

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Dissertação de Mestrado

Uma Abordagem para Orquestração do  
Conhecimento com Suporte ao Planejamento e à  
Avaliação Curricular em Ciência da Computação

Anderson Felinto Barbosa

Campina Grande, Paraíba, Brasil

Fevereiro - 2016

Universidade Federal de Campina Grande  
Centro de Engenharia Elétrica e Informática  
Coordenação de Pós-Graduação em Ciência da Computação

Uma Abordagem para Orquestração do  
Conhecimento com Suporte ao Planejamento e à  
Avaliação Curricular em Ciência da Computação

Anderson Felinto Barbosa

Dissertação submetida à Coordenação do Curso de Pós-Graduação em  
Ciência da Computação da Universidade Federal de Campina Grande -  
Campus I como parte dos requisitos necessários para obtenção do grau  
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Sistemas de Informação

Ulrich Schiel

(Orientador)

Campina Grande, Paraíba, Brasil

©Anderson Felinto Barbosa, 29/02/2016

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

- B238a      Barbosa, Anderson Felinto.  
              Uma abordagem para orquestração do conhecimento com suporte ao planejamento e à avaliação curricular em Ciência da Computação / Anderson Felinto Barbosa. – Campina Grande, 2016.  
              84f. : il. color.
- Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática.  
              "Orientação: Prof. Ulrich Schiel".
1. Ciência da Computação - Currículo. 2. Ciência da Computação - Ensino - Brasil. 3. Processos de Aprendizagem. 4. Planejamento Curricular. I. Schiel, Ulrich. II. Título.

CDU 004:37.016(043)

**"UMA ABORDAGEM PARA ORQUESTRAÇÃO DO CONHECIMENTO COM SUPORTE  
AO PLANEJAMENTO E AVALIAÇÃO CURRICULAR EM CIÊNCIA DA COMPUTAÇÃO"**

**ANDERSON FELINTO BARBOSA**

**DISSERTAÇÃO APROVADA EM 29/02/2016**



**ULRICH SCHIEL, Dr., UFCG  
Orientador(a)**



**JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, D.Sc, UFCG  
Examinador(a)**



**ARTURO HERNANDEZ DOMINGUEZ, Dr., UFAL  
Examinador(a)**

**CAMPINA GRANDE - PB**

*Dedico esta dissertação  
aos meus pais, base da minha educação.*

## **Agradecimentos**

Primeiramente, agradeço a Deus, que me concedeu forças para a realização desse mestrado e enfrentar diversos momentos difíceis durante essa caminhada.

À minha família, pelo carinho, compreensão e incentivo durante os momentos difíceis; um agradecimento especial aos meus pais.

Ao meu orientador Ulrich Schiel, agradeço pela oportunidade de trabalharmos juntos, pelos ensinamentos e incentivos que recebi durante todo o processo de orientação, também pela compreensão e companheirismo nos bons e maus momentos vivenciados durante esses período de pós-graduação.

Agradeço aos professores que ministraram disciplinas durante o mestrado e aos companheiros de laboratório e disciplinas com os quais pude compartilhar dúvidas e opiniões que contribuíram para a minha formação.

Agradeço às amigas e ex-professoras Vera Medeiros e Isabel Nunes que contribuíram para o meu ingresso na pós-graduação.

Aos amigos conquistados, dentro e fora da UFCG, que torceram pela concretização desse projeto, vivenciando comigo as angústias e alegrias de um pós-graduando.

Agradeço aos funcionários do PPGCC pela assistência prestada, sempre que necessária, durante o período do mestrado.

Por fim, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

*“Quem não age, perde até mesmo a pouca  
compreensão que já tem ou pensa ter.”*

*(Marcos 4:21-25)*

## Resumo

O Currículo Referência (CR) para Cursos de Bacharelado em Ciência da Computação e Engenharia da Computação é um documento desenvolvido pela Sociedade Brasileira de Computação que descreve um conjunto de informações relacionadas ao ensino da computação no Brasil. O corpo do conhecimento, parte do CR descreve o conhecimento que pode ser ensinado ao aluno durante o processo de aprendizagem, pode ser útil para o processo de Planejamento Curricular, porém, diferentemente das estruturas curriculares criadas, o conhecimento descrito no CR apresenta, apenas, relações hierárquicas entre suas partes. Diante disso, neste trabalho, é proposta uma abordagem que visa a orquestrar o conhecimento descrito na granularidade “Matéria” do CR, neste trabalho denominada como “Categoria de Conhecimento”. Para isso, a abordagem baseia-se no princípio *bottom-up* e identifica as possíveis relações entre as 57 Categorias de Conhecimento a partir das disciplinas e das relações de pré-requisitos presentes em estruturas curriculares dos cursos para, assim, identificar o Grau de Dependência entre Categorias (*GDep*) e o Grau de Ocorrência da Categoria (*GOC*) relevantes para o processo de planejamento e avaliação das estruturas curriculares. Como forma de avaliar a abordagem foi realizado um estudo de caso envolvendo dados de 474 disciplinas, de 7 cursos de graduação, no qual foi mensurado do *GRel* e o *GOC* das 57 Categorias de Conhecimento. Além disso, também foi verificada a aplicação das métricas identificadas no processo de planejamento e avaliação das estruturas curriculares que, no contexto observado, apresentou resultados que comprovaram o uso de ambas para as finalidades propostas.

**Palavras-chave:** Orquestração, Planejamento, Avaliação, Estrutura Curricular, Currículo Referência.



## Abstract

*The Reference Curriculum for Degree Courses in Computer Science and Computer Engineering (RC) is a document developed by the Brazilian Computer Society that aims to describe a set of information related to computing education in Brazil. The body of knowledge, a part of RC that describes the knowledge that should be taught to a student during the learning Process, can be useful on the Curricular Planning process, however, unlike the curriculum structure created, which describes relations between disciplines, the knowledge described on the Reference Curriculum only shows hierarchical relations between different levels of this knowledge. To this situation, in this dissertation is proposed an approach that aimed orchestrating the knowledge described on granularity “Subject” of RC, called, in this dissertation, “Knowledge Categories”. For this, the approach is based on the bottom-up and identifies possible relationships between the 57 Knowledge Categories induced from the disciplines of graduate courses and their prerequisites relationships, to thereby identify the Dependence Degree Between Categories (GDep) and Category Occurrence Degree (GOc) relevants for the process of curriculum planning and evaluation. For evaluate approach, ion purposes a case study was conducted involving 474 disciplines from 7 undergraduate courses. It was possible to observe the measurement of GRel and GOC of the 57 Knowledge Categories from RC. In addition, the application were verified of the metrics identified in the planning process and evaluation of curriculum structures that, on the context observe, showed results that confirmed the use of such metrics for aims proposed.*

**Keywords:** *Orchestration, Planning, Evaluation, Curricula Structure, Reference Curriculum.*

# Lista de Siglas e Abreviaturas

*BCC – Bacharelado em Ciência da Computação*

*BoK – Body of Knowledge*

*CC - Ciência da Computação*

*CC2001 – Computer Curricula 2001*

*CK – Content Knowledge*

*CR – Currículo Referência*

*CR05 – Currículo Referência SBC 2005*

*CS2008 – Computer Science Curricula 2008*

*CS2013 – Computer Science Curricula 2013*

*CSCL – Computer-Supported Collaborative Learning*

*CT – Categorização de Textos*

*DI – Design Instrucional*

*EI – Extração da Informação*

*ENADE – Exame Nacional de Avaliação do Desempenho do Estudante*

*GDep – Grau de Dependência entre Categorias*

*GOC – Grau de Ocorrência da Categoria*

*IA – Inteligência Artificial*

*IEEE – Institute of Electrical and Electronics Engineers*

*IES – Instituições de Ensino Superior*

*KA – Knowledge Areas*

*KE – Knowledge Engineering*

*KU – Knowledge Units*

*ML – Machine Learning*

*MD – Mineração de Dados*

*MT – Mineração de Textos*

*PCK – Pedagogical Content Knowledge*

*PLN – Processamento de Linguagem Natural*

*POSCOMP – Exame Nacional para Ingresso na Pós-Graduação em Computação*

*PPC – Plano Pedagógico Curricular*

*RI – Recuperação da Informação*

*SBC – Sociedade Brasileira de Computação*

*SHA – Sistemas Hiperídia Adaptativos*

*STI – Sistemas Tutores Inteligentes*

*SVM – Support Vector Machine*

*TEL – Technology-Enhanced Learning*

*TF-IDF – Term Frequency-Inverse Document Frequency*

*TIC – Tecnologia da Informação e Comunicação*

*TK – Technological Knowledge*

*TPCAK – Technological Pedagogical Content Knowledge*

*WEKA – Waikato Environment for Knowledge Analysis*

# Lista de Figuras

1.1	Fragmento de uma Esturura Curricular de um Curso de Bacharelado em Ciência da Computação . . . . .	3
2.1	Framework Conceitual TPACK. . . . .	11
2.2	Detalhamento da Unidades de Conhecimento <i>Basic Analysis</i> em Tópicos. . .	14
2.3	Framework Conceitual do Processo de Orquestração da Aprendizagem. . .	17
2.4	Processo de Stemming. . . . .	22
2.5	Representação de uma Matriz de Confusão. . . . .	25
2.6	Árvore de Decisão para o Problema da Classificação de Mamíferos. . . . .	26
2.7	<i>Support Vector Machine</i> . . . . .	26
3.1	Representação das Relações entre as Diferentes Granularidades do Currículo CS2013 . . . . .	29
3.2	Representação de Dependências Cruzadas entre o Conhecimento . . . . .	29
4.1	Diagrama da Abordagem para Orquestração do Conhecimento Curricular . .	35
4.2	Exemplo de uma Ementa Acadêmica . . . . .	36
4.3	Diagrama de Atividades Ilustrando o fluxo do Processo de Orquestração do Conhecimento Curricular . . . . .	39
4.4	Diagrama de Atividade da Subatividade “Selecionar Estrutura Curricular” .	40
4.5	Diagrama de Atividades da Subatividade “Extrair Conhecimento da Ementa”	41
4.6	Diagrama de Atividades da Subatividade “Construir Conjunto Treinamento”	43
4.7	Diagrama de Atividades da Subatividade “Categorizar Ementas” . . . . .	44
4.8	Exemplo de Dependência Nula entre Categorias. . . . .	48
4.9	Exemplo de Dependência Válida entre Diferentes Categorias. . . . .	48

---

4.10	Exemplo de Dependência Válida entre mais de uma Categorias Diferentes. .	48
5.1	Distribuição das disciplinas observadas por curso . . . . .	56
5.2	Distribuição dos registros dos cursos utilizados na construção do Conjunto Treinamento . . . . .	57
B.1	Modelo do Arquivo ARFF . . . . .	80

# Lista de Tabelas

5.1	Cursos Utilizados no Processo de Orquestração do Conhecimento Curricular	54
5.2	Distribuição dos Cursos do Conjunto Treinamento . . . . .	55
5.3	Resultado Bruto da Categorização de Ementas . . . . .	59
5.4	Resultado da Avaliação da Categorização das Ementas . . . . .	59
5.5	Detalhamento de Categorias por Curso e Núcleo de Conhecimento . . . . .	60
5.6	Mensuração do <i>GDep</i> das Categorias de Conhecimento do Núcleo Matemática	63
5.7	Mensuração do <i>GDep</i> das Categorias de Conhecimento entre os Núcleos Fundamentos da Computação e Matemática. . . . .	64
5.8	Relacionamentos Sugeridos para as Diferentes Categorias de Conhecimento utilizadas na Proposta de Planejamento . . . . .	66
5.9	Relações entre Categorias Apresentadas pelas Estruturas Curricular Analisada	67
A.1	Sumarização dos Algoritmos de Categorização . . . . .	78
C.1	Resultado <i>GOC</i> do núcleo Matemática . . . . .	81
C.2	Resultado <i>GOC</i> do núcleo Ciências Naturais . . . . .	81
C.3	Resultado <i>GOC</i> do núcleo Eletrônica . . . . .	82
C.4	Resultado <i>GOC</i> do núcleo Fundamentos da Computação . . . . .	82
C.5	Resultado <i>GOC</i> do núcleo Tecnologia da Computação . . . . .	82
C.6	Resultado <i>GOC</i> do núcleo Contexto Social e Profissional . . . . .	83
C.7	Resultado <i>Dep</i> Geral . . . . .	84

# Lista de Quadros

2.1	Áreas de Conhecimento do CC2013. . . . .	13
2.2	Detalhamento da Área de Conhecimento <i>Algorithms and Complexity</i> . . . . .	14
3.1	Quadro de Pesquisas Relacionadas . . . . .	32
5.1	Mensuração do <i>GOC</i> das Categorias de Conhecimento Mais Representativas. . . . .	61
5.2	Mensuração do <i>GOC</i> das Categorias de Conhecimento Menos Representativas . . . . .	62
5.3	Categorização das Disciplinas utilizadas no processo de Planejamento e Avaliação. . . . .	65

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivos . . . . .	4
1.3	Relevância . . . . .	5
1.4	Estrutura da Dissertação . . . . .	6
<b>2</b>	<b>Fundamentação Teórica</b>	<b>8</b>
2.1	Representação do Conhecimento Curricular . . . . .	8
2.1.1	Formas de Representação do Conhecimento Curricular . . . . .	9
2.1.2	Currículos Referência da Ciência da Computação . . . . .	12
2.2	O Processo de Orquestração . . . . .	16
2.3	Categorização de Textos Baseada em Mineração de Textos . . . . .	19
2.3.1	Tarefa de Categorização . . . . .	20
2.3.2	Pré-processamento de Documentos . . . . .	21
2.3.3	Aprendizagem de Máquina . . . . .	23
2.4	Considerações Finais do Capítulo . . . . .	27
<b>3</b>	<b>Pesquisas Relacionadas</b>	<b>28</b>
3.1	Orquestração do Conhecimento Curricular . . . . .	28
3.2	Planejamento e Avaliação Curricular . . . . .	30
3.3	Categorização de Texto . . . . .	31
3.4	Considerações Finais do Capítulo . . . . .	32
<b>4</b>	<b>Orquestração do Conhecimento Curricular em Ciência da Computação</b>	<b>33</b>
4.1	Descrição da Abordagem . . . . .	33



4.1.1	Estrutura Curricular . . . . .	35
4.1.2	Categorias de Conhecimento Curricular . . . . .	36
4.1.3	Processo de Orquestração das Categorias de Conhecimento . . . . .	38
4.2	Uso da Orquestração para Planejamento e Avaliação Curricular . . . . .	50
4.2.1	Proposta para uso no Planejamento Curricular . . . . .	50
4.2.2	Proposta para uso na Avaliação Curricular . . . . .	51
4.3	Considerações Finais do Capítulo . . . . .	51
<b>5</b>	<b>Avaliação da Abordagem Proposta</b>	<b>52</b>
5.1	Estudo de Caso . . . . .	52
5.1.1	Seleção das Estruturas Curriculares . . . . .	53
5.1.2	Processo de Orquestração de Conhecimento . . . . .	55
5.1.3	Avaliação da Proposta de Uso no Planejamento e Avaliação Curricular	64
5.2	Considerações Finais do Capítulo . . . . .	67
<b>6</b>	<b>Considerações Finais</b>	<b>69</b>
6.1	Conclusões . . . . .	69
6.2	Limitações e Sugestões para Pesquisas Futuras . . . . .	70
<b>A</b>	<b>Avaliação dos Algoritmos de Categorização</b>	<b>77</b>
<b>B</b>	<b>Categorização das Ementas</b>	<b>79</b>
B.1	Arquivos ARFF dos Cursos Orquestrados . . . . .	79
B.2	Resultado Bruto e Avaliação da Categorização dos Cursos . . . . .	79
B.3	Modelo de Arquivo ARFF . . . . .	80
<b>C</b>	<b>Resultados Detalhados: <i>GOC</i> e <i>GDep</i></b>	<b>81</b>
C.1	Detalhamento <i>GOC</i> . . . . .	81
C.2	Detalhamento <i>GDep</i> . . . . .	83

# Capítulo 1

## Introdução

A expansão do ensino superior proporcionou o acesso das pessoas aos conhecimentos das mais diversas áreas. Um processo inerente dessa expansão foi a criação de novos cursos e, com isso, há a necessidade da realização do processo de Planejamento Curricular objetivando estruturar o conhecimento que deve ser adquirido pelo estudante em estruturas, comumente, denominadas de disciplinas.

Portanto, nesse processo, observa-se que há um particionamento de um conhecimento maior em estruturas menores, mais simplificadas, organizadas entre si. Logo, pode-se inferir que o conhecimento adquirido durante o processo de ensino pode ser representado em estruturas de diferentes granularidades, estruturadas a partir da delimitação de relacionamentos.

Contudo, a delimitação dessas relações pode ser uma tarefa complexa, uma vez que demanda conhecimento sobre o domínio de ensino e suas simplificações em diferentes partes, e os possíveis relacionamentos entre essas partes. Devido a abrangência de um curso de graduação, tal delimitação pode se tornar uma tarefa multidisciplinar, dependendo, assim, da presença de vários especialistas (CARVALHO, 2010).

### 1.1 Motivação

O contexto apresentado no início desse capítulo descreve, de forma simplificada, a perspectiva da necessidade do processo de Planejamento Curricular e como esse caracteriza-se como complexo, já que sua realização demanda tempo e conhecimentos técnicos e

pedagógicos por parte dos profissionais responsáveis pela sua realização.

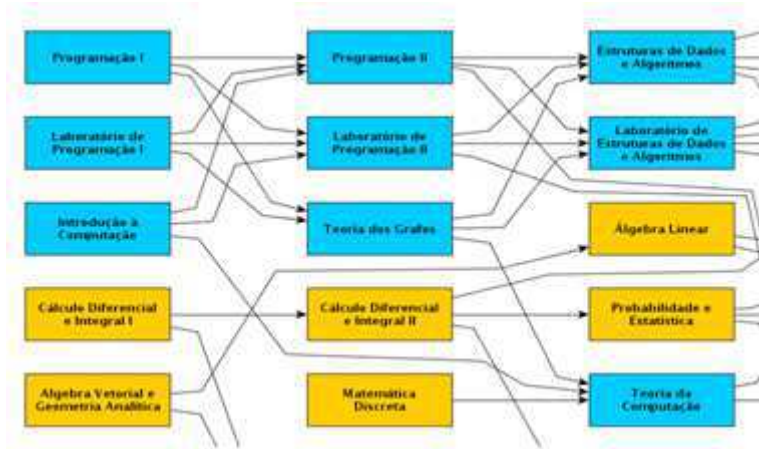
As estruturas curriculares dos cursos descrevem, de modo geral, uma visão do conhecimento educacional como sendo o insumo à ser entregue aos alunos durante o processo de aprendizagem, caracterizando, assim, “o quê” e “como” o aluno deve aprender. Por sua vez, esse conhecimento pode ser representado em diferentes granularidades. Tal fato, pode ser verificado nas diferentes áreas de conhecimento, tais como, Artes, Matemática, Português, dentre outras; nas quais observa-se que o conhecimento é organizado em partes menores, que podem apresentar uma sequência entre seus componentes.

Diante disso, verifica-se que o processo de Planejamento Curricular compreende uma atividade na qual o conhecimento maior (domínio) é particionado em unidades simplificadas (disciplinas) orquestradas de modo que se possa alcançar os objetivos de aprendizagem propostos. Comumente a representação desse conhecimento é dependente da presença de especialistas, porém, alguns domínios possuem documentos que descrevem a organização desse conhecimento, minimizando, assim, a necessidade da presença dos mesmos. Por outro lado, os documentos podem apresentar limitações como, por exemplo, a não delimitação dos relacionamentos entre as diferentes partes do conhecimento descrito neles.

No contexto do planejamento do ensino da computação no Brasil, verifica-se que a Sociedade Brasileira de Computação (SBC) disponibiliza o Currículo Referência (CR) para Ciência da Computação e Engenharia da Computação. Este documento descreve além de informações gerais sobre planejamento educacional, um conjunto de informações referentes ao corpo do conhecimento que pode ser utilizado no processo de planejamento das disciplinas que compõem as estruturas curriculares dos cursos de graduação no país. Porém, diferentemente das estruturas criadas, na qual é possível identificar relações entre as disciplinas (Figura 1.1), as informações descritas no CR não apresentam as possíveis relações de dependências entre as partes do conhecimento, expressando, assim, a ideia de sequenciamento entre esses componentes.

Assim, a motivação para a realização da pesquisa surgiu da observação da relevância do CR para o processo de Planejamento Curricular e, além disso, da inexistência dos relacionamentos entre as partes do conhecimento no CR. Deste modo, acredita-se que a identificação dessas relações possa auxiliar os processos de planejamento e avaliação das estruturas curriculares para verificação da conformidade com o “padrao” identificado.

Figura 1.1: Fragmento de uma Estrutura Curricular de um Curso de Bacharelado em Ciência da Computação



Fonte: Departamento de Sistemas de Computação (2010)

A partir das estruturas curriculares dos cursos existentes poderia ser induzida a orquestração das categorias do CR e serem determinadas as categorias mais representativas no contexto? Questionamento como esse foi determinante para as seguintes hipóteses de pesquisa:

- H1-0:** É possível orquestrar o Conhecimento Curricular descrito no CR a partir das estruturas curriculares de cursos reais;
- H1-1:** Não é possível orquestrar o Conhecimento Curricular descrito no CR a partir das estruturas curriculares de cursos reais;
- H2-0:** É possível identificar quais os Conhecimentos apresentados no CR são mais relevantes no contexto do Planejamento Curricular.
- H2-1:** Não é possível identificar quais os Conhecimentos apresentados no CR são mais relevantes no contexto do Planejamento Curricular.
- H3-0:** É possível utilizar a Orquestração entre Conhecimentos para auxiliar o processo de planejamento e avaliação de estruturas curriculares.
- H3-1:** Não é possível utilizar a Orquestração entre Conhecimentos para auxiliar o processo de planejamento e avaliação de estruturas curriculares.

A próxima seção apresenta os objetos de pesquisa estipulados para responder à questão de pesquisa e verificar as hipóteses formuladas anteriormente.

## 1.2 Objetivos

Considerando que o CR apresenta o conhecimento da área da Ciência da Computação distribuído em diferentes granularidades, porém, sem apresentar relações de dependências entre as essas partes, mas que, quando utilizado no contexto de Planejamento Curricular, são expressos os relacionamentos entre essas diferentes granularidades, o objetivo geral desta pesquisa consistiu na proposta de uma abordagem para orquestração do conhecimento curricular descrito no Currículo Referência.

A proposta visa, a partir da metodologia *bottom-up*, utilizar as estruturas curriculares existentes para orquestrar os conhecimentos descritos no CR e, além disso, utilizar o resultado desse processo de orquestração para auxiliar no planejamento e avaliação de estruturas curriculares existentes. A partir deste objetivo geral, foram elencados os seguintes objetivos específicos:

1. Realizar levantamento do estado da arte dos temas abordados nesta pesquisa;
2. Propor uma abordagem para a orquestração do Conhecimento Curricular em Ciência da Computação. Para isso, foram elencados os seguintes subobjetivos:
  - (a) Especificar o conjunto de categorias de conhecimento com base nas matérias descritas no Currículo Referência da SBC que deverão ser orquestradas;
  - (b) Definir um mapeamento entre os dados de cursos reais, obtidos a partir das ementas acadêmicas e/ou planos pedagógicos, e as categorias descritas no tópico anterior;
  - (c) Propor uma metodologia para a determinação das relações de dependências entre as categorias de conhecimento; e
  - (d) Propor métricas para mensurar o grau de ocorrência das categorias de conhecimento e das suas relações de dependências com outras categorias identificadas.

3. Propor a utilização do “padrão” de orquestração de conhecimento identificado para auxiliar o processo de planejamento e avaliação de estruturas curriculares; e
4. Avaliar a abordagem proposta e seu uso para planejamento e avaliação curricular a partir da realização de estudo de caso.

## 1.3 Relevância

A relevância da pesquisa é verificada a partir da complexidade apresentada pelo processo de Planejamento Curricular e pela inexistência das relações entre as partes do conhecimento descritos no CR que, se explicitados, poderiam auxiliar no processo de planejamento e avaliação das estruturas curriculares criadas para os cursos de graduação. Diante disso, verifica-se que a orquestração do conhecimento apresenta-se como relevante para a comunidade, em especial para àqueles envolvidos no processo de ensino da Ciência da Computação no Brasil.

Todo o conhecimento da Ciência da Computação utilizado neste trabalho baseia-se nas recomendações propostas pela SBC e descritas no CR05<sup>1</sup> (SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2005). Logo, verifica-se a relevância das informações apresentadas pelas Categorias de Conhecimento, denominação utilizada para representar o conhecimento descrito na granularidade “Matéria” do CR.

O uso do termo Orquestração fundamentou-se na ideia da realização de um arranjo. Este termo, comumente utilizado no contexto da educação *online* para modelar o sequenciamento da entrega de atividades aos alunos (KATSAMANI; RETALIS, 2013), nesta abordagem, foi utilizado para orquestrar as Categorias de Conhecimento do CR05. Ressalta-se que, diferentemente do CS2013<sup>2</sup> (ASSOCIATION FOR COMPUTING MACHINERY, 2013), o CR05 não apresenta relacionamentos cruzados entre disciplinas ou tópicos, o que possibilita a identificação das relações entre diferentes conhecimentos.

Neste trabalho, o processo de orquestração baseia-se nos dados contidos nas estruturas curriculares de IES brasileiras, apenas para disciplinas que possuem pressupostos teóricos. Dessa forma, espera-se que as relações identificadas representem ao máximo a organização

---

<sup>1</sup>Denominação utilizada para referenciar o Currículo Referência da SBC de 2005.

<sup>2</sup>Denominação utilizada para referenciar o *Computer Science Curricula* 2013.

do conhecimento no ensino da computação. Não se pode generalizar os relacionamentos identificados, porém, no contexto do estudo de caso realizado, representa o padrão de relacionamentos entre as Categorias de Conhecimento utilizado pelas IES.

O uso de um conjunto de dados reais para orquestrar as Categorias de Conhecimento, proporcionou o uso e técnicas de Mineração de Textos (FELDMAN; SANGER, 2007), especificamente da tarefa de Categorização de Textos (CT) para realizar o mapeamento entre as categorias e as estruturas curriculares dos cursos. Além disso, a CT reduziu a carga de trabalho demandada para o processo de mapeamento, quando comparado a realização do mesmo processo de forma manual. Para a categorização foi utilizada a ferramenta de mineração de dados WEKA<sup>3</sup>, conforme apresentado no Capítulo 4.

A mensuração do relacionamento entre duas Categorias de Conhecimento e, além disso, a identificação das categorias mais representativas no contexto do ensino em computação, respectivamente, foram identificadas a partir da mensuração do Grau de Dependência entre Categorias (*GDep*) e do Grau de Ocorrência das Categorias (*GOC*). Acredita-se que ambas as métricas podem fornecer informações úteis para o processo de planejamento e avaliação de estruturas curriculares.

Por fim, espera-se que esta pesquisa além de auxiliar os processos de planejamento e avaliação curricular, com base na identificação das métricas *GOC* e *GDep*, possa contribuir para a modelagem do domínio de sistemas que necessitam dessa representação de conhecimento, tais como, os Sistemas Tutores Inteligentes (STI).

## 1.4 Estrutura da Dissertação

Além deste capítulo introdutório, esta dissertação possui outros 5 (cinco) capítulos organizados da seguinte forma: no Capítulo 2 são apresentados os pressupostos teóricos referente à representação de conhecimento curricular, ao processo de orquestração e a tarefa de Categorização de Textos baseada em Mineração de Dados, essenciais para a compreensão da pesquisa.

No Capítulo 3 são apresentadas pesquisas identificadas na literatura que estão relacionadas à pesquisa ora apresentada.

---

<sup>3</sup>Do termo inglês, *Waikato Environment for Knowledge Analysis*.

---

No Capítulo 4 é apresentada a abordagem para Orquestração do Conhecimento Curricular proposta. Além disso, são apresentadas as propostas de uso das métricas identificadas pelo processo para auxiliar o planejamento e avaliação curricular.

No Capítulo 5 é apresentada a validação da abordagem a partir da realização de um estudo de caso, destacando, também a metodologia utilizada na condução do estudo.

No Capítulo 6 são apresentadas as considerações finais da pesquisa, bem como as limitações e propostas para pesquisas futuras identificadas durante a sua condução.



# Capítulo 2

## Fundamentação Teórica

Neste capítulo serão apresentados os pressupostos teóricos necessários para o entendimento da pesquisa. Na Seção 2.1 são apresentados os aspectos relacionados à Representação do Conhecimento Curricular. A fundamentação teórica sobre o processo de Orquestração é descrita na Seção 2.2. Posteriormente, a Seção 2.3, aborda conceitos e técnicas referentes à tarefa de Categorização de Textos baseada em Mineração de Dados Textuais. Por fim, na Seção 2.4, são apresentadas as considerações finais deste capítulo.

### 2.1 Representação do Conhecimento Curricular

O termo “Representação do Conhecimento Curricular” fundamenta-se nos termos “Conhecimento” e “Currículo” que, no escopo desta pesquisa, é utilizado para referenciar o conhecimento que deve ser transmitido aos aluno durante o processo de aprendizagem, ou seja, representa o conjunto de informações (conteúdos) que devem ser repassados ao aluno durante o processo de ensino.

A perspectiva do termo “Conhecimento” utilizado na fundamentação, baseia-se na conceituação apresentada por Takeuch e Nonaka (2004), que este definido como sendo um conjunto de informações organizadas em termos, fatos e premissas. O termo “Currículo” apresenta-se em diversas perspectivas, porém, foi utilizada a apresentada por Su (2012), que descreve o mesmo como sendo o resultado de um processo que envolve diferentes fatores, tais como: tempo, recursos de aprendizagem e de ensino, características dos alunos, objetivos de aprendizagem, conteúdos e seus sequenciamentos, dentre outros.

Um dos aspectos apresentados por Su (2012) e reforçado por Simons e Porter (2015) que deve ser considerado na composição do currículo, é o conteúdo curricular. Logo, a representação desses conteúdos e, conseqüentemente, do conhecimento inerente a esses, apresenta-se como importante, pois a qualidade do produto final desenvolvido (currículo) depende da qualidade das informações representadas. Ressalta-se que esta representação não faz referência à metodologia de ensino, ou seja, aos aspectos pedagógicos utilizados para entrega dos conteúdos, mas, sim, às informações repassadas.

A representação do conhecimento curricular pode ser verificada em ferramentas computacionais que visam auxiliar o processo de ensino. Essas ferramentas, por exemplo, os Sistemas Tutores Inteligentes<sup>1</sup> (STI) e Sistemas Hiperímia Adaptativos<sup>2</sup> (SHA) objetivam a entrega, de forma adaptativa, de uma seqüência de conteúdos educacionais ao aluno a partir das necessidades apresentadas por eles e objetivos de aprendizagem delimitados (GIRAFFA, 1998; MORNAR, 2001).

A adaptação e entrega dos conteúdos, nessas ferramentas, é possível devido a arquitetura básica desses sistemas possuírem um modelo de domínio com a finalidade de representar o conteúdo. Nesse modelo é representado o conhecimento a ser ensinado, em diferentes granularidades, e os relacionamentos entre eles. Segundo Filatro (2008) a definição dos relacionamentos é responsável pela seqüência de entrega. Comumente, a representação desse modelo é realizada por especialistas no domínio ou por meio de documentos que descrevam o conhecimento necessário (ver Subseção 2.1.2).

### 2.1.1 Formas de Representação do Conhecimento Curricular

Diferentes termos são utilizados para definir a Representação de Conhecimento Curricular. Todos fundamentam-se no princípio de representar um conjunto de informações que descrevam um domínio de conhecimento que possa auxiliar o processo de Planejamento Curricular<sup>3</sup>. Nesta pesquisa, serão apresentados alguns termos utilizados pela comunidade, tais como: Conteúdo Curricular, Conhecimento do Conteúdo e Corpo do Conhecimento.

---

<sup>1</sup>Do termo inglês, *Intelligent Tutoring Systems*

<sup>2</sup>Do termo inglês, *Adaptative Hiperímia System*

<sup>3</sup>Do termo inglês, *Curriculum Planning*.

### Conteúdo Curricular

O Conteúdo Curricular<sup>4</sup> é definido como o conhecimento que deve ser transmitido ao aluno durante o processo de aprendizagem. Considerado um fator importante, este conteúdo representa o meio para a transmissão do conhecimento de áreas ou assunto aos alunos (UNESCO, 2015). Além disso, Miel (1956) ressalta a importância do Conteúdo Curricular quando destaca o termo como assuntos, processos, abordagens e sentimentos que podem ser utilizados no Planejamento Curricular.

### Conhecimento do Conteúdo

O termo Conhecimento do Conteúdo<sup>5</sup> é definido como sendo um conjunto de informações sobre uma Área de Conteúdo<sup>6</sup>, ou seja, um domínio de conhecimento, que devem ser ensinadas pelo professor e organizadas de acordo com a necessidade e abrangência necessária (KOEHLER; MISHRA, 2009). Além disso, espera-se que os alunos tenham conhecimento sobre essas informações após o processo de aprendizagem (THE GLOSSARY OF EDUCATION REFORM, 2016).

Observa-se que a representação do Conhecimento do Conteúdo é um aspecto importante para o processo do Planejamento Curricular, uma vez que esta representação é utilizada no processo em questão. Logo, planejar essa representação pode ser considerado um fator de sucesso para a aprendizagem. Por sua vez, esse planejamento é modelado nos *frameworks Pedagogical Content Knowledge* (PCK) e no *Technological Pedagogical Content Knowledge* (TPACK) (KOEHLER; MISHRA, 2009), versão aprimorada do PCK que inclui o uso das Tecnologias de Informação e Comunicação (TIC) ao processo de planejamento educacional.

Conforme a Figura 2.1, o *framework* TPACK, selecionado para representar graficamente a metodologia de planejamento é composto pelo: Conhecimento Pedagógico<sup>7</sup>, o Conhecimento Tecnológico<sup>8</sup> e Conhecimento do Conteúdo, abordado nesta Subseção.

---

<sup>4</sup>Do termo inglês, *Curriculum Content*.

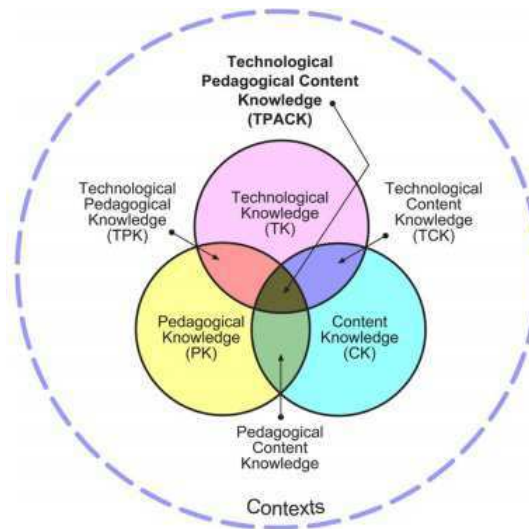
<sup>5</sup>Do termo inglês, *Content Knowledge - CK*.

<sup>6</sup>Do termo inglês, *Content Area*.

<sup>7</sup>Do termo inglês, *Pedagogical Knowledge - PK*.

<sup>8</sup>Do termo inglês, *Technological Knowledge - TK*.

Figura 2.1: Framework Conceitual TPACK.



Fonte: Koehler e Mishra (2009).

### Corpo do Conhecimento

O Corpo do Conhecimento<sup>9</sup> consiste em um conjunto de informações, comumente representado em documentos desenvolvidos por órgãos importantes nos respectivos domínios de aplicação, que visa organizar e descrever o conhecimento de uma área em diferentes níveis de abstração (granularidades).

As informações contidas nestes documentos podem ser utilizadas no planejamento educacional e curricular (FILATRO, 2008), pois tais especificações podem auxiliar tanto os especialistas, como também prover informações sobre o conhecimento à ser modelado quando na ausência dos especialistas.

Diversos domínios, tais como, a Ciência da Computação (SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2005; ASSOCIATION FOR COMPUTING MACHINERY, 2013), Sistemas de Informação (SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2003; ASSOCIATION FOR COMPUTING MACHINERY, 2010), Engenharia de Software (IEEE COMPUTER SOCIETY, 2014), Engenharia Civil (AMERICAN SOCIETY OF CIVIL ENGINEERS, 2008), entre outras; apresentam informações acerca do conhecimento que deve ser ensinado aos alunos em documentos contendo seu Corpo do Conhecimento.

Considerando o escopo deste trabalho, optou-se por apresentar apenas os documentos que descrevam o Corpo do Conhecimento do ensino da computação como atividade fim, ou

<sup>9</sup>Do termo inglês, *Body of Knowledge – BoK*.

seja, os cursos de bacharelado em Ciência da Computação, conforme detalhado na Subseção 2.1.2.

### 2.1.2 Currículos Referência da Ciência da Computação

Os Currículos Referência, de maneira geral, são documentos desenvolvidos por especialistas de um domínio de conhecimento que detalham um conjunto de informações visando auxiliar o processo de planejamento educacional e curricular do domínio em questão.

Comumente, esses documentos, além de apresentar o Corpo do Conhecimento, importante para o planejamento curricular, apresentam informações relevantes para o planejamento educacional. Dentre essas informações, são destacados aspectos relacionados ao perfil de formação do estudante, objetivos de aprendizagem, competências, dentre outras informações.

Ressalta-se que, não necessariamente, os CR devem ser utilizados na sua totalidade para o processo de Planejamento Curricular, pois, o conhecimento descrito nesses documentos visa prover um suporte para a modelagem de diferentes disciplinas, não limitando-as aos conteúdos especificados nos documentos.

No contexto do ensino em Ciência da Computação, observa-se que dois documentos apresentam-se como relevantes para a comunidade brasileira: o *Computer Science Curricula* e o Currículo Referência da SBC.

#### Computer Science Curricula

O *Computer Science Curricula*<sup>10</sup>, documento desenvolvido pela *Association for Computing Machinery* (ACM) em conjunto com o *IEEE Computer Society*, contém um conjunto de orientações curriculares para programas de graduação em Ciência da Computação (CC).

Com um histórico de versões CC2001<sup>11</sup>, CS2008<sup>12</sup> e CS2013<sup>13</sup> (ASSOCIATION FOR COMPUTING MACHINERY, 2001, 2008, 2013), este documento descreve um conjunto de informações que podem auxiliar o processo de Planejamento Curricular. Para isso, é

---

<sup>10</sup>Documento disponível em: <<http://goo.gl/oamkCM>>

<sup>11</sup>*Computer Science Curricula 2001*

<sup>12</sup>*Computer Science Curricula 2008*

<sup>13</sup>*Computer Science Curricula 2013*

apresentado o Corpo do Conhecimento que pode ser utilizado na estruturação curricular de cursos de graduação.

A representação do Corpo do Conhecimento é organizado em três granularidades (níveis de abstração): Áreas de Conhecimento<sup>14</sup>, Unidades de Conhecimento<sup>15</sup> e Tópicos. Assim, o mais alto nível (macro) de complexidade de conhecimento é apresentado nas Áreas de Conhecimento, sendo detalhado no nível meso pelas Unidades de Conhecimento, e estas detalhadas no nível micro pelos Tópicos (conteúdos).

O Corpo do Conhecimento apresentado no CS2013 é composto por 18 Áreas de Conhecimento (Quadro 2.1), detalhados em um conjunto de Unidades de Aprendizagem (Quadro 2.2) que, por sua vez, são detalhadas em um conjunto de Tópicos (Figura 2.2). Alguns tópicos apresentam relações cruzadas, possibilitando, assim, a identificação de relacionamento entre diferentes Áreas de Conhecimento.

Quadro 2.1: Áreas de Conhecimento do CC2013.

<b>Knowledge Area</b>
AL-Algorithms and Complexity
AR-Architecture and Organization
CN-Computational Science
DS-Discrete Structures
GV-Graphics and Visualization
HCI-Human-Computer Interaction
IAS-Information Assurance and Security
IM-Information Management
IS-Intelligent Systems
NC-Networking and Communication
OS-Operating Systems
PBD-Platform-based Development
PD-Parallel and Distributed Computing
PL-Programming Languages
SDF-Software Development Fundamentals
SE-Software Engineering
SF-Systems Fundamentals
SP-Social Issues and Professional Practice

<sup>14</sup>Do termo inglês, *Knowledge Areas* - KA

<sup>15</sup>Do termo inglês, *Knowledge Units* - KU

Quadro 2.2: Detalhamento da Área de Conhecimento *Algorithms and Complexity*.

AL. Algorithms and Complexity
AL/Basic Analysis
AL/Algorithmic Strategies
AL/Fundamental Data Structures and Algorithms
AL/Basic Automata, Computability and Complexity
AL/Advanced Computational Complexity
AL/Advanced Automata Theory and Computability
AL/Advanced Data Structures, Algorithms, and Analysis

Figura 2.2: Detalhamento da Unidades de Conhecimento *Basic Analysis* em Tópicos.**Topics:**

- Differences among best, expected, and worst case behaviors of an algorithm
- Asymptotic analysis of upper and expected complexity bounds
- Big O notation: formal definition
- Complexity classes, such as constant, logarithmic, linear, quadratic, and exponential
- Empirical measurements of performance
- Time and space trade-offs in algorithms
- Big O notation: use
- Little o, big omega and big theta notation
- Recurrence relations
- Analysis of iterative and recursive algorithms
- Some version of a Master Theorem

Fonte: Association for Computing Machinery (2013)

**Currículo Referência da SBC**

O CR05, assim como o CS2013, descreve um conjunto informações relacionadas ao contexto do ensino da Ciência da Computação. Tais informações, podem ser úteis para os processos de planejamento curricular e educacional dos cursos de graduação e, além disso, também são utilizadas como base para a delimitação de diretrizes curriculares de cursos de graduação (CONSELHO NACIONAL DE EDUCAÇÃO, 2012), na especificação dos conteúdos e competências que devem ser apresentados pelos estudantes de computação em testes como o ENADE (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2014) e do POSCOMP.

O Corpo do Conhecimento apresentado no CR05 é estruturado em três níveis de

granularidades: Núcleo, Matérias e Tópicos. Tal modelagem é semelhante a apresentada no CS2013, porém, os Tópicos do CR05 não apresentam as relações de referências cruzadas entre diferentes Áreas de Conhecimento, dificultando, assim, a identificação dos relacionamentos entre elas.

O detalhamento do conhecimento apresentado no CR05 é composto por 6 (seis) Núcleos, 58 (cinquenta e oito) Matérias e, associada a cada matéria, há um conjunto de Tópicos. Os núcleos e matérias descritas no CR05, são:

- **Matemática (M)**

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

- **Ciências Básicas (CB)**

- Física (CB1).

- **Eletrônica (E)**

- Circuitos Eletrônicos (E1).

- **Fundamentos da Computação (F)**

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

- **Tecnologia da Computação (T)**

- Análise de Desempenho (T1); Bancos de Dados (T2); Circuitos Integrados (T3); Compiladores (T4); Computação Gráfica (T5); Automação e Controle (T6); Engenharia de Software (T7); Inteligência Artificial (T8); Interação Humano-



Computador (T9); Matemática Computacional (T10); Métodos Formais (T11); Modelagem e Simulação (T12); Processamento Digital de Sinais (T13); Processamento de Imagens (T14); Programação Paralela (T15); Redes de Computadores (T16); Segurança e Auditoria de Sistemas (T17); Sistemas Digitais (T18); Sistemas Distribuídos (T19); Sistemas Embarcados (T20); Sistemas Multimídia (T21); Tolerância a Falhas (T22) e Telecomunicações (T23).

- **Contexto Social e Profissional (P)**

- Administração (P1); Computadores e Sociedade (P2); Comunicação e Expressão (P3); Contabilidade e Custos (P4); Direito e Legislação (P5); Economia (P6); Empreendedorismo (P7); Estágio (P8); Filosofia (P9); Informática na Educação (P10); Inglês (P11); Métodos Quantitativos Aplicados à Administração de Empresas (P12); Sociologia (P13) e Psicologia (P14).

Ressalta-se que, além de apresentar o conhecimento referente à Ciência da Computação nos núcleos Fundamentos da Computação e Tecnologia da Computação, o CR05 também apresenta um conjunto de conhecimentos relacionados à Matemática, Ciências Básicas, Eletrônica e Contexto Social e Profissional, visando contribuir para a construção de um conhecimento sólido não apenas da computação, mas também, da formação adicional que estudante deve apresentar.

## 2.2 O Processo de Orquestração

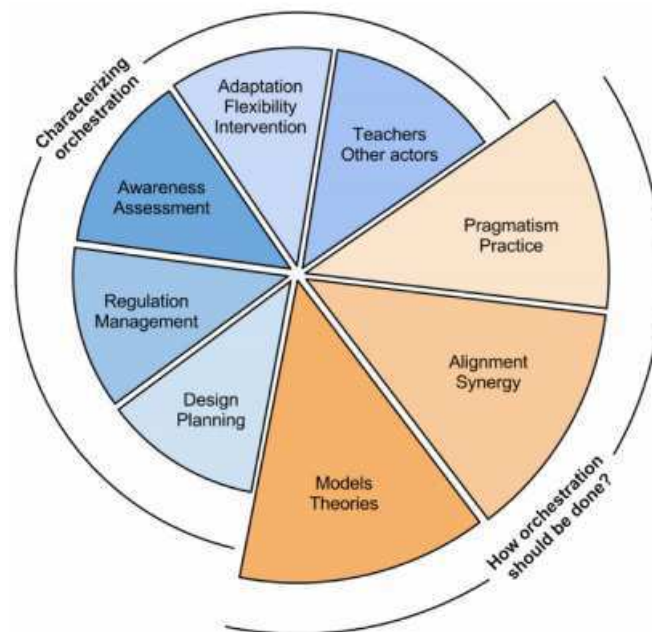
O termo “Orquestração” conceitualmente é definido como o ato de orquestrar, ou seja, de combinar ou harmonizar elementos de modo que se tenha um efeito desejado (DICIONÁRIO MICHAELIS, 2016). Tal termo, comumente, é utilizado no contexto musical, mais especificamente para fazer referência a uma orquestra, porém, ele também apresenta aplicações no contexto educacional.

Na perspectiva educacional, este termo é utilizado pelas áreas da *Technology-Enhanced Learning* (TEL) e *Computer-Supported Collaborative Learning* (CSCL) e seu uso, de modo geral, está relacionado ao processo de gerenciamento das atividades de aprendizagem, em

tempo real, que conduza o estudante ao sucesso (ROSCHELL; DIMITRIADIS; HOPPE, 2013; DILLENBOURG, 2011).

Em uma visão geral, o processo de orquestração visa arranjar recursos e aspectos educacionais com a finalidade de alcançar as metas e objetivos propostos durante o planejamento educacional (SANTOS, 2010). Este arranjo, ainda pode ser dividido em duas perspectivas e diferentes aspectos, compondo, assim, o framework conceitual do processo de orquestração.

Figura 2.3: Framework Conceitual do Processo de Orquestração da Aprendizagem.



Fonte: Prieto et al. (2011)

Na estrutura do framework conceitual da orquestração (Figura 2.3) é possível verificar um conjunto de diferentes aspectos organizados em duas perspectivas que descrevem as características do processo, e como ele deve ser realizado. Os aspectos relacionados à perspectiva que caracteriza o processo de orquestração são:

- *Design/Planning* – Este aspecto refere-se ao processo de planejar as atividades de aprendizagem que serão orquestradas durante o processo de ensino. Para isso, podem ser utilizados aspectos de *Design* Instrucional (FILATRO, 2008), com destaque para a modelagem do fluxo das atividades;

- *Regulation/management* – O processo de orquestração visa maximizar o sucesso do aluno durante o processo de aprendizagem. Logo, se faz necessário a realização de um processo de gerenciamento e regulação do processo. Tal regulação pode basear-se em fatores externos ao processo ou internos a ele, caracterizando uma autorregulação;
- *Adaptation/flexibility/intervention* – Refere-se à necessidade de que o processo de orquestração seja adaptativo, de modo que o plano ou *design* desenvolvido esteja de acordo com o contexto da aprendizagem;
- *Awareness/assessment* – Este aspecto está relacionado à delimitação e especificação do processo de avaliação do processo de aprendizagem, de modo que os envolvidos tenham ciência do funcionamento do processo, e
- *Roles of the teacher and other actors* – Neste aspecto são definidos os papéis dos atores do processo de orquestração, com a finalidade de especificar e delegar funções aos mesmos.

Verifica-se também que o processo de orquestração de aprendizagem além dos aspectos apresentados no framework, faz referência a fatores como (DILLENBOURG, 2011):

- Currículo – Tem a finalidade de descrever “o quê” deve ser entregue aos alunos durante o processo de aprendizagem, norteando, assim, as decisões que serão tomadas pelo professores;
- Conteúdos – Refere-se a forma de entrega efetiva da instrução aos alunos, com destaque para as propriedades do conteúdo, didática, e o *design* de entrega;
- Como os alunos aprendem - Visa observar a forma com que os alunos aprendem, identificando aspectos relacionados à cognição, às questões motivacionais, às dificuldades e limitações apresentadas pelos alunos, dentre outros; e,
- Alunos – serão os atores do processo que receberão as instruções, conteúdos e conhecimentos do processo de aprendizagem. Logo, se faz necessária a sua observação e análise.

Nesta pesquisa, foi observado o aspecto de *design/planning* de modo a verificar a forma de organização do currículo no processo de ensino e aprendizagem.

## 2.3 Categorização de Textos Baseada em Mineração de Textos

A Mineração de Textos<sup>16</sup> (MT) pode ser descrito como um processo que visa extrair informações úteis a partir de conjuntos de dados textuais. Diferentemente do processo de Mineração de Dados<sup>17</sup> (MD), que utiliza dados estruturados e armazenados em banco de dados, os dados utilizados na MT são não-estruturados, representados nos textos dos documentos (FELDMAN; SANGER, 2007; ORKESON, 2013).

Análogo ao processo de MD, a MT visa extrair informações úteis de fonte dados de 2 (duas) formas: baseada na análise semântica, na qual é verificado o significado dos termos contidos no texto; e baseada na análise estatística que baseia-se na frequência com que os termos aparecem no texto. Independente do tipo de análise, o processo de MT pode ser dividido em 4 (quatro) fases (GONÇALVES, 2002), sucintamente descritas como:

- **Coleta de Documentos** – Etapa responsável pela coleta dos documentos relacionados ao contexto da aplicação que serão analisados;
- **Pré-processamento** – Etapa na qual são realizadas ações com a finalidade de preparar o conjunto de documentos coletados na fase anterior para a extração do conhecimento;
- **Extração de Conhecimento** – Nesta etapa ocorre a extração do conhecimento a partir dos documentos pré-processados, e
- **Avaliação e Interpretação dos Resultados** – Esta etapa tem a finalidade de avaliar os resultados obtidos na fase de extração com o objetivo de facilitar o entendimento dos mesmo e, assim, auxiliar em uma possível tomada de decisão.

Ressalta-se que, para Balakrishnan e Kaniimozhi (2015), além de técnicas de MD, a MT utiliza-se das técnicas de Recuperação da Informação (RI), Processamento de Linguagem Natural (PLN) e Extração da Informação (EI). Cada área contribui de alguma forma, por exemplo, a RI auxilia a recuperação de documentos em uma coleção, colaborando, assim, para a coleta dos documentos que serão utilizados; o PLN e a EI auxiliam no processo de

---

<sup>16</sup>Do termo inglês, *Text Mining*.

<sup>17</sup>Do termo inglês, *Data Mining*.

pré-processamento dos documentos coletados, gerando, assim, os dados que serão utilizados na MD para extração do conhecimento.

Na MT a Extração do Conhecimento é realizada a partir de tarefas que consistem na aplicação de algoritmos para a identificar a informação desejada. Diferentes tarefas podem ser aplicados aos dados pré-processados, tais como: a tarefa de agrupamento de textos<sup>18</sup>, de Mineração de Conceitos<sup>19</sup>, Recuperação da Informação<sup>20</sup>, Extração da Informação<sup>21</sup> e a tarefa de Categorização de Textos<sup>22</sup>. Apenas a tarefa de Categorização de Textos foi detalhada, conforme Subseção 2.3.1.

### 2.3.1 Tarefa de Categorização

A Categorização de Textos (CT), também conhecida como Classificação de Textos, é uma tarefa da MT que visa classificar dados em um conjunto de categorias pré-especificadas (VIDHYA; AGHILAI, 2010).

Segundo Feldman e Sanger (2007), a CT é formalmente descrita por uma função  $F : D \times C \rightarrow \{0, 1\}$ , onde  $D$  representa uma coleção de documentos possíveis e  $C = \{c_1, \dots, c_{|m|}\}$  representa o conjunto de categorias pré-definidas. Partindo desse conceito inicial, para cada par  $(d_i, c_j) \in F \times C$  será atribuído valor 1 quando o documento  $d_i$  pertencer a categoria  $c_j$ , e 0 nos casos em que o documento não pertence a categoria.

Igualmente a outras técnicas de Inteligência Artificial (IA), a CT apresenta duas abordagens para a sua realização. A primeira forma, baseia-se no processo descrito na Engenharia do Conhecimento<sup>23</sup>, no qual um conjunto de regras são definidas por um especialista do domínio e estas são utilizadas no processo de categorização, por sua vez, esse processo é considerado custoso por demandar tempo e conhecimento para a definição das regras de categorização. A segunda forma, fundamenta-se na utilização de técnicas de Aprendizagem de Máquina<sup>24</sup> objetivando a identificação das categorias de forma automática.

Segundo Sebastiani (2002), as diferentes abordagens destacam as fases do histórico do

---

<sup>18</sup>Do termo inglês, *Text Clustering*

<sup>19</sup>Do termo inglês, *Concept Mining*

<sup>20</sup>Do termo inglês, *Information Retrieval*

<sup>21</sup>Do termo inglês, *Information Extraction*

<sup>22</sup>Do termo inglês, *Text Categorization*

<sup>23</sup>Do termo inglês, *Knowledge Engineering - KE*

<sup>24</sup>Do termo inglês, *Machine Learning - ML*

CT, uma vez que o estudo dessa tarefa iniciou-se na década de 60. Na década de 80 surgiram os estudos relacionando a CT à Engenharia do Conhecimento, já a associação entre CT a técnica de Aprendizagem de Máquina (ver Subseção 2.3.3) só começou a ser estudada na década de 90.

### Tipos de Categorizadores

Segundo Sebastiani (2002) e Feldman e Sanger (2007), os categorizadores podem apresentar diferentes comportamentos, que variam de acordo com o contexto de aplicação. Neste caso, a função  $F$  da CT pode categorizar documentos de duas formas: em monocategorias<sup>25</sup> ou em multicategorias<sup>26</sup>.

A categorização de forma monocategorial, significa que ao documento  $d_i \in D$  é atribuída apenas uma categoria pré-definida no conjunto  $c_j \in C$ . Por outro lado, quando a categorização é realizada de forma multicategorial pode ser atribuído ao documento  $d_i \in D$  mais de uma categoria  $c_j \in C$ .

Além disso, a CT pode ser modelada de duas formas: Categorização Orientada a Documentos<sup>27</sup> e Categorização Orientada a Categoria<sup>28</sup>. A modelagem Categorização Orientada a Documentos ocorre quando a categorização dos documentos se dá a partir do documento, ou seja, a partir de um documento  $d_i \in D$  é identificada a categoria  $c_j \in C$  do mesmo. Por outro lado, a Categorização Orientada à Categoria ocorre de forma inversa, ou seja, a partir de uma  $c_j \in C$  são identificados todos os documentos  $d_i \in D$ .

### 2.3.2 Pré-processamento de Documentos

Esta etapa corresponde à fase de preparação dos documentos para que seja possível realizar a Extração do Conhecimento.

---

<sup>25</sup>Do termo inglês, *single-label categorization*

<sup>26</sup>Do termo inglês, *multilabel categorization*

<sup>27</sup>Do termo inglês, *document-pivoted categorization*

<sup>28</sup>Do termo inglês, *category-pivoted categorization*

### Tokenização

O processo de Tokenização<sup>29</sup> visa particionar o texto dos documentos em unidades mínimas, denominado de tokens. Esses tokens, representam as palavras do texto que será utilizado para extrair o conhecimento, portanto, estes devem apresentar significado semelhante ao texto original. Ressalta-se que esse processo é facilitado devido às palavras nos textos apresentarem caracteres de formatação, tais como espaços ou sinais, que podem ser utilizados como delimitadores (FELDMAN; SANGER, 2007).

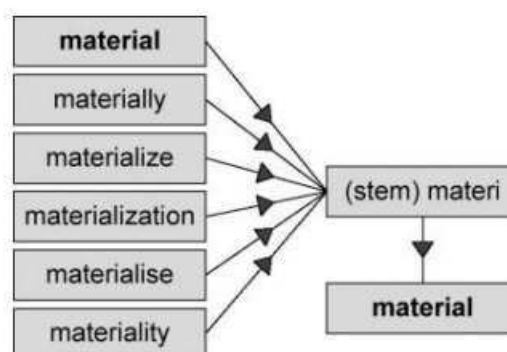
### Remoção das Stopwords

As *Stopword* correspondem a um conjunto de palavras comuns à qualquer linguagem, não apenas inglês. As palavras contidas nesse conjunto são caracterizadas pela alta frequência no idioma. Portanto, a retirada destes termos não traz impactos ao resultado final da análise realizada. Geralmente, fazem parte desse conjunto as preposições, artigos, adjetivos, conjunções (GANESAN, 2016; CHOY, 2012).

### Stemming

O método de stemming ou lematização visa a redução dos termos extraídos do documento. Para isso, os termos sofrem mudanças que visam reduzi-los aos seu radical (RAMASUBRAMANIAN; RAMYA, 2013), conforme a Figura 2.4.

Figura 2.4: Processo de Stemming.



Fonte: Ramasubramanian e Ramya (2013).

<sup>29</sup>Do inglês, *Tokenization*

O funcionamento deste método baseia-se na eliminação das variações morfológicas dos termos, com isso, são removidas vogais temáticas, desinências, prefixos e sufixos dos termos, sendo mais comum a retirada dos sufixos (SILVA, 2004).

### **Corte de Palavras Baseado na Frequência**

Refere-se a outra forma de redução de dimensionalidade que é utilizada para identificar os termos (tokens) mais representativos dentre os existentes, baseando-se na frequência de vezes que o termo aparece. Deste modo, após a contagem dos termos é estipulado um ponto de corte e, a partir desse, são identificados os termos que serão descartados e aqueles que devem ser analisados (SOARES; PRATI; MONARD, 2008).

### **Valor do Atributo**

Cada documento  $d_i \in D$  corresponde a um vetor  $d_i = (a_{i,1}, a_{i,2}, \dots, a_{i,M})$ , no qual o valor de  $a_{i,j}$  refere-se ao  $j$ -ésimo termo (token) do documento  $d_i$ . O valor desse termo pode ser calculado de diferentes medidas que consideram a frequência do termo no documento. Dentre as diferentes medidas, optou-se, neste trabalho, por utilizar a medida *Term Frequency - Inverse Document Frequency (tf-idf)* que, sucintamente, pondera a frequência dos termos de uma coleção de documentos, de modo que os termos que aparecem com mais frequências tenham um peso de representação menor e, conseqüentemente, os termos com menor frequência apresentam um peso maior de representação (SOARES; PRATI; MONARD, 2008).

### **2.3.3 Aprendizagem de Máquina**

Conforme descrito anteriormente, a CT também pode ser realizada com o uso das técnicas de Aprendizagem de Máquina. Nestes casos, essas técnicas proporcionam a criação de um classificador automático a partir de um conjunto treinamento contendo uma coleção de registros pré-categorizados. Por sua vez, esse conjunto treinamento é utilizado na aprendizagem do categorizador, consistindo, assim, no processo de aprendizagem supervisionada (FELDMAN; SANGER, 2007).

A aprendizagem do categorizador proporcionada pelo uso do conjunto treinamento



precisa ser avaliada. Tal avaliação visa assegurar a acurácia da aprendizagem e, conseqüentemente, do processo de categorização que será realizado. As diferentes formas de avaliação dessa aprendizagem são descritas na próxima Subseção.

### **Avaliação da Aprendizagem**

De maneira geral, a avaliação da aprendizagem visa assegurar a qualidade da aprendizagem do categorizador. Diferentes formas podem ser utilizadas para essa verificação, tais como:

- *Ressubstituição* – nessa mensuração todos os dados utilizados na aprendizagem (treinamento) são utilizados para os testes. Essa medida apresenta-se como falsa, pois, comumente apresenta valores de acurácia elevados devido a utilização de todo o conjunto treinamento para testes.
- *Holdout* – neste método o conjunto treinamento é particionado em porcentagens fixas, comumente, 2/3 do conjunto para treinamento e 1/3 para teste. Ressalta-se que esses valores podem ser modificados, uma vez que um bom treinamento é dependente da quantidade de instâncias utilizadas.
- *Cross-validation* – este método visa assegurar a representatividade máxima das classes de teste e treinamento. Para isso, o conjunto treinamento é particionado aleatoriamente em  $k$  partes mutuamente exclusivas, que são utilizadas para treinamento e teste.

Para avaliar os resultados gerados com o objetivo de verificar os erros e acertos, para cada documento categorizado, é gerada uma Matriz  $M_{n,n}$ , conhecida como Matriz de Confusão, conforme Figura 2.5. As linhas e colunas dessa matriz correspondem aos elementos do conjunto  $C = \{c_1, c_2, \dots, c_n\}$  correspondente às categorias utilizadas na tarefa de CT. Deste modo, cada elemento  $M(c_i, c_j)$  representa o número de exemplos da categoria  $c_i$  que foram categorizados como sendo da classe  $c_j$ . Assim, entende-se que a diagonal  $M(c_i, c_j)$  corresponde ao número de acertos, já os demais valores apresentam-se como erros.

### **Algoritmos de Categorização**

A literatura apresenta que diferentes algoritmos de Aprendizagem de Máquina podem ser utilizados na tarefa de CT. Nesta Subseção serão detalhados 3 (três) destes algoritmos, o algoritmo de Redes Bayesianas, de Árvores de Decisão e *Support Vector Machine*.

Figura 2.5: Representação de uma Matriz de Confusão.

Classe	Predita $C_1$	Predita $C_2$	...	Predita $C_k$
verdadeira $C_1$	$M(C_1, C_1)$	$M(C_1, C_2)$	...	$M(C_1, C_k)$
verdadeira $C_2$	$M(C_2, C_1)$	$M(C_2, C_2)$	...	$M(C_2, C_k)$
⋮	⋮	⋮	⋮	⋮
verdadeira $C_k$	$M(C_k, C_1)$	$M(C_k, C_2)$	...	$M(C_k, C_k)$

O algoritmo probabilístico de **Redes Bayseanas** baseia-se na verificação da probabilidade de que  $c_j \in C$  seja da categoria do documento  $d_i \in D$ . Tal verificação, é fundamentada no teorema de Bayes (FELDMAN; SANGER, 2007), conforme Equação 2.1:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \quad (2.1)$$

Dado que cada documento  $d_i$  é composto por um vetor  $d = (w_1, w_2, \dots, w_n)$ , e que a Rede Bayseana fundamenta-se na suposição de que esses termos são condicionalmente independente, o cálculo da probabilidade da categorização deve ser realizado para cada token representado no vetor. Logo, o valor de saída da função considera que a probabilidade  $P(d|c)$  é igual ao produto da probabilidade de cada termo do vetor de forma independente, conforme Equação 2.2:

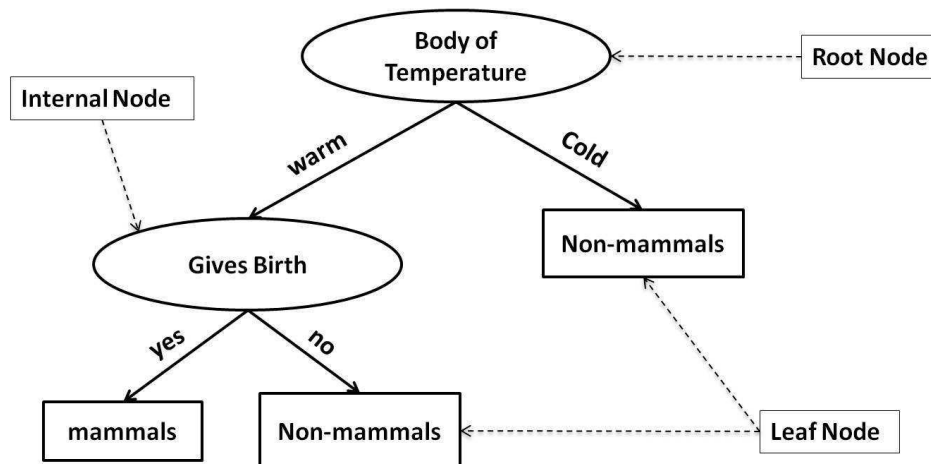
$$P(d|c) = \prod_i P(w_i|c) \quad (2.2)$$

O algoritmo de **Árvore de Decisão** é uma estrutura de dados recursiva, onde os nós indicam o rótulo de uma classe ou um nó decisão que contém um teste baseado nos valores dos atributos de entrada. Cada nó da árvore representa um teste ou uma decisão, cujo o resultado são as suas sub-árvores (GONÇALVES, 2002).

Na Figura 2.6 é apresentado um exemplo de um problema de classificação utilizando Árvores de Decisão. Neste exemplo, para a classificação de mamíferos ou não, é possível observar que primeiramente, um teste é realizado no nó raiz (*root node*), e dependendo do resultado esses testes se ramificarão para a direita ou esquerda, onde serão realizados novos testes nos nós internos (*internal nodes*), até que o nó folha ou terminal (*leaf node*) seja identificado, identificando, assim, a solução para o problema da categorização.

O *Support Vector Machine* (SVM) é um algoritmo considerado rápido e efetivo para

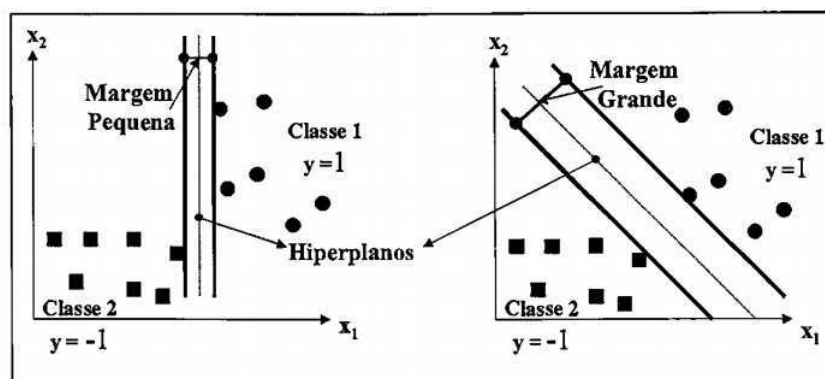
Figura 2.6: Árvore de Decisão para o Problema da Classificação de Mamíferos.



Fonte: Adaptado de Tan, Steinbach e Kumar (2006)

o problema de CT (FELDMAN; SANGER, 2007). De modo geral, esse algoritmo visa identificar um hiperplano que separe um conjunto de dados em classes distintas (GONÇALVES, 2002), conforme Figura 2.7.

Figura 2.7: Support Vector Machine.



Fonte:(GONÇALVES, 2002)

O funcionamento e entendimento desse algoritmo é facilitado quando os dados são linearmente separáveis, conforme a Figura 2.7 onde é possível verificar as Classes 1 e 2 como sendo linearmente separáveis. Porém, existem casos em que os dados não são apresentados desta forma, nestes casos, o hiperplano identificado deve ser aquele que apresente a menor taxa de erro.

## **2.4 Considerações Finais do Capítulo**

Este capítulo apresentou temas importantes para o entendimento da abordagem proposta na pesquisa. Para isso, inicialmente, ressaltou-se a importância da representação do conhecimento curricular e como o processo de orquestração pode auxiliar no processo de organização das partes desse conhecimento. Posteriormente, foram apresentados os Currículos Referência, objetivando destacar a importância do corpo do conhecimento descrito nestes documentos que, por vezes, não são organizados entre si, de modo que possam ser identificadas as relações entre as partes que compõem o conhecimento descrito nos documentos. Por fim, foi apresentada a tarefa de Categorização de Texto baseada em Mineração de Textos, essencial para a realização do mapeamento entre a estrutura curricular dos cursos e o corpo do conhecimento apresentado no CR.

# Capítulo 3

## Pesquisas Relacionadas

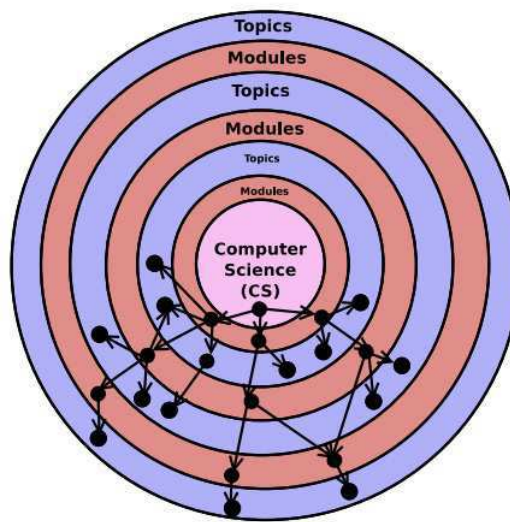
Este capítulo apresenta, com base na literatura identificada, pesquisas que relacionam-se com a descrita neste trabalho. Para isso, na Seção 3.1 são apresentadas as pesquisas relacionadas ao processo de Orquestração do Conhecimento Curricular. Na Seção 3.2 são descritas pesquisas que abordam a problemática do planejamento e avaliação curricular. Na Seção 3.3 são destacadas pesquisas relacionadas à tarefa de categorização de textos. Por fim, a Seção 3.4 apresenta as considerações finais deste capítulo.

### 3.1 Orquestração do Conhecimento Curricular

Conforme apresentado no capítulo anterior, a ideia de Orquestração do Conhecimento Curricular corresponde ao arranjo das diferentes partes do conhecimento que deve ser ensinado ao aluno durante sua formação acadêmica. Ressalta-se que essas partes do conhecimento correspondem as diferentes granularidades de representação do mesmo, tais como: módulos, tópicos, atividades, entre outros.

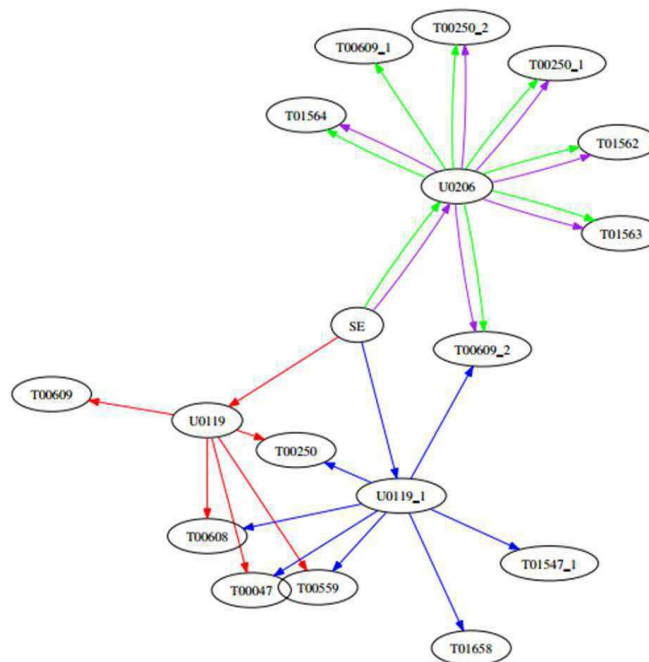
A orquestração do conhecimento diferentes granularidades foi apresentado por Marshall (2014) quando, em sua pesquisa, foi proposta uma organização curricular composta por elementos de diferentes granularidades com relacionamentos entre si. Tal ideia foi exemplificada utilizando as especificações descritas no Corpo do Conhecimento do CS2013, com destaque para as diferentes granularidades de representação e, além disso, pela especificação das relações de pré-requisito entre as diferentes granularidades, conforme verificado na Figura 3.1.

Figura 3.1: Representação das Relações entre as Diferentes Granularidades do Currículo CS2013



Fonte: Marshall (2014)

Figura 3.2: Representação de Dependências Cruzadas entre o Conhecimento



Fonte: Marshall (2014)

Em Marshall (2014) também foi detalhado a ideia de dependências cruzadas entre diferentes conhecimento, explicitando, assim, a multidisciplinaridade existente nos currículos. Para isso, foi utilizada a granularidade em nível de Tópicos, conforme Figura 3.2, na qual é possível verificar a existência de relações de dependências cruzadas entre

diferentes partes de conhecimento.

Diferentemente dos exemplos apresentados, existem casos em que o processo de orquestração é aplicado às atividades de aprendizagem, como apresentado por Katsamani e Retalis (2013). Neste caso, além do conhecimento curricular podem ser abordados detalhes pedagógicos relacionados à forma de transmissão de determinado conhecimento aos alunos, contribuindo, assim, para o processo de gerenciamento da aprendizagem, em especial, das atividades de aprendizagem.

Diante contexto apresentado, observa-se que o processo de Orquestração pode ser aplicável à diferentes contextos que necessitem de um arranjo de suas partes, porém, não há uma metodologia que descreva como este deve ser realizado. Assim, verifica-se a importância da explicitação de uma metodologia que proporcione um suporte para a realização dessa atividade, conforme proposto neste trabalho.

## **3.2 Planejamento e Avaliação Curricular**

Os processos de planejamento e avaliação curricular apresentados pelos trabalhos descritos nessa seção baseiam-se no uso de documentos referências, em especial, nos currículos referência para sua realização.

O uso do CR como suporte ao processo de planejamento curricular e educacional é apresentado em diversos Plano Pedagógicos Curriculares, conforme descrito na Fundamentação Teórica deste documento. Porém, existem na literatura algumas pesquisas que utilizam esses documentos como um importante insumo para o processo de Planejamento Curricular devido a descrição de um conjunto de diretrizes que podem ser utilizadas no processo em questão. Um exemplo desse uso é apresentado por Batista et al. (2015) que utilizou o CR para reestruturar um curso de graduação em Ciência da Computação. Para isso, foram realizadas algumas análises entre as disciplinas do curso e os núcleos de conhecimento descrito no CR, na qual, foi possível detectar algumas inconsistências que culminaram na mudanças propostas para as grades curriculares existentes.

Por outro lado, há pesquisas que utilizam-se das recomendações descritas no CR para avaliar as estruturas curriculares dos cursos existentes e, assim, auxiliar o processo

de planejamento. Um exemplo desse uso é apresentado por Lousnak et al. (2002), no qual os autores realizam um conjunto de análises das estruturas curriculares de cursos de graduação em Ciência da Computação com base nas recomendações descritas no CR05 com a finalidade verificar o comportamento dos cursos perante às diretrizes propostas pela SBC. Como resultado dessas análises, foram identificados, dentre os cursos observados, diferentes comportamentos (distribuição de disciplinas) entre os diferentes núcleos de conhecimento e, desta forma, foi possível identificar casos onde o ensino da Matemática, área complementar ao ensino da computação, apresentava uma quantidade de créditos superior às disciplinas dos núcleos de computação (Fundamentos da Computação e/ou Tecnologia da Computação).

Também foram identificados estudos que não fundamentavam-se diretamente no CR, mas que apresentam uma organização curricular semelhante aos núcleos descritos no Currículo Referência. Um exemplo desses estudos foi apresentado por Gerab, Bueno e Gerab (2014) que, através da aplicação de técnicas de mineração de dados, conseguiu identificar um conjunto de áreas de conhecimento em uma estrutura curricular, identificando, assim, as áreas de hipermídia, Sistemas e Informação, Banco de Dados, Compiladores, Humanidade, dentre outras.

### **3.3 Categorização de Texto**

Diversos trabalhos relacionados à Categorização de Textos podem ser encontrados na literatura, uma vez que esta é uma área bastante estudada pela comunidade. Porém, não foram identificadas pesquisas que se relacionam diretamente com a categorização de ementas acadêmicas. Logo, serão apresentadas pesquisas em outros contextos de aplicação.

Um exemplo de aplicação da CT a ementas não acadêmicas, foi apresentado por Ferauche (2011), que categorizou (classificou) ementas em um contexto judicial. Para tanto, foram utilizadas técnicas de Mineração de Textos (MT) para categorizar as ementas da Jurisprudência da Justiça do Trabalho de São Paulo, por assunto. Como resultado, foi mostrado que a aplicação das técnicas de MT realmente auxiliam o processo de categorização, porém, no escopo do estudo, não foi possível identificar qual a melhor das técnicas utilizadas e, neste caso, optou-se pela junção delas objetivando construir um classificador mais acurado.



A Categorização de Textos também teve suas técnicas aplicadas ao contexto empresarial, conforme mostra a pesquisa de Rizzi et al. (2000), que apresenta o uso da MT como um meio para obtenção de informações estratégicas, a partir de fontes textuais e, assim, acredita-se que as informações coletadas possam apresentar vantagens competitivas para as empresas.

### 3.4 Considerações Finais do Capítulo

Esse capítulo apresentou, de forma sucinta, algumas pesquisas que relacionam-se, mesmo que indiretamente, com a proposta de orquestração ora citada. Dessa forma, objetivou-se destacar o contexto do uso das informações descritas nos CR como fonte de informação para o processo de planejamento e, além disso, descrever o processo de orquestração do conhecimento descrito nestes documentos. Foram ressaltados alguns trabalhos utiliza o CR para o processo de planejamento e avaliação curricular, e, por fim, foi destacado a CT como uma técnica viável para a identificação de informações úteis em documentos textuais.

Quadro 3.1: Quadro de Pesquisas Relacionadas

	Orquestração do Conhecimento	Avaliação Curricular	Planejamento Curricular	Categorização de Textos	Uso do Currículo Referência
Rizzi et al. (2000)				x	
Ferauche (2011)				x	
Gerab, Bueno e Gerab (2014)			x		
Lousnak et al. (2002)		x			x
Batista et al. (2015)			x		x
Marshall (2014)	x				
Katsamani e Retalis (2013)	x				
<b>Abordagem Proposta</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>	<b>x</b>

O Quadro 3.1, destaca a abrangência de cada pesquisa relacionada com os fatores abrangidos no escopo dessa pesquisa (Orquestração do Conhecimento, Avaliação Curricular, Planejamento Curricular, Categorização de Textos e Uso do Currículo Referência). A última linha do Quadro 3.1 mostra o escopo da abordagem de orquestração do conhecimento descrito no Capítulo 4. Assim, observa-se que a proposta descrita nesse trabalho fundamenta-se na junção de diferentes fatores, coberto por pesquisas distintas, mas que quando juntas, podem fornecer um conjunto de informações relevantes para a comunidade da educação em computação.

# Capítulo 4

## Orquestração do Conhecimento

### Curricular em Ciência da Computação

Este capítulo está diretamente relacionado ao objetivo geral desta pesquisa, uma vez que será apresentada a abordagem proposta para a orquestração do conhecimento curricular em Ciência da Computação e, posteriormente, a proposta de uso das métricas identificadas no processo de orquestração para auxiliar no planejamento e avaliação de estruturas curriculares.

#### 4.1 Descrição da Abordagem

Conforme descrito na Motivação e na Fundamentação Teórica, o conteúdo curricular é um dos aspectos que devem ser considerados durante o processo de Planejamento Curricular e/ou Planejamento Educacional.

Neste trabalho, o agrupamento dos conteúdos a ser ensinado foi denominado de “Conhecimento Curricular”, devido a sua finalidade de descrever “o quê” o aluno deve aprender no processo de ensino-aprendizagem, representando, assim, o conhecimento a ser ensinado pelo professor. Tais agrupamentos baseiam-se na granularidade “Matéria” do CR05.

Verificou-se, no contexto do ensino da Ciência da Computação, a existência de documentos que descrevem um conjunto de orientações educacionais que visam auxiliar os processos de planejamento educacional e curricular. Tais documentos, por exemplo, CS2013

e o CR05 detalham, em diferentes níveis de abstração, o corpo do conhecimento que deve ser transmitido ao aluno. Por sua vez, esse corpo detalha um conjunto de conteúdos que devem ser abordados nas disciplinas dos cursos de graduação. Ressalta-se que estes documentos descrevem apenas recomendações que podem ou não serem acatadas em sua totalidade.

Ao analisar os documentos CS2013 e CR05, verificou-se semelhanças quando observado a organização do corpo de conhecimento dos mesmos. Ambos detalham o conhecimento da Ciência da Computação em 3 (três) diferentes níveis de abstração (granularidades), o CS2013 detalha o corpo do conhecimento nas granularidades: Área de Conhecimento, Unidade de Conhecimento e Tópicos, que respectivamente correspondem às granularidades Núcleos, Matérias e Tópicos do CR05.

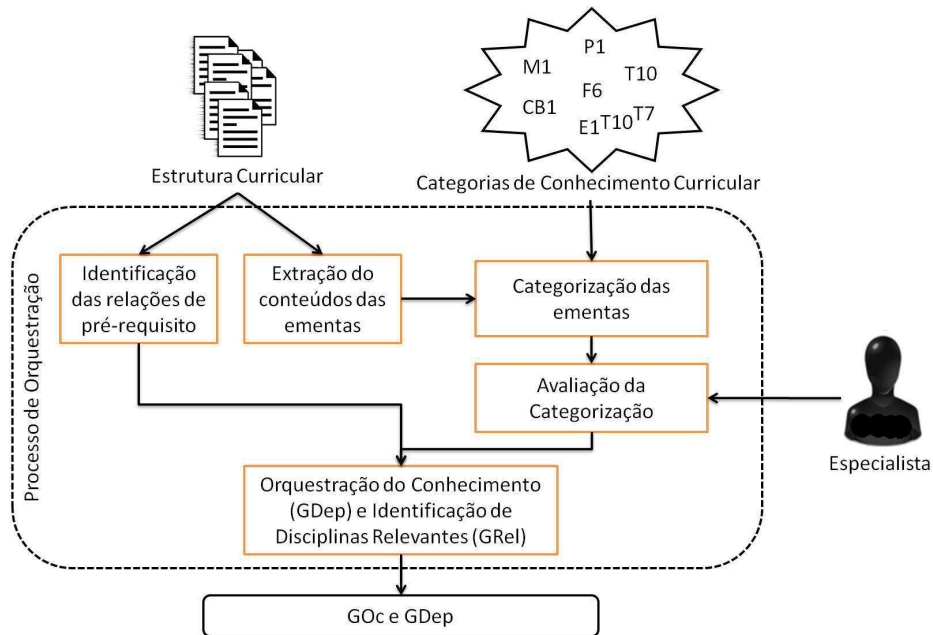
No Brasil, o CR05 comumente é utilizado no desenvolvimento de Planos Pedagógicos dos Cursos (PPC) de graduação em Ciência da Computação (UNIVERSIDADE FEDERAL DO CEARÁ, 2013; UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2010). Como no CR5 não há especificação de relações de pré-requisitos entre as Matérias estas relações tiveram que ser definidas exclusivamente pelos responsáveis pela criação da estrutura curricular do respectivo curso.

Diante disso, verifica-se que a explicitação das relações entre as Matérias do CR05 poderia auxiliar o processo de Planejamento Curricular, facilitando, assim, a determinação das relações de pré-requisitos entre as disciplinas criadas para o curso. Logo, verifica-se a necessidade dessa organização/arranjo entre as partes do conhecimento descrito no CR05. Ressalta-se que a ideia apresentada pelo processo de Orquestração é utilizada na TEL e CSCL para gerenciar atividades de aprendizagem, porém, na abordagem proposta, esse processo será utilizado com foco no planejamento do conhecimento a ser ensinado.

Nesse contexto, foi proposta a abordagem descrita neste trabalho que objetiva orquestrar o conhecimento curricular descrito no CR, a partir dos relacionamentos apresentados nas estruturas curriculares de cursos de IES brasileiras. Para isso, foi desenvolvida uma abordagem *bottom-up* composta por: o conjunto de documentos que descrevem as estruturas curriculares de cursos, detalhado na Subseção 4.1.1; um conjunto de categorias de conhecimento curricular baseadas na granularidade Matéria do CR05 que serão orquestradas, conforme descrito na Subseção 4.1.2; e o processo de orquestração composto por 5 (cinco) fases, manuais e automáticas, conforme descrito na Subseção 4.1.3. O modelo da abordagem

de Orquestração do Conhecimento proposta está representado na Figura 4.1.

Figura 4.1: Diagrama da Abordagem para Orquestração do Conhecimento Curricular



Fonte: O Autor.

Espera-se que com a orquestração do conhecimento seja possível auxiliar o processo de Planejamento Curricular, como também avaliar estruturas curriculares existentes, tomando como base o “padrão” identificado, conforme especificado na Seção 4.2.

### 4.1.1 Estrutura Curricular

A estrutura curricular refere-se às informações dos cursos de IES brasileiras que serão utilizadas no processo de orquestração do conhecimento. Essas estruturas, comumente, estão descritas no PPC dos cursos de graduação que, além de informações burocráticas dos cursos, apresentam o ementário das disciplinas ofertadas pelo mesmo.

O ementário descreve um conjunto de informações sobre as disciplinas do curso, tais como: carga horária, tipo de disciplina (obrigatória ou eletiva), descrição de ementas, pré-requisitos, dentre outras informações, conforme Figura 4.2.

Ressalta-se que, dentre as informações descritas nos PPCs, as informações relacionadas ao ementário das disciplinas são as mais relevante para esse trabalho, principalmente as que descrevem as relações de pré-requisitos e conteúdos das disciplinas. Logo, todos documentos dos cursos submetidos ao processo de orquestração devem apresentar tais informações.

Figura 4.2: Exemplo de uma Ementa Acadêmica

Código: <b>INF1007</b>	Nome: <b>PROGRAMAÇÃO II</b> Abreviado: <b>PROGRAMAÇÃO II</b>
Ementa Linguagens imperativas. Recursão. Arrays e algoritmos básicos: pesquisa seqüencial e pesquisa binária, ordenação por seleção, "heapsort", "mergesort", "quicksort", ordenação por caixas. Tipos estruturados: Estruturas lineares: listas, representação por arrays e por encadeamento, pilhas e filas. Alocação dinâmica de memória. Cadeia de caracteres. Introdução à programação modular.	
Pré-requisitos: INF1005	
Bibliografia Básica: DEPARTAMENTO DE INFORMÁTICA, Introdução à Ciência da Computação, DI/PUC-Rio, 2007. CELES, W. et al., Introdução a Estruturas de Dados, Ed. Campus, 2004.	
Bibliografia Complementar:	
Categoria de Avaliação 4	

Fonte: Pontifícia Universidade Católica do Rio de Janeiro (2008)

As estruturas curriculares foram selecionadas com duas finalidades: a primeira, refere-se às estruturas que tiveram suas disciplinas submetidas ao processo de orquestração do conhecimento, para, assim, serem mensurados as métricas propostas na abordagem; a segunda, refere-se as estruturas, conseqüentemente, os conteúdos das disciplinas que foram utilizadas na construção do conjunto treinamento utilizado no processo de categorização.

### 4.1.2 Categorias de Conhecimento Curricular

Devido a proposta da abordagem apresentar como base o ementário das disciplinas descritas nas estruturas curriculares, optou-se por utilizar a granularidade "matéria" descritas no CR05, por entender que esta é a que mais se aproxima do nível de abstração do conhecimento apresentado nos ementários. Para distinguir as Matérias do CR05 das matérias/disciplinas especificadas nas estruturas curriculares dos cursos, foi utilizada a denominação "Categoria de Conhecimento Curricular".

Foram utilizadas 57 Categorias de Conhecimento Curricular, das 58 categorias possíveis. A categoria Estágio (P8), do contexto Social e Profissional (P), foi desconsiderada devido a não especificação de conteúdo, informação fundamental para a realização do mapeamento entre categoria e disciplina de curso de forma automática. A seguir, estão listadas as Categorias de Conhecimento utilizadas que deverão ser orquestradas ao término do processo proposto com a abordagem.

- **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. Arq. e Org. de Computadores

F4. Circuitos Digitais

F5. Fundamentos de Sistemas

F6. Linguagens de Programação

F7. Ling. For. Aut. 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. Seg. e Auditoria de Sistemas

T18. Sistemas Digitais

T19. Sistemas Distribuídos	T22. Tolerância a Falhas
T20. Sistemas Embarcados	T23. Telecomunicações
T21. Sistemas Multimídia	

- **Contexto Social e Profissional (P)**

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

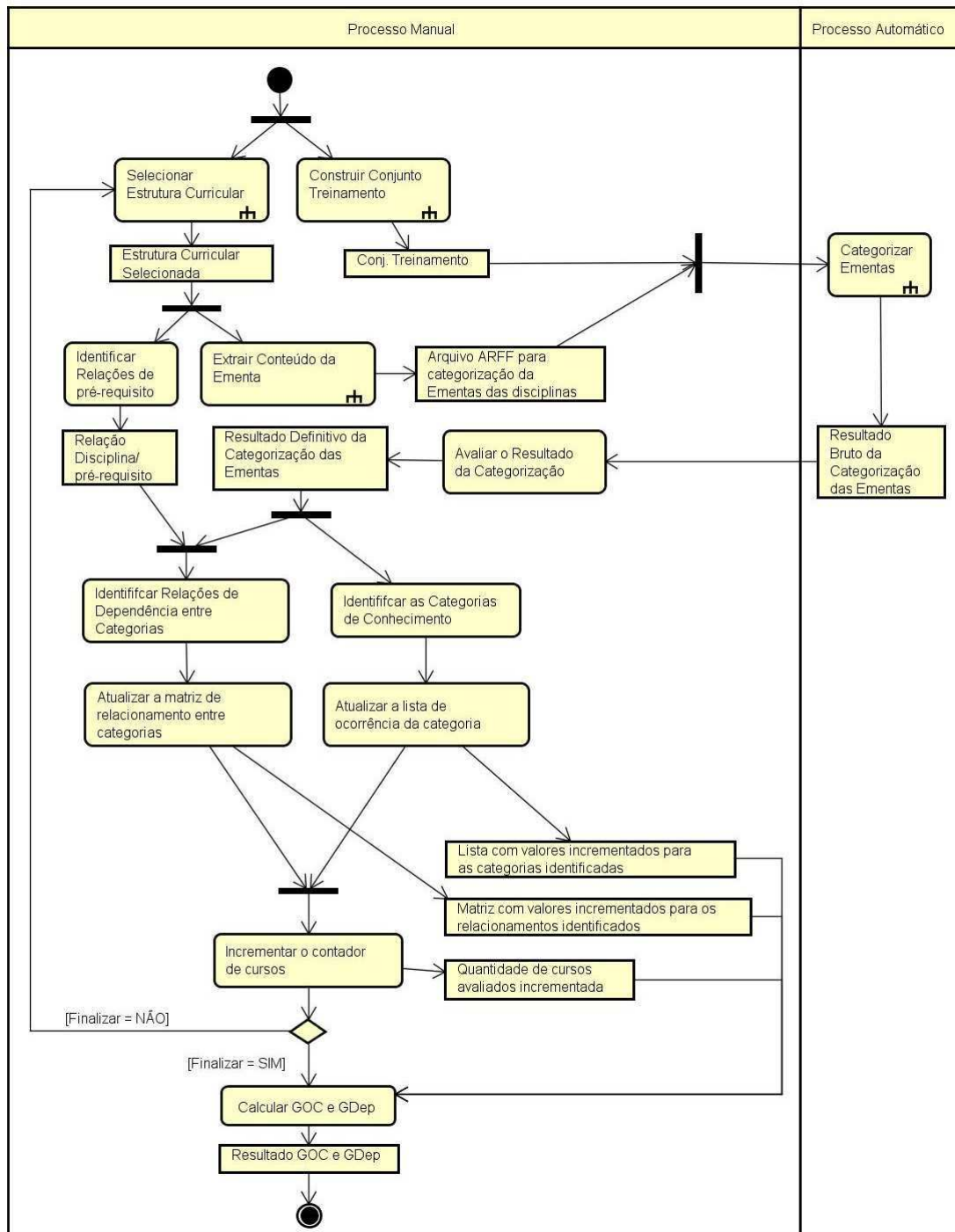
### 4.1.3 Processo de Orquestração das Categorias de Conhecimento

Nesta subseção, é descrito o processo de orquestração das Categorias de Conhecimento Curricular utilizando-se da abordagem *bottom-up* proposta, conforme a Figura 4.1.

O detalhamento do fluxo do processo de orquestração do conhecimento é apresentado no diagrama de atividades da Figura 4.3, no qual é possível verificar o sequenciamento das ações, subatividades e dos objetos relacionados as fases apresentadas na abordagem de orquestração proposta. Ressalta-se que, foi adicionado ao diagrama (Figura 4.3) a subatividade “Selecionar Estrutura Curricular” para descrever o processo de seleção de uma estrutura curricular adequada para ser utilizada no processo de orquestração, conforme detalhado no diagrama de atividades dessa subatividade (Figura 4.4).

As etapas que compõe a Abordagem de Orquestração do Conhecimento e, conseqüentemente, as ações, subatividades e objetos relacionadas a elas estão descritas a seguir, da seguinte forma: identificação das relações de pré-requisitos, extração dos conteúdos das ementas, categorização das ementas, avaliação da categorização, e a orquestração do conhecimento para mensuração do Grau de Ocorrência das Categorias (*GOC*) e do Grau de Dependência entre Categorias (*GDep*).

Figura 4.3: Diagrama de Atividades Ilustrando o fluxo do Processo de Orquestração do Conhecimento Curricular



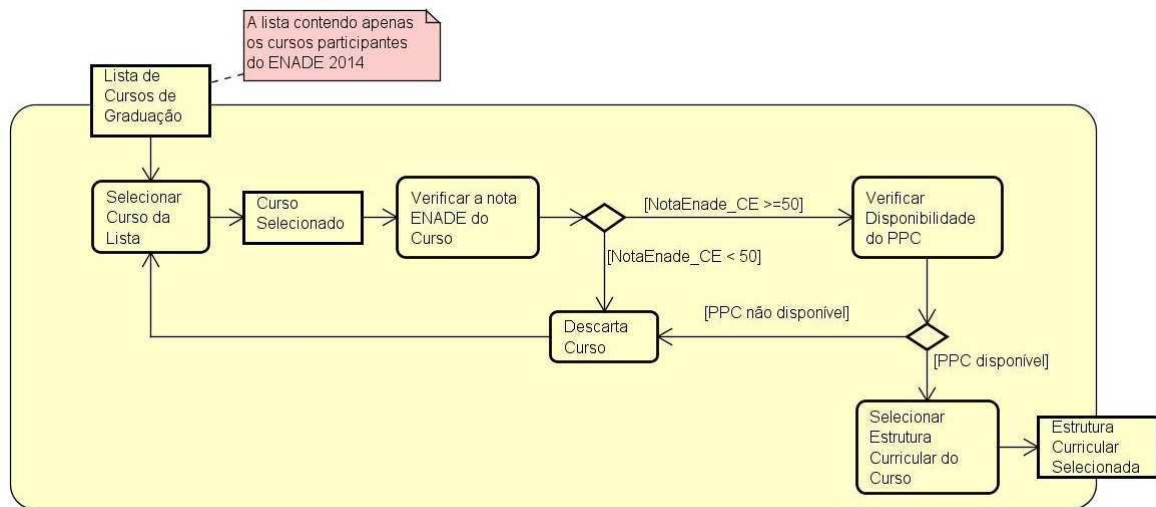
Fonte: O autor.

### Identificar Relações de Pré-requisitos

As informações sobre as relações de pré-requisitos são obtidas a partir da observação das estruturas curriculares dos cursos que, comumente, apresentam as disciplinas organizadas



Figura 4.4: Diagrama de Atividade da Subatividade “Selecionar Estrutura Curricular”



Fonte: O autor.

em uma estrutura contendo as relações de dependências entre si.

Estas relações são fundamentais para a orquestração do conhecimento descrito no CR05, especificamente da granularidade “Matéria”, uma vez que as informações descritas nesse nível de abstração podem ser utilizadas para o planejamento de disciplinas dos cursos de graduação. Assim, ao observar as relações entre as disciplinas de cursos, será possível identificar as possíveis relações entre as Categorias de Conhecimento, e conseqüentemente a orquestração dessas.

Ressalta-se que devido a não padronização dos PPCs dos cursos, esta etapa é realizada de forma manual. Desta forma, essa fase demanda tempo para sua realização, uma vez que deverão ser identificadas, para cada curso avaliado a partir da observação da estrutura curricular do mesmo, as relações de dependência entre as disciplinas que compõem a grade curricular. Como resultado dessa fase, é gerado um documento contendo o nome da disciplina (identificador da mesma) e o conjunto de disciplinas pré-requisitos da mesma, conforme representado na ação “Identificar Relações de Pré-requisito” que resulta o objeto “Relação Disciplina/Pré-requisito” apresentado no diagrama de atividade (Figura 4.3).

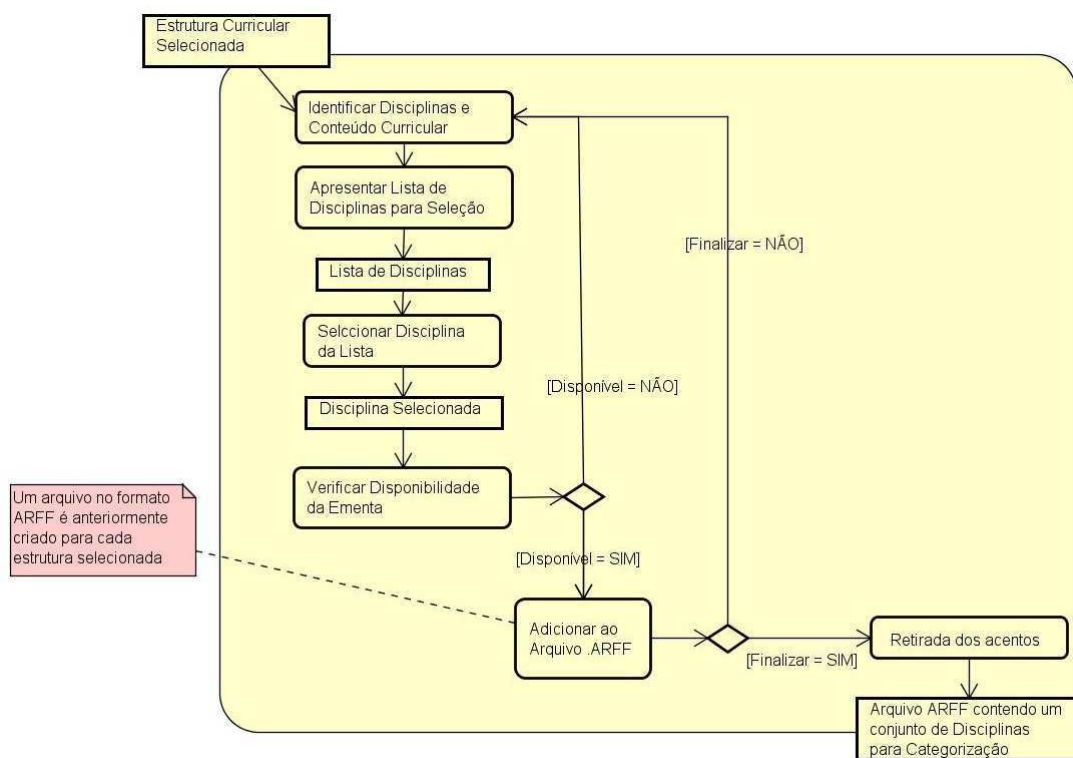
### Extração do Conteúdo das Ementas

Nesta etapa são extraídas das ementas acadêmicas um conjunto de informações sobre os conteúdos abordados em cada disciplina do curso. O resultado dessa fase consiste em

um documento no formato ARFF, para cada estrutura curricular selecionada e analisada, contendo apenas a descrição dos conteúdos abordados nas disciplinas contidas na estrutura curricular. Ressalta-se que a organização do arquivo é baseada na quantidade de disciplinas, uma vez que cada linha do arquivo corresponde a uma disciplina que deverá ser categorizada. Todo esse processo realizado nesta fase está representado no diagrama de atividade da subatividade “Extrair Conhecimento da Ementa”, conforme a Figura 4.5.

A extração do conteúdo e criação do documento é importante para realização do mapeamento entre os dados dos cursos reais e as Categorias de Conhecimento utilizadas na abordagem, processo, esse, automatizado com o uso da tarefa de Categorização de Textos. Destaca-se que todo o processo de extração do conteúdo e criação do arquivo ARFF foi realizado de forma manual, devido a não padronização dos dados.

Figura 4.5: Diagrama de Atividades da Subatividade “Extrair Conhecimento da Ementa”



Fonte: O autor

### Categorização das Ementas

Essa etapa tem finalidade de identificar, a partir das informações obtidas das estruturas curriculares, descritas nos documentos criados na fase anterior, uma das Categorias de

Conhecimento, conforme especificado no Objetivo Específico 2b (Seção 1.2).

Esse processo de identificação foi automatizado com o uso da Categorização de Textos (CT) baseado em Mineração de Textos. Assim, as informações das disciplinas contidas no documento foram mapeadas para uma das 57 Categorias de Conhecimento, deste modo, foi possível identificar a qual a categoria a disciplina pertence. Inicialmente, foi estudada a categorização baseada na Engenharia do Conhecimento, porém, observou-se que esse processo demandaria tempo e esforço para sua realização. Portanto, optou-se pela realização utilizando-se técnicas de Aprendizagem de Máquina, conforme descrito por Linden (2008).

Para a realização da fase de Categorização das Ementas foi utilizada a CT por meio da análise estatística dos termos contidos nos documentos gerados na fase de Extração do Conteúdo das Ementas. Cada linha deste documento representa o ementário da disciplina contida na estrutura curricular, ou seja, cada registro  $d_i$  da coleção de documento  $D = \{d_1, d_2, \dots, d_{|n|}\}$  representa uma disciplina que deve ser classificada em uma das categorias  $c_j$  do conjunto de Categorias de Conhecimento  $C = \{c_1, c_2, \dots, c_{57}\}$ .

A categorização automática foi possível com o uso da ferramenta de Mineração de Dados WEKA<sup>1</sup>, comumente utilizada pela comunidade de MD, mas que também oferece suporte para a MT (HALL et al., 2009). Além disso, para a realização da tarefa de categorização baseada na MT, se faz necessário os seguintes tarefas:

- **Construção do Conjunto Treinamento** – Este conjunto contém registros pré-categorizados de forma manual que serão utilizados como modelo para a fase de Categorização das Ementas. Portanto, a criação desse conjunto está relacionada às Categorias de Conhecimento que deverão ser orquestradas. Sabe-se que o CR05 apresenta um conjunto de informações sucintas sobre o conhecimento que deve ser ensinado ao aluno, porém, se utilizado apenas esse documento para a criação do conjunto, poderá não se obter resultados satisfatórios quando aplicado à CT. Dessa forma, se faz necessária a identificação e uso de outros documentos para a criação desse conjunto treinamento, por exemplo, o ementário de cursos existentes e/ou outros documentos que possam apresentar as informações sobre os conteúdos abordados nas Categorias de Conhecimento. A Figura 4.6 apresenta o diagrama de atividades da subatividade “Construir Conjunto Treinamento”, responsável pela criação do conjunto

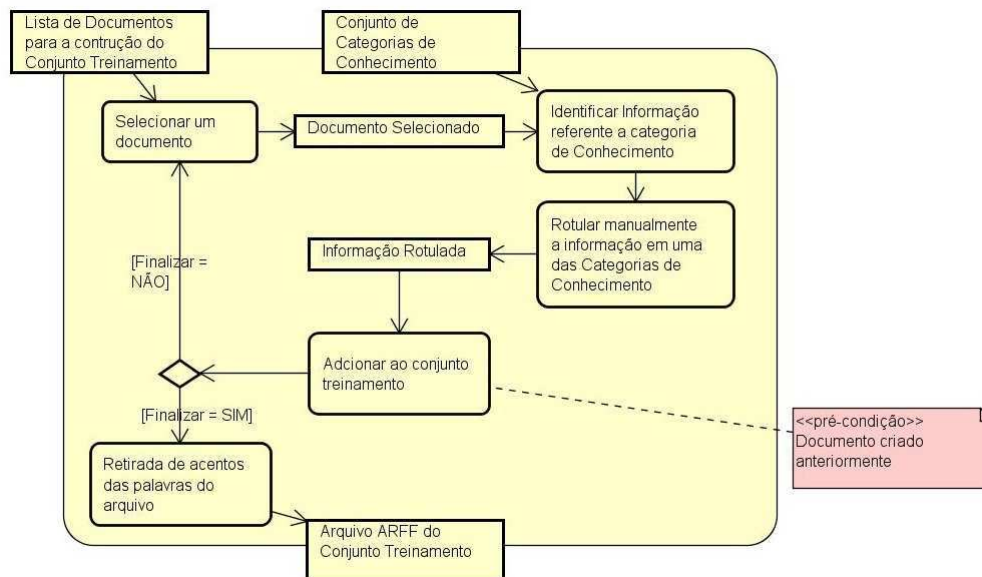
---

<sup>1</sup>Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>

descrito nesse item.

- **Representação dos Documentos** – Aos documentos devem ser aplicadas técnicas de redução da dimensionalidade, tais como: *stemming* das palavras, uso de *stoplists* e definição da frequência mínima de termos.
- **Algoritmo de Categorização** – Diversos algoritmos podem ser utilizados para a tarefa de CT. Nesta pesquisa, para a etapa de categorização das ementas foi selecionado, dentre os algoritmos analisados (ver Apêndice B), o algoritmo probabilístico de Redes Bayesianas.

Figura 4.6: Diagrama de Atividades da Subatividade “Construir Conjunto Treinamento”



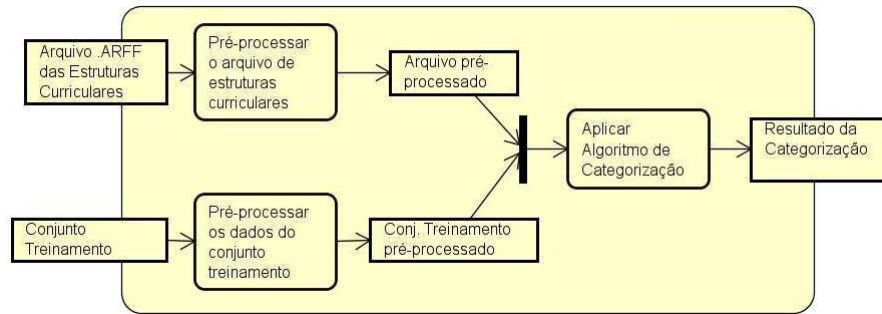
Fonte: O autor

Ressalta-se que o conjunto de dados utilizados para treinamento se diferenciará dos dados dos novos cursos que serão categorizados. Portanto, neste caso, será utilizada a técnica de Casamento de Padrão, no qual é estabelecido um grau de semelhança entre os dados utilizados para treinamento e as disciplinas do novo curso a ser categorizado (FILHO; SILVA; BRAGA, 2015).

O uso da técnica citada, na ferramenta WEKA, foi possível com o uso da implementação *FilteredClassifier* presente na ferramenta, na qual é possível categorizar documentos que contenham tokens diferentes do utilizado para o conjunto treinamento. Como resultado

da Categorização de Ementas, é gerado um documento contendo os registros  $d_i \in D$  classificados em uma das categorias  $c_j \in C$ , ou seja, é criado o par  $(d_i, c_j)$ . O detalhamento de todo o processo de categorização está descrito no diagrama de atividades da subatividade responsável pela categorização (Figura 4.7)

Figura 4.7: Diagrama de Atividades da Subatividade “Categorizar Ementas”



Fonte: O autor

Conforme descrito, a categorização de um documento em uma categoria é realizada com base na semelhança entre os termos descritos no documento  $d_i$  e categoria  $c_j$ . Porém, alguns termos podem estar contidos no conjunto de dados utilizados no treinamento de mais de uma categoria, resultando, assim, na classificação do documento em mais de uma categoria. Nestes casos, se faz necessária a realização de uma análise da categorização com a finalidade de identificar a categoria correta, conforme apresentado na próxima Subseção.

### Avaliação da Categorização das Ementas

Essa etapa tem a finalidade de assegurar a qualidade do resultado do processo de categorização de ementas, ou seja, verificar possíveis erros da etapa de categorização que podem causar impactos negativos na mensuração dos graus de ocorrência e dependência entre categorias.

Além disso, essa fase visa identificar documentos não pertencentes às categorias descritas no conjunto de categorias  $C$ , uma vez que entende-se que possam existir documentos (disciplinas) que não fazem referência às categorias  $c_j \in C$ , mas que foram categorizados devido semelhança entre os termos do documento e da categoria predita.

Esta avaliação é realizada de forma manual, caracterizando-se, assim, como uma atividade subjetiva. Os dados resultantes do processo de categorização são apresentados

na forma de distribuição de probabilidade com valores entre 0 (zero) e 1 (um) que, respectivamente, representam a ausência e a presença total de semelhança entre os termos identificados no documento e conjunto treinamento utilizado como modelo.

Ressalta-se que a forma de apresentação da distribuição de probabilidade é dependente do algoritmo de classificação utilizado. Esta abordagem faz uso do algoritmo probabilístico de Redes Bayesianas, com isso, a identificação dos registros que necessitam de avaliação foi facilitada. De modo geral, a metodologia utilizada para avaliação da tarefa de categorização das ementas é apresentada como:

- Para todo registro de disciplina categorizada:
  - Se a distribuição de probabilidade ( $D \times C$ ) for igual a 1 (um):

**Aceitar a categoria predita pelo algoritmo**
  - Caso contrário, se a distribuição de probabilidade ( $D \times C$ ) apresentar valor entre 0 (zero) e 1 (um), o registro deve ser analisado seguindo as orientações:
    1. Observar nome e ementa da disciplina;
    2. Verificar as categorias que apresentaram termos semelhantes;
    3. Possíveis ações a serem realizadas:
      - \* **Aceitar a categoria predita pelo algoritmo** – Será aceita a categoria predita pelo algoritmo, ou seja, àquela que apresenta maior semelhança entre dos termos do documento e da categoria.
      - \* **Recusar a categoria predita e recategorizar em outra** – Entende-se que os registros podem apresentar termos utilizados em diferentes categorias. Portanto, essa ação visa recategorizar o registro de forma manual utilizando-se de uma das categorias informadas pelo algoritmo e escolhida pelo especialista.
      - \* **Recusar a categoria e recategorizar como excedente** – Conforme mencionado, poderão existir registros de ementas de disciplinas que não possuem relação com as Categorias de Conhecimento utilizadas no processo de categorização de ementas.
  - As distribuições de probabilidade que apresentam valor igual a 0 (zero) não são analisadas.

Esse processo de avaliação complementa o processo de categorização de ementas. Na pior das hipóteses, essa análise deve ser realizada em todos os registros das disciplinas dos cursos categorizado. No diagrama de atividades do processo de orquestração, essa fase refere-se a ação “Avaliar Resultado da Categorização” que tem como *input* o objeto “Resultado Bruto da Categorização das Ementas” e gera o *output* “Resultado Definitivo da Categorização das Ementas”.

### **Orquestração do Conhecimento Curricular**

Essa etapa visa orquestrar o conhecimento da Ciência da Computação com base na observação das relações de pré-requisitos entre os registros de disciplinas categorizadas. Se observado o diagrama de atividades do processo de orquestração do conhecimento (Figura 4.3) as ações que devem ser realizadas nessa fase, são:

- “Identificar Relações de Dependência entre Categorias” – Esta ação fundamenta-se na identificação das relações entre duas ou mais Categorias de Conhecimento. Para isso, nesta ação, além do resultado do objeto “Resultado Definitivo da Categorização das Ementas”, resultado da ação “Avaliar o Resultado da Categorização”, são utilizadas as relações de pré-requisitos identificadas a partir da ação “Identificar Relações de Pré-requisitos”;
- “Identificar Categoria de Conhecimento” – Ação que fundamenta-se na identificação das categorias de Conhecimento que foram identificadas no processo “Categorização de Ementas” e avaliadas no processo de “Avaliar o Resultado da Categorização”;
- “Atualizar a Matriz de Relacionamento entre Categorias” – Neste ação, a matriz que corresponde a soma dos relacionamentos entre duas Categorias de Conhecimento terá seu valor incrementado em 1 (um) a medida que novos cursos sejam submetidos ao processo de orquestração e estes apresentem um mesmo relacionamento observado entre duas categorias;
- “Atualizar a Lista de Ocorrência da Categoria” – Ação que corresponde a soma da ocorrência de cada Categoria de Conhecimento identificada após o processo de categorização;

- “Incrementar o Contador de Cursos” – Ação que tem apenas a finalidade de representar o incremento do contador de cursos que terá seu resultado utilizado no cálculo do item a seguir, e
- “Cálculo do GOC e GDep” – Nesta ação, que é executada após o processamento de todos cursos considerados para a orquestração, é realizado o cálculo que mensura, respectivamente, a relevância da Categoria de Conhecimento no contexto do planejamento curricular, e as possíveis relações de dependência entre as categorias identificadas com o uso das estruturas curriculares dos cursos, considerando a matriz de dependência e a lista de ocorrência das categorias, juntamente com a quantidade de estruturas que foram analisadas.

O detalhamento das ações citadas, incluindo a metodologia para realização das mesma e as equações utilizadas para mensurar os índices, é mostrado a seguir.

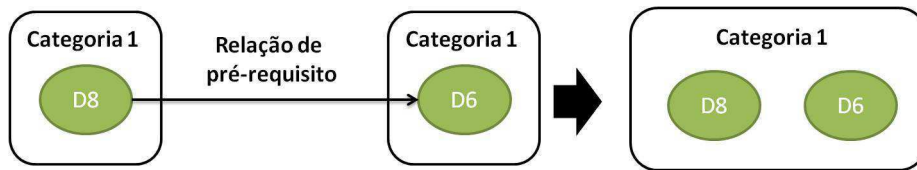
Conforme citado na descrição do processo de Categorização das Ementas, não há uma quantidade fixa de disciplinas estipuladas para os cursos, ou seja, o tamanho do conjunto de documentos  $D$  é variável, com isso, poderá haver mais de um registro  $d_i \in D$  categorizados em uma única categoria  $c_j \in C$ . Dessa forma, se faz necessária a aplicação de um algoritmo para sintetização e a identificação dos possíveis relacionamentos entre as diferentes categorias identificadas para a coleção de documentos  $D$ .

O algoritmo proposto visa agrupar os registros de  $d_i \in D$  que possuem a mesma categoria  $c_i \in C$  e, a partir disso, identificar as relações de dependências nulas e/ou válidas destes, considerando as seguintes situações:

1. Se uma disciplina categorizada apresentar uma relação de pré-requisito com outra disciplina da mesma categoria, essa relação de dependência entre disciplinas e, conseqüentemente, a relação entre as categorias serão anuladas, conforme a Figura 4.8.
2. Se uma disciplina apresentar como pré-requisito uma disciplina de categoria diferente a dela, existirá uma relação correspondente entre as duas categorias, conforme a Figura 4.9.

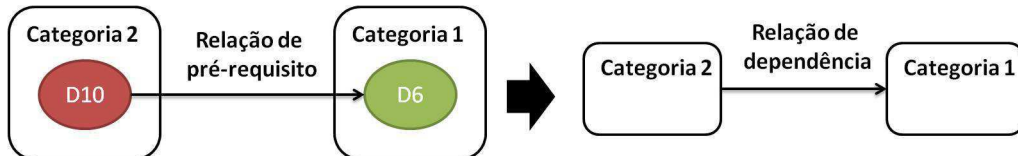


Figura 4.8: Exemplo de Dependência Nula entre Categorias.



Fonte: O autor.

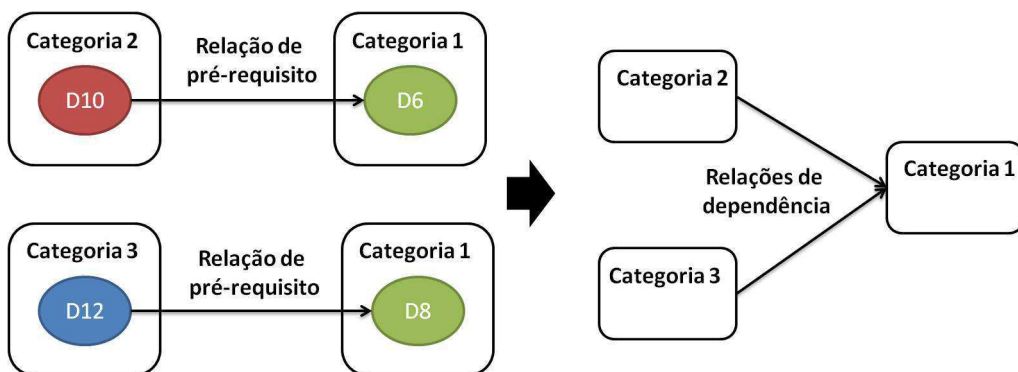
Figura 4.9: Exemplo de Dependência Válida entre Diferentes Categorias.



Fonte: O autor.

Ressalta-se que a metodologia descrita é válida para uma ou mais disciplinas, ou seja, a relação de dependência de uma categoria  $c_j \in C$  será o conjunto de todas as relações de dependências das disciplinas nela categorizada, conforme Figura 4.10.

Figura 4.10: Exemplo de Dependência Válida entre mais de uma Categorias Diferentes.



Fonte: O autor.

Após esse processo de sintetização e identificação das relações de dependências entre diferentes categorias, é proposta a orquestração do conhecimento propriamente dita. No processo de orquestração, será mensurado entre duas diferentes categorias o **Grau de Dependência entre Categorias** ( $GDep$ ) que quantificará a intensidade da relação observada entre as categorias. Observou-se que informações relevantes sobre as disciplinas seriam desperdiçadas caso apenas o  $GDep$  fosse mensurado, dessa forma, optou-se em também mensurar o **Grau de Ocorrência da Categoria** ( $GOC$ ) que identificará quais as Categorias de Conhecimento mais relevantes no contexto do planejamento curricular dos cursos.

A mensuração do *GDep* baseia-se na média aritmética da relação de dependências entre duas categorias, sobre todos os cursos. O resultado do *GDep* é apresentado em uma matriz  $M_{n \times n}$ , com  $n = 57$ . Ressalta-se que cada elemento  $M_{i \times j}$  da matriz  $M$  corresponde ao *GDep* do relacionamento entre duas diferentes categorias do curso em questão.

Ressalta-se que o cálculo do *GDep* deve apresentar valores  $0 \leq M_{i \times j} \leq 1$ . Portanto, para satisfazer essa condição, deve ser realizado o somatório dos elementos  $M_{i \times j}$  de todos os cursos e, posteriormente, a divisão dessa soma pelo valor  $n$  que representa a quantidade de cursos utilizados na mensuração. Portanto, para que esse cálculo seja realizado. O valor de cada elemento  $M_{i \times j}$ , para cada curso, será 0 (zero) quando não houver relação de dependência entre as categorias, e 1 (um) quando houver a relação de dependência.

A fórmula utilizada para o cálculo do *GDep* é apresentada na Equação 4.1, que utiliza a função  $gd(k, M_{i,j})$  para determinar o grau de relevância do elemento  $M_{i,j}$ , representando o grau do relacionamento entre as categorias  $c_i$  e  $c_j$ , do  $k$ -ésimo curso analisado. Portanto, para  $i$  e  $j$ , variando de 1 a 57, correspondente a quantidade de Categorias de Conhecimento utilizado na abordagem de orquestração, e  $n$  sendo a quantidade de cursos utilizados na orquestração, temos:

$$GDep(i, j) = \frac{\sum_{k=1}^n (gd(k, M_{i,j}))}{n} \quad (4.1)$$

A mensuração do *GOC* baseia-se na frequência de ocorrência da Categoria de Conhecimento nos cursos utilizados na abordagem. A necessidade da mensuração desse grau foi verificada a partir da observação de que com apenas a análise do *GDep* não seria possível identificar quais categorias são mais representativas no processo de ensino, ou seja, as Categorias de Conhecimento mais presentes nos cursos de graduação. Outra motivação identificada, foi que algumas categorias não apresentam relações de dependência com outras, e, desta forma, não seria possível identificá-las apenas observando as relações de dependências.

O cálculo desse grau também baseia-se na representatividade booleana da Categoria de Conhecimento nos cursos analisados. Portanto, quando a categoria estiver representada na estrutura curricular do curso, o valor será 1 (um), caso contrário, o valor será 0 (zero). Após o processamento de todos os cursos analisados, deverá ser calculado, para cada Categoria de Conhecimento, a média aritmética da mesma utilizando-se da quantidade  $n$  de cursos

avaliados, conforme a equação 4.2, que utiliza a função  $goc(k, c_i)$  para determinar se a categoria  $c_i$ , ocorre no curso  $k$  analisado. Portanto, com  $i$  variando de 1 a 57, correspondente a quantidade de Categorias de Conhecimento, temos:

$$GOC(c_i) = \frac{\sum_{k=1}^n (goc(k, c_i))}{n} \quad (4.2)$$

Ressalta-se que tanto o  $GDep$  quanto o  $GOC$  apresentam valores entre 0 e 1. Neste caso, o valor 1 representa que a situação observada está presente em todos os cursos analisados. Portanto, tal situação pode ser considerado como representativa no contexto observado.

## 4.2 Uso da Orquestração para Planejamento e Avaliação Curricular

A mensuração do  $GDep$  e o  $GOC$  podem ser insumos importantes para o processo de planejamento e avaliação de estruturas curriculares. Assim, nesta seção, serão discutidos alguns aspectos relacionados ao uso dessas duas métricas no contexto apresentado.

### 4.2.1 Proposta para uso no Planejamento Curricular

Acredita-se que os índices  $GDep$  e  $GOC$  podem ser úteis para o processo de Planejamento Curricular, uma vez que a partir de ambas métricas podem ser identificadas quais as Categorias de Conhecimento, dentre o conjunto apresentado no CR, são mais representativas e, além disso, é possível verificar as relações de dependência existente entre as Categorias de Conhecimento.

Tais informações, podem ser utilizadas na determinação do conhecimento que deve ser ensinado ao aluno, e na verificação e especificação de relações de dependências entre as diferentes Categorias de Conhecimento que compõe o conhecimento da computação.

Como forma de avaliar esse uso no planejamento, foi proposto a realização de um planejamento curricular utilizando-se um conjunto de disciplinas, conforme descrito no Capítulo 5. Nessa avaliação, serão descritas apenas as Categorias de Conhecimento e, com base nos valores do  $GDep$ , deverão ser apresentadas as possíveis relações entre as diferentes Categorias de Conhecimento analisadas. Além disso, com base no  $GOC$ , será observado se

as Categorias de Conhecimento descritas no estudo, apresentam-se como relevantes.

### 4.2.2 Proposta para uso na Avaliação Curricular

A proposta de uso do resultado do processo de orquestração na Avaliação da Estrutura Curricular, visa verificar a conformidade dessa estrutura com o padrão identificado. Esta verificação, será baseada na observação das relações de dependências entre diferentes Categorias de Conhecimento e da existência dessas no currículo do curso, utilizando-se, respectivamente, da mensuração *GDep* e *GOC*.

Para tanto, a estrutura do curso a ser analisado, deverá passar por um conjunto de etapas referentes ao processo de Orquestração do Conhecimento, conforme Figura 4.1. Ressalta-se que, como apenas um curso será analisado, não será calculada a média referente ao *GDep* e *GOC*. Logo, para o curso avaliado, esses valores serão booleanos (0 ou 1). Quando confrontados os valores booleanos com o padrão identificado, será possível analisar a conformidade das relações de dependência e ocorrência das Categorias de Conhecimento com o “padrão” identificado.

No Capítulo 5, é apresentada um exemplo de uso com a finalidade de validar a proposta de uso do resultado da orquestração no processo de Avaliação Curricular. Para isso, será verificada a cobertura das Categorias de Conhecimento e as relações entre as diferentes Categorias descritas na estrutura curricular analisada, com base no padrão identificado.

## 4.3 Considerações Finais do Capítulo

Este Capítulo apresentou os aspectos fundamentais para o entendimento da abordagem para orquestração proposta neste trabalho. Para isso, foram detalhados os processos a serem realizados para a identificação da Orquestração das Categorias de Conhecimento a partir dos dados de cursos reais. Além disso, foram descritas as métricas, resultado do processo de orquestração, e que podem ser utilizadas no processo de Planejamento e Avaliação Curricular, conforme proposto na Seção 4.2. No Capítulo 5 será apresentado o Estudo de Caso realizado com a finalidade de avaliar a abordagem proposta.

# Capítulo 5

## Avaliação da Abordagem Proposta

Este capítulo está diretamente relacionado ao objetivo específico 4, descrito na Seção 1.2, uma vez que é apresentada a avaliação da abordagem para orquestração do conhecimento proposta. Para isso, será descrita a metodologia utilizada na condução do estudo de caso, sua condução e os resultados obtidos.

Este capítulo está organizado da seguinte forma: na Seção 5.1, são apresentadas algumas considerações acerca do estudo de caso realizado, com destaque para a metodologia proposta para a condução do estudo. A seleção dos cursos que serão submetidos à avaliação para identificação da Orquestração do Conhecimento, é descrita na Subseção 5.1.1; o processo de Orquestração do Conhecimento Curricular e os resultados obtidos, são descritos na Subseção 5.1.2; o uso das métricas *GDep* e *GOC* no processo de Planejamento Curricular e na avaliação de uma estrutura curricular, é descrito na Subseção 5.1.3. Por fim, na Seção 5.2, são apresentadas as considerações finais deste capítulo.

### 5.1 Estudo de Caso

Para a realização desse estudo de caso foi proposta uma metodologia com a finalidade de auxiliar a condução do mesmo. 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.

A seleção das estruturas curriculares objetivou identificar, dentre os cursos de graduação

em Ciência da Computação no Brasil, aqueles que tiveram os dados das suas estruturas curriculares utilizados no processo de construção do conjunto treinamento e aqueles utilizados no processo de orquestração das Categorias de Conhecimento Curricular. Após essa seleção, os dados foram submetidos ao processo de Orquestração do Conhecimento Curricular e, conseqüentemente, a um conjunto etapas que resultou na mensuração do *GOC* e *GDep*.

### 5.1.1 Seleção das Estruturas Curriculares

A seleção das estruturas curriculares é uma atividade de fundamental importância para a abordagem ora proposta, uma vez que estas serão os insumos utilizados para a identificação da orquestração entre as Categorias de Conhecimento.

Tal seleção, realizada de forma manual, visa identificar um conjunto de dados para o processo de mapeamento entre os cursos reais e as categorias, pois, tal processo baseia-se nos conteúdos apresentados no ementário das disciplinas contidos nessas estruturas. As diferentes finalidades dessa seleção são descritas nas Subseções a seguir.

#### **Estruturas Selecionadas para o Processo de Orquestração**

As estruturas curriculares utilizadas no processo de Orquestração do Conhecimento (Subseção 5.1.2) são oriundas dos cursos de Bacharelado em Ciência da Computação (BCC). Sabe-se que, no Brasil, há diversos cursos de BCC, por isso, para o uso na abordagem foram selecionados os cursos de maior “representatividade” com base no critério de notas do ENADE<sup>1</sup>.

Inicialmente, foram selecionados cursos que tiveram alunos submetidos ao ENADE do ano de 2014 e obtiveram pontuação na área de Conhecimento Específico superior a 50,0 pontos. Tal critério, baseou-se no acerto mínimo da metade das questões de conhecimento específico, ou seja, o conhecimento mais aproximado dos representados nas Categorias de Conhecimento. Após essa fase inicial, foi verificado, em ordem decrescente de seleção, se os cursos identificados disponibilizavam seu PCC e, assim, foi realizado um refinamento da seleção anterior.

---

<sup>1</sup>Conceito ENADE – Disponível em: <http://goo.gl/eA5iFM>

Sabe-se que o cálculo da nota final do ENADE é baseado na pontuação dos alunos na provas de conhecimento específico e geral. Portanto, não necessariamente, os cursos utilizados nessa abordagem para a orquestração do conhecimento apresentarão o maior conceito possível na prova, ou seja, igual a 5 (cinco). Assim, os 7 (sete) cursos selecionados para serem submetidos à abordagem proposta, estão apresentados na Tabela 5.1.

Tabela 5.1: 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

Ressalta-se que, para a seleção dos cursos utilizados no processo de identificação da orquestração, não foram considerados àqueles utilizados para compor o conjunto treinamento, pois, para estes foram aplicados outros critérios para seleção, conforme apresentado na próxima Subseção.

### **Estruturas Selecionadas para a Construção do Conjunto Treinamento**

A seleção de estruturas curriculares utilizadas na construção do conjunto treinamento baseou-se no critério ENADE, não apenas os cursos que obtiveram pontuação superior a 50 pontos, mas também na região geográfica da IES, ambos os critérios buscaram diversificar as estruturas selecionadas. Assim, foram selecionados 8 (oito) cursos de graduação de diferentes regiões geográficas e conceitos ENADE, conforme Tabela 5.2

Além das ementas das disciplinas contidas nas estruturas curriculares selecionadas, para a criação do conjunto treinamento também foram utilizados dados extraídos de documentos referência, conforme apresentado no tópico **Construção do Conjunto Treinamento** da Subseção 5.1.2.

Tabela 5.2: Distribuição dos Cursos do Conjunto 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

### 5.1.2 Processo de Orquestração de Conhecimento

Conforme descrito no Capítulo 4, esse processo baseou-se nas fases **Extração de Conteúdo das Ementas, Identificação das Relações de Pré-requisitos, Categorização das Ementas, Avaliação da Categorização e Mensuração do *GOc* e *GRel***. A seguir, serão realizados alguns comentários sobre cada uma das fases e seu comportamento durante a condução do estudo de caso.

#### **Extração do Conteúdo da Ementas e Identificação das Relações de Pré-requisitos**

As fasea de Extração do Conteúdo das Ementas e Identificação das Relações de Pré-requisitos, apesar de serem duas fases distintas, foram agrupadas nesta Subseção devido a sua realização a partir das estruturas curriculares selecionadas (ver Subseção 5.1.1).

Ressalta-se que, devido a não padronização dos documentos que descrevem o ementário das disciplinas de um curso, não foi possível a automatização de ambos os processos, exigindo, assim, que fosse demandada uma carga de trabalho manual para a sua realização.

O resultado desses processos consistiu na extração dos conteúdos das disciplinas, comumente, descritos nas ementas das mesmas e, além disso, foram identificadas as relações de pré-requisitos entre as disciplinas da estrutura curricular, importantes para o processo de Orquestração do Conhecimento Curricular.

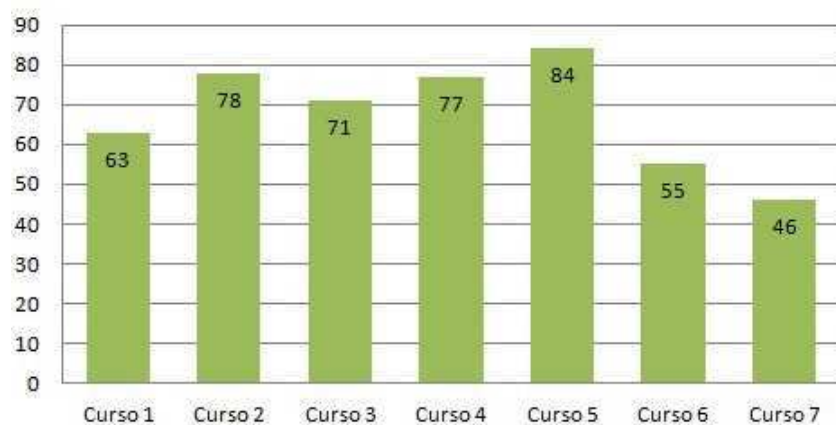
Os dados selecionados na fase de Extração de Conhecimento das Ementas foram organizadas para, assim, serem utilizadas no processo de categorização como modelo para reconhecimento das categorias de forma automática. Devido a utilização da ferramenta WEKA para o processo de categorização, as informações extraídas foram organizadas em



um arquivo no formato .ARFF<sup>2</sup>, no qual cada linha representa a ementa de uma disciplina. O detalhamento do arquivo será realizado no processo de construção do conjunto treinamento, pois, a estrutura de ambos é semelhante. Ressalta-se que os documentos criados foram submetidos ao processo de retirada da acentuação das palavras dos textos que descrevem o conteúdo curricular.

De modo geral, esse processo contabilizou 474 registros de disciplinas com quantidades distribuídas conforme a Figura 5.1, organizados em diferentes arquivos .ARFF que foram submetidos ao processo de Categorização, conforme Apêndice B.1.

Figura 5.1: Distribuição das disciplinas observadas por curso



Fonte: O autor.

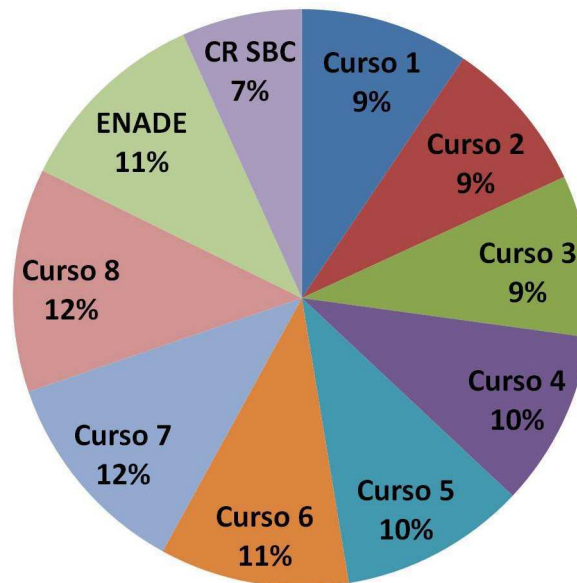
### Categorização das Ementas

Esse processo é realizado com base nas informações extraídas da Estrutura Curricular do curso à ser avaliado, o conjunto treinamento e a aplicação do algoritmo de classificação probabilístico Rede Bayesianas.

Conforme descrito anteriormente, para a construção do **Conjunto Treinamento** foram utilizadas informações da granularidade Matéria do CR05, das diretrizes do ENADE e 8 (oito) estruturas curriculares diferente das submetidas ao processo de Orquestração do Conhecimento, resultando em 519 registros de disciplinas (texto das ementas) pré-classificadas manualmente em uma das 57 categorias de conhecimento do CR, distribuídos conforme a Figura 5.2.

<sup>2</sup>Tipo de arquivo compatível com a ferramenta WEKA

Figura 5.2: Distribuição dos registros dos cursos utilizados na construção do Conjunto Treinamento



Fonte: O Autor.

Os dados utilizados no Conjunto Treinamento não apresenta distribuições igualitárias entre cada Categoria do Conhecimento por curso. Isso ocorre devido a existência de mais de uma disciplina categorizada em uma mesma categoria, situação aceitável, pois o CR descreve apenas orientações, não limitando a quantidade de disciplinas do currículo à quantidade apresentada no documento.

Diante disso, as Categorias de Conhecimento que mais pontuaram registros no conjunto treinamento, foram: Linguagem de Programação (34 registros), Engenharia de Software (31 registros), Organização e Arquitetura de Computadores (22 registros) e Cálculo Diferencial e Integral e Redes de Computadores, que pontuaram 21 registros cada. Por outro lado, as categorias Variáveis Complexas, Métodos Qualitativos Aplicado à Administração e Psicologia, pontuaram apenas 1 (um) registro cada, obtidos com o uso do CR.

Após o processo de criação do conjunto treinamento, o documento foi utilizado na fase de **Categorização das Ementas**. Nessa fase, além do conjunto treinamento, foi utilizada a ferramenta WEKA e o conjunto de dados extraído das estruturas curriculares (conforme apresentado na Seção 5.1.1). Aos dados coletados, foram aplicadas as técnicas de *stemming* das palavras, uso de *stoplist*<sup>3</sup> e a definição da frequência mínima de termos igual a 4 quatro,

<sup>3</sup>Utilizada a *stoplist* da ferramenta PreText. Disponível em: <http://sites.labc.icmc.usp.br/pretext2/>

devido a distribuição não igualitária dos dados.

Como resultado do processo de Categorização de Ementas, para cada estrutura curricular analisada, foi gerada uma Matriz  $M_{n \times 57}$ , com  $n$  correspondendo ao número de disciplinas categorizadas no curso. Essa matriz apresenta, para cada par categoria e disciplina, a distribuição de probabilidade do documento (registro de disciplina) pertencer à Categoria de Conhecimento e, com base nessa distribuição, é predita a categoria correspondente ao registro. Os documentos citados após a execução do processo de categorização de ementas estão citados no Apêndice B.2.

Por sua vez, esse resultado apresenta-se como bruto, ou seja, todos os registros categorizados apresentam uma categoria correspondente, especificada no conjunto treinamento. Porém, conforme descrito no Capítulo 4, podem existir erros nesse processo de categorização devido, por exemplo, aos termos comuns utilizados na descrição das ementas e conjunto treinamento, acarretando, assim na necessidade do processo de Avaliação da Categorização realizada.

O processo de **Avaliação da Categorização das Ementas**, conforme já mencionado, analisou o resultado do processo de categorização, com a finalidade de assegurar o resultado obtido. Foram avaliados apenas os registros que apresentaram distribuição de probabilidade com valor entre 0 e 1.

No resultado bruto do processo de Categorização das Ementas, descrito na Tabela 5.3, é possível verificar a quantidade total de registros de disciplinas utilizadas, as disciplinas que apresentaram categorização com distribuição de probabilidade igual a 1 (um), ou seja, disciplinas não serão submetidas às análises, e a quantidade dos registros de disciplinas que apresentaram distribuição de probabilidade diferente de 1 (um), ou seja, foram submetidas às análises da categorização.

Vale ressaltar que, não necessariamente, a quantidade de registros analisados representam erros na categorização, eles apenas destacam registros onde os termos da disciplina estão apresentados em mais de uma categoria ou, no pior caso, disciplinas que não possuam correspondentes no CR. Outra informação que pôde ser obtida com essa análise preliminar foi que o processo de categorização de forma automática reduziu a carga de trabalho de uma categorização manual, uma vez que do conjunto de 474 disciplinas, 166 necessitaram de análises, significando 35% do total dos registros de disciplinas.

Curso	Quantidade de Disciplinas	Disciplinas analisadas	não	Disciplinas Analisadas
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

Tabela 5.3: Resultado Bruto da Categorização de Ementas

Após a identificação dos registros de disciplinas que necessitaram de análises para verificação da categoria predita, eles foram submetidos ao algoritmo de avaliação, conforme destacado no capítulo anterior. O resultado desse processo de avaliação é apresentado na Tabela 5.4, na qual é possível verificar a quantidade de registros que tiveram as categorias aceitas, recusadas e recategorizadas em outra categorias ou como excedentes.

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

Tabela 5.4: Resultado da Avaliação da Categorização das Ementas

Com base na Tabela 5.4, é possível observar que do total de ementas categorizadas e analisadas, 49% tiveram as categorias previstas aceitas, 34% foram recategorizadas em outra categorias do CR e 17% foram recategorizadas como excedentes. Ao analisar alguns registros recategorizados, foi verificado que a não exatidão do algoritmo foi devido à semelhança entre os conteúdos de diferentes Categorias de Conhecimento, situação verificada também nas categorias: Sistemas Distribuídos e Redes de Computadores, Sistemas Operacionais e Programação Paralela, Inglês e Comunicação e Expressão, entre outras.

Em outros casos, observou-se que as disciplinas recategorizadas como excedentes referenciavam, por exemplo, conteúdos de bioinformática, questões indígenas, libras,

geoprocessamento, que não apresentam referência direta no CR05 e, conseqüentemente, nas Categorias de Conhecimento utilizada na construção do conjunto treinamento.

Por fim, verificou-se que todo o processo de Categorização das Ementas, no contexto dessa pesquisa, apresentou-se como uma técnica computacional viável, pois, de modo geral, 81,65% de categorização das ementas foi realizada de forma correta, considerando, neste cálculo, os registros que necessitaram de análises.

A última fase do processo descrito corresponde à **Orquestração do Conhecimento Curricular**, para isso, serão apresentados os resultados referentes às análises das Categorias de Conhecimento para verificação das relações de dependência e a mensuração dos índices *GOC* e *GDep*.

A análise das Categorias de Conhecimento para verificação das relações de dependência entre categorias e também para sumarização da quantidade de categorias, resultou em uma média de 33 Categorias de Conhecimento, por curso analisado. Por sua vez, essas categorias estão organizadas nos núcleos descritos no CR05, conforme Tabela 5.5.

Curso	Quant. Cat.	M	CB	E	F	T	P
1	31	6	1	0	7	12	5
2	35	6	1	1	7	13	7
3	32	6	1	0	9	12	4
4	37	6	1	0	9	15	6
5	35	6	1	0	8	14	6
6	33	6	1	0	8	13	4
7	29	7	0	0	8	14	2

Tabela 5.5: Detalhamento de Categorias por Curso e Núcleo de Conhecimento

Ao analisar os dados da Tabela 5.5, verifica-se que, sem distinção entre disciplinas obrigatórias e eletivas, os currículos dos cursos sempre apresentam categorias das áreas da Matemática (M), Fundamentos da Computação (F), Teoria da Computação (T) e Contexto Profissional (P), respectivamente pontuando uma média de 6, 8, 13 e 5 Categorias de Conhecimento. Além disso, a área da Ciência Básica (CB) também é bastante presente nos currículos, diferentemente da Eletrônica (E) que apresentou uma representatividade baixa, tendo pontuado em apenas 1 (um) dos cursos analisados.

Observa-se que, com base apenas nas relações de dependência entre as diferentes Categorias de Conhecimento, não há como identificar quais as Categorias de Conhecimento

mais relevantes no contexto do ensino de computação. Tal identificação, só será possível a partir da mensuração do Grau de Ocorrência da Categoria *GOC*.

Quadro 5.1: Mensuração do *GOC* das Categorias de Conhecimento Mais Representativas.

Categoria de Conhecimento	GOC
F3. Arquitetura e Organização de Computadores	1
F4. Circuitos Digitais	1
F6. Linguagens de Programação	1
F7. Linguagens Formais, Autômatos e Computabilidade	1
F9. Sistemas Operacionais	1
M1. Álgebra Linear	1
M7. Matemática Discreta	1
M8. Probabilidade e Estatística	1
P2. Computadores e Sociedade	1
T16. Redes de Computadores	1
T19. Sistemas Distribuídos	1
T2. Bancos de Dados	1
T5. Computação Gráfica	1
T7. Engenharia de Software	1
T9. Interação Humano- Computador	1
CB1. Física	0,86
F10. Teoria dos Grafos	0,86
F2. Algoritmos e Estrutura de Dados	0,86
M3. Cálculo Diferencial e Integral	0,86
M5. Geometria Analítica	0,86
M6. Lógica Matemática	0,86
P7. Empreendedorismo	0,86
T4. Compiladores	0,86
F1. Análise de Algoritmos	0,71
P10. Informática na Educação	0,71
T10. Matemática Computacional	0,71
T15. Programação Paralela	0,71
T20. Sistemas Embarcados	0,71
F5. Fundamentos de Sistemas	0,57
P1. Administração	0,57
T14. Processamento de Imagens	0,57
T17. Segurança e Auditoria de Sistemas	0,57
T21. Sistemas Multimídia	0,57
T8. Inteligência Artificial	0,57

A Quadro 5.1 apresenta a mensuração do *GOC* identificada a partir das estruturas curriculares selecionadas para esse fim, independente do tipo da disciplinas ser obrigatórias e eletivas. Neste caso, é possível observar que algumas Categorias de Conhecimento apresentam-se como relevantes, uma vez que elas estão presentes em grande parte dos cursos analisados. Ressalta-se que com valor de  $n = 7$ , isso implica dizer que as categorias com  $GOC = 1$  foram verificadas em todos os cursos avaliados. Por outro lado, foram identificadas Categorias de Conhecimento que apresentaram representatividade

baixa, conforme o Quadro 5.2. O detalhamento do *GOC* considerando os Núcleos de Conhecimento e a diferenciação entre disciplinas obrigatórias e eletivas, esta descrito no Apêndice C.1.

Quadro 5.2: Mensuração do *GOC* das Categorias de Conhecimento Menos Representativas

Categoria de Conhecimento	GOC
M4. Equações Diferenciais	0,43
P11. Inglês	0,29
P12. Métodos Quantitativos Aplicados à Administração de Empresas	0,29
P9. Filosofia	0,29
T1. Análise de Desempenho	0,29
T6. Automação e Controle	0,29
E1. Circuitos Eletrônicos	0,14
F8. Organização de Arquivos e dados	0,14
P13. Sociologia	0,14
P14. Psicologia	0,14
P3. Comunicação e Expressão	0,14
P4. Contabilidade e Custos	0,14
P5. Direito e Legislação	0,14
P6. Economia	0,14
T11. Métodos Formais	0,14
T12. Modelagem e Simulação	0,14
T18. Sistemas Digitais	0,14
T22. Tolerância a Falhas	0,14
T23. Telecomunicações	0,14
T3. Circuitos Integrados	0,14
M2. Análise Combinatória	0
M9. Variáveis Complexas	0
T13. Processamento Digital de Sinais	0

Após a mensuração da relevância de cada Categoria do Conhecimento, foram mensuradas as relações de dependência entre estas categorias. Para isso, foi calculado o **Grau de Dependência entre Categorias** ( $GD_{ep}$ ), fundamental para o processo de Orquestração.

O cálculo  $GD_{ep}$  utilizando as estruturas curriculares descritas na Seção 5.1.1, resultou na identificação de 45 Categorias de Conhecimento que apresentaram relacionamentos com outras 32 categorias. Um exemplo desses relacionamentos e, conseqüentemente, do cálculo do  $GD_{ep}$  na Tabela 5.6, na qual são explicitados os relacionamentos entre as Categorias de Conhecimento do Núcleo da Matemática.

Na mensuração do  $GD_{ep}$  para o Núcleo da Matemática é possível observar que 6 (seis) Categorias de Conhecimento (M1, M3, M4, M6, M7 e M8) apresentam relações com outras Categorias do mesmo Núcleo. Observa-se também que a proporção entre as “Categorias que

	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

Tabela 5.6: Mensuração do  $GDep$  das Categorias de Conhecimento do Núcleo Matemática

se relacionam” e as “Categorias que são relacionadas” não necessariamente são iguais, uma vez que as 6 (seis) categorias relacionam-se com outras 5 (cinco) categorias.

Os maiores valores de  $GDep$  foram vistos nos relacionamentos entre as Categorias “Probabilidade e Estatística” com “Cálculo Diferencial e Integral” e “Arquitetura e Organização de Computadores” com “Circuitos Diigitais”, ambos os relacionamentos apresentaram valor igual a 0,86. Outros exemplos com alta representatividade foi mensurado entre as categorias “Computação Gráfica” com “Álgebra Linear” e “Engenharia de Software” com “Linguagem de Programação”, mensurado em 0,71. O resultado completo do  $GDep$  é apresentado no Apêndice C.2.

Também foram identificados relacionamentos entre categorias de Núcleos distintos. Na Tabela 5.7 são exemplificadas algumas relações entre as Categorias de Conhecimento dos Núcleo Fundamentos da Computação (F) e Matemática (M). Entre esses relacionamentos, por exemplo, destaca-se a relação de dependência entre Linguagens Formais e Autômatos e Computabilidade (F7) com Matemática Discreta (M7), mensurado em 0,57 que representa, no contexto estudado, um relacionamento significativo.

De modo geral, entende-se que o valor do  $GDep$  será mais preciso a medida que a quantidade de cursos avaliados for incrementada. Portanto, não se pode generalizar as análises realizadas, mas, no contexto observado, verificou-se que parte das Categorias de Conhecimento que não possuem relações correspondem as disciplinas que, comumente, são



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

Tabela 5.7: Mensuração do *GDep* das Categorias de Conhecimento entre os Núcleos Fundamentos da Computação e Matemática.

lecionadas no início dos cursos como, por exemplo, Geometria Analítica. Por outro lado, verificou-se também que as Categorias de Conhecimento do Núcleo do Contexto Social e Profissional não apresentaram relacionamentos de dependências com outras disciplinas e nem foram relacionadas por outras.

### 5.1.3 Avaliação da Proposta de Uso no Planejamento e Avaliação Curricular

Para exemplificar a proposta de uso da abordagem no processo de Planejamento Curricular e Avaliação das estruturas curriculares foram utilizados os mesmos conjunto de dados identificados a partir de um curso real. Para isso, foram selecionadas de uma estrutura curricular de um curso real, um conjunto contendo 22 disciplinas referente às disciplinas do 3º e 4º período e as disciplinas pré-requisitos dessas associadas à elas.

As disciplinas selecionadas foram submetidas ao processo de categorização para, assim, serem utilizadas no processo de Avaliação e Planejamento Curricular. Este processo baseou-se no processo de Orquestração descrito no Capítulo 4. Assim, foram retiradas as disciplinas de laboratórios que complementam o ensino da teoria representado por outra disciplina e disciplinas introdutórias. Além disso, conjunto de disciplinas resultantes foram submetidos ao processo de sumarização, para, então, serem identificadas as respectivas Categorias de Conhecimento das mesmas, o resultado desse processo é apresentado no Quadro 5.3.

Quadro 5.3: Categorização das Disciplinas utilizadas no processo de Planejamento e Avaliação.

Disciplinas da Estrutura Curricular	Categoria de Conhecimento Correspondente
Álgebra Linear	Álgebra Linear
Álgebra Vetorial e Geometria Analítica	Geometria Analítica
Fundamentos de Física	Física
Teoria da Computação	Linguagens Formais Autômatos e Computabilidade
Matemática Discreta	Matemática Discreta
Teoria dos Grafos	Teoria dos Grafos
Estrutura de Dados e Algoritmos	Algoritmo e Estrutura de Dados
Gerência da Informação/ Sistemas de Informação	Fundamentos de Sistemas
Probabilidade e Estatística/ Métodos Estatísticos	Probabilidade e Estatística
Paradigmas de Linguagem de Programação	Linguagem de Programação
Lógica Matemática	Lógica Matemática
Organização e Arquitetura de Computadores	Arquitetura e Organização de Computadores
Engenharia de Software	Engenharia de Software

### Uso no Planejamento Curricular

Baseando-se nas informações apresentadas na coluna **Categoria de Conhecimento Correspondente** do Quadro 5.3, objetivamos planejar as Categorias de Conhecimento: Álgebra Linear (M5), Física (CB1), Linguagens Formais Autômatos e Computabilidade (F7), Algoritmos e Estrutura de Dados (F2), Fundamentos de Sistemas (F6), Probabilidade e Estatística (M8), Linguagem de Programação, Lógica Matemática (M6), Arquitetura e Organização de Computadores (F3) e Engenharia de Software (T7).

Ao analisar as Categorias de Conhecimento selecionadas, com base nos valores apresentados pelo *GOC*, verifica-se que todas as categorias propostas, no contexto dessa pesquisa, apresentam-se como relevantes no ensino da Ciência da Computação, pois apresentam mensuração com valores acima de 0,5. Após essa análise, a próxima etapa, compreende a verificação das Categorias de Conhecimento com as quais elas se relacionam, tal mensuração foi realizada de acordo com os valores do *GDep* identificados.

Ao analisar a mensuração do *GDep* identificado a partir das estruturas curriculares utilizadas, verificou-se que, forma decrescente de apresentação, os relacionamentos

sugeridos para as Categorias de Conhecimento utilizadas descritas anteriormente, pode ser verificado na Tabela 5.8.

Tabela 5.8: Relacionamentos Sugeridos para as Diferentes Categorias de Conhecimento utilizadas na Proposta de Planejamento

Categorias de Conhecimento	Categoria de Conhecimento Pré-requisito
Álgebra Linear	Geometria Analítica
Física	Calculo Diferencial e Integral
Linguagens Formais Autômatos e Computabilidade	Matemática Discreta
	Algoritmo e Estrutura de Dados
	Lógica Matemática
Algoritmo e Estrutura de Dados	Linguagem de Programação
	Matemática Discreta
Fundamentos de Sistemas	Banco de Dados
	Linguagem de Programação
	Engenharia de Software
Probabilidade e Estatística	Calculo Diferencial e Integral
	Matemática Discreta
Linguagem de Programação	Algoritmo e Estrutura de Dados
	Engenharia de Software
	Linguagens Formais Autômatos e Computabilidade
Lógica Matemática	Matemática Discreta
	Linguagens Formais Autômatos e Computabilidade
Arquitetura e Organização de Computadores	Circuitos Digitais
	Linguagem de Programação
Engenharia de Software	Linguagem de Programação
	Lógica Matemática
	Algoritmo e Estrutura de Dados

O resultado apresentado identifica apenas as relações existentes entre as Categorias de Conhecimento propostas para o planejamento. Acredita-se que estas sugestões dos relacionamentos podem ser úteis para determinar possíveis relações de pré-requisito entre as disciplinas que serão planejadas a partir das informações descritas no CR.

### Uso para Avaliação da Estrutura Curricular

O uso do *GDel* no processo de avaliação de uma estrutura curricular confrontou os dados do curso real analisado e as relações de dependências identificadas a partir da mensuração utilizando-se da abordagem de orquestração proposta.

Ao comparar os resultados obtidos a partir do processo de Orquestração de Conhecimento (Tabela 5.8) e os dados do curso real analisado (Tabela 5.9), verificou-se que há semelhanças e diferenças entre ambos. Ao observar os relacionamentos entre as

Tabela 5.9: Relações entre Categorias Apresentadas pelas Estruturas Curricular Analisada

Categorias de Conhecimento	Categoria de Conhecimento Pré-requisito
Álgebra Linear	Geometria Analítica
Física	Calculo Diferencial e Integral
Linguagens Formais Autômatos e Computabilidade	Matemática Discreta
	Algoritmo e Estrutura de Dados
	Teoria dos Grafos
Algoritmo e Estrutura de Dados	Linguagem de Programação
	Teoria dos Grafos
Fundamentos de Sistemas	<i>Não apresentou relações de dependência com outras categorias</i>
Probabilidade e Estatística	Calculo Diferencial e Integral
	Álgebra Linear
Linguagem de Programação	Algoritmo e Estrutura de Dados
	Linguagens Formais Autômatos e Computabilidade
Lógica Matemática	Linguagens Formais Autômatos e Computabilidade
Arquitetura e Organização de Computadores	Física
	Algoritmos e Estrutura de Dados
Engenharia de Software	Linguagem de Programação
	Probabilidade e Estatística

Categorias de Conhecimento do curso real, comparado ao “padrão” identificado, verificam-se que, com base nas estruturas curriculares utilizadas, os seguintes relacionamentos não foram identificados:

- Linguagens Formais, Autômatos e Computabilidade & Teoria dos Grafos;
- Probabilidade e Estatística & Álgebra Linear;
- Engenharia de Software & Probabilidade e Estatística;
- Arquitetura e Organização de Computadores & Física e Estrutura de Dados e Algoritmos.

## 5.2 Considerações Finais do Capítulo

Neste capítulo foi apresentado o estudo de caso realizado que visou demonstrar, de forma aplicada, a abordagem para Orquestração do Conhecimento Curricular proposta. Dentre os resultados apresentados, destacam-se o uso da Categorização de Textos aplicada aos dados extraídos das ementas contidas nas Estruturas Curriculares dos cursos. Este uso, proporcionou uma redução da carga de trabalho por parte do especialista, caso a mesma

tarefa tivesse sido realizadas de forma manual, conforme apresentado na Subseção 5.1.2. Além disso, foi verificado que o uso das métricas *GOC* e *GDep* podem auxiliar os processos de planejamento e avaliação de estruturas curriculares, uma vez que, respectivamente elas podem ser utilizadas para identificar quais as Categorias de Conhecimento mais relevantes e as relações de dependência entre essas categorias, conforme apresentado na Subseção 5.1.3.

De forma geral, com base no resultado apresentado neste capítulo foi possível responder ao questionamento inicial que motivou a realização dessa pesquisa e foram verificadas as hipóteses formuladas na Seção 1.1. Com isso, foi verificado que, no contexto do estudo de caso realizado, as hipóteses não foram refutadas. Logo, verifica-se que a abordagem proposta conseguiu alcançar os objetivos estipulados inicialmente.

# Capítulo 6

## Considerações Finais

Neste capítulo, são apresentadas as considerações finais relacionadas a abordagem de orquestração de conhecimento proposta neste trabalho e as limitações identificadas. Além disso, são propostas algumas sugestões para pesquisas futuras.

### 6.1 Conclusões

A abordagem ora descrita visou propor uma forma de orquestração do conhecimento curricular da Ciência da Computação no Brasil a partir dos dados extraídos de estruturas curriculares de cursos reais e um conjunto de Categorias de Conhecimento baseada no CR.

Inicialmente, observou-se que o conhecimento curricular no contexto da Ciência da Computação no Brasil é descrito, em diferentes níveis de granularidade, no Currículo Referência da SBC. Por sua vez, este documento podem ser utilizado no planejamento de disciplinas dos cursos de bacharelado em Ciência da Computação, conforme verificado durante o processo de mapeamento entre as disciplinas de cursos reais e as categorias de conhecimento curriculares descritas no CR, processo este que apresentou aproximadamente 81% de acertos.

Em decorrência desse processo de CT, também foi possível identificar disciplinas não relacionadas às Categorias de Conhecimento utilizadas na abordagem, ou seja, disciplinas que fazem parte de alguns currículos de cursos analisados, mas que foram criadas fora do escopo determinado pelo CR, para estas disciplinas foram dada a denominação de “disciplinas excedentes”. A importância dessa identificação é verificada quando, a partir dela, foi possível

identificar um conjunto de disciplinas e, conseqüentemente, um conjunto de assuntos estudados que não fazem parte do “padrão” apresentado pela SBC, mas que, em alguns cursos, fazem parte do conhecimento entregue durante o processo de formação do bacharel em Ciência da Computação.

Outro resultado satisfatório foi obtido a partir da verificação das características dos cursos de Ciência da Computação no Brasil, no que diz respeito a organização destes com base nas recomendações descritas pela SBC, mensurados a partir das métricas Grau de Ocorrência da Categoria (*GOC*) e Grau de Dependência entre Categorias (*GDep*).

Entende-se que, neste trabalho, o *GOC* mede o grau de ocorrência de cada Categorias de Conhecimento partir das estruturas curriculares analisadas. O resultado dessa mensuração e, conseqüentemente, a relevância das informações identificadas estão apresentadas nos Quadros 5.1 e 5.2, nos quais é possível identificar a representatividade de cada Categorias de Conhecimento Curricular.

O cálculo do *GDep* também é considerado um aspecto relevante, pois mede grau do relacionamento entre diferentes Categorias de Conhecimento. Nesse trabalho, esta mensuração foi utilizada para orquestrar as 57 Categorias de Conhecimento, conforme descrito na Seção 5.1.2 e detalhado no Apêndice C.2. Esta mensuração também pode ser utilizada para recomendar as relações de dependências entre categorias, auxiliando, assim, o processo de Planejamento Curricular.

O resultado da abordagem ora citada objetivou uma aplicação direta no processo de planejamento e avaliação das estruturas curriculares de cursos, conforme descrito na Subseção 5.1.3. Para a verificação desse uso foi proposto alguns estudos de caso utilizando-se dos índices *GOC* e *GDep* para auxiliar os processos em questão. Os resultados desses estudos apresentaram resultados que atestaram a possibilidade de uso de ambas as métricas nos processos em questão.

## 6.2 Limitações e Sugestões para Pesquisas Futuras

Uma das limitações identificadas nesse processo foi a necessidade da realização de trabalhos manuais para a extração dos dados das ementas, pré-categorização do conjunto treinamento e construção dos conjuntos de dados de testes, uma vez que não há um padrão entre as ementas

para facilitar o uso de técnicas computacionais para extração das informações.

Outra limitação, foi a não verificação da semântica no processo de categorização de ementas, pois foi observado que algumas ementas apresentavam termos sinônimos. Para estes casos, foi necessária a realização da fase da Avaliação da Categorização, conforme apresentado no Capítulo 4, para verificar a categorização da disciplina. Um exemplo dessa não observação semântica pôde ser verificado nos casos em que “Sistemas Embarcados” foi escrito como “Sistemas Embutidos”, formas diferentes, mas com mesmo significado semântico. Além disso, ressalta-se a importância do incremento do conjunto treinamento utilizado na tarefa de Categorização da Ementas.

Como trabalho futuro, destaca-se a importância da automatização dos processos de Análise e Sumarização das Categorias, apresentado no processo de Orquestração do Conhecimento Curricular (Seção 4.1), bem como, no processo de auxílio do planejamento e verificação da conformidade das estruturas curriculares (Seção 4.2), beneficiando, assim, a maximização do uso da abordagem proposta.

Por fim, espera-se que a abordagem ora citada contribua não apenas para o processo de planejamento e avaliação de estruturas curriculares, mas também, para a modelagem do domínio de sistemas que necessitam de uma representação curricular. Além disso, acredita-se que outros domínios de conhecimento poderão utilizar a abordagem proposta, desde que a sua representação do conhecimento possa ser descrita em um conjunto de categorias e termos relacionadas à elas, para seja possível a aplicação das técnicas de Mineração de Textos proposta.

Contudo, para as áreas que não possuam uma estrutura de representação bem definida, semelhante ao CR da SBC, pode ser utilizada uma adaptação da abordagem apresentada, de modo que a identificação das categorias seja também realizada de forma *bottom-up* utilizando-se técnicas de agrupamento. Assim, poderiam ser agrupadas disciplinas semelhantes e, a partir da análise desses agrupamentos, ser identificado um termo comum entre que possa representar a Categoria de Conhecimento descrita nos *clusters*. Com isso, acredita-se que o resultado desse processo seria um conjunto formado por  $n$  agrupamentos, cada um, representando uma possível categoria de conhecimento obtida a partir da fusão de diferentes estruturas curriculares.



# Bibliografia

AMERICAN SOCIETY OF CIVIL ENGINEERS. *Civil Engineering Body of Knowledge for the 21st Century*. 2008. <[http://www.asce.org/uploadedFiles/Education\\_and\\_Careers/Body\\_of\\_Knowledge/Content\\_Pieces/body-of-knowledge.pdf](http://www.asce.org/uploadedFiles/Education_and_Careers/Body_of_Knowledge/Content_Pieces/body-of-knowledge.pdf)>. [Online; Último acesso em 16 de janeiro de 2016].

ASSOCIATION FOR COMPUTING MACHINERY. *Computing Curricula 2001 Computer Science*. 2001. <[https://www.acm.org/education/curric\\_vols/cc2001.pdf](https://www.acm.org/education/curric_vols/cc2001.pdf)>. [Online; Último acesso em 02 de janeiro de 2016].

ASSOCIATION FOR COMPUTING MACHINERY. *Computer Science Curricula*. 2008. <<https://www.acm.org/education/curricula/ComputerScience2008.pdf>>. [Online; Último acesso em 02 de janeiro de 2016].

ASSOCIATION FOR COMPUTING MACHINERY. *Curriculum Guidelines for Undergraduate Degree Programs in Information Systems*. 2010. <<https://www.acm.org/education/curricula/IS%202010%20ACM%20final.pdf>>. [Online; Último acesso em 17 de dezembro de 2015].

ASSOCIATION FOR COMPUTING MACHINERY. *Computing Science Curricula*. 2013. <<https://www.acm.org/education/CS2013-final-report.pdf>>. [Online; Último acesso em 03 de janeiro de 2016].

BALAKRISHNAN, R.; KANIIMOZHI, M. *Overview of Text Mining*. 2015. <[http://ijarcsse.com/docs/papers/Volume\\_5/9\\_September2015/V5I9-0210.pdf](http://ijarcsse.com/docs/papers/Volume_5/9_September2015/V5I9-0210.pdf)>. [Online; Último acesso em 21 de novembro de 2015].

BATISTA, D. et al. *Nova grade curricular do BCC-IME-USP*. 2015. <<http://www.lbd.dcc.ufmg.br/colecoes/wei/2015/015.pdf>>.

CARVALHO, V. N. de. Planejamento educacional: Organização de estratégias e superação de rotinasou protocolo instrucional? 2010.

CHOY, M. *Effective Listings of Function Stop words for Twitter*. 2012. <<http://goo.gl/kccM3v>>. [Online; Último acesso em 10 de janeiro 2016].

CONSELHO NACIONAL DE EDUCAÇÃO. *Diretrizes Curriculares Nacionais para os cursos de graduação em Computação*. 2012. <[http://portal.mec.gov.br/component/docman/?task=doc\\_download&gid=11205&Itemid=>](http://portal.mec.gov.br/component/docman/?task=doc_download&gid=11205&Itemid=>)>. [Online; Último acesso em 15 de janeiro de 2016].

- DEPARTAMENTO DE SISTEMAS DE COMPUTAÇÃO. *Plano de Curso*. 2010. <<http://www.computacao.ufcg.edu.br/graduacao/plano-de-curso>>. [Online; Último acesso em 12 de dezembro de 2015].
- DICIONÁRIO MICHAELIS. *Orchestration*. 2016. <<http://michaelis.uol.com.br/moderno/portugues/index.php>>. [Online; Último acesso em 21 de janeiro de 2016].
- DILLENBOURG, P. *Design for Orchestration*. 2011. <[http://www.teleurope.eu/mod/file/download.php?file\\_guid=93714](http://www.teleurope.eu/mod/file/download.php?file_guid=93714)>. [Online; Último acesso em 21 de janeiro de 2016].
- FELDMAN, R.; SANGER, J. *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. [S.l.]: Cambridge University Press, 2007. ISBN-13 978-0-511-33507-5.
- FERAUCHE, T. *Aplicações de Técnicas de Mineração de Textos para Classificação de Ementas da Jurisprudência da Justiça do Trabalho de São Paulo*. 2011. <<http://www.centropaulasouza.sp.gov.br/pos-graduacao/trabalhos-academicos/dissertacoes/formacao-tecnologica/2011/thiago-ferauche.pdf>>. CEETEPS Programa de Pós-graduação.
- FILATRO, A. *Design Instrucional na Prática*. [S.l.]: Pearson Prentice Hall, 2008. ISBN 978-85-7605-188-6.
- FILHO, M. P. A.; SILVA, F. A. T. F. D.; BRAGA, A. P. S. *Operador Morfológico Adaptativo de Casamento de Padrões – Proposte a Aplicação nas Imagens de Satélites*. 2015. <[www.sbai2015.dca.ufrn.br/download/artigo/450](http://www.sbai2015.dca.ufrn.br/download/artigo/450)>. [Online; Último acesso em 16 de dezembro de 2015].
- GANESAN, K. *A Brief Note on Stop Words for Text Mining Retrieval*. 2016. <<https://goo.gl/O2vgy>>. [Online; Último acesso em 10 de janeiro 2016].
- GERAB, F.; BUENO, I. A. M.; GERAB, I. F. da S. *Análise das Interações Curriculares em um Curso de Ciência da Computação: buscando subsídios para aprimoramento curricular*. 2014. <<http://dx.doi.org/10.5753/RBIE.2014.22.01.30>>. [Online; Último acesso em 12 de dezembro de 2015].
- GIRAFFA, L. M. M. *Fundamentos de Sistemas Tutores Inteligentes*. 1998. <<http://docplayer.com.br/11577556-Fundamentos-de-sistemas-tutores-inteligentes.html>>. [Online; Último acesso em 19 de dezembro de 2015].
- GONÇALVES, L. S. M. *Categorização em Text Mining*. 2002. <<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-22062015-202748/publico/LeaSilviaMGoncalves.pdf>>. [Online; Último acesso em 21 de novembro de 2015].
- HALL, M. et al. *The WEKA data mining software: an update*. 2009. <<http://dl.acm.org/citation.cfm?id=1656278>>.
- IEEE COMPUTER SOCIETY. *Software Engineering Body of Knowledge*. 2014. <[www.swebok.org](http://www.swebok.org)> [Online; Último acesso em 17 de janeiro de 2016].
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. *Portaria Inep nº 238*. 2014. <[http://download.inep.gov.br/educacao\\_superior/enade/legislacao/2014/diretrizes\\_cursos\\_diplomas\\_bacharel/diretrizes\\_bacharel\\_computacao.pdf](http://download.inep.gov.br/educacao_superior/enade/legislacao/2014/diretrizes_cursos_diplomas_bacharel/diretrizes_bacharel_computacao.pdf)>. [Online; Último acesso em 15 de janeiro de 2016].

- KATSAMANI, M.; RETALIS, S. Orchestrating learning activities using the cadmos learning design tool. 2013. [Online; Último acesso em 20 de dezembro de 2015 ].
- KOEHLER, M. J.; MISHRA, P. *What Is Technological Pedagogical Content Knowledge?* 2009. <<http://www.citejournal.org/vol9/iss1/general/article1.cfm>>. [Online; Último acesso em 13 de janeiro de 2016].
- LINDEN, G. S. *Combinação de classificadores na categorização de textos*. 2008. <<http://hdl.handle.net/10923/1454>>. [Dissertação de Mestrado em Ciência da Computação - Pontifícia Universidade Católica do Rio Grande do Sul].
- LOUSNAK, A. L. de A. et al. *Currículo de Referência da SBC e a Grade Curricular do BCC*. 2002. <<http://bcc.ime.usp.br/principal/documentos/bccs.pdf>>.
- MARSHALL, L. *A graph-based framework for comparing curricula*. 2014. Department of Computer Science University of Pretoria.
- MCCALLUM, A.; NIGAM, K. *A Comparison of Event Models for Naive Bayes Text Classification*. 1998. <<http://www.cs.cmu.edu/~knigam/papers/multinomial-aaaiws98.pdf>>. [Online; Último acesso em 15 de dezembro 2015].
- MIEL, A. *A view of Curriculum Content*. 1956. <[http://www.ascd.org/ASCD/pdf/journals/ed\\_lead/el\\_195603\\_miel.pdf](http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_195603_miel.pdf)>. [Online; Último acesso em 19 de janeiro de 2016].
- MORNAR, N. H.-B. e V. *Navigation Support in a web-based Adaptive Educational Hypermedia System*. 2001. <<http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1465&context=amcis2001>>. [Online; Último acesso em 20 de janeiro de 2016].
- ORKESON, A. *Text & Data Mining - A Librarian Overview*. 2013. <<http://library.ifla.org/252/1/165-okerson-en.pdf>>. [Online; Último acesso em 11 de janeiro de 2016].
- PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO. *Projeto Pedagógico do Curso de Graduação - Ciência da Computação*. 2008. <<http://www.inf.puc-rio.br/wp-content/uploads/2009/05/projeto-pedagogico-ciencia-computacao-2008.pdf>>.
- PRIETO, L. P. et al. *Orchestrating technology enhanced learning: a literature review and a conceptual framework*. 2011. <[http://www.gsic.uva.es/~lprisan/Prieto2011\\_OrchestratingTELLiteratureReviewConceptualFramework.pdf](http://www.gsic.uva.es/~lprisan/Prieto2011_OrchestratingTELLiteratureReviewConceptualFramework.pdf)>. [Online; Último acesso em 21 de janeiro de 2016].
- RAMASUBRAMANIAN, C.; RAMYA, R. *Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm*. 2013. <[http://www.ijarcce.com/upload/2013/december/IJARCCE1G-S-worldnow-\\_EFFIECTIVE.pdf](http://www.ijarcce.com/upload/2013/december/IJARCCE1G-S-worldnow-_EFFIECTIVE.pdf)>. [Online; Último acesso em 10 de janeiro 2016].
- RIZZI, C. B. et al. *Fazendo uso da categorização de textos em atividades empresariais*. 2000. <[https://www.academia.edu/1523364/Fazendo\\_uso\\_da\\_categoriza%C3%A7%C3%A3o\\_de\\_textos\\_em\\_atividades\\_empresariais](https://www.academia.edu/1523364/Fazendo_uso_da_categoriza%C3%A7%C3%A3o_de_textos_em_atividades_empresariais)>. Último acesso em 10 de fevereiro de 2016.

ROSCHELL, J.; DIMITRIADIS, Y.; HOPPE, U. *Classroom orchestration: Synthesis*. 2013. <doi:10.1016/j.compedu.2013.04.010>. [Online; Último acesso em 22 de janeiro de 2016].

SANTOS, L. P. P. *The many faces of orchestration: Towards a (more) operative definition*. 2010. <<https://wikis.gsic.uva.es/gs2/images/d/d4/20100402-Orchestration-enviado.pdf>>. [Online; Último acesso em 10 de dezembro de 2015].

SEBASTIANI, F. *Machine Learning in Automated Text Categorization*. 2002. <<http://nmis.isti.cnr.it/sebastiani/Publications/ACMCS02.pdf>>. [Online; Último acesso em 114 de janeiro de 2016].

SILVA, C. F. da. *Uso de informações linguísticas na etapa de pré-processamento em Mineração de textos*. 2004. <"<http://osorio.wait4.org/alumni/mestrado/Dissertacao-Cassiana-Silva.pdf>">. [Online; Último acesso em 11 de janeiro 2016].

SIMONS, J.; PORTER, N. *Knowledge and the Curriculum*. [S.l.]: Policy Exchange, 2015. ISBN: 978-1-910812-01-3.

SOARES, M. V. B.; PRATI, R. C.; MONARD, M. C. *PreText II: Descrição da Reestruturação da Ferramenta de Pré-Processamento de Textos*. 2008. <[http://www.icmc.usp.br/CMS/Arquivos/arquivos\\_enviados/BIBLIOTECA\\_113\\_RT\\_333.pdf](http://www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_333.pdf)>. [Online; Último acesso em 10 de janeiro 2016].

SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. *Currículo Referência - Sistemas de Informação*. 2003. <<http://www.sbc.org.br/documentos-da-sbc/summary/131-curriculos-de-referencia/764-curriculo-de-referencia-is-versao-2003>>. [Online; Último acesso em 15 de janeiro de 2016].

SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. *Currículo de Referência da SBC para Cursos de Graduação em Bacharelado em Ciência da Computação e Engenharia de Computação*. 2005. <<http://www.sbc.org.br/documentos-da-sbc/send/131-curriculos-de-referencia/760-curriculo-de-referencia-cc-ec-versao2005>>. [Online; Último acesso em 18 de fevereiro de 2016].

SU, S.-W. *The Various Concepts of Curriculum and the Factors Involved in Curricula-making*. 2012. <<http://ojs.academypublisher.com/index.php/jltr/article/view/jltr0301153158>>. [Online; Último acesso em 13 de dezembro de 2015].

TAKEUCH, H.; NONAKA, I. *Gestão do Conhecimento*. [S.l.]: Bookman, 2004. ISBN 0-470-82074-8.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to Data Mining*. [S.l.]: Pearson, 2006. ISBN 0-321-32136-7.

THE GLOSSARY OF EDUCATION REFORM. *Content Knowledge*. 2016. <<http://edglossary.org/content-knowledge/>>. [Online; Último acesso em 20 de janeiro de 2016].

UNESCO. *Curriculum*. 2015. <<http://www.unesco.org/new/en/education/themes/strengthening-education-systems/quality-framework/core-resources/curriculum>>. [Online; Último acesso em 19 de janeiro de 2016].

---

UNIVERSIDADE FEDERAL DE SÃO CARLOS. *Projeto Pedagógico - Ciência da Computação*. 2010. <[http://www.prograd.ufscar.br/projetoped/ppc\\_CIENCIA\\_COMPUTACAO.pdf](http://www.prograd.ufscar.br/projetoped/ppc_CIENCIA_COMPUTACAO.pdf)>.

UNIVERSIDADE FEDERAL DO CEARÁ. *Projeto Pedagógico do Curso de Ciência da Computação*. 2013. <<http://www.campusrussas.ufc.br/grades/projeto-pedagogico-CC.pdf>>.

VIDHYA; AGHILAI, G. Text mining process, techniques and tools an overview. *International Journal Of Information Technology And Knowledge Management*, v. 2, p. 613–622, 2010.

# Apêndice A

## Avaliação dos Algoritmos de Categorização

Para a avaliação do modelo de categorização e, conseqüentemente, do algoritmo de categorização foram utilizados os dados do conjunto treinamento<sup>1</sup>. Essa avaliação visou identificar, dentre os algoritmos utilizados, qual aquele que apresenta melhor resultado e que será utilizado no processo de Categorização das Ementas.

Há uma diversidade de algoritmos para CT, porém, nesta abordagem, optou-se pelo uso e avaliação dos algoritmos Redes Bayesianas, *Support Vector Machine* e Árvores de Decisão. Os documentos utilizados foram submetidos à etapa de pré-processamento, as quais foram aplicados os processos de redução da dimensionalidade de dados, utilizando-se do *stemming*, retirada das *stopwords*, menor frequência do termo igual a 4 (quatro), e o IDF-ITF Transform.

Para o algoritmo probabilístico de Rede Bayesiana, foi utilizada a Rede Bayesiana Multinomial devido a literatura que apresentou uma redução de erros quando comparada a versão simplificada (MCCALLUM; NIGAM, 1998). Assim, nessa avaliação foram utilizadas as implementações Naive Bayes Multinomial, J48 e SVO presentes na ferramenta WEKA. Além disso, foram verificadas o *holdout*, *cross-validation*, com  $k = 10$ , e a ressubstituição. O resultado da execução é apresentado na Tabela A.1.

Com base nos resultados apresentados na Tabela A.1 é possível observar que em todos os casos, a porcentagem de precisão, recall e F-measure, apresentou um valor alto, porém, esta medida apresenta-se como falsa, uma vez que, neste caso, todos o conjunto treinamento

---

<sup>1</sup>Disponível em: <<https://goo.gl/jwludW>>

Tabela A.1: Sumarização dos Algoritmos de Categorização

Algoritmo	Avaliação	Precisão	Recall	F-measure
Naive Bayes Multinomial	Ressubstituição	0,98	0,981	1
	k-cross validation	0,828	0,846	0,829
	holdout	0,755	0,791	0,763
SMO	Ressubstituição	1	1	1
	k-cross validation	0,795	0,794	0,786
	holdout	0,75	0,695	0,693
J48	Ressubstituição	0,851	0,856	0,847
	k-cross validation	0,605	0,588	0,579
	holdout	0,489	0,458	0,438

é utilizado para teste.

Se observadas as outras metodologias de avaliação, verifica-se que o algoritmo Naive Bayes Multinomial é aquele que, neste estudo, apresenta-se como sendo o mais acurado, pois, apresentou valores superiores aos identificados em outros algoritmos em todas as métricas observadas.

Portanto, o algoritmo Naive Bayes Multinomial será utilizado na fase de Categorização de Ementas. Ressalta-se que o modelo do arquivo ARFF utilizado nesta avaliação, está apresentado no Apêndice C.3.

# Apêndice B

## Categorização das Ementas

### B.1 Arquivos ARFF dos Cursos Orquestrados

Foi criado, para cada Estrutura Curricular utilizada no processo de Orquestração, um arquivo ARFF contendo, em cada linha, a descrição de uma disciplina. Os arquivos de cada Estrutura Curricular estão disponibilizados nos endereços abaixo:

- **Curso 1** – <<https://goo.gl/bZDokd>>
- **Curso 2** – <<https://goo.gl/9k13BJ>>
- **Curso 3** – <<https://goo.gl/UrKbLd>>
- **Curso 4** – <<https://goo.gl/9IDgl0>>
- **Curso 5** – <<https://goo.gl/CbBdpn>>
- **Curso 6** – <<https://goo.gl/QPqkmt>>
- **Curso 7** – <<https://goo.gl/XcWQR4>>

### B.2 Resultado Bruto e Avaliação da Categorização dos Cursos

As distribuições de probabilidade da CT de cada curso utilizado, bem como a avaliação desses, podem ser verificados de forma eletrônica. Para isso, é apresentado para cada curso



o endereço correspondente ao resultado.

Ressalta-se que a planilha eletrônica apresentada nos links é composta por 3 (três) abas. Nessas planilhas é possível verificar o conjunto de disciplinas submetidas à CT e as dependências entre elas (aba **Lista de Disciplinas**), o resultado da CT e as análises realizadas (aba **Categorização**) e os dados utilizados para análise do *GOC* e *GDep* (aba **Análise da Grade Curricular**).

Quanto à Análise da Categorização, ressalta-se apenas as linhas destacadas com preenchimento de cor cinza foram submetidas ao processo de Avaliação e o resultado dessa fase foi inserido na coluna **avaliação\_cat**, conforme pode ser verificado nos links abaixo:

- **Curso 1** – <<https://goo.gl/GpJGyS>>
- **Curso 2** – <<https://goo.gl/LH8vGS>>
- **Curso 3** – <<https://goo.gl/sdbAHk>>
- **Curso 4** – <<https://goo.gl/WZkDZf>>
- **Curso 5** – <<https://goo.gl/3hbAor>>
- **Curso 6** – <<https://goo.gl/pptc1i>>
- **Curso 7** – <<https://goo.gl/L6kYhH>>

## B.3 Modelo de Arquivo ARFF

Figura B.1: Modelo do Arquivo ARFF

```
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, artmetica:
15 "logica, tecnicas de demonstracao e validacao d
```

# Apêndice C

## Resultados Detalhados: *GOC* e *GDep*

### C.1 Detalhamento *GOC*

Nesta Seção será detalhado o cálculo do *GOC* com distinção entre disciplinas obrigatórias e optativas. O detalhamento por núcleo de conhecimento é apresentado nas Tabelas C.1, C.2, C.3, C.4, C.5 e C.6.

Tabela C.1: Resultado *GOC* do núcleo Matemática

Categoria de Conhecimento	Obrigatoria	Optativa	Geral	GOC obrigatório	GOC Optativa	GOC
M1. Álgebra Linear	7		7	1,00	0,00	1,00
M2. Análise Combinatória			0	0,00	0,00	0,00
M3. Cálculo Diferencial e Integral	6		6	0,86	0,00	0,86
M4. Equações Diferenciais	1	2	3	0,14	0,29	0,43
M5. Geometria Analítica	5	1	6	0,71	0,14	0,86
M6. Lógica Matemática	6		6	0,86	0,00	0,86
M7. Matemática Discreta	7		7	1,00	0,00	1,00
M8. Probabilidade e Estatística	7		7	1,00	0,00	1,00
M9. Variáveis Complexas			0	0,00	0,00	0,00

Tabela C.2: Resultado *GOC* do núcleo Ciências Naturais

Categoria de Conhecimento	Obrigatória	Optativa	Geral	GOC obrigatório	GOC Optativa	GOC
CB1. Física	4	2	6	0,57	0,29	0,86

Tabela C.3: Resultado GOC do núcleo Eletrônica

Categoria de Conhecimento	Obrigatoria	Optativa	Geral	GOC obrigatório	GOC Optativa	GOC
E1. Circuitos Eletrônicos		1	1	0,00	0,14	0,14

Tabela C.4: Resultado GOC do núcleo Fundamentos da Computação

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

Tabela C.5: Resultado GOC do núcleo Tecnologia da Computação

Categoria de Conhecimento	Obrigatori a	Optativa	Geral	GOC obrigatório	GOC Optativa	GOC
T1. Análise de Desempenho		2	2	0,00	0,29	0,29
T2. Bancos de Dados	7		7	1,00	0,00	1,00
T3. Circuitos Integrados		1	1	0,00	0,14	0,14
T4. Compiladores	6		6	0,86	0,00	0,86
T5. Computação Gráfica	6	1	7	0,86	0,14	1,00
T6. Automação e Controle		2	2	0,00	0,29	0,29
T7. Engenharia de Software	7		7	1,00	0,00	1,00
T8. Inteligência Artificial	4		4	0,57	0,00	0,57
T9. Interação Humano- Computador	5	2	7	0,71	0,29	1,00
T10. Matemática Computacional	5		5	0,71	0,00	0,71
T11. Métodos Formais		1	1	0,00	0,14	0,14
T12. Modelagem e Simulação	1		1	0,14	0,00	0,14
T13. Processamento Digital de Sinais			0	0,00	0,00	0,00
T14. Processamento de Imagens	2	2	4	0,29	0,29	0,57
T15. Programação Paralela	2	3	5	0,29	0,43	0,71
T16. Redes de Computadores	7		7	1,00	0,00	1,00
T17. Segurança e Auditoria de Sistemas	1	3	4	0,14	0,43	0,57
T18. Sistemas Digitais		1	1	0,00	0,14	0,14
T19. Sistemas Distribuídos	3	4	7	0,43	0,57	1,00
T20. Sistemas Embarcados	1	4	5	0,14	0,57	0,71
T21. Sistemas Multimídia	2	2	4	0,29	0,29	0,57
T22. Tolerância a Falhas		1	1	0,00	0,14	0,14
T23. Telecomunicações		1	1	0,00	0,14	0,14

Tabela C.6: Resultado *GOC* do núcleo Contexto Social e Profissional

Categoria de Conhecimento	Obrigatori a	Optativa	Geral	GOC obrigatório	GOC Optativa	GOC
P1. Administração		4	4	0,00	0,57	0,57
P2. Computadores e Sociedade	7		7	1,00	0,00	1,00
P3. Comunicação e Expressão	1		1	0,14	0,00	0,14
P4. Contabilidade e Custos		1	1	0,00	0,14	0,14
P5. Direito e Legislação	1		1	0,14	0,00	0,14
P6. Economia		1	1	0,00	0,14	0,14
P7. Empreendedorismo	1	5	6	0,14	0,71	0,86
P9. Filosofia	2		2	0,29	0,00	0,29
P10. Informática na Educação		5	5	0,00	0,71	0,71
P11. Inglês	2		2	0,29	0,00	0,29
P12. Métodos Quantitativos Aplicados à Administração de Empresas		2	2	0,00	0,29	0,29
P13. Sociologia	1		1	0,14	0,00	0,14
P14. Psicologia		1	1	0,00	0,14	0,14

## C.2 Detalhamento *GDep*

A Tabela C.7 apresenta o resultado geral do Grau de Dependência entre Categorias (*GDep*). Na Tabela, foram excluídas as linhas e colunas das categorias que não apresentaram nenhuma relação de dependência. O resultado completo, incluindo todas as Categorias de Conhecimento, pode ser consultado no endereço: <<https://goo.gl/KY9Vlu>>.

