Uma Abordagem para Orquestração do Conhecimento como Suporte ao Planejamento Curricular em Ciência da Computação

Anderson Felinto Barbosa[1,2], **Janderson Jason Barbosa Aguiar**[1], **Ulrich Schiel**[1]

[1]Universidade Federal de Campina Grande (UFCG)
Campina Grande – PB – Brasil

[2]Instituto Federal da Paraíba (IFPB) – Campus Monteiro
Monteiro – PB – Brasil

{anderson.fbarbosa, janderson.jason}@gmail.com,

ulrich@computacao.ufcg.edu.br

Abstract. *This paper presents an approach for Orchestration of Curricular Knowledge in Computer Science based on the subjects (Knowledge Categories) of the Reference Curriculum for Computer Science and curricular structures of undergraduate courses. For this, semiautomatically, the disciplines of the structures are categorized in on the of the Categories of Knowledge, with this it is possible to measure the representativeness of each Category and to identify the possible relations of dependencies between them. A case study was carried out with 457 subjects, making it possible to identify, in the context of the 7 bachelor's courses in Computer Science, the representativeness of the Knowledge Categories and, in addition, the dependency relations between them.*

Resumo. *Este artigo apresenta uma abordagem para Orquestração do Conhecimento Curricular em Ciência da Computação baseado nas matérias (Categorias de Conhecimento) do Currículo Referência para a Ciência da Computação e em estruturas curriculares de cursos de graduação. Para isso, de forma semiautomática, as disciplinas das estruturas curriculares são categorizadas em uma das Categorias de Conhecimento, com isso é possível mensurar a representatividade de cada Categoria e identificar as possíveis relações de dependências entre elas. Um estudo de caso foi realizado com 457 disciplinas, possibilitando identificar, no contexto dos 7 cursos de bacharelado em Ciência da Computação utilizados, a representatividade das Categorias de Conhecimento e, além disso, as relações de dependência entre elas.*

1. Introdução

Na literatura, o termo *Design* Instrucional (DI) está associado ao processo de planejamento de um conjunto de métodos, técnicas e atividades que serão entregues aos alunos durante o processo de aprendizagem [Filatro 2008]. Com diferentes níveis de granularidade (Macro, Meso e Micro), o *Design* Instrucional pode ser utilizado em diferentes contextos inerentes ao processo de planejamento educacional [Barbosa et al. 2015].

O DI na granularidade meso, por exemplo,pode ser utilizado para estruturar cursos e, consequentemente, definir qual será o conjunto de disciplinas necessárias para a

formação do perfil do aluno esperado pelas instituições de ensino, além dos objetivos que devem ser alcançados pelas disciplinas, seus conteúdos e as relações de dependências (pré-requisitos) entre elas.

Para algumas áreas de conhecimento são criados documentos para auxiliar o processo de definição das disciplinas, conteúdos e objetivos educacionais. Com isso, o processo de DI que utiliza esses documentos está alinhado aos parâmetros estabelecidos por comitês, sociedades e/ou outras entidades responsáveis ou que colaboram para a educação. No Brasil, a Ciência da Computação (CC) é uma área que possui um conjunto de documentos com esta finalidade, seja no contexto nacional, por meio dos documentos criados pela Sociedade Brasileira de Computação (SBC), ou internacional, criados pela *Association for Computing Machinery* (ACM).

O documento das recomendações curriculares para a Ciência da Computação descreve de forma clara e detalhada o conjunto de conteúdos necessários para a formação dos alunos. No contexto internacional, o *Computer Science Curricula 2013* (CS2013) [Joint Task Force on Computing Curricula e Society 2013] destaca os conteúdos organizados em três diferentes granularidades, os objetivos e relações entre os conteúdos. No Brasil, a SBC disponibiliza o Currículo Referência para cursos de Ciência da Computação e Engenharia da Computação (CR05) [SBC 2005] contendo um conjunto de matérias que podem ser abordados nos cursos de graduação do país. Contudo, as relações entre os diferentdes conteúdos (matérias) não são destacadas, deixando tal definição a critério das Instituições de Ensino Superior (IES).

Diante disso, neste artigo, é apresentada uma abordagem para Orquestração do Conhecimento Curricular em Ciência da Computação com base no CR05 e em estruturas curriculares de cursos de bacharelado em CC. Para isso, de forma semiautomática e *bottom-up*, as disciplinas dos cursos de graduação das IES são categorizadas em uma das 57 Categorias de Conhecimento (matérias do CR05) utilizando-se de técnicas de categorização textual estatística com Redes Bayesianas, possibilitando a mensuração do Grau de Ocorrência da Categoria (*GOc*) e o Grau de Dependência entre Categorias (*GDep*), ambos variando entre 0 e 1.

Um estudo de caso foi realizado com finalidade de validar a abordagem proposta. Inicialmente, foi criado um conjunto de treinamento contendo os conteúdos descritos no CR05, nas diretrizes do Enade 2014 (Exame Nacional de Desempenho dos Estudantes) [INEP 2014] e nas estruturas curriculares de 8 IES brasileiras de diversas regiões e com diferentes conceitos, juntamente com a Categoria de Conhecimento correspondente. Após isso, outras 7 estruturas curriculares, totalizando 457 disciplinas, foram submetidas à abordagem para identificação da categoria correspondente e aplicação das métricas *GOc* e *GDep*.

Os resultados mostraram que das 57 Categorias possíveis, 34 apresentaram *GOc* com representatividade entre 0.5 e 1.0; 20 apresentaram valor maior que 0.1 e menor que 0.5; e 3 não pontuaram nesse quesito. Quando analisado o *GDep*, verificou-se a existência de algumas relações entre diferentes Categorias de Conhecimento, principalmente naquelas que possuem conteúdos abordados em disciplinas a partir do 2º período dos cursos analisados. Logo, com base nos resultados e identificação de ambas as métricas, é possível afirmar que a abordagem possibilita a Orquestração do Conhecimento Curricular

em Ciência da Computação.

O restante do artigo está organizado da seguinte forma: na seção 2, são comentados os conceitos fundamentais para este trabalho; na seção 3, é descrita a abordagem proposta; na seção 4, são apresentados e discutidos os resultados; e, na seção 5, estão as considerações finais.

2. Referencial Teórico

Inicialmente, o termo Orquestração pode remeter-se a uma orquestra musical na qual os integrantes precisam se combinar ou arranjar para alcançar um objetivo. Por sua vez, este termo já é utilizado em pesquisas aplicadas à educação, principalmente nas que objetivam orquestrar as atividades em Ambientes de Aprendizagem Online [Oliveira et al. 2017] ou processos didáticos [da Silva e Cavalcante 2016]. Porém, neste trabalho, a Orquestração será empregada com a finalidade de organizar o Conhecimento Curricular, ou seja, como o conhecimento deverá entregue ao aluno durante a sua formação.

Diversas pesquisas objetivam modelar o Conhecimento Curricular que deve ser entregue aos alunos durante o processo de aprendizagem. Tal modelagem, comumente, é realizada em diferentes níveis de granularidade, a fim de maximizar seu uso em diferentes aplicações. Conforme descrito na seção anterior, algumas áreas de conhecimento apresentam documentos bem fundamentados que descrevem o conhecimento necessário para cada área.

A organização do Conhecimento Curricular para a área da Ciência da Computação, pode ser auxiliada por um conjunto de documentos que descrevem o conhecimento necessário para a formação do aluno. O CS2013 [Joint Task Force on Computing Curricula e Society 2013] descreve o conhecimento curricular da CC organizado em um Corpo de Conhecimento (do inglês *Body of Knowledge*) composto por: Áreas de Conhecimento, Unidades de Conhecimento e Tópicos. O CR05 [SBC 2005], outro documento, descreve o conhecimento da CC organizada hierarquicamente em: 6 Núcleos, 58 Matérias, e um conjunto de Tópicos associados às matérias.

4. Fundamentos da Computação (F)

F1. Análise de Algoritmos
F2. Algoritmos e Estrutura de Dados
F3. Arquitetura e Organização de Computadores
F4. Circuitos Digitais
F5. Fundamentos de Sistemas
F6. Linguagens de Programação
F7. Linguagens Formais, Autômatos e Computabilidade
F8. Organização de Arquivos e dados
F9. Sistemas Operacionais
F10.Teoria dos Grafos

F1. Análise de Algoritmos

Medidas de Complexidade, Análise Assintótica de Limites de Complexidade, Técnicas de Prova de Cotas Inferiores. Notação "Big O", "Little o", "Omega" e "Theta". Medidas Empíricas de Performance. O Uso de Relações de Recorrência para Análise de Algoritmos Recursivos. Análise de Algoritmos Iterativos e Recursivos.

Figura 1. Fragmento do CR05 - Núcleo Fundamentos da Computação [SBC 2005]

Na Figura 1, é apresentada a modelagem em diferentes granularidades, no qual o Núcleo **Fundamentos da Computação** é composto por 10 Matérias. No exemplo, a Matéria **F1 - Análise de Algoritmos** é detalhada em uma série de **Tópicos** que podem auxiliar o processo de planejamento curricular das disciplinas de cursos de graduação ou

pesquisas que objetivam identificar dependências entre diferentes níveis do conhecimento, conforme a pesquisa de [Marshall 2014], que estruturou o conhecimento do CS2013 utilizando grafos e utilizou a estrutura criada para comparar os currículos de cursos em CC.

3. Abordagem para Orquestração do Conhecimento Curricular

A abordagem proposta decorre do fato de que o CR05 não apresenta relações de pré-requisitos entre os diferentes conhecimentos. Porém, tais relações podem ser úteis para o processo de planejamento curricular dos cursos de graduação e/ou estruturação do conhecimento da CC em sistemas automatizados, uma vez que, comumente, os conhecimentos não estão isolados em cursos.

Diante disso, é proposta uma abordagem *bottom-up* para Orquestrar o Conhecimento Curricular em Ciência da Computação a partir das Categoria de Conhecimento, denominada neste trabalho para referenciar as Matérias do CR05, e das relações de dependência existente entre as disciplinas dos cursos de graduação [Barbosa 2016], conforme a Figura 2.

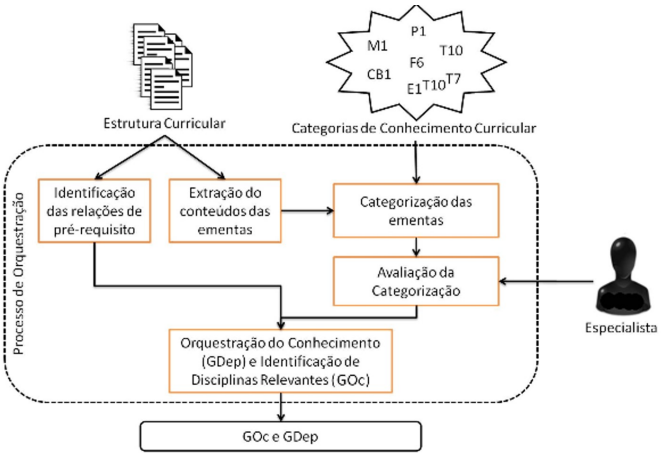

Figura 2. Arquitetura da Abordagem de Orquestração do Conhecimento

Partindo do pressuposto de que os cursos de bacharelado em CC do país baseiam-se na estruturação do conhecimento descrita no CR05, acredita-se que será possível mapear as disciplinas das IES para uma das 57 Categorias de Conhecimento. A partir disso, para cada curso analisado, deve ser realizada uma análise das diferentes Categorias identificadas para cada disciplina mapeada, e a Categoria da disciplina de pré-requisito da mesma, quando existir. Consequentemente, essa análise identificará a(s) relação(ões) de dependência(s) entre diferentes Categorias.

Como as Matérias do CR05 apresentam Conteúdos que podem ser estruturadas em mais de uma disciplina dos cursos de graduação, também foi proposto um processo de sumarização das Categorias, com a finalidade de agrupar as categorias semelhantes. Com isso, é mensurado o Grau de Ocorrência das Categorias (*GOc*) e do Grau de Dependência entre Categorias (*GDep*), fundamentais para a Orquestração. As subseções a seguir descrevem os principais conceitos e técnicas empregadas na abordagem.

3.1. Categorias de Conhecimento e Estruturas Curriculares

As Categorias de Conhecimento $C = \{c_1, c_2, ..., c_{57}\}$ utilizadas na abordagem correspondem às Matérias do CR05 (Figura 3). Das 58 possíveis, a Categoria **P8 - Estágio** foi desconsiderada por não apresentar Tópicos que pudessem ser detalhados em disciplinas dos cursos de graduação.

Matemática (M)
M1. Álgebra Linear
M2. Análise Combinatória
M3. Cálculo Diferencial e Integral
M4. Equações Diferenciais
M5. Geometria Analítica
M6. Lógica Matemática
M7. Matemática Discreta
M8. Probabilidade e Estatística
M9. Variáveis Complexas

Ciências Básicas (CB)
CB1. Física

Eletrônica (E)
E1. Circuitos Eletrônicos

Fundamentos da Computação (F)
F1. Análise de Algoritmos
F2. Algoritmos e Estrutura de Dados
F3. Arquitetura e Organização de Computadores
F4. Circuitos Digitais
F5. Fundamentos de Sistemas
F6. Linguagens de Programação
F7. Linguagens Formais, Autômatos e Computabilidade
F8. Organização de Arquivos e dados
F9. Sistemas Operacionais
F10.Teoria dos Grafos

Tecnologia da Computação (T)
T1. Análise de Desempenho
T2. Bancos de Dados
T3. Circuitos Integrados
T4. Compiladores
T5. Computação Gráfica
T6. Automação e Controle
T7. Engenharia de Software
T8. Inteligência Artificial
T9. Interação Humano- Computador
T10. Matemática Computacional
T11. Métodos Formais
T12. Modelagem e Simulação
T13. Processamento Digital de Sinais
T14. Processamento de Imagens
T15. Programação Paralela
T16. Redes de Computadores
T17. Segurança e Auditoria de Sistemas
T18. Sistemas Digitais
T19. Sistemas Distribuídos
T20. Sistemas Embarcados
T21. Sistemas Multimídia
T22.Tolerância a Falhas
T23. Telecomunicações

Contexto Social e Profissional (P)
P1. Administração
P2. Computadores e Sociedade
P3. Comunicação e Expressão
P4. Contabilidade e Custos
P5. Direito e Legislação
P6. Economia
P7. Empreendedorismo
P9. Filosofia
P10. Informática na Educação
P11. Inglês
P12. Métodos Quantitativos Aplicados à Administração de Empresas
P13. Sociologia
P14. Psicologia

Figura 3. Categorias de Conhecimento [SBC 2005]

Outro documento utilizado na abordagem é a estrutura curricular dos cursos de graduação em CC das IES. Comumente, essas estruturas descrevem o conjunto de disciplinas do curso $D = \{d_1, d_2, ..., d_{n-1}, d_n\}$, as ementas, conteúdos, relações de pré-requisitos de cada $d_i \in D$.

3.2. Categorização de Textos

O processo de Categorização de Textos (CT) ou Classificação de Textos foi utilizado com a finalidade de classificar as disciplinas das estruturas curriculares $d_i \in D$ em apenas uma das Categoria de Conhecimento $c_i \in C$. Para isso, foi utilizado o método de Categorização de Textos estatístico baseado em Redes Bayesianas [Feldman e Sanger 2007, Witten e Frank 2005] e a ferramenta WEKA[1].

A CT utiliza os conteúdos/tópicos descritos nas ementas das disciplinas para classificá-la automaticamente em uma Categoria do Conhecimento $D \times C$. Para isso, deve ser criado um conjunto de treinamento contendo o máximo de dados textuais possíveis para cada Categoria. A abordagem propõe a criação do conjunto a partir dos dados apresentados no Corpo do Conhecimento do CR05, dos conteúdos descritos nas diretrizes do Enade, e das ementas de disciplinas dos cursos de graduação que apresentaram nota máxima no exame do Enade. Devido à utilização do software de Mineração de Dados WEKA, optou-se por criar um único arquivo com extensão .arff contendo os dados extraídos de todos os documentos. O arquivo criado não possui um tamanho máximo, porém, o layout do arquivo (Figura 4) é composto por um atributo do tipo *string* que

[1]Disponível em: https://www.cs.waikato.ac.nz/ml/weka/

deve os conter conteúdos abordados na Categoria (atributo @document_ementa) e outro para o rótulo da Categoria de Conhecimento (atributo @disc_sbc) que deve ser marcado manualmente.

```
1  @relation Modelo_ARFF
2
3  @attribute document_ementa string
4  @attribute disc_sbc{algebraLinear, analiseComb,
5
6  @data
7
8  "limite e continuidade, derivada, integral inde:
9  "vetores no plano e no espaco; retas no plano e
10 "algebra booleana e portas logicas; simplificac:
11 "introducao e conceitos; logida da programacao;
12 "estruturas lineares e encadeadas: listas,matri:
13 "sistemas de equacoes lineares; espacos vetoria:
14 "componentes de computadores; numeros, artmetic:
15 "logica, tecnicas de demonstracao e validacao d
```

Figura 4. Formato do arquivo .arff utilizado na CT

Após a criação do conjunto treinamento, devem ser criados os documentos que serão categorizados. Cada disciplina das Estruturas Curriculares dos cursos teve seu conteúdo organizado em um arquivo no formato .arff com layout semelhante ao apresentado na Figura 4, porém, neste caso, o atributo @disc_sbc deve conter uma "?" (interrogação), pois o mesmo será submetido ao processo de CT que irá mensurar o valor desse atributo, ou seja, a Categoria de Conhecimento. Devido ao uso da técnica de Redes Bayesianas, o resultado da CT indicará a distribuição de probabilidade da disciplina (termos dos conteúdos) pertencer às Categorias de Conhecimento $D \times C$ variando entre 0 (zero) e 1 (um).

Devido à natureza da técnica empregada, uma disciplina $d_i \in D$ pode ter seu conteúdo categorizado em mais de um $c_i \in C$. Diante disso, todo o resultado da CT $(D \times C)$ deve ser submetido a um processo de validação que consiste em:

1. Se $D \times C = 1 \rightarrow$ Aceitar a Categoria de Conhecimento resultante;
2. Se $0 < D \times C < 1 \rightarrow$ Deve-se analisar o nome da disciplina e conteúdo com a finalidade de:
   (a) Aceitar o resultado da CT;
   (b) Recusar o resultado e recategorizar manualmente em outra;
   (c) Recusar o resultado e recategorizar como categoria excedente, ou seja, não apresenta correspondente válido;
3. Se $D \times C = 0 \rightarrow$ Não aceitar a Categoria de Conhecimento.

Devido a não padronização dos documentos utilizados, a extração dos textos utilizados no conjunto treinamento ou das disciplinas que serão categorizadas ocorreu de forma manual. Ressalta-se que, após a extração e criação dos arquivos .arff, estes passaram por uma etapa de pré-processamento com a finalidade de remover caracteres especiais, além da redução dos documentos com a criação de *tokens*, definição da frequência mínima dos termos e remoção das *stopwords* em português.

3.3. Orquestração do Conhecimento Curricular

O processo de Orquestração do Conhecimento consiste em um conjunto de ações com a finalidade de mensurar o Grau de Ocorrência das Categorias (*GOc*) e o Grau de Dependência entre Categorias (*GDep*) identificadas após o processo de categorização.

3.3.1. Grau de Ocorrência das Categorias (*GOc*)

O resultado da CT pode mostrar diversas disciplinas $d_i \in D$ categorizadas em uma mesma $c_i \in C$ ou apresentar Categorias que não foram abordadas em uma estrutura curricular D, o que resulta em uma representação variável das Categorias de Conhecimento nas estruturas curriculares. Diante disso e objetivando identificar a relevância de determinada $c_i \in C$ nas diferentes estruturas de cursos $D_k = \{D_1, D_2, ..., D_n\}$ utilizadas na abordagem, é proposto o Grau de Ocorrência das Categorias (*GOc*).

$$GOc(c_i) = \frac{\sum_{k=1}^{n}(goc(k, c_i))}{n} \tag{1}$$

O cálculo do *GOc* (Equação 1) baseia-se na representatividade booleana da Categoria de Conhecimento na estrutura analisada. Portanto, quando houver representatividade da Categoria na estrutura do curso, o valor será 1 (um), caso contrário, o valor será 0 (zero). Após o processamento de todas as estruturas, é calculada para cada $c_i \in C$ a média aritmética da quantidade de Categorias identificadas. Para isso, utiliza-se a quantidade n de estruturas curriculares analisadas e a função $goc(k, c_i)$ que verifica o valor de cada $c_i \in C$ representada na estrutura k analisada. Devido à natureza da fórmula, o resultado será um $0 \leq GOc \leq 1$.

3.3.2. Grau de Dependência entre Categorias (*GDep*)

As relações de dependências entre as Categorias de Conhecimento são identificadas a partir da análise das relações dos pré-requisitos entre as disciplinas submetidas ao processo de CT. Devido à natureza da categorização, vários registros $d_i \in D$ podem ser categorizados em uma única categoria $c_j \in C$. Logo, poderão podendo existir registros $d_i \in D$ que apresentem relações de pré-requisito com disciplinas categorizadas na mesma Categoria de Conhecimento. Para esta situação, foi criado um algoritmo para sintetizar e identificar os relacionamentos entre as Categorias identificadas na coleção de documentos D, considerando as seguintes situações:

1. Se a disciplina categorizada apresentar relação de pré-requisito com outra disciplina de mesma categoria. Então, a relação entre as categorias será anulada (Figura 5).
2. Se a disciplina apresenta como pré-requisito uma disciplina de Categoria diferente a dela. Então, existirá uma relação correspondente entre as duas Categorias (Figura 6).
3. Se houver mais de um pré-requisito para d_i categorizada em c_i. Então, a dependência de c_i será a união de todas as Categorias de Conhecimento identificadas por meio das relações de dependência válida, conforme é apresentado na Figura 7.

Após o processo de sintetização e identificação das relações de dependências válidas, é calculado o Grau de Dependência entre Categorias ($GDep$) com a finalidade de quantificar a intensidade da relação de dependência entre diferentes Categorias do Conhecimento. Inicialmente, para cada estrutura D analisada, cria-se uma matriz $N_{57\times57}$,

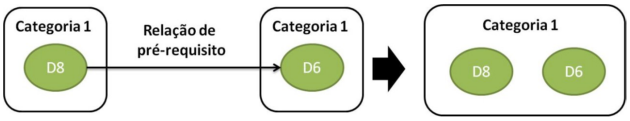

Figura 5. Exemplo de Dependência Nula entre Categorias

Figura 6. Exemplo de Dependência Válida entre diferentes Categorias

em que $N_{i,j}$ apresenta valor 0 ou 1, indicando, respectivamente, a ausência ou presença da relação de dependência entre as Categorias representadas pela linha e coluna da matriz.

As matrizes individuais N são utilizadas no cálculo do $GDep$ das Categorias de Conhecimento. Representado por uma matriz quadrada $M_{57\times57}$ e com estrutura semelhante à matriz N, o resultado do $GDep$ mostra o elemento $M_{i,j}$, com i e j variando até 57, apresentando valor $0 \leq GDep(i,j) \leq 1$. Este valor é calculado na média aritmética da função $gd(k, N_{i,j})$ que retorna o grau de relevância do elemento $N_{i,j}$, ou seja, o grau do relacionamento entre as categorias c_i e c_j, da $k - ésima$ estrutura analisada, conforme a Equação 2.

$$GDep(i,j) = \frac{\sum_{k=1}^{n}(gd(k, N_{i,j}))}{n} \tag{2}$$

Ressalta-se que o uso de ambas as métricas foi necessário, pois, se fosse considerada apenas uma delas, haveria perda de informações que poderiam ser úteis para comunidade. Por exemplo, se apenas o GOc fosse mensurado, não seria possível a identificação e mensuração das relações entre as Categorias do Conhecimento. Por outro lado, se apenas o $GDep$ fosse mensurado, seriam perdidas informações sobre a relevância das Categorias de Conhecimento nas estruturas da curriculares, pois, provavelmente algumas não seriam identificadas por não se relacionar com outras.

4. Resultados e Discussão

Conforme descrito na Seção 3, a abordagem proposta mescla atividades manuais e automatizadas baseadas em um conjunto de documentos, ferramentas e análises. De forma geral, a condução do estudo de caso baseou-se na **seleção das estruturas curriculares; na execução do processo de orquestração; e na verificação do uso das métricas identificadas no processo de planejamento e avaliação das estruturas curriculares**.

Diante disso, a seleção de estruturas curriculares utilizadas na criação do conjunto de treinamento e no processo de CT para a Orquestração baseou-se no conceito Enade 2014[2] dos cursos de graduação das IES. As estruturas selecionadas para o processo de Orquestração basearam-se na ordem decrescente das IES que obtiveram nota do conhecimento especifico (CE) do Enade 2014 superior a 50,0 pontos e que apresentavam suas

[2]Conceito Enade. Disponível em:http://portal.inep.gov.br/web/guest/conceito-enade

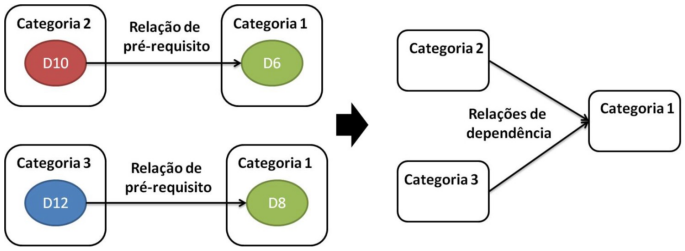

Figura 7. Exemplo de Dependência Válida entre mais de uma Categoria

estruturas curriculares disponíveis na web (Tabela 1). Para o conjunto treinamento, foram selecionadas 8 IES de diferentes regiões do país que apresentaram CE superior a 50,0 pontos e suas estruturas disponíveis na web (Tabela 2). Tal escolha deu-se por achar necessário abordar não apenas IES com nota máxima, além de diversificar o padrão de escrita dos documentos.

Tabela 1. Distribuição dos Cursos Utilizados no Processo de Orquestração do Conhecimento Curricular

Identificador do Curso	Nota Bruta - CE	Conceito ENADE
Curso 1	61,4	5
Curso 2	60,2	5
Curso 3	60,0	5
Curso 4	55,8	5
Curso 5	59,9	5
Curso 6	55,5	4
Curso 7	58,2	5

Tabela 2. Distribuição dos Cursos Utilizados no Conjunto de Treinamento

Identificador do Curso	Região	Conceito ENADE
Curso 1	Sul	5
Curso 2	Sudeste	5
Curso 3	Norte	5
Curso 4	Nordeste	4
Curso 5	Centro-oeste	3
Curso 6	Sudeste	3
Curso 7	Nordeste	3
Curso 8	Norte	2

Após a seleção, as estruturas foram analisadas para a extração dos textos e criação dos arquivos .arff (Figura 4) do conjunto de treinamento. Todo esse processo ocorreu manualmente devido a não padronização dos documentos das diferentes IES, e os dados extraídos foram estruturados em um único arquivo com 519 registros rotulados com as diferentes Categorias de Conhecimento, com distribuição conforme a Figura 8. Por outro lado, os arquivos utilizados no processo de Orquestração foram organizados em sete arquivos .arff (um para cada estrutura de IES analisada), com tamanhos variáveis, sem o rótulo da possível Categoria de Conhecimento, distribuídos conforme a Figura 9.

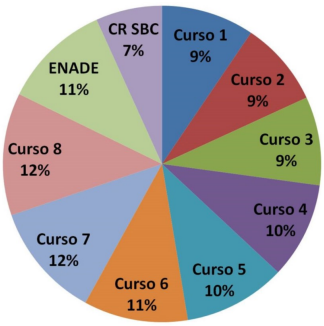

Figura 8. Distribuição dos documentos utilizados no Conjunto de Treinamento

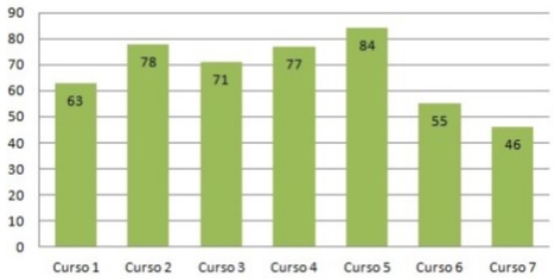

Figura 9. Distribuição da quantidade de disciplinas por curso utilizado na Orquestração

A distribuição das disciplinas por curso mostra 457 disciplinas que devem ser categorizadas em uma Categoria de Conhecimento. O uso da Mineração de Textos possibilitou a automatização desse processo. Na Tabela 3, é apresentado o resultado do processo de categorização indicando a quantidade de registros categorizados que foram aceitos ou que precisaram de recategorização (Tabela 4), conforme o algoritmo descrito na Seção 3.

Tabela 3. Resultado do Processo de Categorização

Curso	Quant. Disc.	Disc. Categ. em 1	Disc. submetidas à análise
1	63	47	16
2	71	55	16
3	78	42	36
4	77	48	29
5	84	45	39
6	55	39	16
7	46	32	14

Com as disciplinas das estruturas categorizadas, foram mensurados o GOc e o $GDep$ das Categorias de Conhecimento fundamentais para a Orquestração do Conhecimento Curricular. Ambas as métricas devem considerar todas as 57 categorias propostas, porém, conforme destacado anteriormente, as estruturas podem não abordar algumas categorias. Nesse contexto, ressalta-se que, quão mais próximo de 1 (um) for o resultado do GOc e o $GDep$, respectivamente, isso indica que a Categoria apresenta um alto grau de representatividade nas estruturas analisadas, e que a relação dela com outras também

Tabela 4. Resultado da Avaliação da Categorização dos Registros que Necessitavam de Análise

Curso	Quant. Disciplinas.	Disciplinas Aceitas	Disciplinas Recategorizadas	Disciplinas Excedentes
1	16	8	6	2
2	16	7	8	1
3	36	20	11	5
4	29	13	10	6
5	39	17	15	7
6	16	7	4	5
7	14	9	3	2

apresenta-se como representativa.

Na Figura 10, é apresentado o resultado do Grau de Ocorrência da Categoria de Conhecimento do Núcleo Fundamentos da Computação. Nesse exemplo, é possível verificar diferentes GOc variando entre $0,5 - 1$, destacando-se como representativa no estudo realizado, até os menos representativos que variam entre $0 - 0,49$. O resultado detalhado de todas as Categorias de Conhecimento pode ser visto neste link[3].

Categoria de Conhecimento	Obrigatoria	Optativa	Geral	GOC obrigatório	GOC Optativa	GOC
F1. Análise de Algoritmos	5		5	0,71	0,00	0,71
F2. Algoritmos e Estrutura de Dados	6		6	0,86	0,00	0,86
F3. Arquitetura e Organização de Computadores	7		7	1,00	0,00	1,00
F4. Circuitos Digitais	6	1	7	0,86	0,14	1,00
F5. Fundamentos de Sistemas	1	3	4	0,14	0,43	0,57
F6. Linguagens de Programação	7		7	1,00	0,00	1,00
F7. Linguagens Formais, Autômatos e Computabilidade	7		7	1,00	0,00	1,00
F8. Organização de Arquivos e dados	1		1	0,14	0,00	0,14
F9. Sistemas Operacionais	7		7	1,00	0,00	1,00
F10.Teoria dos Grafos	5	1	6	0,71	0,14	0,86

Figura 10. GOc **do Núcleo - Fundamentos da Computação**

Das 57 Categorias de Conhecimento utilizadas na abordagem, o $GDep$ identificou que apenas 45 apresentam relações de pré-requisitos com outras 32. Ressalta-se que uma Categoria de Conhecimento $c_i \in C$ pode relacionar-se com outras 56 Categorias, pois uma auto-relação é considerada nula, conforme descrito no processo de sintetização (Figura 5).

O resultado do $GDep$ é apresentado em uma Matriz com dimensão 57×57, na qual as linhas e colunas referem-se às Categorias de Conhecimento utilizadas[4], o que impossibilita a inserção dela no artigo. Diante disso, optou-se por apresentar o resultado $DGep$ de Categorias pertencentes ao mesmo Núcleo de Conhecimento. Na Figura 11, mostra-se o $GDep$ calculado entre as Categorias do núcleo de conhecimento Matemática, sendo possível identificar relações de pré-requisito entre as Categorias, tendo como mais representativa a relação entre **M3 - Cálculo Diferencial e Integral** e **M8 - Probabilidade e Estatística**, mensurada em 0.86, (ou seja, essa mesma relação foi identificada em dife-

[3]Resultado Completo do GOc - Disponível em https://goo.gl/5T28ii

[4]Resultado completo $GDep$ - Disponível em: https://goo.gl/bEN6PH

	M1	M2	M3	M4	M5	M6	M7	M8	M9
M1. Álgebra Linear	0,00	0,00	0,00	0,00	**0,29**	0,00	0,00	0,00	0,00
M2. Análise Combinatória	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
M3. Cálculo Diferencial e Integral	**0,14**	0,00	0,00	0,00	**0,14**	0,00	0,00	0,00	0,00
M4. Equações Diferenciais	0,00	0,00	0,00	0,00	0,00	0,00	**0,14**	0,00	0,00
M5. Geometria Analítica	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
M6. Lógica Matemática	0,00	0,00	0,00	0,00	0,00	0,00	**0,29**	0,00	0,00
M7. Matemática Discreta	0,00	0,00	0,00	0,00	0,00	**0,14**	0,00	0,00	0,00
M8. Probabilidade e Estatística	0,00	0,00	**0,86**	0,00	0,00	0,00	**0,14**	0,00	0,00
M9. Variáveis Complexas	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Figura 11. $GDep$ **entre Categorias do núcleo Matemática**

	M1	M2	M3	M4	M5	M6	M7	M8	M9
F1. Análise de Algoritmos	0,00	0,00	0,00	0,00	**0,14**	0,00	**0,14**	**0,14**	0,00
F2. Algoritmos e Estrutura de Dados	0,00	0,00	0,00	0,00	0,00	0,00	**0,29**	0,00	0,00
F3. Arquitetura e Organização de Computadores	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F4. Circuitos Digitais	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F5. Fundamentos de Sistemas	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F6. Linguagens de Programação	0,00	0,00	0,00	0,00	0,00	0,00	**0,14**	0,00	0,00
F7. Linguagens Formais, Autômatos e Computabilidade	0,00	0,00	0,00	0,00	0,00	**0,43**	**0,57**	0,00	0,00
F8. Organização de Arquivos e dados	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
F9. Sistemas Operacionais	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**0,14**	0,00
F10.Teoria dos Grafos	0,00	0,00	0,00	0,00	0,00	0,00	**0,14**	0,00	0,00

Figura 12. $GDep$ **entre Categorias do núcleo Fundamentos da Computação e Matemática**

rentes estruturas curriculares analisadas). Por outro lado, na Figura 12, apresenta-se uma relação $GDep$ representativa que envolve os núcleos Fundamentos da Computação e Matemática, conforme é verificado na relação entre **F7 - Linguagens Formais, Autômatos e Computabilidade** e **M7 - Matemática Discreta**, mensurada em 0.57.

Os resultados apresentados com a mensuração do GOc e $GDep$ comprovam que o conhecimento curricular em Ciência da Computação pode ser orquestrado a partir da abordagem proposta. Os resultados podem auxiliar o processo de planejamento curricular, uma vez que, dentre as Categorias de Conhecimento estabelecidas, foram identificadas aquelas que apresentam maior representatividade e, além disso, foram identificadas relações entre elas. Por fim, acredita-se que a orquestração do conhecimento também pode contribuir para pesquisas que necessitem de uma organização curricular com base nas matérias do CR05, como também em análises que verifiquem a conformidade da estrutura curricular dos cursos com o CR05.

5. Considerações Finais

A partir da mensuração das métricas GOc e $GDep$, alcançadas por meio da abordagem para Orquestração do Conhecimento proposta, foi constatado o que foi pressuposto no início do estudo: (i) as disciplinas das grades curriculares dos cursos de CC do país baseiam-se no conhecimento descrito no CR05, e (ii) é possível Orquestrar o Conhecimento Curricular com base nos pré-requisitos das disciplinas analisadas.

Com valores variando entre $0 - 1$, para o Grau da Ocorrência da Categoria (GOc) e o Grau de Dependência entre Categorias ($GDep$), conseguiu-se mensurar quão representativa é cada Categoria de Conhecimento, e como se dá a relação dela com as demais,

observando o contexto de pré-requisito. Ressalta-se que, quando mensurada a representatividade das categorias, verificou-se que mais de 50%, das 57 Categorias possíveis, apresentaram-se com representatividade acima de 0.5. Por outro lado, no contexto do estudo realizado, constatou-se que, do mesmo total de Categoria possíveis, 47 delas se relacionaram com outras 32.

Os números destacados no parágrafo anterior e detalhados nos resultados gerados mostram que, para o contexto analisado, a abordagem conseguiu responder aos questionamentos realizados e, além disso, agrega valor ao conhecimento descrito no CR05, uma vez que possibilitou a Orquestração do Conhecimento Curricular (na granularidade Matéria), podendo ser útil para o planejamento e avaliação da conformidade de um curso com o "padrão" identificado.

Contudo, ressalta-se que limitações de automatização do processo foram identificadas, porém, não inviabilizaram o objetivo final do trabalho. Ressalta-se ainda que o uso da abordagem não se restringe apenas em identificar um padrão que possa ser útil para sequenciar disciplinas de um curso de graduação em CC, ou avaliar a disposição das disciplinas no curso, mas, também, acredita-se que o resultado seja de grande valia para a comunidade que estuda o ensino em computação ou que necessita estruturar o conhecimento em outras área do conhecimento.

Referências

Barbosa, A., Nunes, I., Menezes, D., Schiel, U. (2015). O design instrucional e seu uso como arquitetura pedagógica: Uma análise das publicações em informática na educação no brasil. In *Anais dos Workshops do IV Congresso Brasileiro de Informática na Educação (CBIE)*. Maceió, Alagoas.

Barbosa, A. F. (2016). Uma abordagem para orquestração do conhecimento com suporte ao planejamento e avaliação curricular em ciência da computação. Dissertação de Mestrado, Universidade Federal de Campina Grande.

da Silva, P. A. Cavalcante, P. S. (2016). Orquestrando processos didáticos com design thinking. In *Anais dos Workshops do V Congresso Brasileiro de Informática na Educação (CBIE)*. Uberlândia, Minas Gerais.

Feldman, R. Sanger, J. (2007). *Text Mining Handbook: Advanced Acpproaches in Analyzing Unstructured Data.* Cambridge University Press, New York, NY, USA.

Filatro, A. (2008). *Design Instrucioinal na Prática*. Pearson Education do Brasil, 1ª edição.

INEP (2014). Portaria inep nº 238, de 02 de junho de 2014. Disponível em: http//download.inep.gov.br/educacao_superior/enade/legislacao/2014/diretrizes_cursos_diplomas_bacharel/diretrizes_bacharel_computacao.pdf. Acessado: 31 mar. 2018.

Joint Task Force on Computing Curricula, A. f. C. M. A. Society, I. C. (2013). *Computer Science Curricula 2013: Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. ACM, New York, NY, USA. 999133.

Marshall, L. (2014). A grapf-based framework for comparing curricula. Dissertação de Mestrado, University of Pretoria. Disponível em: https://repository.up.ac.za/handle/2263/37060. Acessado em 30 mar. 2018.

Oliveira, I. V. P. D., Gomes, A. S., Brito, J. A., Filho, I. J. M. (2017). Learning orchestration in distributed learning environments scenarios. In *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–4.

SBC (2005). Currículo de referência da SBC para cursos de graduação em bacharelado em ciência da computação e engenharia de computação. Disponível em: http://www.sbc.org.br/documentos-da-sbc/send/131-curriculos-de-referencia/760-curriculo-de-referencia-cc-ec-versao2005/. Acessado: 31 mar. 2018.

Witten, I. H. Frank, E. (2005). *Data mining : practical machine learning tools and techniques*. Elservier, San Francisco, CA.