na criação de subfluxos, podendo definir o número de subfluxos a serem criados. Esse campo foi denominado *num_subflows*, o que permite criar a chamada na interface conforme exemplo:

```
setsockopt(fd, SOL_TCP, TCP_MULTIPATH_ADD, num_subflows);
```

Para um teste inicial destas operações, utilizamos 2 máquinas virtuais *VMs* com *VMWare*, ambas rodando em um *Ubuntu Linux Server 14.04* com kernel versão 3.14.33 e, finalmente, com suporte ao protocolo *MPTCP* em sua versão de implementação 0.89. Uma máquina foi utilizada como servidor e a outra como cliente e através de duas aplicações escritas em *Python* simulamos o cliente e o servidor 4. Para fins de testes, criamos 4 subfluxos em uma mesma conexão *MPTCP*.

Atualmente, temos uma interface simples desenvolvida na linguagem *Python* para efetuar a chamada da API em *user mode* e um conjunto de funções desenvolvidas em linguagem C estendendo a API de *Sockets* em *Kernel Mode* no *Linux*. As funções disponíveis são: *TCP_MULTIPATH_ENABLE*, *TCP_MULTIPATH_CONNID* e *TCP_MULTIPATH_ADD*.

5.1. Caso de uso da RFC 6897: tratamento de fluxos elefantes

Como caso de uso, desenvolvemos um módulo que, junto ao *end-host*, tem a capacidade de verificar se um determinado fluxo de saída contém características de um fluxo elefante. Quando um fluxo elefante é detectado, a aplicação solicita a criação de subfluxos conforme a necessidade. A Figura 4 ilustra uma visão completa e genérica da solução. A seguir, temos uma descrição da arquitetura do módulo e a definição do que é um fluxo elefante para o mesmo.



Figura 4. Visão da Arquitetura do Módulo.

A arquitetura do módulo basicamente consiste nas seguintes fases:

- *Fase de detecção do fluxo elefante*: Nessa fase, todo o tráfego é monitorado através de um serviço que será executado no *end-host*, permitindo assim a detecção de um possível fluxo elefante;

- *Tratamento do fluxo elefante*: Após o fluxo elefante ser detectado, o módulo notifica o *kernel MPTCP* transformando esse fluxo elefante em vários fluxos camundongos;
- *Encaminhamento*: Conforme o número de subfluxos que foram estabelecidos, os fluxos camundongos que foram gerados são encaminhados. Eles chegam até o seu destino através do espalhamento que é feito utilizando os subfluxos criados.

A arquitetura se baseia em um algoritmo que define todo processo de atuação no *end-host*, passando pela detecção, tratamento e o encaminhamento final. A validação do algoritmo é determinada através de valores considerados mínimo e máximo para atuação do mesmo. Esses valores são utilizados como parâmetros limites de atuação para a detecção de um fluxo elefante, que hoje é definido a partir de 100MB até o limite de 300MB, abrindo um novo subfluxo a cada 50MB até atingir o limite máximo de 4 subfluxos, baseando se na definição de fluxo elefante encontrada em [Curtis et al. 2011].

O algoritmo consiste em um método contínuo que fica monitorando todos os fluxos *TCP* até que o valores estabelecidos como limite sejam atingidos. Quando o valor definido para o limite é alcançado, o módulo de controle de conexões *MPTCP* é ativado, recebendo a conexão sinalizada e a transformando em diversos fluxos camundongos. Esses fluxos camundongos são distribuídos através dos subfluxos criados e múltiplos caminhos existentes na rede conforme observado na Figura 5.

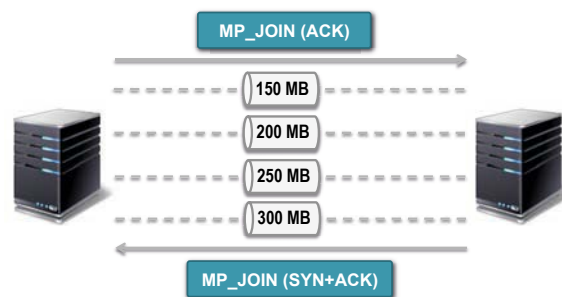


Figura 5. Visão de Criação e Controle de Subfluxos pelo Módulo.

O módulo trabalha em modo de serviço (*daemon UNIX*) que consiste na comunicação direta colocando a interface de rede em modo promíscuo, permitindo com isso que o repasse de fluxos e gerenciamento de chamadas ao protocolo *MPTCP* seja estabelecido.

As fases de tratamento e encaminhamento dependem de um *kernel* do *linux* que tenha suporte ao *MPTCP*. Importante lembrar que para o funcionamento do módulo se faz necessário o controle de conexões e criação de subfluxos através de rotinas que serão controladas pela aplicação, justificando a necessidade de implementação da API.

6. Conclusão

O *Multipath TCP (MPTCP)* foi desenvolvido para que fluxos *TCP* tradicionais usem com maior eficiência as diferentes interfaces de rede dos dispositivos, assim como os diferentes caminhos disponíveis em uma topologia de rede.

Entretanto, todo potencial fornecido pelo *MPTCP* só pode ser explorado se as aplicações fizerem uso do protocolo de maneira consciente (*MPTCP-awareness*), ideia esta especificada na *RFC 6897* que propõe uma *API* para que um controle fino dos subfluxos *MPTCP* seja realizado pelas aplicações que desejem usar tal protocolo.

Este trabalho, portanto, apresenta uma implementação desta *RFC* em nível de *kernel* para que aplicações façam uso do protocolo *MPTCP* conforme suas necessidades, definindo quando adicionar e remover fluxos assim como quando ligar e desligar o protocolo. Com esta implementação, os desenvolvedores e usuários do protocolo *MPTCP* podem encontrar caminhos para estender as suas aplicações de forma a utilizar os recursos existentes no protocolo com a finalidade de se obter os melhores resultados em sua utilização.

Atualmente, com os testes que foram realizados, conseguimos criar até 4 subfluxos simultâneos por conexão *MPTCP*. Dando continuidade ao trabalho, os próximos passos são: desenvolvimento das funções de *TCP_MULTIPATH_REMOVE* e *TCP_MULTIPATH_SUBFLOWS*, realização de avaliações experimentais da solução e discussão com o WG do *MPTCP*. Por fim, pretendemos disponibilizar a implementação para a comunidade.

Referências

- Al-Fares, M., Radhakrishnan, S., Raghavan, B., Huang, N., and Vahdat, A. (2010). Hedera: Dynamic flow scheduling for data center networks. In *NSDI*, volume 10, pages 19–33.
- Bocconi, L., Fayed, M. M., and Marina, M. K. (2013). Binder: a system to aggregate multiple internet gateways in community networks. In *Proceedings of the 2013 ACM MobiCom workshop on Lowest cost denominator networking for universal access*, pages 3–8. ACM.
- Casado, M. (2013). Of mice and elephants.
- Curtis, A. R., Kim, W., and Yalagandula, P. (2011). Mahout: Low-overhead datacenter traffic management using end-host-based elephant detection. In *INFOCOM, 2011 Proceedings IEEE*, pages 1629–1637. IEEE.
- Eardley, P. (2013). Survey of mptcp implementations.
- Ford, A., Raiciu, C., Handley, M., Barre, S., Iyengar, J., et al. (2011). Architectural guidelines for multipath tcp development. *IETF, Informational RFC*, 6182:2070–1721.
- Ford, A., Raiciu, C., Handley, M., Bonaventure, O., et al. (2009). Tcp extensions for multipath operation with multiple addresses. *IETF MPTCP proposal-http://tools.ietf.org/id/draft-ford-mptcp-multiaddressed-03.txt*.
- Greenberg, A., Hamilton, J. R., Jain, N., Kandula, S., Kim, C., Lahiri, P., Maltz, D. A., Patel, P., and Sengupta, S. (2009). V12: a scalable and flexible data center network. In *ACM SIGCOMM Computer Communication Review*, volume 39, pages 51–62. ACM.
- Hesmans, B., Detal, G., Bauduin, R., Bonaventure, O., et al. (2015a). Smapp: Towards smart multipath tcp-enabled applications. In *CoNEXT'15*.
- Hesmans, B., Tran-Viet, H., Sadre, R., and Bonaventure, O. (2015b). A first look at real multipath tcp traffic. In *Traffic Monitoring and Analysis*, pages 233–246. Springer.
- Lim, Y.-s., Chen, Y.-C., Nahum, E. M., Towsley, D., and Gibbens, R. J. (2014a). How green is multipath tcp for mobile devices? In *Proceedings of the 4th workshop on All things cellular: operations, applications, & challenges*, pages 3–8. ACM.
- Lim, Y.-s., Chen, Y.-C., Nahum, E. M., Towsley, D., and Lee, K.-W. (2014b). Cross-layer path management in multi-path transport protocol for mobile devices. In *INFOCOM, 2014 Proceedings IEEE*, pages 1815–1823. IEEE.

- Oh, B.-H. and Lee, J. (2015). Constraint-based proactive scheduling for mptcp in wireless networks. *Computer Networks*, 91:548–563.
- Paasch, C., Barre, S., et al. (2013). Multipath tcp in the linux kernel.
- Paasch, C. and Bonaventure, O. (2014). Multipath tcp. *Communications of the ACM*, 57(4):51–57.
- Paasch, C., Detal, G., Duchene, F., Raiciu, C., and Bonaventure, O. (2012). Exploring mobile/wifi handover with multipath tcp. In *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, pages 31–36. ACM.
- Raiciu, C., Barre, S., Pluntke, C., Greenhalgh, A., Wischik, D., and Handley, M. (2011). Improving datacenter performance and robustness with multipath tcp. *ACM SIGCOMM Computer Communication Review*, 41(4):266–277.
- Raiciu, C., Handley, M., and Bonaventure, O. (2013). Tcp extensions for multipath operation with multiple addresses. Technical report, RFC 6824, Jan.
- Raiciu, C., Paasch, C., Barre, S., Ford, A., Honda, M., Duchene, F., Bonaventure, O., and Handley, M. (2012). How hard can it be? designing and implementing a deployable multipath tcp. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, pages 29–29. USENIX Association.
- Scharf, M. and Ford, A. (2013). Multipath tcp (mptcp) application interface considerations. Technical report, RFC 6897, March.
- Schmidt, P. S., Enghardt, T., Khalili, R., and Feldmann, A. (2013). Socket intents: Leveraging application awareness for multi-access connectivity. In *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pages 295–300. ACM.

Otimizando a Transmissão TCP para Veículos Aéreos Não Tripulados

Rafael de Amorim Silva¹, Leandro de Melo Sales¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Caixa Postal 57.072-900 – Maceió – AL – Brasil

{rafael, leandro}@ic.ufal.br

Abstract. *This paper proposes a disruption-tolerant communication system for unmanned aerial vehicles that utilizes information from lower layers to control the transmission of TCP in disruptive channels. We adopt the suite ATN/IPS as a baseline, developing entities in the mobile router and in the controller host to perform this control. We discuss some technical details that must be considered to develop a draft for the IETF.*

Resumo. *Este artigo propõe um sistema de comunicação tolerante a interrupções para veículos aéreos não tripulados que utilizam informações provenientes das camadas inferiores da arquitetura TCP/IP para controlar a transmissão deste TCP em canais intermitentes. Utiliza-se a suíte de protocolos do padrão ATN/IPS como base, desenvolvendo e colocando entidades no roteador móvel e no host controlador para desempenhar tal controle. Também são discutidos detalhes técnicos que devem ser considerados para se desenvolver um esboço de padrão para o IETF.*

1. Introdução

Nos últimos anos, há um crescente interesse da indústria e dos centros de pesquisa no desenvolvimento e na manutenção dos veículos aéreos não tripulados (*Unmanned Aerial Vehicles* ou UAV). Basicamente, um UAV consiste em uma aeronave sem pilotos, sendo controlada a distância por controladores ou de forma autônoma. Tais veículos podem ser utilizados para diversos fins, como por exemplo, em operações militares, na transmissão de eventos esportivos, no sensoriamento de florestas, na prática agro-pecuária, entre outros. Em muitas dessas aplicações, há uma demanda crescente pela oferta de transmissão multimídia de qualidade.

Segundo [Chao et. al, 2010], os principais protocolos utilizados para suprir tal demanda são derivados da arquitetura de Internet, sendo o *Transmission Control Protocol* ou TCP (oferecendo uma entrega confiável dos dados) e protocolos construídos na camada de aplicação utilizando o UDP (transmissão sem entrega confiável) como base. Os mecanismos de retransmissão permitem que um protocolo de transporte tolere interrupções durante uma comunicação, evitando um rápido encerramento de sua sessão. O protocolo TCP, por exemplo, implementa um mecanismo de retransmissão que tolera um certo período de interrupção. Este mecanismo implementa um temporizador para adiar a entrega, tolerando atrasos ou interrupções em uma comunicação entre UAV e controlador.

Entretanto, o uso deste mecanismo em cenários com grandes interrupções pode reduzir o desempenho de transmissão desta rede, afetando a vazão do sistema, aumentando as retransmissões desnecessárias e o consumo de energia além de deixar o sistema mais ocioso. Por exemplo, um UAV está programado para transmitir um dado evento esportivo e deve sobrevoar uma determinada área para realizar o serviço de filmagem. Entretanto, o deslocamento da aeronave aliado a cobertura insuficiente da rede compromete a qualidade na transmissão do sinal, causando uma transmissão cortada ou com períodos de interrupção. Neste caso, é importante que a UAV tenha um protocolo de transporte sensível a perda de conectividade e que saiba adequadamente lidar com os efeitos negativos da interrupção na transmissão do TCP ou de outro protocolo de transporte.

Alguns trabalhos reduzem os efeitos indesejáveis da perda momentânea de conectividade de um enlace utilizando uma técnica que congela temporariamente a janela de congestão do TCP, impedindo que pacotes sejam retransmitidos desnecessariamente. Dois trabalhos se destacam na aplicação destas técnicas: (i) Freeze-TCP [Goff et.al, 2000] e (ii) DTP [Go et. al, 2012]. A primeira técnica é utilizada especialmente no processo de *handoffs*, isto é, quando um host móvel sai da sua estação atual e entra em outra cobertura de uma nova estação. Durante esta transição, o TCP zera a sua janela de congestão e aguarda o processo de *handoff* ser concluído, para retomar a sua transmissão. A segunda técnica implementa um mecanismo próprio de mobilidade em sua camada de transporte para mascarar as falhas na rede preterida. Neste caso, o *host* móvel efetua o congelamento da transmissão e o nó correspondente executa o mecanismo de retransmissão do TCP. Embora efetivos em alguns cenários, o ambiente aeronáutico é altamente dinâmico e exige novas adaptações na arquitetura TCP/IP para suportar apropriadamente transmissões de qualidade em cenários com conectividade intermitente.

Portanto, este artigo propõe uma abordagem para mitigar tais efeitos durante o voo de uma UAV dentro de uma determinada estação aéreo-terrestre (*Air-to-Ground* ou ATG), descrevendo uma abordagem que controla transmissões intra-celulares em cenários com longas interrupções na comunicação entre estações ATG e aeronaves. Especificamente, a abordagem insere 2 entidades na rede para realizar o controle da transmissão e a notificação sobre os estados de conectividade entre aeronave e estação terrestre. O TCP, sendo controlado por tais entidades, congela sua janela de congestão durante uma interrupção de enlace em uma célula ATG, permitindo que não haja retransmissões desnecessárias. A abordagem consegue obter um desempenho superior em relação ao uso de um tradicional TCP em cenários de transmissão aeronáutica.

O resto deste trabalho é apresentado da seguinte maneira: Seção 2 descreve os trabalhos relacionados com a proposta deste artigo. Seção 3 descreve a proposta deste trabalho, detalhando o mecanismo de retransmissão do TCP e as mudanças efetuada na arquitetura para permitir uma tolerância mais acurada nas interrupções aeronáuticas. Seção 4 apresenta as propostas para padronização desta arquitetura para o IETF. A Seção 5 apresenta as considerações finais deste trabalho.

2. Fundamentação

A técnica de congelamento da janela de congestão do TCP tem sido uma técnica efetiva para diversas aplicações na literatura, embora o seu uso exija cuidados importantes no projeto e na implantação de uma arquitetura de rede apropriada para suportá-la. Neste caso, é importante conhecer a natureza das interrupções aeronáuticas para eficazmente projetar uma arquitetura capaz de lidar com estas. Esta seção relata alguns problemas associados à estas interrupções e descreve alguns trabalhos relacionados que tentam resolver parte destes problemas.

2.1. Relato do Problema

Redes intermitentes consistem em interrupções no canal aeronáutico devido a perdas momentâneas de conectividade do enlace aéreo-terrestre. As interrupções podem ocorrer devido aos seguintes fatores [birrane et. al, 2011]:

- mobilidade das aeronaves: uma das causas mais prováveis de interrupção deve-se as mudanças frequentes do endereço IP de uma aeronave e a inexistência de antenas e filtros apropriados para tolerar altas taxas de erro de transmissão por causa da alta velocidade destas aeronaves;
- posicionamento inadequado das antenas: antenas devem ser direcionadas para cima para cobrir o espaço aéreo e melhorar o desempenho do canal de comunicação. Um posicionamento inadequado destas antenas pode criar buracos na cobertura da rede, permitindo o surgimento das interrupções;
- interferências: dentro e entre estações onde há cobertura e, portanto, transmissão. Podem derrubar a conectividade de um enlace e criar interrupções;
- restrições de energia: Falhas na energia e problemas com cobertura da rede devido à qualidade do sinal reduz a capacidade de transmissão das redes aéreo-terrestres;
- distribuição irregular das estações.

O mecanismo de controle do congestionamento do TCP é sensível a perda de pacotes na rede devido à interrupções na comunicação, reduzindo sua capacidade de transmissão precipitadamente quando houver uma longa exposição de uma interrupção aeronáutica. Este mecanismo controla o fluxo de pacotes na rede, evitando que a rede se torne congestionada. A idéia é introduzir um pacote na rede apenas quando os pacotes antigos já terem sido entregues. O TCP implementa uma janela de congestão e dinamicamente altera o seu tamanho. Isto ocorre porque o TCP supõe que a principal razão de um estouro de temporizador é devido à problemas de congestionamento da rede. Para controlar o fluxo, o TCP (versão *NewReno*) implementa 4 algoritmos que modificam o valor de sua janela de congestão ao longo de sua sessão, sendo eles [Floyd et. al, 2004]:

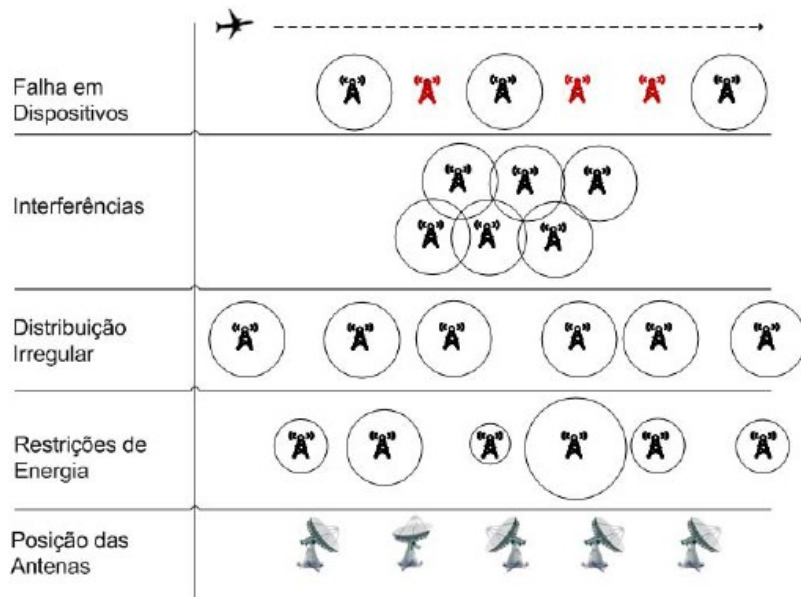


Figura 1. Tipos de interrupção aeronáutica.

- *Slow Start* (SS): este algoritmo é responsável por aumentar exponencialmente a janela de congestão quando um ACK é recebido até alcançar um determinado patamar pré-estabelecido. Isto permite controlar o fluxo de dados que está sendo inserido na rede, evitando que a rede se sobrecarregue com fluxos elevados.
- *Congestion Avoidance* (CA): este algoritmo aumenta linearmente a janela do transmissor quando um ACK é recebido até alcançar o patamar ou limite de largura de banda definido pelo receptor (*Advertised Window* ou *AWND*). Quando o congestionamento é detectado e pacotes são perdidos na rede, o estado deste controle retorna para o estado SS.
- *Fast Retransmit* (FRET): este algoritmo aprimora o controle de congestionamento por considerar ACKs duplicados como uma forma de alertar que um certo segmento está fora de ordem, replicado ou foi descartado pela rede. Depois que 3 ACKs forem duplicados, o algoritmo reenvia os pacotes antes que um timeout ocorra. Isto ajuda a reduzir o impacto dos longos timeouts em uma sessão TCP.
- *Fast Recovery* (FREC): algoritmo que governa a transmissão dos dados até que um ACK não duplicado seja recebido. A ideia é não forçar a mudança de estado para SS, pois o recebimento de ACKs duplicados indicam que a rede não está confiável, mas que ainda chegam pacotes ao destino.

2.1.1. Consequências para a transmissão TCP

A ocorrência de interrupções em ambientes aeronáuticos influencia protocolos como o TCP, limitando sua capacidade de transmissão, causando um maior consumo de energia para as aeronaves e causando o surgimento de retransmissões desnecessárias. Isto

ocorre por causa dos estouros do temporizador de retransmissão e os atrasos na transmissão aeronáutica devido a estes estouros. Em primeiro lugar, os estouros de temporização ocorrem quando o mecanismo de retransmissão não recebe nenhuma informação do receptor, estourando o limite de tempo de espera. Neste caso, há um ajuste na janela de congestão (cai pela metade) e uma nova retransmissão é realizada. Quanto maior for tal interrupção, maior será o número de retransmissões ocorridas durante esta fase, gerando perdas na capacidade de transmissão do TCP, baixa eficiência energética e acréscimos no tempo de transferência de dados ao longo desta rede. Por outro lado, os atrasos de reinício ocorrem toda vez que o algoritmo *backoff* é executado, pois a transmissão apenas ocorre quando houver um estouro de temporização, levando a rede esperar por tal *timeout*. Caso a conectividade com o enlace aéreo-terrestre seja retomada, o mecanismo irá aguardar até o estouro acontecer, permitindo a incidência deste tipo de atraso. Em outras palavras, a utilização de probes periódicos implicam em longos tempos de espera para que o TCP possa realizar a transmissão dos dados. Tais problemas afetam os protocolos da camada de aplicação como o FTP, HTTP, SMTP e aplicações não interativas como transferência de arquivos, e-mails eletrônicos e aplicações Web, sendo essencial a utilização de uma arquitetura de rede que permita o retorno imediato da transmissão assim que o enlace ATG for reestabelecido, aumentando, portanto, a eficiência e a capacidade de transmissão da rede.

2.2. Trabalhos Relacionados

Esta seção descreve os trabalhos correlatos a técnica de congelamento da janela de congestão do TCP. O uso desta técnica para transmissões de UAVs em células ATG permite que retransmissões desnecessárias sejam mitigadas, o consumo de energia seja reduzido e o retorno de conectividade seja instantâneo. Dois trabalhos se destacam na implementação desta técnica: (i) Freeze-TCP [Goff et.al, 2000]; e (ii) DTP [Go et. al., 2012].

O Freeze-TCP proporciona o congelamento da janela de congestão quando handoffs são efetuados em uma rede de comunicação. Quando um *host* móvel sai de sua área de cobertura, um *handoff* na saída da estação é realizado com o transmissor, neste caso, zerando a sua janela de congestão e somente a descongelando após um novo handoff de entrada em uma nova estação ser feito. Isto reduz perdas de pacotes durante a fase de *handoff*.

O DTP é um protocolo de transporte desenvolvido para mascarar as falhas na rede preterida. O DTP possui um esquema de tolerância à interrupção através da implementação de um mecanismo próprio de mobilidade em seu núcleo de transporte. Assim como o Freeze-TCP, este protocolo também utiliza a técnica de congelamento da janela de congestão do TCP. No momento de interrupção da rede, o *host* móvel efetua o congelamento de sua transmissão e o nó correspondente executa o algoritmo tradicional de retransmissão, até que o parâmetro máximo de retransmissão seja alcançado e a conexão encerrada. A percepção de retorno de conectividade é feita por probes periódicos ou por mecanismos de captura como o *netlink socket* ou o *broadcast receiver*. Durante o *handoff* em uma estação, o mecanismo reconhece o novo IP adquirido pelo *host* móvel utilizando um processo de autenticação entre os *hosts*. A

sessão é encerrada quando o *timeout* do *keep alive* estourar.

2.3. Discussão

Os trabalhos descritos anteriormente apresentam os seguintes problemas: (i) ambas as técnicas não consideram o ambiente em que um UAV está inserido, já que frequentes interrupções ocorridas dentro de uma célula de transmissão não necessariamente exige um processo de *handoff* cada vez que a conectividade for restabelecida. Embora permita transparência para as aplicações, para que esta técnica funcione, será necessário mudar a camada de transporte dos nós finais; (ii) A sessão TCP não pode ser descontinuada por causa destas interrupções, comprometendo seriamente a transmissão de grande quantidade de dados neste cenário. O DTP, em particular, pode não manter as variáveis de controle de sua sessão; (iii) o retorno de conectividade não pode ser feito por probes periódicos. Por isso, é essencial a elaboração de um protocolo de transmissão confiável que leve em conta as interrupções na conectividade e que aumente a capacidade da transmissão dos dados em uma rede aeronáutica utilizada para UAVs.

3. TCP modificado para UAV

Um protocolo de transmissão confiável localizado na camada de transporte da arquitetura ATN/IPS [ICAO, 2011] deve tolerar perdas de pacotes ou interrupções durante uma comunicação através do seu mecanismo de retransmissão. Um dos protocolos de transmissão confiável mais utilizados na Internet é o TCP [Jacobson et. al, 2014], o qual implementa um algoritmo de retardo da transmissão (*backoff*) toda vez que uma comunicação estiver com largos erros de pacote, tolerando um certo nível de interrupção na rede. Esta seção descreve uma arquitetura de rede que permita que o protocolo TCP possa receber informações *cross-layer* das camadas inferiores para controlar sua transmissão, utilizando a técnica de congelamento da janela de congestão.

3.1. Descrição do Protocolo TCP

O TCP é um protocolo de transporte robusto que provê um serviço de fluxo contínuo de *bytes* confiável. A transmissão é feita de forma bidirecional (a sessão permite o tráfego de dados tanto do transmissor ao receptor quanto do receptor ao transmissor). As principais funções deste protocolo são [Floyd et. al, 2004]:

- A entrega em ordem dos pacotes através do mecanismo de janela deslizante;
- O controle de congestionamento;
- O controle de fluxo dos dados, limitando a quantidade de dados que o transmissor pode enviar;
- Um mecanismo de retransmissão, garantindo a entrega de todos os pacotes a serem transmitidos.

O mecanismo de retransmissão do TCP implementa um algoritmo *backoff* para permitir uma maior tolerância em situações onde há perdas de pacotes devido a

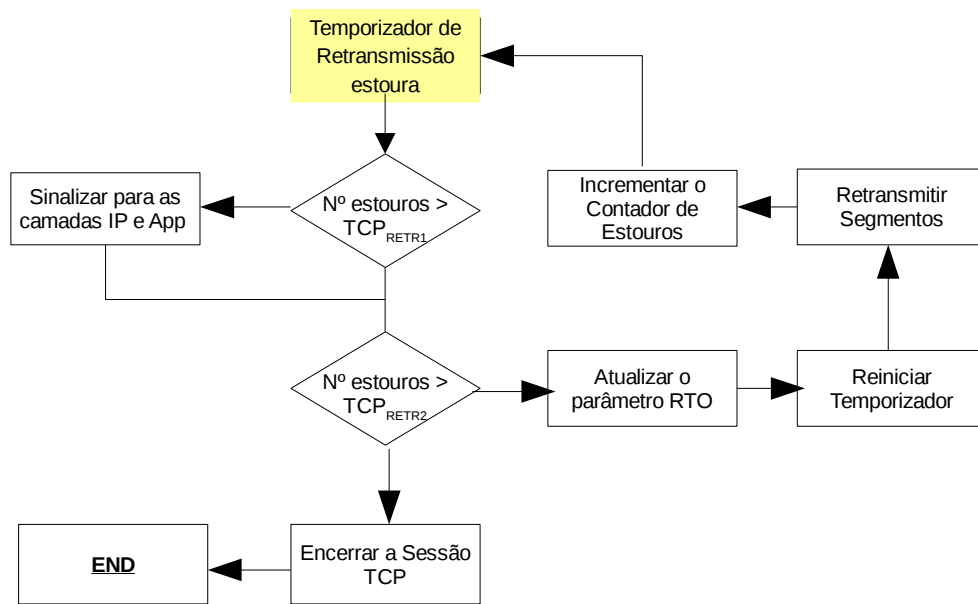


Figura 2. Fluxograma do Mecanismo de retransmissão do TCP.

redes congestionadas ou intermitências no canal. Este algoritmo possui um comportamento exponencial, o qual o retardo é aumentado exponencialmente até alcançar um patamar máximo (RTO_{Max}) e um comportamento linear, aumentando linearmente o retardo depois que este patamar for alcançado. Além disso, um parâmetro (n_{Max}) determina o número máximo de retransmissões suportadas pelo mecanismo, encerrando a sessão do TCP assim que tal parâmetro for alcançado. A Equação 1 define esta tolerância e é baseada no trabalho [Caini et.al, 2009].

A Equação

$$RTO(r) = \begin{cases} RTO_{ini} & \text{se } r = 0 \\ 2 \times RTO(r - 1) & \text{se } r < n_{expo} \\ RTO_{max} & \text{se } n_{expo} \leq r \leq n_{max} \end{cases}$$

representa o valor do estouro de retransmissão (RTO) ao longo das (r) retransmissões ocorridas em um trecho interrompido, sendo que a primeira linha da equação representa o parâmetro inicial do RTO, a segunda linha representa o comportamento exponencial do mecanismo TCP e a terceira linha representa o seu comportamento linear.

O TCP instancia um temporizador para garantir que todos os pacotes sejam corretamente transmitidos, assegurando uma entrega de dados confiável. Este temporizador é iniciado com o parâmetro RTO_{ini} toda vez que o TCP enviar um pacote pela rede. Caso o transmissor receba uma confirmação de entrega pelo receptor (ACK), o temporizador é interrompido e reiniciado com o mesmo valor RTO_{ini} . Caso todos os pacotes sejam devidamente entregues com suas respectivas confirmações, o temporizador será desligado. Todavia, se existir pacotes na rede sem confirmação de

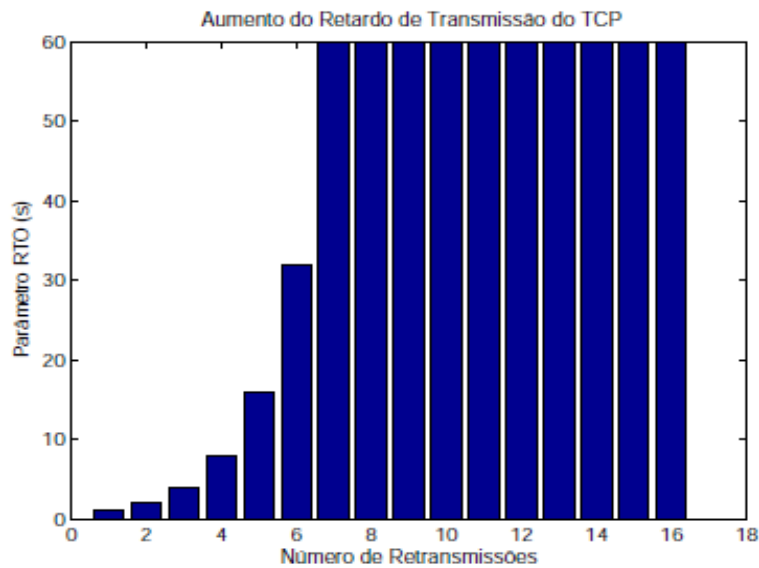


Figura 3. Algoritmo de Backoff do TCP.

entrega e este temporizador expirar, então o TCP retransmitirá os pacotes à espera de um ACK (segundo padrão *Go Back n*) e reiniciará o temporizador exponencialmente segundo a equação descrita anteriormente.

Esta operação acontece cada vez que o temporizador expira até o momento que este temporizador alcance o patamar máximo RTO_{Max} . A partir deste ponto, o temporizador assume o valor RTO_{Max} constante até que o parâmetro n_{Max} seja alcançado. O parâmetro TCP_{Retr1} é um patamar definido pelo TCP para alertar a camada IP sobre a interrupção, permitindo que esta camada inicie procedimentos específicos para tratar tal interrupção. O parâmetro TCP_{Retr2} é o valor a ser associado ao n_{max} , representando o número máximo de retransmissões suportadas. Quando isto acontecer, o TCP encerra sua sessão e somente é possível reiniciar a transmissão com a intervenção da aplicação.

A Figura 3 ilustra um exemplo do comportamento obtido pela Equação 1 ao longo de uma interrupção. Ao longo de um trecho interrompido com $n_{Max} = 16$, utiliza-se a Equação $RTO_{(n)} = 2 \times RTO_{(r-1)}$ até que r seja igual ao parâmetro $RTO_{Max} = 60$. A partir deste ponto, cada retransmissão é calculada utilizando a Equação $RTO_{(r)} = RTO_{Max}$. Depois que r for igual a n_{Max} , a sessão do TCP é encerrada e é necessário que uma aplicação estabeleça uma nova sessão para continuar transmitindo dados. O eixo X representa o número de retransmissões ocorridas durante o trecho interrompido e o eixo Y representa o crescimento do parâmetro RTO ao longo deste trecho.

4. Proposta de Adaptação do TCP para UAV

Este artigo propõe uma arquitetura de rede adaptada para aumentar a capacidade de transmissão entre aeronaves e estações ATG. A arquitetura suporta a troca *cross layer* de mensagens entre camadas inferiores e camadas superiores da pilha MIPv6, fornecendo uma camada específica para controlar a transmissão de pacotes do TCP em cenários de interrupção. Esta arquitetura aumenta o desempenho de comunicação na

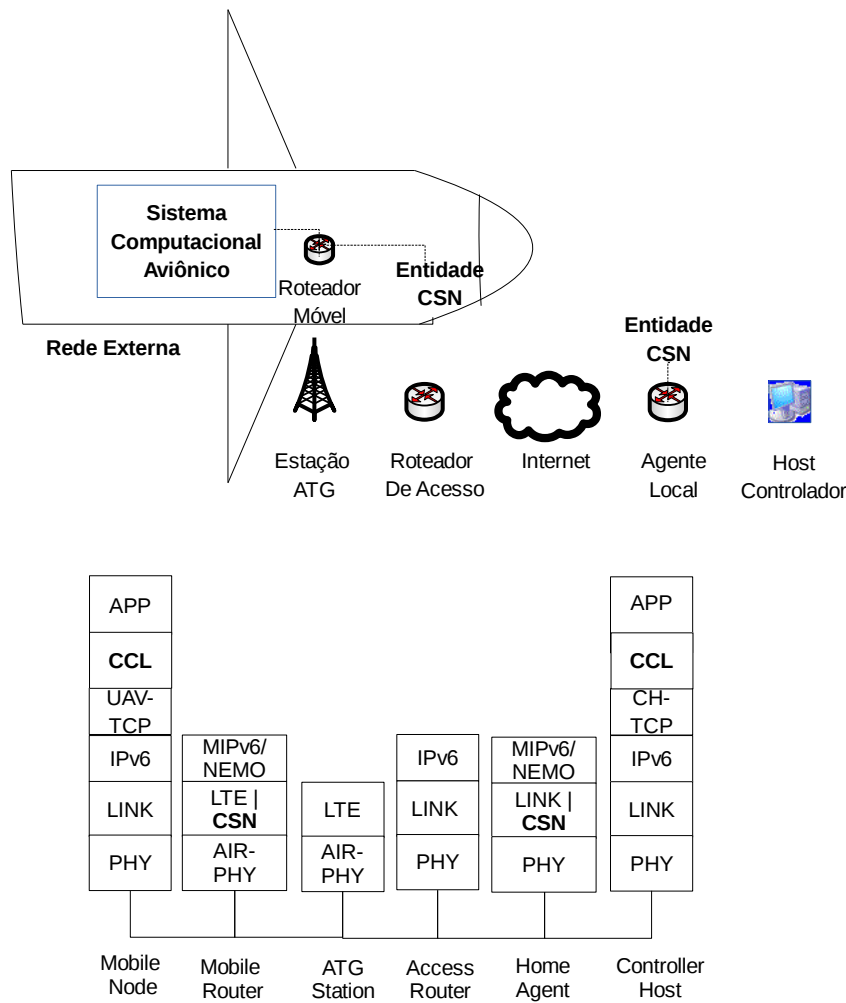


Figura 4. Arquitetura Proposta Neste Artigo para UAV.

rede, garantindo a qualidade de serviço e um alto padrão de transmissão. Diferentemente de trabalhos que consideram trajetórias de aeronaves baseadas em plano de voo [silva et. al, 2014], esta arquitetura direciona o ganho de desempenho para transmissões intra-celulares, onde não se tem informações sobre quando a conectividade entre aeronave e estação ATG será reestabelecida. Neste caso, uma arquitetura mais eficiente deve ser implementada para lidar com este cenário.

4.1. Arquitetura de Rede

A arquitetura proposta neste artigo utiliza o padrão ATN/IPS como sendo o seu modelo de referência [Ayaz et. al., 2009]. Os protocolos PMIPv6 [Johnson et. al., 2004], NEMO [Devarapalli et. al., 2005], *Diffserv* [Blake et. al., 1998], IEEE 802.21 [Taniuchi et. al, 2009] e Global HAHA (otimização na rede) são adotados nesta arquitetura segundo [Ayaz et. al, 2009]. Além disso, a tecnologia LTE é adotada como o padrão de comunicação entre a aeronave e as estações. Tal tecnologia é escolhida devido a sua elevada largura de banda e boa capacidade de transmissão com veículos de alta velocidade. Além destes protocolos, a arquitetura é composta por uma entidade de rede que notifica as camadas superiores sobre o status da conectividade do enlace e uma camada de convergência que implementa uma máquina de estados para controlar a

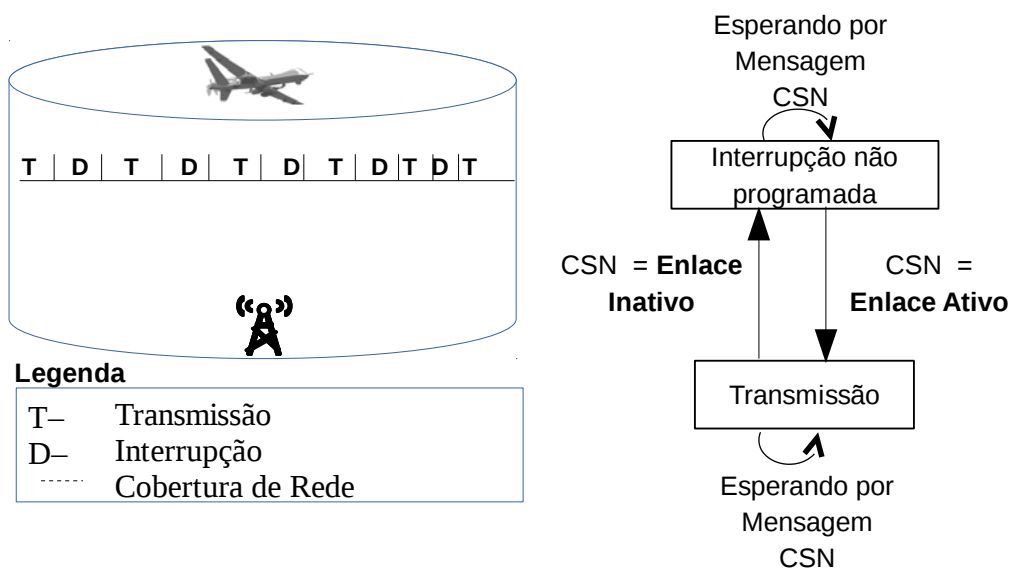


Figura 5. Comportamento da Arquitetura Proposta neste Artigo durante Transmissão intra-celular.

transmissão do TCP. A Figura 4 ilustra a pilha de protocolos utilizada na arquitetura proposta. Note que a camada de convergência controla o TCP da UAV (UAV-TCP) e o TCP do *Host* Controlador (CH-TCP). Tanto o roteador da aeronave quanto o agente local da pilha PMIPv6 implementa a entidade notificadora (CSN), sendo essencial para a identificação de interrupções durante o voo de uma UAV. Os demais protocolos são detalhados em [Ayaz et. al, 2009].

4.1.1. Notificador do estado de conectividade

O notificador de conectividade (*Connectivity Status Notifier* ou CSN) é uma entidade da rede que permite que uma camada de convergência saiba sobre o atual estado de conectividade de enlace do seu *host*. O foco desta entidade é informar, para as camadas superiores, se há ou não interrupções na comunicação. Tanto a aeronave quanto os agentes locais são modificados para suportar esta operação. Esta entidade provê o estado atual do enlace (i.e. se está ativo ou não) e o endereço MAC (*Medium Access Control*) da atual estação. Este endereço permite que aplicações da camada superior da pilha de protocolos saibam em qual estação a aeronave se encontra, servindo também para estimar a localização atual da aeronave.

A entidade CCL utiliza as informações processadas pela entidade CSN para realizar a mudança de estado em sua máquina de estados. Para tal, a entidade CCL muda o estado de interrupção para transmissão quando a mesma recebe uma informação positiva de conectividade da entidade CSN ou muda o estado de transmissão para interrupção quando a mesma recebe uma informação negativa de conectividade da entidade CSN. A entidade CSN envia suas mensagens ICMP através da técnica *cross-layer*, estabelecendo uma comunicação direta entre as outras entidades. O protocolo IEEE 802.21 [Taniuchi et. al, 2009] é utilizado como base para prover os serviços ofertados por esta entidade.

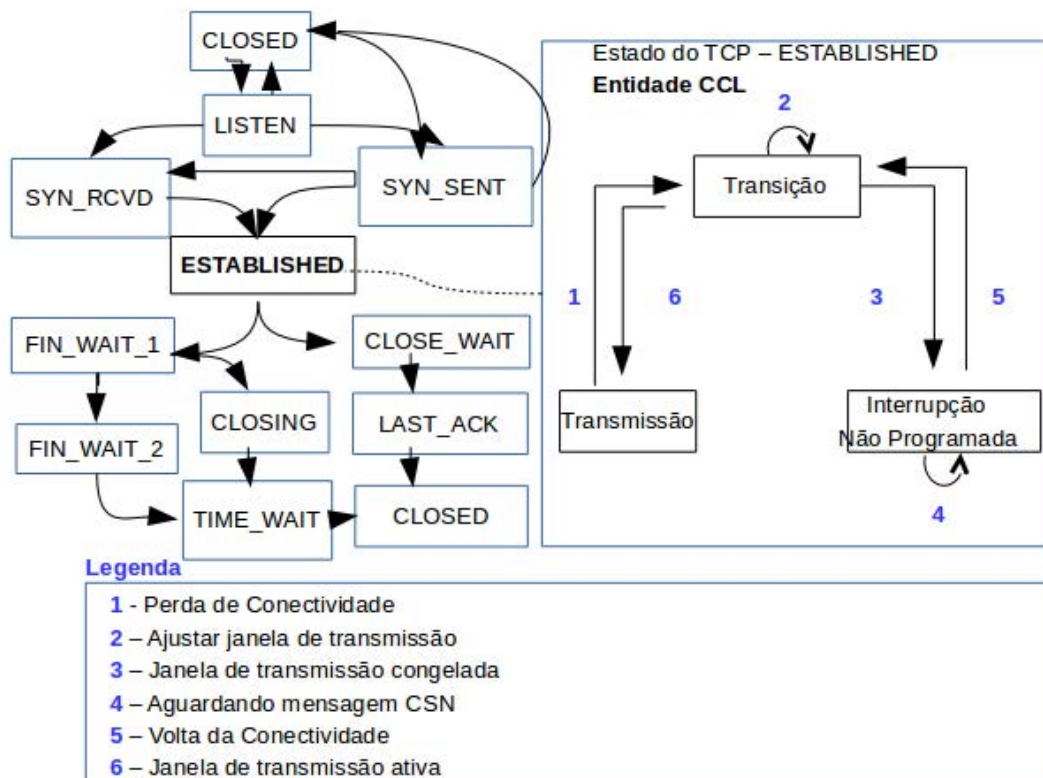


Figura 6. Máquina de Estados do TCP (à esquerda) e o uso da camada de convergência para Controle do TCP (à direita).

4.1.2. Controlador da Camada de Convergência

O controlador da camada de convergência (CCL) é responsável por controlar a transmissão do TCP durante o voo de uma UAV. Esta entidade modifica o TCP para minimizar os impactos das interrupções aeronáuticas, estabelecendo um sincronismo de transmissão entre os TCPs finais para aumentar a capacidade de transmissão da rede. Tal entidade é dividida em três módulos principais:

- **Máquina de Estados:** controla a transmissão, permitindo reduzir as retransmissões e os atrasos de reinício na comunicação. Esta entidade instancia e controla os temporizadores de transmissão e interrupção ao longo de cada segmento de rota, obtendo regularmente informações sobre o estado de conectividade do enlace. A informação do estado de enlace é necessária para realizar a transição entre os estados de transmissão e interrupção.
 - **Ajustador da Janela de Congestão:** este módulo é ativado no estado de transição do TCP, efetuando alterações na janela de transmissão do TCP para otimizar o desempenho da rede aeronáutica.
 - **Gerenciador de Mensagens:** este módulo recebe, envia e processa mensagens de sinalização entre os TCPs modificados para sincronizar os tempos de transmissão e interrupção do sistema aviônico com o sistema terrestre. O módulo recebe mensagens ICMP da entidade CSN sobre o estado de conectividade do enlace, necessário para se realizar a mudança entre os estados implementados na máquina de estados.
- As mudanças de estado devem ocorrer quando ou o temporizador de transmissão

expirar e quando a notificação de estado do enlace for recebida. A máquina de estados define 3 estados para controlar o TCP em cenários aeronáuticos: (i) estado de transmissão; (ii) estado de transição; e (iii) estado de interrupção não programada.

- **No estado de transmissão**, permitem-se que os dados sejam transmitidos durante o sobrevoo de uma aeronave em uma estação ATG. O estado de transmissão utiliza os parâmetros convencionais de retransmissão do TCP e efetua a comunicação com a rede de acesso. Além disso, permite a troca de mensagens de sinalização entre os módulos primários e secundários.
- **No estado de transição**, a janela de congestão é modificada para congelar ou reiniciar a transmissão, sendo o seu valor armazenado como uma variável de sessão do TCP.
- **No estado de interrupção não programada**, O estado só muda quando o enlace detectar novamente a conectividade. Este estado é ativado quando o enlace perde sua conectividade durante a fase de transmissão. Nesta técnica, não há necessidade de sincronismo, pois a janela de congestão, em um trecho interrompido, estará congelada até quando a máquina de estados receber uma notificação do estado de enlace, decidindo pela mudança de estado. A eficiência, neste caso, estará ligada ao agente CSN, pois este decidirá quando a janela de congestão será alterada. Os pacotes *keep alive* devem ser controlados para o correto funcionamento.

4.2. Operação

A operação da rede é descrita como segue. A CCL é responsável por controlar os parâmetros da sessão, decidindo quando a janela de congestão é congelada ou estabelecida. A camada de enlace de uma célula ATG envia mensagens de sinalização do Ipv6 contendo o estado do enlace ATG atual (i.e. ativo, interrompido, falho) para a camada de convergência situada no UCH-TCP. Neste caso, o temporizador de transmissão ainda está rodando e uma informação de não conectividade é enviada pela entidade CSN. A partir do momento em que a entidade CSN enviar tal informação, o TCP deverá mudar do estado de transmissão para estado de transição, zerando sua janela de congestão e mudando novamente para o estado de interrupção não programada, ficando no aguardo do retorno de conectividade. Este estado permite que o TCP obtenha desempenhos superiores quando as interrupções não forem previstas durante a fase de transmissão. Isto permite que não haja retransmissão e que a janela de congestão permaneça inalterada após o retorno da transmissão.

5. Propostas para padronização pela IETF

A Força tarefa de engenharia da Internet consiste em um grupo de pessoas que colaboram para a engenharia e a evolução das tecnologias de Internet. Este grupo elabora as especificações de novos padrões de Internet, não havendo uma corporação por trás nem membros ou diretores. Seu objetivo é identificar e propor soluções para problemas técnicos da Internet, especificando o desenvolvimento ou o uso de protocolos.

A padronização das entidades de rede inseridas na arquitetura discutida neste artigo deve ser realizada da seguinte forma:

1. Padronização da Arquitetura → A IETF tem documentos do tipo informativo que servem para formalizar arquiteturas, plataformas, entre outros. Neste caso, a arquitetura proposta para UAV poderia ser padronizada utilizando suas entidades de rede e a integração entre os vários protocolos IETF envolvidos.
2. Padronizar a entidade notificadora → Esta entidade requer regras para atualização de estados da máquina TCP e para solicitação e recebimento de mensagens vindas de outras entidades.
3. Padronizar a entidade controladora → Esta entidade requer regras para realizar a comunicação entre a entidade notificadora e uma aplicação controladora.

6. Considerações Finais

Este artigo propôs uma arquitetura de rede para UAV projetada para suportar longas interrupções na comunicação aeronáutica. Esta abordagem consistiu em adicionar uma entidade de rede que controla a transmissão através de notificações emitidas por outra entidade de rede instalada na camada de enlace, mitigando as retransmissões e os atrasos na transmissão de dados ocasionados pelo inadequado funcionamento do TCP em cenários de comunicação intermitente. Embora este artigo tenha investigado o uso da arquitetura em um cenário de transmissão UAV intra-celular, é possível estendê-los para outros tipos de cenários e serviços como:

- Comunicação Aeronáutica: A arquitetura proposta neste trabalho pode ser adaptada para fornecer serviços de comunicação aeronáutica. Para tal, é necessário que suas aplicações não possuam restrições de tempo (o que acontece com serviços de mensagens, típicos dos serviços de tráfego aéreo e controle operacional da aeronave).
- Redes de alta altitude: é possível adaptar a arquitetura proposta para sincronizar as transmissões por meio do uso de balões.
- Redes de satélites: pode-se utilizar informações referentes a navegação de uma constelação de satélites para realizar transmissão de dados. No entanto, altos RTTs e baixo desempenho da comunicação devido a ruídos prejudicam a qualidade do canal.
- Redes veiculares: A arquitetura proposta pode aumentar a capacidade de transmissão adequando-se os parâmetros do TCP, sendo útil para inúmeras aplicações em trens, ônibus ou carros, por exemplo.

As seguintes contribuições são destacadas: (i) transmissão otimizada, sendo o controle realizado por uma camada *overlay*, não sendo obrigatória a modificação do TCP; (ii) economia de energia de uma UAV, já que retransmissões são mitigadas; (iii) ao longo de uma transmissão intra-celular, a vazão de pacotes é amplificada, melhorando o desempenho da rede. Portanto, a arquitetura de rede proposta proporciona uma melhora na capacidade de transmissão e na transferência de dados ao longo de uma comunicação UAV, aumentando significativamente a eficiência deste sistema.

Referências

- H. Chao, Y. Cao, and Y. Chen, "Autopilots for small unmanned aerial vehicles: a survey", *Int. J. Control, Automation and Systems*, Springer, vol. 8, pp. 36-44, 2010.
- S. Ayaz, C. Bauer, C. Kissling, F. Schreckenbach, F. Arnal, C. Baudoin, K. Leconte, M. Ehammer, and T. Graeupl, "Architecture of an IP-based Aeronautical Network", *Integrated Communications, Navigation and Surveillance Conference*, IEEE, pp. 1-9, 2009.
- Jacobson, V., Braden, B., e Borman, D. (2014) "TCP extensions for high performance", RFC 7323, IETF Standard.
- ICAO (2011) "Manual for the ATN using IPS Standards and Protocols", Doc 9896, ICAO draft.
- Caini, C., Cornice, P., Firrincieli, R., Livini, M. e Lacamera, D. (2009) "Analysis of TCP and DTN retransmission algorithms in presence of Channel Disruptions", In *International Conference on Advances in Satellite and Space Communications*, IEEE Publisher, Piscataway.
- Silva, R.A., Netto, J.E., Paiva, M.A.C., Anzaloni, A. "Uma abordagem de Transmissões Programadas Atenta a Interrupções para Redes Aéreo-Terrestres", *IEEE Latin America Transactions*, v. 12, n. 8, p. 1409-1417, 2014.
- Devarapalli, V., Wakikawa, R., Petrescu, A. e Thubert, P. (2005) "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, IETF Standard.
- Floyd, S., Gurtov, A. e Henderson, T. (2004) "The NewReno modification to TCP's fast recovery algorithm", RFC 2582, IETF Standard.
- Johnson, D., Perkins, C. e Arkko, J. (2004) "Mobility support in IPv6", RFC 3775, IETF Standard.
- Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z e Weiss, W (1998), "An Architecture for Differentiated Services", RFC 2475, IETF Standard.
- Goff, T., Moronski, J., Phatak, D., Gupta, V. "Freeze-TCP: A true end-to-end TCP enhancement mechanism for mobile environments", *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, v.3, pp 1537-1545, 2000.
- Go, Y., Moon, Y., Nam, G., Park, K. A disruption-tolerant transmission protocol for practical mobile data offloading, *Proceedings of the third ACM international workshop on Mobile Opportunistic Networks*, pp 61-68, ACM, 2012.
- Birrane, III., Edward, J., Burleigh, S.C., Cerf, V. Defining tolerance: impacts of delay and disruption when managing challenged networks, *Proceedings of AIAA Infotech@ Aerospace Conference*, 2011.
- Taniuchi, K., Ohba, Y., Fajardo, V., Das, S., Tauil, M., Cheng, YH, Dutta, A., Baker, D., Yajnik, M., Famolari, D. IEEE 802.21: Media independent handover: Features, applicability, and realization, *Communications Magazine*, IEEE, v. 47, nº 1, pp 112-120, 2009.

Protocolo de comunicaciones para control de la generación distribuida de flujo multimedia

C. F. Perez-Monte^{1,2}, M. F. Piccoli², M. Perez¹, C. Luciano^{3,4}, S. Rizzi⁵

¹GridTICs – Universidad Tecnológica Nacional - Facultad Reg. Mendoza (UTN-FRM)
Mendoza, Argentina

²LIDIC – Universidad Nacional de San Luis (UNSL)
San Luis, Argentina

³Dept. of Biomedical and Health Information Sciences

⁴Dept. of Bioengineering - University of Illinois at Chicago (UIC)
Chicago, USA

⁵Argonne Leadership Computing Facility – Argonne National Laboratory (ANL)
Chicago, USA

cristian.perez@gridtics.frm.utn.edu.ar, mpiccoli@unsl.edu.ar,
mauricio.david.perez@gmail.com, clucial@uic.edu, srizzi@alcf.anl.gov

Resumen. *Aplicaciones de tiempo real como realidad virtual generan un flujo multimedia de tipo de máquina-humano que puede ser dividido tanto espacial como temporalmente para distribuir recursos de procesamiento, memoria o red. Este trabajo describe un protocolo para control de flujo de datos multimedia, donde su generación debe ser distribuida para incrementar el rendimiento. Este protocolo no describe la comunicación del flujo multimedia, sino solamente la comunicación de control para la generación de un único flujo multimedia entre múltiples nodos teniendo en cuenta un esquema de mejor esfuerzo. Para lograr estos objetivos se propone un protocolo de capa de aplicación, el cual utiliza UDP o UDP-lite en capa de transporte, IPv6 multicast en capa de red y Gigabit Ethernet con control de flujo en capa de enlace de datos.*

1. Introducción

Diversos protocolos se han propuesto para el transporte de información multimedia requerida para la comunicación humano-humano, ejemplos son el video o el audio. Existe un área específica de IETF denominada ART [IETF 2016] con grupos de trabajo dedicados al continuo desarrollo de los mismos. En la actualidad, aplicaciones como la realidad virtual, permiten el transporte de información multimedia para la comunicación humano-máquina, utilizado ampliamente en la industria del entretenimiento [Huang 2013]. En dichas aplicaciones, generalmente existe una comunicación de ancho de banda reducido, correspondiente a la comunicación humano a máquina y una comunicación de ancho de banda elevada, correspondiente al flujo multimedia de video de máquina a humano. Dicho flujo multimedia, para ser foto-realista, debe poseer técnicas de generación de la escena complejas y, además, poseer resolución y cantidad de cuadros por unidad de tiempo elevadas para una fluidez aceptable. Para cumplir con estos requisitos se necesita una elevada capacidad de procesamiento, una apropiada

solución es la utilización de numerosos nodos para la generación del flujo de video multimedia.

En el presente trabajo se propone un protocolo de comunicaciones para la sincronización de tareas en la generación de un flujo de video multimedia. El desarrollo de un protocolo de comunicación para controlar un flujo multimedia en tiempo real y en un sistema de múltiples nodos no es una tarea trivial, por lo cual se sugiere su discusión en grupos de IETF para su estandarización. En este trabajo se propone sentar las bases para iniciar su discusión, por ello el protocolo de comunicación fue simplificado, optimizado y generalizado para su utilización en un mayor número de aplicaciones. Representa una actualización de un trabajo anterior [Perez-Monte 2014] donde se mostró su viabilidad a través de resultados experimentales.

El presente trabajo se divide de la siguiente manera: en primer lugar, en esta sección desarrollamos conceptos básicos de la renderización distribuida y trabajos relacionados. En la sección 2, describimos un sistema de procesamiento paralelo donde los nodos tienen diferentes roles, y un modelo de distribución de tareas, los cuales son el punto de partida para el diseño del protocolo de comunicaciones. Finalmente, en la sección 3 exponemos el protocolo de comunicaciones.

1.1. Renderización en Tiempo Real y Foto-realista

Un modelo 3D es una representación de un objeto 3D, el cual puede ser descrito a través de polígonos o de un volumen compuesto de voxels u otra técnica. Renderización es un término frecuentemente utilizado para describir el proceso por el cual a un modelo 3D se le realiza un procesamiento por el cual, mediante determinados parámetros de entradas, se genera una imagen 2D desde un punto de vista del observador. Tanto la ubicación del observador como su orientación visual son los parámetros de entrada más importantes, sin embargo pueden existir otras como la ubicación de la iluminación. En la renderización en tiempo real, los parámetros de entrada pueden variar constantemente debido al cambio constante de posición del observador o las luces, implicando una generación de imágenes 2D en el tiempo. Además de los parámetros de entrada, el modelo 3D determinará un estado interno del método de renderización, el cual puede cambiar si la escena cambia o los objetos constitutivos sufren desplazamientos o deformaciones.

Así como la renderización tiene como finalidad generar una imagen para la visualización de un modelo 3D, la renderización foto-realista tiene como objetivo, además, generar una imagen con una realidad visual lo más cercana a la realidad usando técnicas visuales de iluminación y sombreado demandantes de un procesamiento elevado. La misma tiene gran aplicación en áreas como el entretenimiento, entrenamiento médico, planeamiento quirúrgico entre otras.

Una mayor capacidad de procesamiento puede obtenerse a partir de sistemas paralelos constituidos por grupos de computadoras conectadas en red. Para poder aprovechar la capacidad de dichos sistemas se debe recurrir a técnicas de renderización paralelas, descritas a continuación.

1.2. Renderización Paralela Distribuida

En renderización paralela distribuida, la taxonomía de Molnar [Molnar 1994] [Cox 1995] establece dos tipos de técnicas basadas en la manera en cómo distribuye el trabajo

entre los diferentes nodos de un sistema distribuido. Estas dos técnicas se conocen con el nombre de *Sort first* o *2D* y *Sort last*, sus principales características son:

- La renderización paralela distribuida *Sort first* o *2D* propone la división de la pantalla en áreas no solapadas, asignando una a cada nodo o unidad de cómputo disponible. De esta forma, cada nodo renderiza una sección determinada de la pantalla completa. Ha sido aplicado en numerosos sistemas [Schwarz 2010] [Moloney 2011].
- En la técnica *Sort last*, la renderización es hecha dividiendo sectores de volúmenes o primitivas, aplicando la renderización y realizando la composición de los resultados parciales para obtener la imagen final. También ha sido ampliamente aplicado. [Eilemann 2007][Marchesin 2008].

Una tercera técnica de renderización distribuida fuera de la taxonomía de Molnar es *Alternate Frame Rendering (AFR)*, muy utilizada en ambientes con varias GPUs en una única PC [Diard 2009][Monfort 2009]. Se basa en la distribución de cuadros de pantalla.

De las tres técnicas citadas, *2D* y *AFR* pueden ser aplicadas con el sistema propuesto en este trabajo, constituyendo una buena solución para la aplicación en la renderización distribuida foto-realista en tiempo real. También es posible aplicar otras técnicas, propias de algunos algoritmos de renderización. Además, todas pueden combinarse y definir nuevas políticas aplicables a ambientes distribuidos-paralelos para resolver problemas con gran demanda computacional.

1.3. Sistemas de Renderización Causales

Un sistema de renderización es causal o determinístico cuando, ante parámetros iniciales de entrada y de estado interno, es perfectamente determinable la imagen de salida renderizada. El sistema de renderización propuesto en la próxima sección tiene estas características. Además podemos determinar dos tipos de sistemas de renderización causales, ellos son:

- Determinista externo: Donde los parámetros de entrada no modifican el estado interno del objeto 3D por lo cual su estado final (y por lo tanto también la imagen de salida renderizada) estará determinado exclusivamente por el último parámetro de entrada. Este es el caso de sistemas en los cuales los parámetros de entrada son, por ejemplo, posición del observador o iluminación.
- Determinista global: En los cuales los parámetros de entrada modifican el estado interno del objeto 3D a renderizar, por lo cual su estado final estará determinado no solo por los últimos parámetros de entrada sino también por el estado interno que, al ser un sistema determinista, es equivalente a todos los parámetros de entrada previos con la información del momento del tiempo de ser aplicados. Este es el caso de sistemas en las cuales existen, además de los anteriormente mencionados, parámetros de entrada tales como fuerza sobre un objeto.

El sistema propuesto en la próxima sección puede funcionar en ambas modalidades.

1.4. Trabajos relacionados

El protocolo presentado en el siguiente trabajo, si bien guarda relación con otros protocolos de flujo multimedia, está más fuertemente relacionado con protocolos destinados al cómputo paralelo. Entre ellos, destaca el protocolo de comunicaciones utilizado en *Message Passing Interfase* (MPI) [Gropp, 1999], el cual es empleado en computación concurrente aportando la sincronización entre procesos mediante el paso de mensajes. El mismo supone una valiosa contribución a la computación paralela. Sin embargo, está diseñado para aplicaciones en las cuales generalmente el resultado es un proceso computacional que se desea que se procese rápidamente mediante el uso de múltiples nodos. Sin embargo raramente es requerimiento que el resultado de dicho proceso sea para uso en tiempo real. MPI generalmente se implementa sobre la capa de transporte TCP, obteniendo todas las ventajas del mismo, pero al tratarse de un protocolo bidireccional no es adecuado para los casos de procesamiento destinados a tiempo real.

Por ello es apropiada la propuesta de un protocolo, que al igual que el utilizado en MPI, aporte sincronización entre procesos, pero sin la necesidad de múltiples comunicaciones ni comunicación de respuesta. Dicha propuesta parte de la determinación de que un único nodo en el sistema contendrá el estado del sistema y que la propagación del estado será unidireccional y compacta a cada nodo de procesamiento. De esta forma resulta en un sistema con la menor latencia posible para el caso específico de generación de flujos multimedia en tiempo real. Además se obtiene la ventaja adicional de funcionar en condiciones de saturación del canal de comunicaciones.

Por lo tanto, en la próxima sección abordaremos este tema con más detalle, aportando una propuesta de sistema de renderizado, a partir del cual planteamos el protocolo planteado en el trabajo.

2. Sistema de Renderización Distribuida en Tiempo Real

El sistema propuesto está compuesto por 3 tipos de nodos, cada uno de los cuales tiene un rol definido. Los roles que puede asumir un nodo en el sistema son:

- *Nodo Administrador (NA)*: Es el nodo que recibe y administra los parámetros de entrada recibidos desde un dispositivo de entrada para transmitirlos por red a los *Nodos Procesador*. Adicionalmente debe anunciar a los *Nodos Procesador* qué nodo o *Nodos Integrador* les corresponde. Opcionalmente puede determinar, o no, que procesa cada *Nodo Procesador*.
- *Nodo Procesador (NP)*: Es el nodo que realiza el renderizado y envía el trabajo finalizado por red al *Nodo Integrador*. Puede asumir, en el caso de que el *NA* no lo haga, la función de auto asignarse la tarea a procesar.
- *Nodo Integrador (NI)*: Es el nodo que recibe todas las tareas procesadas y las integra pudiendo ser el encargado también de realizar la visualización en tiempo real de la escena renderizada. Pueden existir múltiples *NI*s pero existe una limitación, cada *NI* tendrá asignado determinados segmentos de la imagen que no podrán estar asignados simultáneamente a otros *NI*s.

Considerando los distintos tipos de nodos, el sistema de renderización distribuido consta de un *NA* para coordinar las tareas a realizar por los demás nodos, el

cual puede estar separado del resto del sistema en una red alejada permitiendo el control del sistema remotamente; varios *NPs*, quienes realizan la renderización propiamente dicha, y uno o varios *NIs* responsables de reunir y realizar la visualización en tiempo real del volumen. Los *NPs* y *NIs* deben estar ubicados físicamente cercanos para su conexión a través de una red de alto rendimiento y baja latencia. Múltiples *NAs* pueden ser soportados para funcionamiento multi-usuario o multi-vista [Hubner 2007]. Además múltiples *NI* son soportados para incrementar la resolución de salida mediante la utilización de muros de pantallas [Johnson 2012][Jiang 2015] o para incrementar el ancho de banda de tráfico de datos de salida unicast cuando el limitante es la interfaz del *NI*.

En la siguiente sección analizamos el modelo de comunicación propuesto para poder gestionar este sistema.

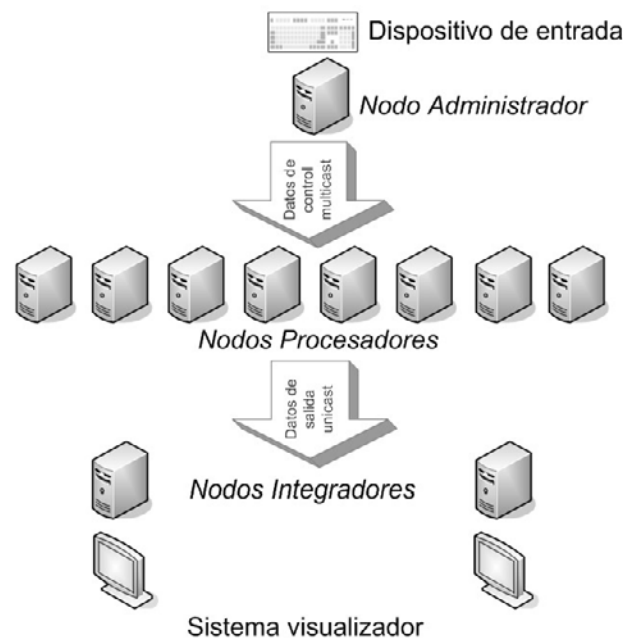


Figura 1. Esquema del sistema

2.1. Modelo de Comunicación Propuesto

Los nodos en el sistema interactúan entre sí de una manera determinada (ver figura 1), estableciendo de este modo dos tipos de comunicaciones, las cuales son:

- **Información de control :** Esta comunicación tiene lugar desde el *NA* a los *NPs* para enviar la información de los parámetros de entrada e información de pertenencia de los segmentos de tareas que le permite conocer a los *NPs* qué segmento de tarea debe procesar y a cuál *NI* se debe enviar cada segmento de tarea procesada. Dado los escasos requisitos necesarios de ancho de banda, esta información puede ser transferida a través de enlaces WAN a redes remotas. La utilización de multicast disminuye los efectos que la latencia de los enlaces WAN pueda ocasionar en el funcionamiento del sistema.
- **Información de salida:** Esta comunicación se lleva a cabo entre los *NPs* y los *NIs*. Cada una de ellas contiene la información del flujo multimedia distribuido, el cual está compuesto por los sucesivos cuadros en el tiempo generados por cada uno de los *NPs*. Los requisitos de gran ancho de banda y baja

latencia requieren de un enlace LAN exclusivo para la comunicación de salida entre los *NPs* y los *NIs*.

Como se indicó anteriormente, el presente trabajo está orientado exclusivamente al protocolo de la información de control. En las siguientes secciones se detallan las principales características y el modo de funcionamiento del modelo de comunicación.

2.2. Características Destacadas del Modelo de Comunicación

El modelo de comunicación diseñado tiene las siguientes características:

- **Segmentado Jerárquico:** El protocolo divide el flujo multimedia a procesar en segmentos en el tiempo y en el espacio para facilitar su procesamiento distribuido. Definimos dos niveles de segmentación, el primero dentro del flujo multimedia, cada bloque de información multimedia que comparte la misma etiqueta de tiempo la denominamos bloque de datos, *Data Block (DB)*. Luego se realiza la división espacial de dicho *DB* a lo que denominamos segmento espacial de datos, *Spatial Data Segment (SDS)*. Esta característica es similar a la utilizada por modelos de programación para GPUs [Kirk 2010] [Perez-Monte 2014]. Un ejemplo común de *DB* es un cuadro o imagen en una sucesión de cuadros de un flujo de video multimedia, mientras que un *SDS* es una porción o subdivisión espacial de dicho cuadro o imagen renderizada.
- **Mejor esfuerzo:** El protocolo al igual que el sistema en su conjunto deben tener una filosofía del “mejor esfuerzo” en el sentido de que ante limitaciones reales tales como ancho de banda de enlace, *NPs* con diferente potencia de cómputo, nodos ocupados, etc., el sistema debe hacer el mejor esfuerzo posible para llevar a cabo la tarea. Por esa razón que el protocolo elegido es IPv6-UDP, el cual es orientado a datagramas siguiendo también un modelo de entrega de mejor esfuerzo.
- **Tolerante a saturación de enlace:** De la mano del concepto anterior, el protocolo debe ser capaz de aprovechar al máximo el ancho de banda de red y ante la saturación del mismo debe ser capaz de continuar en operación normal. La pérdida de información no repercute en el sistema ya que cada paquete de datos o datagrama incluye toda la información necesaria sobre el estado del sistema.
- **Energéticamente eficiente:** El procesamiento de información insume un consumo de energía considerable sobre los *NPs* y como el sistema de comunicaciones realiza el máximo esfuerzo por transportar los datos de salida pero no siempre puede lograrlo, es conveniente procesar solo la cantidad de datos que el enlace entre los *NPs* y los *NIs* sea capaz de transportar.
- **Tolerante a fallas de *NP* o *NI*:** Cuando la cantidad de *NPs* crece, las posibilidades de que al menos un nodo falle se incrementa, es por ello que, si la política de renderización lo acompaña, el protocolo está preparado para tolerar la falla de un nodo y que el sistema continúe en operación.
- **Sin Estado Interno en *NP* :** El estado completo del sistema es enviado en la información de control incluso si los parámetros de entrada modifican el estado interno, como es el caso de sistemas deterministas globales.

Estas características de diseño del modelo de comunicación determinan tanto el

modo de funcionamiento como el diseño del protocolo, ambos descriptos a continuación.

2.3. Modo de Funcionamiento

El modo de funcionamiento a través de tiempo consta de tres etapas en donde los estados generales del sistema compuestos por los parámetros de entrada y sus correspondientes parámetros de tiempo son administrados por el *NA*.

En una primera etapa el *NA* debe obtener información sobre los identificadores únicos de *NP* y *NI*. Dicho método, tanto de identificación individual, grupal y de descubrimiento no está cubierto por el presente trabajo y se espera utilizar protocolos de comunicación existentes para dicha etapa.

En una segunda etapa, en la cual está definido el protocolo, el *NA* envía, por multicast a todos los *NP*, los datos de control que constan de:

- Especificación de presencia obligatoria de cómo generar el flujo multimedia.
- Especificación opcional de qué *NP* deben realizar procesamiento por cada *SDS*.
- Especificación opcional de los *NI* a los cuales enviar el procesamiento realizado por los *NPs*.
- Especificación de presencia obligatoria de los parámetros de entrada capturados de un dispositivo de entrada con información del tiempo al momento de la captura.

Así los *NPs* reciben del *NA* qué *SDS* procesar y a qué *NI* enviar el resultado o, por el contrario, determinan internamente ambas decisiones. El envío se realiza periódicamente con lapsos de tiempo inferiores al tiempo de refresco del sistema visualizador asegurando así que siempre los *NPs* tengan tarea para realizar. Se puede considerar que el *NA* gestiona la creación de un flujo multimedia virtual, de forma tal que los *NPs* generan un flujo multimedia real de menor rendimiento de acuerdo a la capacidad de cómputo global permitida. Cada *NP* recibe los datos de control, eligiendo el más actual posible y descartando el resto, y a partir del cual realiza el procesamiento adecuado de cada *SDS* de cada *DB*.

En una tercera etapa, no cubierta aún en el presente protocolo, los *NPs* envían los *SDS* procesados por tráfico unicast a él o los *NIs* correspondientes. El envío de esta información se realiza tan rápido como cada *NP* es capaz de hacerlo, siendo el límite impuesto por el procesamiento de los *NPs* o por el ancho de banda de la red de entrada de los *NIs*. Los *NIs* tienen como función reunir el flujo multimedia distribuido y finalmente visualizarlo en un sistema visualizador compuesto de tantas pantallas como *NIs* existan o reenviarlo a un único nodo visualizador local o remoto para que realice la visualización en un único dispositivo de salida.

2.4. Principios de Diseño del Protocolo.

El protocolo debe contar con determinados principios de diseño, las cuales han sido determinados para un adecuado funcionamiento y flexibilidad en el uso:

- Existirá un identificador único de tamaño compacto para identificar a :

- Cada uno de los *SDSs* dentro de un *DB*.
 - Grupos de *SDSs* dentro de un *DB* incluido un grupo todos los *SDSs* de un *DB* y ningún *SDS* del *DB*.
 - Cada uno de los *NPs*.
 - Grupos de *NPs* incluido un grupo todos los *NPs*.
 - Cada uno de los *NIs*.
- Los *NPs* pueden procesar múltiples *SDSs* de cada *DB*. Incluso más de un *NP* pueden procesar el mismo *SDS* de un mismo *DB* (de forma diferente o no). Métodos de renderización iterativos [Perez-Monte 2013] o tolerantes a fallas nodales pueden justificar este modo de funcionamiento.
 - Diferentes *SDSs* de un mismo *DB* pueden ser enviados a diferentes *NIs* pero un *SDS* de un *DB*, si bien puede ser procesado por más de un *NP*, solo puede ser enviado a un único *NI*.
 - Por defecto todos los *NPs* no especificados procesan todos los *SDSs* de cada *DB*. En el caso de que el *NA* lo especifique, *NPs* especificados solo procesan los *SDSs* que son explícitamente especificados.
 - Por defecto los *NPs* no envían las tareas de los *SDSs* asignados a ningún *NI*, por lo cual el *NA* debe indicar al menos un *NI* por cada *SDS* procesado para que el sistema funcione adecuadamente ya que los *NPs* interrumpirán el procesamiento en caso de no poder realizar el envío de dicho procesamiento a los *NIs*.
 - El *NA* podrá especificar que:
 - Un *SDS* específico o un grupo de *SDSs* específico sea procesado por un *NP* o grupo de *NPs*.
 - Un *SDS* específico ya procesado o un grupo de *SDSs* específico ya procesados sea enviado a un *NI* especificado. No deberán especificarse grupos de *NIs*.
 - Todos los *SDSs* o grupos de *SDSs* ya procesados por especificados *NPs* o grupos de *NPs* sean enviado a un *NI* especificado.
 - Todos los *SDSs* o grupos de *SDSs* ya procesados y enviados a especificados *NIs* sea procesados anteriormente por especificados *NPs* o grupos de *NPs*.
 - Todos los *SDSs* o grupos de *SDSs* ya procesados por especificados *NPs* o grupos de *NPs* sean procesados también por otro nodo o grupo de *NPs* especificados.
 - En caso de que un *SDS* sea especificado para procesarse por más de un *NP*, será procesado por cada uno de ellos y por todos los *NPs* no especificados. Esto no se aplica a grupos de *NPs*. Esta característica se denomina redundancia de procesamiento automática.
 - En el caso de que un *SDS* procesado sea especificado para enviar a más

de un *NI*, será enviado solo al de mayor privilegio. En caso de no poder hacerlo, se enviará al subsiguiente en privilegio de los especificados. Esta característica se denomina redundancia especificada de integración.

- Si un *NP* es asignado más de una vez a diferentes *SDSs*, estará asignado al conjunto de los *SDSs* de todas las asignaciones (operación OR).

Todos los principios y características del modelo de comunicaciones descriptos previamente permiten la especificación de una propuesta de protocolo como se describe en la próxima sección.

3. Protocolo SDCP

El protocolo diseñado es denominado Streaming Distributed Control Protocol (SDCP) y se basó en trabajos previos para los cuales se incorporaron los principios considerados en la sección previa. A lo largo de la presente sección describiremos los protocolos de capa inferior requeridos para su funcionamiento, la descripción del protocolo y los dos modos fundamentales de funcionamiento.

3.1. Arquitectura de la Pila de Protocolos

La arquitectura de la pila de protocolos que se utiliza para cumplir con los objetivos precedentes, es la siguiente:

- Capa de enlace de datos: En el caso de los enlaces entre los *NPs* y *NI*s se utiliza Ethernet con control de flujo (opcional) [IEEE 1997]. En caso de tener control de flujo, se permite que ante saturación de enlace los *NPs* solo procesen lo que son capaces de enviar por la red, lo cual reduce el consumo energético. Una implementación con esta capa de enlace ha sido validada experimentalmente; se presentó además un análisis de rendimiento en un trabajo previo [Perez-Monte 2014].
- Capa de red: Protocolo IPv6 [Deering 1998], el estándar de red de alcance mundial en su versión más reciente que lo hace adecuado para computación distribuida y con multicast para reducir el tráfico de información de control.
- Capa de transporte: Protocolo UDP [Postel 1980] o UDP-Lite [Larzon 2004], implementando un protocolo con un modelo de entrega de mejor esfuerzo y tolerante a saturación de enlace. En el caso de utilizar UDP-Lite se exige que el checksum abarque al menos los primeros 256 bits del protocolo correspondientes al encabezado obligatorio.

De esta forma SDCP será un protocolo definido en la pila de protocolos en capa de aplicación. A continuación se detallará la estructura y descripción del protocolo.

3.2. Descripción de SDCP

En los sistemas de renderizado descriptos se distinguen dos tipos de comunicación, las cuales son Información de Control Multicast e Información de Salida Unicast. SDCP describe la primera.

SDCP está constituido por datagramas que tienen como función principal el envío de los datos de control de entrada. Incluyen 2 cabeceras, una principal y la

siguiente opcional. La estructura es la siguiente:

Tabla 1. Protocolo de Control

Tipo de Estructura	Campos	Descripción
Cabecera IPv6 + Cabeceras opciones	Definidos por protocolo	Definidos por el protocolo según RFC 2460
	Dir. IPv6 de destino	Dirección IPv6 de destino multicast del grupo de los NP según RFC 3306.
Cabecera UDP	Definidos por protocolo	Definidos por el protocolo según RFC 768.
Cabecera de control obligatoria	Control Data type	8 bits. Tipo de control.
	M	1 bit. Bit de modo de control. Instantaneo o Historial.
	RD	3 bits. Reservado para uso futuro.
	Streaming Generator Id	20 bits. Identificador de fuente generadora (modelo 3D) de flujo multimedia.
	Var DB Counter	32 bits . Campo de contador de número de variaciones de payload enviado.
	SDCP Counter	32 bits. Campo de número de paquete enviado.
	Timestamp	64 bits. Marca de tiempo de los datos de control.
	DB type	16 bits. Descriptor de tipo de DB.
	DB size	32 bits. En múltiplos de cantidad de SDSs.
	SDS size	32 bits. En cantidad de pixels vertical y horizontal.
	Next Header Counter	16 bits. Número de cabeceras opcionales. Equivalente a tamaño de cabecera opcional en múltiplos de 16 octetos.
Optional Header Cabecera opcional de asignación de recursos.	First Assigned work	64 bits. Número de identificación de tarea asignada.
	First Resource	64 bits. Número de identificador de recurso a asignarle tareas.
	Next Assigned work	128 bits. Tantos como lo indique Next Header Counter.
	Next Resource	
Payload: Cuerpo con datos de control.	Last data control	Tamaño especificado por tipo de control . Datos de control correspondientes a la marca de tiempo actual del encabezado. Es un campo obligatorio.
	Historical timestamp Historical data control	Tamaño especificado por tipo de control + 64 bits. Solo si Modo historial activado . Tantos como lo indique Var DB Counter

- Cabecera de control obligatoria: Posee la información obligatoria general.

- Cabecera de control opcional: Información opcional para anunciar a los *NP*s, dicha información puede ser sobre los propios *NP*s (que *SDS* o grupo de *SDS*s debe procesar cada uno) o sobre los *NI*s (que *SDS* o grupo de *SDS*s tiene asignado cada *NI* recibir).
- Cuerpo del paquete: Información obligatoria necesaria para la generación del flujo multimedia por parte de los *NP*s.

El datagrama de la información de control se muestra en la tabla (ver tabla 1).

A partir de la estructura y descripción realizada, a continuación presentamos los modos de renderización que definen el contenido y estructura del cuerpo del protocolo.

3.3. Modo de Renderización

El objetivo fundamental del protocolo es permitir enviar los parámetros de entrada de renderización conocidas por el *NA* a todos los *NP*.

Cuando los parámetros de renderización no generan cambios internos en el modelo 3D cada *NP* puede renderizar la escena simplemente conociendo el último Dato de Control conocido y el sistema funcionará en Modo Instantáneo indicado en el bit M del encabezado principal y correspondiente a un sistema de renderización determinista externo. Cuando los parámetros de renderización generan cambios internos en el modelo 3D, cada *NP* sin estado interno solo puede renderizar adecuadamente la escena si conoce todos los datos de entrada desde el inicio de generación del flujo multimedia y el momento del tiempo en que cada uno de ellos fue generado. En este caso el sistema funcionará en Modo histórico y será así indicado en el bit M del encabezado principal. Por otro lado, en el cuerpo del paquete existirán tantos grupos de datos de entrada con su correspondiente marca de tiempo como lo indique el campo Var DB counter del encabezado principal. De esta forma el sistema de renderización será determinista global.

Finalmente, a continuación, se describe el contenido de los encabezados opcionales que definen el funcionamiento del sistema de acuerdo a las necesidades del *NA*.

3.4. Modo de Asignación de Trabajo a los Recursos

Los modos de asignación de tareas (en donde una tarea puede ser tanto procesar un *SDS* por un *NP* o que un *NI* reciba un determinado *SDS* ya procesado) están determinados por la información contenida en el encabezado opcional.

Los *NP*s actúan en un modo libre, por defecto todos los *NP*s pueden procesar todos los *SDS*s (la política de elección es internamente elegida por dichos nodos) y más de uno puede hacerlo simultáneamente (a menos que se explicita). Por el contrario los *NI*s funcionan en un modo restringido, por defecto ningún *NP* envía *SDS*s a ningún *NI* a menos que se explicita lo contrario. Si se asignan los mismos *SDS*s procesados a más de un *NI*, los *NP*s envían éstos al de mayor prioridad (a menos que falle, en tal caso se envía al siguiente en prioridad). Los modos de asignación son los expresados en la tabla 2.

Tabla 2. Modos de asignación de recursos

Modo de asignación	Identificador de trabajo	Identificador de Recurso
Asig. para procesamiento del SDS por NP	Id. de un SDS	Id. de un NP
Asig. para procesamiento del grupo de SDSs por NP	Id. de grupo de SDSs	Id. de un NP
Asig. para procesamiento de SDS simultáneamente por todo el grupo de NP	Id. de un SDS	Id. de un grupo de NPs
Asig. para procesamiento de todo el grupo de SDSs simultáneamente por todo el grupo de NP	Id. de un grupo de SDSs	Id. de un grupo de NP
Asig. de envío de SDS procesado a NI	Id. de un SDS	Id. de un NI
Asig. de envío de un grupo de SDSs procesados a un NI	Id. de grupo de SDSs	Id. de un NI
Asig. de los SDS procesados por el NP 1 para que sean procesados también por el NP 2	Id. de NP 1	Id. de NP 2
Asig. de todos los SDSs procesados por el grupo NP 1 para que sean procesados también por grupo NP 2	Id. de grupo de NP 1	Id. de grupo de NP 2
Asig. de los SDSs procesados por NP para que sean enviados al NI	Id. de NP	Id. de NI
Asig. de todos los SDSs procesados por el grupo de NP para que todos sean enviados al NI	Id. de grupo de NP	Id. de NI
Asig. de todos los SDSs que serán enviados al NI para que sean procesados previamente por el NP	Id. de NI	Id. de NP
Asig. de todos los SDSs que serán enviados al NI para que sean procesados previamente por todos los Nodos del grupo de NP	Id. de NI	Id. de grupo de NP
Prohibido	Id. de NI 1	Id. de NI 2
Prohibido	Id. de SDS 1	Id. de SDS 2
Reservado para Uso Futuro	Id. de un SDS	Id. de grupo de SDSs
Reservado para Uso Futuro	Id. de grupo 1 de SDSs	Id. de grupo 2 de SDSs
Sin efecto	Id.	Mismo Id.
Comportamiento opuesto al por defecto. Dicho NP o grupo de NP no realiza procesamiento alguno	Id. de Grupo de ningún SDS	Id. de NP o Grupo de NP presente una sola vez.
Comportamiento opuesto al por defecto. Dicho NI recibe todos los SDSs procesados	Id. de Grupo de todos los SDSs.	Id. de NI presente en posición de máximo privilegio.
Comportamiento por defecto. Asignación de todos los SDSs para procesar por dicho NP	-	Un id. de NP no está presente en este campo y además no está presente ningún id de grupo al cual dicho NP pertenece
Comportamiento por defecto. Asignación de ningún SDS a enviar a dicho NI	-	Un id. de NI no está presente en este campo
Asignación del SDS o grupo de SDSs al NI de más prioridad	SDS o grupo de SDSs a través de cualquier tipo de Id.	Más de un NI presente en este campo
Asignación del SDS o grupo de SDSs al NI siguiente en prioridad	SDS o grupo de SDSs a través de cualquier tipo de Id.	Más de un NI presente en este campo y el de más prioridad falla (NPs detectan NI inalcanzable)
Descartar	Id. desconocido	Cualquier Id válido
Descartar	Cualquier Id. válido	Id desconocido

De esta forma y, para finalizar, la utilización de segmentación de cómputo gracias al uso de *SDSs* permite una gestión adecuada de los recursos gracias a la implementación de identificadores únicos de 64 bits que permiten representar inequívocamente tanto recursos como segmentos de cómputo. Además el uso de 64 bits representa una longitud de código lo suficientemente grande como para permitir un pseudo-checksum implícito,

ya que los errores de transmisión tendrán la probabilidad varios órdenes de magnitud superior de generar un identificador desconocido que transformarse en otro identificador válido.

4. Conclusiones y Trabajos Futuros

El protocolo diseñado cumple con los requisitos necesarios para el control de un flujo multimedia distribuido y permite la gestión tanto de tareas como de recursos de procesamiento. Como se indicó, el protocolo posibilita una implementación ecoeficiente, ya que ante la ausencia o daño del *NA*, el comportamiento por omisión de los *NP*s es procesamiento de todos los *SDS*s y envío a ningún *NI*, por lo cual el flujo de procesamiento se detiene por falta de destino y se evita un consumo de energía innecesario. Además permite un funcionamiento redundante resistente a fallas de *NP*s, fallas de *NAs*, fallas de *NI*s y fallas en la transmisión entre *NA* y *NP*.

El presente trabajo está orientado exclusivamente al estudio del protocolo de control de la generación multimedia distribuida para su discusión en grupos de tareas específicos de IETF. En un futuro se realizará un estudio de la asignación de grupos ya sea de *NP*s o de *SDS*s y se propondrá un protocolo para los datos de salida en el caso de no existir soluciones estándar previas.

5. Agradecimientos

Se agradece el asesoramiento y soporte de: Dept. of Bioengineering y Dept. of Biomedical and Health Information Sciences pertenecientes a University of Illinois at Chicago, Argonne National Laboratory, GridTICs de UTN-FRM, LICPaD de UTN-FRM y LIDIC de UNSL.

El trabajo es sostenido económicamente gracias al financiamiento de los proyectos 25J084 “SARA Operation” y PICT2010/29 “Procesamiento para visualización utilizando algoritmos paralelos en GPU y distribuidos en red” ambos de UTN-FRM, PROICO-30310 de UNSL y de otras fuentes de financiamiento de grupo GridTICs.

Finalmente, se agradece especialmente el aporte económico de UTN con una beca de doctorado y el excelente soporte académico de la carrera de doctorado de UNSL.

Referencias

- Cox, M. (1995) “Algorithms for Parallel Rendering”. PhD thesis, Department of Computer Science, Princeton University.
- Deering, S. and Hinden, R. (1998) “Internet Protocol, Version 6 (Ipv6) Specification” IETF STANDARDS TRACK Lo dos Santos, and Dirk Reiners (Eds.). Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 29-36. DOI=<http://dx.doi.org/10.2312/EGPGV/EGPGV07/029-036ywhere>: an open cloud gaming system. In *Proceedings of th*
- Gropp, W., Lusk, E., & Skjellum, A. (1999). “Using MPI: portable parallel programming with the message-passing interface” (Vol. 1). MIT press.
- Hubner, T. and Pajarola, R. (2007) “Single-pass multi-view volume rendering”. In: *Proceedings of International Conference Computer Graphics and Visualization*.

- IEEE Standards for Local and Metropolitan Area Networks: (1997) “Supplements CSMA/CD Access Method and Physical Layer Specifications” - 100BASE-T2, IEEE Std 802.3x-1997 and IEEE Std 802.3y-1997 (Supplement to ISO/IEC 8802-3: 1996; ANSI/IEEE Std 802.3, 1996 Edition), vol., no., pp.0_1,324, 1997”.
- IETF Datatracker: ART (2016) - “Applications and Real-Time Area” <https://datatracker.ietf.org/wg/#art>.
- Jiang, J., Hereld, M., Insley, J., Papka, M., Rizzi, S. and Uram, T. (2015) “Streaming ultra high resolution images to large tiled display at nearly interactive frame rate with v13”. In: Large Data Analysis and Visualization (LDAV), 2015 IEEE 5th Symposium on. 2015, p. 133–4. doi: 358 10.1109/LDAV.2015.7348084.
- Johnson, G., Abram, G., Westing, B., Navr’til, P. and Gaither, K..(2012), “Displaycluster: An interactive visualization environment for tiled displays” . In: Cluster Computing (CLUSTER), 2012 IEEE International Conference on. 2012, p. 239–47. doi:10.1109/CLUSTER.2012.78.
- Kirk, D. and Hwu, W. (2010) “Programming Massively Parallel Processors, A Hands on Approach” , Elsevier, Morgan Kaufmann.
- Larzon, L., Degermark, M., Pink, S., Jonsson, L., Ed., and Fairhurst, G. Ed. (2004), "The Lightweight User Datagram Protocol (UDP-Lite)", RFC 3828, DOI 10.17487/RFC3828, IETF Standard Protocol.
- Marchesin, S., Mongenet, C., and Dischler, J. (2008). Multi-gpu sort-last volume visualization. In Eurographics Symposium on Parallel Graphics and Visualization (EGPGV08).
- Molnar, S., Cox, M., Ellsworth, D., and Fuchs, H. (1994) “A Sorting Classification of Parallel Rendering.” IEEE Computer Graphics and Algorithms, pages 23-32.
- Moloney B., Ament M., Weiskopf D. and Moller T. (2011) "Sort-First Parallel Volume Rendering," in IEEE Transactions on Visualization and Computer Graphics, vol. 17, no. 8, pp. 1164-1177, Aug. 2011.
- Monfort, J. and Grossman, M. (2009) “Scaling of 3D Game Engine Workloads on Modern Multi-GPU Systems” - HPG-09.
- Perez-Monte, C., Piccoli, F., Luciano, C., Rizzi, S., Bianchini, G. and Scutari, P. (2013) “Estimation of volume rendering efficiency with GPU in a parallel distributed environment”. Procedia Computer Science 2013;18(0):1402 –11. ICCS 2013
- Perez-Monte, C., Mercado, G. , Taffernaberry, C. and Piccoli, F. (2014) - “Protocolo de comunicaciones para renderización distribuida en tiempo real” CSBC 2014 - IWPIETF LAC - 28 al 31 de julio del 2014 – Brasilia – Brasil.
- Postel, J., (1980) “User Datagram Protocol”, STD 6 , RFC 768, DOI 10.17487/RFC0768, IETF Standard Protocol.
- Schwarz, N. and Leigh, J. (2010) “Distributed Volume Rendering for Scalable High-resolution Display Arrays”, Grapp 2010

Problema de Alocação de Controladores SDN: Uma Análise do Custo em Redes de Backbone

Marcelo Santos¹, Stênio Fernandes¹

¹Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brasil

{mabs, sflf}@cin.ufpe.br

Abstract. *The SDN Placement Problem is an unsolved challenge, and it is considered an NP-Hard problem similarly to the set covering problems. Hence, a mathematical model based on Linear Programming is proposed to find a suitable controller placement solution to minimize the controller deployment cost. In order to validate the proposed model, the placement cost was analyzed through 112 backbone networks from the topology-zoo dataset. After that, the controller deployment cost was estimated and analyzed which topology characteristics affect the cost.*

Resumo. *O problema de alocação de controladores em Redes Definidas por Software (SDN) é um desafio em aberto por ser um problema de alocação NP-Completo análogo aos problemas clássicos de recobrimento de rede. Para solucionar esse problema, nós propomos um modelo matemático baseado em Programação Linear que realiza a distribuição de controladores SDN minimizando o custo de alocação de todos os controladores. Para validar nossa proposta foram estimados os custos de recobrimento de 112 redes de backbone. Em seguida foram analisadas quais características são determinantes no custo de implantação de controladores SDN de acordo com as características de cada rede.*

1. Introdução

Redes Definidas por Software (SDN) criou um paradigma relativamente recente que vem sendo discutido dentro das reuniões da IETF/IRTF em vários *work groups* (WGs) e *research groups* (RGs), com destaque para o grupo SDNRG. Em resumo, SDN consiste em remover o plano de controle de cada switch e colocá-los em uma entidade externa e centralizada (ou seja, um controlador SDN) a fim de permitir novas funcionalidades, tais como, por exemplo, manipular fluxos de rede através de uma interface programável. Consequentemente há uma visão global da rede pelo controlador possibilitando atender novas demandas tanto por parte dos usuários quanto provedores de infraestrutura. Além disso, SDN é vista como uma das tecnologias que vai servir como base para a Internet do Futuro, buscando ajudar a superar barreiras como a ossificação da Internet (Esteves, Granville, & Boutaba, 2013). Nesse contexto, juntamente com seus benéficos, emergem diversos desafios de otimização que vem sendo atacados pela indústria e pela academia, como por exemplo: (1) Problema de alocação de controladores SDN; (2) Problemas de alocação de regras de controladores SDN e (3) Problemas de alocação de Hypevisors SDN.

Decidir qual o tipo de controlador SDN, quais protocolos devem ser utilizados, a capacidade ideal de cada controlador, bem como a localização destes novos controladores são questões fundamentais que influenciam diretamente aspectos como desempenho, custo, resiliência, tempo de instalação de regras nos switches, tempo de convergência entre controladores e sua capacidade de escalabilidade. Tais questões foram discutidas pelos autores na reunião 95 da IETF/IRTF realizada em Buenos Aires na Argentina no grupo SDNRG¹. Assim, este artigo tem como uma de suas contribuições estender tal discussão na comunidade brasileira do IETF /IRTF com ênfase na análise do problema de alocação de controladores SDN.

Um controlador SDN é um ponto estratégico importante de controle na rede. Decidir onde alocar os controladores SDN, quais switches são controlados por cada controlador SDN (estimando sua capacidade) e quantos controladores são necessários para que não haja um ponto limitante são questões importantes a fim de minimizar o custo de recobrimento da rede e garantir um bom desempenho. Note que a disposição dos controladores SDN pode afetar o tempo de configuração de um novo fluxo que chega a um switch. Consequentemente, devido a restrições de tempo de comunicação mínimo entre controlador e switch, em alguns casos, torna-se impossível cobrir toda a rede com um único controlador, mesmo que um controlador possuía capacidade suficiente para controlar toda a rede. Seguindo esse raciocínio, decidimos investigar redes de *backbone* levantando questões como: (1) quais características de uma rede de *backbone* podem influenciar no custo da implantação de controladores de uma rede SDN? (2) como as restrições de tempo de comunicação podem afetar o custo e a quantidade final de controladores? (3) qual a relação entre tamanho da topologia e o custo de implantação dos controladores? e (4) qual a relação entre o atraso dos enlaces da rede e o custo de implantação dos controladores?

Nesse contexto, abordamos o problema de alocação de controladores SDN sob a perspectiva do custo financeiro de implantação considerando que existem capacidades variáveis do controlador. O primeiro passo da solução é baseado no uso de programação linear. A partir da modelagem realizada utilizamos a ferramenta CPLEX² da IBM para analisar topologias de rede reais fornecidas pelo Topology-Zoo (Knight, Nguyen, Falkner, Bowden, & Roughan, 2011) e determinar a melhor alocação de controladores para cada rede. Dessa forma, as contribuições deste artigo são três: (1) um modelo formal para o problema de alocação de controladores SDN que minimiza os custos de implantação; (2) Investigação de quais características de topologias afetam o custo de alocação de controladores SDN e (3) Discussão de propostas de drafts para o SDNRG.

Nosso modelo pode ser usado, por exemplo, quando um provedor de infraestrutura decide migrar para uma solução SDN. Nesse cenário é necessário determinar qual conjunto de switches será atribuído a cada controlador SDN (domínio do controlador), respeitando as restrições de capacidade e o tempo de configuração de fluxo (atraso entre os switches até o controlador). Portanto, na perspectiva de provedor de infraestrutura, é possível definir este problema como um problema de alocação que busca determinar o número de controladores SDN, capacidade de cada controlador, localização na rede de cada controlador e conjunto de switches controlados de forma a minimizar o custo total de implantação.

¹ <https://www.ietf.org/proceedings/95/slides/slides-95-sdnrg-2.pdf>

² <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>

2. Referencial Teórico

2.1. Redes Definidas por Software: Conceitos Básicos

SDN surgiu há alguns anos como uma alternativa para superar problemas enfrentados no ambiente de redes de computadores. A grande inovação baseia-se na dissociação do plano de controle do plano de dados. Dessa forma, switches tem seu plano de dados gerenciado por um controlador SDN (entidade que age como o plano de controle). Conseqüentemente, ao contrário das redes tradicionais, esses switches não implementam o plano de controle e as decisões de encaminhamento são decididas por fluxos encaminhados do switch para o controlador.

Um plano de controle centralizado num controlador SDN permite uma visão única e global da rede, trazendo benefícios como um conjunto de políticas menos propensas a erros, configurações de dispositivos de baixo nível mais simples e adaptações dinâmica da rede. O controlador SDN, ou eventuais aplicações executadas sobre o controlador, são responsáveis por atualizar a tabela de encaminhamento de cada roteador/switch da rede. Essa abordagem permite que novas funcionalidades sejam implementadas simplesmente pela adição de novas peças de software (aplicações) que realizam o gerenciamento da rede através do controlador SDN. Conseqüentemente, uma rede SDN tem uma camada de abstração que permite a implementação de uma variedade de algoritmos como balanceamento e classificação de tráfego com base em restrições definidas pelo operador de rede. Portanto, é comum se referir a controladores SDN como o sistema operacional de rede (Nunes *et al.*, 2014).

Note que quando um switch não possui uma regra para encaminhar determinado pacote, o switch deve encaminhar o pacote para seu controlador que baseado em alguma aplicação de gerenciamento decidirá qual o melhor caminho para rotear este fluxo novo na rede. Por esta razão, o tempo de comunicação (atraso) entre o controlador e switches pode ser um fator importante no desempenho de uma rede SDN. Atualmente, a comunicação entre controladores SDN e dispositivos de encaminhamento é realizada principalmente através do protocolo OpenFlow (Jarraya, Madi, & Debbabi, 2014). Uma boa referência para aprofundamento sobre SDN é a survey intitulada: “Software-Defined Networking: A Comprehensive Survey” (Kreutz *et al.*, 2015).

2.2. Trabalhos Relacionados

Há várias propostas da academia e indústria sobre arquiteturas distribuídas de controladores SDN bem como discussões sobre técnicas de alocação. É consenso na comunidade científica que os controladores SDN devem ser distribuídos, ao invés de se ter apenas um único controlador para toda a rede. A ideia por trás da distribuição de controladores é obter maior escalabilidade, espalhar pontos de falha tanto quanto possível e reduzir o atraso entre o controlador e switches, como discutido em (Bari *et al.*, 2013; Curtis *et al.*, 2011; Hassas Yeganeh e Ganjali, 2012; Phemius, Bouet e Leguay, 2014).

Heller, Sherwood, & McKeown (2012) investigaram o impacto da latência entre controladores SDN e switches de acordo com a posição dos controladores na rede. No entanto, este artigo não propõe nenhum algoritmo ou técnica de otimização aplicada para o mapeamento de controladores da SDN. Os autores testam exaustivamente todas as possíveis soluções, assumindo que precisam alocar um número fixo de controladores em uma topologia específica. Os resultados indicam que controladores bem distribuídos

podem minimizar a latência entre os switches e controladores. Os autores não levam em conta características importantes, tais como a capacidade do controlador, balanceamento de carga e o custo de cada controlador.

Krishnamurthy, Chandrabose, & Gember-Jacobson (2014) criam uma estratégia para atribuir switches e aplicações SDN para um determinado controlador SDN, buscando minimizar a comunicação entre controladores e o consumo de recursos. Os autores reduzem o consumo de memória RAM e de CPU em um controlador através do gerenciamento de aplicativos que são executados. Consequentemente, pode haver uma redução de custo devido ao uso de controladores de menor capacidade. Em resumo, a proposta é reduzir o consumo de recursos em cada controlador (CPU e memória RAM), além de tentar minimizar o tempo de configuração de fluxo.

Xiao *et al.* (2014) propuseram um algoritmo de alocação baseado em clusterização afim de determinar quantos controladores SDN são necessários para cobrir uma topologia e a posição de cada controlador. Neste trabalho, Xiao *et al.* não levam em consideração a heterogeneidade dos controladores com diferentes capacidades e os custos financeiros envolvidos. Além disso, a avaliação da solução proposta foi realizada em uma única topologia, que limita a validação da solução.

3. Uma solução ótima para minimização do custo de controladores SDN através do uso de Programação Linear (PL)

Nós modelamos o problema de alocação de controladores SDN usando programação linear inteira (PLI) com o objetivo de minimizar o custo final de alocação de todos os controladores de forma que todos os switches sejam cobertos por pelo menos um controlador. Programação linear permite obter soluções ótimas geralmente com um alto custo computacional quando temos um problema NP-difícil (Lange et al., 2015) com muitas variáveis envolvidas. Em problemas de cobertura de redes, como o apresentado neste trabalho, é computacionalmente dispendioso encontrar uma solução ótima diante de redes com um grande número de nós.

Em nosso modelo o número máximo de controladores SDN que podem ser alocados é igual ao número de switches da rede. Assumimos que cada controlador adicionado é obrigatoriamente ligado a um switch existente na rede. Em outras palavras, cada switch da rede pode ser visto como um potencial ponto de alocação de um controlador. A seguir temos uma descrição do modelo:

A) Notação e constantes:

N = Número de nós (switches)

M = Um número grande (Constante)

S_i = número de fluxos gerado por um switch i

A = conjunto de arcos existentes (i, j)

D_{ij} = atraso do enlace entre os nós (i, j)

K = quantidade de tipos de controladores SDN (índice k)

F_k = custo de um controlador do tipo k

C_k = capacidade de um controlador do tipo k

β = restrição de atraso entre um controlador e um switch

δ = restrição de atraso entre controladores

B) Variáveis de decisão:

$$y_{ij} = \begin{cases} 1, & \text{se um nó } i \text{ é atribuído a um controlador } j \\ 0, & \text{caso contrário} \end{cases}$$

$$x_{kj} = \begin{cases} 1, & \text{se o controlador do tipo } k \text{ é alocado em } j \\ 0, & \text{caso contrário} \end{cases}$$

C) Função Objetivo:

A função objetivo é definida como a minimização do custo de cobertura da rede por controladores SDN considerando que podem haver diferentes tipos de controladores. Conseqüentemente, cada controlador tem uma capacidade e um custo específico:

$$\text{Min} \sum_{k=1}^K \sum_{j=1}^M x_{kj} F_k \quad (1)$$

D) Restrições:

Equação 2 garante que todo switch é associado a um controlador alocado em j :

$$\sum_{j=1}^M y_{ij} = 1, \quad \forall i \quad (2)$$

Considerando que cada controlador tem uma capacidade e custo diferente, temos uma restrição que permite existir apenas um controlador alocado em j :

$$\sum_{k=1}^K x_{kj} \leq 1, \quad \forall j \quad (3)$$

Nós assumimos que um switch i gera um número de fluxos S_i . Desta forma, a capacidade de um controlador SDN deve ser igual ou maior a soma de todos os fluxos gerados pelos switches que lhe foram atribuídos:

$$\sum_{j=1}^M y_{ij} S_i \leq \sum_{k=1}^K x_{kj} C_k, \quad \forall i \quad (4)$$

Como exemplo, se todos os switches gerarem 10 mil fluxos/s ($S_i = 10K$) e um controlador tiver a capacidade para lidar com 1000 mil fluxos/s, logo esse controlador pode suportar no máximo 100 switches.

Para garantir que haja um atraso máximo entre um switch i e um controlador alocado na posição j , temos:

$$y_{ij} D_{ij} \leq \beta, \quad \forall (i, j) \in A \quad (5)$$

A equação (6) garante um atraso máximo entre quaisquer controladores da rede alocados numa posição i e j . Note que D_{ij} é uma matriz de atraso entre todos os nós da topologia e é uma das entradas da simulação.

$$\sum_{k=1}^K x_{kj} \cdot D_{ij} \leq \delta + \left(1 - \sum_{k=1}^K x_{ki}\right) M, \quad (i, j) \in A \quad (6)$$

Condições de integralidade e não negatividade:

$$x_{kj} \in \{0,1\} \quad \forall i, j \quad (7)$$

$$y_{ij} \in \{0,1\} \quad \forall i, j \quad (8)$$

4. Metodologia de Avaliação

4.1. Topologia

Foram analisadas 112 topologias do Topology-Zoo (Knight et al., 2011) aplicando o modelo exibido na seção 4. Anomalias foram removidas em algumas topologias como, por exemplo, os nós com grau zero (sem enlaces). Algumas das topologias disponíveis têm mais de uma versão, nós usamos somente as versões mais recentes para não viesar os resultados.

Através da localização geográfica de nós, estimamos o atraso entre os diferentes nós da topologia. As Equações de 9 e 10 foram usadas para calcular a distância entre os nós e o atraso.

$$D = R * \arccos \left(\frac{\sin(a) \sin(c) + \cos(a) * \cos(c) * \cos(b - d)}{\cos(a) * \cos(c) * \cos(b - d)} \right) \quad (9)$$

D é a distância em Km; R é igual a 6371 (raio médio da terra); (a, b) é a latitude e a longitude do switch i e (c, d) é a latitude e a longitude do switch j. Assim, o atraso é estimado através da seguinte equação (JIN & JIN, 2009):

$$RTT = P * D \quad (10)$$

P é uma métrica relacionada à condição de rede, segundo (JIN e JIN, 2009), P = 20us/km. Quando a latitude e longitude não estão disponíveis em alguns nós das topologias do Topology-zoo, assumimos que o atraso do enlace é de 15 ms. A matriz de caminho mais curto entre todos os nós é obtida usando a biblioteca JGraphT³ e o algoritmo de Dijkstra. Esta matriz é usada como entrada do modelo para determinar o atraso entre cada controlador SDN e os switches que compõe a rede. Para solucionar o problema de programação linear modelado foi utilizado o CPLEX da IBM (Cplex, 2007).

4.2. Custo e Capacidade do Controlador SDN

Grandes fornecedores de dispositivos de rede, como HP, Alcatel, Juniper e VMware entraram forte no mercado de SDN fornecendo switches OpenFlow, controladores SDN ou soluções completas de SDN. O custo atrelado a uma migração para uma rede SDN exige um custo inicial de investimento (CAPEX - despesas de capital) devido à grande mudança realizada na rede. Incidindo sobre o custo do controlador SDN podemos destacar duas categorias importantes: (1) hardware e software (2).

Estimar o custo de um controlador é uma tarefa difícil devido a vários atributos que podem compor esse valor. Há diversas opções de software gratuitos para

³ JGraphT - <http://jgrapht.org/>

controladores SDN tais como NOX, POX⁴, Floodlight⁵ e OpenDayLight⁶, bem como soluções proprietárias como o controlador HP VAN SDN. O problema relacionado a controladores gratuitos é a falta de suporte e atualizações. Por outro lado, uma solução proprietária como a oferecida pela HP possui custos variáveis de acordo com a quantidade de switches gerenciados pelo controlador. Ou seja, à medida que aumenta o tamanho da rede, o custo também sobe. A HP recomenda que os requisitos mínimos do sistema sejam: 16 GB de RAM; 64 GB de HD e adaptador de rede Ethernet Gigabit que acarretam custos de hardware⁷. Obviamente, o desempenho do controlador SDN é dependente da capacidade do hardware subjacente que está sendo utilizado. Assim, é razoável estimar que o controlador SDN custa alguns milhares de reais (custo de hardware e software).

O desempenho do controlador varia de acordo com a combinação de hardware e software. Por conseguinte, nós optamos por usar único número para representar todos os custos atribuídos a um controlador. Shah *et al.* (2013) mostra que dentre os controladores avaliados a capacidade de manipular fluxos varia de 1 a 4 milhões de fluxos/s. Nós assumimos que é possível optar por um controlador que possua uma menor capacidade e um menor custo ao invés de considerar todos os controladores com uma única capacidade e custo, possibilitando assim uma maior capacidade de otimização de custos. Podemos ver na Tabela 1 a capacidade do controlador representada como a quantidade de fluxos que o controlador pode manipular e o seu respectivo custo em reais. Note que tais entradas no modelo são facilmente modificadas para representar custo e capacidade de controladores atendendo a outros cenários.

4.3. Tempo de Configuração do Fluxo (*Flow setup time*) e Geração de Fluxos

Quando um switch recebe um novo fluxo e não possui uma regra de como encaminhá-lo, deve-se então requisitar ao controlador uma regra de encaminhamento adequada de acordo com o que o controlador considera mais adequado. O tempo requerido para executar essa operação é conhecido como tempo de configuração do fluxo (*flow setup time*). Um *flow setup time* de 10ms pode adicionar um atraso de 10% na comunicação em fluxos de curta duração (Tootoonchian, Gorbunov, Ganjali, Casado, & Sherwood, 2012). Sendo assim, um fator importante a ser analisado.

Segundo Benson et al. (Benson, Akella, & Maltz, 2010), um conjunto de 100 switches pode gerar picos de mais de 10 milhões de fluxos/s considerando um cenário de congestionamento da rede. Consequentemente, nós consideramos aceitável estimar que cada switch em nossa simulação gerasse uma carga de 70.000 fluxos/s sobre o controlador ao qual foi atribuído.

Pensando ainda no tempo de sincronização entre controladores e na convergência sobre as decisões de encaminhamento tomadas, definimos uma restrição de atraso máximo de 200ms (Bari, Roy, Chowdhury, & Zhang, 2013) entre todos os controladores alocados em uma mesma topologia. Assim, podemos ver nas Tabelas 1 e 2 os valores de todas as restrições assumidas em relação ao *flow setup time*, atraso entre controladores, parâmetros, fatores e níveis.

⁴ openflow.stanford.edu/display/ONL/POX+Wiki

⁵ <http://www.projectfloodlight.org/>

⁶ www.opendaylight.org

⁷ http://h20564.www2.hp.com/hpsc/doc/public/display?docId=emr_na-c03967703

4.4. Fatores, Níveis e Entradas do Modelo

Efetuamos vários experimentos de acordo com as diferentes características e parâmetros apresentados na Tabela 1 e Tabela 2. Variamos a restrição de atraso entre um switch e um controlador, além de considerar uma restrição de atraso entre controladores.

Tabela 1. Fatores e Níveis

Fatores	Unidade	Níveis	Valores
Restrição entre switch e controlador	ms	8	25, 50, 75, 100, 125, 150, 175, 200
Topologia (backbone)	-----	112	Redes do Topology-Zoo ¹
Capacidade do Controlador	Fluxos (M)	3	1, 2 or 4
Custo do Controlador	Reais (R\$)	3	10.000, 15.000 e 20.000

Tabela 2. Parâmetros

Parâmetro	Valor
Fluxos gerados por um switch j (S_j)	70K (fluxos)
Atraso dos enlaces	Definido pela equação 9 and 10
Restrição entre controladores	200ms

Em essência, as entradas são: (1) topologia composta por nós, enlaces e atraso dos enlaces e (2) Especificação de características como o atraso entre controladores e switch, capacidade que um controlador pode manipular fluxos oriundos de um conjunto de switches e taxa de geração de fluxos de um switch. Como saída de modelo, se houver pelo menos uma solução válida, temos a saída da resolução do problema de recobrimento de uma rede de *backbone* para uma rede SDN. O modelo de programação linear foi expresso em Java e resolvido pelo software CPLEX versão 12.61.

Os experimentos foram realizados em uma máquina com as seguintes configurações: Intel® Core™ i7-4500U CPU @1.8GHz; 8GB RAM; 1TB HD e Sistema Operacional Windows 10 Professional 64bits. Aproximadamente 900 experimentos foram executados.

5. Resultados e Discussão

Nesta seção, vamos apresentar e comparar os resultados da solução proposta. Como dito anteriormente, nós executamos diferentes combinações de fatores e níveis, com o objetivo de compreender o impacto no custo de implantação de uma rede SDN.

5.1. Como a restrição de atraso entre controladores SDN e switches afetam o custo de alocação?

Na Figura 1 temos as médias das topologias analisadas. As barras representam o número médio de controladores necessários para cobrir as topologias analisadas. Já a linha tracejada representa o custo médio desses controladores. O intervalo de confiança de 95% foi calculado e exibido na mesma figura.

Podemos notar que a variação da restrição de atraso entre o controlador SDN e os switches pertencentes ao seu domínio não teve um grande impacto no número de controladores SDN e conseqüentemente no custo de implantação dos controladores. Com exceção para a restrição de 25ms e 50ms que reduzem a capacidade de um controlador

alcançar um switch que esteja geograficamente mais distante mesmo que tal controlador possua capacidade suficiente para manipular mais fluxos de redes. Vemos na Figura 1 que ao considerar uma restrição de 50ms ou maior 2 controladores são, em média, suficientes para cobrir toda a rede com um custo entre 20 e 25 mil reais. Nota-se uma estabilização do custo e número de controladores SDN quando a restrição de atraso entre controlador e switch é maior ou igual a 50ms. Este comportamento se deve às características intrínsecas das topologias avaliadas que possuem baixo atraso entre seus nós. Em outras palavras, como o atraso entre os nós é baixo, a restrição de atraso não é um fator de grande impacto no custo durante o processo de alocação.

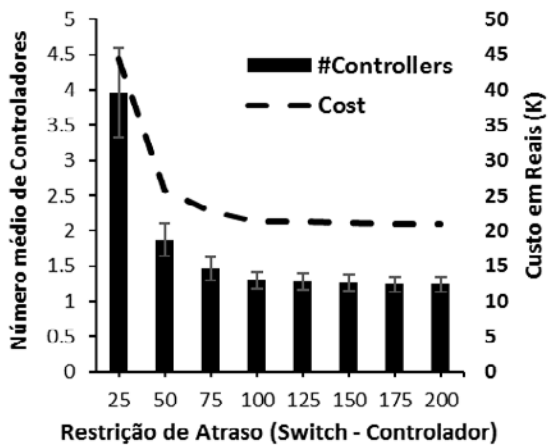


Figura 1. Custo e Número de controladores SDN

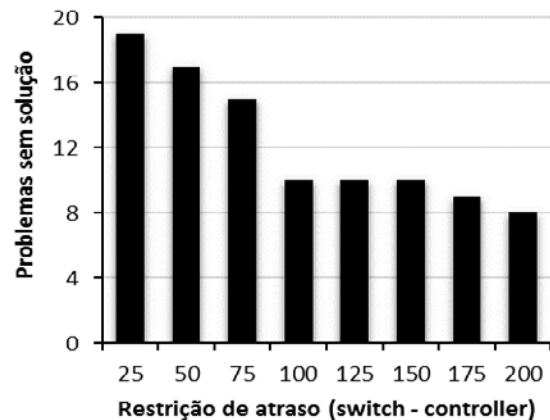


Figura 2. Casos sem solução viável encontrados

Devido às restrições para alocação dos controladores nem sempre foi possível encontrar uma solução viável. Obviamente que quanto mais rígida a restrição, o espaço de soluções viáveis tende a ser menor. Assim, podemos ver na Figura 2 que dentre as 112 topologias analisadas considerando uma restrição de 25ms entre o controlador e o switch temos 19 casos em que não conseguimos encontrar nenhuma solução viável. Quando relaxamos essa restrição para 200ms o número de casos cai para 8. Cabe ressaltar que a restrição de atraso entre todos os controladores foi sempre de 200ms em todas as simulações executadas. Dessa forma, quando temos tais restrições atuando em conjunto torna-se impossível em certos casos encontrar uma solução devido ao grande número de nós da topologia e o alto atraso entre seus nós.

É importante salientar que embora a média de controladores necessários para cobrir uma rede seja próximo de dois, não necessariamente todas as topologias são satisfeitas por 2 controladores que possuem a mesma capacidade ou até mesmo que apenas 2 controladores seja suficiente. Por isso, analisamos nas subseções abaixo a correlação entre: (1) custo x tamanho da topologia e (2) custo x atraso médio entre os nós que compõe a topologia.

5.2. Qual a correção entre o custo e atraso com a média dos enlaces que compõe uma topologia?

Uma das questões que surgem quando consideramos restrições de atraso entre controlador e switch e entre os próprios controladores são quais características da topologia podem influenciar no custo de alocação. Para esclarecer tal questão, analisamos a correlação

entre o atraso médio dos enlaces de uma topologia e o custo final de alocação dos controladores SDN em nossos experimentos. Podemos ver na Figura 3 que a correlação do custo com o atraso médio dos enlaces de uma topologia é fraca.

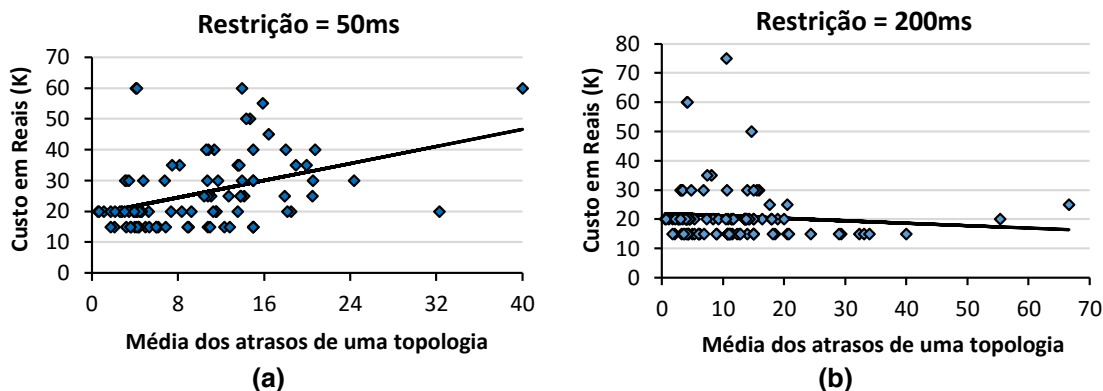


Figura 3. Correlação entre custo e atraso médio de uma topologia

Na figura 4(a) a correlação foi de apenas 0.4 contra 0.09 na Figura 4(b). No geral a correlação variou de 0.09 a 0.56 considerando as diferentes restrições entre switch e controlador consideradas. Por conseguinte, podemos afirmar que não há uma correlação forte entre custo e atraso dos enlaces das topologias analisadas.

5.3. Qual a correlação entre o custo e o número de nós de nós que compõe uma topologia?

Analisando a correlação entre o tamanho da topologia e o custo, nós estamos investigando basicamente a quantidade de controladores exigidos para cobrir uma rede de acordo com a quantidade de switches controlados por cada controlador. Em outras palavras, se um controlador não possui capacidade suficiente para manipular a quantidade de fluxos gerados pela quantidade de switches em seu domínio, será então necessário a implantação de outro controlador que acarretará um maior custo.

Podemos ver na Figura 4 há uma correlação forte entre essas métricas. Na figura 5(a) temos uma correlação de 0.68 e na Figura 5(b) uma correlação de 0.97. No geral a correlação variou de 0.38 a 0.97 considerando as diferentes restrições entre switch e controlador consideradas. Por conseguinte, podemos afirmar que há uma correlação forte entre custo e tamanho das topologias analisadas influenciada pela restrição de atraso definida entre do controlador e switch.

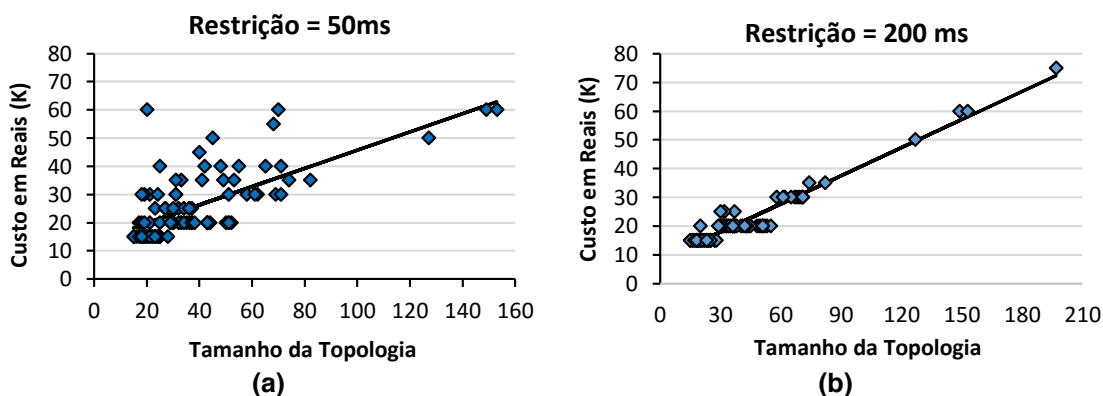


Figura 4. Correlação entre custo e tamanho da topologia

O fato da correlação ser menor quando a restrição de atraso é mais rígida (Figura 4a) entre controlador e switch deve-se ao fato do controlador não poder cobrir um número maior de switches devido ao atraso e não necessariamente pela capacidade do controlador.

6. Conclusão e Trabalhos Futuros

Neste artigo apresentamos uma solução para o problema de alocação de controladores SDN onde o objetivo é minimizar o custo financeiro de migração de uma rede tradicional para uma rede SDN. Uma solução ótima é obtida utilizando programação linear. Analisamos o modelo proposto em 112 topologias reais classificadas como redes de *backbone*. Através dos resultados obtidos realizamos uma análise de quais características influenciaram no custo de final de alocação dos controladores.

Os resultados revelam que em média dois controladores são suficientes para cobrir uma rede classificada como *backbone*. Vimos que devido as características das topologias analisadas a restrição de atraso entre controlador e switch provocou pouco impacto no custo de alocação dos controladores SDN. Analisando a correlação de algumas métricas chegamos à conclusão de que há uma fraca correlação entre o custo e a média do atraso dos enlaces de uma topologia do Topology-Zoo. Por outro lado, identificamos uma tendência de correlação entre o tamanho da topologia e o custo de alocação dos controladores, chegando em determinados cenários a atingir uma correlação de 0.97.

Possíveis drafts podem surgir do trabalho realizado como, por exemplo, um *informational draft* sobre os tipos de problema de alocação em SDN (Alocação de Controladores, Alocação de Regras e Alocação de *Hypervisors*), suas características e desafios. Outro *informational draft* que pode surgir desse trabalho diz respeito as técnicas para resolução desses problemas e suas respectivas vantagens e desvantagens. De forma geral, o trabalho desenvolvido serve como guia sobre quais características são importantes em uma topologia a ponto de afetar o custo de implantação de controladores SDN, bem como quais possíveis restrições podem acarretar um custo maior nesse processo.

Cabe ressaltar que este é um problema NP-difícil, e devido à sua complexidade, algumas topologias com muitos nós são impossíveis de serem analisadas por um método exato como programação linear devido a quantidade de tempo e recursos computacionais exigidos. Portanto, como trabalho futuro pretende-se atacar o problema através do uso de uma meta-heurística. Além disso, novas restrições deverão ser consideradas como, por exemplo, a consideração de atributos de dependabilidade.

Referências

- Bari, M., Roy, A., Chowdhury, S., & Zhang, Q. (2013). Dynamic Controller Provisioning in Software Defined Networks. International Conference on Network and Service Management.
- Benson, T., Akella, A., & Maltz, D. A. (2010). Network traffic characteristics of data centers in the wild. In Proceedings of the 10th annual conference on Internet measurement - IMC '10 (p. 267). New York, New York, USA: ACM Press.
- Esteves, R. P., Granville, L. Z., & Boutaba, R. (2013). On the management of virtual networks. IEEE Communications Magazine, 51(7), 80–88.
- Heller, B., Sherwood, R., & McKeown, N. (2012). The controller placement problem.

ACM SIGCOMM Computer Communication Review, 42(4), 473.

- Jarraya, Y., Madi, T., & Debbabi, M. (2014). A Survey and a Layered Taxonomy of Software-Defined Networking. *IEEE Communications Surveys & Tutorials*, 16(1), 1–1.
- JIN, T., & JIN, H. (2009). Network Delay Model for Overlay Network Application. *International Journal of Communications, Network and System Sciences*, 02(05), 400–406.
- Knight, S., Nguyen, H. X., Falkner, N., Bowden, R., & Roughan, M. (2011). The internet topology zoo. *Selected Areas in Communications, IEEE Journal on*, 29(9), 1765–1775.
- Kreutz, D., Ramos, F. M. V., Verissimo, P. E., Rothenberg, C. E., Azodolmolky, S., & Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. *Proceedings of the IEEE*, 103(1), 14–76.
- Krishnamurthy, A., Chandrabose, S. P., & Gember-Jacobson, A. (2014). Pratyaaatha: an efficient elastic distributed SDN control plane. In *Proceedings of the third workshop on Hot topics in software defined networking - HotSDN '14* (pp. 133–138). New York, New York, USA: ACM Press.
- Lange, S., Gebert, S., Zinner, T., Tran-Gia, P., Hock, D., Jarschel, M., ... Cho, H. (2015). Heuristic Approaches to the Controller Placement Problem in Large Scale SDN Networks. *IEEE Transactions on Network and Service Management*, 12(1), 4–17.
- Nunes, B. A. A., Mendonca, M., Nguyen, X., Obraczka, K., & Turetli, T. (2014). A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. *IEEE Communications Surveys & Tutorials*, 16(3), 1617–1634.
- Shah, S. A., Faiz, J., Farooq, M., Shafi, A., & Mehdi, S. A. (2013). An architectural evaluation of SDN controllers. In *2013 IEEE International Conference on Communications (ICC)* (pp. 3504–3508). IEEE.
- Tootoonchian, A., Gorbunov, S., Ganjali, Y., Casado, M., & Sherwood, R. (2012). On controller performance in software-defined networks. *Proceeding Hot-ICE'12 Proceedings of the 2nd USENIX Conference on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services*, 10–10. Retrieved from https://www.usenix.org/system/files/conference/hot-ice12/hotice12-final33_0.pdf
- Xiao, P., Qu, W., Qi, H., Li, Z., & Xu, Y. (2014). The SDN controller placement problem for WAN. In *2014 IEEE/CIC International Conference on Communications in China (ICCC)* (pp. 220–224).

Agradecimentos

Os autores gostariam de agradecer a FACEPE (IBPG-1321-1.03/11 para Marcelo Santos), CAPES e CNPq (206410/2014-0, 304422/2013-4, 482831/2012-0 para Stenio Fernandes). Por fim, gostaríamos de agradecer ao NIC.BR e ao CGI pelo suporte na participação das reuniões da IETF/IRTF.

Uma Proposta para a Autenticação de Estações de Trabalho em Redes Definidas por Software com Utilização de Certificados Auto-Assinados

Osiel O. Souza ¹, Jeferson C. Nobre ²

¹Instituto de Informática – Universidade do Vale do Rio dos Sinos (UNISINOS)
Caixa Postal 15.064 – 91.501-970 – São Leopoldo – RS – Brazil

osielolivera@gmail.com

Abstract. *At the same time the architecture of the Software Defined Networking - SDN has to be promising, there are some security challenges to be overcome in the implementation of such technology. The need for authentication of network components becomes a key issue in SDN due to centralization of logic controllers. An attacker could host station malicious applications and launch attacks against the control plane, making the network vulnerable. This paper proposes the implementation of a public key infrastructure combined with the use of self-signed certificates as a possible solution to the authentication problem of the origin in a SDN OpenFlow.*

Resumo. *Ao mesmo tempo em que a arquitetura das redes definidas por software (Software Defined Networking - SDN) apresenta-se promissora, existem alguns desafios de segurança a serem transpostos na implementação de tal tecnologia. A necessidade de autenticação dos componentes da rede torna-se uma questão fundamental em SDN devido a centralização lógica dos controladores. Uma estação atacante poderia hospedar aplicações maliciosas e disparar ataques contra o plano de controle, tornando a rede vulnerável. Este trabalho propõe a implementação de uma infraestrutura de chaves públicas aliada a utilização de certificados auto-assinados como uma possível solução ao problema de autenticação da origem em uma SDN OpenFlow.*

1. Introdução

A abstração de rede proporcionada pela arquitetura SDN aliada a programabilidade e visão centralizada chamaram a atenção de profissionais de TI e pesquisadores do mundo todo, movendo rapidamente o conceito para uma realidade [Jain et al. 2013]. O paradigma SDN desacopla o controle de encaminhamento dos dados e estreita a interação entre aplicações, dispositivos e serviços de rede. A divisão entre o plano de controle e plano de dados permite que a tomada de decisões antes realizada nos dispositivos, seja efetuada em outro ponto da rede através de software [Guedes et al. 2012]. Os equipamentos tais como roteadores e comutadores estabelecem o plano de dados e são responsáveis pelo encaminhamento de pacotes. A inteligência para tomada de decisões está concentrada no plano de controle, executada por uma entidade de rede chamada de controlador [Kim and Feamster 2013]. Uma das primeiras abordagens que fundamentou o paradigma SDN foi a definição do protocolo OpenFlow [OpenFlow Specification-Version 2013]. O Openflow pode ser utilizado para prover diversas funcionalidades nas infraestruturas de

rede, como por exemplo a implementação gradual de recursos e funcionalidades relacionadas a segurança.

As redes de computadores, tanto as tradicionais quanto as SDN, necessitam de mecanismos de segurança. Um dos mecanismos de segurança aplicado as rede de computadores é a autenticação de portas, a qual restringe o acesso não autorizado de dispositivos a uma rede local [Barros and Foltran Junior 2008]. As estações finais que se conectam a uma rede de computadores nem sempre serão confiáveis, podendo apresentar inúmeras vulnerabilidades. O protocolo OpenFlow, apesar de largamente utilizado em SDN, não possui um mecanismo nativo para autenticação segura da origem. Por padrão, as estações finais se autenticam a uma rede OpenFlow através da validação de seus endereços MAC (*Media Access Control*) ou IP (*Internet Protocol*), com base na lógica de comutação de pacotes definida pelo administrador da topologia. Um atacante poderia, por exemplo, obter o endereço MAC ou IP de uma estação legítima e configurar esses dados em uma estação maliciosa. Posteriormente o atacante poderia realizar uma tentativa de autenticação na rede. Esse problema ao ocorrer em uma SDN, pode comprometer a segurança da topologia, devido a centralização lógica do controlador. Um atacante poderia explorar vulnerabilidades presentes no controlador e suas aplicações, ou disparar um ataque de negação de serviço (*Denied of Service - DOS*), provocando a indisponibilidade de recursos na rede.

Existem algumas propostas para autenticação de portas em SDN. Uma delas é o mecanismo de autenticação chamado AuthFlow, que realiza autenticação das estações finais diretamente na camada enlace e associa suas credenciais a porta do computador na qual a estação está conectada [Mattos et al. 2014]. Apesar de prover uma importante contribuição para segurança SDN, o mecanismo AuthFlow realiza apenas a autenticação de portas em uma SDN OpenFlow, não garante que a origem é realmente que diz ser. Outra proposta sugere que a estação final receba um endereço IP temporário e seja direcionada para um sítio Web, onde terá as credenciais de acesso validadas [Casado et al. 2007]. Tal proposta autentica as estações na rede através de um usuário e senha conhecido pelo controlador, porém não implementa autenticação mútua ou negociação de um seguro método de criptografia entre as partes.

O presente artigo propõe a implementação de uma infraestrutura de chaves públicas aliada a utilização de certificados auto-assinados como possível solução ao problema de autenticação da origem em uma SDN OpenFlow. Para isso, um mecanismo de autenticação de portas que utilize o padrão 802.1x em conjunto com certificados digitais deverá ser desenvolvido. Esse mecanismo será integrado a um controlador OpenFlow de modo a assegurar autenticação segura para as estações finais.

A estrutura deste trabalho encontrasse no seguinte formato: na seção 2 é apresentada uma introdução ao paradigma SDN e funcionamento do protocolo OpenFlow; A seção 3 detalha os principais desafios de segurança em SDN e apresenta o problema de autenticação da origem em uma SDN OpenFlow; A seção 4 detalha a proposta de um mecanismo de autenticação e seus aspectos técnicos; A seção 5 sugere uma discussão sobre trabalhos relacionados; Por fim, a seção 6 conclui o artigo e apresenta as considerações finais.

2. O Paradigma das Redes Definidas por Software

O paradigma SDN desacopla o controle de encaminhamento dos dados e estreita a interação entre aplicações, dispositivos e serviços de rede, sejam estes reais ou virtualizados. A inteligência para tomada de decisões está concentrada no controlador de rede, que por meio de interfaces de programação, realiza a mediação entre aplicações que desejam comunicar-se com elementos da rede e elementos da rede que desejam transmitir informações na topologia [Nadeau and Gray 2013]. Uma SDN possibilita a seu administrador o desenvolvimento de uma lógica centralizada, implementação gradual de recursos e visão global da rede, consolidando as ações em um único ponto de controle. A rede torna-se independente dos fabricantes de equipamentos, pois os dispositivos são responsáveis apenas por executar as decisões previamente definidas pelo controlador [Guedes et al. 2012].

Uma SDN é caracterizada pela presença de um software de controle, o qual através de uma interface de programação supervisiona o encaminhamento de pacotes realizado pelos elementos de comutação. O software de controle recebe dos comutadores uma interface de programação que o permite inspecionar, alterar e definir entradas na tabela de roteamento dos elementos de comutação [Nadeau and Gray 2013]. O controlador possibilita ainda a utilização de um divisor de visões, permitindo que pacotes que se identifiquem com um determinado padrão, sejam associados a múltiplos comportamentos e executados em diferentes controladores de uma mesma rede física [Guedes et al. 2012]. Na Figura 1, podemos vislumbrar a arquitetura de uma SDN em suas principais camadas e interfaces de comunicação. Em uma SDN, interfaces *Southbound* são usadas para a comunicação entre o controlador e o plano de dados, possibilitando alterações na rede em tempo real. Tais interfaces podem ser de código aberto ou proprietárias e constituem uma camada de abstração para o controle e gerenciamento da topologia.

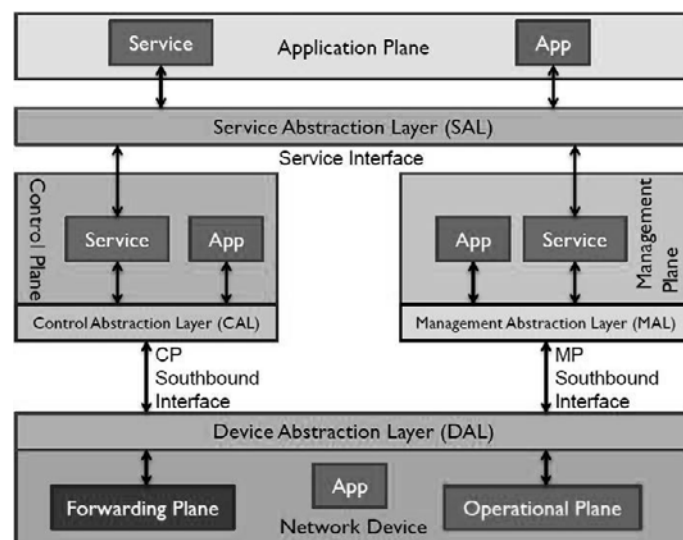


Figura 1. Arquitetura SDN Modelo em Camadas
[Haleplidis et al. 2014]

De uma forma geral, o plano de controle é responsável por estabelecer um conjunto de dados utilizado para a criação de uma tabela de encaminhamento. Essa tabela é utilizada pelo plano de dados para encaminhar o tráfego no destino correto

[Guedes et al. 2012]. O conjunto de dados construído a partir da topologia da rede é chamado de *Routing Information Base - RIB*. Muitas vezes a RIB é mantida consistente através da troca de informações entre controladores de uma mesma rede. As tabelas de encaminhamento *Forwarding Information Base - FIB* contém as informações necessárias sobre os fluxos de dados e as ações necessárias para cada um deles, como por exemplo, sair por determinada porta de um comutador. A FIB é criada contanto que o RIB seja considerado consistente e estável [Nadeau and Gray 2013].

O plano de dados é responsável pela comutação e repasse dos datagramas na rede. Tal plano opera em nível de link coletando os datagramas entrantes via inúmeros meios físicos, tais como fibra óptica, cabeado ou sem fio [Nadeau and Gray 2013]. Ao receber os datagramas, o plano de dados realiza uma verificação de integridade. Caso o datagrama seja considerado consistente, este é processado pelo plano de dados de acordo com uma consulta realizada na FIB (recebida do plano de controle). Sendo o destinatário conhecido e presente na FIB, o datagrama será encaminhado de acordo com os parâmetros estipulados para a entrega do dado. Se o destinatário for considerado desconhecido pelo plano de dados, este encaminha os datagramas para o plano de controle, que executa novamente o processo de criação da RIB e conseqüentemente a geração de uma nova FIB [Rothenberg et al. 2010].

2.1. O Protocolo OpenFlow

Um das formas de implementar interfaces Southbound é através do protocolo OpenFlow. O OpenFlow foi desenvolvido a partir de um projeto na rede de estudos da universidade de Stanford. O propósito inicial era o desenvolvimento de novos protocolos que pudessem auxiliar em pesquisas e experimentos [McKeown et al. 2008]. Para atingir o objetivo a universidade precisou criar seu ambiente de experimentação a partir do zero. Mediante a criação do núcleo da rede, pesquisadores sugeriram que o OpenFlow poderia substituir as camadas 2 e 3 do modelo TCP/IP em equipamentos de rede comerciais. Em 2011, um consórcio sem fins lucrativos chamado (*Open Networking Foundation - ONF*) foi constituído por um grupo de prestadores de serviços com objetivo de comercializar, padronizar e promover a utilização do OpenFlow em redes produtivas [Nadeau and Gray 2013].

Os elementos chave do protocolo OpenFlow tornaram-se parte da definição comum de SDN. Tais elementos são a separação do plano de controle e plano de dados. O OpenFlow estabelece um protocolo padrão, que age como mediador entre o controlador e os equipamentos da rede, por meio de uma API moderna e extensível [Rothenberg et al. 2010]. É importante salientar que o OpenFlow é um conjunto de protocolos e uma API. Não devemos categorizar o OpenFlow como um produto por si só, ou uma única característica de um produto, pois o controlador não executa nada sem uma API para determinar as regras como os pacotes devem ser comutados [Nadeau and Gray 2013].

No OpenFlow, consultas as tabelas de encaminhamento são realizadas nos próprios dispositivos de rede, garantindo o desempenho do tráfego. Decisões de como cada pacote deve ser tratado na topologia não são mais tarefas dos equipamentos de rede e sim do controlador [McKeown et al. 2008]. Interfaces programáticas tem o domínio sobre os dispositivos de rede para a implementação de novas funcionalidades e aplicação

instantânea de regras. Os comutadores recebem uma tabela de fluxo interno e possuem uma interface padronizada para adicionar ou remover entradas conforme a rede vai sendo programada [Nadeau and Gray 2013]. Atualmente o protocolo OpenFlow está dividido em duas partes:

- *Wire Protocol*: Um protocolo utilizado para o estabelecimento de sessões de controle, define uma estrutura para troca de mensagens de modificações no fluxo e realiza a coleta de dados estatísticos. Define ainda a estrutura básica do switch como, por exemplo, o conjunto de portas e tabelas nos dispositivos.
- *Of-Config*: É um protocolo de configuração e gerenciamento, seu funcionamento foi baseado no protocolo NETCONF e sua principal função é alocação de portas físicas de um switch para um controlador em especial. Além disso, é utilizado para definição de parâmetros de alta disponibilidade (ativo / passivo) e ações que o plano de dados deve tomar em caso de falha no controlador.

O protocolo Wire introduz o conceito de substituição de estado efêmero, ou seja, as entradas de fluxo não são armazenadas de forma permanente nos equipamentos. O conceito de estado efêmero dispensa o uso de dispositivos com alto poder de processamento uma vez que os equipamentos são responsáveis apenas pela comutação de pacotes [Rothenberg et al. 2010].

Em uma entrada de fluxo OpenFlow, todo o cabeçalho do pacote está disponível para modificações [McKeown et al. 2008]. Através da utilização do protocolo OpenFlow o controlador pode adicionar, atualizar e excluir entradas nas tabelas de fluxo, tanto de forma proativa como de forma reativa. Na forma proativa, o controlador popula previamente as flow tables podendo descartar fluxos não previstos. Na forma reativa, o primeiro pacote do fluxo aciona o controlador para inserir uma entrada na flow table. O modelo reativo apresenta um uso eficiente, porém gera maior fluxo de mensagens entre controlador e os switches [OpenFlow Specification-Version 2013].

3. Desafios de Segurança em SDN

Ao mesmo tempo em que a arquitetura SDN apresenta-se promissora, existem alguns desafios de segurança a serem transpostos na implementação de tal tecnologia. A centralização do plano de controle traz inúmeras vantagens, como programabilidade, lógica centralizada e visão global da rede [Feamster et al. 2014]. Tais atributos representam significativos benefícios, porém aumentam a exposição do controlador e suas aplicações a ataques de DOS e interceptação de fluxos. Em uma rede OpenFlow, todo pacote é analisado, estando um pacote com o cabeçalho não associado aos fluxos existentes, é enviado para inspeção do controlador. Caso algum comutador da topologia envie uma quantidade incomum de novos cabeçalhos de pacotes para o controlador, tal entidade poderia ter seus recursos de processamento esgotados [Kreutz et al. 2013].

Outra questão pertinente refere-se a falta de uma relação de confiança entre os elementos de rede e os controladores. Por exemplo, as aplicações que rodam no controlador podem apresentar um comportamento malicioso, o controlador deveria ser capaz de identificar ações anômalas, porém, não o faz nativamente [Kreutz et al. 2013]. Tal relação de confiança também deveria existir entre os controladores e comutadores da rede, pois uma vulnerabilidade em um comutador poderia permitir acesso indevido a dados da topologia. Não menos importante é a forma como as estações finais se autenticam em uma SDN, pois

podem hospedar códigos maliciosos destinados a exploração de vulnerabilidades presentes no controlador [Porras et al. 2012].

Existem duas linhas de pensamento sobre a segurança em redes SDN, a primeira diz que melhorias podem ser desenvolvidas explorando a capacidade de programação e visão global nativa de tal arquitetura [Porras et al. 2012]. Por exemplo, a centralização lógica do plano de controle torna possível um monitoramento de segurança altamente reativo, viabilizando a implementação de métodos para detecção de anomalias, os quais poderiam gerar dados a partir de uma visão global da topologia e enviá-los para o controlador. Com base nas informações recebidas via análise, uma política de segurança poderia ser constantemente atualizada e propagada na rede [Scott-Hayward et al. 2013]. A segunda linha de pensamento diz que os atributos de programabilidade e centralização expõe a rede a uma nova gama de ataques, em virtude das diversas possibilidades de customização. Segundo [Kreutz et al. 2013] a criação de aplicações de segurança em uma SDN torna-se um desafio uma vez que a própria segurança de uma SDN é questionável, visto que questões como a autenticação de dispositivos não possui padrão estabelecido. Sem um mecanismo seguro de autenticação, um *host* atacante poderia alcançar o controlador e explorar vulnerabilidades presentes em tal entidade [Scott-Hayward et al. 2013].

3.1. Problema de Autenticação de Portas em SDN

A necessidade de autenticação dos elementos de rede não é exclusividade de uma SDN, tão pouco um novo paradigma, redes convencionais sem separação do plano de dados e controle apresentam o mesmo tipo de problema [Porras et al. 2012]. A questão peculiar em uma SDN é o aumento da criticidade desse fato, pois estando um nó da rede comprometido, alvos como o controlador e estação de gerenciamento podem ser alcançados, tornando a rede vulnerável [Kreutz et al. 2013]. Um controlador OpenFlow deveria ser capaz de identificar aplicações maliciosas sob sua gestão, porém não executa essa tarefa por si só. Para detecção de fluxos maliciosos é necessário o desenvolvimento de um módulo de segurança no próprio controlador [Rothenberg et al. 2010]. Tal fato reforça a necessidade de um mecanismo para autenticação da origem, pois caso uma estação maliciosa conseguisse se autenticar, poderia enviar fluxos ao controlador sem ser detectada.

O protocolo OpenFlow provê programabilidade a rede, porém, não possui um mecanismo nativo para autenticação de portas [Rothenberg et al. 2010]. Por padrão, as estações finais se autenticam a uma rede OpenFlow através da validação de seus endereços MAC (*Media Access Control*) ou IP (*Internet Protocol*), com base na lógica de comutação de pacotes definida pelo administrador da topologia. Um atacante poderia obter o endereço MAC ou IP de uma estação legítima e configurar esses dados em uma estação invasora. Posteriormente, o atacante realizaria uma tentativa de autenticação na rede. Caso o acesso fosse bem sucedido, a estação maliciosa poderia explorar aplicações vulneráveis no plano de controle ou disparar um ataque de DOS contra o controlador, consumir recursos de link, memória e processamento dos comutadores, gerando indisponibilidade de serviços na topologia.

Nativamente, o OpenFlow implementa uma conexão protegida através do protocolo TLS entre os comutadores e os controladores da rede. O TLS não garante a autenticação, o TLS garante a privacidade dos dados, pois estes são criptografados [McKeown et al. 2008]. Para haver autenticação entre controlador e comutador, ou en-

tre os *hosts* e o controlador OpenFlow, os dispositivos de rede deveriam ter sua origem certificada.

4. Proposta para Autenticação de Portas em SDN

Em um ambiente de rede onde o meio físico é compartilhado ou aberto, como nas redes sem fio e redes cabeadas, a necessidade de confiança nas estações finais torna-se um aspecto fundamental na topologia [Congdon et al. 2003]. Uma forma encontrada para resolução desse problema foi o desenvolvimento de protocolos para autenticação de portas, provendo controle de acesso a uma rede computacional. Através da autenticação de portas é possível aprimorar a segurança do ambiente tecnológico e aplicar políticas de controle de acesso a todos os dispositivos que precisam acessar uma determinada rede [Barros and Foltran Junior 2008]. O padrão 802.1X é amplamente utilizado em mecanismos para autenticação de portas, tal padrão provê autenticação entre clientes de rede e os equipamentos nos quais estão conectados.

O padrão IEEE 802.1X possibilita o acesso autenticado em redes Ethernet, Token Ring e redes sem fio padrão 802.11, também oferece suporte ao protocolo RADIUS (*remote authentication dial in user service support*). O padrão IEEE 802.1X define porta como sendo um ponto de conexão à rede, podendo ser uma porta física, em redes cabeadas ou uma porta lógica, quando existe associação entre um dispositivo sem fio e o ponto de acesso [Congdon et al. 2003]. Um Servidor de autenticação RADIUS pode autenticar cada estação conectada a uma porta antes que esta possa acessar qualquer serviço oferecido pela rede. Até que o cliente esteja devidamente autenticado, o controle de acesso 802.1X habilita somente o tráfego do protocolo EAP *Extensible Authentication Protocol* na porta onde a estação estiver conectada. Caso a autenticação obtenha sucesso, o tráfego de pacotes será integralmente permitido [Barros and Foltran Junior 2008]. A utilização do protocolo EAP torna possível a independência quanto a mecanismos de autenticação limitados apenas por senha, como por exemplo, o protocolo de autenticação PPP (*Point-to-Point Protocol*). O EAP apresenta uma alternativa para interligação de redes devido a sua capacidade de adaptação a novos mecanismos de autenticação e pode, por exemplo, ser utilizado em conjunto com o protocolo TLS para implementações onde sejam utilizados certificados digitais [Congdon et al. 2003]. O EAP-TLS usa certificados padrão X.509 para verificar a identidade do usuário, aplicação ou estação de trabalho. Suporta autenticação mútua, onde o cliente deve confiar no certificado de um servidor e o servidor deve confiar no certificado do cliente. Através da utilização do EAP-TLS, também pode ser definido um algoritmo de criptografia antes da transmissão dos dados.

Um modo de prover autenticação, integridade e confidencialidade através da utilização do protocolo EAP-TLS, é a implementação de uma infraestrutura de chaves públicas. Tal infraestrutura é uma composição de segurança cujos serviços são executados e entregues utilizando conceitos e técnicas de criptografia assimétrica [Adams and Lloyd 2003]. Isso significa que para a encriptação de uma mensagem serão necessárias uma chave pública e uma chave privada. As chaves públicas não devem ser mantidas em segredo, porém devem ser protegidas contra falsificação. As chaves privadas entretanto, devem ser mantidas permanentemente em segredo [Buchmann 2002]. As infraestruturas de chaves públicas são responsáveis pela distribuição e gerenciamento de tais chaves. Podemos dizer ainda que tais infraestruturas foram desenvolvidas para autenticar e identificar usuários e serviços, garantindo que as informações trocadas estejam

disponíveis apenas as entidades autorizadas, assegurando que se uma entidade realizar uma ação, não poderá negar que a realizou [Adams and Lloyd 2003].

O presente artigo propõe o desenvolvimento de um mecanismo de autenticação para estações finais em uma SDN OpenFlow através da implementação de uma infraestrutura de chaves públicas. A Figura 2 ilustra uma proposta de arquitetura para autenticação de estações de trabalho em uma SDN OpenFlow através da utilização de certificados auto-assinados.

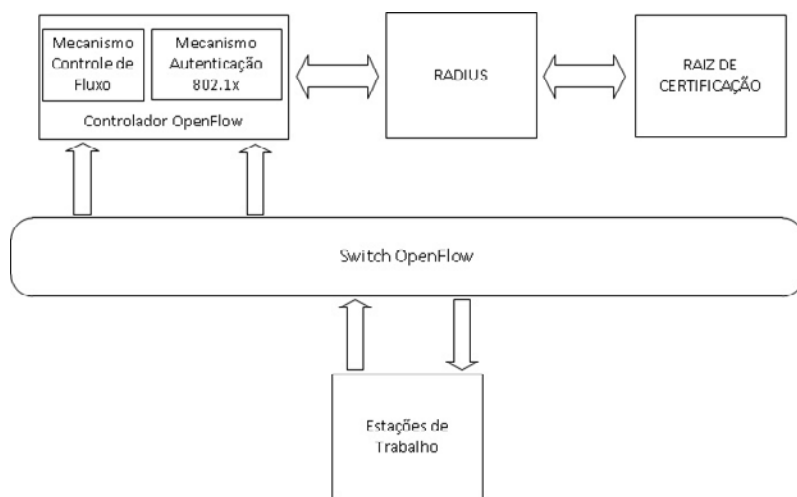


Figura 2. Arquitetura proposta

O mecanismo proposto irá adotar o padrão 802.1x para realizar a troca de mensagens com as estações de trabalho na camada de enlace. As estações deverão enviar um pacote 802.1x com o método de autenticação EAP-TLS ao controlador. O mecanismo de autenticação deverá realizar a troca de mensagens 802.1x com a estação e receber um pacote com conteúdo EAP proveniente da estação suplicante. Posteriormente o mecanismo de autenticação deverá encaminhar a requisição para o servidor RADIUS, que por sua vez, irá consultar a raiz de certificação e identificar se a estação possui um certificado válido. A raiz de certificação é responsável pela emissão, gerenciamento e revogação de certificados para a SDN. A estrutura de um certificado é basicamente um nome previamente comprovado, uma chave pública e uma assinatura digital. A assinatura é geralmente realizada pela autoridade certificadora, que atesta a autenticidade do certificado através de sua chave privada. Os certificados auto-assinados são produzidos de uma forma que é possível identificar se um atacante substituiu a chave pública, nome, ou qualquer outra informação. Um formato de certificado amplamente utilizado é o padrão X.509 v3, atualmente descrito na RFC5280 [Dang et al. 2010]. Caso as credenciais estejam corretas, a estação de trabalho recebe uma mensagem de sucesso e o processo de autenticação é concluído. Por fim, as tabelas de encaminhamento serão criadas pelo mecanismo de controle de fluxo e repassadas ao comutador que passará a tratar os pacotes oriundos dos dispositivos previamente autenticados.

5. Trabalhos Relacionados

Existem algumas propostas que buscam prover segurança a uma SDN, uma delas é o FortNOX, um mecanismo que verifica os fluxos gerados em uma rede OpenFlow que pos-

sua um controlador NOX. Tal mecanismo identifica contradições causadas por eventuais inserções de fluxos maliciosos. Mesmo relevante a segurança de uma SDN OpenFlow, o FortNOX é um mecanismo de autorização, não autenticando os elementos da rede, age após a estação estar autenticada na rede concedendo acesso com base nos fluxos dos dispositivos [Porras et al. 2012].

Certas propostas estão diretamente relacionadas a autenticação de estações finais em uma SDN, uma delas é o AuthFlow. O AuthFlow é um mecanismo de autenticação de portas e controle de acesso para redes OpenFlow. A ideia principal do AuthFlow é realizar a autenticação utilizando-se de protocolos da camada de enlace, fazendo o mapeamento da identidade usada na autenticação, em fluxos criados por uma dada estação autenticada. A troca de informações oriundas do processo de autenticação é feita por meio do padrão IEEE 802.1X e o protocolo EAP, o qual encapsula a troca de mensagens de autenticação entre a estação suplicante e um servidor de autenticação RADIUS [Mattos et al. 2014]. Apesar de oferecer um considerável avanço na segurança de uma SDN o AuthFlow não garante que a estação final é realmente quem diz ser. Um atacante com acesso físico ao comutador poderia identificar a porta na qual a estação final está conectada, copiar o endereço MAC e realizar uma tentativa de acesso à rede.

Outra proposta para autenticação de estações finais é a concessão de um endereço IP temporário e o posterior redirecionamento dos *hosts* para um sítio Web. Dessa forma, as estações que pretendem acessar a rede seriam isoladas e deveriam apresentar suas credenciais de acesso antes de serem autenticadas [Casado et al. 2007]. Tal proposta tem como requisito a presença de um navegador Web instalado nas estações finais, apresentando limitação para estações que não possuam interface gráfica instalada. Outra questão a ser analisada é o método de autenticação, que fica limitado a utilização de usuário e senha.

6. Conclusão

Pela observação dos aspectos analisados nesse artigo, entende-se que a segurança de uma SDN necessita de mecanismos eficientes para autenticação e controle de acesso. O protocolo OpenFlow fez com que o paradigma SDN evoluísse rapidamente de um conceito para uma realidade adotada por diversas empresas. Tal evolução trouxe consigo importantes benefícios como a separação eficiente entre os planos de dados e controle, porém apresentou alguns desafios de segurança a serem transpostos. A centralização lógica dos controladores e a criticidade de suas aplicações nos fazem entender quão importante é a autenticação da origem em uma arquitetura SDN. A falta de um mecanismo seguro para autenticação dos dispositivos finais poderia permitir que uma estação maliciosa disparasse ataques contra o controlador e provocasse indisponibilidade de recursos na rede. O desenvolvimento de um mecanismo de autenticação com base na utilização de certificados auto-assinados poderia garantir que uma estação final é realmente quem diz ser, dificultando o acesso de estações maliciosas na topologia.

Referências

- Adams, C. and Lloyd, S. (2003). *Understanding PKI: concepts, standards, and deployment considerations*. Addison-Wesley Professional.
- Barros, L. G. and Foltran Junior, D. C. (2008). Autenticação iee 802.1 x em redes de computadores utilizando tls e eap.

- Buchmann, J. A. (2002). *Introdução à criptografia. São Paulo: Berkeley Brasil.*
- Casado, M., Freedman, M. J., Pettit, J., Luo, J., McKeown, N., and Shenker, S. (2007). Ethane: taking control of the enterprise. In *ACM SIGCOMM Computer Communication Review*, volume 37, pages 1–12. ACM.
- Congdon, P., Aboba, B., Smith, A., Zorn, G., and Roese, J. (2003). Ieee 802.1 x remote authentication dial in user service (radius) usage guidelines. *RFC3580, September.*
- Dang, Q., Polk, T., and Brown, D. R. (2010). Internet x. 509 public key infrastructure: Additional algorithms and identifiers for dsa and ecDSA.
- Feamster, N., Rexford, J., and Zegura, E. (2014). The road to sdn: an intellectual history of programmable networks. *ACM SIGCOMM Computer Communication Review*, 44(2):87–98.
- Guedes, D., Vieira, L., Vieira, M., Rodrigues, H., and Nunes, R. (2012). Redes definidas por software: uma abordagem sistêmica para o desenvolvimento de pesquisas em redes de computadores. *Minicursos do Simpósio Brasileiro de Redes de Computadores-SBRC 2012*, 30(4):160–210.
- Haleplidis, E., Denazis, S., Pentikousis, K., Salim, J. H., Meyer, D., and Koufopavlou, O. (2014). Sdn layers and architecture terminology. *Internet Engineering Task Force, Internet Draft, Aug.*
- Jain, S., Kumar, A., Mandal, S., Ong, J., Poutievski, L., Singh, A., Venkata, S., Wanderer, J., Zhou, J., Zhu, M., et al. (2013). B4: Experience with a globally-deployed software defined wan. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pages 3–14. ACM.
- Kim, H. and Feamster, N. (2013). Improving network management with software defined networking. *Communications Magazine, IEEE*, 51(2):114–119.
- Kreutz, D., Ramos, F., and Verissimo, P. (2013). Towards secure and dependable software-defined networks. In *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*, pages 55–60. ACM.
- McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., and Turner, J. (2008). Openflow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 38(2):69–74.
- Nadeau, T. D. and Gray, K. (2013). *SDN: Software Defined Networks.* "O'Reilly Media, Inc."
- OpenFlow Specification-Version, O. S. (2013). 1.4. 0.
- Porras, P., Shin, S., Yegneswaran, V., Fong, M., Tyson, M., and Gu, G. (2012). A security enforcement kernel for openflow networks. In *Proceedings of the first workshop on Hot topics in software defined networks*, pages 121–126. ACM.
- Rothenberg, C. E., Nascimento, M. R., Salvador, M. R., and Magalhães, M. F. (2010). Openflow e redes definidas por software: um novo paradigma de controle e inovação em redes de pacotes. *Cad. CPqD Tecnologia, Campinas*, 7(1):65–76.
- Scott-Hayward, S., O'Callaghan, G., and Sezer, S. (2013). Sdn security: A survey. In *Future Networks and Services (SDN4FNS), 2013 IEEE SDN for*, pages 1–7. IEEE.

Um Modelo Formal de Propósito Geral para Simulações de Redes Definidas por Software

Rafael Souza¹, Marcelo Santos¹, Braulio Mello², Stênio Fernandes¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 181 - CEP 89802-112 – Recife – PE – Brasil

²Ciência da Computação – Universidade Federal da Fronteira Sul (UFFS)
Chapecó, Brasil. Caixa Postal 181 - CEP 89802-112 Chapecó - SC - Brasil

{rrs4, mabs, sflf}@cin.ufpe.br, braulio@ufffs.edu.br

Abstract. *Software Defined Networking (SDN) arises as a new paradigm for network management due to the clear separation between the control plane and the data plane. In this context, in order to measure which characteristics affect network performance is an arduous task considering different scenarios, several requirements, and different optimization objectives. Hence, we developed a model based on DEVS formalism that enables to execute SDN simulations quickly and in a flexible way. The proposed model makes it possible to analyze performance issues that may impact on SDN deployment. Moreover, it is also possible to assess optimization strategies that can be integrated into DEVS opening up the opportunity to validate other published papers.*

Resumo. *Redes Definidas por Software (SDN) permitem uma maior flexibilidade na gerência de fluxos de rede devido a clara separação entre o plano de controle e o plano de dados. Nesse contexto, mensurar as características que afetam, por exemplo, o desempenho da rede é uma tarefa árdua diante de diferentes cenários, requisitos e objetivos distintos de otimização. Assim, desenvolvemos um modelo baseado no formalismo de especificação de sistemas a eventos discretos (DEVS) para que seja possível simular cenários de rede SDN de forma flexível e rápida a fim de auxiliar no entendimento do funcionamento da rede e de aspectos que podem impactar no seu desempenho. Além disso, é possível ainda avaliar estratégias de otimização que podem ser integradas ao DEVS abrindo a oportunidade para validar diversos outros trabalhos.*

1. Introdução

Atualmente, o funcionamento de redes de computadores depende de milhões de dispositivos de comunicação, o que acarreta uma complexidade da rede. Consequentemente, há diversos desafios devido à dificuldade de realizar modificações no núcleo da rede ocasionando, por exemplo, o problema conhecido como ossificação da *Internet*. Diante desses problemas, bem como da necessidade de redução das despesas de capital (*Capital Expenditure* - CapEX) e de custos operacionais (*Operational Expenditure* - OpEX), diversas abordagens surgiram como uma possível solução para os desafios atuais enfrentados na *Internet*, entre as quais podemos destacar duas grandes vertentes: (1) evolucionária - *evolutionary* e (2) limpa - *clean slate* [Rexford and Dovrolis 2010]. Neste sentido, surgiram

as redes definidas por software (*Software Defined Networking* - SDN) que permitem desacoplar o plano de controle (lógico) do plano de dados, propiciando um cenário evolutivo das redes atuais.

Técnicas de Modelagem e Simulação (M&S) têm-se mostrado capazes de contribuir com a redução do esforço e custo para planos de avaliação de cenários de rede [Wainer 2009]. Todo o campo de redes de computadores passa pela análise/avaliação de desempenho para conceber novos cenários, arquiteturas, protocolos, entre outros. Assim, uma das contribuições deste trabalho é disponibilizar modelos hierárquicos e genéricos para que seja possível explorar eficientemente detalhes sobre o comportamento e estrutura de uma SDN. Para tal, é interessante utilizar mecanismo que permita acoplamento hierárquico para coexistência de modelos em representações mais simples de subsistemas. Os modelos devem considerar todas as interconexões e interações entre os mesmos, a fim de permitir um conjunto de modelos acoplados que representam o sistema complexo [Fishwick 1995]. Alguns trabalhos têm proposto técnicas de simulação para superar limites de escalabilidade [Curtis et al. 2011] [Yu et al. 2011] [Bari et al. 2013]. Em [Handigol et al. 2012] são apresentadas melhorias para o desenvolvimento e teste de novas aplicações do controlador por meio do Mininet, mas ainda tem limitações de escalabilidade, pois não pode lidar com grande quantidade de tráfego, isto porque a interface de auto-retorno (*loopback*) não é capaz de processá-lo em tempo hábil. Em [Gupta et al. 2013] foi proposto o *fsSDN*, uma ferramenta para criação de protótipos e avaliação de novas aplicações baseada em SDN. No entanto, nosso trabalho vai além de uma ferramenta específica, por permitir que o modelo possa ser genérico o suficiente para abordar diversas características de SDN, baseado na robustez e grande escalabilidade do formalismo DEVS, o qual detalhamos na seção 3.

A construção de modelos em DEVS segue uma estrutura hierárquica e as transições de eventos e o avanço de tempo são formalmente definidos. Essas características fornecem vantagens para a modelagem e simulação, tais como: (i) facilidade para experimentação, (ii) testes, (iii) manutenção, e (iv) escalabilidade. Assim, o modelo proposto permite alterações de acordo com especificação, comportamento e métricas da rede em análise. As nossas contribuições são em duas vertentes: (i) construção de um modelo genérico para avaliação de desempenho de SDN, (ii) definição de uma metodologia para planejamento e avaliação de desempenho de SDN.

Dentro da IETF/IRTF, o *Software Defined Networking Research Group* (SDNRG) investiga SDN sob várias perspectivas com o objetivo de identificar as abordagens que podem ser definidas, implementadas e utilizadas no curto prazo, além de também identificar os futuros desafios de pesquisa. Em particular, as principais áreas de interesse incluem soluções escaláveis, abstrações, linguagens de programação e paradigmas particularmente úteis no contexto da SDN, bem como identificação de futuros desafios de pesquisa na área. Assim, acreditamos que criar um modelo genérico que propicie a execução de simulações para investigar o comportamento de cenários complexos em SDN pode ser uma contribuição para o SDNRG que tem investigado, por exemplo, a performance de controladores SDN [Bhuvan et al. 2016a] [R. Gu et al. 2016] [Bhuvan et al. 2016b]. Além da comunidade da IETF/IRTF, consideramos uma contribuição para comunidade acadêmica que poderá utilizar um modelo existente para executar ou replicar algum experimento relacionado a SDN.

2. Fundamentação Teórica

2.1. Redes Definidas Por Software

Uma análise da arquitetura atual dos roteadores permite observar que se trata de um modelo formado basicamente por duas camadas bem distintas: o software de controle (*control plane*) e o hardware dedicado ao encaminhamento de pacotes (*data plane*) [Feamster et al. 2013]. SDN tem como base a dissociação dos planos de controle e de dados dos roteadores e *switches*, visando tornar os *switches* mais baratos, rápidos e flexíveis. SDN pode ser visto como um paradigma recente que promete fornecer um amplo controle sobre fluxos da rede, simplificando o gerenciamento através da utilização de um controlador externo. Desta forma, o tráfego pode ser explicitamente roteado e inspecionado por elementos de rede de controle com segurança e políticas de gerenciamento eficientes.

Desafios neste campo de pesquisa vêm do fato de que a lógica de projeto da arquitetura de uma SDN foi baseada em uma abordagem de cima para baixo (ou seja, da camada de plano de controle para os dispositivos da camada de plano de dados subjacente em uma infraestrutura), a fim de resolver problemas de eficácia ou eficiência da rede. No entanto, novos cenários de implantação de serviços SDN estão se tornando mais complexos e estão migrando para as infraestruturas distribuídas e heterogêneas, exigindo uma nova concepção lógica. Desafios de pesquisa para uma arquitetura SDN escalável podem ser abordados em vários aspectos, desde otimizações do plano de dados até a distribuição de funcionalidades do plano de controle. Como um novo paradigma para a computação em nuvem e ambientes de datacenters, mecanismos escaláveis que envolvem SDN são esperados para alavancar oportunidades de negócios potencialmente transformadoras.

2.2. Especificação Formal de Eventos Discretos

O formalismo de especificação de evento discreto (*Discrete Event System Specification* - DEVS) foi originalmente proposto por Zeigler [Zeigler et al. 2000] no final dos anos 70 para modelagem de sistemas de eventos discretos. Devido à sua capacidade de modelar sistemas complexos de maneira concisa e sem ambiguidades, o DEVS tem sido usado em diversas aplicações da engenharia (por exemplo: sistemas de manufatura, sistemas de comunicação, projeto de hardware) e ciência (por exemplo: biologia). Desde a formulação inicial de Zeigler, diversas variantes foram propostas para facilitar a modelagem e expandir as classes de sistemas que podem ser representadas pelo DEVS, tais como: *Classic DEVS System Specification*, *Parallel DEVS System Specification*, *Hierarchical Models*, *Object-Oriented Implementations of DEVS* [Zeigler et al. 2000]. Além de propostas de extensões para DEVS com o objetivo de simular sistemas com restrições temporais, tais como em [Mello and Wainer 2016].

Os modelos básicos DEVS, podem ser visualizados como caixas pretas que apenas interagem por meio de uma interface bem definida e hierárquico. O seu comportamento interno é definido por um conjunto de funções. Modelos ditos acoplados combinam os modelos básicos em uma estrutura hierárquica que define como ocorre a comunicação entre os modelos básicos. O comportamento de um sistema é descrito através das portas de entrada e saída dos modelos de seu estado e transições. O modelo mais básico, ou modelo atômico, é definido da seguinte forma [Zeigler et al. 2000]:

$$MA = (X, Y, S, \delta_{ext}, \delta_{int}, \lambda, t_a)$$

onde,

- X : é o conjunto de valores de entrada;
- Y : é o conjunto de valores de saída;
- S : é o conjunto de estados;
- $\delta_{ext}: Q \times S \rightarrow Q$ é a função de transição externa, onde $Q = (s, e) | s \in S, 0 \leq e = t_a(s)$ é o conjunto total de estados e e é o tempo decorrido desde a última transição;
- $\delta_{int}: S \rightarrow S$ é a função de transição interna;
- $\lambda: S \rightarrow Y$ é a função de saída;
- $t_a: S \rightarrow \mathbb{R}_0^+ \cup \infty$ é a função de avanço de tempo (com $\mathbb{R}_0^+ \cup \infty$) o conjunto de reais positivos com 0 e ∞ .

Um conjunto de modelos atômicos conectados, forma o modelo acoplado. Assim, o DEVS permite que os modelos acoplados possam ser usados como modelo atômico dentro de um modelo maior, construindo dessa forma, uma estrutura hierárquica. O modelo acoplado de uma forma modular é definido formalmente da seguinte maneira [Zeigler et al. 2000]:

$$MC = (X, Y, D, \{M_i\}, \{I_i\}, \{Z_{i,d}\}, \text{Select})$$

onde,

- X : é o conjunto de valores de entrada;
- Y : é o conjunto de valores de saída;
- D : é o componente de referência do conjunto, de modo que para cada $d \in D$, é um modelo DEVS;
- M_i : índice do componente do modelo acoplado;
- I_i : é o conjunto de influência do modelo i e $\forall j \in I_i$;
- $Z_{i,j}: Y_i \rightarrow X_j$ tem i para j como função de transição;
- $Select$: é o seletor *tie-break*.

A Figura 1 ilustra a estrutura de um modelo DEVS, em que podemos interpretar os elementos da seguinte forma: a qualquer momento o sistema está em algum estado "S". Nesse caso, se não acontecer nenhum evento externo, o sistema ficará no estado "S" por um período de tempo $t_a(s)$. Vale a pena ressaltar que $t_a(s)$ é um número real, mas também pode ser um número entre 0 e ∞ . Assim, na primeira situação, o sistema ficará no estado "S" para sempre, isso porque ele aguarda até que aconteça algum evento externo para que ocorra uma transição interna. Quando o tempo de *resting* finalizar, o tempo decorrido de *elapsed time* $e = t_a(s)$, o sistema vai apresentar as saídas nas portas *output*, por meio da função de $\lambda(s)$ e, com isso, altera o estado através da função interna δ_{int} . Agora, se um evento ($x \in X^b$) externo acontecer antes de finalizar o tempo, o sistema fica em um estado (s,e) com $e \leq t_a(s)$. O sistema vai executar a transição de estado, através da função externa $\delta_{ext}(s,e,x)$. Portanto, isso implica que a função de transição externa estabelece um novo estado para o sistema [Zeigler et al. 2000].

Em resumo, para construção de modelos em DEVS, primeiro se definem os modelos básicos a partir dos quais modelos maiores podem ser construídos. A partir disso, os modelos são conectados de modo a formar uma hierarquia. Teoremas, provas e detalhes adicionais sobre os formalismos podem ser obtidos em [Zeigler et al. 2000].

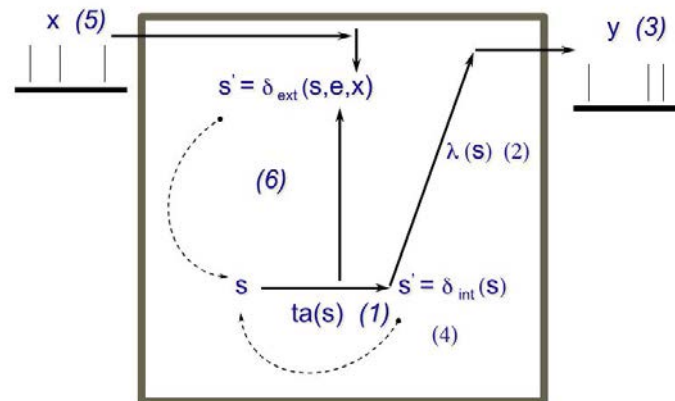


Figura 1. Estrutura de um modelo DEVS. [Zeigler et al. 2000]

3. DEVS na Simulação de Redes SDN

Este trabalho propõe um modelo genérico para representação do comportamento de redes SDN, baseado no formalismo DEVS. Existem outros formalismos na literatura que modelam e simulam sistemas de eventos discretos, como redes de Petri [Molloy 1982] [Murata 1989] e cadeias de Markov [Puterman 2014]. No entanto, esses formalismos têm o problema de explosão de espaço de estado que dificulta a modelagem dos sistemas grandes e complexos [Valmari 1998]. Além desse fator, a escolha do formalismo DEVS foi porque o mesmo traz simplicidade no desenvolvimento e diversas vantagens em representar o comportamento de redes SDN, bem como facilidade para especificação dos parâmetros do modelo de simulação de eventos discretos. Entre as vantagens, podemos citar as seguintes: fornece um *framework* formal para M&S; suporta uma completa capacidade de representação de sistema dinâmico; suporta hierárquica para o desenvolvimento modular do modelo; separa modelagem de simulação; deriva-se do formalismo genérico de sistemas dinâmicos; fornece acoplamento bem definido de componentes; suporta a construção hierárquica e reuso de modelos a partir de repositório.

Um outro ponto importante que ajudou na adoção do formalismo foi a disponibilidade de ferramentas em código aberto, como o CD++ Builder, que é um conjunto de ferramentas DEVS para modelagem e simulação que fornece uma biblioteca em C++ para especificar os modelos em vários formalismos DEVS (por exemplo, Cell DEVS, PDEVS) [Wainer 2009] [Wainer 2002]. Este trabalho adota a variante denominada Classic DEVS System Specification. As simulações podem ser realizadas localmente ou remotamente, através do envio de especificações do modelo para o servidor de simulação. O Kernel do CD++ Builder fornece um *plug-in* que permite edição de modelos, tanto textual quanto graficamente, bem como a visualização de resultados de simulação.

4. Modelos DEVS e Avaliação

4.1. Metodologia

Esta seção apresenta a metodologia concebida para avaliar o modelo em um cenário SDN simples. A execução é compreendida em 5 passos, que são: especificação do sistema, modelo atômico, modelo acoplado, construção final do modelo, e avaliação. A modelagem é feita por meio do formalismo DEVS, com uma estratégia hierárquica para representar o

modelo de desempenho genérico para arquiteturas SDN. Na figura 2, é descrito o processo realizado pela adoção do formalismo DEVS para modelagem hierárquica.

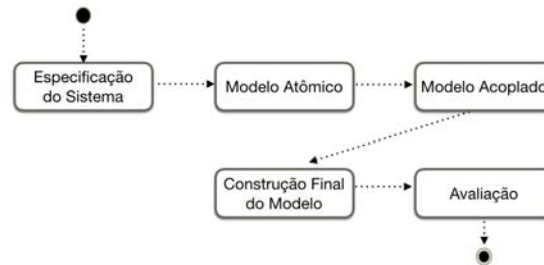


Figura 2. Metodologia para Avaliação de Redes SDN, usando o Modelo DEVS.

- Tem finalidade de compreender todos os parâmetros que são utilizados na avaliação do sistema, tais como a funcionalidade e interação entre o sistema e subsistema. Nesta fase, é importante identificar o problema e detalhes a serem analisados em um ambiente SDN;
- Modelo Atômico: Define os parâmetros do submodelo a serem analisados, a partir das informações do passo anterior são criados os modelos atômicos (submodelos) de SDN;
- Modelo Acoplado: É construída a lógica de relação e comunicação entre componentes (*switches* e controladores);
- Construção do Modelo Final: A partir deste modelo, é possível extrair as métricas desejadas. Este trabalho, adota métricas de *Flow Setup Time* (FST) e *Flow Count*;
- Avaliação: O modelo do sistema é avaliado pelas métricas e produz os resultados de interesse.

4.2. Descrição do Sistema

O cenário modelado compreende uma rede WAN, com *switches OpenFlow* [McKeown et al. 2008]. Além disso, há apenas um único controlador e consequentemente um único domínio, dado que o objetivo é apenas validar a usabilidade do modelo proposto. O controlador tem como responsabilidade a criação de caminhos entre os *switches* dentro do seu próprio domínio, onde periodicamente coleta o fluxo e estatísticas de portas a partir dos *switches*. Quando um novo fluxo chega em um *switch*, primeiro é verificado se já existe uma regra de encaminhamento para esse fluxo. Se a regra existe, então o pacotes desse fluxo são encaminhados sem haver nenhuma consulta ao controlador, seguindo, assim, a definição da regra existente. Caso contrário, o *switch* solicita ao controlador por meio de mensagem *PacketIn Openflow* que seja informado como realizar o encaminhamento do fluxo, ocasionando um atraso no encaminhamento do fluxo até que se receba a resposta do controlador. O tempo necessário para realizar esta operação é conhecido como *flow setup time*.

4.3. Formulação do Problema

Consideramos um grafo não direcionado $G(S, E)$, sendo "S" o conjunto de vértices (*switches*), e "E" as arestas que interligam os *switches*. O custo do menor caminho entre os *switches* i e j é expresso em termos de números de saltos. Assumimos que não temos um conjunto de implantação de controlador, visto que é tratado o problema apenas com

um único controlador que, por sua vez, tem a capacidade necessária para o número de *switches*.

A capacidade do controlador é definida por $\langle u_1, u_2, \dots, u_{|F|} \rangle$, bem como seu número máximo de requisições por segundo expresso em u_m . Assim, o número máximo permitindo entre o switch e o controlador é representado por d_{ij} . Também consideramos o *Flow Setup Time*, que é o custo de configuração de uma nova regra de fluxo, o qual é a soma dos saltos do switch até o controlador e da volta do controlador até o switch. O custo de FST é representado pela equação 1, que mostra o custo para o pedido de configuração da rota do fluxo.

$$C_p^R = \sum_{i \in S} \sum_{i \in S} \tau_{ii} x_{im} d_{im} \quad (1)$$

para maiores informações sobre a equação, ver [Bari et al. 2013].

5. Modelo Genérico para Representação de Redes SDN

Este é um modelo genérico que permite representar várias características de SDN, tais como taxa de chegada de fluxos, distribuições de probabilidade e quantidade de *switches*, sem alterar a estrutura do modelo e o percentual de *Flow Setup Time* que vai ocorrer na rede (por exemplo, 40% de chance de ocorre *Flow Setup Time*). Com isso, pode-se representar diferentes características do SDN. Portanto, para criar o cenário 1 no CD++, foi desenvolvido em Python um gerador de topologia, que permite uma maior flexibilidade na definição do tamanho da rede, onde podemos setar a quantidade de nós e links, bem como definir os modelos atômico e acoplado. Além disso, foi considerado o algoritmo de menor caminho entre os switch e controlador.

Algoritmo 1: Gerador de Topologia

```

1  $G \leftarrow nx.dense_{gnm_r} \text{and} om_{graph}(Node, Link)$ 
2  $isolated \leftarrow True$ 
3 initialization;
4 while  $isolated$  do
5   for  $i$  in  $G.nodes()$  : do
6     if  $(nx.is\_isolated(G, i))$  then
7        $isolated \leftarrow False$ ;
8       Break;
9      $isolated \leftarrow False$ ;
10  if  $(isolated)$  then
11     $G \leftarrow nx.dense_{gnm_r} \text{and} om_{graph}(Node, Link)$ 

```

A Figura 3 mostra um exemplo do modelo DEVS para SDN. As áreas sombreadas são destinadas apenas para auxiliar o leitor a identificar as diferentes estrutura do modelo. O modelo proposto é um modelo acoplado composto pelos modelos atômicos, que são: um nó fonte (*Generator*), três nós switch; e um nó *sink(host)*.

Cada modelo atômico tem um conjunto de portas de entrada (in_1, \dots, in_n) e saída (out_1, \dots, out_n) unidirecionais para realizar a comunicação com outros modelos

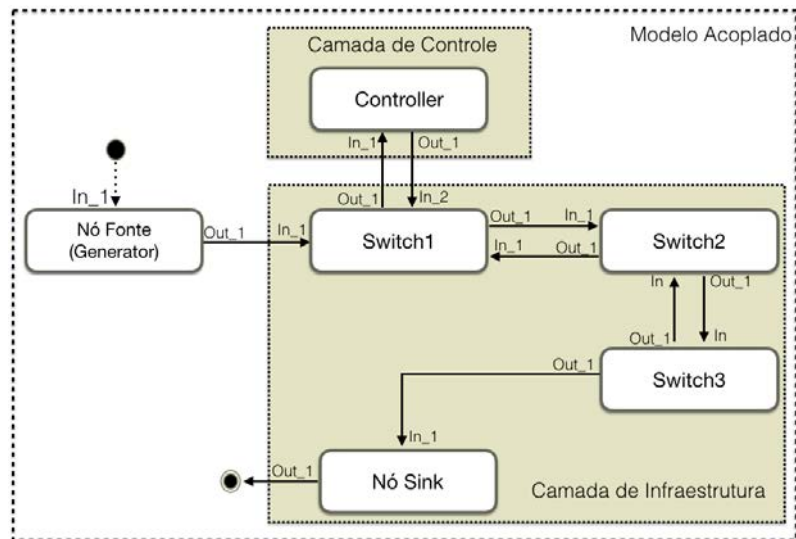


Figura 3. Topologia do Modelo DEVS de um Ambiente SDN.

atômicos. Os eventos de entrada e de saída, utilizam ambas as portas para interagir com os modelos subsequentes. Os modelos atômicos são implementados na linguagem C++ e são relacionados pelas funções herdadas da classe atômica (*InitFunction*, *externalFunction*, *InternalFunction*, *OutputFunction*), as quais são ativadas pelo método *holdIn* (*active*, *sigma*). Maiores informações sobre funções, ver [Zeigler et al. 2000].

O nó fonte (Generator) fornece o fluxo de tráfego de dados na rede através da porta de saída para o *switch*. Ele gera a uma taxa constante, mas que é controlada pelo gerador de eventos em períodos pré-determinados. O tempo entre chegadas é simulado usando taxa de distribuição de probabilidade e, para determinar o intervalo entre chegadas de pacotes na rede, foi implementado um gerador em Python para criar os períodos de entrada. O algoritmo 2 simboliza que temos 4 períodos definidos pelos números de linhas (2, 3 e 4), e as colunas representam os seguintes detalhes de configuração: horas, minuto, segundo, milissegundo, intervalo.

Algoritmo 2: Gerador de Tráfego em Períodos de Eventos Específicos

```

1 Period ← [
2 [0, 0, 0, 100, 50],
3 [0, 0, 0, 200, 50],
4 [0, 0, 0, 300, 15],
5 for i to range(len(period)) : do
6     #Finalsimulation
7     endHour ← period[i][0] ;
8     endMinuts ← period[i][1] ;
9     endSeconds ← period[i][2] ;
10    endMiliseconds ← period[i][3] ;
11    incrementMiliseconds ← period[i][4]
    File.write(format(period + str(random.randint(0, 20)))

```

O componente *switches* consiste de duas portas de entrada e duas portas de saída. O *switch*₁, nesse exemplo, é responsável pelo acesso direto ao controlador, a partir do momento em que ele obtém as regras de tráfego, as mesmas são distribuídas para os outros *switches*. Para refinar o modelo e evitar redundância de componentes (Host) na topologia da rede, foram implementadas em cada *switch* funções de probabilidade de ocorrência de solicitação de acesso à rede. A solicitação representa o momento em que o *switch* ativa a função Host, seguidamente a solicitação, considerando os parâmetros configurados, chama a função *switch* que, por sua vez, verifica se tem a rota solicitada em sua tabela de endereçamento. Se tiver, ele encaminha o pacote para o destino (Nó Sink). Caso contrário, ele calcula a distância de acesso até o controlador para obtenção da nova rota. Este cálculo caracteriza o *flow setup time*.

O componente controlador recebe o tráfego do *switch*₁ e aguarda o período de tempo definido no *preparation time*, que representa o tempo de processamento de cálculo da rota. Seguidamente, envia a mensagem para todos os *switches*, contendo informação da rota, que é entregue inicialmente pela porta *Out*₁ para porta *In*₂ do *switch*₁, que subsequentemente repassa a mensagem. Vale ressaltar que as mudanças de estado de um componente e, conseqüentemente do modelo, são dependentes da ocorrência de eventos. Pois é por meio da especificação dos eventos que o projetista modela o comportamento de um modelo de simulação em DEVS.

O componente Nó Sink (Host) tem uma porta de entrada e uma porta de saída. Ele recebe os pacotes a partir do switch 3, e verifica se tem o endereço do pacote. Seguidamente, espera um período de tempo t_a . O modelo permanece no seu estado atual. Por um período de tempo definido por *Preparation Time*. Quando o tempo da função expirar, a função de saída é invocada. A função de saída envia os eventos para porta *Out*₁.

Todos componentes têm a seguinte ordem de processamento interno: quando a função de saída é execução, a função de transição interna será chamada para determinar o novo estado do modelo. A função de período de tempo é chamada antes de cada execução do funcionamento interno e externo, uma vez que cada estado deve ser associado a um valor de tempo único. Um outro ponto importante do modelo genérico DEVS foi a implementação do algoritmo 3, o qual realiza a avaliação de *flow setup time*, considerando o *Flow Count*. O mesmo procedimento pode ser utilizado para outras análises, como novos comportamentos que venham a ser representado no modelo.

6. Avaliação de Desempenho

Para avaliar o modelo proposto, é definido um simples estudo de caso para análise de desempenho, considerando a métrica número de fluxo - *flow count* e *flow setup time* do ambiente SDN. O principal objetivo é a validação do modelo genérico DEVS. O cenário de avaliação de SDN inclui um único controlador para toda a rede.

6.1. Geração de Topologia

Construímos nossa proposta baseada na estratégia de um único controlador que é usado para toda a rede. Esta estrutura de geração de topologias foi implementada com NetworkX [Net], mas é possível obter topologia de outras fontes.

Assim, nossos experimentos são baseados na topologia ISP com 108 nós e 306 links. A topologia referente foi definida e construída por meio de um *framework* de-

Algoritmo 3: Avaliação de *Flow Setup Time* - (AFST)

```

1 time ← []
2 flow ← []
3 dx ← value
4 for i to range(value) : do
5     count ← 0
6     File ← open(flowSetupTimeData" + str(i), "rb")
7     read ← cvs.read(File)
8     for row in read : do
9         if countdivdx = 0 then
10            time ← time + [str2time(row[0])] ;
11            flow ← flow + [str2ms(row[2])];
12     count ← count + 1 ;
13 File.close( )

```

envolvido em Python, que engloba várias partes de criação do modelo. Uma dessas partes foi a criação da topologia da rede, que foi baseada na biblioteca *NetworkX* [Schult and Swart 2008] e no algoritmo de menor caminho para definir o ponto exato do controlador, para auxiliar na relação de distância (saltos) exatos entre o *switch* e controlador. Esta biblioteca foi escolhida por ser implementada em Python e de fácil criação de manipulação de grafos, além de fornecer ferramentas úteis para manipular a topologia.

A comunicação entre todos os componentes permite transferir e sincronizar o status do *switch* e status de nível de porta. O controlador faz cálculo do caminho para o novo fluxo de entrada, a partir das informações que tem sobre a rede na sua base de dados local e define o caminho até o nó destino.

6.2. Resultados

Os resultados a seguir consideram um ambiente com um controlador para toda a rede. Assim, coletamos as métricas *Flow Setup Time*. Além disso, usamos interpolação para suavizar os dados. Na interpolação, é definido o tipo como *Cubic*. A Figura 4 ilustra um controlador ativo apenas na rede. O percentual permite obter uma melhor compreensão sobre o impacto da quantidade de configurações de rotas no controlador. Para os dois cenários, a simulação foi executada em um período de 48 horas com um fluxo de tráfego variando entre (0~20(k)), mas com uma diferença por haver percentual diferente de *Flow Setup Time*. A Figura 4(a) mostra o resultado influenciado pela variação de tráfego na rede. A análise foi feita utilizando o percentual de *Flow Setup Time* igual a 50%. Já na figura 4(b) com a mesma variação de tráfego, podemos observar uma diferença na quantidade de configurações de rotas, devido ao percentual de ocorrência de *Flow Setup Time* ser 30%. Com isso, a alta variação de tráfego pode levar a ultrapassar os limites aceitáveis por um único controlador na rede.

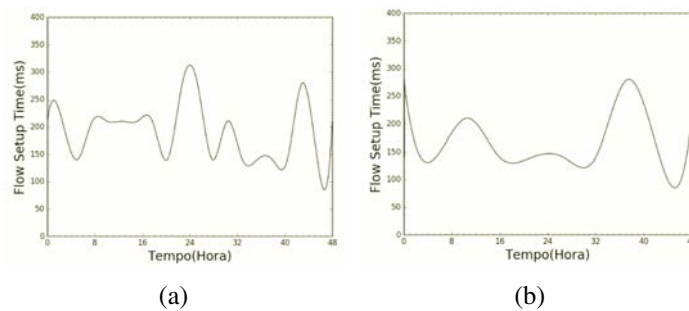


Figura 4. Flow Setup Time

7. Conclusão

Este trabalho concebeu um modelo genérico de SDN, baseado no formalismo DEVS. A partir do mesmo, é possível avaliar diversos tipos de experimentos e permite a comparação entre diferentes abordagens de gerenciamento de fluxos. Além disso, o modelo genérico proporciona uma diminuição do tempo de aprendizado para realização de testes por não necessitar do desenvolvimento, além de fornecer abstrações e ausência de protocolos e camadas de rede de baixo nível. Estes resultados mostram que as medições são precisas e escaláveis para grandes volumes de tráfego. A nossa ferramenta fornece robustez sobre a ferramenta *fs-SDN* [Gupta et al. 2013], pois também é concebido gerador de Tráfego, gerador de topologia, analisador de métricas, modelos genérico DEVS para SDN integrado ao CD++. Os experimentos são facilmente replicáveis, além disso, só é necessário utilizar uma única máquina para simular uma rede de maior escala. Um outro ponto importante, é que o trabalho é genérico o suficiente para tratar topologia com um único controlador, mas pretendemos ampliar em trabalhos futuros o desenvolvimento para múltiplos controladores, bem como adicionar meta-heurísticas.

Referências

- Networkx python package website. <https://networkx.lanl.gov/wiki>. Acessado: 2016-01-14.
- Bari, M. F., Roy, A. R., Chowdhury, S. R., Zhang, Q., Zhani, M. F., Ahmed, R., and Boutaba, R. (2013). Dynamic controller provisioning in software defined networks. In *Network and Service Management (CNSM), 2013 9th International Conference on*, pages 18–25. IEEE.
- Bhuvan, V., Basil, A., Veryx, T., Mark, T., Hewlett-Packard, Manral, V., Sec, N., Banks, S., and Monitoring, V. (2016a). Benchmarking methodology for sdn controller performance draft-ietf-bmwg-sdn-controller-benchmark-meth-01. <https://tools.ietf.org/html/draft-ietf-bmwg-sdn-controller-benchmark-meth-01>.
- Bhuvan, V., Basil, A., Veryx, T., Mark, T., Hewlett-Packard, Manral, V., Sec, N., Banks, S., and Monitoring, V. (2016b). Terminology for benchmarking sdn controller performance draft-ietf-bmwg-sdn-controller-benchmark-term-01. <https://tools.ietf.org/html/draft-ietf-bmwg-sdn-controller-benchmark-term-01>.
- Curtis, A. R., Mogul, J. C., Tourrilhes, J., Yalagandula, P., Sharma, P., and Banerjee, S. (2011). Devoflow: scaling flow management for high-performance networks. In *ACM SIGCOMM Computer Communication Review*, volume 41, pages 254–265. ACM.

- Feamster, N., Rexford, J., and Zegura, E. (2013). The road to sdn. *Queue*, 11(12):20.
- Fishwick, P. A. (1995). *Simulation Model Design and Execution: Building Digital Worlds*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition.
- Gupta, M., Sommers, J., and Barford, P. (2013). Fast, accurate simulation for sdn prototyping. In *Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking*, pages 31–36. ACM.
- Handigol, N., Heller, B., Jeyakumar, V., Lantz, B., and McKeown, N. (2012). Reproducible network experiments using container-based emulation. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 253–264. ACM.
- McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., and Turner, J. (2008). Openflow: Enabling innovation in campus networks. *SIGCOMM Comput. Commun. Rev.*, 38(2):69–74.
- Mello, B. A. and Wainer, G. A. (2016). Scheduling predictability in i-devs by schedulability analysis. Spring Simulation Multi-Conference / Symposium on Theory of Modeling and Simulation (TMS/DEVS), pages 622–629. SCS/ACM.
- Molloy, M. K. (1982). Performance analysis using stochastic petri nets. *Computers, IEEE Transactions on*, 100(9):913–917.
- Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580.
- Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.
- R. Gu, E., Li, C., and Wang, R. (2016). Problem statement of sdn and nfv co-deployment in cloud datacenters draft-gu-sdnrg-problem-statement-of-sdn-nfv-in-dc-01. <https://tools.ietf.org/html/draft-gu-sdnrg-problem-statement-of-sdn-nfv-in-dc-01>.
- Rexford, J. and Dovrolis, C. (2010). Future internet architecture: Clean-slate versus evolutionary research. *Commun. ACM*, 53(9):36–40.
- Schult, D. A. and Swart, P. (2008). Exploring network structure, dynamics, and function using networkx. In *Proceedings of the 7th Python in Science Conferences (SciPy 2008)*, volume 2008, pages 11–16.
- Valmari, A. (1998). The state explosion problem. In *Lectures on Petri nets I: Basic models*, pages 429–528. Springer.
- Wainer, G. (2002). Cd++: a toolkit to develop devs models. *Software: Practice and Experience*, 32(13):1261–1306.
- Wainer, G. A. (2009). *Discrete-event modeling and simulation: a practitioner’s approach*. CRC Press.
- Yu, M., Rexford, J., Freedman, M. J., and Wang, J. (2011). Scalable flow-based networking with difane. *ACM SIGCOMM Computer Communication Review*, 41(4):351–362.
- Zeigler, B. P., Praehofer, H., and Kim, T. G. (2000). *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. Academic press.