PUYW: Uma ontologia para o gerenciamento de perfis de aprendiz integrando YouTube e Wikipédia para apoiar um sistema de recomendação ubíqua de conteúdos

Sidney D. F. Bezerra Francisco M. M. Neto Bruno S. Monteiro
Salatiel D. Silva Patrício A. Silva Antônio V. T. Costa

Programa de Pós-Graduação em Ciência da Computação - Departamento de Computação – Universidade Federal Rural do Semi-Árido (UFERSA) Caixa Postal 59.625-900 – Mossoró – RN – Brasil

{sidneypereiro, tieldantas, avictorcosta}@gmail.com, {miltonmendes, patricio.alencar, brunomonteiro}@ufersa.edu.br

ABSTRACT

Recommendation systems have become an increasingly present element in the purpose of detecting and recommending important features to users. The challenge in such systems is to present information that is in accordance with interest of the user. For that, they make use of techniques for filtering content to associate them to users. In this sense, this article aims to demonstrate an ontology called Profiler User YouTube Wikipedia, which aims to represent and store users' information, and their digital traits relative to their preferences about videos, channels, YouTube playlists and Wikipedia texts. This ontology will support the Youubi recommendation system in the search and selection of content associated with the characteristics of the learner.[1]

KEYWORDS

Digital traits, Filtering content, Ontology, Recommendation systems.

1 INTRODUÇÃO

A Internet tornou-se um dos principais mecanismos para o acesso a informação, devido, principalmente, à interatividade que oferece através do uso de sons, imagens, vídeos e textos. A mesma ultrapassa as dimensões do espaço-tempo das aulas tradicionais, uma vez que está disponível a todo momento [5].

Além da gama de materiais multimídia e hipermídia disponíveis na Internet, é possível usar mecanismos que permitam criar ambientes virtuais para educação, que consideram a mobilidade dos estudantes e o contexto em que estão inseridos. Aplicações sensíveis ao contexto podem aperfeiçoar e amplificar as capacidades físicas e intelectuais dos aprendizes [9]. Em outras palavras, é importante o desenvolvimento de ferramentas que ofereçam elementos para auxiliar na aprendizagem do aluno, na medida em que cria ambientes adaptativos e com conteúdos apropriados para diferentes perfis de estudante [8].

O uso da internet propiciou também o crescimento exponencial de dados na Web, tornando-se necessário desenvolver tecnologias computacionais com o intuito de processar, gerenciar e armazenar conteúdos vindos da Web com objetivo de associá-los às características de seus usuários. Dentre essas tecnologias, destacam-se os Sistemas de recomendação (SRs) e as ontologias.

Os SRs operam com conjuntos de informações dos interesses dos usuários com o objetivo de encaminhar conteúdos de acordo com suas características e interesses [4]. O principal desafio em sistema de recomendação é fornecer conteúdos de forma precisa ao usuário por meio de técnicas de filtragem que adeque o conteúdo ao seu perfil [4]. As ontologias, por sua vez, podem relacionar conceitos de um domínio específico, permitindo o uso por pessoas, bases de dados e aplicações de software [13]. Dessa maneira, quando outros sistemas ou aplicações necessitarem das informações da ontologia, podem reutilizar essas abstrações, sem a necessidade de criar outras ontologias.

Diante do cenário apresentado, o objetivo desse trabalho é demonstrar uma ontologia de suporte a um sistema de aprendizado ubíquo, sendo esta ontologia capaz de associar dados do perfil estudantil de usuários à conteúdos presentes em plataformas como Youtube e Wikipédia. Deste modo o sistema pode recomendar conteúdos destas plataformas considerando o perfil de cada usuário.

O corrente trabalho está organizado da seguinte forma: A Seção 2 apresenta os trabalhos relacionados encontrados na literatura. A Seção 3 demonstra a metodologia seguida na realização da pesquisa. A Seção 4 traz em detalhes o desenvolvimento da ontologia PUYW, demonstrando os elementos utilizados na criação desta. A Seção 5 apresenta os resultados parciais obtidos na pesquisa. Por fim, a Seção 6 traz nossas considerações finais e os próximos passos a serem executados no desenvolvimento do projeto.

2 TRABALHOS RELACIONADOS

In: XIV Workshop de Trabalhos de Iniciação Científica (WTIC 2017), Gramado, Brasil. Anais do XXIII Simpósio Brasileiro de Sistemas Multimídia e Web: Workshops e Pôsteres. Porto Alegre: Sociedade Brasileira de Computação, 2017.©2017 SBC – Sociedade Brasileira de Computação. ISBN: 978-85-7669-380-2. ISBN

Em [7], é descrita uma abordagem para a construção de perfis de usuários baseada na interação com dispositivos móveis no contexto de notícias. O perfil é traçado em relação às categorias das notícias, interesses a longo prazo e o contexto no qual estão inseridas. Essa aplicação usa mecanismo de recomendação de notícias de jornais Noruegueses. O sistema de recomendação de notícias filtra a notícia e mostra ao usuário uma lista de notícias classificadas em relação à sua relevância para o usuário. As ações realizadas pelo usuário no aplicativo são armazenadas. Estes dados incluem o ID do usuário, o tempo de utilização, a localização da notícia, a categoria e palavras-chaves que descrevem o conteúdo do artigo. O contexto é construído em função do que o usuário realiza no sistema, como tópicos que o usuário selecionou ou leu.

Em [13], é descrita uma ontologia nomeada de PERSONNA que tem como objetivo modelar os perfis e os contextos de alunos em ambientes *e-learning*, para que assim possa se padronizar informações e com isso simplificar recomendações de objetos de aprendizagem.

Em [3], é apresentado um sistema de recomendação de objetos de aprendizagem para estudantes baseado em três aspectos: assunto de interesse que o aluno deseja aprender, preferências pessoais e estilo de aprendizagem.

[14] propõe um sistema de recomendação no domínio da saúde para disponibilizar vídeos ao usuário de forma confiável através da base de dados *Medline Plus*. A partir dos dados do vídeo do YouTube o sistema identifica o título do vídeo e o fatora enriquecendo o termo para filtragem do vídeo. É realizada uma inserção de filtros com o objetivo de identificar termos médicos que tenham relação com o título do vídeo, com o propósito de recomendar vídeos com informações médicas presentes no SNOMED CT.

O UbiGroup é um projeto de recomendação ubíqua de conteúdo para grupos de estudantes, que tem como objetivo recomendar objetos de aprendizagem (OAs) de acordo com o perfil do aprendiz e o contexto no qual está inserido, auxiliando o professor na busca e seleção de conteúdos educacionais [2].

Em [12], é apresentado um sistema de recomendação baseada nas interações sociais, recomendando recursos educacionais em relação ao perfil e ao contexto do aprendiz. O contexto e o perfil são extraídos de sistemas colaborativos (redes sociais e ambientes virtuais de aprendizagem) automaticamente. São utilizados Dados Ligados (ou *Linked Data*) como complemento dos conteúdos educacionais (Objetos de Aprendizagem), de modo que o seu uso em sistema de recomendação amplia as chances de encontrar conteúdos em diferentes fontes. Neste trabalho foram utilizadas a DBPEDIA e a *Open University*.

3 METODOLOGIA

Para alcançar o objetivo de desenvolver um sistema de recomendação ubíquo integrando hipermídia baseada em ontologia, esta seção apresenta o método de pesquisa adotado, que é composto pelas seguintes atividades: revisão de literatura, formalização e prototipação.

Revisão de literatura: esta atividade teve como objetivo investigar os conceitos e projetos correlatos referentes a sistemas de recomendação semânticos de apoio ao aprendizado e os fundamentos das seguintes tecnologias: técnicas de recomendação de conteúdo, modelo de representação de aprendiz e ontologias. A revisão de literatura referente à modelagem do perfil do aprendiz seguiu os principais modelos computacionais de representação do estudante. Para as técnicas de recomendação de conteúdo, foram analisados algoritmos e práticas de assimilação entre conteúdo e usuário. Sobre as ontologias, foram analisados projetos correlatos de representação de um aprendiz e como os conteúdos são modelados.

Formalização: uma ontologia é um modelo formal em que a corretude e a consistência podem ser verificadas automaticamente. Esta atividade teve como objetivo modelar o perfil do aprendiz, os metadados das APIs do Youtube [6] e da MediaWiki [10] em ontologias e utilizá-los como recursos para aplicações. A ontologia do perfil do aprendiz é responsável pela manipulação das características do estudante, tais como: identificação, preferências, dados geográficos, entre outras. A ontologia dos metadados da API do YouTube e mediawiki, apresenta informações sobre os vídeos, tais como: título, quantidade de *likes*, duração, quantidade de visualizações, entre outras, e sobre a enciclopédia Wikipédia, informações como título, descrição, URL, entre outros.

Prototipação: esta atividade teve como objetivo conceber um modelo conceitual da aplicação e analisar as diferentes técnicas de prototipação, com o objetivo de verificar suas propriedades ao contexto ao qual será inserido e demostrar que é possível construir o sistema proposto.

4 ONTOLOGIA PROFILE USER YOUTUBE WIKIPEDIA (PUYW)

Para definir o perfil do aprendiz, foram utilizadas informações acerca dos dados pessoais, relacionamentos, preferencias, entre outros. Tornou-se necessário utilizar um modelo ontológico para representar tais informações.

Parte da ontologia desse trabalho é baseada na proposta de [13] que realizou uma análise das especificações e padronizações dos dados sobre contexto, perfil do aluno, e trabalhos que utilizaram estes padrões no desenvolvimento de ontologias para definição do perfil do aprendiz.

As classes, atributos e propriedades foram selecionados de acordo com a proposta deste trabalho e com isso algumas classes foram desconsideras, como também foram adicionadas novas classes para alcançar o objetivo de formar um perfil de aprendiz com as características dos conteúdos e das hipermídias utilizadas.

4.1 Perfil do Aprendiz

O Perfil do aprendiz contém características distintas sobre suas preferencias, interesses, relacionamento, entre outras. No entanto, estudantes possuem tais características de formas diferentes. Um dos desafios dos sistemas na manipulação dos dados é gerenciar de forma correta essas informações. A Fig. 1 exibe parte da estrutura da ontologia que demonstra as relações entre as classes que caracterizam o perfil do usuário.

Os dados pessoais do aprendiz são representados pela Classe *LearnerPersonalInformation*. Por meio de atributos de

dados (*data properties*) os indivíduos desta classe dispõem das seguintes informações: 1) *name*: nome do aprendiz; 2) *age*: idade do aprendiz; 3) *sex*: gênero do aprendiz; 4) *cellphone*: telefone do aprendiz; 5) *country*: país do aprendiz; 6) *email*: endereço eletrônico do aprendiz; 7) *occupation*: ocupação ou profissão.

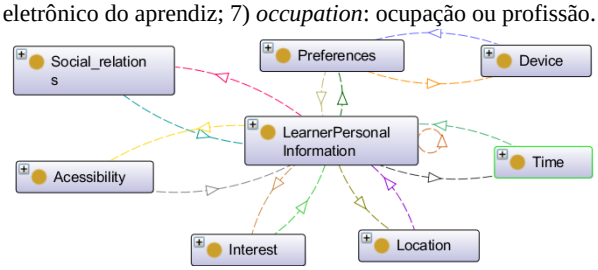

Figure 1: Parte da Estrutura da PUYW que contém a descrição dos elementos do Perfil do Aprendiz.

Os relacionamentos entre aprendizes estão armazenados pela classe *Social_relations*, que contém as seguintes informações: 1) *socialIdentifier*: identificador dos aprendizes que apresentam algum relacionamento com o aprendiz do perfil; 2) *relationToMe*: como os alunos desse perfil estão relacionados; 3) *relationToThem*: como outros alunos estão associados a esse perfil; 4) *relationsLabel*: detalha o tipo de relacionamento (amigo, inimigo, amigo do trabalho, entre outros).

A classe *Time* está relacionada com as datas e o momento da realização de um evento do aprendiz, suas informações armazenadas são: 1) *dateNow*: data atual; 2) *timeNow*: tempo atual.

As informações relacionadas à localização geográfica do aprendiz estão apresentadas na classe *Location*, são elas: 1) *locationName*: nome do local; 2) *GPS*: coordenados capturadas pelo *GPS* (*Global Positioning System*); 3) *speed*: velocidade do aprendiz no instante da captura das coordenadas.

Os interesses do aprendiz em determinados conteúdos são representados e armazenados na classe *Interest*, que detém informações das buscas e temas relevantes do aprendiz, são elas: 1) *typeName*: tipo de interesse (profissional, acadêmico, hobbies, etc.); 2) *tag*: relacionado aos conteúdos relevantes do aprendiz.

Para facilitar a aproximação e aquisição de usuário neste ambiente, estão presentes mecanismos que gerenciam tais informações de acessibilidade. A classe *Acessibility* contém o gerenciamento dos dados acerca de: 1) *acessibilityComment*: comentário sobre leitura, escrita e fala do aluno; 2) *acessibilityContentType*: tipo de conteúdo de acessibilidade; 3) *disabilities*: nome da inaptidão.

A classe *Preferences* compreende as informações das preferências do aprendiz em relação às mídias do conteúdo. Esta classe contém as seguintes informações: 1) *language*: idioma do conteúdo; 2) *media*: lista de mídias como vídeo, texto, áudio, entre outras; 3) *order*: ordem da preferencias da mídia.

A classe *Device* armazena os dados a respeito dos dispositivos utilizados pelos estudantes. Esta classe contém as seguintes informações: 1) idPerson: número de identificação do aprendiz; 2) *bandWidth*: largura de banda (alta, baixa); 3) *Devices*: lista dos dispositivos usados pelo aprendiz (tablet, celular, PC, notebook, TV digital); 4) *networkOperatorNow*: operadora utilizada no dispositivo no momento; 5) *productModel*: modelo do dispositivo; 6) *Resolution*: dimensões do tamanho da tela do aparelho; 7) *soVersion* : versão do sistema operacional utilizado pelo dispositivo; 8) *type* : tipo do dispositivo (móvel/ fixo).

4.2 YouTube e Wikipédia

Para que o sistema de recomendação apresentasse suporte a vídeos e textos, desenvolveu-se na ontologia PUYW classes e propriedades que descrevem os aspectos relativos aos vídeos do YouTube e os conteúdos da Wikipédia. Para o YouTube levou-se em conta os elementos disponibilizados na API do mesmo. Já os elementos da Wikipedia são disponibilizados pela MediaWiki. A API do YouTube fornece diversos metadados acerca dos vídeos, tais como: título, categoria, número de visualizações, número de *likes*, *deslikes*, entre outros. De forma semelhante, os textos do Wikipédia dispõem de metadados que possibilitam recuperar informações como título, descrição, entre outros. Cada um desses elementos foi convertido para a ontologia formando assim um conjunto de classes e relações que descrevem os vídeos e textos. A Fig. 2 apresenta parte da estrutura da ontologia que demonstra as relações entre as classes que representam o usuário e as classes que representam a estrutura do YouTube e da Wikipédia.

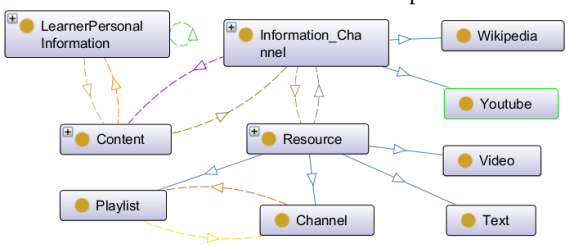

Figure 2: Parte da Estrutura da PUYW que contém a descrição dos elementos do YouTube e Wikipedia.

Como pode ser visto na Fig. 2, existem quatro classes essenciais para representar e manter informações dos textos e vídeos, sendo estas: *Playlist*, *Video*, *Channel* e *Text*. As três primeiras fazem jus à estrutura de apresentação dos conteúdos do YouTube, ou seja, nesta ferramenta o usuário pode visualizar um vídeo que pertence a um determinado canal ou *playlist*. Estas relações são mantidas na ontologia por meio de *Object Propertiers*. Uma destas é a *isVideoOf* que possibilita o armazenamento da relação "*video isVideoOf Channel or Playlist*". Em outros termos, é possível associar para qualquer vídeo, qual canal e/ou *playlist* este pertence. Esta propriedade possui ainda sua forma inversa *hasVideo* que permite associar quais vídeos pertencem a um canal ou *playlist*. De forma semelhante, é possível arquivar a informação sobre qual canal uma playlist faz parte, por meio das propriedades *isPlaylistOf* e *hasPlaylist*.

Existem também as *Datatype Properties*, que associam os indivíduos da ontologia a dados obtidos tanto da API do YouTube

como da Wikipédia. Há, por exemplo, as propriedades *hasTitle*, *hasLikeCount*, *hasViewCount*, *hasDescription*, entre outras. A primeira delas relaciona indivíduos da classe *Text*, *Video* e *Playlist* à informação de título do conteúdo. Já a segunda permite armazenas as informações de quantidade de *likes* registrados que indivíduos da classe *Video* e *Text* contém.

De acordo com a Fig. 2, é possível notar que a classe *LearnerPersonalInformation* detém relações com a classe *Content* (que formaliza um conteúdo escolar). Estas relações se dão por meio das propriedades *Access* e *isAccessed* que possibilitam relacionar um usuário a um conteúdo e vice-versa. Há também relações entre a classe *Content* e a classe *Information_Channel* (que representa um canal de informação por onde os conteúdos podem ser transmitidos). Para isso foi-se desenvolvido as propriedades *isTransmitted* e *Transmitts* dando à ontologia a capacidade de associação os conteúdos acessados ao canal de informação utilizado (YouTube e/ou Wikipedia). A classe *Information_Channel* mantém relacionamentos com a classe *Resource* (que representa os recursos que um canal pode ter, tais como vídeo, texto, canais, entre outros).

Consequentemente, por meio desta representação de estrutura, é possível armazenar, para os usuários da aplicação, quais conteúdos estes acessaram, por quais canais de informações estes conteúdos foram transmitidos, que recursos foram utilizados e os dados associados a cada recurso (quantidade de visualizações, *likes*, entre outros). Deste modo, os algoritmos de recomendação dispõem de diversas informações acerca do perfil de acesso ao YouTube e Wikipédia para quaisquer usuários, podendo, através destes traços, recomendar conteúdos para os mesmos a partir das informações dos demais usuários.

5 RESULTADOS PARCIAIS

Até o momento foi concebida uma arquitetura composta por três webservices: Comunicador, Mediador e Buscador. O Comunicador é responsável por realizar a comunicação com o ambiente de aprendizagem ubíqua Youubi [11] e fornecer dados relevantes sobre o usuário. Esses dados serão tratados na ontologia de perfil de aprendiz que está inserida neste webservice. Este módulo ainda tem o objetivo de armazenar informações dos YouTube e Wikipédia que os usuários do Youubi manipularam. O Mediador é responsável por tratar as informações recebidas do Comunicador e manipular os conteúdos por meio de técnicas de recomendação, para que os recursos sejam retornados para o usuário. Este webservice tem como objetivo rastrear conteúdos no Youtube e na Wikipédia e retornar tais conteúdos para o Comunicador, levando em conta as características do aprendiz. O Buscador o objetivo de capturar e selecionar vídeos e textos nas bases de dados do YouTube e Wikipédia, de acordo com as características do usuário. Na realização dessa tarefa, este *webservice* se comunica com as APIs do YouTube [6] e da MediaWiki [10] recuperando os metadados e armazenando-os na ontologia PUYW.

Para realizar a manipulação dos dados e da ontologia utilizou-se a API Jena [1]. A mesma permite manipular a ontologia PUYW sendo possível modificar sua estrutura quando necessário, realizar inserção dos dados advindos dos canais de comunicação e armazenar as relações existentes entre os dados. A partir desta API é possível também recuperar as informações armazenadas via consultas SPARQL ou por varreduras na ontologia.

Com a finalização desses três componentes, o sistema de recomendação possui mecanismos apropriados para executar os algoritmos de recomendação.

6 CONCLUSÃO

Este artigo tem como objetivo apresentar uma ontologia de suporte a um sistema de aprendizado ubíquo. Esta ontologia possui em sua estrutura elementos capazes de representar e armazenar os traços digitais e o perfil de acesso de usuários a conteúdos do YouTube e Wikipédia. Estas informações dão suporte direto ao sistema de recomendação do Youubi, para que esta ferramenta disponibilize conteúdos aos usuários de maneira mais adequada, fornecendo um serviço mais eficaz.

Os próximos passos dessa pesquisa consistem em implementar os algoritmos para o sistema de recomendação, utilizando mecanismos de filtragem por conteúdo, colaboração e híbrida.

REFERÊNCIAS

[1] Apache Software Foundation. 2017. Apache Jena A free and open source Java framework for building Semantic Web and Linked Data applications. Disponível em: https://jena.apache.org/ Acesso em: 02 de Agosto de 2017

[2] Araujo F., Luis Gustavo et al. 2015 UbiGroup: Um Modelo de Recomendação Ubíqua de Conteúdo para Grupos Dinâmicos de Aprendizes. Revista Brasileira de Informática na Educação, v. 23, n. 3, 2015.

[3] Borges, Grace; STIUBIENER, Itana. 2014. Modelo De Recomendação Com Base No Estilo De Aprendizagem Do Aluno. Em: 2014 - XI Congresso Brasileiro de Ensino Superior a Distância (ESUD), 2014. p. 214-227

[4] Costa, A. A. L. 2015. "Mecanismo de Recomendação Personalizada de Conteúdos para apoiar um Ambiente de Aprendizagem Informal no contexto da Saúde." 2015. 92 f. Dissertação (Mestrado) - Curso de Pós-graduação em Ciência da Computação, Universidade Federal Rural do Semi-Árido e Universidade do Estado do Rio Grande do Norte, Mossoró.

[5] DA SILVA, Edna M. Oliveira. 2016 A Webquest na Internet: o novo material didático. Revista da FAE, v. 11, n. 2, 2016.

[6] Google Developers. 2017. Youtube Data API (V3). Disponível em: https://developers.google.com/youtube/v3/docs/ Acesso em: 02 de Agosto de 2017

[7] Gulla, J. Atle et al. 2014. Implicit User Profiling in News Recommender Systems. In: WEBIST (1). 2014. p. 185-192.

[8] Levis, Darci; BARBOSA, Jorge LV; PINTO, S. Crespo CS. 2008. Aperfeiçoamento automático do perfil do aprendiz em ambientes de educação ubíqua. Brazilian Journal of Computers in Education, v. 16, n. 01, 2008

[9] Marinho, Carlos SS; JUNIOR, Messias; VIANA, Windson. 2016. Notificação sensível ao contexto: uma análise de aceitação. In: Workshop de Trabalhos em Iniciação Científica (WTIC), 13., 2016, Teresina. Anais do XXII Simpósio Brasileiro de Sistemas Multimídia e Web. Porto Alegre: Sociedade Brasileira de Computação, 2016. v. 2

[10] MediaWiki. 2017. MediaWiki action API documentation Disponível em: https://www.mediawiki.org/wiki/API:Main_page/pt-br. Acesso em: 02 de Agosto de 2017

[11] Monteiro, B. de Sousa et al. 2015 "Youubi: Ambiente de Aprendizagem Ubíqua." In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). p. 111.

[12] Pereira, C. Kelle et al. 2015. Explorando Dados Ligados através de um Sistema de Recomendação Educacional. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2015. p. 1042.

[13] Rezende, P. A. Almeida et al. 2015. "PERSONNA: proposta de ontologia de contexto e perfil de alunos para recomendação de objetos de aprendizagem". Revista Brasileira de Informática na Educação, v. 23, n. 01, p. 70, 2015.

[14] Rivero R., A. et al. 2013. A health information recommender system: Enriching YouTube health videos with Medline Plus information by the use of SnomedCT terms. In: Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on. IEEE, 2013. p. 257-261.