

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO
CONHECIMENTO

WEBERT JÚNIO ARAÚJO

**PROPOSTA METODOLÓGICA PARA ENRIQUECIMENTO DE
ONTOLOGIAS DE DOMÍNIO**

Belo Horizonte

2021

Webert Júnio Araújo

PROPOSTA METODOLÓGICA PARA ENRIQUECIMENTO DE ONTOLOGIAS DE DOMÍNIO

Tese apresentada ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Doutor em Gestão e Organização do Conhecimento.

Área de concentração: Ciência da Informação.

Linha de Pesquisa: Arquitetura e Organização do Conhecimento

Orientadora: Profa. Dra. Gercina Ângela de Lima

Belo Horizonte

2021

A663p

Araújo, Webert Júnio.

Proposta metodológica para enriquecimento de ontologias de domínio [recurso eletrônico] / Webert Júnio Araújo. - 2021.

1 recurso online (271 f. : il., color.) : pdf.

Orientadora: Gercina Ângela de Lima

Tese (Doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 209-220.

Apêndices: f. 221-263.

Anexos: f. 264-271.

Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da informação – Teses. 2. Ontologias (Recuperação da informação) – Teses. 3. Organização da informação – Teses. I. Título. II. Lima, Gercina Ângela Borém de Oliveira. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4.03

Ficha catalográfica: Rosimeire Silva Campos de Lima CRB:6/3145

Biblioteca Profª Etelvina Lima, Escola de Ciência da Informação da UFMG.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

ATA DA DEFESA DE TESE DO ALUNO

WEBERT JUNIO ARAUJO

Realizou-se, no dia 17 de setembro de 2021, às 14:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *Proposta metodológica para enriquecimento de ontologias de domínio*, apresentada por WEBERT JUNIO ARAUJO, número de registro 2017659554, graduado no curso de BIBLIOTECONOMIA, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Gercina Ângela de Lima - ECI/UFMG (Orientadora), Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG, Prof(a). Célia da Consolação Dias - ECI/UFMG, Prof(a). Maria Luiza de Almeida Campos - UFF, Dr(a). Ivo Pierozzi Júnior - Embrapa, Prof(a). Cristiane Mendes Netto - UNIVALE, Prof(a). Lucinéia Souza Maia - UFOP.

A Comissão considerou a tese:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 17 de setembro de 2021.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Gercina Angela de Lima, Professora do Magistério Superior**, em 17/09/2021, às 23:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Celia da Consolação Dias, Coordenador(a) de curso de pós-graduação**, em 20/09/2021, às 11:15, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Mendes Netto, Usuário Externo**, em 20/09/2021, às 14:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Luiza de Almeida Campos, Usuário Externo**, em 20/09/2021, às 14:45, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 20/09/2021, às 14:54, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ivo Pierozzi Junior, Usuário Externo**, em 21/09/2021, às 08:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Lucinéia Souza Maia, Usuário Externo**, em 21/09/2021, às 15:13, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0967899** e o código CRC **8177A0F2**.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

FOLHA DE APROVAÇÃO

Proposta metodológica para enriquecimento de ontologias de domínio

WEBERT JUNIO ARAUJO

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

Aprovada em 17 de setembro de 2021, por videoconferência, pela banca constituída pelos membros:

Prof(a). Gercina Ângela de Lima (Orientadora)
ECI/UFMG

Prof(a). Benildes Coura Moreira dos Santos Maculan
ECI/UFMG

Prof(a). Célia da Consolação Dias
ECI/UFMG

Prof(a). Maria Luiza de Almeida Campos
UFF

Dr(a). Ivo Pierozzi Júnior
Embrapa

Prof(a). Cristiane Mendes Netto
UNIVALE

Prof(a). Lucinéia Souza Maia
UFOP

Horizonte, 17 de setembro de 2021.



Documento assinado eletronicamente por **Gercina Angela de Lima, Professora do Magistério Superior**, em 17/09/2021, às 23:56, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Celia da Consolação Dias, Coordenador(a) de curso de pós-graduação**, em 20/09/2021, às 11:15, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristiane Mendes Netto, Usuário Externo**, em 20/09/2021, às 14:01, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Luiza de Almeida Campos, Usuário Externo**, em 20/09/2021, às 14:44, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 20/09/2021, às 14:54, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ivo Pierozzi Junior, Usuário Externo**, em 21/09/2021, às 08:00, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Lucinéia Souza Maia, Usuário Externo**, em 21/09/2021, às 15:13, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0967915** e o código CRC **CE7A602B**.

DEDICATÓRIA

*Dedico esta tese à minha mãe, a minha
Rosa.*

AGRADECIMENTOS

A Deus, merecedor de toda honra e Glória, que torna tudo isto realizável, que me dá forças para vencer todos os obstáculos e não desanimar frente às dificuldades.

Aos meus pais, que sempre acreditaram na importância dos estudos e que não mediram esforços para que eu pudesse ter uma boa educação. Obrigado por cada gota de suor que derramaram para que eu pudesse chegar até aqui. Em especial a minha mãe, mulher de força, guerreira que me ensinou todos os valores, mesmo com dificuldades nunca deixou de priorizar a educação, se esforçando para suprir todas as necessidades. Mulher responsável por colocar o primeiro da família em um curso superior. Meu maior exemplo de vida.

À minha esposa, pelo carinho, por acreditar nos meus sonhos e motivar-me sempre por estar sempre ao meu lado em todos os momentos (muitos deles difíceis). Obrigado pelo seu amor e cuidado e por todo incentivo. Obrigado por entender minhas ausências e momentos de ansiedade. Te amo!

Ao meu filho Isaac, por ter trazido tanta alegria e leveza, apesar de todas as provações e lutas. Um amor imensurável. Obrigado pela compreensão... por brincar, “estudar” e “trabalhar” ao meu lado.

À minha Tia Glau (in memoriam), que sempre me encorajou a perseguir os meus sonhos, não importando com o tamanho dos obstáculos. Obrigado por me ensinar, ainda que em outras palavras, que “quem é dono de uma vontade firme modela o mundo ao seu feitio.”

Aos meus irmãos, Kel, Tim e Taka, que sempre se orgulharam de mim e que me motivam a continuar em busca dos meus sonhos.

À minha orientadora Gercina Lima, que acreditou no meu potencial e me guiou nos caminhos complexos da pesquisa desde a graduação. Obrigado, Professora, pela dedicação e paciência, pela forma como você acredita nos seus orientandos e os motiva a fazer sempre o melhor. Obrigado por ensinar e orientar com o coração. Obrigado por ver cada orientando como um ser humano!

À professora Benildes, exemplo de pesquisadora, e que muito me auxiliou na seleção do mestrado. Obrigado por sempre aceitar participar das minhas bancas e por toda paciência.

À professora Célia Dias, por todo incentivo desde às aulas da graduação e por ter aceito participar das minhas bancas.

Às professoras Maria Luiza, Cristiane e Lucinéia por terem aceito participar da banca prontamente e que são fontes de inspiração.

Ao Ivo Pierozzi Júnior, pela paciência e disponibilidade para as conversas por videoconferência.

Ao Caio, que é uma fonte de inspiração para que eu acredite na realização de todos os meus sonhos.

A todos os meus familiares, por acreditarem sempre no meu potencial e por todas as mensagens de incentivo.

Ao Júnio Lopes por todo seu empenho em me ajudar na normalização do trabalho.

Aos colegas do CEFET-MG Contagem.

Aos professores da ECI, nos quais me inspirei para buscar o conhecimento sempre.

Aos companheiros do Grupo MHTX, pelas prévias, discussões e todas as experiências que vivemos juntos.

Aos funcionários da Escola de Ciência da Informação.

Às agências de fomento.

*“O que sabemos é uma gota; o que
ignoramos é um oceano”
(Isaac Newton)*

RESUMO

Introdução: As ontologias de domínio são construídas com objetivos determinados por meio da captura de conhecimento de especialistas de domínio, assim como na extração de conteúdo de fontes de informação que explicitam o conhecimento do domínio. Porém o conhecimento é algo que muda com descobertas, com mudanças de paradigmas ou até mesmo com refutações de teorias consolidadas. Diante da provisoriedade do conhecimento humano, os desenvolvedores de ontologias precisam garantir que esses instrumentos passem por processo de atualização sempre que houver necessidade. **Objetivo:** Esta pesquisa teve como objetivo norteador propor uma metodologia para o processo de enriquecimento de ontologias de domínio com base em construtos da engenharia ontológica e da Biblioteconomia e Ciência da Informação e que superasse as lacunas dos atuais métodos para enriquecimento de ontologias de domínio. **Motivação/Problema:** A escolha desse tema se respalda nos problemas relacionados à incipiência de estudos sobre manutenção e atualização de ontologias, a provisoriedade do conhecimento humano e conseqüentemente a necessidade de atualizar as ontologias, assim como a inexistência de um guia metodológico sobre como conduzir o processo de enriquecimento de ontologias de domínio. **Pressupostos:** os seguintes pressupostos nortearam o desenvolvimento desta investigação: (1) o enriquecimento de ontologias pode contribuir para que as ontologias representem o domínio com mais completude; (2) as metodologias de enriquecimento facilitam a atualização de conhecimento em ontologias de domínio; (3) a Biblioteconomia e Ciência da Informação possui arcabouços teórico-metodológicos que podem auxiliar no enriquecimento de ontologias de domínio. **Fundamentação:** na fundamentação teórica e metodológica abordaram-se as temáticas manutenção e atualização de Sistemas de Organização do Conhecimento, Teoria do Conceito, Ontologias, Engenharia Ontológica e Enriquecimento de Ontologia. **Metodologia:** A pesquisa caracteriza-se como aplicada, quanto à natureza, exploratória, quanto aos objetivos e bibliográfica no que concerne aos procedimentos metodológicos. As etapas propostas na metodologia desta investigação são cinco: (1) pesquisa bibliográfica; (2) revisão sistemática da literatura; (3) seleção de documentos sobre manutenção e atualização de sistemas de organização do conhecimento; (4) aplicação da Análise de Conteúdo no material empírico selecionado; (5) desenvolvimento e validação da metodologia de enriquecimento. **Resultados:** como resultado, apresenta-se a METHODOE (Methodology for Domain Ontology Enrichment). A METHODOE possui cinco fases que possibilitam a realização de todas as atividades relacionadas ao processo de enriquecimento, desde o seu planejamento até o pós-enriquecimento. Entre as funcionalidades e características dessa metodologia, destacam-se: (1) apresentação de uma visão sistemática do processo de enriquecimento; (2) independência de domínio e de ferramentas específicas, possibilitando a aplicação em ontologias de domínios diversos; (3) independência de fontes de informação específicas e de difícil acesso; (4) apresentação de forma detalhada da maioria das etapas e com algumas exemplificações; (5) possibilidade da realização de todos os tipos de enriquecimento (lexical, conceitual, de relações e de axiomas).

Palavras-chave: Enriquecimento de ontologia. Atualização de ontologia. Ontologia de domínio. Sistema de Organização do Conhecimento.

ABSTRACT

Introduction: Domain ontologies are built with determined goals through the capture of knowledge from domain experts, as well as the extraction of content from information sources that make domain knowledge explicit. However, knowledge is something that changes with discoveries, with paradigm shifts, or even with rebuttals of consolidated theories. Given the temporary nature of human knowledge, ontology developers need to ensure that these instruments go through an update process whenever necessary.

Objective: This research aimed to propose a methodology for the process of domain ontologies enrichment based on constructs from ontological engineering and Library and Information Science and that would overcome the gaps of current methods for enriching domain ontologies.

Motivation/Problem: The choice of this theme is based on problems related to the incipience of studies on maintenance and updating of ontologies, the temporary nature of human knowledge, and consequently the need to update ontologies, as well as the lack of a methodological guide on how to conduct the domain ontologies enrichment process.

Assumptions: the following assumptions guided the development of this investigation: (1) the ontology enrichment can contribute so that ontologies represent the domain with more completeness; (2) enrichment methodologies facilitate the updating of knowledge in domain ontologies; (3) Library and Information Science has theoretical-methodological frameworks that can help to enrich domain ontologies.

Referential: the theoretical and methodological foundation addressed the maintenance and updating of Knowledge Organization Systems, Concept Theory, Ontologies, Ontological Engineering and Ontology Enrichment.

Methodology: The research is characterized as applied, in terms of nature, exploratory, in terms of objectives, and bibliographic in terms of methodological procedures. The steps proposed in the methodology of this investigation are five: (1) bibliographic research; (2) systematic literature review; (3) selection of documents on maintenance and updating of knowledge organization systems; (4) application of Content Analysis to selected empirical material; (5) development and validation of the enrichment methodology.

Results: as a result, the METHODOE (Methodology for Domain Ontology Enrichment) is presented. The METHODOE has five phases that allow you to carry out all activities related to the enrichment process, from planning to post-enrichment. Among the features and characteristics of this methodology, the following stand out: (1) presentation of a systematic view of the enrichment process; (2) domain and specific tools independence, allowing the application in ontologies of different domains; (3) independence from specific and difficult-to-access sources of information; (4) detailed presentation of most steps and some examples; (5) possibility of carrying out all types of enrichment (lexical, conceptual, relations and axioms).

Keywords: Ontology enrichment. Ontology update. Domain ontology. Knowledge Organization System.

LISTA DE FIGURAS

Figura 1 – Esquema com as etapas da revisão de literatura	32
Figura 2 – Esquema do processo de seleção dos estudos	39
Figura 3 – Origem dos autores dos documentos selecionados	42
Figura 4 – Modelo para construção de conceito	71
Figura 5 – Triângulo do Conceito	72
Figura 6 – Relação entre processos e atividades da engenharia ontológica	85
Figura 7 – Mapa conceitual sobre o funcionamento do processo de enriquecimento	90
Figura 8 – Camadas para o processo de Aprendizagem de Ontologia	94
Figura 9 – Relação entre o tipo de enriquecimento e componentes de ontologia	96
Figura 10 – Procedimentos metodológicos	101
Figura 11 – Esquema com material empírico para aplicação da Análise de Conteúdo	106
Figura 12 – Procedimentos da categoria Planejamento do Enriquecimento	127
Figura 13 – Procedimentos da categoria Aquisição do Conhecimento	128
Figura 14 – Procedimentos da categoria Tratamento do Conhecimento	128
Figura 15 – Procedimentos da categoria Implementação do Enriquecimento	129
Figura 16 – Procedimentos da categoria Pós-Enriquecimento	130
Figura 17 – Fases e etapas da METHODOE	136
Figura 18 – Fluxograma da Fase 1 – Planejamento do Enriquecimento	138
Figura 19 – Fluxograma da Fase 2 – Aquisição de Conhecimento	147
Figura 20 – Fluxograma da Fase 3 – Organização do Conhecimento	159
Figura 21 – Fluxograma da Fase 4 – Implementação do Enriquecimento	166
Figura 22 – Fluxograma da fase de Pós-Enriquecimento	170

Figura 23 – Estrutura hierárquica da OntoAgroHidro com as classes e subclasses principais	175
Figura 24 – Ausência de anotações para a classe Environmental Transformation e suas subclasses	177
Figura 25 – Presença dos conceitos fenômenos meteorológicos, evaporation e precipitation na OntoAgroHidro	189
Figura 26 – Mapa conceitual da classe Natural Environmental Transformation	193
Figura 27 – Mapa conceitual da classe Anthropic Environmental Transformation	194
Figura 28 – Classe Environmental Transformation enriquecida	196
Figura 29 – Hierarquia de classe Environmental Transformation enriquecida	197

LISTA DE TABELAS

Tabela 1 – Relação de procedimentos candidatos para a metodologia de enriquecimento a partir da literatura	111
Tabela 2 – Relação de procedimentos candidatos para a metodologia de enriquecimento a partir de documentos sobre manutenção e atualização de SOCs.....	122
Tabela 3 – Exemplo 1 de documento resultante desta etapa	157

LISTA DE GRÁFICOS

Gráfico 1 – Quantitativo de estudos recuperados por base de dados	36
Gráfico 2 – Resultado da primeira triagem (seleção após análise dos títulos).....	37
Gráfico 3 – Resultado da segunda triagem (seleção após análise dos resumos)	38
Gráfico 4 – Quantidade de publicações por ano.....	40
Gráfico 5 – Quantitativo de publicação por tipologia documental	41
Gráfico 6 – Frequência das fontes de informação utilizadas para enriquecimento.....	58
Gráfico 7 – Frequência dos tipos de enriquecimentos realizados.....	59
Gráfico 8 – Frequência dos tipos de técnicas utilizadas para extração da informação	60
Gráfico 9 – Detalhamento das etapas do método nos estudos	61
Gráfico 10 – Quantitativo de estudos que tratam dos procedimentos.....	118

LISTA DE QUADROS

Quadro 1 – Strings de busca por base de dados	34
Quadro 2 – Critérios de Inclusão e Exclusão adotados para seleção dos documentos.....	35
Quadro 3 – Categorias e suas respectivas subcategorias propostas por Dahlberg	72
Quadro 4 – Expressões de busca da pesquisa bibliográfica	103
Quadro 5 – Relação entre objetivos específicos e procedimentos metodológicos da pesquisa	108
Quadro 6 – Plano de ação para desenvolvimento da metodologia de enriquecimento de ontologias de domínios	131
Quadro 7 – Modelo de documento referente à avaliação da necessidade de enriquecimento.....	142
Quadro 8 – Pitfalls para diagnóstico de ontologia aplicáveis ao processo de enriquecimento.....	143
Quadro 9 – Exemplo de relatório com resultado do diagnóstico (documento 3).....	144
Quadro 10 – Exemplo do documento de escopo do enriquecimento.....	146
Quadro 11 - Exemplo de documento com o resultado da seleção das fontes de informação	149
Quadro 12 – Recorte dos dados apresentados no Apêndice D.....	152
Quadro 13 - Exemplo de documento resultante da etapa de escolha da técnica de extração	153
Quadro 14 – Exemplo 2 de documento resultante desta etapa	157
Quadro 15 – Modelo de quadro para organização do conhecimento adquirido	162
Quadro 16 – Identificação da ontologia OntoAgroHidro	174
Quadro 17 – Resultado da avaliação da necessidade de enriquecimento na classe Environmental Transformation.....	176

Quadro 18 – Relatório do diagnóstico da OntoAgroHidro.....	178
Quadro 19 – Documento de escopo do enriquecimento da OntoAgroHidro	179
Quadro 20 – Fontes de informação selecionadas	181
Quadro 21 – Técnicas de extração da informação	182
Quadro 22 – Resultado da análise dos SOCs	183
Quadro 23 – Amostra do resultado da análise dos documentos	184
Quadro 24 – Organização do conhecimento adquirido.....	185
Quadro 25 – Representação intermediária em formato hierárquico	191

LISTA DE ABREVIATURAS

AC	–	Análise de Conteúdo
ACM	–	Association for Computing Machinery
AGRO	–	Agronomy Ontology
ANSI	–	American National Standards Institute
AO	–	Aprendizagem de Ontologia
API	–	Application Programming Interface
BARTOC	–	Basel Register of Thesauri, Ontologies and Classification
BCGO	–	Breast Cancer Grading Ontology
BCI	–	Biblioteconomia e Ciência da Informação
BDTD	–	Biblioteca Digital de Teses e Dissertações
BFO	–	Basic Formal Ontology
BRAPCI	–	Base de Dados de Periódicos em Ciência da Informação
CHEBI	–	Chemical Entities of Biological Interest
CI	–	Ciência da Informação
CIDOC-CRM	–	Conceptual Reference Model
CLOE	–	Cross-Lingual Ontology Enrichment
CO	–	Cell Ontology
CRG	–	Classification Research Group
CTO	–	Cell Typo Ontology
DALOS	–	Drafting Legislation with Ontology Based Support
DOLCE	–	Descriptive Ontology for Linguistics and Cognitive Engineering
EMBRAPA	–	Empresa Brasileira de Pesquisa Agropecuária
ENVO	–	Environment Ontology
EO	–	Enriquecimento de Ontologia
FCA	–	Formal Concept Analysis

FMA	_	Foundational Model of Anatomy
FOIS	_	Formal Ontology in Information Systems
GA	_	Google Acadêmico
GACS	_	Global Agricultural Concept Scheme
GO	_	Gene Ontology
HPO	_	Human Phenotype Ontology
HTML	_	HyperText Markup Language
IA	_	Inteligência Artificial
IBICT	_	Instituto Brasileiro de Informação em Ciência e Tecnologia
ICBO	_	International Conference on Biomedical Ontology
IEE	_	Institute of Electrical and Electronic Engineers
ISO	_	International Organization for Standardization
ISTA	_	Information Science & Technology Abstracts
KOS	_	Knowledge Organization System
LISA	_	Library and Information Science Abstracts
LMF	_	Lexical Markup Framework
LN	_	linguagem natural
LOD	_	Linked Open Data
LSP	_	Hearst Lexico Sintatic Pattern
LT4eL	_	Language Technology for eLearning
METHODOE	_	Methodology for Domain Ontology Enrichment.
NCIT	_	National Cancer Institute Thesaurus
NER	_	Named Entity Recognition
NISO	_	National Information Standards Organization
NPMI	_	Normalized Pointwise Mutual Information
OAH	_	OntoAgroHidro
OBO	_	Open Biological Ontology

ODP	_	templates Ontology Design Patterns
OEM	_	Ontology Enrichment Method
OOPS	_	Ontology Pitfall Scanner
OPPL	_	Ontology Preprocessor Language
OWL	_	Ontology Web Language
PACTOLE	_	Property and Class Characterization From Text for Ontology Enrichment
PLN	_	Processamento de Linguagem Natural
POS	_	Part of Speech
PRO	_	Protein Ontology
QCs	_	Questões de Competência
RadLex	_	Radiology Lexicon
RDFS	_	Resource Description Framework Schema
RL	_	Revisão de Literatura
RO	_	Relation Ontology
RSL	_	Revisão Sistemática da Literatura
RSS	_	Really Simple Syndication
SDGIO	_	Sustainable Development Goals Interface Ontology
SOC	_	Sistema de Organização do Conhecimento
Tags	_	etiquetas
TF-IDF	_	Term Frequency
TF-IDF	_	Term Frequency – Inverse Document Frequency
TI	_	Tecnologia da Informação
UFO	_	Unified Foundational Ontology
WoS	_	Web of Science
WSD	_	Word Sense Disambiguation
WWW	_	World Wide Web

SUMÁRIO

1	INTRODUÇÃO	22
1.1	Problema e justificativa	25
1.2	Pressupostos	28
1.3	Objetivos	29
1.3.1	<i>Objetivo geral</i>	29
1.3.2	<i>Objetivos específicos</i>	29
1.4	Estrutura da tese	30
2	REVISÃO DE LITERATURA SOBRE ENRIQUECIMENTO DE ONTOLOGIAS DE DOMÍNIO	31
2.1	Procedimentos adotados na revisão	31
2.2	Análise quantitativa dos estudos	40
2.3	Análise qualitativa dos estudos	43
2.4	Análise da revisão de literatura	57
3	FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA	64
3.1	Manutenção e atualização de Sistemas de Organização do Conhecimento	64
3.2	Uso da Teoria do Conceito no contexto de Sistemas de Organização do Conhecimento	70
3.3	Ontologias	74
3.2.1	<i>Tipologia, componentes e linguagens da representação das ontologias</i>	75
3.4	Engenharia Ontológica	80
3.4.1	<i>Ciclo de vida das ontologias</i>	81
3.4.2	<i>Processos e atividades relacionados à manutenção e atualização de ontologias</i>	83
3.5	Enriquecimento de ontologias	87
3.5.1	<i>Relação do processo de enriquecimento de ontologia com a aprendizagem de ontologia</i>	91
3.5.2	<i>Tipos e fontes de Enriquecimento de Ontologia</i>	94

3.5.3	<i>Fases do Enriquecimento de Ontologia</i>	96
3.5.4	<i>Técnicas de extração da informação usadas para Enriquecimento de Ontologia</i>	97
4	METODOLOGIA	99
4.1	Caracterização da pesquisa	99
4.2	Delimitação do universo da pesquisa	99
4.3	Método	100
4.4	Procedimentos metodológicos	100
4.5	Relação entre os objetivos específicos e os procedimentos metodológicos	108
5	DESCRIÇÃO DOS PROCEDIMENTOS METODOLÓGICOS E RESULTADOS	109
5.1	Análise de conteúdo dos documentos selecionados na revisão de literatura (Grupo 1)	109
5.2	Análise de conteúdo dos documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento (Grupo 2)	120
5.3	Categorização dos procedimentos extraídos do material empírico	126
5.4	Plano de ação para subsidiar o desenvolvimento da metodologia de enriquecimento	130
6	METODOLOGIA PARA ENRIQUECIMENTO DE ONTOLOGIAS DE DOMÍNIO (METHODOE)	135
6.1	Planejamento do enriquecimento (Fase 1)	137
6.2	Aquisição de Conhecimento (Fase 2)	146
6.3	Organização do conhecimento (Fase 3)	158
6.4	Implementação do enriquecimento (Fase 4)	165
6.5	Pós-Enriquecimento (Fase 5)	169
7	APLICAÇÃO DA METHODOE E DISCUSSÃO DOS RESULTADOS	173
7.1	Planejamento do enriquecimento	173
7.2	Aquisição de conhecimento	180

7.3	Organização do conhecimento	185
7.4	Implementação do enriquecimento.....	191
7.5	Pós-enriquecimento.....	197
7.6	Considerações gerais sobre a METHODOE e sobre a aplicação na OntoAgroHidro	198
8	CONSIDERAÇÕES FINAIS	203
8.1	Limitações da pesquisa.....	207
8.2	Trabalhos futuros.....	208
	REFERÊNCIAS	209
	APÊNDICE A – PROTOCOLO DE REVISÃO SISTEMÁTICA DE LITERATURA	221
	APÊNDICE B – DADOS SOBRE DOCUMENTOS DA REVISÃO DE LITERATURA	225
	APÊNDICE C – COLETA DE DADOS REFERENTES À APLICAÇÃO DA ANÁLISE DE CONTEÚDO NOS 29 DOCUMENTOS DA REVISÃO DE LITERATURA	231
	APÊNDICE D – COLETA DE DADOS REFERENTES À APLICAÇÃO DA ANÁLISE DE CONTEÚDO NOS DOCUMENTOS SOBRE MANUTENÇÃO E ATUALIZAÇÃO DE SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO	247
	APÊNDICE E – RESULTADO DA ANÁLISE DOS DOCUMENTOS.....	256
	APÊNDICE F – GLOSSÁRIO COM OS CONCEITOS PARA ENRIQUECIMENTO DA ONTOAGROHIDRO	261
	ANEXO A – PITFALL 8 DE POVEDA-VILLALÓN (2016).....	263
	ANEXO B – PITFALL 9 DE POVEDA-VILLALÓN (2016).....	264
	ANEXO C – PITFALL 10 DE POVEDA-VILLALÓN (2016).....	265
	ANEXO D – PITFALL 11 DE POVEDA-VILLALÓN (2016).....	266
	ANEXO E – PITFALL 12 DE POVEDA-VILALLÓN (2016).....	267

ANEXO F – PITFALL 13 DE POVEDA-VILLALÓN (2016)	268
ANEXO G – PITFALL 14 DE POVEDA-VILALLÓN (2016)	269
ANEXO H – PITFALL 30 DE POVEDA-VILLALÓN (2016).....	270

1 INTRODUÇÃO

A sociedade atual está inserida em um contexto de produção vertiginosa de dados e informações, que traz uma gama de desafios para áreas que lidam com a organização da informação e do conhecimento. Entre os principais desafios, estão: a heterogeneidade e o volume colossal de dados, a preservação desses dados, a ambiguidade da linguagem natural para representação e recuperação da informação e do conhecimento, a interoperabilidade entre sistemas e a busca pela concretização do sonho da Web Semântica, em que as máquinas tenham a capacidade de interpretar informações e apoiar as tarefas dos humanos. Assim, estudos para superar esses desafios são essenciais, principalmente em uma área que lida diretamente com a informação e os processos associados a ela, assim como seu efetivo uso, como é o caso da Ciência da Informação (CI).

A CI possui um arcabouço teórico-metodológico que pode contribuir para a solução de vários problemas referentes aos desafios ora mencionados; porém essa área sozinha não dispõe de todas as soluções, dada a complexidade de vários desses problemas, que torna necessária a colaboração com diferentes áreas para solucioná-los, sendo preciso a aplicação de teorias, modelos, técnicas e estudos provenientes de outros campos do conhecimento. Entre os campos com os quais a CI interage, estão: a Ciência da Computação, fazendo uso de modelos de dados, linguagens formais de representação, modelos de recuperação da informação; a Linguística, utilizando sobretudo as técnicas da área de Processamento de Linguagem Natural para extração da informação de forma automática ou semiautomática; a Inteligência Artificial, com os estudos sobre aprendizagem de máquina, representação formal do conhecimento e inferência computacional; a Terminologia, com estudos sobre os termos para melhorar a comunicação em áreas específicas e desenvolvimento de vocabulários controlados; a Biblioteconomia, usando teorias da classificação, processos e práticas de tratamento da informação.

Entre as temáticas que confirmam a importância desse aspecto de interdisciplinaridade da CI estão as ontologias. Enquanto artefato de representação do conhecimento, as ontologias têm sido estudadas no contexto da Inteligência Artificial desde a década de 1970 e, a partir de então, despertaram o interesse de outras comunidades científicas. A CI é uma dessas comunidades, e o interesse dos

pesquisadores pela temática teve origem com o intuito de solucionar problemas no contexto da organização e recuperação da informação.

Na CI, as ontologias são classificadas como um tipo de Sistema de Organização do Conhecimento, ou seja, são instrumentos utilizados para representar o conhecimento. Desde os estudos iniciais de Vickery (1997) e Soergel (1999), que foram os precursores dos estudos na área, o interesse de cientistas da informação pelas temáticas associadas a esses artefatos cresceu, frente à sua capacidade de representar o conhecimento de forma menos ambígua e mais próxima da realidade, se comparada com outros SOC's com os quais a CI lida. As ontologias possibilitam novos cenários para a Organização da Informação e do Conhecimento.

As ontologias também são essenciais para o contexto da Web Semântica devido às suas características de representação do conhecimento de modo formal, que propiciam à representação a atribuição de significado livre de ambiguidade e assim possibilitam a interpretação por computadores. Junto a isso, as ontologias permitem a integração semântica entre fontes heterogêneas. Ora, a Web Semântica visa justamente que os dados façam sentido para as máquinas. Em suma, as ontologias auxiliam para que o sonho da Web Semântica se torne cada vez mais real. No contexto desta pesquisa, o foco está em um tipo particular de ontologias, as denominadas Ontologias de Domínio. Essas ontologias lidam com a representação do conhecimento em um domínio particular.

O uso de Ontologias de Domínio pode ser uma solução para organização e comunicação do conhecimento em diferentes contextos e áreas, uma vez que ontologias têm a capacidade de contribuir no compartilhamento do conhecimento, por meio da criação de um consenso terminológico e conceitual, estabelecendo uma *língua franca* em um domínio. E, em um cenário em que os dados estão sendo produzidos em uma escala inimaginável, é cada vez mais necessária a apropriação de ferramentas e tecnologias para auxiliar na interoperabilidade e na estruturação desses dados. Assim diversos campos do conhecimento têm utilizado as ontologias nessa perspectiva.

A construção de ontologias de domínio não é tarefa simples, e passa por várias etapas, conhecidas também como o ciclo de vida das ontologias. A área responsável por todos os processos associados ao desenvolvimento e manutenção de ontologias é a Engenharia Ontológica. Todos os processos da Engenharia Ontológica podem ser segmentados em três grupos: (1) pré-desenvolvimento, (2) desenvolvimento e (3) pós-

desenvolvimento. No pré-desenvolvimento, são previstos os processos e as atividades referentes ao planejamento e à preparação para o desenvolvimento das ontologias. Nessa fase, são definidos os objetivos da ontologia, seus usuários, questões que ela irá responder, entre outros. No desenvolvimento, estão os processos e as atividades que lidam com a implementação da ontologia, assim são criados os componentes ontológicos, como classes, relacionamentos, instâncias, axiomas. E, por fim, no pós-desenvolvimento, estão os processos e as atividades responsáveis pela manutenção, atualização e avaliação da ontologia, ou seja, depois de desenvolvida, para saber se ela está de acordo com o que foi planejado no pré-desenvolvimento, a ontologia precisa ser verificada.

Considerando os três grupos de processos da Engenharia Ontológica, o foco, nesta pesquisa, é no terceiro grupo, isto é, no pós-desenvolvimento. Essa fase é composta de vários processos, a saber: reengenharia, reestruturação, evolução, versionamento, customização, população, modificação, enriquecimento. Dentre esses processos, a investigação desta pesquisa será empenhada no processo de Enriquecimento de Ontologia, pois acredita-se que ele pode contribuir para a atualização de ontologias existentes.

O processo de Enriquecimento de Ontologia trata da expansão e especialização da estrutura de uma ontologia existente, por meio da inclusão de novos conceitos e seus símbolos linguísticos alternativos (sinônimos, por exemplo), relações (taxonômicas e não taxonômicas) e suas propriedades restritivas e axiomas. Esses novos elementos devem ser incluídos nos locais corretos da ontologia, ou seja, devem estar de acordo com o modelo de mundo que a ontologia objetiva representar. O processo de enriquecimento visa cooperar justamente para que as ontologias representem o domínio de modo mais completo e fidedigno. Portanto, o enriquecimento é uma das formas de se realizar a atualização da estrutura da ontologia.

Dado o contexto geral desta pesquisa, no qual se percebe a importância das ontologias para o âmbito da Organização da Informação e do Conhecimento e do estudo de todo o ciclo de vida desses instrumentos, a próxima subseção apresenta a problemática na qual esta pesquisa debruçar-se-á, bem como a justificativa para realização deste estudo.

1.1 Problema e justificativa

As ontologias de domínio são desenvolvidas em diversos campos do conhecimento, como medicina, biologia, agricultura, administração, em virtude das possibilidades de aplicações e do potencial que elas apresentam, constado por vários autores, na representação e compartilhamento do conhecimento (GRUBER, 1995; NOY, MCGUINNESS, 2001) e na interoperabilidade semântica entre diferentes sistemas ou bases de dados (SMITH, 2003; FARINELLI, 2017). A comunidade biomédica é a que mais se destaca nos estudos sobre ontologias, pois possui uma comunidade de desenvolvimento sólida, financiamento de grandes projetos e um importante repositório de ontologias, o Open Biological and Biomedical Ontology (OBO) Foundry¹.

Contudo percebe-se, em alguns campos, a existência de ontologias de domínio desenvolvidas de modo incompleto, ou seja, que não possuem em sua estrutura conhecimento para representar de forma completa o domínio que desejam representar. Entre algumas incompletudes, podem-se citar: (1) graus de granularidade superficiais, (2) relações não formalizadas, (3) falta de definições para alguns conceitos, (4) falta de especificação de axiomas.

Essas incompletudes impactam diretamente nas possibilidades de reuso dentro do domínio do conhecimento e na capacidade de representar um domínio, além do uso potencial na inferência de conhecimento. Uma das razões para as incompletudes das ontologias é o fato de o desenvolvimento de ontologias ser uma tarefa desgastante, consumidora de tempo (DRUMOND; GIRARDI, 2008), difícil e de alto custo, haja vista que é um processo quase todo manual, que exige mão de obra intensiva. Ainda que existam várias propostas (CIMIANO, HOTH, STAAB, 2005; CIMIANO, VAOLKER, 2005; HAZMAN, EL-BELTAGY, RAFAA, 2009) de automatização de algumas etapas do desenvolvimento de ontologias, principalmente na fase de aquisição de conhecimento, a tarefa de construir ontologias é árdua e dependente de mão de obra especializada.

As ontologias de domínio precisam apresentar uma estrutura mais completa possível, com todos os conceitos associados ao domínio de desenvolvimento, definições completas para as classes e relações, estabelecimento das propriedades

¹ <http://www.obofoundry.org/>

concernentes, e atribuição de axiomas, que ajudam a restringir as interpretações de determinado conceito, pois isso está diretamente associado com as capacidades das ontologias.

As ontologias de domínio são construídas com objetivos determinados por meio da captura de conhecimento de especialistas de domínio, assim como na extração de conteúdo de fontes de informação que explicitam o conhecimento do domínio. Porém o conhecimento é algo que muda com as novas descobertas, com mudanças de paradigmas ou até mesmo com refutações de teorias consolidadas.

Diante da provisoriedade do conhecimento humano, e como as ontologias são construídas com base nas abstrações que os especialistas fazem da realidade, os desenvolvedores de ontologias precisam garantir que esses instrumentos passem por processo de atualização sempre que houver necessidade, isto é, sempre que as ontologias não estiverem representando o conhecimento para o qual se propõem representar. Em outras palavras, atividades de atualização precisam fazer parte do ciclo de vida das ontologias, visto que o conhecimento sobre a realidade e a realidade em si não são estanques, são dinâmicos. Conforme pontuado por Tennis (2012) em um artigo sobre análise de domínio, a nossa visão de um domínio muda e os domínios mudam.

Sendo assim, vislumbra-se na temática Enriquecimento de Ontologia uma solução para as ontologias já desenvolvidas, mas que ainda se encontram incompletas e desatualizadas. O Enriquecimento de Ontologias (EO) trata-se de um processo associado a uma área mais ampla “atualização de ontologia”, e tem o propósito de expandir uma ontologia existente de acordo com seus componentes (classes, relações, axiomas).

O Enriquecimento de Ontologias pode aumentar a capacidade das ontologias em distintos aspectos, por exemplo: a ontologia pode representar um domínio de forma mais específica com a inclusão de conceitos característicos do domínio, pode melhorar a inferência automática de conhecimento com a adição de axiomas, pode ser utilizada no contexto da anotação semântica com a inclusão de sinônimos (realizações alternativas, como siglas, por exemplo) para um conceito. Ademais enriquecimento é um processo menos desgastante (tempo e recurso pessoal, econômico) do que construir uma ontologia do ponto inicial. Portanto, ontologias enriquecidas têm o potencial ampliado para resolução de problemas em várias áreas das atividades humanas.

Destaca-se que algumas das principais metodologias para desenvolvimento de ontologias de domínio (Método 101 de NOY e MCGUINNES, 2001; Sensus de SWARTOUT et al, 1996) não preveem o processo de atualização de ontologias. Essas limitações dessas metodologias impactam diretamente na reutilização desses instrumentos. O processo de atualização de ontologias está previsto nas metodologias (Methontology de FERNÁNDEZ et al., 1997; On-to-Knowledge de SURE, STAAB E STUDER, 2004; NeON de SUÁREZ-FIGUEROA et al., 2008), porém, na maioria delas, não está bem descrito como essa atualização deve ocorrer. Isso se deve principalmente ao grau de maturidade da área de Engenharia de Ontologia, que ainda é baixo, se comparado com a Engenharia de Software (SUÁREZ-FIGUEROA, 2010).

Os atuais métodos para enriquecimento de ontologias, apesar de produzirem resultados relevantes em estudos de caso, não podem, em sua maioria, ser generalizados para aplicação em diferentes domínios do conhecimento. Dessa forma, uma das lacunas é a falta de visão holística sobre o processo de enriquecimento. Outra lacuna é a dependência de ferramenta ou tipo de fonte de informação específica. Ademais falta o desenvolvimento de um modelo (ou metodologia) genérico, que integre os procedimentos relevantes de vários estudos em um único local e facilite assim a aplicação do processo de enriquecimento por ontologistas.

Metodologias contribuem no sentido de fazer com que processos e atividades se tornem sistematizados e menos uma atividade intuitiva, como acontecia no início dos estudos sobre construção de ontologias. Cada ontologista ou grupo de ontologistas construía seus artefatos com base em princípios próprios (*ad hoc*), estabeleciam etapas de desenvolvimento a partir da experiência (empiricamente), sem nenhum respaldo metodológico; assim, tornava-se uma tarefa artesanal. Portanto, para que o processo de manutenção e atualização de ontologias também não passe pelo mesmo problema, é preciso trabalhar em propostas metodológicas que guiem todo o processo de forma clara, objetiva e eficiente.

Dessa forma, a presente pesquisa se justifica por buscar contribuir com os estudos sobre ontologias na perspectiva da Biblioteconomia e Ciência da Informação (BCI), propondo uma metodologia de Enriquecimento de Ontologias de domínio que contemple todos os componentes e que possa ser utilizada em diferentes domínios. Uma metodologia tem a capacidade de integrar diferentes métodos e criar um padrão que possa direcionar o processo de EO. Vale ressaltar que se propõe o desenvolvimento de uma metodologia original para Enriquecimento de Ontologias,

afinal a literatura aponta a existência de propostas associadas ao enriquecimento de, no máximo, três componentes da ontologia, carecendo, assim, de uma proposta que abranja todos os componentes e que defina detalhadamente as várias atividades e tarefas de cada etapa.

Diante da problemática e justificativas apresentadas, as questões que norteiam o presente estudo são: 1) como desenvolver uma metodologia para enriquecimento de ontologias de domínio que supere as lacunas dos atuais métodos? 2) O desenvolvimento de uma metodologia pode simplificar o processo de enriquecimento de ontologias de domínio? Frente a esses questionamentos, e visando oferecer uma metodologia para enriquecimento de ontologias de domínio, apresentam-se a seguir os pressupostos desta pesquisa.

1.2 Pressupostos

Esta pesquisa parte dos seguintes pressupostos: (1) o enriquecimento de ontologias pode contribuir para que as ontologias representem o domínio com mais completude; (2) as metodologias de enriquecimento facilitam a atualização de conhecimento em ontologias de domínio; (3) a BCI possui arcabouço teórico-metodológico que pode auxiliar no enriquecimento de ontologias de domínio.

Com relação ao pressuposto (1) **o enriquecimento de ontologias pode contribuir para que as ontologias representem o domínio com mais completude**, cabe dizer que o processo de enriquecimento visa expandir uma ontologia já desenvolvida com novos componentes (conceitos, relações, propriedades e axiomas); assim, o potencial de representação do domínio é ampliado. Isso vai ao encontro do que disse Soergel (2009) a respeito de Sistemas de Organização do Conhecimento: “adicionando mais um pouco de informação a um SOC, vai expandir tremendamente a sua utilidade”. Ademais a explicitação das propriedades das relações e a definição de axiomas permitem que o modelo de mundo seja mais bem representado, contribuindo para a restrição de interpretações. A diminuição de interpretações sobre o domínio está diretamente associada à sua aproximação com o modelo que se almeja representar.

Já no que concerne ao pressuposto (2) **as metodologias de enriquecimento facilitam a atualização de conhecimento em ontologias de domínio**, argumenta-se que as metodologias objetivam descrever quais os procedimentos para a

realização de um processo, portanto descrevem quando e como uma atividade deve ocorrer. Assim elas coadjuvam para que o processo de enriquecimento seja menos subjetivo e mais coordenado, principalmente quando a metodologia provê os distintos cenários possíveis para o desenvolvimento do processo. O processo de enriquecimento se baseia no conhecimento atualizado de um domínio, seja conhecimento de especialistas de domínio, seja de fontes como bases de dados, corpus textuais ou outros Sistemas de Organização do Conhecimento. Dessa forma, o conhecimento atualizado extraído dessas fontes é utilizado para enriquecer as ontologias.

Por fim, no que tange ao pressuposto (3) **a BCI possui arcabouço teórico-metodológico que pode auxiliar no enriquecimento de ontologias de domínio**, deduz-se que teorias e técnicas empregadas no contexto da BCI têm potencial para assistir o processo de enriquecimento, dada a vasta experiência no desenvolvimento de outros SOCs, como sistemas de classificação, tesouros e taxonomias, conhecidos como SOCs tradicionais e que já existiam bem antes da Informática e, conseqüentemente, da formalização desses modelos, para que as máquinas interpretem o conhecimento representado. Portanto acredita-se que alguns dos aportes utilizados para o desenvolvimento desses SOCs tradicionais podem contribuir no contexto do enriquecimento de ontologias de domínio.

1.3 Objetivos

Esta tese tem o propósito de atender ao objetivo geral e aos objetivos específicos apresentados a seguir.

1.3.1 *Objetivo geral*

Propor uma metodologia para o processo de enriquecimento de ontologias de domínio com base em construtos da engenharia ontológica e da BCI e que supere as lacunas dos atuais métodos para enriquecimento de ontologias.

1.3.2 *Objetivos específicos*

- 1) Caracterizar o processo de enriquecimento e desenvolver um fluxo para ele.

- 2) Identificar insumos na área de engenharia ontológica e na Biblioteconomia e Ciência da Informação para o desenvolvimento da metodologia de enriquecimento.
- 3) Demonstrar as principais lacunas nos métodos existentes para enriquecimento de ontologias de domínio.
- 4) Dar subsídios aos desenvolvedores de ontologias para melhoria do processo de enriquecimento.
- 5) Contribuir com os estudos sobre atualização de ontologias de domínio.

1.4 Estrutura da tese

O restante desta tese está estruturado da seguinte forma: no [capítulo 2](#), descreve-se o estado da arte concernente ao Enriquecimento de Ontologias de domínio; no [capítulo 3](#), abordam-se os fundamentos teóricos e metodológicos nos quais este estudo se embasa, a saber: Sistemas de Organização do Conhecimento, Ontologias, Engenharia Ontológica e Enriquecimento de Ontologia; já o [capítulo 4](#) traz a metodologia da pesquisa, caracteriza-a e, nele, relatam-se os procedimentos metodológicos adotados para a execução do estudo; no [capítulo 5](#), faz-se o detalhamento de alguns dos procedimentos metodológicos e apresentam-se os resultados; o [capítulo 6](#) apresenta o resultado principal desta pesquisa, que se refere à metodologia de enriquecimento desenvolvida, a METHODOE; no [capítulo 7](#), valida-se a METHODOE, por meio de exercício de enriquecimento de uma ontologia do domínio agrícola, denominada OntoAgroHidro; e, por fim, o [capítulo 8](#) contém as considerações finais, as limitações desta pesquisa, bem como as possibilidades de trabalhos futuros que podem derivar desta investigação.

2 REVISÃO DE LITERATURA SOBRE ENRIQUECIMENTO DE ONTOLOGIAS DE DOMÍNIO

Neste capítulo, apresentam-se os resultados da revisão de literatura. Realizou-se uma revisão de literatura do tipo sistemática, pois é um tipo de revisão que descreve claramente todos os procedimentos realizados para seleção dos estudos, possibilitando, assim, que o resultado seja auditado.

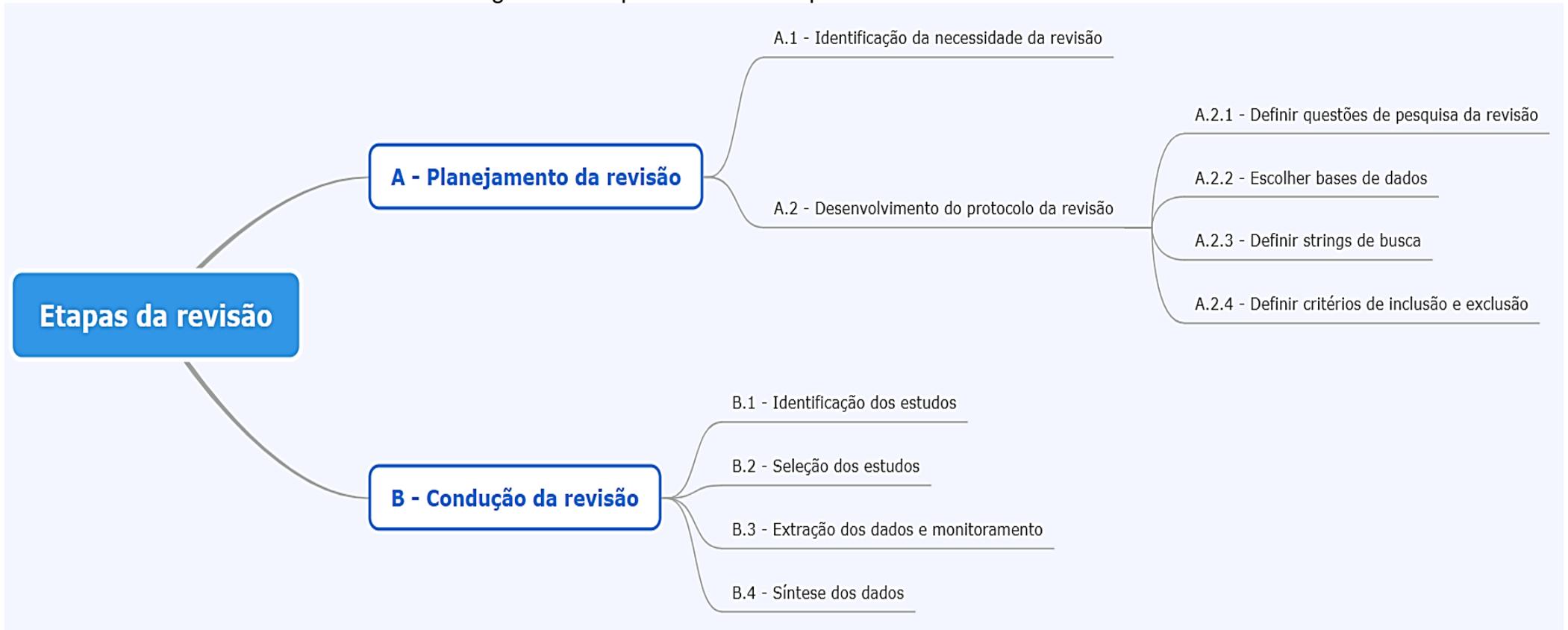
2.1 Procedimentos adotados na revisão

Para a execução da revisão, adaptaram-se as diretrizes para revisão sistemática de Kitchenham (2004) e utilizou-se o *software* Start² para gestão e organização dos estudos. Sendo assim, as etapas da revisão foram divididas em duas grandes fases: A) Planejamento da Revisão e B) Condução da Revisão. A fase de Planejamento da Revisão é composta de duas etapas: 1) Identificação da necessidade da revisão e 2) Desenvolvimento do protocolo da revisão. E a fase de Condução da Revisão é composta das etapas³: 1) Identificação de estudos; 2) Seleção dos estudos; 3) Extração dos dados e monitoramento; e 4) Síntese dos dados.

² O software Start (State of the Art through Systematic Review) foi desenvolvido por pesquisadores do Laboratório de Pesquisa em Engenharia de Software da Universidade Federal de São Carlos (UFSCAR). Esse software permite a gestão de todas as etapas da revisão sistemática da literatura, desde o planejamento, seleção de estudos, até a etapa de redação do relatório da revisão.

³ Na proposta original de Kitchenham (2004), são previstas cinco etapas na fase de Condução da Revisão. Nesta tese, não foi aplicada a etapa de Avaliação da qualidade dos estudos, pois se chegou a um número razoável de estudos, não havendo assim a necessidade de filtragem dos mesmos.

Figura 1 – Esquema com as etapas da revisão de literatura



Fonte: elaborada pelo autor.

Na Etapa A.1 (Identificação da necessidade de revisão), realizou-se uma análise exploratória da literatura sobre enriquecimento de ontologias e, diante do grande número de estudos, percebeu-se a necessidade da revisão para coletar de forma organizada os estudos publicados sobre a aplicação de métodos de enriquecimento de ontologias de domínio.

Na Etapa A.2 (Desenvolvimento do protocolo da revisão), desenvolveram-se as (i) questões de pesquisa da revisão: (ii) seleção das bases de dados, (iii) criação das *strings* de busca e (iv) criação dos critérios de inclusão e exclusão. O protocolo completo encontra-se no [Apêndice A](#). Vale ressaltar que, antes do preenchimento do protocolo, realizou-se uma pesquisa exploratória no Google Acadêmico (GA) com o intuito de identificar as bases de dados nas quais havia uma representatividade maior de estudos sobre a temática, bem como identificar as palavras-chave mais utilizadas para representar esses estudos.

As questões de pesquisa que a revisão de literatura objetiva responder são: (1) Quais são os estudos que aplicaram métodos para enriquecimento de ontologias de domínio publicados entre 2000 e 2019? (2) Quais as características dos métodos para enriquecimento de ontologias de domínio? (3) Quais as lacunas existentes nesses métodos? (4) Quais são os aspectos positivos dos métodos existentes? O período temporal foi definido com base na pesquisa exploratória no Google Acadêmico, que revelou que os primeiros estudos surgiram a partir do ano 2000.

As bases de dados selecionadas para levantamentos dos estudos foram as seguintes: ACM Digital Library, Google Acadêmico (GA), IEEE, Scopus e Web of Science (WoS). Também foram realizadas consultas pré-testes nas bases Library and Information Science Abstracts (LISA) e Information Science & Technology Abstracts (ISTA). No entanto a quantidade de registros recuperados foi ínfima, e eles já estavam contemplados nas bases de dados selecionadas anteriormente. Sendo assim, essas duas bases não foram incluídas na seleção.

Após realizar a seleção das bases de dados, passou-se à elaboração das *strings* de busca, que continham os termos mais relevantes para recuperação dos estudos. Observa-se, portanto, que cada base de dados demandou uma *string* que se adequasse aos parâmetros específicos daquela base. Contudo, a diretriz da *string* se manteve, ou seja, buscar por estudos que continham as palavras ontologia e enriquecimento, bem como suas variações no campo "título". No Quadro 1, apresentam-se as *strings* de busca separadas por base de dados.

Quadro 1 – Strings de busca por base de dados	
BASE DE DADOS	STRING
ACM	Publication title (ontolog*) AND publication title (enrich*)
GA	Allintitle: ontology enrichment
IEEE	Document title "ontolog*" AND document title "enrich**"
Scopus	title(ontolog* AND enrich*)
WoS	TI=(ontolog* AND enrich*)

Fonte: dados de pesquisa (2021).

Na última atividade do desenvolvimento do protocolo, definiram-se os critérios de exclusão (E) e inclusão (I). Essa definição dos critérios foi feita com base no que se almeja com a revisão, ou seja, no seu objetivo, que se refere à identificação de estudos que aplicaram o processo de enriquecimento em ontologias de domínio. A seguir, no Quadro 2, estão os quatro critérios de inclusão e os sete critérios de exclusão.

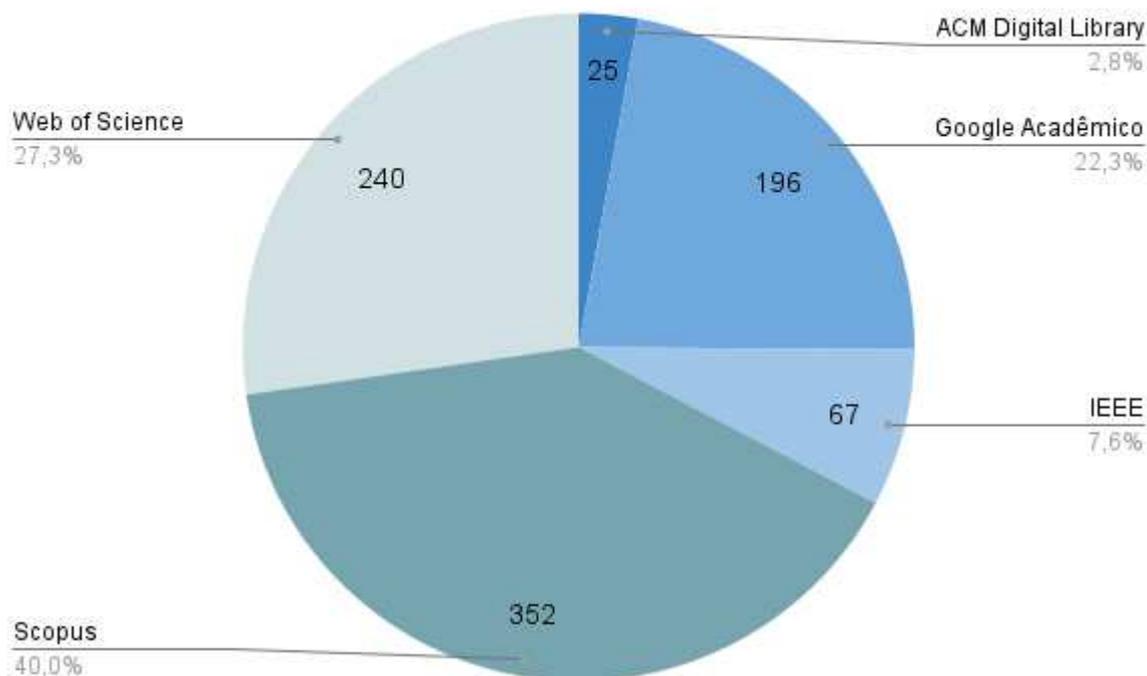
Quadro 2 – Critérios de Inclusão e Exclusão adotados para seleção dos documentos

Critérios de inclusão (I)	Critérios de exclusão (E)
<ol style="list-style-type: none"> 1. Título possui relação com a temática enriquecimento de ontologia de domínio; 2. Resumo possui relação com a temática enriquecimento de ontologia de domínio; 3. Serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas; 4. Serão incluídos trabalhos que tenham aplicado um método de enriquecimento em ontologia de domínio; 	<ol style="list-style-type: none"> 5. Título não possui relação com a temática enriquecimento de ontologias de domínio; 6. Resumo não possui relação com a temática enriquecimento de ontologia de domínio; 7. Serão excluídos trabalhos que não apresentem resumo/<i>abstract</i>; 8. Serão excluídos <i>short papers</i> e trabalhos disponibilizados em bases <i>pre-print</i>; 9. Serão excluídos documentos que não estejam em português ou inglês; 10. Serão excluídos trabalhos que não estejam disponíveis em texto completo; 11. Serão excluídos os trabalhos que não apliquem o processo de enriquecimento em ontologias de domínio.

Fonte: dados de pesquisa (2021).

Na fase B – Condução da Revisão, **Etapas B.1 – Identificação de estudos**, realizaram-se pesquisas nas cinco bases de dados selecionadas e chegou-se ao total de 880 estudos, a partir da aplicação do protocolo desenvolvido na fase anterior. O Gráfico 1 apresenta o resultado da identificação de estudos separados por base de dados.

Gráfico 1 – Quantitativo de estudos recuperados por base de dados

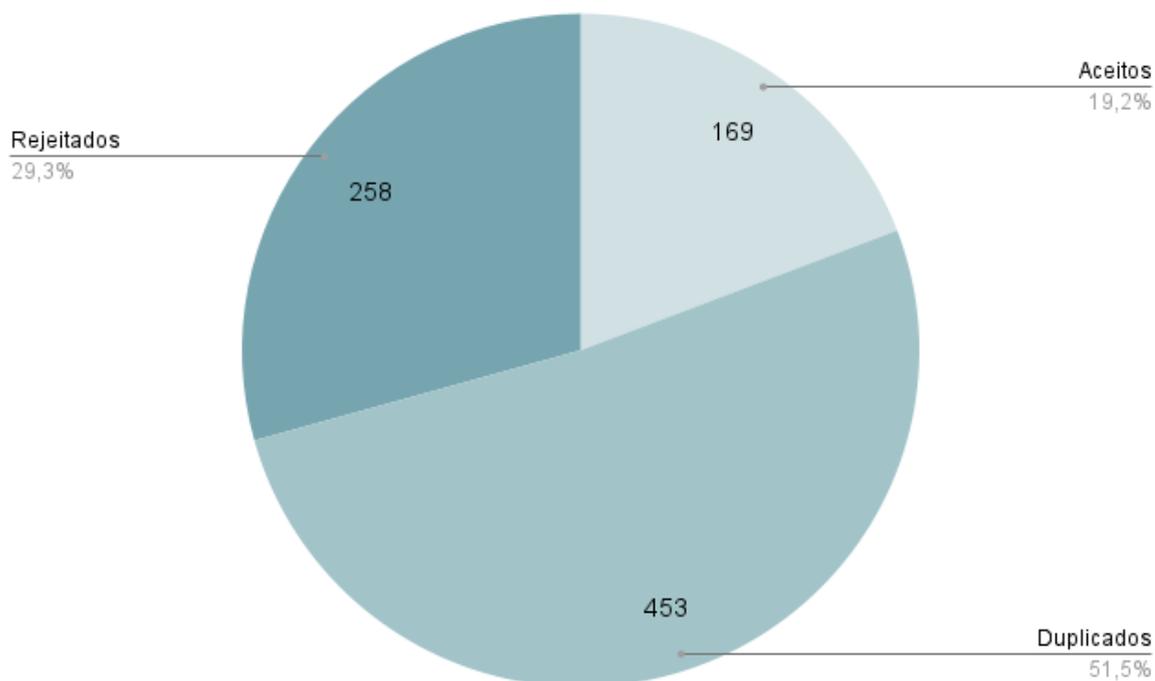


Fonte: dados de pesquisa (2021).

Na **Etapa B.2 – Seleção dos estudos**, com o uso do software Start, realizou-se a análise/filtragem dos 880 registros recuperados nas bases de dados ACM, Google Acadêmico, IEEE, Scopus e Web of Science. A primeira atividade executada foi a exclusão de 454 duplicatas. Em seguida, procedeu-se à análise do título dos 427 registros restantes. Assim, foram realizadas a leitura do título⁴ desses estudos e a classificação de acordo com os seguintes critérios de inclusão e exclusão: (Inclusão) título possui relação com a temática enriquecimento de ontologia de domínio; (Exclusão) título não possui relação com a temática enriquecimento de ontologias de domínio. O Gráfico 2 mostra os números após essa primeira triagem dos estudos.

⁴ Para aqueles estudos que continham títulos muito genéricos ou não eram suficientemente claros, realizou-se a leitura do resumo.

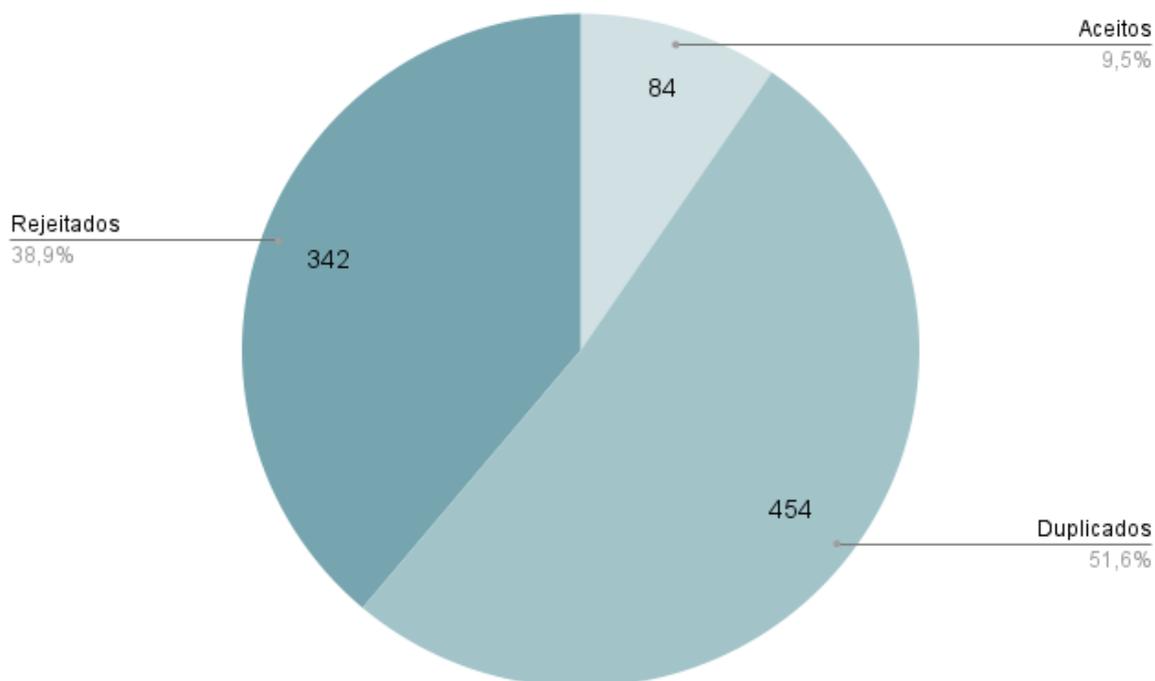
Gráfico 2 – Resultado da primeira triagem (seleção após análise dos títulos)



Fonte: dados de pesquisa (2021).

Realizada a análise do título, foram excluídos 260 estudos, restando 167 para a próxima filtragem, análise do resumo. Nessa atividade, realizou-se a leitura do resumo dos 167 estudos e classificação de acordo com os seguintes critérios de inclusão e exclusão: (I) Resumo possui relação com a temática enriquecimento de ontologia de domínio; (E) Serão excluídos trabalhos que não apresentem resumo/abstract; (E) Resumo não possui relação com a temática enriquecimento de ontologia de domínio. Após a análise do resumo, foram excluídos 83 estudos, restando, assim, 84 estudos para a próxima análise. O Gráfico 3 mostra os números após a segunda triagem dos estudos.

Gráfico 3 – Resultado da segunda triagem (seleção após análise dos resumos)

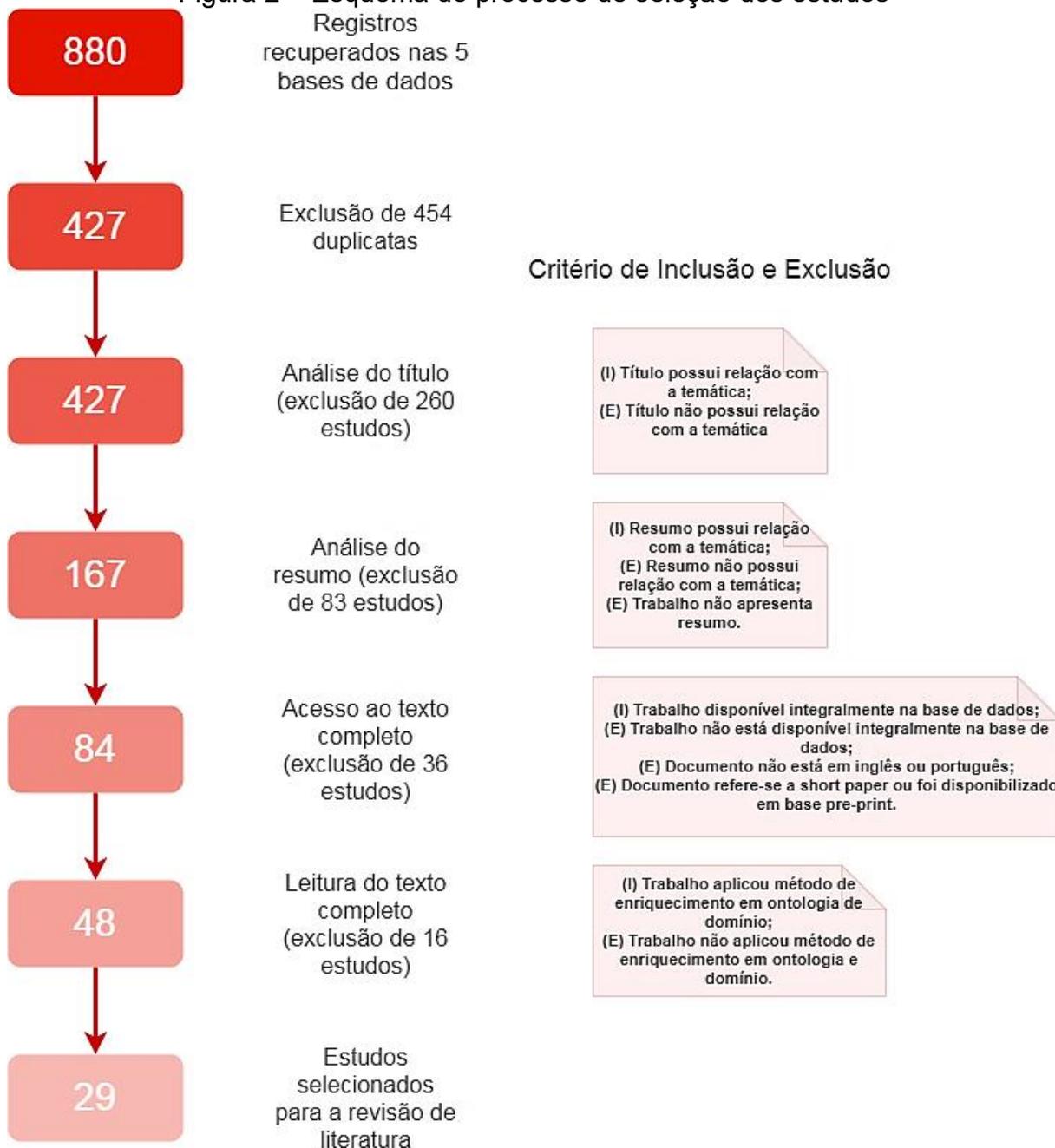


Fonte: dados de pesquisa (2021).

A próxima análise refere-se à leitura do texto completo dos 84 estudos restantes. Os trabalhos foram analisados com base nos seguintes critérios: (I) Serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas; (E) Serão excluídos trabalhos que não estejam disponíveis em texto completo; (E) Serão excluídos trabalhos que não estejam em português ou inglês; (E) Serão excluídos *short papers* e trabalhos disponibilizados em bases *pre-print*. Assim, após essa análise, 36 estudos foram excluídos, de acordo com os critérios de exclusão mencionados; restaram 48 estudos para a última triagem.

Os textos completos dos 48 estudos restantes foram analisados e classificados de acordo com os seguintes critérios: (I) Serão incluídos trabalhos que tenham aplicado um método de enriquecimento em ontologia de domínio; (E) Serão excluídos os trabalhos que não aplicaram o processo de enriquecimento em ontologias de domínio. Após essa análise, restaram 29 estudos, considerados o *corpus* para a revisão de literatura e dos quais foram extraídos os dados que caracterizam os estudos. A Figura 2 apresenta todo o processo de seleção dos estudos em forma de esquema.

Figura 2 – Esquema do processo de seleção dos estudos



Fonte: dados de pesquisa (2021).

Na **Etapa B.3 – Extração dos dados e monitoramento**, com o uso das funcionalidades do Software Start, desenvolveram-se formulários de extração de dados que apresentavam os seguintes campos: fonte de informação, tipo de enriquecimento, técnica aplicada, detalhamento das etapas, principais lacunas e aspectos favoráveis. Sendo assim esses campos serviram de guia para a extração dos dados dos estudos selecionados. Essas análises serão apresentadas na seção

2.3 Análise qualitativa dos estudos. Ademais extraíram-se os dados quantitativos para caracterização dos estudos, que serão apresentados na próxima seção.

A **Etapa B.4 – Síntese dos dados** refere-se à organização e escrita dos dados que foram extraídos na etapa anterior. Conforme já apresentado, os dados foram separados em dados quantitativos e qualitativos e serão apresentados nas próximas subseções.

2.2 Análise quantitativa dos estudos

Nesta seção, apresentam-se dados quantitativos dos 29 estudos selecionados conforme parâmetros apresentados na seção anterior. Discorre-se sobre os seguintes aspectos: número de publicações por ano; tipologia documental; país de origem dos autores.

No que se refere ao número de publicações por ano, percebe-se que existe uma certa equivalência e distribuição das publicações entre os anos. Os anos 2010, 2015 e 2019 possuem três publicações. Somente o ano de 2017 teve uma maior ocorrência, com quatro publicações.

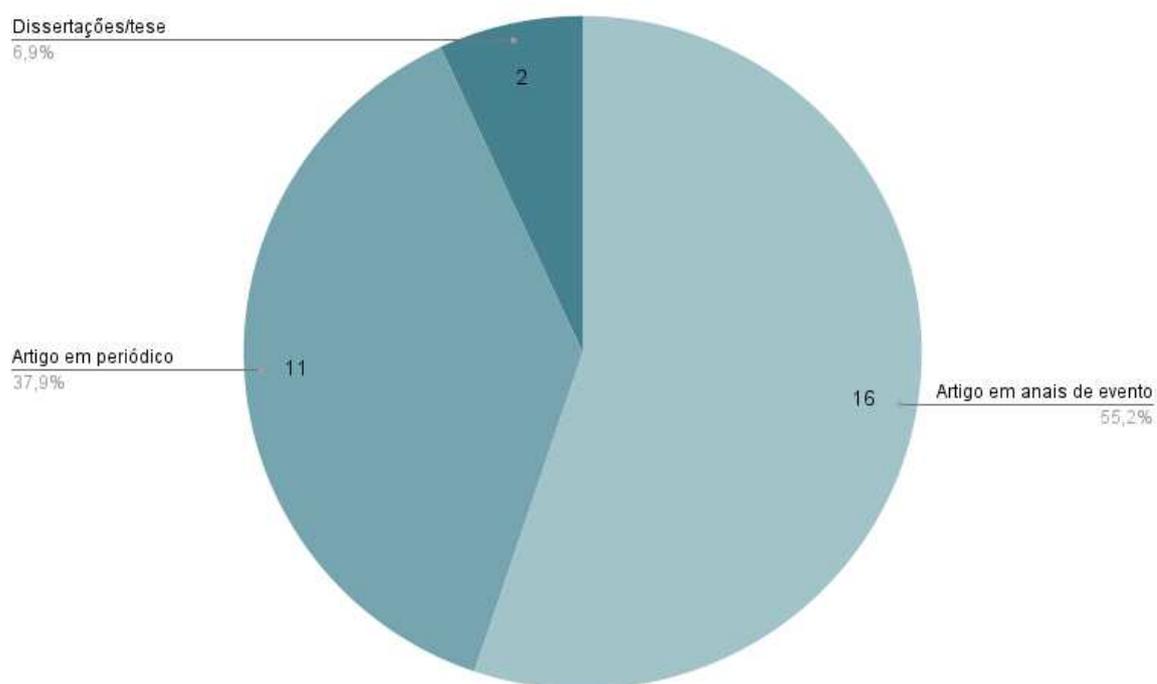


Fonte: dados de pesquisa (2021).

Os dados referentes à tipologia documental, apresentados no Gráfico 5, revelam que 37,9% das publicações são originárias de artigos de periódicos. No entanto, a maioria das publicações (52,2%) é proveniente de anais de eventos

científicos, enquanto 6,9% referem-se a teses e dissertações. A explicação para esses dados pode ter relação com o fato de a temática enriquecimento de ontologia ainda ser um assunto emergente, que está em desenvolvimento. Assim, é mais comum a publicação dessas temáticas em anais de eventos.

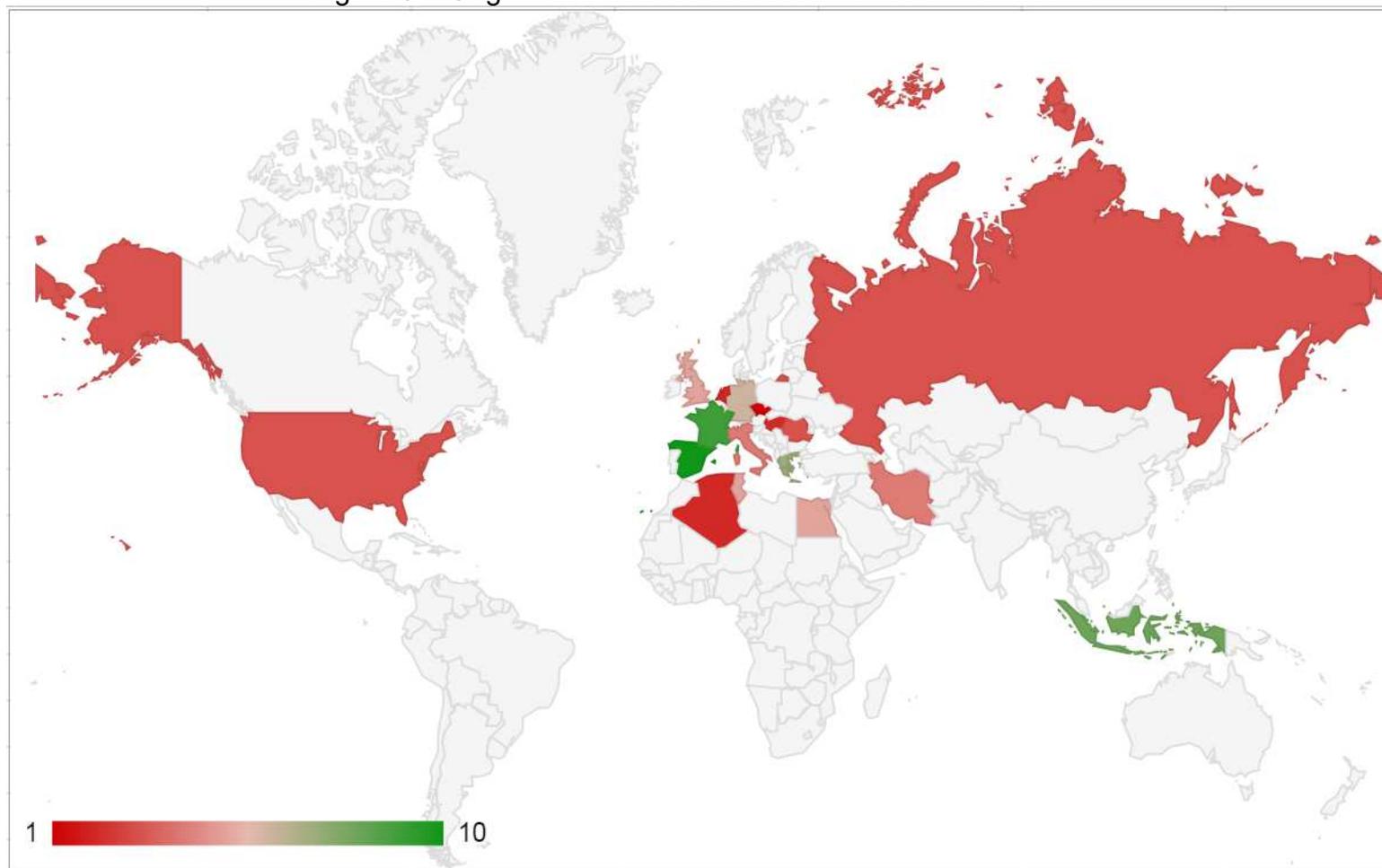
Gráfico 5 – Quantitativo de publicação por tipologia documental



Fonte: dados de pesquisa (2021).

Com relação ao país de origem dos autores, há uma distribuição em 17 países. No entanto não foi selecionado nenhum artigo de pesquisadores brasileiros, considerando os parâmetros adotados na presente pesquisa, o que reforça a importância do presente estudo. A Figura 3 mostra uma concentração de publicações nos países europeus, sendo que a Espanha é o país que mais possui autores (10) que publicaram sobre enriquecimento, seguido da França, com 9 autores. A Indonésia possui 8 autores; e a República Tcheca, apenas 1 autor.

Figura 3 – Origem dos autores dos documentos selecionados



Espanha (10); França (9); Indonésia (8); Grécia (7); Alemanha (6); Egito (5); Reino Unido (5); Tunísia (5); Irã (4); Itália (4); Estados Unidos (3); Romênia (3); Rússia (3); Argélia (2); Holanda (2); Hungria (2); República Tcheca (1)

Fonte: dados de pesquisa (2021).

Na próxima seção, abordam-se as características e os objetivos principais de cada método tratado nos 29 estudos selecionados.

2.3 Análise qualitativa dos estudos

Apesar de a literatura sobre métodos para aquisição de conhecimento de forma automática e semiautomática (MAEDCHE, 2002; SHAMSFARD, BARFOUROUSH, 2004; BUITELAAR, CIMIANO, 2005; CIMIANO, VAOLKER, 2005; CIMIANO, 2006) ser profícua em estudos (também conhecidos como métodos de Aprendizagem de Ontologia), nesta pesquisa não se considerou esse tipo de trabalho, haja vista que esses estudos focam em adquirir conhecimento para o desenvolvimento de uma ontologia do início ou apenas focam na tarefa de extrair conhecimento, mas não discorrem propriamente sobre o enriquecimento da ontologia. O propósito desta revisão de literatura foi a identificação de métodos⁵ e metodologias⁶ de enriquecimento de ontologias, que realmente foram aplicados, não apenas estudos que focam na técnica de aquisição de conhecimento.

Os primeiros estudos sobre Enriquecimento de Ontologias surgiram a partir dos anos 2000, quando se identificou a necessidade de atualizar, corrigir ou aperfeiçoar as ontologias já desenvolvidas. Assim foram desenvolvidas várias atividades vinculadas à manutenção de ontologias.

Mediante a análise da literatura, percebe-se um grande volume de propostas e métodos para enriquecimento de ontologias provenientes, principalmente, de estudiosos da área de Ciência da Computação. A seguir, dedica-se a descrever cada uma dessas propostas, apresentando-as com mais detalhes. Os 29 trabalhos selecionados foram apresentados em ordem cronológica entre os anos de 2001 e 2019. O propósito é descrever as características principais dos métodos existentes.

Os pesquisadores Velardi, Fabriani e Missikoff (2001) do Departamento de Ciência da Informação da Universidade de Roma e do Laboratório italiano para Conhecimento e Sistemas Empresariais apresentam o estudo *Using text processing techniques to automatically enrich a domain ontology*, sendo um dos primeiros estudos da temática.

A proposta dos autores é o enriquecimento conceitual e de relações a partir da aplicação de técnicas de análise linguística em documentos textuais. Entre as técnicas

⁵Método: é um conjunto de processos ou procedimentos ordenados usados na engenharia de um produto ou na realização de um serviço (MENDONÇA, 2015).

⁶Metodologia é “uma série integrada de técnicas ou métodos criando uma teoria geral de sistemas de como uma classe de pensamento pode ser executada” (MENDONÇA, 2015).

aplicadas no método OntoLearn (nome escolhido pelos autores), estão: *Named Entity Recognition* (NER); *Parsing*; *POS tagging*; *Chunking (sentence segmentation)*; análise gramatical. Aplica-se o OntoLearn em uma ontologia da área de turismo, e os resultados da avaliação demonstraram que o método apresenta resultados similares para documentos textuais com conhecimento genérico quanto para documentos com conhecimento especializado, porém o método é sensível ao tamanho do corpus disponível para extração da informação.

Faatz e Steinmetz (2002), no artigo *Ontology enrichment with texts from the WWW*, oferecem um método para enriquecimento semiautomático de ontologias com base no uso de medidas de similaridades entre conceitos (distância semântica). A partir dos conceitos da própria ontologia que se almeja enriquecer, realizaram-se consultas em um corpus textual por conceitos similares. Aplicou-se o método em uma ontologia médica, porém não há uma descrição dos passos para realização do enriquecimento. O foco dos autores foi na avaliação do método com base em dois experimentos. O primeiro experimento apoiou-se em um *corpus* geral com 28.700.000 sentenças, e os resultados se mostraram pobres, de acordo com os autores. O segundo experimento usou um *corpus* especial de 15.500 sentenças e apresentou resultados mais consistentes.

O artigo *A Name-Matching Algorithm for Supporting Ontology Enrichment*, publicado por Valarakos *et al.* (2004), apresenta um método semiautomático para enriquecimento lexical de ontologias de domínio com base em um algoritmo computacional. As etapas do método são: (1) usar a ontologia de domínio para anotar semanticamente um *corpus* específico do domínio; (2) usar o *corpus* anotado para treinar um modelo estatístico (Modelo Escondido de Markov); (3) extrair novos termos candidatos do *corpus*; (4) usar especialistas do domínio para validar os novos termos e adicioná-los manualmente à ontologia. Para fins de validação, aplicou-se o método em uma ontologia que representa produtos de computador portátil.

Navigli e Velardi (2006), no artigo *Ontology Enrichment Through Automatic Semantic Annotation of On-Line Glossaries*, apresentam um método automático para enriquecimento de ontologias com base em um glossário existente no mesmo domínio da ontologia. As etapas do método são: (1) análise da marcação gramatical de cada definição do glossário; (2) reconhecimento do nome de entidades nas definições; (3) anotação de segmentos de sentenças com propriedades da ontologia; (4)

formalização do glossário. Aplica-se o método no domínio de Patrimônio Cultural, mais especificamente na ontologia CIDOC-CRM.

No estudo *Enrichment of OBO ontologies*, Bada e Hunter (2007) propõem um método para enriquecimento lexical, conceitual e de relações das ontologias *Gene Ontology* (GO), *Cell Type Ontology* (CTO) e *Chemical Entities of Biological Interest* (CHEBI). Para isso, os autores se baseiam no conhecimento modelado nas próprias ontologias como fonte de informação. O método de enriquecimento se respalda principalmente na recuperação de padrões (ou seja, expressões regulares encontradas nas anotações da classe) nos conceitos da GO e nas outras duas ontologias. Esses padrões são ranqueados por frequência e validados manualmente por especialista do domínio. Como resultado, chega-se a uma ontologia enriquecida resultante da integração das três ontologias do domínio biomédico. Ademais os resultados da aplicação do método mostram que a grande maioria do conteúdo utilizado para enriquecer as ontologias foi considerada válida por especialista do domínio, posto que foram inseridas mais de 4500 relações entre conceitos nas ontologias.

O pesquisador Wim Peters (2009), do Departamento de Ciência da Computação da Universidade de Sheffield, no estudo *Text-based legal ontology enrichment*, apresenta uma proposta de enriquecimento semiautomática de conceitos e relações de uma ontologia da área de Direito, mais especificamente na subárea de proteção do consumidor, denominada *Drafting Legislation with Ontology Based Support* (DALOS). O método se respalda em várias técnicas para extrair informação de documentos textuais do domínio legal. Entre as técnicas, estão: Análise sintática (*tokenização, sentence splitting, POS tagging, lematização*); Análise estatística (TF-IDF, *Mutual Information*); Padrão léxico sintático (*Headword matching, Hearst*). Os resultados de testes empíricos revelaram que algumas dessas técnicas de extração produziram resultados mais interessantes do que outras, mostrando a necessidade de aprimoramento de algumas delas.

Em 2008, no artigo nomeado *PACTOLE: A methodology and a system for semi-automatically enriching an ontology from a collection of texts*, Bendaoud, Toussaint e Napoli propõem a metodologia PACTOLE (Property And Class Characterization from Text for Ontology Enrichment) para enriquecimento de ontologia do domínio astronômico a partir do conhecimento extraído de textos do domínio modelado pela ontologia, assim como de recursos de representação do domínio (como tesouros e

bases de dados). O processo de enriquecimento se baseia em técnicas de PLN (Parsing) e Formal Concept Analysis (FCA); e a metodologia possui cinco passos:

1. extração da informação/análise de textos: envolve técnicas de PLN para extrair objetos e suas propriedades do domínio. Assim consideram-se as expressões “verbo/sujeito”, “verbo/objeto”, “verbo/complemento” e “verbo/preposição” (*prepositional phrase dependencies*), visto que são boas sugestões sintáticas para atribuir uma propriedade a um objeto. O resultado dessa etapa são pares de objetos e suas propriedades;
2. classificação dos objetos usando *Formal Concept Analysis* (FCA): constrói-se uma estrutura de conceito a partir dos pares (objeto, propriedade). O resultado dessa etapa é hierarquia de conceitos;
3. conversão do conhecimento presente nos recursos do domínio (tesauros, bases de dados) em uma estrutura usando FCA. O resultado dessa etapa são hierarquias atribuídas pela base de dados;
4. mesclagem das duas estruturas desenvolvidas nos passos 2 e 3. O resultado dessa etapa são a mesclagem e o enriquecimento das hierarquias;
5. representação da estrutura final em lógica formal, mais especificamente em Lógica Descritiva. Por fim, o resultado é a Ontologia em Lógica Descritiva.

A metodologia PACTOLE foi aplicada em 11591 resumos do periódico *Astronomy e Astrophysics* entre os anos 1994 e 2002. Os resultados da avaliação da metodologia mostraram um índice de precisão de 74.71%, que significa que os objetos estão classificados nas classes adequadas da ontologia. Por outro lado, a revocação é baixa, devido ao número pequeno de propriedades associadas aos objetos. Além dessa deficiência no estudo, citam-se outras: não é previsto o enriquecimento para todos os componentes da ontologia, haja vista que não se considera o enriquecimento de axiomas; a metodologia não é detalhada, o que dificulta para o ontologista a execução de alguns passos.

Reiter e Buitelaar (2008), no estudo *Lexical enrichment of a human anatomy ontology using WordNet*, partem do pressuposto de que as ontologias de domínio representam o conhecimento com menos ênfase nas realizações linguísticas (palavras) das entidades (conceitos), enquanto recursos léxicos, tais como WordNet definem as palavras, seus significados linguísticos e características morfossintáticas; porém apresentam pouca ênfase no contexto do conhecimento associado a essas

palavras. Sendo assim, o método se respalda na base léxica WordNet, verbetes sobre anatomia humana na Wikipedia e no *corpus* especializado *British National Corpus* para extração de realizações linguísticas (sinônimos) com o intuito de enriquecer a ontologia *Foundational Model of Anatomy* (FMA). O método dos pesquisadores se respalda principalmente na técnica de *Word Sense Disambiguation* (WSD). O teste empírico foi realizado em uma amostra de 50 termos da FMA, e os resultados demonstraram uma performance superior da ontologia enriquecida em comparação com a ontologia original, no que se refere às medidas de precisão, revocação e medida F.

Os pesquisadores da Universidade Utrecht, na Holanda, Monachesi e Markus (2010), no artigo *Using social media for ontology enrichment*, abordam um método automático para enriquecer léxico (sinônimos), conceitos e relações de ontologias de domínio a partir de dados extraídos de aplicações de mídia social (*Delicious*, *Youtube* e *Slideshare*). O método se baseia em etiquetas (*tags*) extraídas dessas mídias com o uso de um *crawler* e compara essas etiquetas com termos já existentes na ontologia que se deseja enriquecer. Para comparação, utilizam-se técnicas de medida de similaridade, como co-ocorrência, similaridade de Cosine⁷. Ademais utiliza-se o *dataset DBpedia* como fundamento para identificação de relações taxonômicas e não taxonômicas. Aplica-se o método na ontologia de computação *Language Technology for eLearning* (LT4eL) e, para fins de validação do método, realiza-se uma comparação entre três ontologias: 1) a ontologia LT4eL original, ou seja, que, depois de construída, não passou por um processo de enriquecimento; 2) a ontologia original enriquecida manualmente (tomada no estudo como *Gold Standard*, ou Modelo Padrão); 3) a ontologia enriquecida automaticamente, como o método ora proposto. O resultado da comparação mostra que, dos léxicos e conceitos candidatos para enriquecimento adquiridos com o método, 80% são relevantes. Com relação às relações, a taxa é de 77%, o que revela a eficiência do método para enriquecimento.

No artigo *Enriching the Gene Ontology via the dissection of labels using the ontology pre-processor language*, Fernandez-Breis *et al.* (2010) abordam como proposta principal o enriquecimento axiomático de uma parte da *Gene Ontology*, conhecida como *Molecular Function*. O método se baseia na exploração semântica

⁷*Cosine similarity*. Given two vectors, cosine similarity is used to compute the similarity between two tags. The vectors can be computed in different ways, which leads to the distinction of different approaches (MONACHESI, MARKUS, 2010, p.171).

dos rótulos das classes e nas convenções de nomes da GO e de ontologias relacionadas, como a *Relation Ontology* (RO), *Chemical Entities of Biological Interest* (CHEBI), *Protein Ontology* (PRO), que são as fontes de informação. Posto que a estrutura da ontologia apresenta certa regularidade com relação aos rótulos, extraem-se padrões que podem tornar a semântica explícita. Para a execução do método, os autores se respaldam nos *templates Ontology Design Patterns* (ODP) e *Ontology Preprocessor Language* (OPPL), que se refere a uma linguagem de *script* para aplicar os *templates* ODP. Os resultados mostram que o método conseguiu enriquecer a ontologia de forma considerável e com o custo relativamente baixo.

Harb e Hajlaoui (2010), no estudo intitulado *Enhanced semantic automatic ontology enrichment*, apresentam uma proposta automática para enriquecimento conceitual em uma ontologia sobre competências na indústria mecânica. As técnicas utilizadas para extração da informação se baseiam em análise morfossintática, na técnica de *Data Mining* nomeada *Association Rules* e na medida estatística *Mutual Information*. Apesar de preliminares, os resultados são considerados promissores pelos pesquisadores com base no teste realizado.

Na tese *Ontology enrichment from free-text clinical documents: a comparison of alternative approaches*, Liu (2011) apresenta um estudo completo e detalhado com a comparação de três métodos (um método associado à linguística e dois métodos relacionados a medidas estatísticas) geralmente utilizados para enriquecimento de ontologias. Para comparação dos métodos, a autora propõe o enriquecimento lexical, conceitual e de relações de duas ontologias do domínio biomédico, uma na área de Patologia (NCIT – *National Cancer Institute Thesaurus*⁸) e outra na área de Radiologia (*Radiology Lexicon* – RadLex). Para tanto, baseia-se em documentos clínicos como fonte de informação e aplicam-se as seguintes técnicas: *Hearst Lexico Sintatic Pattern* (LSP); POS *tagging*; NER; Análise estatística; Medidas de similaridade. Os resultados dos testes empíricos são avaliados, e a autora chega à conclusão de que, de forma geral, os três métodos têm potencial para enriquecimento de ontologias no domínio biomédico, apesar de algumas limitações e da importância de especialistas do domínio para a validação do conhecimento para enriquecimento.

No âmbito do enriquecimento axiomático, tipo de enriquecimento mais raro na literatura, o francês Ferré e o alemão Rudolph (2012), no artigo *Advocatus Diaboli:*

⁸ Apesar de receber o nome de tesouro, o NCIT apresenta várias características de ontologias.

exploratory enrichment of ontologies with negative constraints, propuseram um método original para o enriquecimento de axiomas com base em restrições negativas. O objetivo do método é mostrar as lacunas com relação às restrições formais da ontologia. Dessa forma, um especialista do domínio explora toda a estrutura da ontologia e tenta inserir uma declaração falsa, ou seja, que não tem relação com nenhum fato da realidade. Caso a ontologia permita a inserção de declarações absurdas, isso revela que existe nela uma falha e, dessa forma, adiciona-se uma restrição proibindo a inserção daquele tipo de declaração absurda. O método é aplicado na conhecida ontologia sobre *pizza* (ontologia que geralmente já vem instalada nas configurações originais do Software Protégé) com o propósito de demonstração.

A dissertação intitulada *Ontology enrichment based on unstructured text data*, de Luksová (2013), apresenta um método semiautomático para enriquecimento conceitual e de relações de uma ontologia com base em documentos textuais. A abordagem da autora se respalda em técnicas *Machine Learning* e de Processamento de Linguagem Natural, como lematização, análise morfológica e *parsing*. O método foi aplicado no domínio de descrição de trabalho, e os resultados do teste empírico revelaram que o método atingiu uma taxa de 71% na métrica Medida-F para os conceitos extraídos e 68% para as relações extraídas. Os resultados são considerados satisfatórios para uso prático do método proposto.

No mesmo ano, Booshehri *et al.* (2013), no artigo *Ontology enrichment by extracting hidden assertional knowledge from text*, trazem uma proposta automática para enriquecimento de relações não taxonômicas de ontologias a partir do conhecimento implícito em documentos textuais. Utiliza-se *Linked Data* para explorar o conteúdo de documentos e identificar relações candidatas para a ontologia. Aplica-se a proposta em uma ontologia do domínio geográfico para fins de validação.

O estudo *Domain ontology enrichment based on the semantic component of LMF-standardized dictionaries*, de Ben Amar, Gargouri e Ben Hamadou (2013), apresenta uma proposta automática para enriquecimento de ontologia de domínio para a língua árabe com base na extração de padrões léxicos sintáticos em dicionários LMF (*Lexical Markup Framework*). O método apresentado enriquece os conceitos e relações de uma ontologia no domínio astronômico. Os autores consideraram os conceitos, e as relações obtidas para enriquecimento da ontologia se mostraram relevantes, provando, assim, a eficiência do método.

Em prosseguimento, no artigo *OntoRich: a support tool for semi-automatic ontology enrichment and evaluation*, Barbour, Blaga e Groza (2013) desenvolveram um sistema denominado Ontorich, que tem como objetivo o enriquecimento de ontologias a partir de RSS Feeds. O enriquecimento é realizado com o uso de OpenNLP API, que é um repositório de PLN e o recurso WordNet. Para extrair relações e conceitos de RSS Feeds, métodos estatísticos e sintáticos são aplicados com o uso do OpenNLP API. As fases propostas pela ferramenta Ontorich para realizar o enriquecimento de ontologias são:

1. extração automática de conteúdo da Web com base na tecnologia RSS;
2. extração de termos do documento RSS. Nessa etapa, o ontologista tem a opção de selecionar vários métodos para processar o documento e extrair termos. A primeira categoria de métodos refere-se àqueles baseados em estatística, ou seja, frequência do termo e frequência inversa no documento (TF-IDF);
3. aplicação de métodos relacionados ao WordNet. O poder semântico do WordNet é explorado com o uso de métodos para recuperar hipônimos e merônimos. Assim, cada palavra exibida na árvore hipônimo pode ser selecionada e adicionada à ontologia como subclasse de um conceito especificado.

A ferramenta é testada em duas ontologias, uma de vinhos e outra sobre Tecnologia da Informação (TI). Depois da fase de enriquecimento, os autores usam várias métricas para medir como a ontologia inicial modificou-se. Ademais, o sistema Ontorich é comparado a duas outras ferramentas de enriquecimento, Kaon e Neon. Os resultados mostram que a ferramenta Ontorich se mostra superior às outras nas categorias elencadas pelos autores para análise de comparação.

Os pesquisadores do Instituto para Programação de Sistemas da Academia de Ciências Russa, Astrakhantsev, Fedorenko e Turdakov (2014), na pesquisa *Automatic enrichment of informal by analyzing a domain-specific text collection*, desenvolveram um método para enriquecimento automático. Os autores se baseiam na fonte de informação mais comum para a aquisição de conhecimento, que são os *corpora* textuais. O método dos autores se respalda em várias técnicas do Processamento de Linguagem Natural, tais como detecção de sentenças, *tokenização*, *POS tagging*, *Lematização*, *Word Sense Disambiguation*, além de algoritmos *Machine Learning* e

técnicas estatísticas, como IF-IDF, frequência, contagem de palavras, etc. O método é aplicado no contexto de jogos de tabuleiro, e os resultados não são considerados satisfatórios, revelando que a aplicação do método de forma totalmente automática não produziu bons resultados, visto que apenas cerca de 33% da terminologia pode ser reconhecida (o reconhecimento da terminologia é uma das primeiras tarefas do enriquecimento de ontologias). Assim, os autores concluem que a interferência de um especialista é necessária para validar os termos incluídos na ontologia, ou seja, um método semiautomático seria o desejável para um melhor desempenho nos resultados.

Quesada-Martínez *et al.* (2015) publicaram o artigo *Approaching the axiomatic enrichment of the Gene Ontology from a lexical perspective*, no qual o principal objetivo é entender como padrões léxicos (ou regularidades léxicas) em rótulos de ontologias do domínio biomédico podem ser usados para o enriquecimento axiomático, uma vez que os rótulos facilitam o entendimento humano, mas não são significativos para as máquinas, que necessitam de representações formais em forma de axiomas. O método é executado por um software chamado OntoEnrich, que executa uma análise léxica da ontologia que será enriquecida (ontologia fonte). Em seguida executa a análise léxica das ontologias fonte. Assim os arquivos gerados de cada análise são comparados com base em uma métrica específica que revela as classes nas quais a métrica é verdadeira, ou seja, em que há relação entre as ontologias. Os autores aplicam o método proposto na Gene Ontology (GO) e utilizam como fontes de informação as ontologias *Chemical Entities of Biological Interest* (CHEBI) e *Cell Ontology* (CO). Por fim os pesquisadores afirmam que os resultados suportam a hipótese de que a abordagem léxica pode contribuir para o enriquecimento axiomático no contexto das ontologias biomédicas.

As pesquisadoras húngaras Gillani e Ko (2015) apresentam um estudo intitulado *Incremental ontology population and enrichment through semantic-based text mining: an application for IT audit domain*, no qual propõem um método semiautomático denominado ProMine para enriquecimento conceitual de uma ontologia do domínio de auditoria de tecnologia da informação. Para tanto, os autores se baseiam em documentos textuais e em recursos léxicos (*Wiktionary*, *WordNet*) como fonte de informação e nas técnicas de análise sintática (*tokenização*, *POS tagging*, *stemming*, remoção de *stop words*, lematização), análise estatística (cálculo de frequência), expansão de termos e similaridade semântica. Foram realizados

experimentos para validação do método, e os resultados demonstraram que o desempenho do método ProMine é superior a dois outros métodos similares que foram levantados na revisão de literatura com relação às métricas de precisão e revocação. Ademais as pesquisadoras afirmam que os resultados mostram que a abordagem pode apoiar especialistas de domínio na construção de ontologias em domínios específicos de forma satisfatória.

Taboada *et al.* (2017) publicaram um estudo denominado *A new synonym-substitution method to enrich the Human Phenotype Ontology*, no qual o propósito é o enriquecimento lexical de forma automática da ontologia *Human Phenotype Ontology* (HPO). De forma mais específica, objetiva-se atribuir novos sinônimos para a ontologia. Para tanto, o método proposto pelos autores se baseia na própria estrutura hierárquica da ontologia que se objetiva enriquecer e nas propriedades lexicais dos conceitos dela para geração de sinônimos. Além disso, o método é aprimorado com o uso da base de dados MEDLINE, que contribui para garantir que determinado sinônimo está associado ao domínio da ontologia HPO. São aplicadas as técnicas de *Named Entity Recognition* (NER) e similaridade léxica. Os resultados da validação de método demonstraram que a capacidade da ontologia em reconhecer conceitos do domínio melhorou com os novos sinônimos adicionados.

No estudo *Enriching semantic relations of basic sciences ontology*, Ejei *et al.* (2017) desenvolveram uma proposta semiautomática para enriquecimento de relações semânticas entre conceitos de uma ontologia sobre ciência básica. O método faz uso de ontologias (de domínio e de alto nível) e tesouro como fonte de informação e se respalda na técnica de identificação de padrões na ontologia de alto nível BFO (*Basic Formal Ontology*). Os autores aplicam o método em uma parte de uma ontologia sobre ciência básica, mais especificamente na parte que trata sobre conceitos básicos de química, porém não é realizada nenhuma análise dos resultados desse enriquecimento.

No estudo *Dwipa Ontology III: implementation of ontology method enrichment on tourism domain*, Kuntarto *et al.* (2017) desenvolveram um método semiautomático para enriquecimento de conceitos e relações de ontologia na área de turismo na Indonésia. O método utiliza como fonte de informação páginas Web sobre turismo e aplica técnicas de análise estatística (TF-IDF) e de análise sintática (POS *tagging*). A ontologia na qual a proposta de enriquecimento foi aplicada denomina-se DWIPA e atualmente encontra-se em sua segunda versão (DWIPA II). A ontologia foi

enriquecida com seis novas subclasses e relações (não fica claro no artigo o número de novas relações) e foi avaliada com base no método de comparação a uma fonte de informação (foi comparada com páginas Web do domínio de turismo) e com base na avaliação por um especialista do domínio. Os autores concluem que a expansão da ontologia com novos componentes está de acordo com a realidade do domínio do turismo.

Em *Automatic domain-specific learning: towards a methodology for ontology enrichment*, Gómez-Moreno e Mestre-Mestre (2017) propõem o método OEM (*Ontology Enrichment Method*), que tem o propósito de expandir automaticamente ontologias com base em um *corpus* de dados terminológicos. Os autores destacam que em sua fase atual o método foi desenhado para enriquecimento de relações taxonômicas do tipo “é_um”. As principais técnicas utilizadas referem-se àquelas usadas no PLN, como *tokenização*, por exemplo. O OEM consiste de três etapas:

1. Identificação da ontologia: o primeiro passo é ter uma ontologia preestabelecida. Essa ontologia tem o propósito de fornecer a base lexicográfica e/ou a informação de cada conceito superordenado, incluindo uma descrição dos eventos principais, qualidades e entidades relacionadas. Esses descritores serão usados como marcos conceituais para orientar o sistema no enriquecimento da ontologia com os novos termos adquiridos. Outro propósito da ontologia é funcionar como um modelo padrão contra o qual irá se comparar a ontologia enriquecida por meio das medidas de precisão e revocação;
2. Compilação do *corpus*: refere-se à seleção e tratamento do *corpus* representativo da ontologia de domínio que está sendo enriquecida. A partir das informações obtidas dos conceitos da ontologia na etapa 1, identifica-se um conjunto de atributos associado a esses conceitos. Assim, para aplicação do método OEM em áreas específicas do conhecimento, faz-se necessário o uso de um *corpus* especializado. Em seguida, aplicam-se filtros no *corpus* para exclusão de palavras sem importância (conjunções, símbolos, *stopwords* em geral) e identificação de palavras de alta frequência no *corpus* e que possuam

mais relação com o termo da ontologia. Essa relação é medida por meio de NPMI⁹ (Pontual de Informação Mútua Normalizada);

3. Classificação dos dados: por meio do resultado do NPMI da etapa 2, os atributos que possuem valor mais alto têm mais chance de terem relação com determinado conceito da ontologia; assim eram associados a um conceito superordenado da ontologia.

Para fins de validação, o método foi aplicado no domínio da Virologia. O método funcionou como uma prova de conceito sobre expansão automática de ontologia. Portanto restringiu-se a apenas uma palavra-chave (dengue) dentro da ontologia escolhida e extraíram-se no corpus selecionado palavras relacionadas. Resultados mostram que o método funciona de forma satisfatória, haja vista que as palavras extraídas do *corpus* (como febre, vírus, mosquito) possuíam relação com a palavra dengue, possibilitando a classificação automática de novos conceitos para a ontologia. Porém a validação do método ocorreu em um cenário muito limitado; assim o desempenho do sistema com uma gama maior de dados pode trazer resultados diferentes. Ademais o método propõe o enriquecimento apenas de uma parte da ontologia, ou seja, a inclusão de novos conceitos na hierarquia, obedecendo às relações taxonômicas.

Idoudi *et al.* (2018), no artigo *Ontology knowledge mining for ontology conceptual enrichment*, propõem o enriquecimento conceitual de uma ontologia no domínio mamográfico a partir do uso da técnica de *clustering*. Os autores apontam primeiramente as limitações de estudos que aplicam métodos estatísticos ou intervenção humana para realizar o enriquecimento de ontologias, bem como as limitações de estudos que aplicam outras técnicas de *clustering*. O método se baseia na geração de *cluster* (grupo de conceitos similares) de duas ontologias (uma ontologia alvo, ou seja, a que será enriquecida; e uma ontologia fonte, a ontologia que fornece os insumos para enriquecimento). Em seguida, esses *clusters* gerados nas duas ontologias são comparados e os similares são usados para enriquecer a ontologia alvo. A ontologia no qual o método foi aplicado denomina-se *Breast Cancer Grading Ontology* (BCGO) e faz parte de um projeto para análise de imagens de

⁹NPMI (Normalized Pointwise Mutual Information) é uma métrica de associação usada na Linguística para identificar a atração semântica entre duas variáveis. Essa medida vai de 0 a 1.

mamografia para auxílio no diagnóstico de câncer. Os resultados da aplicação empírica mostram que o método é efetivo no enriquecimento da ontologia.

Os gregos Kokla e Tomai (2018), com o estudo *Enrichment and population of a geospatial ontology for semantic information extraction*, apresentaram uma proposta de enriquecimento lexical, conceitual e de relações de uma ontologia do domínio geográfico. O método dos autores toma como fonte de informação uma ontologia de domínio (*Geonames*), base léxica (*BabelNet*) e conteúdo de páginas Web. As técnicas utilizadas para extração da informação são análises das definições dos conceitos providas pelas fontes de informação, principalmente a base léxica *BabelNet* e a ontologia *Geonames*, além de técnicas de Processamento de Linguagem Natural (PLN), tais como *tokenização*, *part-of-speech (POS) tagging*, divisão de sentenças e lematização. Um exemplo da aplicação do método na classe "terremoto" é apresentado, porém os resultados da aplicação do método não são avaliados.

Guerram e Mellal (2018), no artigo *A domain independent approach for semantic ontology enrichment*, propõem uma abordagem geral para enriquecimento semântico de ontologias. O objetivo da proposta é enriquecer uma ontologia inicial, no que concerne aos seus conceitos e relações taxonômicas e não taxonômicas, por meio de textos do domínio alvo. Para isso, utilizam-se análise sintática e semântica no *corpus* de texto relacionado ao domínio. A análise sintática é realizada usando-se ferramentas de Processamento de Linguagem Natural para obter a marcação gramatical (*part of speech tags – POS*) e o reconhecimento de nome de entidade (*Named Entity Recognition*) no *corpus*. As fases da proposta são:

1. Anotação dos textos utilizando análise morfossintática para prover o primeiro nível de entendimento dos textos dados (apontados). Os textos são analisados (com a técnica *parsing*) para extrair relações sintáticas entre termos, assim como a marcação gramatical (*part of speech tags – POS*) desses termos;
2. Simplificação de sentenças complexas para condições simples, que consiste na análise semântica do texto anotado, no qual as condições obedecem à forma: Verbo (Sujeito, Objeto);
3. Enriquecimento semântico, que se refere à comparação de cada relação extraída com o conteúdo da ontologia inicial usando uma medida de similaridade. De acordo com essa comparação, decide-se se a relação extraída pode ser candidata para enriquecer a ontologia inicial. A medida de

similaridade é baseada no WordNet, e o processo de enriquecimento objetiva identificar novos conceitos, novas relações ou apenas instâncias de conceitos e relações que já existem. Nessa etapa, estuda-se, para cada relação Verbo (Sujeito, Objeto) extraída do texto, a medida semântica dessa relação com as relações existentes na ontologia para identificar novos conceitos e relações que podem ser relevantes para enriquecer a ontologia.

Na proposta de Guerram e Mellal (2018), não é realizado nenhum tipo de avaliação (esse é um trabalho futuro que os autores conjecturam), o que impossibilita identificar claramente as vantagens da mencionada proposta. Uma contribuição da proposta dos autores é que pode ser aplicada a qualquer domínio do conhecimento. A proposta de enriquecimento não se aplica a todos os componentes das ontologias e também não detalha muito bem como realizar cada etapa para alcançar o enriquecimento.

Em 2019, os indonésios Kuntarto, Isyahrani e Gunawan publicaram o artigo intitulado *Performance of K-Means Clustering Algorithm in enriching a new concept of amenities into Dwipa Ontology III within indonesia Tourism Domain*. O propósito desse artigo foi o enriquecimento de uma ontologia da área de turismo com novos conceitos a partir da extração de dados do site TripAdvisor. Os pesquisadores aplicaram técnicas da área de Mineração de Dados, mais especificamente o Algoritmo de Clustering K-Means e técnicas de medida estatística como TF-IDF (Term Frequency – Inverse Document Frequency) para atribuição de pesos. Os resultados foram considerados satisfatórios pelos autores, visto que a ontologia foi enriquecida com mais quatro classes principais e 29 subclasses. Porém pode-se destacar as seguintes lacunas no método: enriquecimento de apenas um componente, fonte de informação muito específica, aplicação do método em uma amostra pequena e falta de detalhamento das etapas do método.

Ali *et al.* (2019), no artigo *CLOE: a cross-lingual ontology enrichment using multi-agent architecture*, abordam um método inovador para enriquecimento conceitual e de relações de ontologias multilíngues. As fontes de informação são textos em pelo menos dois idiomas e a ontologia que se deseja enriquecer. O método é independente de domínio e de idioma e faz uso das seguintes técnicas para extração da informação: *POS tagging*, *sentence splitter*, *tokenização*, *stop words*, *machine learning*. Para validação do método, os pesquisadores fazem um experimento usando

textos de três idiomas distintos (inglês, alemão e árabe) e o aplicam em uma ontologia do domínio da Tecnologia da Informação (TI). Por fim, o método de enriquecimento e a ontologia enriquecida são avaliados com base na aplicação de um questionário de satisfação e com base na comparação da ontologia contra um Modelo Padrão (*Gold Standard*). Com relação ao questionário, 79% dos participantes (ontologistas) que responderam a ele disseram estar satisfeitos com a completude das relações e conceitos candidatos para enriquecimento da ontologia. No que tange à avaliação com base na comparação a um Modelo Padrão, os resultados também foram satisfatórios, visto que o método CLOE superou três outros métodos.

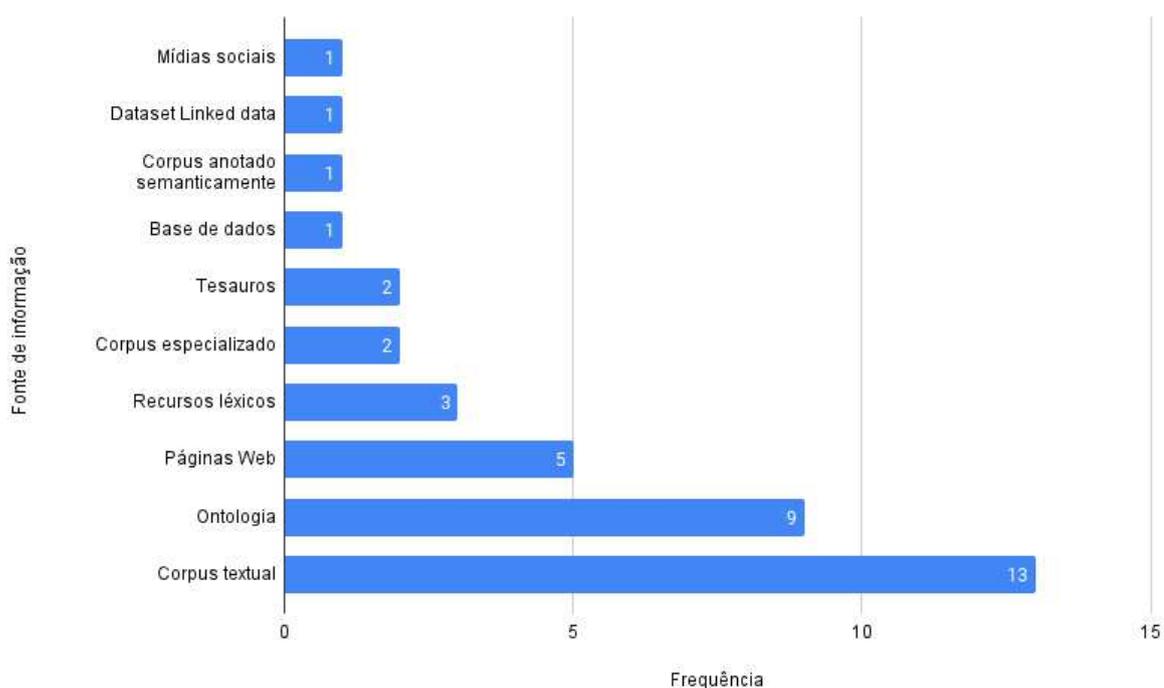
2.4 Análise da revisão de literatura

Percebe-se, com base nas características apresentadas de cada um dos 29 estudos selecionados nesta revisão de literatura, que existe uma grande heterogeneidade de métodos para enriquecimento de ontologias. Todos eles apresentam uma característica própria e não existe nenhum método igual. No entanto também existem algumas similaridades entre os métodos: todos partem de uma ontologia inicial, todos se baseiam em alguma fonte de informação para extração de informações para enriquecimento (ainda que os tipos de fonte de informação escolhidas sejam diferentes); todos propõem pelo menos o enriquecimento de um componente da ontologia; todos eles baseiam-se em uma ou mais técnicas para extração da informação. Esses são os elementos básicos do processo de enriquecimento, posto que não existe enriquecimento sem uma ontologia já desenvolvida; também é preciso extrair as informações de alguma fonte (ainda que a fonte de informação seja a própria ontologia). Por conseguinte, utiliza-se, no mínimo, uma técnica para extração da informação e, por fim, é preciso enriquecer ao menos um tipo de componente da ontologia.

Com o objetivo de coletar dados referentes aos 29 estudos selecionados e descritos nesta revisão de literatura, foram tabulados dados das seguintes categorias: (1) fonte de informação, (2) tipo de enriquecimento (com as subcategorias lexical, conceitual, de relações taxonômicas, de relações não taxonômicas e de axiomas) e técnica(s) aplicada(s), (3) detalhamento das etapas (com os campos sim, não e parcialmente) e (4) principais lacunas dos métodos e aspectos favoráveis.

O Gráfico 6 – Frequência das fontes de informação utilizadas para enriquecimento apresenta a sistematização dos métodos quanto à fonte de informação utilizada para extração dos dados. Percebe-se que a fonte de informação mais comum para extração da informação para enriquecimento são os *corpus* textuais. Isso tem relação com o fato de ser o tipo de fonte de mais fácil acesso, com maior disponibilidade em vários domínios do conhecimento. Ademais os documentos textuais são importantes fontes de onde podem-se extrair aspectos linguísticos que podem ser transformados em aspectos semânticos na estrutura da ontologia por meio dos relacionamentos entre os conceitos. O Gráfico 6 mostra, também, que as ontologias são muito utilizadas para se extrair informações passíveis de contribuir com o enriquecimento.

Gráfico 6 – Frequência das fontes de informação utilizadas para enriquecimento

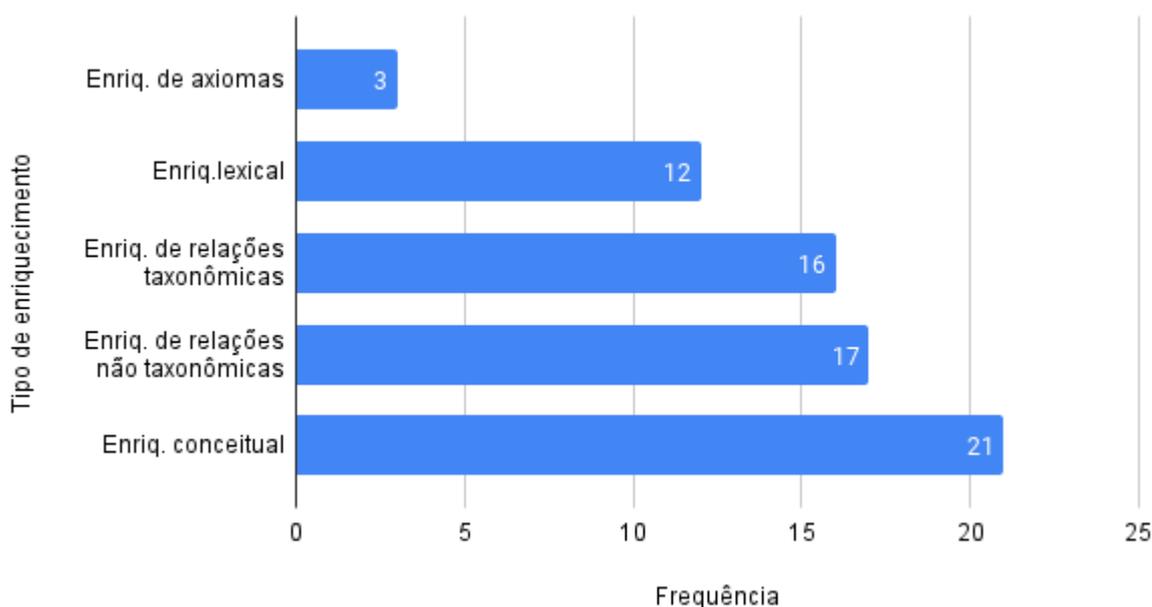


Fonte: dados de pesquisa (2021).

Percebeu-se, nos estudos, que o uso de ontologias como fonte de informação é muito comum no domínio biomédico, mais especificamente com relação às ontologias que fazem parte do Consórcio OBO (*Open Biomedical Ontologies Foundry*). Essas ontologias possuem características comuns e se fundamentam em uma mesma ontologia de alto nível, a *Basic Formal Ontology* (BFO), o que facilita a aplicação de algumas técnicas de extração da informação.

O Gráfico 7 – Frequência dos tipos de enriquecimentos realizados revela que o enriquecimento conceitual é o tipo mais comum nos estudos analisados. Isso tem relação com o fato de esse tipo de enriquecimento exigir técnicas menos complexas e mais comuns na literatura; além disso é mais simples a identificação de elementos linguísticos e a transformação desses elementos em conceitos (classes e subclasses na ontologia), se compararmos com o enriquecimento de outros componentes, que exigem técnicas mais aprimoradas. Já o enriquecimento de axiomas é o que menos aparece nos estudos, com apenas três estudos que propõem o enriquecimento de tal componente. Argumenta-se que as razões para isso também têm relação com a complexidade, posto que o enriquecimento desse tipo de componente dificilmente é realizado de forma automática, uma vez que exige mais capacidade de análise, necessitando de um especialista para adicionar os axiomas.

Gráfico 7 – Frequência dos tipos de enriquecimentos realizados

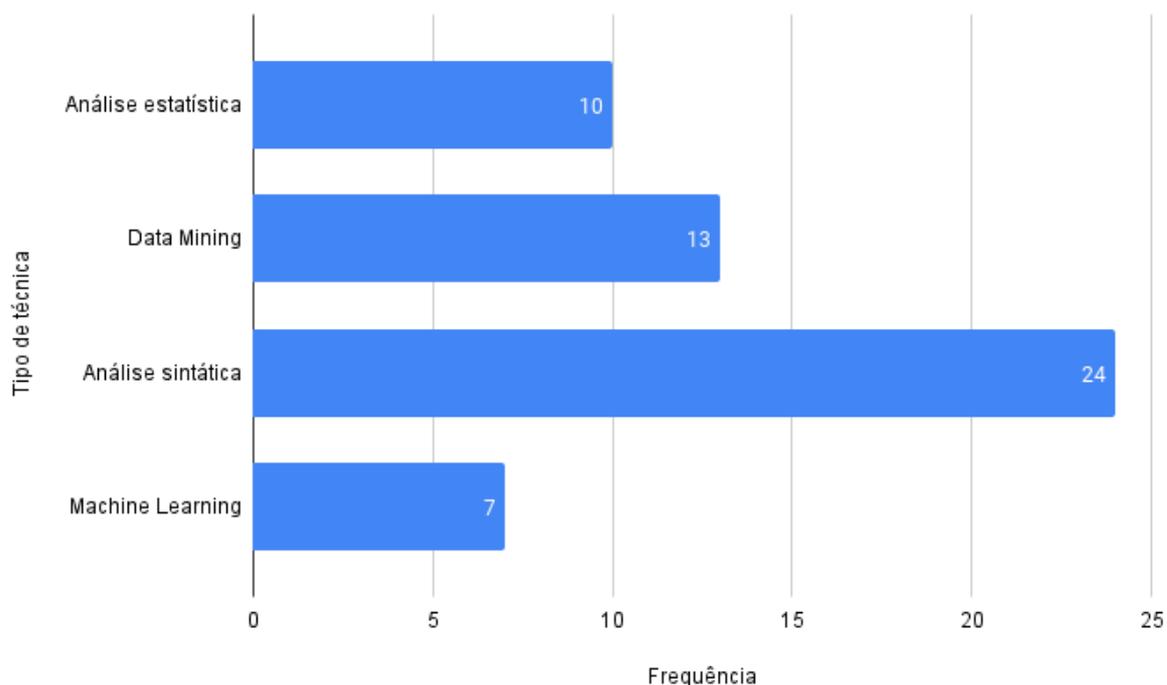


Fonte: dados de pesquisa (2021).

O Gráfico 8 – Frequência dos tipos de técnicas utilizadas para extração da informação revela que existe uma grande diversidade de técnicas. Portanto, para facilitar a identificação, podemos classificar essas técnicas em: análise sintática, análises estatísticas, *Data Mining* e *Machine Learning*. Percebe-se que as técnicas do tipo análise sintática são as mais recorrentes. Isso se deve à existência de um grande número de técnicas ligadas à análise sintática (ou técnicas de Processamento de Linguagem Natural), tais como *tokenização*, *parsing*, *part-of-speech tagging*,

lematização, etc. O [Apêndice B – Dados sobre documentos das Revisão de Literatura](#) apresenta todas as técnicas empregadas.

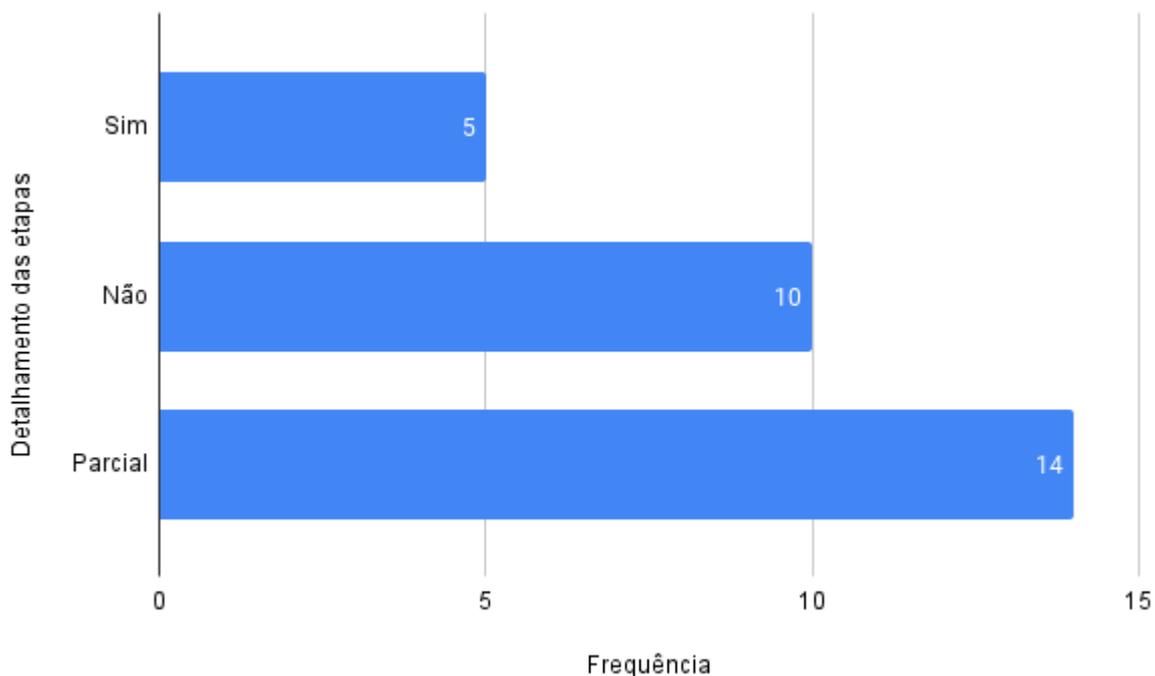
Gráfico 8 – Frequência dos tipos de técnicas utilizadas para extração da informação



Fonte: dados de pesquisa (2021).

Com relação ao detalhamento das etapas de cada proposta, o Gráfico 9 revela que a maioria (48,3%) delas apresenta um detalhamento parcial sobre as etapas do método proposto. Destaca-se também que uma grande parcela (34,5%) das propostas não detalha como o método de enriquecimento foi aplicado, impossibilitando, assim, a sua replicação.

Gráfico 9 – Detalhamento das etapas do método nos estudos



Fonte: dados de pesquisa (2021).

A partir da revisão de literatura, perceberam-se algumas limitações dos métodos propostos para enriquecimento de ontologias. Entre as principais, estão:

- I. **Técnicas que exigem a supervisão de um especialista:** as técnicas empregadas no método exigem um linguista computacional para sua aplicação.
- II. **Métodos muito específicos para o domínio biomédico:** as técnicas e as fontes utilizadas no método não podem ser generalizadas para outros domínios, visto que foram construídas especificamente para aplicação no domínio biomédico.
- III. **Resultado insatisfatório do enriquecimento totalmente automático:** métodos que não possuem a intervenção de um especialista tendem a não produzir resultados satisfatórios.
- IV. **Falta de previsão do enriquecimento de todos os componentes:** é importante que uma metodologia possibilite o enriquecimento de todos os componentes das ontologias de domínio.

- V. **Ausência de detalhamento dos passos para execução do método:** as formas como as atividades devem ser desenvolvidas precisam ser claras e objetivas para que facilitem o trabalho do ontologista.
- VI. **Dificuldade de acesso às fontes de informação por serem muito específicas:** as fontes de informação não podem ser de difícil acesso. A metodologia deve oferecer opções de fontes de informação.
- VII. **Validação em um cenário limitado:** a atividade de validação ou avaliação do método de enriquecimento deve possibilitar avaliar todos os aspectos do método.

No entanto outros dados que emergiram da análise dos documentos referem-se aos aspectos favoráveis, que se destacam como critérios almejados pela metodologia de enriquecimento que aqui se objetiva. Entre os aspectos favoráveis, aparecem:

- I. **Independência de software ou ferramenta específica:** como as tecnologias da informação e comunicação evoluem rapidamente, softwares podem ser descontinuados ou podem se tornar obsoletos.
- II. **Independência de domínio:** uma característica importante que as metodologias devem possuir é a possibilidade de serem aplicadas em diferentes domínios do conhecimento para ocorrência da generalização do método.
- III. **Uso de fonte de informação de fácil acesso:** as fontes de onde são extraídas as informações para enriquecimento das ontologias devem ser de fácil acesso em qualquer domínio do conhecimento, visto que se trata da matéria-prima para realização do processo de enriquecimento.
- IV. **Descrição das etapas do método bem descritas:** o método apresenta de forma detalhada o passo a passo para a sua execução, possibilitando, assim, a replicação.
- V. **Enriquecimento da ontologia por completo:** o método possibilita o enriquecimento de todo o conteúdo da ontologia e não apenas de uma parcela.
- VI. **Aplicação do enriquecimento de axiomas:** o método propõe o enriquecimento de axiomas, proposta de enriquecimento raro na literatura.

A literatura apresenta alguns estudos autodenominados propostas de enriquecimento de ontologias, porém o enriquecimento é feito em bases léxicas, como

o *WordNet* (AGIRRE *et al.*, 2000); (GHARIB *et al.*, 2012); em *datasets* de *Linked Data*, como a DBpedia (TÖPPER; KNUTH; SACK, 2012); ou propõe o enriquecimento com instâncias (VALARAKOS *et al.*, 2001), que foge da proposta deste trabalho, haja vista que aqui se considera a tarefa de adicionar instâncias a uma ontologia como População de Ontologias.

3 FUNDAMENTAÇÃO TEÓRICA E METODOLÓGICA

Neste capítulo, são apresentados os conceitos das linhas gerais que dão embasamento teórico a esta tese. Assim, na seção [3.1 Manutenção e atualização de Sistemas de Organização do Conhecimento](#), definem-se esses sistemas e os caracteriza, pois eles contextualizam as ontologias no âmbito da CI, assim como os estudos sobre manutenção e atualização desses sistemas são importantes para a área específica de Engenharia Ontológica. A seção [3.2 Uso da Teoria do Conceito no contexto de Sistemas de Organização do Conhecimento](#), versa sobre o que é a teoria e a importância no contexto dos SOCs. Na seção [3.3 Ontologias](#), trata-se especificamente desse SOC, pois é parte inerente do objeto de estudo. A seção [3.4 Engenharia Ontológica](#) se dedica ao estudo dos processos associados ao ciclo de vida das ontologias, com o intuito de contextualizar a temática de estudo e algumas temáticas que possuem relação com a manutenção e a atualização de ontologias. Por fim, a seção [3.5 Enriquecimento de Ontologia](#) trata especificamente da temática do presente estudo, na qual se discute a relação com o assunto aprendizagem de ontologia e são apresentados os tipos de enriquecimento, técnicas utilizadas e fontes de dados usadas para extrair o conhecimento.

3.1 Manutenção e atualização de Sistemas de Organização do Conhecimento

A expressão Sistema de Organização do Conhecimento (em inglês *Knowledge Organization System*), de acordo com Hodge (2000), foi proposta em 1998 pelo Networked Knowledge Organization Systems Working Group, na primeira conferência da ACM Digital Libraries, em Pittsburgh, Pensilvânia. A expressão é utilizada para abranger todos os tipos de esquemas para organização da informação e promoção da gestão do conhecimento, tais como sistemas e esquemas de classificação, dicionários, bases de dados lexicais, taxonomias, tesouros, ontologias (HODGE, 2000). Até a expressão ser cunhada, muitos desses instrumentos eram conhecidos na comunidade da Biblioteconomia e Ciência da Informação como vocabulários controlados, linguagens documentárias e/ou linguagens de indexação, mas a história dos SOCs é bem mais antiga; atribui-se aos filósofos o desenvolvimento dos primeiros sistemas para organizar o conhecimento. Aristóteles categorizou o conhecimento em

grupos, tais como Física, Política, Psicologia. Essa divisão, segundo Hodge (2000), influenciou nossa linguagem, educação e ciência. Nesse sentido, de acordo com essa autora, os trabalhos filosóficos de Thomas Bacon influenciaram a criação do Sistema de Classificação da Biblioteca do Congresso Americano.

Para definir Sistema de Organização do Conhecimento, recorre-se aos principais autores da BCI que abordam o assunto, como Hodge (2000), Zeng (2008), Hjørland (2008) e Soergel (2009), juntamente com pesquisadores que fizeram estudos relevantes sobre a temática (SOUZA; TUDHOPE; ALMEIDA, 2012; BRATKOVÁ, KUČEROVÁ, 2014; HJØRLAND, 2013, 2015; LIMA, 2020). A definição de Hodge (2000), conforme já abordada no início desta seção, é abrangente e envolve todos os tipos de esquemas para organização da informação. A autora complementa dizendo que SOCs são usados para organizar materiais com o objetivo de recuperar e gerenciar uma coleção. Além disso, esses sistemas servem como uma ponte entre a necessidade de informação dos usuários e a representação dos documentos de uma coleção.

Encontram-se na literatura diferentes propostas de tipologias para os diversos SOCs existentes (HODGE, 2000; ZENG, 2008; SOERGEL, 2009; SOUZA, TUDHOPE E ALMEIDA, 2012; BRATKOVÁ E KUČEROVÁ, 2014). Definem-se, brevemente, dentre os tipos de SOCs, aqueles que serão mencionados ao longo do presente estudo. São eles: glossários, dicionários, taxonomias, tesouros e ontologias (esse SOC será descrito com mais detalhes na [seção 3.3](#), por ser parte do objeto de estudo). Esclarece-se que se objetiva tão somente apresentar definições breves desses SOCs por eles não serem o objeto do presente estudo; outrossim existem muitos trabalhos de aprofundamento em cada um deles.

Glossários são listas alfabéticas de termos, geralmente com definições e que abrangem um ou mais domínios do conhecimento. Alguns glossários são restritos a um trabalho particular, como um livro, um relatório, um documento da área jurídica, etc. (HODGE, 2000; HJORLAND, 2008; ZENG, 2008; SOERGEL, 2009). Soergel (2009) diz que, apesar de esses instrumentos focarem no acesso ao termo, eles proveem informações sobre conceitos.

Zeng (2008) define **dicionários** como listas alfabéticas de termos e suas definições proveem sentidos variantes para cada termo em outros contextos e/ou áreas do conhecimento. Hjørland (2008) e Soergel (2009) acrescentam que esses instrumentos podem prover outras informações, tais como forma ortográfica correta,

sinônimos, equivalentes em outros idiomas (no caso dos dicionários multilíngues), morfologia e origem do termo. Os dicionários não possuem nenhuma forma de estruturação hierárquica ou tentativa de agrupar termos por conceitos (SOERGEL, 2009).

O termo **Taxonomia** surgiu no contexto da Biologia e era usado para classificação de organismos. Popularizou-se em outras áreas e é usado para se referir à classificação de coisas em diferentes contextos, por exemplo, classificação do conteúdo de páginas Web, em empresas para organização de intranets e outros sistemas corporativos (SOERGEL, 2009). Como um tipo de SOC, pode-se definir Taxonomia como “uma lista de termos preferenciais com estrutura hierárquica, que utiliza a relação de generalização (é_um) ou tipo de.” (MOREIRO-GONZÁLEZ, 2011). A definição apresentada vai ao encontro da apresentada na ANSI/NISO (2005, p. 9, tradução própria¹⁰) “uma coleção de termos de vocabulário controlado organizados em uma estrutura hierárquica. Cada termo em uma taxonomia é um ou mais relacionamentos pai/filho (mais amplo/mais estreito) com outros termos da taxonomia.”

As taxonomias estão presentes na estrutura das ontologias. De fato, a taxonomia é o esqueleto principal da ontologia, em que só existem as relações de generalização-especialização (ou relação é_um) e parte-todo entre as suas classes. Na literatura da área, é comum encontrar a expressão “relação taxonômica” para referir-se ao tipo de relação “é_um”, assim como encontra-se a expressão “relação não taxonômica” para tratar dos outros tipos de relações.

Tesouro é definido na norma ANSI/NISO (2005, p. 18) como “um vocabulário controlado, organizado em uma ordem conhecida e estruturada para que as várias relações entre termos (tesauros apresentam três tipos de relações: de equivalência, hierárquica e associativa) sejam exibidas claramente e identificadas por indicadores padronizados de relacionamento”. Essa definição é acompanhada por Zeng (2008).

Soergel (1999) diz que tesouro é uma estrutura que gerencia a complexidade da terminologia por meio da indicação dos vários significados de homônimos e relacionamentos entre termos idênticos ou que possuem similaridade no significado

¹⁰A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.

(como sinônimos e quase sinônimos). Além disso, ainda na linha desse autor, os tesouros proveem relacionamentos conceituais.

Antes mesmo da expressão Sistema de Organização do Conhecimento ser cunhada, Soergel (1974) já escrevia sobre a manutenção de Linguagens de Indexação e de Tesouros no capítulo nomeado *Updating and maintenance of indexing languages and thesauri*. O autor aborda (1) **tipos de mudanças** que podem ocorrer, como introdução de um novo descritor, eliminação de um descritor, subdivisão de um descritor geral em descritores mais específicos, mudança na definição, adição ou eliminação de relacionamentos, etc; (2) **fontes** de onde novos termos, conceitos e relacionamentos podem ser extraídos, como fontes dentro do próprio Sistema de Recuperação da Informação que utiliza a linguagem de indexação (registros das expressões de busca utilizadas pelos usuários, por exemplo); outras fontes podem ser os documentos de indexação, relatórios sobre novidades na área, discussões informais, reuniões, etc.; (3) **procedimentos para atualizações periódicas**, que envolvem o uso de formulários de tesouros para as atualizações, o processamento das atualizações de informações na estrutura das linguagens, o registro das datas e alterações realizadas, entre outros; (4) **revisão da linguagem de indexação em intervalos longos**; (5) **observações sobre a flexibilidade das linguagens de indexação**; (6) **problemas da reindexação devido à atribuição de novos termos e alterações nos termos existentes**; (7) **problemas comuns da atualização e compatibilização de tesouros**.

Hodge (2000, p. 27) também aborda a manutenção de SOCs. Segundo a autora, a manutenção deve ser considerada no planejamento do instrumento e é responsabilidade dos profissionais que elaboram o instrumento e não do bibliotecário encarregado da biblioteca digital. Hodge (2000) destaca a importância da atualização dos SOCs ao dizer que o uso desses instrumentos desatualizados no contexto de uma biblioteca digital traz mais entraves do que benefícios. A autora levanta algumas questões sobre a manutenção de SOCs; entre elas: Se o provedor mantiver o SOC, como a biblioteca digital descobrirá as alterações que possam ser feitas nele?; Existe um mecanismo de notificação em vigor?; Com que frequência as informações devem ser atualizadas para beneficiar os usuários da biblioteca digital?; A manutenção será autoevidente ou o contrato deve incluir requisitos de notificação?; O que o proprietário fará se a manutenção não puder mais ser executada?

A Norma ANSI/NISO Z39.19 (2005) destaca que os SOCs são reflexões da linguagem e são instrumentos dinâmicos. Assim são necessários procedimentos e políticas para revisão periódica desses instrumentos para a criação de novos termos, substituição de termos obsoletos. Nessa Norma, encontram-se orientações gerais para realização da atualização de um SOC, sempre que houver necessidades (a Norma trata especificamente de Vocabulário Controlado, que também é um tipo de SOC), seja para adição de termos, seja para alteração de termos existentes, seja para exclusão de termos. Além disso, a norma orienta sobre os impactos (negativos ou positivos) que a atualização de um SOC pode causar nas bases de dados ou sistemas que utilizam esse SOC em questão. Recomenda-se a criação de uma nota com a data de cada mudança e identificação do responsável pela mudança. Sobre a adição de termos, sugere-se que, sempre que um termo ou combinação de termos apropriados não puder ser encontrado em um vocabulário controlado, o indexador ou pesquisador deve nomear um novo termo como um termo candidato. Para cada termo, devem-se tomar decisões pautadas nas seguintes questões: i) Esse termo deve ser adicionado ao vocabulário controlado ou ele já é coberto por um termo existente ou pela combinação de outros termos? Se é coberto, há a necessidade de inserir uma remissiva?; ii) É necessária a inclusão de nota de escopo ou outro tipo de nota?; iii) Se o termo deve ser adicionado, qual a forma correta, considerando sintaxe, número e forma escrita?; iv) Como o termo deve ser relacionado com os termos existentes?

Ainda sobre a Norma ANSI/NISO Z39.19-2005, no que concerne à modificação de termos existentes, recomenda-se que indexadores e usuários devem ser capazes de propor modificações a termos existentes ou aos relacionamentos deles, desde que forneçam explicações de base lógica e forneçam documentação que suporte as mudanças propostas. Se o termo for modificado, deve-se registrar a data da alteração na nota histórica, e as remissivas (USE reference) precisam ser alteradas do termo antigo para o termo novo. Sobre a atividade "exclusão de termos", sugere-se na Norma que termos pouco utilizados na atividade de indexação devem ser considerados candidatos para modificação ou exclusão. Menciona-se ainda que termos subutilizados podem ser substituídos por dois ou mais termos de maior especificidade. Na seção sobre manutenção, a Norma faz referência às seções sobre construção de vocabulários controlados, visto que algumas atividades da manutenção são exatamente iguais a algumas atividades de construção; por exemplo, na adição

de um termo, é preciso verificar as orientações sobre a escolha de termos (descrita na seção de construção de vocabulários controlados).

A Norma ISO 25964-1 (2011) aborda a construção e a manutenção de tesouros em uma mesma seção. Esse documento recomenda que a manutenção de tesouros deve ser ativa durante o tempo de vida do produto. Além disso, menciona que a necessidade de atualização deve nascer com o planejamento do tesouro. A Norma traz uma lista de diferentes cenários de atualização que podem ocorrer em um tesouro. Destaca-se que a necessidade de atualização geralmente foca na adição de termos, mas que os outros cenários da lista também podem ocorrer. Aborda ainda aspectos sobre a comunicação/disseminação das mudanças realizadas no vocabulário, e diz que a melhor forma de comunicar as alterações para os usuários é por meio do lançamento de uma nova versão.

Hjorland (2013) faz uma referência ao Falibilismo para argumentar sobre a necessidade de atualização do conhecimento presente nos SOCs. Falibilismo é uma doutrina filosófica que discorre sobre a possibilidade de falha das nossas afirmações baseadas no conhecimento científico. Assim, de acordo com essa corrente, o conhecimento é sempre provisório e passível de ser refutado com a descoberta de um novo conhecimento. É oportuno dizer que o Falibilismo não argumenta que o conhecimento científico é inacessível (HJORLAND, 2013). Para ilustrar um exemplo de conhecimento passível de falha, Hjorland (2013) cita o exemplo de Plutão, que até poucos anos era considerado um planeta, mas que devido à redefinição do conceito de planeta pela International Astronomical Union, foi rebaixado para a categoria de planeta anão.

A realização da manutenção e atualização dos SOCs depende de suas características e funções específicas, sendo que, em alguns casos, pode ser necessário o uso de um ou mais métodos. Entretanto alguns procedimentos podem ser comuns a todos os sistemas; por exemplo, a adição, a alteração e a exclusão de termos, assim como os procedimentos para controle dessas mudanças.

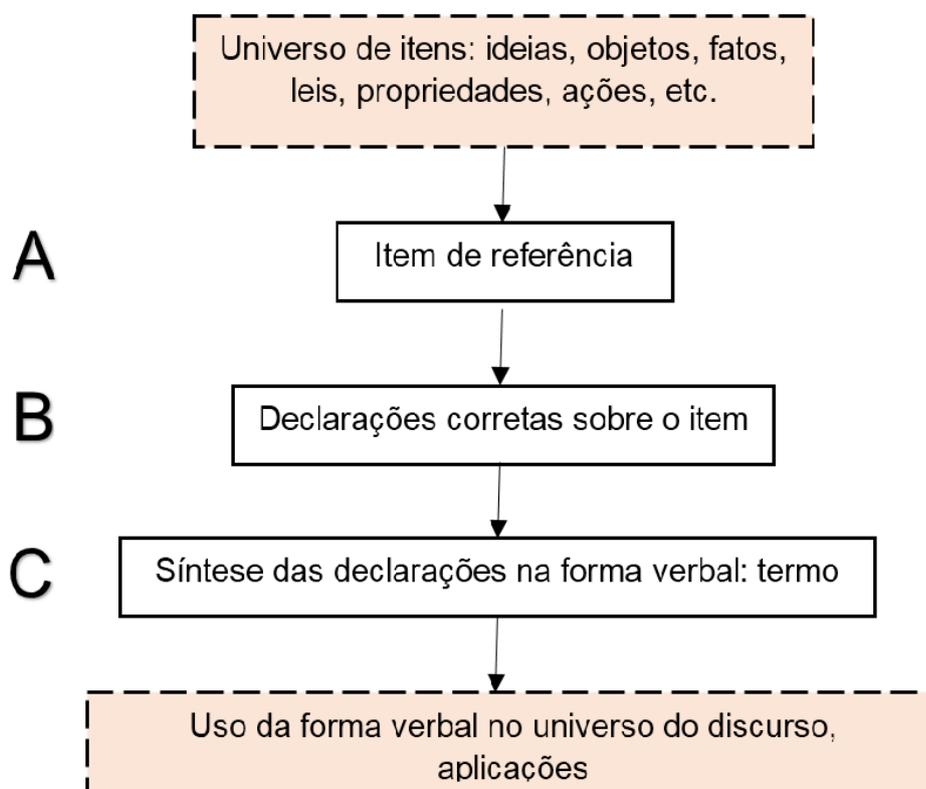
3.2 Uso da Teoria do Conceito no contexto de Sistemas de Organização do Conhecimento

A Teoria do Conceito foi concebida por Ingetraut Dahlberg e está sistematizada em um artigo de 1978 intitulado “*A referent-oriented, analytical concept theory for INTERCONCEPT*”. A ideia da teoria proposta por Dahlberg (1978, p. 142) é “clarear a natureza e estrutura dos conceitos, e respaldar procedimentos apropriados para seus estudos.”

Na concepção de Dahlberg (1978), conceito é uma unidade de conhecimento e sua formação se dá pela reunião e compilação de enunciados verdadeiros a respeito de determinado objeto. Esses enunciados são utilizados para definir a forma verbal (termo) que expressa o conceito. A autora propõe um modelo para a formação do conceito (representado na Figura 4), que inicia a partir de um universo de itens, que pode ser ideias, objetos, fatos, leis, propriedades, ações, etc. A partir desse universo seleciona-se um (A) item de referência (ou referente). Esse referente selecionado pode ser um único objeto (como uma árvore que está em determinado jardim) ou um conjunto de objetos considerados como uma unidade (como todas as árvores existentes). Uma vez escolhido o referente (item de referência), faz-se (B) declarações verdadeiras sobre ele, ou seja, declarações que possam ser verificadas por meio de alguma evidência ou por meio de um acordo intersubjetivo¹¹. Por último, sintetiza-se as declarações verdadeiras sobre o referente em uma (C) forma verbal (termo), visto que isso possibilita a utilização do conceito para comunicação (na fala ou escrita).

¹¹ Esse acordo intersubjetivo está associado aos conceitos que não encontram um referente no mundo físico, ou seja, que existe como um acordo entre mentes conscientes. Exemplo: um dragão que solta fogo.

Figura 4 – Modelo para construção de conceito

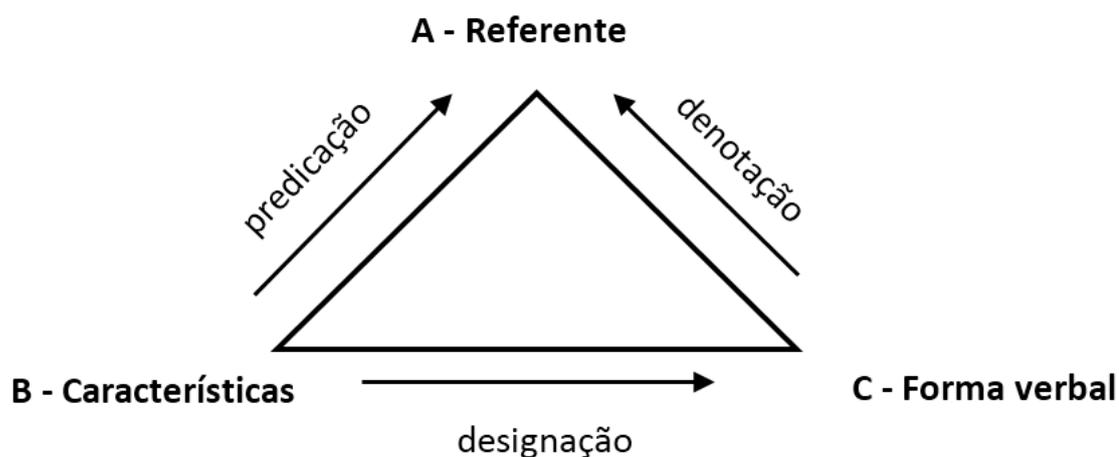


Fonte: Dahlberg (1978, p. 143).

Em síntese, os componentes para formação de conceitos são: (A) seu referente ou item de referência; (B) declarações corretas ou julgamentos sobre o referente; (C) forma verbal para representar o referente. De acordo com Dahlberg (1978), cada declaração correta sobre o referente gera um elemento de conhecimento (ou seja, característica, atributo) do referente e a soma de todas as declarações fornecem uma unidade de conhecimento (ou seja, o conceito).

Para simplificar os passos envolvidos na construção de conceitos, Dahlberg (1978) fala em três passos: (A) referencial, que versa sobre a identificação do item de referência no mundo físico ou mental; (B) predicacional, que trata das características que identificam o referente; (C) representacional, que trata da forma como o referente será comunicado, ou seja, o termo. A autora propõe então um triângulo (representado na Figura 5) para representar graficamente os componentes do conceito.

Figura 5 – Triângulo do Conceito



Fonte: Dahlberg (1978, p. 144)

A partir da Figura 5 fica claro o tipo de relação/atividade entre os componentes. Entre B e A é predicação; B e C é designação; C e A é denotação. (DAHLBERG, 1978).

Uma vez que se compreende como se constrói os conceitos, pode-se respaldar em suas características para identificação de conceitos de mesma natureza e assim formar categorias. Sobre a categorização, Dahlberg (1978) se respalda nas categorias aristotélicas e propõe quatro categorias que se desdobram em 12 subcategorias (sendo três subcategorias para cada categoria) para aglutinação de conceitos. O Quadro 3 apresenta as categorias e subcategorias da autora.

Quadro 3 – Categorias e suas respectivas subcategorias propostas por Dahlberg

Categorias	Subcategorias
Entidades	Princípios Objetos imateriais Objetos materiais
Propriedades	Quantidades Qualidades Relações
Atividades	Operações Estados Processos
Dimensões	Tempo Posição Espaço

Fonte: elaborado pelo autor.

A partir desse sistema de categorias pode-se sistematizar conceitos provenientes de qualquer campo do conhecimento e conforme Campos (2004) pode auxiliar na diminuição de erros lógicos na relação entre conceitos, por exemplo, sendo mais improvável confundir uma atividade com uma entidade por exemplo. Segundo Dahlberg (1978)

A determinação final dos referentes por meio da predicação pode, portanto, assumir a forma de uma "escada de características" ou "escada de conceitos", uma vez que características também são conceitos. Partindo de qualquer base e predicando os predicados até uma subcategoria de forma final, pode-se estabelecer a classe de forma para uma das características de um conceito. O procedimento pode ser ilustrado por qualquer referente selecionado aleatoriamente, como, por exemplo, um "jornal semanal". (DAHLBERG, 1978, p. 145. tradução própria¹²).

Em sua teoria, Dahlberg (1978) fala também da definição de conceitos. De acordo com ela, para que se utilize os conceitos e seus termos corretamente é preciso saber exatamente o que eles são. Sendo assim, a definição é o instrumento que contém a soma de características que representam a intensão do conceito.

Os princípios da Teoria do Conceito são utilizados no âmbito da BCI para desenvolvimento de SOCs, como tesouros, sistema de classificação e ontologias. Posto que a forma como possibilita a compreensão e sistematização dos conceitos e suas características, facilita o entendimento e as correlações entre essas unidades de conhecimento, permitindo assim o desenvolvimento de sistemas de conceitos.

Vários pesquisadores da BCI (CAMPOS, 2001, 2004; SILVA, 2008; MACULAN, LIMA, 2017; FERREIRA, 2020) reconhecem a importância e aplicabilidade dos conceitos na concepção de SOC. E é com base nesse respaldo que se percebe como as ontologias podem se beneficiar dos princípios da Teoria do Conceito, principalmente no estabelecimento de definições consistentes, que no contexto das ontologias podem contribuir para o estabelecimento de relações entre classes e subclasses e na configuração de axiomas.

¹² The final determination of referents through predication may thus take the form of a "ladder of characteristics" or "ladder of concepts", since characteristics are concepts too. Starting from any base and predicating the predicates up to an ultimate form subcategory one can establish the form class for one of the characteristics of a concept. The procedure can be illustrated by any randomly selected referent, such as, for example, a "weekly paper"

3.3 Ontologias

Segundo Smith e Welty (2001), o primeiro uso do termo "**ontologia**" foi realizado em 1613 por dois filósofos, Rudolf Gockel e Jacob Lorhard; e a primeira representação do termo em inglês, reportado no Oxford English Dictionary, foi no Dicionário de Bailey de 1721.

A ontologia já é objeto de estudo da Filosofia há séculos. Nesse campo, é tida com um ramo da Metafísica que trata da natureza do ser ou a essência das coisas. Ontologia, nesse aspecto, é uma disciplina filosófica que se preocupa com as categorias de coisas que existem na realidade (SILVA, 2008).

Ainda na concepção filosófica do termo "Ontologia", alguns autores (MARTÍNEZ-ÁVILA e FOX, 2015; ARP, SMITH e SPEAR, 2015; FARINELLI, 2017) argumentam que o conceito de ontologia pode ser perseguido de volta até Aristóteles e sua teoria realística das categorias e hierarquias, capturando a forma como o mundo é da forma que "é".

Saber a origem da temática "ontologia" é importante para entender os aspectos nos quais ela é utilizada atualmente no contexto da Ciência da Informação e Ciência da Computação. Ainda que a abordagem tenha um viés diferente da concepção de ontologia na Filosofia, a essência de entender e representar a realidade permanece.

Ontologia é uma disciplina da metafísica que estuda a realidade e os tipos de coisas que existem nela, mas ontologia enquanto artefato de representação não trata da realidade em si, mas de um modelo que visa representar os fenômenos sociais e físicos, e como qualquer outra representação, não é perfeita. Ontologias são modelos que visam se aproximar ao máximo do mundo real. Portanto, esses modelos são sempre passíveis de aperfeiçoamento, já que a nossa concepção do mundo se altera, bem como o próprio mundo físico e social.

Dessa forma, segundo Fonseca (2007) e outros autores, como Guarino (1998), é importante fazer a distinção entre a disciplina 'Ontologia' (no singular e com "O" maiúsculo) e 'ontologias' (no plural e com "o" minúsculo), artefatos para representação do conhecimento. As ontologias enquanto artefatos são objetos de estudo de outros campos, como Ciência da Computação, Inteligência Artificial, Medicina, Ciência da Informação.

O estudo de ontologias no âmbito da Inteligência Artificial e Ciência da Computação é anterior aos estudos na CI. Segundo Smith (2003), o interesse pela

temática iniciou-se na década de 1970 por pesquisadores de IA, que iniciavam os estudos sobre modelo de dados com base em ontologias. Somente após quase três décadas (final dos anos 1990), com o potencial de contribuir na organização da informação, é que começaram a surgir estudos de ontologias no contexto da CI. Vickery (1997) foi um dos primeiros pesquisadores da área a abordar a temática, juntamente como Soergel (1999) e, já nos anos 2000, Gilchrist (2003).

Há várias definições para ontologias de autores da Filosofia (SMITH, 1998; 2003), Ciência da Computação (GRUBER, 1993; STUDER *et al.*, 1998; GUARINO e GIARETTA, 1995), Ciência da Informação (SØERGEL, 1999; FONSECA, 2007; ALMEIDA, 2013). A mais recorrente na literatura é a definição de Gruber (1993), na qual ele diz que "ontologia é a especificação formal e explícita de uma conceitualização". A palavra "formal" apontada por Gruber quer dizer que a ontologia deve possuir um nível de formalidade que possa ser compreendida pelas máquinas; já a palavra "explícita" significa que todos os recursos empregados na ontologia devem ser claros; quanto à palavra "conceitualização", trata-se de um modelo abstrato de algum fenômeno do mundo por meio da identificação dos conceitos relevantes daquele fenômeno. Studer *et al.* (1998) completaram a definição com o complemento "compartilhada", que quer dizer que aquele conhecimento trazido pela ontologia deve ser consensual e aceito por um grupo de pessoas e não apenas por alguns indivíduos.

Na esfera deste trabalho, fundamenta-se principalmente nos autores Zeng (2008) e Soergel (1999), que contextualizam as ontologias como tipos de SOCs. Assim, ontologias são modelos da realidade representados de maneira formal, com base no consenso entre especialistas sobre os conceitos e relações entre esses conceitos, que são representativos de uma área.

3.2.1 Tipologia, componentes e linguagens da representação das ontologias

Quanto à tipologia, a literatura apresenta propostas diferentes para a classificação das ontologias, que pode ser quanto a: função, grau de formalismo, aplicação, estrutura e conteúdo (ALMEIDA; BAX, 2003). Porém a proposta de Guarino (1997) é uma das mais recorrentes na literatura. Ele propõe a existência de quatro tipos de ontologias, a saber: ontologias de alto-nível, de domínio, de tarefa e de aplicação.

- ontologias de alto nível (ou de fundamentação, como também é encontrado na literatura): descrevem conceitos gerais, como espaço, tempo, evento, matéria, ações, etc. Esses conceitos, tipicamente, são independentes de um problema particular ou domínio. Sendo assim, é bem razoável ter-se uma ontologia de alto-nível compartilhada por grandes comunidades de usuários. Exemplos desse tipo de ontologia são: *Basic Formal Ontology* (BFO), desenvolvida em 2002 por Barry Smith e Pierre Grenon (BFO, 2019); *Descriptive Ontology for Linguistics and Cognitive Engineering* (DOLCE), desenvolvida por Gangemi e colegas (2002); *Unified Foundational Ontology* (UFO), desenvolvida por Guizzardi e Wagner (2005);
- ontologias de domínio: descrevem o vocabulário genérico relacionado a um domínio. O objetivo é representar as entidades e os conceitos de um domínio científico específico. Isso pode ser feito por meio da especialização de conceitos introduzidos nas ontologias de alto nível, quando se parte de uma ontologia de alto nível ou pode ser criada a partir de conceitos do próprio domínio, sem a necessidade de uma ontologia de alto nível. Exemplos desse tipo de ontologia são inúmeros em áreas diversas do conhecimento, como: *Chemical Entities of Biological Interest* (CHEBI), na área Química e Biologia; *Agronomy Ontology* (AGRO), na Agronomia; *Environment Ontology* (ENVO), nas Ciências Ambientais; e inúmeras outras, que podem ser encontradas em repositórios de ontologias, como *OBO Foundry, Basel Register of Thesauri, Ontologies and Classification* (BARTOC). As ontologias de domínio podem ser utilizadas em distintas aplicações, mas uma característica fundamental de todas elas é que representam um consenso entre especialistas do domínio através da representação reduzida de uma realidade por meio de conceitos, atributos e das associações entre esses conceitos;
- ontologias de tarefa: possuem similaridades com as ontologias de domínio, mas objetivam descrever um vocabulário relacionado a uma tarefa ou atividade genérica, como diagnóstico ou compra, segundo Farinelli (2017);
- ontologias de aplicação: especializa conceitos tanto das ontologias de domínio como também das de tarefas, sendo que seus conceitos, geralmente, correspondem a papéis desempenhados por entidades do domínio enquanto executam uma atividade.

As ontologias são constituídas de componentes que podem apresentar nomenclaturas diferentes, a depender da área em que são desenvolvidas, da linguagem de representação e da ferramenta de desenvolvimento e de edição, assim como do tipo de formalismo lógico empregado (Lógica de Primeira Ordem, Lógica Descritiva, por exemplo). Porém há uma concordância entre os pesquisadores (GRUBER, 1993; GRUBER 1995; GUARINO e WELTY, 2000; ALMEIDA e BAX, 2003; GUARINO, OBERLE e STAAB, 2009; ARP, SMITH e SPEAR, 2015; FARINELLI, 2017; MUSTAPHA, AUFAURE e BAAZHAOUI-ZGHAL, 2006) quanto aos componentes gerais das ontologias que possibilitam a elas representarem o conhecimento. São eles:

- classes e subclasses: são coleções de elementos compostos de um conjunto de atributos semelhantes. Compõem as unidades básicas de uma ontologia e formam conceitos que determinam certo objeto. As classes representam os conceitos do domínio;
- propriedades/atributos: descrevem as características e/ou qualidades das classes e subclasses;
- Relações (relacionamento ou propriedade): são responsáveis pelas ligações semânticas entre os conceitos de um dado domínio. Há diversos tipos de relações, podendo segmentá-las em taxonômicas (relação é_um) e não taxonômicas, que se refere a todos os outros tipos de relações;
- axiomas ou restrições: são as regras declaradas sobre as relações. Permitem fazer inferências de conhecimento que não estão indicados nas taxonomias da ontologia. Os axiomas possibilitam que as interpretações possíveis do modelo representado na ontologia sejam restringidas e isso é importante para que não haja ambiguidade, tampouco interpretações que os desenvolvedores não desejem ou que não representem a realidade. Os axiomas possibilitam que a representação da ontologia se aproxime mais do modelo pretendido;
- instâncias (ou indivíduos): representam um determinado objeto de um conceito, ou seja, são os próprios dados da ontologia.

Além dos componentes gerais, podem existir componentes auxiliares (ou específicos) na estrutura das ontologias, como:

- cardinalidade entre classes: medida do número de ocorrências de um objeto (entidade) associado com o número de ocorrências em outro (FARINELLI, 2017).
- atributos das instâncias: são propriedades relevantes que descrevem as instâncias de uma entidade (FARINELLI, 2017).
- constantes: tipo de propriedade que sempre tem o mesmo valor (FARINELLI, 2017).
- propriedades das relações: são os atributos que descrevem a relação entre as propriedades lógicas da relação: reflexividade, transitividade, simetria.

Para expressar todos os componentes de forma que seja legível por computadores e para que seja utilizada em aplicações computacionais, precisa-se de uma linguagem (ou modelos) de representação do conhecimento, entre elas cita-se: RDF (Resource Description Framework), RDFS (Resource Description Framework Schema), OWL (Web Ontology Language).

Resource Description Framework (RDF) é um modelo para descrição de dados e se baseia em três componentes básicos: recurso (sujeito), propriedade (predicado) e valor (objeto). Por meio desses componentes pode-se descrever as relações entre objetos, conceitos, pessoas (que no contexto do modelo RDF são recursos). Cada descrição desses três elementos é conhecida como tripla, por exemplo: *Pelé* (Sujeito) *é* (predicado) *brasileiro* (objeto).

Já o Resource Description Framework Schema (RDFS) é uma extensão do RDF e possibilita a descrição de grupo de recursos e os relacionamentos entre eles. Assim, pode-se criar esquemas para relacionar os recursos, por exemplo: *Professor* (conceito ou classe) *leciona* (propriedade) *disciplina* (conceito ou classe). O modelo RDFS possui construtos que possibilitam descrever uma ontologia sem muita expressividade, já que é possível descrever a hierarquia entre classes (usando o construto `rdfs:subClassOf`), direção de um relacionamento por meio `rdfs:domain` e `rdfs:range`. Por exemplo: na tripla *pessoa – possui - nacionalidade*, o *domain* da relação *possui* é *pessoa* e o *range* é *nacionalidade*.

A OWL (Web Ontology Language) é uma linguagem computacional baseada em lógica e possibilita a representação de conhecimento rico e complexo sobre coisas e grupos de coisas, bem como as relações entre elas, (W3C, 2020). A linguagem possui uma série de construtos que permitem representar vários aspectos de como

as coisas funcionam na realidade de forma que os agentes de software consigam interpretar e inferir novos conhecimentos, a seguir alguns dos construtos:

- OWL:restriction permite restringir quais classes podem ser relacionadas por determinada relação.
- allValuesFrom (restrição universal) é um tipo de restrição declarada em uma propriedade (relacionamento) em relação a uma classe. Por exemplo: a classe Pessoa pode ter uma relação possui_filha e a restrição estaria relacionada à classe Mulher.
- someValuesFrom (restrição existencial) é um tipo de restrição declarada em uma propriedade (relacionamento) em relação a uma classe. Uma classe específica pode ter uma restrição sobre uma propriedade de que pelo menos um valor para essa propriedade seja de um determinado tipo. Por exemplo a classe Livro pode ter uma restrição na relação possui_escritor que declara que o objeto dessa relação (range) deve ser alguma instância da classe Escritor.
- inverseOf é uma característica de relações para descrever relacionamentos inversos. Exemplo: a relação possui_filho é inversa da relação possui_pais.
- transitiveProperty é uma característica de relações. Uma relação é transitiva se B pertence a A e C pertence a B, então C pertence a A. Exemplo: fruta cítrica é_uma fruta, laranja é_uma fruta cítrica, então, laranja é_uma fruta.
- symmetricProperty é uma característica de relações. Uma relação é simétrica quando é inversa dela mesma. Exemplo: Isaac é_parente_de Liz. A relação é_parente_de é simétrica, pois Liz também é_parente_de Isaac.
- functionalProperty é uma característica de relações para declarar um único valor. Exemplo: a relação é_primeiro_empregado só pode ser atribuído a uma instância.
- inverseFunctionalProperty é uma característica de relações para declarar relações com funcionalidade inversa. Se uma relação é inversa funcional, então o inverso da propriedade é funcional. Exemplo: a relação tem_número_de_CPF pode ser declarada com funcionalidade inversa. O inverso é a relação é_número_de_CPF_de tem no máximo um valor para qualquer indivíduo na classe de números de previdência social.

- `intersectionOf` é um construto para criar classes a partir da união classes existentes na ontologia. Exemplo: a classe “livros da loja” é resultado da interseção entre as classes “itens da loja” e “livro”.
- `disjointWith` é um axioma para declarar classes que não podem compartilhar instâncias. Exemplo: a classe homem e a classe mulher são disjuntas.

A seguir, na seção Engenharia Ontológica, apresenta-se a área que lida com todos os processos e atividades que envolvem as ontologias, desde o planejamento, passando pelo desenvolvimento, até a atualização e manutenção, depois de desenvolvidas.

3.4 Engenharia Ontológica

A área de Engenharia Ontológica surgiu da necessidade de transformar o processo de desenvolvimento de ontologias em um processo bem conduzido, com base em métodos, metodologias e princípios, uma vez que, até então, o processo era mais uma arte, em que ontologias eram construídas sem nenhum respaldo metodológico (SUÁREZ-FIGUEROA *et al.*, 2010).

Ainda de acordo com Suárez-Figueroa *et al.* (2010), a Engenharia Ontológica refere-se ao conjunto de atividades que lida com o processo de desenvolvimento de ontologias, o ciclo de vida das ontologias, os métodos e as metodologias para construí-las, assim como as ferramentas e as linguagens que dão suporte ao processo.

Quando se compara a Engenharia Ontológica com a Engenharia de Conhecimento e Engenharia de Software, há uma grande disparidade no que concerne à maturidade das áreas, principalmente com relação à última, que já possui normas consolidadas para desenvolvimento de sistemas. Assim, embora as ontologias não sejam softwares, a área de Engenharia Ontológica pode e tem aproveitado muitos processos, técnicas e princípios da Engenharia de Software, como defendem alguns autores (FERNÁNDEZ-LÓPEZ, 1999; SUÁREZ-FIGUEROA, 2010). A ideia de ciclo de vida, que se apresenta a seguir, foi aproveitada da Engenharia de Software.

3.4.1 Ciclo de vida das ontologias

O ciclo de vida das ontologias refere-se a todos os processos associados à construção e manutenção de ontologias, desde antes de sua criação (planejamento), passando pelo desenvolvimento (conceitualização, implementação, formalização) até chegar à fase de manutenção e atualização. Juntamente com essas atividades, existem outras que dão apoio à execução delas, como Aquisição de Conhecimento, Avaliação, Documentação, conhecidas como atividades de suporte. Objetiva-se, nesta seção, apresentar um panorama do ciclo de vida das ontologias, com suas várias etapas. Em seguida, o foco será nas atividades relacionadas à manutenção e atualização das ontologias, que é um dos interesses desta tese.

Apesar de não haver um consenso geral dos desenvolvedores (GRUNINGER e FOX, 1995; USCHOLD e KING, 1995; GÓMEZ-PEREZ, FERNANDEZ-LOPES e VICENTE, 1996; NOY e MCGUINESS, 2001; SURE, STAAB e STUDER, 2004; DE NICOLA, MISSIKOFF e NAVIGLI, 2005; SUARÉZ -FIGUEROA, 2010) das principais metodologias, percebe-se que as fases mais recorrentes dessas metodologias são: 1) **identificação do propósito da ontologia** (também denominado, nas distintas metodologias, como especificação, cenário de motivação, escopo da ontologia, *workflow* de requisitos, entre outras expressões que culminam no mesmo objetivo); 2) **aquisição de conhecimento** (identificação de termos relevantes do domínio, listagem de termos, *workflow* de análise); 3) **conceitualização** (criação de taxonomia, definição e organização de classes); 4) **formalização e implementação** (codificação, *workflow* de implementação); 5) **avaliação** (verificação, validação); 6) **documentação** (SILVA, 2008; MENDONÇA, 2015).

A **identificação do propósito da ontologia** refere-se à etapa em que são estabelecidas as atividades de planejamento que irão direcionar todo o restante do trabalho de desenvolvimento de ontologias. Assim, nessa etapa, define-se o escopo da ontologia, objetivo de desenvolvimento de determinada ontologia, grau de granularidade, quem serão os usuários. As questões de competência, que tratam das perguntas cujas respostas a ontologia pretende ter, são os meios adotados por grande parte das metodologias para auxiliar nessa etapa de identificação do propósito.

A **aquisição de conhecimento** é a etapa em que se obtém a matéria-prima para desenvolvimento das ontologias, que pode acontecer concomitantemente a outras etapas. Nessa etapa, o ontologista adquire conhecimento do domínio para

construção da ontologia. Existem algumas fontes em que o ontologista pode conseguir esse conhecimento; entre elas, estão livros, artigos, figuras, tabelas, manuais, outros SOCs (como glossários, tesouros, taxonomias) ou os especialistas do domínio. Diante dessas fontes, aplicam-se algumas técnicas para obtenção do conhecimento, tais como entrevistas estruturadas ou não estruturadas com especialistas do domínio, análise dos documentos, análise de SOCs que representam o mesmo domínio em que a ontologia está sendo construída. Vale evidenciar que muitas técnicas oriundas da Biblioteconomia e Ciência da Informação podem contribuir na realização dessa etapa, principalmente no que concerne à identificação de fontes de conhecimento para desenvolvimento das ontologias, assim como na extração da informação dessas fontes, na qual pode ser empregada a Análise de Assunto. A Análise de Assunto, de acordo com Naves (1996, p. 215) é o “processo de extrair conceitos que traduzam a essência de um documento”. Nesse sentido, as técnicas empregadas para analisar o assunto de um documento, podem auxiliar na identificação daqueles que contenham insumos para desenvolvimento ou manutenção de uma ontologia.

Na etapa de **conceitualização**, cria-se o modelo conceitual do domínio que se objetiva representar. Com o respaldo da atividade aquisição de conhecimento, cria-se uma lista de termos extraídos do domínio e inicia-se a modelagem por meio da criação de uma taxonomia, em que a atividade de categorização desempenha um papel fundamental. Um dos desafios nessa etapa é determinar qual termo melhor representa um conceito dentro do domínio de conhecimento.

Na etapa de **formalização e implementação**, o modelo conceitual desenvolvido na fase anterior, que até então estava representado em linguagem natural, é transformado em um modelo formal com o intuito de eliminar possíveis ambiguidades e, em seguida, implementado em uma linguagem legível por computador, como RDFS (*Resource Description Framework Schema*), OWL (*Ontology Web Language*). Salienta-se que algumas metodologias abordam essas etapas separadamente, ou seja, primeiro realizam a formalização e depois a implementam. Porém, na prática, percebe-se que essas etapas são realizadas ao mesmo tempo.

A fase de **avaliação** trata de analisar tecnicamente se a ontologia corresponde ao domínio que se pretende representar. A avaliação pode ser dividida em verificação e validação. Na verificação, analisa-se a ontologia com base em alguns critérios, como completude, corretude, consistência, concisão, expansibilidade. Na validação, testa-

se a ontologia desenvolvida com base em alguma tarefa para verificar se ela está modelando o domínio corretamente. Algumas metodologias propõem que a avaliação ocorra no final do processo de desenvolvimento de ontologias, já outras indicam que a avaliação deve ocorrer durante todo o processo, ou seja, após cada etapa, avalia-se o projeto de desenvolvimento da ontologia. O critério de completude possui relação intrínseca com esta pesquisa, pois está associado com a capacidade da ontologia de representar o conhecimento do domínio que se propõe a representar. A completude da ontologia pode ser verificada com base no seu desempenho nas respostas às questões de competência e por meio de outras técnicas da área de avaliação de ontologia.

A fase de **documentação** refere-se ao registro do que se realizou em cada uma das etapas de desenvolvimento da ontologia. É uma etapa de suma importância dentro do ciclo de vida das ontologias, pois corrobora para que outras pessoas possam reusar a ontologia, fazer atualizações e manutenções.

3.4.2 Processos e atividades relacionados à manutenção e atualização de ontologias

Diante do que foi exposto na seção anterior, neste estudo, considera-se que as ontologias possuem um ciclo de vida composto pelas seguintes partes: planejamento (ou pré-desenvolvimento, como prefere SUÁREZ-FIGUEROA, 2010), desenvolvimento, manutenção e atualização (ou pós-desenvolvimento). Concorda-se com Suarez-Figueroa (2010) que existem várias atividades e processos que dão suporte a essas três partes. São eles: aquisição de conhecimento (elicitação), documentação, gerenciamento da configuração, avaliação e verificação ontológica.

De forma mais específica, existem vários processos e atividades que fazem parte do ciclo de vida das ontologias. Suárez-Figueroa (2010) propõe um glossário na Metodologia NeOn para explicá-los. Existem muitos processos e atividades que estão relacionados com a fase de planejamento da ontologia, outros associam-se à fase de desenvolvimento e, por fim, existem aqueles que se relacionam com a fase de manutenção e atualização da ontologia. Este estudo tem uma conexão estreita com essa última fase. Portanto, a seguir, apresentam-se alguns dos processos que têm o objetivo de atualizar ou fazer manutenção em ontologias, a saber: atualização,

diagnóstico, versionamento¹³, personalização¹⁴, população¹⁵, modificação¹⁶, evolução, reúso, reengenharia, reestruturação enriquecimento (descrito na [seção 3.5](#)).

Suarez-Figueroa (2010) define atualização de ontologia como uma atividade que trata de pequenas alterações realizadas em uma ontologia e que não podem ser consideradas um upgrade, ou seja, que represente uma nova versão ou que reformule o modelo conceitual da ontologia. Interpreta-se que a autora quer dizer que essas alterações não têm impacto na estrutura da ontologia a ponto de ser necessário o lançamento de uma nova versão da ontologia. Nesse ponto, discorda-se da autora, já que nesta tese, considera-se que a atualização de ontologia é uma subárea dentro da engenharia ontológica que envolve vários outros processos e atividades que visam aprimorar a estrutura das ontologias com informações atualizadas do domínio de conhecimento.

O diagnóstico de ontologia, segundo Suarez-Figueroa (2010, p. 70) “refere-se à atividade de identificar partes da ontologia diretamente responsáveis pela incorreção e incompletude”. Em outras palavras, o diagnóstico refere-se à inspeção do conteúdo da ontologia para fins de identificação de erros de modelagem ou falta de representação de conhecimento de domínio (incompletude). A autora diz ainda que o diagnóstico de ontologia é acionado pela validação da ontologia (SUAREZ-FIGUEROA, 2010).

Ainda sobre o diagnóstico de ontologia, vale citar o estudo de Poveda-Villalón (2016), em que a autora desenvolveu um dos trabalhos mais completos para diagnóstico de ontologias com base em um catálogo de 41 *pitfalls* (erros na modelagem de ontologias). Alguns desses pitfalls são responsáveis pela incompletude ontológica, de acordo com Poveda-Villalón, Suárez-Figueroa e Gómez-Pérez (2010),

¹³**Versão de Ontologia** refere-se à atividade de lidar com alterações da ontologia, criando e gerenciando diferentes versões da ontologia (STOJANOVIC, 2004).

¹⁴**Personalização de Ontologia** refere-se à atividade de adaptar uma ontologia às necessidades de um usuário específico (SUAREZ-FIGUEROA, 2010).

¹⁵**População de Ontologia** (ou **Povoamento de Ontologia**) é uma atividade de aquisição de conhecimento que se baseia em métodos (semi) automáticos para transformar dados não estruturados (por exemplo, corpus textual), semiestruturados (por exemplo, folksonomias e páginas HTML) e fontes de dados estruturados (por exemplo, bancos de dados) em dados de instância (SUAREZ-FIGUEROA, 2010).

¹⁶**Modificação de Ontologia** refere-se à atividade de alterar a ontologia sem considerar a consistência (STOJANOVIC, 2004).

são eles: pitfall 4 (criar elementos ontológicos sem conexão), pitfall 9 (falta de informação básica da ontologia), pitfall 10 (falta da axioma de disjunção entre as classes), pitfall 11 (falta de definição de domain e range nas relações), pitfall 12 (falta de definição de relações equivalentes), pitfall 13 (falta de definição de relacionamentos inversos), pitfall 16 (falta de estabelecimento de classes primitivas e definidas).

A Figura 6 ilustra a relação entre alguns desses processos e atividades. Vale acentuar que a diferença entre alguns desses processos mencionados é sutil, o que leva a uma certa confusão e falta de consenso na literatura, assim como a uma sobreposição entre o fazer de um processo e o de outro. Além disso, como já apontado, a engenharia ontológica ainda não é uma área madura, portanto é natural que nem todos os conceitos estejam sedimentados, nem que haja um consenso sobre eles.

Figura 6 – Relação entre processos e atividades da engenharia ontológica



Fonte: Adaptado de Suárez-Figueroa (2010).

Evolução de Ontologia, de acordo com Stojanovic (2004), é um processo responsável pela adaptação adequada de uma ontologia às mudanças surgidas e à reprodução dessas mudanças em instrumentos dependentes da ontologia. Esse processo pode ser visto como uma consequência de diferentes outros processos e atividades durante o desenvolvimento de ontologia.

O reúso ontológico, conforme Katsumi e Gruninger (2017), é conceituado como um caso especial de *design*, que se refere à tarefa de pegar uma ontologia existente e manipulá-la de alguma forma com o propósito de satisfazer os requisitos da

ontologia em desenvolvimento. Na perspectiva de Lonsdale *et al.* (2010), reúso envolve a construção de novas ontologias por meio da máxima adoção de ontologias pré-utilizadas ou de componentes dessas ontologias. Suárez-Figueroa (2010) diz que o reúso de ontologia se refere ao processo de usar ontologias disponíveis para resolver diferentes problemas [a atualização e manutenção de ontologias pode ser um deles]. Percebe-se, com base nas definições apresentadas, que elas são consensuais e se completam.

Reengenharia ontológica, segundo Suárez-Figueroa (2010), diz respeito ao processo de recuperar e transformar um modelo conceitual de uma ontologia implementada numa nova ontologia mais correta e com o modelo conceitual mais completo. O processo de Reengenharia Ontológica pode ser visto como algo mais abrangente, que envolve os seguintes subprocessos: Engenharia Ontológica Reversa, Engenharia Ontológica Uniforme (*forward*) e Reestruturação Ontológica.

O processo de Reestruturação Ontológica tem como objetivo corrigir e reorganizar o conhecimento contido em um modelo conceitual inicial, além de detectar a falta de conhecimento no modelo. Esse processo possui as fases: análise e síntese. Na fase de análise, o propósito é avaliar a ontologia tecnicamente para checar se a hierarquia da ontologia, suas classes, instâncias, relações e funções estão completas, consistentes (ou seja, não há contradições na ontologia no que se refere às fontes de informação utilizadas), concisas (ou seja, não existem redundâncias explícitas ou implícitas) e sintaticamente correta. A fase de síntese procura corrigir as falhas encontradas na fase anterior e documentar quaisquer são alterações realizadas na ontologia (GÓMEZ-PÉREZ, ROJAS-AMAYA, 1999). Segundo a classificação de Suárez-Figueroa (2010), a Reestruturação Ontológica possui três atividades, a saber: Modularização Ontológica¹⁷, Podagem (*prunning*) Ontológica¹⁸ e Enriquecimento Ontológico (ou como tratar-se-á nesta tese, Enriquecimento de Ontologias).

¹⁷**Modularização de Ontologia** refere-se à atividade de identificar um ou mais módulos em uma ontologia com o objetivo de oferecer suporte à reutilização ou manutenção (SUAREZ-FIGUEROA, 2010).

¹⁸**Poda de Ontologia** refere-se à atividade de descartar estruturas conceituais de uma determinada ontologia que não são mais relevantes (SUAREZ-FIGUEROA, 2010).

3.5 Enriquecimento de ontologias

Enriquecimento de ontologia é um assunto emergente na área de Engenharia Ontológica, portanto ainda existe pouco embasamento teórico sobre esse processo. Há também uma certa confusão terminológica, uma vez que o termo EO ainda não está totalmente sedimentado.

Na literatura, há um uso indiscriminado de termos para definir a atividade de enriquecimento de ontologia ou atividades que têm relação estreita. Alguns dos termos são: expansão de ontologia (*ontology expansion*¹⁹), ampliação de ontologia (*ontology enlarge*), completude da ontologia (*ontology completeness*), melhoria de ontologia (*ontology improvement*), extensão de ontologia (*ontology extension*²⁰), refinamento de ontologia (*ontology refinement*), especialização de ontologia (*ontology specialization*²¹). Vale ressaltar que alguns termos podem ser usados como sinônimos de enriquecimento (expansão, extensão, alargamento); já outros tratam de processos diferentes (evolução e refinamento, por exemplo). No âmbito deste estudo, utilizaremos o termo “Enriquecimento de Ontologia” em vez de qualquer outro, pois é o termo mais comumente utilizado na literatura.

A literatura mostra que os estudos sobre EO são, em sua maioria, provenientes da área de Ciência da Computação, Inteligência Artificial e Linguística. A existência de estudos na área de Ciência da Informação é ínfima.

Petasis *et al.* (2011) define EO como a atividade de expandir uma ontologia com conceitos, relações e regras adicionais, inserindo-os na posição correta na ontologia. O autor diz ainda que é a atividade responsável por ampliar o conhecimento prévio em determinado domínio. Dessa forma, o processo de EO só faz sentido se a ontologia já existe no domínio.

Já Faatz e Steinmetz (2002) definem EO como a operação baseada em um conjunto de fórmulas que tem o propósito de incluir novos conceitos, propriedades e novas instâncias em ontologias existentes.

¹⁹**Expansão de ontologia** refere-se à atividade de facilitar a modificação da ontologia preservando a sua consistência (SUÁREZ-FIGUEROA, 2010).

²⁰**Extensão de Ontologia** é uma atividade do Enriquecimento de Ontologia responsável por alongar a ontologia em largura (SUÁREZ-FIGUEROA, 2010).

²¹**Especialização de Ontologia** é uma atividade do processo de Enriquecimento Ontológico que visa estender a ontologia em profundidade, ou seja, especificar os conceitos da ontologia.

Guerram e Mellal (2018) dizem que EO é um dos objetivos importantes do processo de Aprendizagem de Ontologia e consiste em adicionar automaticamente novos conceitos e relações a uma ontologia de domínio construída previamente. Eles ainda destacam o fato de essas novas inclusões serem colocadas em locais relevantes (ou certos) na ontologia.

As definições apresentadas por Petasis *et al.* (2011), Faatz e Steinmetz (2002) e Guerram e Mellal (2018) se completam, mesmo que algumas não mencionem alguns componentes das ontologias nos quais o enriquecimento pode acontecer. Porém todas as definições abordam EO em uma mesma perspectiva, a de expandir uma ontologia já existente, seja por meio de novos conceitos, seja por meio de relações, propriedades, axiomas. Com relação à inclusão de novas instâncias, existe uma divergência na literatura, haja vista que a atividade de inclusão de novas instâncias é conhecida como “população ou povoamento de ontologia”. Assim, neste estudo, seguir-se-á o viés que distingue Enriquecimento de Ontologia de População de Ontologia.

Na nossa perspectiva, em complemento às definições apresentadas no início desta seção, EO é um processo, manual ou semiautomático que dá suporte à área de manutenção e atualização de ontologia e é responsável pela expansão e especialização da estrutura da ontologia, por meio da inclusão de novos conceitos e seus símbolos linguísticos alternativos (sinônimos, por exemplo), relações (taxonômicas e não taxonômicas) e suas propriedades restritivas e axiomas; que devem ser incluídos nos locais corretos na ontologia. Seguindo a analogia da “camada de bolo²²”, o Enriquecimento seria os preenchimentos que podem ser feitos no bolo (organizar as fatias) e os complementos que podem ser adicionados a ele sem a necessidade de refazê-lo, ou seja, a inclusão de confetes, chantilly, pasta americana, entre outros acessórios que tornam o bolo mais atrativo e útil para outros contextos.

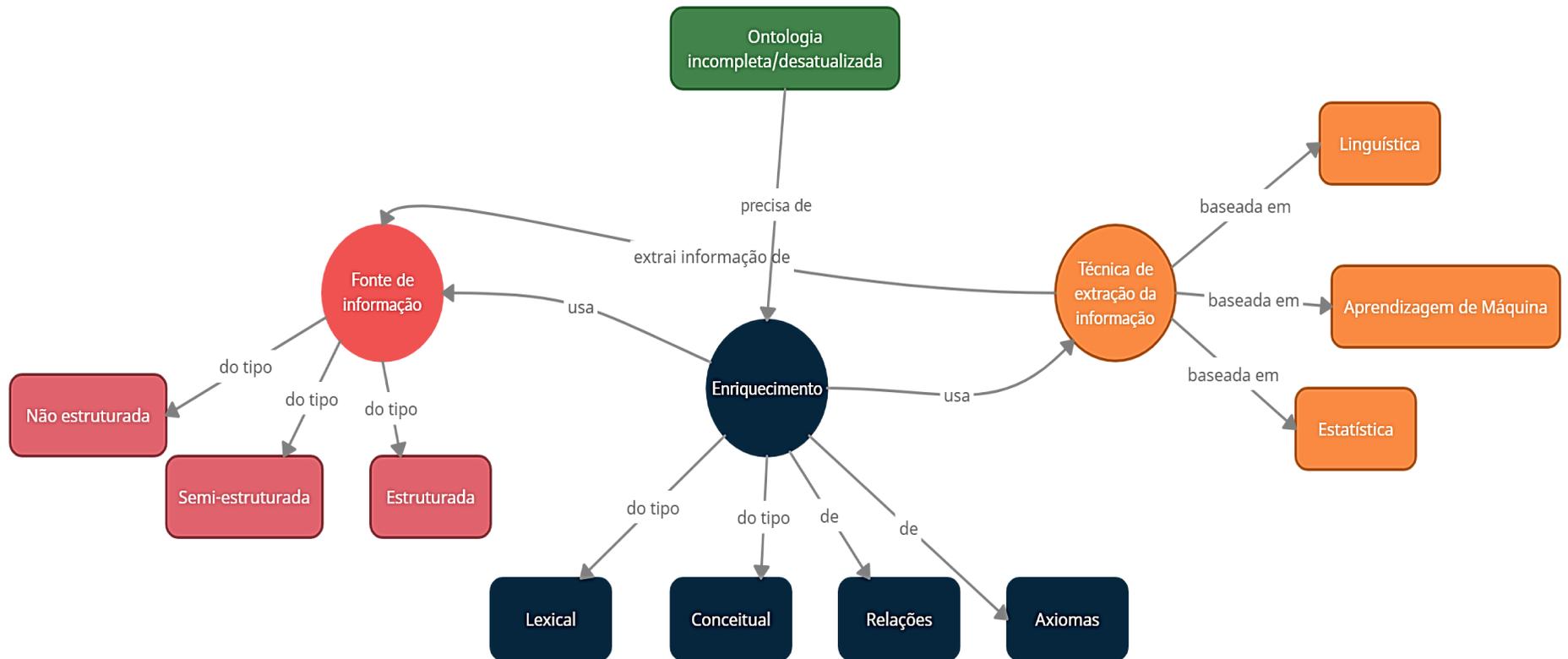
Suárez-Figueroa (2010) associa duas atividades ao processo de Enriquecimento de Ontologia, que são: (1) Extensão da Ontologia é a atividade de alongar a ontologia em largura (horizontalmente), ou seja, trata-se de ampliar a estrutura da ontologia com a inserção de conceitos amplos, que podem configurar uma nova classe ou subclasse na ontologia; (2) Especialização da Ontologia é a

²² Essa analogia está presente nas camadas propostas para a Web Semântica, bem como no contexto da temática Aprendizagem de Ontologias.

atividade de estender a ontologia em profundidade (verticalmente), ou seja, inclusão de conceitos mais específicos.

A Figura 7 apresenta um mapa conceitual que explica simplificadaamente o funcionamento do processo de Enriquecimento de Ontologia. Inicia-se pela ontologia que precisa ser enriquecida; em seguida, analisa(m)-se o(s) tipo(s) de enriquecimento(s) necessário(s), depois aborda(m)-se a(s) técnica(s) de enriquecimento e fontes de informação que serão utilizadas para enriquecer a ontologia.

Figura 7 – Mapa conceitual sobre o funcionamento do processo de enriquecimento



Fonte: elaborada pelo autor.

3.5.1 Relação do processo de enriquecimento de ontologia com a aprendizagem de ontologia

A temática Aprendizagem de Ontologia (AO) surgiu como uma das soluções para o gargalo da Aquisição de Conhecimento para desenvolvimento de ontologias (SHAMSFARD, BARFOROUSH, 2004), posto que o processo de Aquisição de Conhecimento é desgastante e consumidor de tempo. Assim a área de AO propõe que a aquisição ou descoberta de conhecimento seja uma atividade (semi) ou automática (DRUMOND, GIRARDI, 2008), visando à agilidade na construção de ontologias e à redução do custo de desenvolvimento de ontologias.

Gómez-Pérez e Manzano-Macho (2004) explicam que AO trata da aplicação de uma série de técnicas para construir, ampliar, adaptar e melhorar ontologias usando conhecimento e fontes de informação heterogêneos.

Guerram e Mellal (2018), na mesma linha de Petasis *et al.* (2011), dizem que a Aprendizagem de Ontologia é uma área de pesquisa vasta que inclui Enriquecimento de Ontologia, População de Ontologia e Resolução de Inconsistência. Ainda de acordo com esses autores, o desenvolvimento e a manutenção de ontologias é uma tarefa árdua que envolve diretamente a aquisição de conhecimento.

Discorda-se da afirmação dos autores que classificam Enriquecimento de Ontologia como um tipo de atividade da Aprendizagem de Ontologia pelas seguintes razões: (1) a própria definição de Aprendizagem de Ontologia, que diz que é um processo semi e automático. Ora, o Enriquecimento de Ontologias pode ser totalmente manual, assim, por definição, EO não pode ser um tipo de AO; (2) Enriquecimento de Ontologia visa à ampliação da ontologia com novos conceitos, relações, axiomas. Assim, difere-se de Aprendizagem de Ontologia, que foca na aquisição de conhecimento por métodos semi ou automáticos; (3) a aproximação de AO com a Aquisição de Conhecimento e EO está mais conectada com a temática manutenção e atualização de ontologia. Assim, neste estudo, segue-se a mesma linha de Suárez-Figueroa (2010), que trata Enriquecimento de Ontologia como um processo ligado à Reestruturação Ontológica e, conseqüentemente, à Reengenharia Ontológica, conforme Figura 6.

Embora discorda-se da classificação de Petasis *et al.* (2011) e Guerram e Mellal (2018), percebe-se na literatura uma forte influência dos métodos de Aprendizagem de Ontologia nos métodos de Enriquecimento de Ontologia, visto que o processo de

EO também pode utilizar as mesmas técnicas empregadas para aprendizagem ontológica. Resumidamente, declara-se que o *modus operandi* do último se apropria do primeiro. Por essa razão, decidiu-se abordar as relações entre as temáticas em uma seção separada, posto que o entendimento da temática Aprendizagem de Ontologia é relevante para a compreensão de Enriquecimento. Além disso, como dizem Valarakos *et al.* (2001, p. 2, tradução própria²³):

Aprendizagem de Ontologia é uma área de pesquisa pertencente à Engenharia Ontológica que tem o propósito de reduzir o máximo possível o esforço humano nas tarefas desgastantes, factíveis de erros e consumidores de tempo, tais como construção, refinamento, enriquecimento e manutenção de ontologias através de técnicas de aprendizagem de máquinas. (VALARAKOS *et al.*, 2001, p. 2, tradução própria)

Uma vez conceituado, abordam-se os requisitos básicos para que a Aprendizagem de Ontologia ocorra. Conforme Drumond e Girardi (2008), a realização da aprendizagem se baseia na disponibilidade de conhecimento preexistente e o tipo de fonte de onde esse conhecimento é extraído. Assim, segundo esses autores, existem três tipos diferentes de entrada de dados para o processo de AO:

1. **dados estruturados** – o conhecimento é aprendido a partir de informação estrutural, em que os dados já possuem uma certa organização. Assim são extraídos pedaços de informações para a ontologia. O problema central desse tipo de entrada de dados pelo processo de AO é determinar quais partes da fonte de informação podem prover conhecimento relevante. Exemplos de fontes de informação são: esquemas de bases de dados, ontologias existentes e base de conhecimento;
2. **dados semiestruturados** – o conhecimento é aprendido a partir de fontes que não possuem a informação totalmente estruturada, como dicionários, bases de dados. Ex.: Base léxica Word Net;
3. **dados não estruturados** – descobre-se o conhecimento por meio de textos em linguagem natural. Ex. artigos em pdf, livros, etc.

Cada tipo de entrada de dados tem suas vantagens e desvantagens. Os dados estruturados facilitam a extração de conhecimento, que pode ser usado para criação

²³ Ontology learning is a research area in the context of ontology engineering that aims to reduce, as much as possible, the human effort into the labor-intensive, error-prone and time consuming process of ontology building, refinement, enrichment and maintenance by means of machine learning techniques. (VALARAKOS *et al.*, 2001, p. 2)

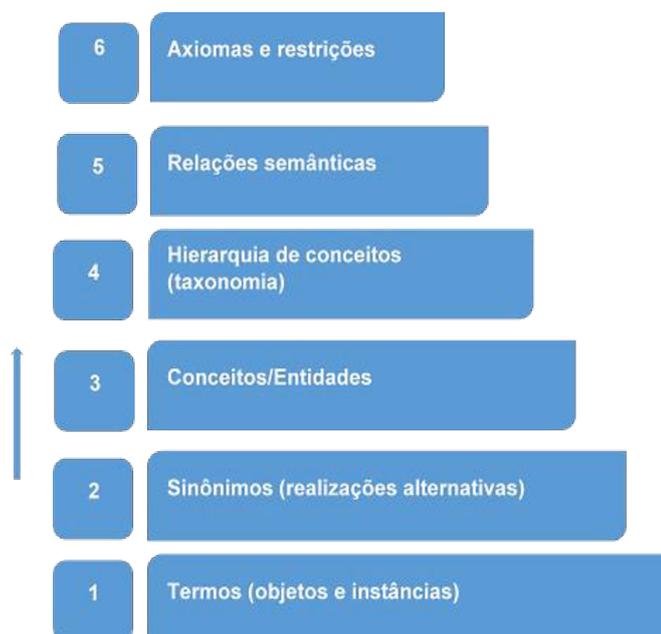
de novas relações semânticas; por outro lado, nem sempre existe uma fonte de dados estruturados em determinado domínio. As fontes com dados semiestruturados também nem sempre estão disponíveis em todos os domínios, porém têm apresentado bons resultados para Aprendizagem de Ontologia (BENZ, 2007). Por fim, as fontes de dados não estruturados são as mais comuns e fáceis de encontrar em qualquer domínio do conhecimento, porém não têm mostrado bons resultados na criação de recursos semânticos numa maneira completamente automatizada.

Cabe um adendo com relação à estruturação da informação. Aqui entende-se que informação estruturada refere-se às informações que estão em alguma linguagem que softwares possam processar. Se o documento de texto está estruturado com título, seções numeradas, etc, isso não configura informação estruturada no contexto da AO.

Dados não estruturados é o formato mais disponível para aplicação da AO. Documentos não estruturados consistem em textos em linguagem natural em formatos variados, como .doc, .docx, .odt, .rtf, .pdf, etc.

O processo de AO pode ser dividido em seis camadas, que formam a “camada de bolo”, de acordo com Buitelaar, Cimiano e Magnini (2005). Essas camadas estão descritas na Figura 8. Cada camada refere-se a componentes das ontologias, nos quais o processo de AO pode ser aplicado. Assim, existem métodos que englobam apenas uma ou duas das camadas (GOMÉZ-MORENO, MESTRE-MESTRE, 2017; GUERRAM, MELLAL, 2008; HASHIMY, KULATHURAMAIYER, 2013), enquanto outros abordam três ou mais camadas (BENDAOU, TOUSSAINT E NAPOLI, 2008; BARBUR, BLAGA, GROZA, 2011).

Figura 8 – Camadas para o processo de Aprendizagem de Ontologia



Fonte: Adaptado de Buitelaar, Cimiano e Magnini (2005).

A analogia da “camada de bolo” é interessante, pois uma camada suporta a outra, assim como em um bolo. Uma camada se apoia no resultado obtido na camada anterior, de acordo com Schiessl e Bräscher (2011). Embora não configure um padrão para o processo de AO, a analogia das camadas propostas pelos Buitelaar, Cimiano e Magnini (2005) foram bem aceitas na área, sendo que vários trabalhos (SCHIESSL e BRÄSCHER, 2011; PETASIS *et al.*, 2011; DRUMOND e GIRARDI, 2008; ZAHRA *et al.*, 2014) citam a figura; alguns com pequenas alterações, principalmente de ordem terminológica, que não impactam na estrutura original.

3.5.2 Tipos e fontes de Enriquecimento de Ontologia

Em busca de uma categorização dos métodos de EO, quanto ao grau de automatização, pode-se caracterizá-los em automáticos, semiautomáticos e manuais (geralmente realizado com o auxílio de um especialista de domínio). Na literatura, são poucos os métodos totalmente automático, sendo mais comuns os métodos semiautomáticos. Ainda nesse assunto, Petasis *et al.* (2011, p. 142) dizem que “um sistema totalmente automatizado é obviamente desejável, mas nem sempre é possível, especialmente com tarefas relacionadas ao enriquecimento ontológico”.

Nesta pesquisa, optou-se pela categorização dos métodos de acordo com o tipo de componente da ontologia que será enriquecido, ou seja, de acordo com o tipo de enriquecimento. Outra possibilidade de categorizar os métodos é de acordo com o tipo de fonte de informação de onde se extraem as informações para enriquecimento. Essas fontes podem ser divididas em estruturadas, semiestruturadas e não estruturadas.

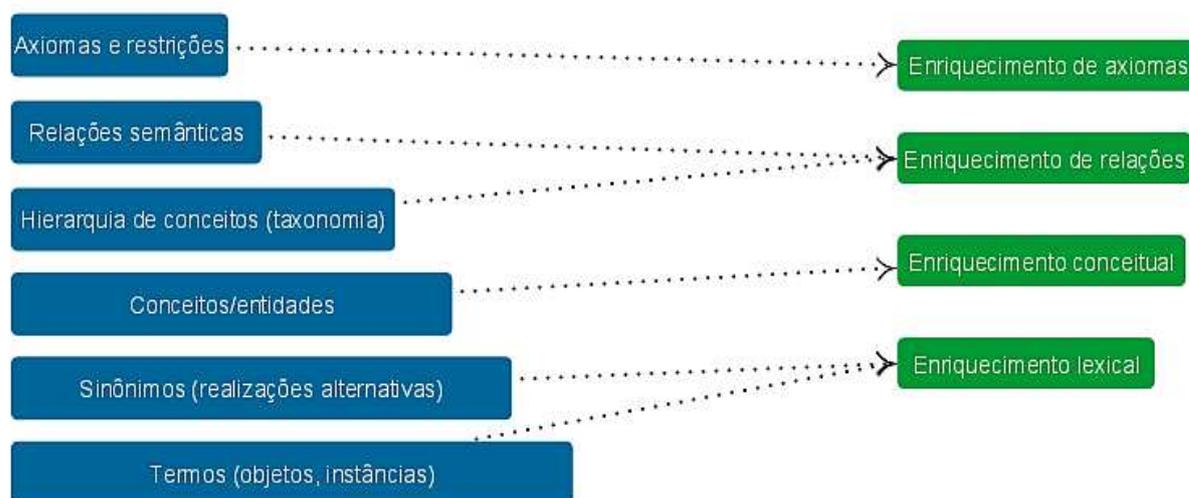
No que tange ao tipo de conhecimento adquirido para enriquecer as ontologias, pode-se dividir os métodos em:

- Enriquecimento lexical: refere-se à identificação de termos, sinônimos e realizações linguísticas alternativas para determinado conceito, que nas ontologias são inseridos como propriedades de anotação;
- Enriquecimento conceitual: trata da descoberta de novos conceitos/entidades, que são inseridos nas ontologias como classes e subclasses;
- Enriquecimento de relações (taxonômicas e não taxonômicas): refere-se à identificação de relacionamentos taxonômicos²⁴ e outros relacionamentos semânticos (ou não taxonômicos) para inclusão na ontologia;
- Enriquecimento de axiomas: trata da inclusão de novos axiomas e/ou restrições na estrutura da ontologia com fins de geração de novos fatos ou inferência de conhecimento.

Uma das contribuições deste trabalho é a tentativa de unificação da nomenclatura para caracterizar os tipos de enriquecimentos de ontologias existentes na literatura. Ainda que esteja claro na literatura que cada um desses tipos de enriquecimento (lexical, conceitual, de relações e de axiomas) esteja associado a um ou mais componentes das ontologias (ou às camadas de bolo da Aprendizagem de Ontologia), como pode ser constatado na Figura 9, essa terminologia ainda não está sedimentada na área de engenharia ontológica.

²⁴ O que está sendo denominado aqui como enriquecimento de relações taxonômicas está intrinsecamente conectado com o enriquecimento conceitual, visto que a própria identificação de classes e subclasses revela a relação taxonômica entre elas. Optou-se por distinguir os dois tipos de enriquecimento (enriquecimento conceitual e enriquecimento de relações taxonômicas), pois no conceitual o objetivo é identificar conceitos (que representam classes e subclasses na ontologia) sem se preocupar com a relação entre eles. A partir do momento que se identifica que determinados conceitos possuem relação de gênero-espécie ou de parte-todo, estar-se-á realizando o enriquecimento de relações taxonômicas.

Figura 9 – Relação entre o tipo de enriquecimento e componentes de ontologia



Fonte: Adaptado de Buitelaar, Cimiano e Magnini (2005).

As fontes de onde extrai-se o conhecimento para enriquecer ontologias são diversas, podem ser especialistas de domínio, artigos e textos científicos do domínio, Sistemas de Organização do Conhecimento, como tesouros, ontologias, taxonomias, dicionários, glossários. Em síntese, todos os materiais que são passíveis de insumos para a obtenção de informação para a realização do processo de enriquecimento.

3.5.3 Fases do Enriquecimento de Ontologia

As fases do processo de EO estão associadas às “camadas de bolo” da Aprendizagem de Ontologias (AO), haja vista que uma das poucas diferenças entre as áreas apontadas na literatura é que alguns autores advogam que EO é um subprocesso da área mais ampla, AO. Ademais AO está associada com a construção de ontologias de forma (semi) automática, e EO trabalha com ontologias já desenvolvidas.

Petasis *et al.* (2011) defende que as etapas de Enriquecimento de Ontologias referem-se a todas as “camadas de bolo” da Aprendizagem de Ontologias, diferente do processo de Povoamento de Ontologia, que se refere apenas às camadas “Termos” e “Sinônimos”. Porém isso não é um consenso, haja vista que alguns autores dizem que o EO está relacionado somente às camadas 3, 4, 5 e 6.

No âmbito deste estudo, concorda-se com aqueles que defendem as etapas de EO como sendo aplicadas a todas as “camadas de bolo” (com exceção da inclusão de instâncias), descritas na [seção 3.5.1](#). Assim, como apontado por Petasis *et al.* (2011), a abordagem principal adotada pela maioria dos métodos de EO se baseia nas seguintes fases:

- 1) identificação dos objetos e de suas realizações alternativas (sinônimos);
- 2) identificação do conceito;
- 3) identificação das relações;
- 4) aquisição de axiomas e restrições;

Na próxima subseção, apresentam-se as técnicas mais utilizadas para a realização do processo de enriquecimento de ontologias, posto que o processo de enriquecimento está invariavelmente associado a uma ou mais técnicas para extração de informação.

3.5.4 Técnicas de extração da informação usadas para Enriquecimento de Ontologia

As técnicas utilizadas para EO se baseiam principalmente no uso de Processamento de Linguagem Natural (PLN), métodos estatísticos e algoritmos de *Machine Learning*. A seguir, explica-se cada uma das técnicas utilizadas em cada categoria.

As técnicas PLN são utilizadas para extrair termos, relações entre termos de um corpus textual e, a partir desses componentes, podem-se propor novos termos, conceitos e relações nas ontologias. Entre as principais técnicas identificadas na literatura, estão:

- tokenização: dada uma sequência de caracteres e uma unidade de documento definida, a tokenização é a tarefa de cortá-la em pedaços, chamadas de *tokens*, talvez ao mesmo tempo eliminando alguns caracteres, como pontuação (STANFORD, 2019).
- análise (*Parsing*): é análise automática de uma frase em relação à sua estrutura sintática.
- análise Morfossintática: consiste em identificar palavras ou expressões isoladas numa frase, sendo este processo feito com base em delimitadores,

como a pontuação e os espaços em branco. As palavras identificadas são classificadas de acordo com a sua categoria gramatical.

- radicalização (*Stemming*): geralmente se refere a um processo heurístico bruto que corta o final das palavras, e geralmente inclui a remoção de afixos derivativos (STANFORD, 2019).
- lematização (*Lemmatization*): é a análise de vocabulário e morfologia das palavras, normalmente com o objetivo de remover apenas terminações flexionadas e retornar a forma básica ou de dicionário de uma palavra, conhecida como lema (STANFORD, 2019).
- *Part of Speech (POS) Tagging*: também chamado de marcação gramatical ou desambiguação de categoria de palavra, é o processo de marcar uma palavra em um texto (*corpus*) como correspondente a uma parte específica da fala, com base em sua definição e seu contexto.
- reconhecimento de Nome de Entidade (*Named Entity Recognition – NER*): também conhecida como identificação de entidade, *chunking* de entidade e extração de entidade, é uma subtarefa de extração de informações, que procura localizar e classificar as menções de entidades nomeadas em texto não estruturado em categorias predefinidas, como nomes de pessoas, organizações, locais, códigos médicos, expressões de tempo, quantidades, valores monetários, percentagens, etc.
- remoção de stopwords: refere-se à exclusão de palavras semanticamente insignificativas, tais como artigos e preposições.
- clusterização: conjunto de técnicas que trata do agrupamento de dados de acordo com o grau de similaridade.

As técnicas estatísticas servem para calcular a possibilidade de associação entre termos com base no número de vezes que aparecem juntas em um *corpus*. Algumas técnicas são: modelo de distribuição de probabilidade (Modelo n-grama); frequência de termos; frequência inversa de termos (TF-IDF).

As técnicas relacionadas à Aprendizagem de Máquina (*Machine Learning*) se baseiam em algoritmos computacionais que visam desempenhar tarefas específicas com o propósito de coletar conhecimento que possa ser útil para o enriquecimento de componentes das ontologias.

4 METODOLOGIA

Neste capítulo, são apresentadas as características da pesquisa, a delimitação do universo, o método, assim como os procedimentos metodológicos que a embasam.

4.1 Caracterização da pesquisa

Gil (2008) apresenta algumas orientações e características dos diversos tipos de pesquisa. Seguindo o viés do autor, essa pesquisa se caracteriza como aplicada, quanto à sua natureza, pois objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos, ou seja, tem o propósito de desenvolver uma metodologia para enriquecimento de ontologias de domínio para resolver o problema de ontologias desatualizadas, incompletas e que não representam o domínio adequadamente, além de sistematizar o processo de enriquecimento, auxiliando ontologistas na aplicação de tal processo. No que tange aos objetivos, caracteriza-se como uma pesquisa exploratória, pois tem a finalidade de apresentar informações sobre o assunto e entendê-lo melhor. Por fim, com relação aos procedimentos metodológicos, qualifica-se como uma pesquisa bibliográfica, pois o desenvolvimento da metodologia de enriquecimento respaldou-se em insumos da literatura.

4.2 Delimitação do universo da pesquisa

Segundo Marconi e Lakatos (2016), pode-se delimitar o universo de uma pesquisa a partir de distintas formas. Nesta investigação, optou-se por delimitar o universo de acordo com o objeto de estudo, campo de investigação e *corpus* da pesquisa. Assim o objeto de estudo refere-se ao processo de enriquecimento de ontologias. A ênfase nesse processo se deve à incipiência de estudos da temática e de metodologias abrangentes que sejam passíveis de aplicação em mais de um domínio.

A segunda delimitação trata-se do campo de investigação, a saber, ontologias de domínio. O foco em ontologias de domínio em vez de em outros tipos de ontologias se deve à existência de um grande número desses instrumentos desenvolvidos em

vários campos do conhecimento, sendo que muitas dessas ontologias não passam por nenhum processo de manutenção e atualização.

A terceira e última delimitação desta investigação trata-se do corpus da pesquisa, que são os 29 estudos selecionados na revisão de literatura, que se referem a pesquisas que aplicaram o processo de enriquecimento em ontologias de domínio. Como complemento, fazem parte do corpus as Normas ISO 25.964 parte1, ANSI/NISO Z39.19 e o capítulo sobre atualização e manutenção de vocabulários controlados do livro *Indexing Languages and Thesauri: construction and maintenance*, do professor Dagobert Soergel.

4.3 Método

O método dessa investigação é o Indutivo, tendo em vista o levantamento e a análise dos estudos identificados na revisão de literatura. Em complemento, emprega-se a Análise de Conteúdo (AC) proposta por Bardin (2011) para atingir os objetivos desta pesquisa. Acredita-se que a AC fornece os subsídios necessários para exploração do material levantado na revisão de literatura.

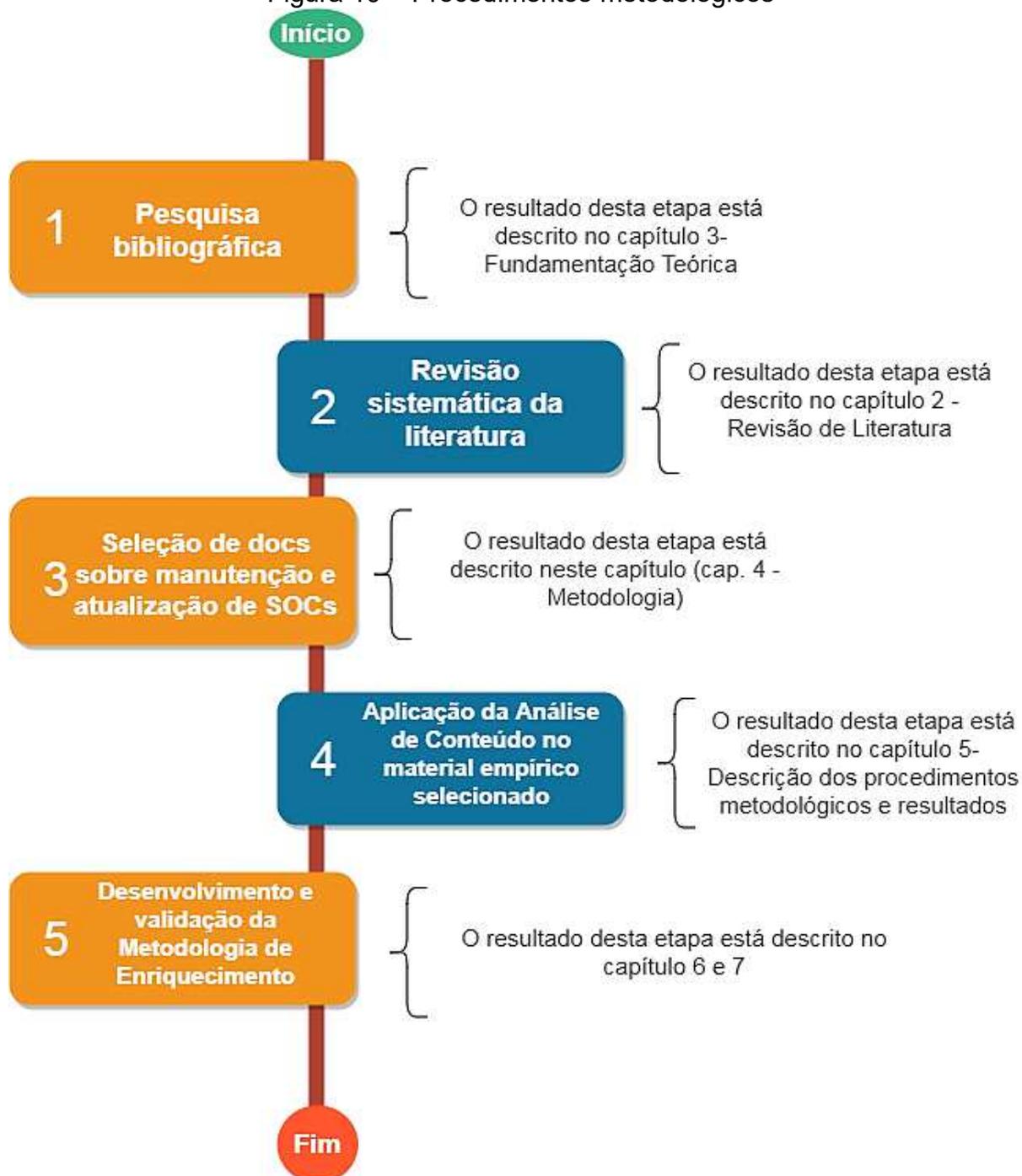
O material empírico dessa pesquisa refere-se aos resultados das etapas 2 e 3 dos procedimentos metodológicos (a ser descrito na [seção 4.4](#)). A Etapa 2 aborda a revisão sistemática da literatura sobre a temática Enriquecimento de Ontologias de domínio. A Etapa 3 trata da seleção de documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento.

Após seleção e aplicação da Análise de Conteúdo no material empírico, espera-se identificar os insumos para elaboração da metodologia de enriquecimento de ontologias de domínio.

4.4 Procedimentos metodológicos

Para o desenvolvimento desta pesquisa, propõem-se as cinco etapas apresentadas na Figura 10 – Procedimentos metodológicos. Em seguida, discorre-se sobre cada uma delas.

Figura 10 – Procedimentos metodológicos



Fonte: elaborada pelo autor.

Etapa 1. Pesquisa bibliográfica

Com o intuito de familiarizar-se com o assunto da pesquisa, nesta etapa, foi realizada uma pesquisa exploratória sobre as temáticas Sistema de Organização do Conhecimento, Ontologia, Engenharia Ontológica, Aprendizagem de Ontologia e Enriquecimento de Ontologia. Foram utilizadas como fonte de pesquisa as bases de dados e periódicos nacionais e internacionais das áreas de Ciência da Informação e Ciência da Computação (visto que grande parte da literatura relacionada às ontologias está nessa área), livros e bibliotecas digitais de teses e dissertações. Entre as bases pesquisadas, estão: Portal de Periódicos da Capes, Google Acadêmico, Base de Dados de Periódicos em Ciência da Informação (BRAPCI), Biblioteca Digital de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Portal ACM, CiteSeer; além de anais de eventos relevantes na área de ontologias, como: Ontobras, *Formal Ontology in Information Systems* (FOIS), *International Conference on Biomedical Ontology* (ICBO). O resultado dessa etapa refere-se à fundamentação teórica deste trabalho, descrita no [Capítulo 3 – Fundamentação teórica e metodológica](#).

Ressalta-se que a separação dessa pesquisa bibliográfica, com propósitos exploratórios, da pesquisa realizada para a revisão de literatura se deu por acreditar-se que facilitará o entendimento da pesquisa como um todo. Na pesquisa bibliográfica, o propósito foi apenas entender a temática Enriquecimento, assim como as temáticas relacionadas para subsidiar a fundamentação teórica e metodológica da pesquisa. Dessa forma, identificaram-se definições de enriquecimento, tipologia, características, e contextualizou-se o processo de enriquecimento dentro das áreas mais abrangentes. Já a revisão de literatura tem como objetivo específico a identificação de estudos que aplicaram o processo de enriquecimento em ontologias de domínio.

Quadro 4 – Expressões de busca da pesquisa bibliográfica

Temática	Expressão utilizada (português)	Expressão utilizada (inglês)
Sistema de Organização do Conhecimento	Sistema de organização do conhecimento, instrumento de representação do conhecimento, manutenção de sistema de organização do conhecimento, atualização de sistema de organização do conhecimento.	<i>Knowledge organization system, knowledge representation tool, maintenance of knowledge organization system, updating of knowledge organization system.</i>
Ontologia	Ontologias, ontologias de domínio, tipologia de ontologias, componentes de ontologias.	<i>Ontologies, domain ontologies, ontology typology, ontology components.</i>
Engenharia Ontológica	Engenharia ontológica, engenharia de ontologias, manutenção de ontologia, atualização de ontologia, ciclo de vida das ontologias.	<i>Ontology engineering, ontology maintenance, ontology update, ontology life cycle.</i>
Enriquecimento de ontologia	de Enriquecimento de ontologia, enriquecimento ontológico, expansão de ontologia, alargamento de ontologia.	<i>Ontology enrichment, ontology expansion, ontology enlarge.</i>
Aprendizagem de Ontologia	de Aprendizagem de ontologia, aprendizagem ontológica.	<i>Ontology learning, ontological learning.</i>

Fonte: elaborado pelo autor.

Etapa 2. Revisão sistemática da literatura sobre enriquecimento de ontologias de domínio

Realizou-se uma revisão sistemática da literatura para identificar as várias propostas para enriquecimento de ontologias de domínio com o intuito de saber como estava o estado da arte da área. O objetivo dessa etapa foi localizar, analisar e sintetizar estudos já publicados sobre o processo de enriquecimento de ontologias, identificar aspectos positivos e lacunas. Os detalhes sobre como os documentos foram selecionados, assim como a sua descrição e análise, estão descritos no [Capítulo 2 – Revisão de literatura sobre enriquecimento de ontologias de domínio](#).

Etapa 3. Seleção de documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento

Foram selecionadas duas normas que abordam a manutenção e a atualização de Sistemas de Organização do Conhecimento, a Norma ISO 25964-1 (2011) e a Norma ANSI/NISO Z39.19 (2005), além do capítulo “*Updating and maintenance of indexing languages and thesauri*” do livro *Indexing Languages and Thesauri: construction and maintenance*, do professor Dagobert Soergel (1974). Esses documentos foram selecionados com base em leituras sobre a temática Sistema de Organização do Conhecimento. As razões para escolhas desses documentos são as seguintes: (1) o fato de as Normas serem documentos consolidados e desenvolvidos com base em estudos detalhados e sugestões de vários pesquisadores experientes do assunto. Portanto as Normas apresentam consenso de uma parcela importante da comunidade científica; (2) já com relação ao capítulo de livro selecionado, respaldou-se na vivência do autor do capítulo, que possui vasta experiência nos estudos da temática. Além disso, o livro ao qual o capítulo pertence é uma referência incontestável sobre construção e atualização de SOCs; (3) as ontologias podem ser classificadas como um tipo de SOC. Dessa forma, possuem similaridades com outros SOCs, por exemplo, taxonomias e tesouros, e, nesse sentido, os procedimentos de manutenção e atualização desses instrumentos podem ser utilizados nas ontologias; (4) a temática construção e manutenção de SOCs ou vocabulários controlados (como também aparece na literatura) é uma área mais consolidada do que a engenharia ontológica. Sendo assim, construtos relevantes podem ser aproveitados no contexto da manutenção de ontologias, mais especificamente no processo de enriquecimento, que é o cerne desta pesquisa.

Uma descrição sobre a que se refere cada um dos documentos selecionados (Normas e capítulo de livro) já foi realizada na [seção 3.1](#) da fundamentação teórica desta pesquisa.

Etapa 4. Aplicação da Análise de Conteúdo no material empírico selecionado

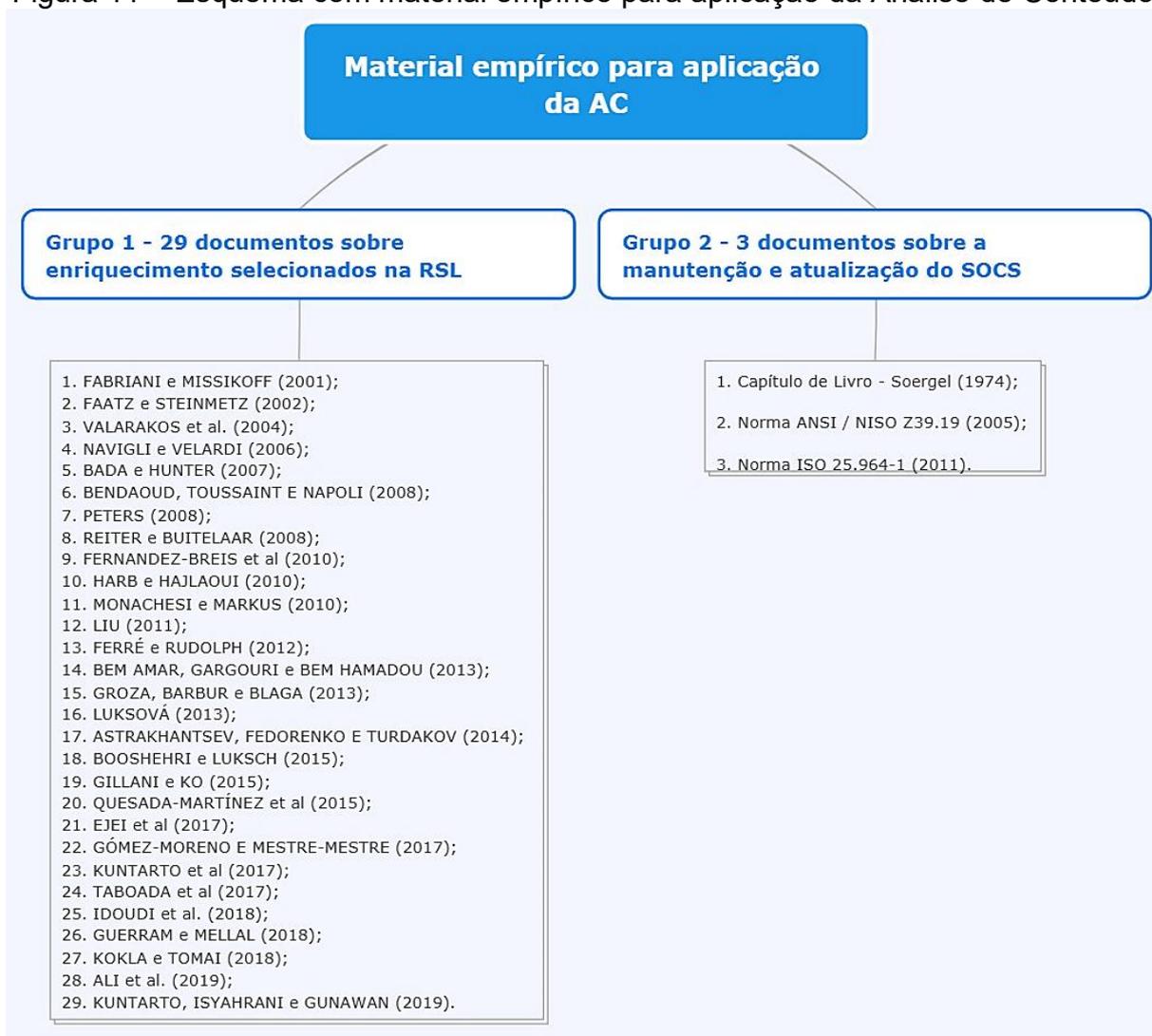
Nesta etapa da metodologia, empregou-se o método de análise de conteúdo proposto por Bardin (2016)²⁵ para explorar os estudos selecionados na **Etapa 2 –**

²⁵ A análise de conteúdo é um conjunto de instrumentos de cunho metodológico em constante aperfeiçoamento, que se aplica a discursos diversificados. (BARDIN, 2016).

Revisão sistemática da literatura sobre enriquecimento de ontologias de domínio e na **Etapa 3 – Seleção de documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento**. Esses documentos se configuram como material empírico para aplicação da Análise de Conteúdo. A finalidade dessa análise é identificar nesses documentos elementos passíveis de serem aplicados na metodologia de enriquecimento que se objetiva desenvolver nesta tese, assim como entender como esses elementos poderão ser utilizados para o desenvolvimento da metodologia de enriquecimento.

Para fins didáticos, separa-se o material empírico para aplicação da técnica de Análise de Conteúdo em dois grupos: Grupo 1 – Documentos oriundos da revisão sistemática da literatura; Grupo 2 – Documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento. Outro motivo para essa divisão refere-se ao fato de os documentos abordarem temáticas diferentes; sendo assim, as estratégias empregadas para tratamento dos elementos extraídos foram específicas para cada grupo. Os documentos do Grupo 1 tratam especificamente do processo de enriquecimento para ontologias de domínio. Já os documentos do Grupo 2 abordam a manutenção e a atualização de SOCs em geral. A Figura 11 apresenta o material empírico dividido nos grupos aos quais pertencem.

Figura 11 – Esquema com material empírico para aplicação da Análise de Conteúdo



Fonte: elaborada pelo autor.

A forma como a Análise de Conteúdo foi empreendida, bem como os resultados da análise em cada grupo de documentos estão detalhados no [Capítulo 5 – Descrição dos procedimentos metodológicos e resultados](#).

Etapa 5. Desenvolvimento e validação da metodologia para enriquecimento de ontologias de domínio

A partir dos insumos identificados e analisados na Etapa 4, bem como respaldado pelos estudos realizados para escrita da fundamentação teórica e revisão de literatura desta tese, desenvolveu-se um plano de ação que serviu de guia para o desenvolvimento da metodologia para enriquecimento de ontologias de domínio (descrito na [seção 5.4](#) – Plano de ação para subsidiar o desenvolvimento da metodologia). A metodologia de enriquecimento e seus detalhes estão descritos no

capítulo 6 – Metodologia para Enriquecimento de Ontologias de Domínio (METHODOE).

Vale destacar que o desenvolvimento de tal metodologia buscou fundamentação também nos construtos teóricos e metodológicos da Biblioteconomia e Ciência da Informação. Assim, sempre que uma atividade pertencente à metodologia se mostrou passível de ser auxiliada com uma técnica ou teoria da BCI, fez-se conexão entre a atividade e a técnica. Como apontado por Marcondes e Campos (2008, p. 122), "ao longo de sua trajetória, a CI desenvolveu um arcabouço teórico e metodológico consubstanciado em obras como as de Otlet (1989), Ranganathan (1967), Dahlberg (1978a), Hjørland (2002) e do CRG (WITSON, 1972), entre outros". Portanto acredita-se que parte desse arcabouço tem muito a contribuir com os estudos sobre ontologias, que também é um objeto de estudo na CI.

A metodologia de enriquecimento foi validada mediante aplicação em uma ontologia desenvolvida no contexto da EMBRAPA Campinas, denominada OntoAgroHidro. Para essa validação, selecionou-se um extrato da ontologia, visto que o enriquecimento de toda ontologia não seria possível dentro dos limites de escopo e prazo desta tese. Ademais o objetivo é testar se a metodologia é capaz de enriquecer todos os componentes da ontologia, bem como se ela apresenta diretrizes sobre como conduzir todo o processo de enriquecimento; sendo assim, um extrato é suficiente. O resultado dessa validação está demonstrado no capítulo 7 – Aplicação da METHODOE e discussão dos resultados.

4.5 Relação entre os objetivos específicos e os procedimentos metodológicos

Os procedimentos metodológicos de uma pesquisa possuem o propósito de apresentar como os objetivos serão alcançados, ou seja, quais serão as estratégias adotadas para se atingir o que se almeja com a pesquisa e assim solucionar os problemas identificados no início do estudo. Com o intuito de deixar evidente a relação entre os objetivos da pesquisa e os procedimentos metodológicos, apresenta-se o Quadro 5 a seguir.

Quadro 5 – Relação entre objetivos específicos e procedimentos metodológicos da pesquisa

Objetivos específicos	Procedimentos metodológicos
1) Caracterizar o processo de enriquecimento e desenvolver um fluxo para ele.	Desenvolvimento da metodologia para enriquecimento de ontologias de domínio.
2) Identificar insumos na área de engenharia ontológica e na Biblioteconomia e Ciência da Informação para o desenvolvimento da metodologia.	Revisão sistemática da literatura sobre enriquecimento de ontologias de domínio. Seleção de documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento. Aplicação da Análise de Conteúdo no material empírico selecionado.
3) Demonstrar as principais lacunas nos métodos existentes para enriquecimento de ontologias de domínio.	Revisão sistemática da literatura sobre a temática enriquecimento de ontologias de domínio.
4) Dar subsídios aos desenvolvedores de ontologias para melhoria do processo de enriquecimento.	Desenvolvimento da metodologia para enriquecimento de ontologias de domínio.
5) Contribuir com os estudos sobre manutenção/atualização de ontologias de domínio.	Revisão sistemática da literatura sobre a temática enriquecimento de ontologias de domínio. Desenvolvimento da metodologia para enriquecimento de ontologias de domínio.

Fonte: elaborado pelo autor.

5 DESCRIÇÃO DOS PROCEDIMENTOS METODOLÓGICOS E RESULTADOS

Neste capítulo, descrevem-se com mais detalhes os procedimentos da Etapa 4 da metodologia de pesquisa, ou seja, a aplicação da análise de conteúdo no material empírico selecionado, bem como os resultados dessa análise.

Este capítulo está dividido de acordo com as seguintes seções: a seção [5.1 – Análise de conteúdo dos documentos selecionados na revisão de literatura \(Grupo 1\)](#) aborda os 29 documentos selecionados na Revisão Sistemática da Literatura para identificação de procedimentos que possam ser úteis para o desenvolvimento da metodologia de enriquecimento de ontologias de domínio; a seção [5.2 – Análise de conteúdo dos documentos sobre manutenção e atualização de SOCs \(Grupo 2\)](#) envolve os 3 documentos selecionados na Etapa 3 da metodologia de pesquisa e visa à extração de procedimentos que também possam ser úteis no contexto do enriquecimento de ontologias; a seção [5.3 – Categorização dos procedimentos extraídos do material empírico](#) discorre sobre a organização dos elementos extraídos dos documentos dos Grupo 1 e 2 em categorias, e a seção [5.4 – Plano de ação para subsidiar o desenvolvimento da metodologia de enriquecimento](#) refere-se ao plano de ação desenvolvido para guiar a construção da metodologia que se almeja nesta pesquisa.

5.1 Análise de conteúdo dos documentos selecionados na revisão de literatura (Grupo 1)

O objetivo desta etapa da pesquisa foi identificar quais procedimentos presentes nos métodos de enriquecimento de ontologias de domínio podem auxiliar no desenvolvimento de uma metodologia para esse mesmo processo. Sendo assim, respaldou-se no método de Análise de Conteúdo proposto por Bardin (2016).

Primeiramente, definiu-se o objetivo da análise, que se refere a identificar procedimentos passíveis de serem aproveitados em uma metodologia de enriquecimento de ontologias de domínio. O pressuposto é que as propostas individuais para enriquecimento aplicadas em domínios específicos possuem insumos úteis para desenvolvimento de uma proposta metodológica ampla e independente de domínio. Os dados para a realização da análise foram provenientes dos estudos selecionados na revisão de literatura.

No segundo passo, realizou-se a leitura flutuante dos estudos selecionados em busca da seção que trata da proposta de enriquecimento, visto que o interesse era identificar a forma como o processo de enriquecimento foi realizado em cada estudo.

No terceiro passo, procedeu-se à exploração do material, do qual extraíram-se trechos dos estudos que pudessem oferecer subsídios sobre como realizar alguma etapa dentro do processo de enriquecimento e, assim, servir de insumo para a metodologia que aqui se almeja.

O último passo foi associar o trecho extraído do estudo a um procedimento generalizado equivalente, ou seja, um procedimento que pudesse ser aplicado de forma abrangente e não apenas naquele domínio do qual o método discorre. Por exemplo, se o trecho extraído de determinado estudo fosse "analisar o tesauro AGROVOC para extração de conceitos sobre recursos hídricos", mapeava-se aquele trecho para um procedimento geral equivalente, tal como "seleção da fonte de informação". A ideia de associar o trecho extraído a um procedimento generalizado vai ao encontro do objetivo desta pesquisa, que, em resumo, visa ao desenvolvimento de uma metodologia de enriquecimento de ontologia de domínio que seja independente de domínio. Portanto o procedimento deveria ser generalizado a ponto de não depender de um domínio específico ou de uma fonte de informação que só existe em certos domínios, como é o caso dos tesouros. Por isso, o procedimento generalizado indicado para o trecho do exemplo é "seleção da fonte de informação" e não "analisar tesouros para extração de termos".

O resultado completo referente à aplicação da Análise de Conteúdo nos 29 documentos da revisão de literatura encontra-se no [Apêndice C](#). Nesse apêndice, apresentam-se todos os trechos extraídos discriminados por estudo, bem como os procedimentos generalizados equivalentes ao trecho. Foram coletados 195 trechos que são representados por 19 procedimentos generalizados equivalentes (considerando a exclusão de duplicatas e a mesclagem de procedimentos muito similares).

Os 19 procedimentos generalizados que foram identificados a partir dos 195 trechos extraídos representam os insumos para a elaboração da metodologia de enriquecimento. Portanto objetiva-se discorrer sobre eles no restante desta seção. Nessa perspectiva, apresentam-se, na Tabela 1, esses procedimentos, assim como uma referência aos estudos do qual o procedimento foi extraído. Ressalta-se que a

ordem de apresentação dos procedimentos na tabela se baseia numa ordem hipotética de execução de cada procedimento no processo de enriquecimento.

Tabela 1 – Relação de procedimentos candidatos para a metodologia de enriquecimento a partir da literatura

			Continua...
#	Procedimento	Fonte de informação	Quantidade de estudos que tratam do procedimento
1	Identificação da ontologia alvo	Faatz e Steinmetz (2002); Bada e Hunter (2007); Monachesi e Markus (2010); Ferré e Rudolph (2012); Luksová (2013); Booshehri e Luksch (2015); Quesada-Martínez (2015); Ejei <i>et al.</i> (2017); Gómez-Moreno e Mestre-Mestre (2017); Guerram e Mellal (2018); Idoudi <i>et al.</i> (2018); Ali <i>et al.</i> (2019)	12
2	Diagnóstico da ontologia	Fernandez-Breis <i>et al.</i> (2010)	1
3	Avaliação da necessidade de inserção de determinado componente	Kuntarto, Isyahrani e Gunawan (2019)	1
4	Definição do escopo do enriquecimento	Luksová (2010); Gómez-Moreno e Mestre-Mestre (2017)	2
5	Identificação de recursos que suportem o enriquecimento	Peters (2009); Ben Amar, Gargouri e Hamadou (2013); Astrakhantsev, Fedorenko e Turdakov (2014); Gillani e Ko (2015); Ejei <i>et al.</i> (2017); Kokla, Papadias e Tomai (2018)	6

Continua...

#	Procedimento	Fonte de informação	Quantidade de estudos que tratam do procedimento
6	Seleção da fonte de informação	Faatz e Steinmetz (2002); Navigli e Velardi (2006); Bada e Hunter (2007); Reiter e Buitelaar (2007); Fernandez-Breis <i>et al.</i> (2010); Monachesi e Markus (2010); Kaihong (2012); Ben Amar, Gargouri e Hamadou (2013); Groza, Barbur e Blaga (2013); Luksová (2013); Astrakhantsev, Fedorenko e Turdakov (2014); Booshehri e Luksch (2015); Gillani e Ko (2015); Ejei <i>et al.</i> (2017); Gómez-Moreno e Mestre-Mestre (2017); Kuntarto (2017); Taboada <i>et al.</i> (2017); Idoudi <i>et al.</i> (2018); Kokla, Papadias e Tomai (2018); Ali <i>et al.</i> (2019); Kuntarto, Isyahrani e Gunawan (2019)	21
7	Escolha da técnica de extração da informação	Navigli e Velardi (2006); Bendadoud, Toussaint e Napoli (2008); Idoudi <i>et al.</i> (2018); Kokla, Papadias e Tomai (2018)	4
8	Preparação da fonte de informação	Valakaros <i>et al.</i> (2004); Peters (2009); Harb e Hajlaoui (2010); Kaihong (2012); Astrakhantsev, Fedorenko e Turdakov (2014); Booshehri e Luksch (2015); Gillani e Ko (2015); Ejei <i>et al.</i> (2017); Gómez-Moreno e Mestre-Mestre (2017); Kuntarto (2017); Taboada <i>et al.</i> (2017); Guerram e Mellal (2018); Ali <i>et al.</i> (2019)	13
9	Aplicação da técnica de extração da informação	Consta em todos os 29 estudos analisados	29

			Continua...
#	Procedimento candidato para a metodologia de enriquecimento	Fonte de informação	Quantidade de estudos que tratam do procedimento
10	Tratamento do conhecimento adquirido	Valakaros <i>et al.</i> (2004); Bada e Hunter (2007); Reiter e Buitelaar (2007); Bendadoud, Toussaint e Napoli (2008); Peters (2009); Fernandez-Breis <i>et al.</i> (2010); Harb e Hajlaoui (2010); Monachesi e Markus (2010); Kaihong (2012); Ben Amar, Gargouri e Hamadou (2013); Groza, Barbur e Blaga (2013); Luksová (2013); Astrakhantsev, Fedorenko e Turdakov (2014); Booshehri e Luksch (2015); Gillani e Ko (2015); Ejei <i>et al.</i> (2017); Gómez-Moreno e Mestre-Mestre (2017); Taboada <i>et al.</i> (2017); Idoudi <i>et al.</i> (2018); Kokla, Papadias e Tomai (2018); Ali <i>et al.</i> (2019)	21
11	Definição das regras para seleção do conhecimento extraído	Navigli e Velardi (2006); Harb e Hajlaoui (2010); Astrakhantsev, Fedorenko e Turdakov (2014); Ejei <i>et al.</i> (2017); Kokla, Papadias e Tomai (2018);	5
12	Seleção do conhecimento candidato para enriquecer a ontologia	Booshehri e Luksch (2015); Gillani e Ko (2015); Ali <i>et al.</i> (2019); Kuntarto, Isyahrani e Gunawan (2019)	4

Continua...

#	Procedimento candidato para a metodologia de enriquecimento	Fonte de informação	Quantidade de estudos que tratam do procedimento
13	Validação do conhecimento adquirido	Valakaros <i>et al.</i> (2004); Bada e Hunter (2007); Bendadoud, Toussaint e Napoli (2008); Harb e Hajlaoui (2010); Kaihong (2012); Ben Amar, Gargouri e Hamadou (2013); Booshehri e Luksch (2015); Quesada-Martínez (2015); Ejei <i>et al.</i> (2017); Guerram e Mellal (2008)	10
14	Desenvolvimento de representação intermediária	Bada e Hunter (2007); Peters (2009)	2
15	Representação do conhecimento usando linguagem formal	Bendadoud, Toussaint e Napoli (2008); Peters (2009)	2
16	Enriquecimento da ontologia	Valakaros <i>et al.</i> (2004); Navigli e Velardi (2006); Bada e Hunter (2007); Bendadoud, Toussaint e Napoli (2008); Peters (2009); Fernandez-Breis <i>et al.</i> (2010); Harb e Hajlaoui (2010); Ferré e Rudolph (2012); Groza, Barbur e Blaga (2013); Luksová (2013); Booshehri e Luksch (2015); Gillani e Ko (2015); Gómez-Moreno e Mestre-Mestre (2017); Taboada <i>et al.</i> (2017); Idoudi <i>et al.</i> (2018); Ali <i>et al.</i> (2019); Kuntarto, Isyahrani e Gunawan (2019)	17
17	Revisão da ontologia enriquecida	Fernandez-Breis <i>et al.</i> (2010)	1

			Conclusão.
#	Procedimento	Fonte de informação	Quantidade de estudos que tratam do procedimento
18	Validação da ontologia enriquecida	Velardi, Fabriani e Missikoff (2001); Kaihong (2012); Kuntarto (2017); Ali <i>et al.</i> (2019)	4
19	Organização dos documentos de entrada e saída de cada etapa	Kaihong (2012); Gillani e Ko (2015); Quesada-Martínez (2015);	3

Fonte: dados de pesquisa.

Passa-se agora a definir cada um dos procedimentos, pois isso facilitará a categorização deles. Ademais entender o que significa cada procedimento facilitará a sua seleção ou descarte da metodologia de enriquecimento em desenvolvimento.

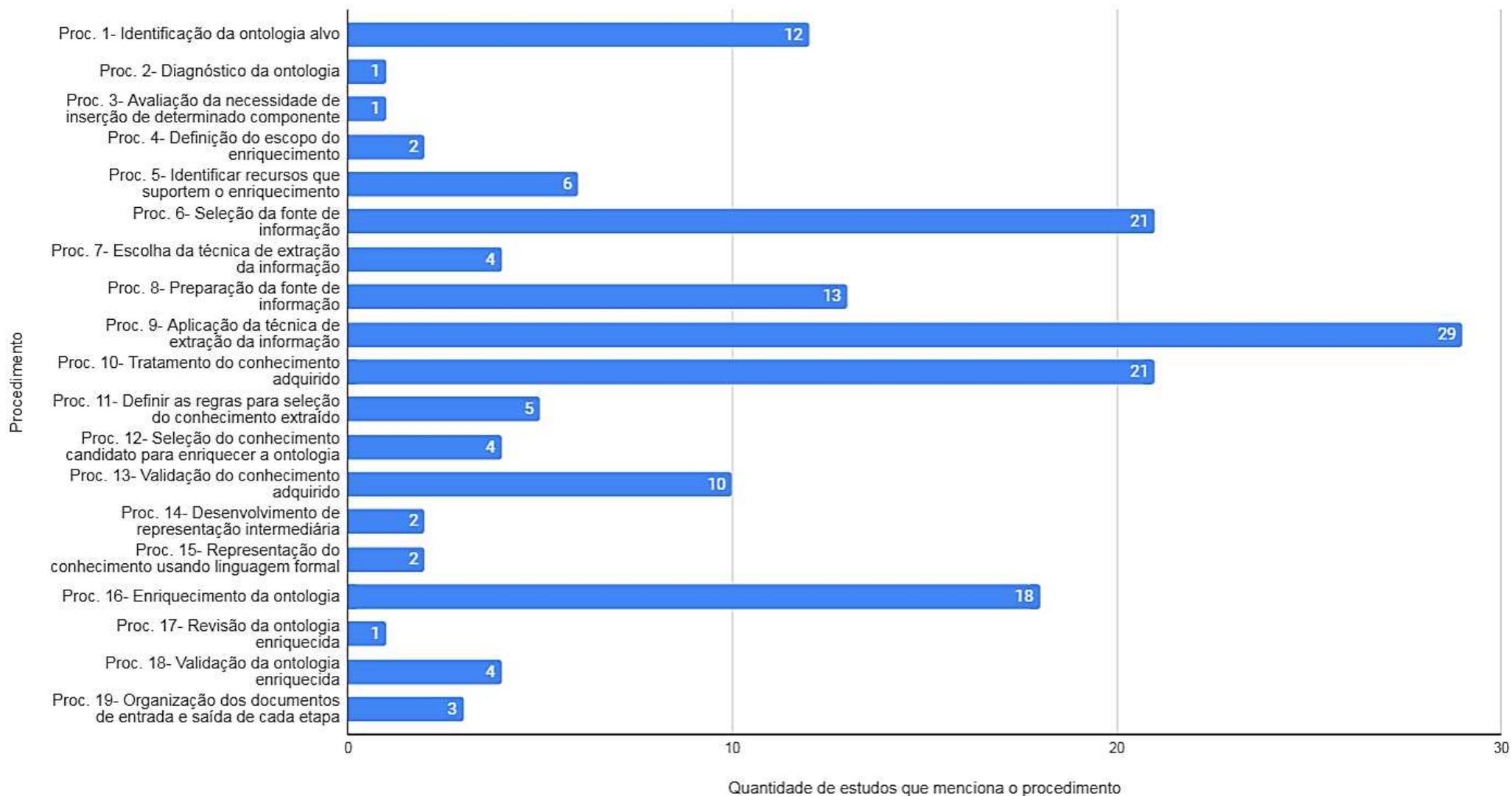
- **Procedimento 1 – Identificação da ontologia alvo:** refere-se à apresentação de uma ontologia inicial na qual o processo de enriquecimento é aplicado.
- **Procedimento 2 – Diagnóstico da ontologia:** trata-se de uma inspeção na estrutura da ontologia para identificar quais partes da ontologia precisam passar pelo processo de enriquecimento.
- **Procedimento 3 – Avaliação da necessidade de inserção de determinado componente:** trata-se de analisar se determinado componente já foi representado na estrutura da ontologia.
- **Procedimento 4 – Definição do escopo do enriquecimento:** aborda o objetivo do processo de enriquecimento, ou seja, quais tipos de enriquecimento são necessários, bem como as partes da ontologia que devem ser enriquecidas.
- **Procedimento 5 – Identificação de recursos que suportem o enriquecimento:** versa sobre todos os tipos de elementos que podem contribuir para a execução do processo de enriquecimento, sejam softwares que auxiliem na organização do conhecimento extraído, sejam repositórios de SOCs de onde se podem recuperar tesouros, taxonomias e ontologias que tenham relação com o domínio da ontologia que está sendo enriquecida.

- **Procedimento 6 – Seleção da fonte de informação:** discorre sobre todos os materiais que contenham conteúdo relevante para enriquecimento da ontologia de domínio.
Procedimento 7 – Escolha da técnica de extração da informação: trata da seleção do método utilizado para se extrair a informação para enriquecimento da ontologia.
- **Procedimento 8 – Preparação da fonte de informação:** aborda a manipulação da fonte de informação de acordo com a técnica de extração da informação, posto que algumas técnicas exigem que a fonte esteja organizada de acordo com um formato de arquivo específico.
- **Procedimento 9 – Aplicação da técnica de extração da informação:** versa sobre a implementação da estratégia de extração da informação, ou seja, é a própria extração do conteúdo presente na fonte de informação.
- **Procedimento 10 – Tratamento do conhecimento adquirido:** refere-se a toda intervenção realizada no conteúdo extraído da fonte de informação, tais como definição, classificação de conceitos, ranqueamento de termos de acordo com quaisquer critérios, análise de similaridade dos conceitos extraídos com o domínio de conhecimento, etc.
- **Procedimento 11 – Definição das regras para seleção do conhecimento extraído:** discorre sobre os critérios que são adotados para selecionar ou descartar o conteúdo extraído da fonte de informação.
- **Procedimento 12 – Seleção do conhecimento candidato para enriquecer a ontologia:** refere-se à escolha do conteúdo relevante para o enriquecimento da ontologia.
- **Procedimento 13 – Validação do conhecimento adquirido:** trata da ratificação de que aquele conteúdo selecionado para enriquecimento da ontologia está realmente adequado.
- **Procedimento 14 – Desenvolvimento de representação intermediária:** refere-se à criação de modelos de representação que contribuem para refletir sobre o conteúdo antes da inserção do conteúdo na ontologia.
- **Procedimento 15 – Representação do conhecimento usando linguagem formal:** versa sobre a representação do conhecimento extraído em lógica, ou seja, numa linguagem não ambígua.

- **Procedimento 16 – Enriquecimento da ontologia:** aborda a inserção do conhecimento adquirido na estrutura da ontologia, possibilitando assim a sua expansão.
- **Procedimento 17 – Revisão da ontologia enriquecida:** refere-se a uma inspeção na estrutura da ontologia para averiguar o impacto do processo de enriquecimento.
- **Procedimento 18 – Validação da ontologia enriquecida:** refere-se a uma análise da ontologia enriquecida para verificar se ela está representando o domínio de forma adequada ou se existe alguma representação inadequada.
- **Procedimento 19 – Organização dos documentos de entrada e saída de cada etapa:** discorre sobre o ordenamento dos documentos gerados em cada etapa e que servem de insumo para uma etapa subsequente.

Na intenção de discutir sobre os procedimentos mais comuns nos estudos analisados, apresenta-se o Gráfico 10 a seguir.

Gráfico 10 – Quantitativo de estudos que tratam dos procedimentos



Fonte: dados de pesquisa.

Percebe-se que os procedimentos que mais aparecem nos estudos são: o procedimento 9 – aplicação da técnica de extração da informação, que está presente em 29 estudos; seguido pelo procedimento 6 – seleção da fonte de informação, com 21 ocorrências; procedimento 10 – tratamento do conhecimento adquirido, também com 21; e o procedimento 16 – enriquecimento da ontologia, que consta em 18 estudos. Os outros procedimentos encontram-se em cerca de 40% dos estudos, tais como: procedimento 8 – preparação da fonte de informação (13); procedimento 1 – identificação da ontologia alvo (12); procedimento 13 – validação do conhecimento adquirido (10). Por fim, têm-se os procedimentos que aparecem em, no máximo, seis estudos, tais como: procedimento 5 – identificação dos recursos que suportem o enriquecimento (6); procedimento 11 – definição das regras para seleção do conhecimento extraído (5); procedimento 7 – escolha da técnica de extração da informação (4); procedimento 12 – seleção do conhecimento candidato para enriquecer a ontologia (4); procedimento 18 – validação da ontologia enriquecida (4); procedimento 19 – organização dos documentos de entrada e saída de cada etapa (3); procedimento 4 – definição do escopo do enriquecimento (2); procedimento 14 – desenvolvimento de representação intermediária (2); procedimento 15 – representação do conhecimento usando linguagem formal (2); procedimento 2 – diagnóstico da ontologia (1); procedimento 3 – avaliação da necessidade de inserção de determinado componente (1); procedimento 17 – revisão da ontologia enriquecida (1).

O fato de ter-se trabalhado com estudos aplicados pode ser uma explicação para alguns desses dados. Conforme foi detalhado no [capítulo 2](#) desta tese, todos os estudos efetivamente aplicaram um método de enriquecimento em uma ontologia de domínio. Sendo assim, é esperado, por exemplo, que o procedimento 9 – aplicação da técnica de extração da informação esteja presente em todos os estudos, posto que não é possível realizar o processo de enriquecimento sem utilizar ao menos uma estratégia de extração da informação. Nessa perspectiva, o procedimento 7 – escolha da técnica de extração da informação também deveria estar presente em todos os estudos, porém ele é mencionado em apenas quatro deles. O motivo é que a maioria dos estudos não explicitam esse procedimento, portanto não foi possível identificá-lo entre os procedimentos adotados nesses estudos, ainda que seja coerente a seleção prévia de determinada técnica de extração da informação para, em seguida, aplicá-la.

Notou-se que os procedimentos mencionados em mais de 60% dos estudos são aqueles que abordam atividades e elementos fundamentais no processo de enriquecimento, tais como **fonte de informação**, posto que a matéria-prima para enriquecimento da ontologia precisa ser extraída de algum material; **técnica de extração**, já que é necessária uma estratégia para se extrair a matéria-prima; **tratamento do conhecimento**, dado que uma vez extraída, é preciso processar e tratar a matéria-prima; e, por fim, o **enriquecimento da ontologia**, pois é preciso inserir os elementos na ontologia para que o processo de enriquecimento seja efetivado.

Já no caso dos procedimentos que ocorrem em menos de 20% dos estudos, percebe-se que alguns deles referem-se a procedimentos que já são contemplados implicitamente em outros procedimentos, como é o caso do procedimento 11 – definição das regras para tratamento do conhecimento adquirido e do procedimento 12 – seleção do conhecimento candidato para enriquecer a ontologia, que, em alguns estudos, são contemplados pelo procedimento 9 – aplicação da técnica de extração da informação. Isso porque alguns estudos possuem métodos semi ou automáticos em que já estão previstos a extração, o tratamento do conhecimento e até a seleção do conhecimento em uma só etapa. A mesma situação se aplica para os procedimentos 14 – desenvolvimento de representação intermediária e 15 – representação do conhecimento usando linguagem formal, que, na maioria dos estudos, são absorvidos pelo procedimento 16 – enriquecimento da ontologia, ou seja, o conteúdo é inserido diretamente na ontologia sem passar por representação intermediária; e a linguagem formal é aquela na qual a ontologia está codificada.

5.2 Análise de conteúdo dos documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento (Grupo 2)

Nesta seção, o objetivo foi aplicar a Análise de Conteúdo nos documentos selecionados na Etapa 3 dos procedimentos metodológicos da pesquisa, que se referem às Normas ISO 25964-1 (2011) e ANSI/NISO Z39.19-2005 (2010) e ao capítulo do livro de Soergel (1974), que configuram o Grupo 2 do material empírico. Esses documentos versam sobre a construção, a manutenção e a atualização de

Sistemas de Organização do Conhecimento; e a análise focou nas seções desses documentos que versam sobre manutenção e atualização.

Assim como a análise empreendida nos documentos do Grupo 1, o objetivo é o mesmo, ou seja, identificar procedimentos passíveis de serem aproveitados em uma metodologia de enriquecimento de ontologias de domínio. Todavia foi necessário converter esses procedimentos para o contexto do enriquecimento de ontologias de domínio, visto que se tratavam de documentos que abordam os Sistemas de Organização do Conhecimento de forma geral, bem como não dissertam sobre o processo de enriquecimento propriamente.

Para levantamento dos procedimentos dos documentos, também realizou-se uma leitura flutuante em busca de insumos relevantes para o processo de enriquecimento de ontologias. Em segundo lugar, na fase de exploração do material, extraíram-se trechos desses documentos que pudessem oferecer subsídios para o processo de enriquecimento e, assim, servir de insumo para a metodologia que aqui se almeja. Em seguida, prosseguiu-se na análise dos trechos com o intuito de associá-los a um procedimento equivalente que fosse aplicável ao processo de enriquecimento de ontologias de domínio. A título de exemplificação, para o trecho extraído “a data em que um descritor foi introduzido deve ser descrita na nota de escopo”, indicou-se o procedimento equivalente “controlar a data em que determinado componente foi inserido na ontologia”. Ao final, somaram-se 57 trechos que representam 20 procedimentos relacionados ao processo de enriquecimento de ontologias.

Todos os trechos selecionados encontram-se no [Apêndice D – Coleta dos dados referentes à aplicação da Análise de Conteúdo nos documentos sobre manutenção e atualização de Sistemas de Organização do Conhecimento](#). Na Tabela 2, apresentam-se os 20 procedimentos e os documentos a partir do qual o procedimento foi extraído. A ordem de apresentação dos procedimentos segue o mesmo critério adotado para apresentar os procedimentos identificados nos documentos do Grupo 1, ou seja, ordem hipotética de execução de cada procedimento no processo de enriquecimento.

Tabela 2 – Relação de procedimentos candidatos para a metodologia de enriquecimento a partir de documentos sobre manutenção e atualização de SOCs

Continua...

#	Procedimento candidato para a metodologia de enriquecimento	Documento (s) de onde o procedimento foi extraído	Quantidade de documentos que tratam do procedimento
1	Identificação da ontologia alvo	ISO 25964-1 (2011)	1
2	Diagnóstico da ontologia	ANSI/NISO Z39.19-2005 (2010); ISO 25964-1 (2011)	2
3	Avaliação da necessidade de enriquecimento	ANSI/NISO Z39.19-2005 (2010)	1
4	Definição do escopo do enriquecimento	ISO 25964-1 (2011)	1
5	Definição dos recursos necessários (humano, ferramentas, financiamento)	ISO 25964-1 (2011)	1
6	Definição das responsabilidades	ISO 25964-1 (2011)	1
7	Identificação dos recursos que suportem o enriquecimento	SOERGEL (1974); ISO 25964-1 (2011)	2
8	Planejamento da frequência do enriquecimento	SOERGEL (1974); ANSI/NISO Z39.19-2005 (2010); ISO 25964-1 (2011)	3
9	Consideração do uso de procedimentos do desenvolvimento de ontologia para o processo de enriquecimento	SOERGEL (1974)	1
10	Seleção da fonte de informação	SOERGEL (1974); ISO 25964-1 (2011)	2
11	Escolha da técnica para extração da informação	ISO 25964-1 (2011)	1

Conclusão.

#	Procedimento candidato para a metodologia de enriquecimento	Documento (s) de onde o procedimento foi extraído	Quantidade de documentos que tratam do procedimento
12	Tratamento do conhecimento adquirido	ISO 25964-1 (2011)	1
13	Validação do conhecimento adquirido	SOERGEL (1974); ANSI/NISO Z39.19-2005 (2010); ISO 25964-1 (2011)	3
14	Enriquecimento da ontologia	SOERGEL (1974); ISO 25964-1 (2011)	2
15	Alteração no conteúdo da ontologia	SOERGEL (1974)	1
16	Possibilidade de propostas de alterações na ontologia por especialistas	ANSI/NISO Z39.19-2005 (2010)	1
17	Controle da data em que o componente foi inserido	SOERGEL (1974); ANSI/NISO Z39.19-2005 (2010); ISO 25964-1 (2011)	3
18	Verificação do impacto da alteração realizada no conteúdo da ontologia	SOERGEL (1974); ANSI/NISO Z39.19-2005 (2010); ISO 25964-1 (2011)	3
19	Controle das alterações realizadas no conteúdo da ontologia	SOERGEL (1974)	1
20	Disseminação da versão da ontologia enriquecida	ISO 25964-1 (2011)	1

Fonte: dados de pesquisa.

A partir da análise dos procedimentos da Tabela 2, verifica-se que nove dos procedimentos identificados nos documentos analisados do Grupo 2 já haviam sido identificados nos documentos do Grupo 1. São os procedimentos: Procedimento 1 –

Identificação da ontologia alvo; Procedimento 2 – Diagnóstico da ontologia; Procedimento 4 – Definição do escopo do enriquecimento; Procedimento 7 – Identificação dos recursos que suportem o enriquecimento; Procedimento 10 – Seleção da fonte de informação; Procedimento 11 – Escolha da técnica para extração da informação; Procedimento 12 – Tratamento do conhecimento adquirido; Procedimento 13 – Validação do conhecimento adquirido; Procedimento 14 – Enriquecimento da ontologia. Dessa forma, apresenta-se, a seguir, a definição apenas dos procedimentos que foram identificados pela primeira vez.

- **Procedimento 1 – Identificação da ontologia alvo:** já definido no Grupo 1.
- **Procedimento 2 – Diagnóstico da ontologia:** já definido no Grupo 1.
- **Procedimento 3 – Avaliação da necessidade de enriquecimento:** trata-se de uma análise para saber se a ontologia precisa passar pelo processo de enriquecimento.
- **Procedimento 4 – Definição do escopo do enriquecimento:** já definido no Grupo 1.
- **Procedimento 5 – Definição dos recursos necessários (humano, ferramentas, financiamento):** refere-se à identificação de todos os meios para que a implantação do enriquecimento seja possível.
- **Procedimento 6 – Definição de responsabilidades:** versa sobre a atribuição de funções para os recursos humanos envolvidos no processo de enriquecimento.
- **Procedimento 7 – Identificação dos recursos que suportem o enriquecimento:** já definido no Grupo 1.
- **Procedimento 8 – Planejamento da frequência do enriquecimento:** discorre sobre a periodicidade do processo de enriquecimento.
- **Procedimento 9 – Consideração do uso de procedimentos do desenvolvimento de ontologia para o processo de enriquecimento:** trata-se do aproveitamento de atividades comuns entre o processo de construção de ontologias e o de enriquecimento.
- **Procedimento 10 – Seleção da fonte de informação:** já definido no Grupo 1.
- **Procedimento 11 – Escolha da técnica para extração da informação:** já definido no Grupo 1.

- **Procedimento 12 – Tratamento do conhecimento adquirido:** já definido no Grupo 1.
- **Procedimento 13 – Validação do conhecimento adquirido:** já definido no Grupo 1.
- **Procedimento 14 – Enriquecimento da ontologia:** já definido no Grupo 1.
- **Procedimento 15 – Alteração no conteúdo da ontologia:** trata-se das mudanças realizadas na ontologia em decorrência do processo de enriquecimento.
- **Procedimento 16 – Possibilidade de propostas de alterações na ontologia por especialistas:** refere-se ao desenvolvimento de meios para que os especialistas possam sugerir conteúdo para o enriquecimento da ontologia.
- **Procedimento 17 – Controle da data em que o componente foi inserido:** trata-se de uma atividade de gestão relevante para o registro temporal de alterações na ontologia.
- **Procedimento 18 – Verificação do impacto da alteração realizada no conteúdo da ontologia:** discorre sobre como o enriquecimento pode repercutir no conteúdo da ontologia.
- **Procedimento 19 – Controle das alterações realizadas no conteúdo da ontologia:** refere-se a um procedimento de gestão para registrar exatamente quais mudanças foram realizadas na ontologia.
- **Procedimento 20 – Disseminação da versão da ontologia enriquecida:** versa sobre a divulgação da nova versão da ontologia para a comunidade interessada.

Constata-se, por meio da Tabela 2, que apenas quatro procedimentos estão presentes nos três documentos analisados, são eles: (8) – planejamento da frequência do enriquecimento, (13) – validação do conhecimento adquirido, (17) – controle da data em que o componente foi inserido, (18) – verificação do impacto da alteração realizada no conteúdo da ontologia. Assim nota-se que os procedimentos identificados foram bem diversos, entre os três documentos analisados.

Observa-se, com base nas definições apresentadas para os procedimentos, que uma boa parte deles tem um viés relacionado com a gestão do processo de enriquecimento, tais como: definir os recursos necessários (humano, ferramentas, financiamento), definir responsabilidades, planejar frequência, controlar a data em que

o componente foi inserido, controlar as alterações realizadas no conteúdo da ontologia. Acredita-se que isso tem relação com as fontes de informação das quais esses procedimentos originaram, que se referem às Normas ISO 25964-1 (2011), ANSI/NISO 39.19-2005 (2010) e ao capítulo do livro de Soergel (1974), uma vez que esses documentos versam sobre Sistemas de Organização do Conhecimento tradicionais (principalmente os tesouros), que já possuem um percurso longo com muitos estudos, tornando-os sólidos. Acredita-se que a gestão tem um importante papel para que todos os processos associados a esses instrumentos tenham êxito, pois não basta apenas dizer o que precisa ser feito, é preciso dizer quando e quem será o responsável.

5.3 Categorização dos procedimentos extraídos do material empírico

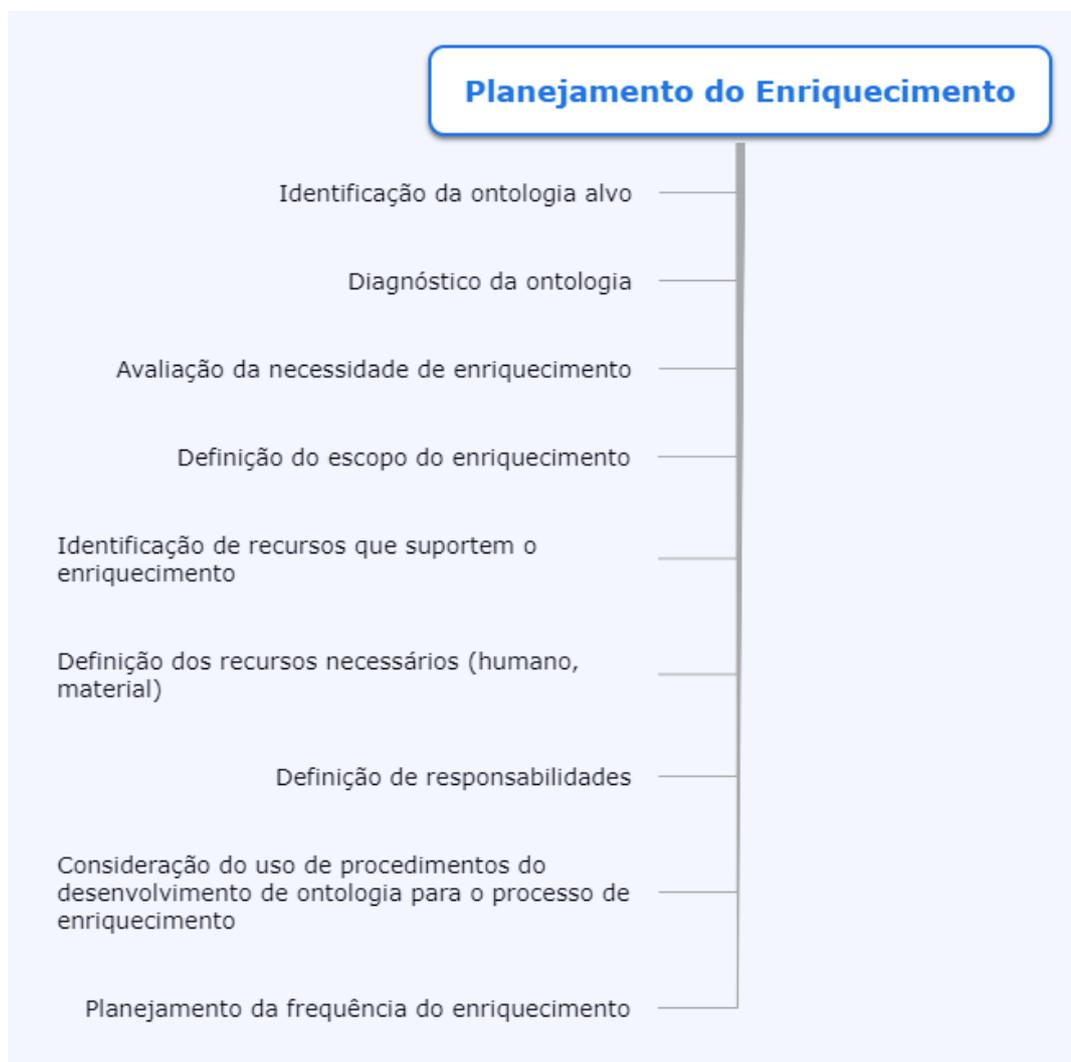
Nesta seção, apresentam-se os procedimentos resultantes da análise do material empírico (Grupos 1 e 2) organizados em categorias. A categorização desses procedimentos visa à integração deles, bem como fornecerá uma perspectiva de organização da metodologia de enriquecimento de acordo com as categorias estabelecidas aqui.

Como apontado anteriormente, para a criação das categorias, respaldou-se na técnica de Análise de Conteúdo. Destarte, as categorias foram desenvolvidas a partir do método indutivo. Após análise de cada procedimento e de sua respectiva definição, foram criadas cinco categorias que integram todos os procedimentos detectados nos documentos do Grupo 1 e do Grupo 2. São elas: **i) planejamento do enriquecimento; ii) aquisição de conhecimento; iii) organização do conhecimento; iv) implementação do enriquecimento; v) pós-enriquecimento.** Essas categorias se respaldam nos critérios de qualidade sugeridos por Bardin (2016, p. 76). Segundo a autora, "um conjunto de boas categorias deve possuir as seguintes qualidades: exclusão mútua, homogeneidade, pertinência, objetividade e fidelidade, produtividade".

Os procedimentos relacionados à categoria **Planejamento do Enriquecimento** são responsáveis pela preparação do processo de enriquecimento. É a partir dos procedimentos dessa categoria que será possível estabelecer as diretrizes para o enriquecimento. Assim, fazem parte dessa categoria os procedimentos: Identificação da ontologia alvo; Diagnóstico da ontologia; Avaliação

da necessidade de enriquecimento; Definição do escopo do enriquecimento; Identificação dos recursos que suportem o enriquecimento; Definição dos recursos necessários (humano e material); Definição de responsabilidades; Consideração do uso de procedimentos do desenvolvimento de ontologia para o processo de enriquecimento; Planejamento da frequência do enriquecimento.

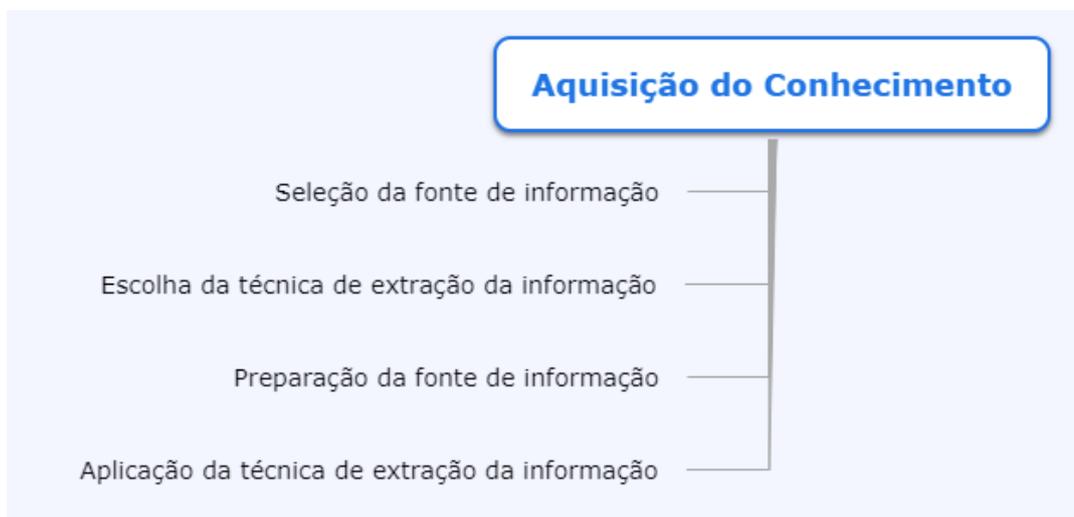
Figura 12 – Procedimentos da categoria Planejamento do Enriquecimento



Fonte: elaborado pelo autor.

A categoria **Aquisição de Conhecimento** trata da seleção e exploração de fonte de informação em busca de matéria-prima para a realização do processo de enriquecimento. Sendo assim, envolve os procedimentos: Seleção da fonte de informação; Escolha da técnica de extração da informação; Preparação da fonte de informação; Aplicação da técnica de extração da informação.

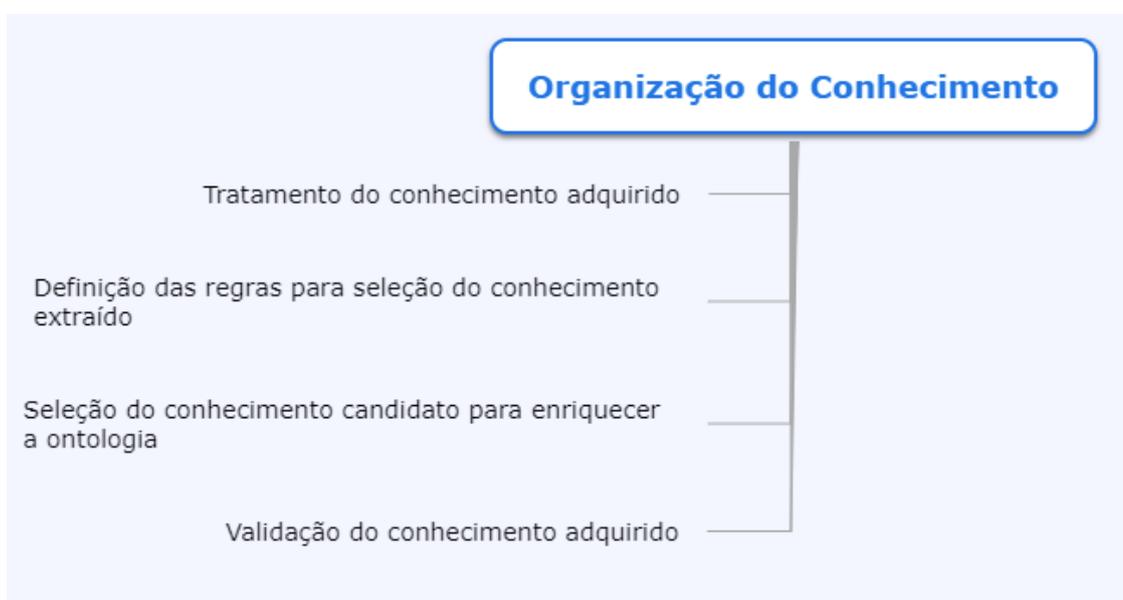
Figura 13 – Procedimentos da categoria Aquisição do Conhecimento



Fonte: elaborado pelo autor.

A categoria **Organização do Conhecimento** aborda os procedimentos concernentes à manipulação da matéria-prima selecionada na categoria Aquisição de Conhecimento e seleção daquele conteúdo que será pertinente para enriquecimento da ontologia. São procedimentos da categoria: Tratamento do conhecimento adquirido; Definição das regras para seleção do conhecimento extraído; Seleção do conhecimento candidato para enriquecer a ontologia; Validação do conhecimento adquirido.

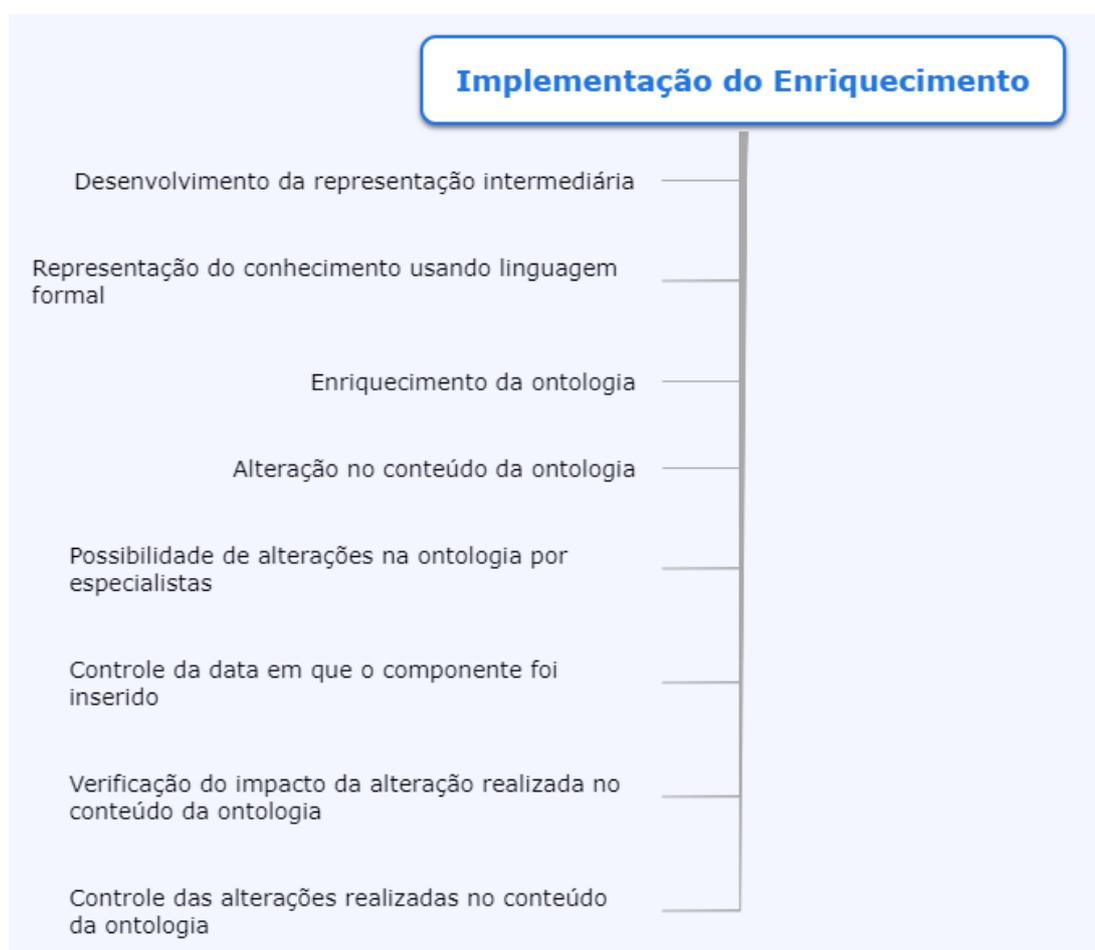
Figura 14 – Procedimentos da categoria Organização do Conhecimento



Fonte: elaborado pelo autor.

A categoria **Implementação do Enriquecimento** versa sobre os procedimentos responsáveis pela expansão da ontologia, ou seja, trata principalmente da inclusão de novos componentes na ontologia e pelo controle das alterações realizadas no conteúdo dela. A categoria é composta pelos procedimentos: Desenvolvimento de representação intermediária; Representação do conhecimento usando linguagem formal; Enriquecimento da ontologia; Alteração no conteúdo da ontologia; Possibilidade de alterações na ontologia por especialistas; Controle da data em que o componente foi inserido; Verificação do impacto da alteração realizada no conteúdo da ontologia; Controle das alterações realizadas no conteúdo da ontologia.

Figura 15 – Procedimentos da categoria Implementação do Enriquecimento

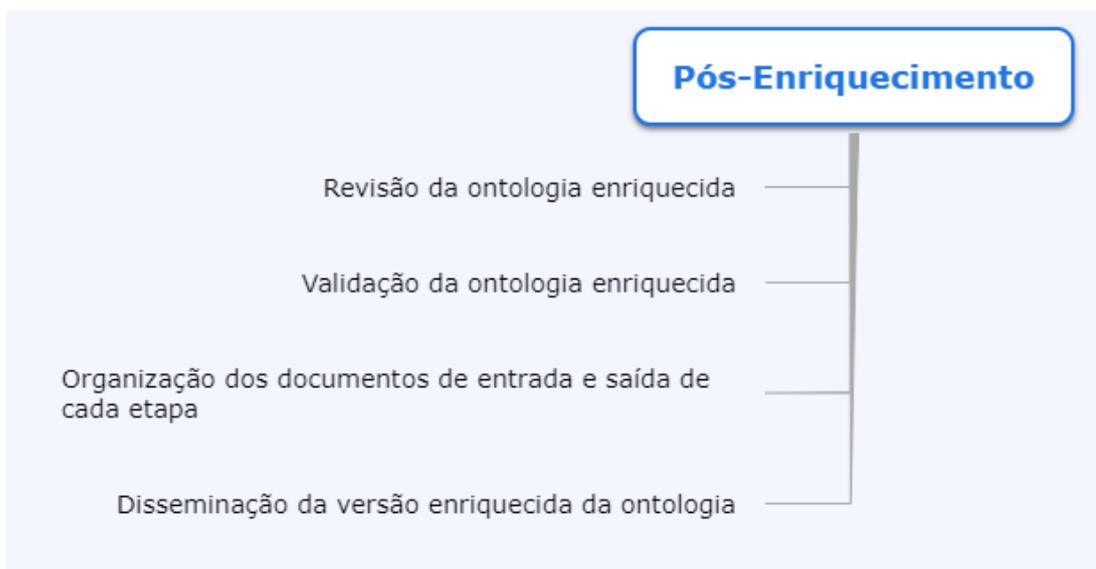


Fonte: elaborada pelo autor.

A categoria **Pós-enriquecimento** refere-se às atividades que devem ser executadas após a inserção de conteúdo na ontologia. São procedimentos dessa categoria: Revisão da ontologia enriquecida; Validação da ontologia enriquecida;

Organização dos documentos de entrada e saída de cada etapa; Disseminação da versão enriquecida da ontologia.

Figura 16 – Procedimentos da categoria Pós-Enriquecimento



Fonte: elaborada pelo autor.

Portanto essas foram as categorias e seus respectivos procedimentos extraídos dos 32 documentos (soma dos documentos dos Grupos 1 e 2) após a aplicação da Análise de Conteúdo. Como esses procedimentos estavam pulverizados em vários estudos, acredita-se que uma coleta daqueles mais pertinentes e a sua organização em uma única metodologia poderá facilitar a operacionalização do processo de enriquecimento, visto que se objetiva desenvolver uma metodologia em que os procedimentos possam ser generalizados para domínios diversos e não apenas focada nas singularidades de determinados domínios. Além disso, aspira-se que a metodologia seja simples para ontologistas e especialistas do domínio aplicarem o processo de enriquecimento nas ontologias.

5.4 Plano de ação para subsidiar o desenvolvimento da metodologia de enriquecimento

Os insumos para a elaboração da metodologia de enriquecimento foram provenientes de três fontes principais: 1) Revisão Sistemática da Literatura, na qual coletaram-se vários dados dos documentos que representam o estado da arte. Entre esses dados, estão: tipos de técnicas/métodos de extração da informação, tipos de

fontes de informação/conhecimento, tipos de enriquecimento, bem como as lacunas e os aspectos positivos identificados nos documentos; 2) Análise de conteúdo dos 29 documentos da revisão de literatura, na qual extraíram-se procedimentos que pudessem ser aplicados na metodologia de enriquecimento; 3) Análise de conteúdo das normas e do capítulo do livro de Soergel (1974), na qual também foram extraídos procedimentos passíveis de serem utilizados na metodologia de enriquecimento.

Para orientar o processo de desenvolvimento da metodologia de enriquecimento, desenvolveu-se um plano de ação com base nos dados e informações oriundos das três fontes principais citadas no parágrafo anterior. O plano de ação está apresentado no Quadro 6 e possui duas colunas: uma que descreve a funcionalidade requerida na ontologia e outra que se refere às ações ou tarefas que serão desenvolvidas para alcançar a funcionalidade equivalente.

A ideia de elaboração desse plano de ação partiu da leitura da pesquisa de Mendonça (2015), na qual o propósito do autor era o desenvolvimento de uma metodologia para construção de ontologias. Como o objetivo principal desta pesquisa também é o desenvolvimento de uma metodologia, acredita-se que a elaboração do plano de ação é um importante auxílio.

Quadro 6 – Plano de ação para desenvolvimento da metodologia de enriquecimento de ontologias de domínios

Continua...

Funcionalidade / Requisito funcional	Ações/tarefas
Especificação de atividades ligadas ao planejamento do enriquecimento	Criar etapa para identificação da ontologia alvo.
	Criar etapa para avaliação da necessidade de enriquecimento.
	Criar etapa para diagnóstico da ontologia.
	Criar etapa para definição do escopo do enriquecimento.
	Definir os recursos necessários para a realização do enriquecimento.
	Organizar os documentos de entrada e saída de cada etapa.
	Definir responsabilidades.

Funcionalidade / Requisito funcional	Ações/tarefas
Independência de ferramenta específica	Desenvolver as etapas da metodologia sempre levando em consideração a não dependência de uma ferramenta específica. Ex.: software. Softwares podem ser utilizados, mas a metodologia não deve depender de um software específico, podendo ser executada mesmo se o software for descontinuado.
Independência de domínio	Desenvolver etapas que possam ser generalizáveis para outros domínios.
Detalhamento das etapas da metodologia	Criar etapas bem detalhadas, se possível, com exemplificações.
Uso de fonte de informação de fácil acesso	Propor o uso de fontes de informação que sejam acessíveis no domínio da ontologia em enriquecimento.
Especificação de atividades de aquisição do conhecimento	<p>Criar etapa para seleção da fonte de informação.</p> <hr/> <p>Definir tipo de informação que deseja extrair.</p> <hr/> <p>Criar etapa para preparação da fonte de informação.</p> <hr/> <p>Criar etapa para escolha da técnica de extração da informação.</p> <hr/> <p>Selecionar ferramentas de suporte para a extração da informação.</p> <hr/> <p>Criar etapa para aplicação da técnica de extração da informação.</p>
Especificação de atividades de organização do conhecimento	<p>Criar etapa para tratamento do conhecimento adquirido.</p> <hr/> <p>Identificar a fonte de onde o conhecimento foi extraído.</p> <hr/> <p>Definir regras para seleção do conhecimento extraído.</p> <hr/> <p>Criar etapa para seleção do conhecimento candidato a enriquecer a ontologia.</p> <hr/> <p>Criar etapa para validação do conhecimento adquirido.</p>

Funcionalidade / Requisito funcional	Ações/tarefas
Enriquecimento de todos os componentes da ontologia	Incluir etapas que possibilitem o enriquecimento de todos os componentes da ontologia.
Especificação de atividades de implementação do enriquecimento	Identificar recursos que suportem o enriquecimento.
	Criar etapa para desenvolvimento de representações intermediárias.
	Criar etapa de inserção dos componentes nos locais corretos da ontologia (enriquecimento da ontologia).
	Verificar o impacto da alteração realizada no conteúdo da ontologia.
	Controlar as alterações realizadas no conteúdo da ontologia.
Especificação de atividades de pós-enriquecimento	Criar etapa de disseminação da versão enriquecida da ontologia.
	Documentar o processo de enriquecimento.
Aplicação de construtos teórico-metodológicos da BCI	Analisar cada etapa proposta na metodologia com o intuito de verificar qual construto da BCI poderia ser relevante.
	Propor o uso de construtos da BCI em cada etapa onde seja relevante.

Fonte: elaborado pelo autor.

Ao comparar os procedimentos extraídos do material empírico (apresentados nas seções 5.1 e 5.2) com as ações propostas no plano de ação, verifica-se a ausência de alguns procedimentos. As razões para isso são: (i) alguns procedimentos não foram considerados pertinentes para a metodologia que aqui se almeja. Como exemplo, o procedimento “Representação do conhecimento usando linguagem formal”, que não foi considerado, pois acredita-se que isso já está subentendido, uma vez que, ao enriquecer a ontologia, os componentes serão inseridos obedecendo à sintaxe de uma linguagem para representação de ontologias, tais como RDFS e OWL; (ii) outros procedimentos são muito específicos e já são contemplados por procedimentos mais amplos, como exemplo, o procedimento “avaliação da necessidade de inserção de determinado componente”, que é contemplado pelo procedimento “avaliação da necessidade de enriquecimento”.

Numa análise quantitativa das ações consideradas no plano de ação, identifica-se que 23 dos 30 procedimentos coletados no material empírico foram considerados. Isso representa uma porcentagem de 76,66% de aproveitamento. O plano de ação aborda ainda cinco ações originárias de lacunas identificadas na revisão sistemática da literatura. Por fim, informa-se que algumas ações foram inseridas com base nos estudos da literatura, nos quais identificou-se a utilidade delas para a metodologia de enriquecimento.

No próximo capítulo, apresenta-se a metodologia de enriquecimento de ontologias de domínio, resultado principal desta tese, e o produto gerado a partir dos insumos abordados neste capítulo.

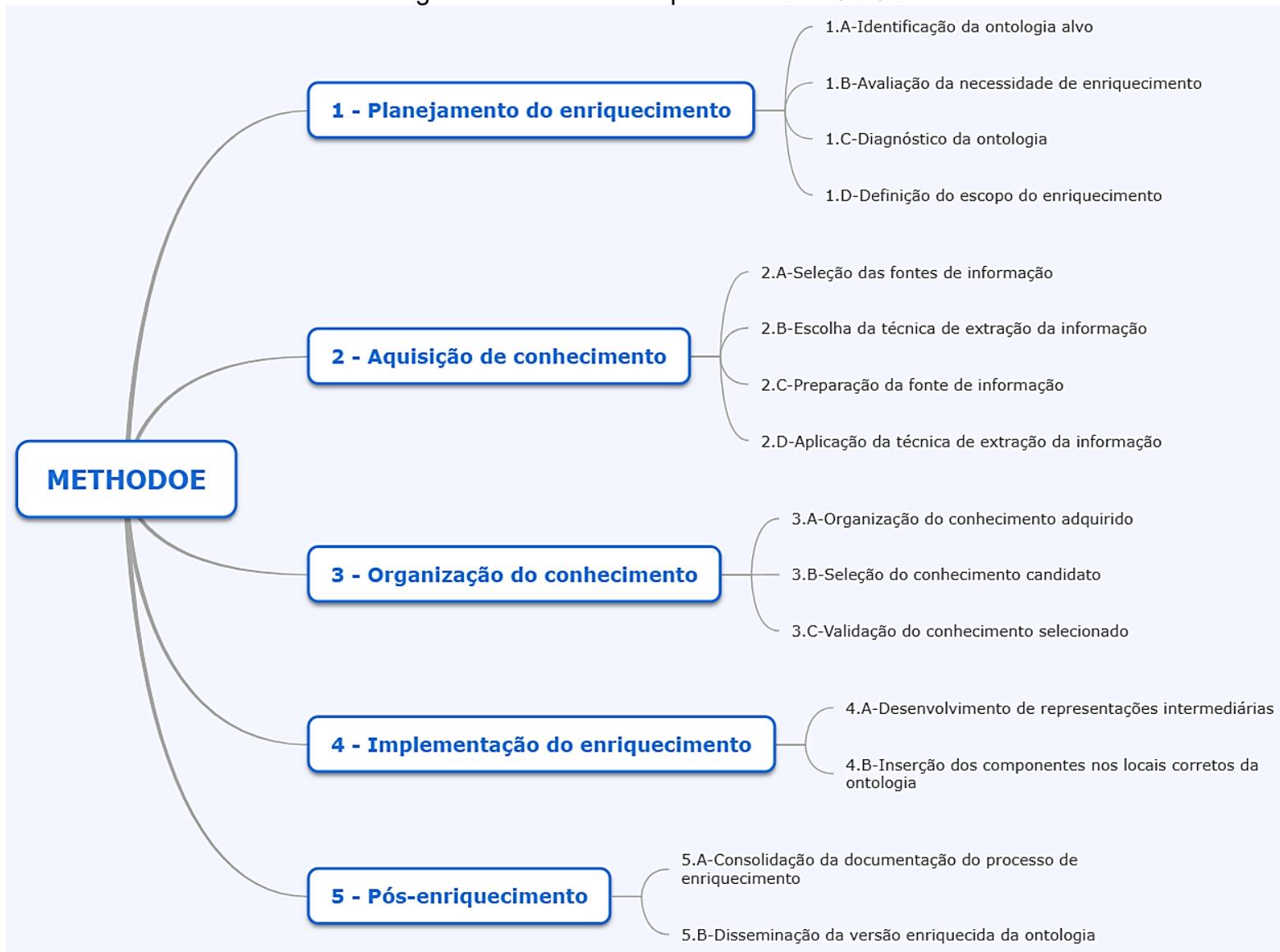
6 METODOLOGIA PARA ENRIQUECIMENTO DE ONTOLOGIAS DE DOMÍNIO (METHODOE)

Neste capítulo, apresenta-se a metodologia para enriquecimento de ontologias de domínio, denominada METHODOE²⁶. A METHODOE é composta de cinco fases principais que se desdobram em 15 etapas, conforme está retratado na Figura 17. O propósito da metodologia é mapear todo o processo de enriquecimento e facilitar a sua operacionalização por ontologistas e especialistas de domínio.

A denominação das cinco fases da METHODOE se derivaram da categorização dos procedimentos extraídos do material empírico ([descrita na seção 5.3](#)) e o desenvolvimento da metodologia de enriquecimento teve aporte no plano de ação elaborado na [seção 5.4](#), além dos estudos apresentados na revisão de literatura e nos fundamentos discutidos no referencial teórico.

²⁶ Methodology for Domain Ontology Enrichment.

Figura 17 – Fases e etapas da METHODOE



Fonte: elaborada pelo autor.

Com o intuito de apresentar a metodologia de forma sistemática, o presente capítulo foi dividido em seções de acordo com as cinco fases da METHODOE. Em cada seção, descrevem-se a fase, as etapas que a compõem e um fluxograma. Cada etapa obedece à seguinte estrutura: descrição da etapa, diretrizes para a sua realização e responsável pela implementação.

6.1 Planejamento do enriquecimento (Fase 1)

A primeira fase da METHODOE lida com a preparação do enriquecimento, sendo a base para identificar a direção desse processo. As etapas que compõem esta primeira fase são: **1.A – Identificação da ontologia alvo**, na qual apresenta-se e caracteriza-se a ontologia que será enriquecida; **1.B – Avaliação da necessidade de enriquecimento**, na qual se verifica se é preciso aplicar o enriquecimento; **1.C – Diagnóstico da ontologia**, que realiza uma análise mais específica da ontologia para identificar em quais partes o enriquecimento precisa ser aplicado e quais componentes precisam ser enriquecidos; **1.D – Definição do escopo do enriquecimento**, na qual se revela quais tipos de enriquecimentos serão realizados e em quais partes (classes ou subclasses) da ontologia. Ao final de cada etapa, são gerados documentos para serem consultados nas etapas seguintes. A seguir, apresenta-se o fluxograma desta fase, seguida do detalhamento das etapas.

Figura 18 – Fluxograma da Fase 1 – Planejamento do Enriquecimento



Fonte: elaborado pelo autor.

Etapa 1.A – Identificação da ontologia alvo

Descrição:

Nesta etapa, apresenta-se a ontologia alvo do processo de enriquecimento. Para isso, recomenda-se a identificação das informações principais referentes à ontologia, tais como: nome, breve descrição, domínio coberto pela ontologia, idioma no qual o conteúdo está descrito, objetivo da ontologia, questões de competência e fonte de onde se obtiveram as informações.

Essa etapa primária traz um conhecimento primordial para a realização das etapas seguintes, pois é nela que se conhecem as características da ontologia. A importância desta etapa pode ser notada por ser ela recorrente em vários estudos sobre enriquecimento de ontologias, tais como em: Guerram e Mellal (2018); Booshehri, Malekpour e Luksch (2013); Idoudi *et al.* (2018); Quesada-Martínez (2015); Ferré e Rudolph (2012).

Diretrizes:

Para identificação das informações, recomenda-se acessar os documentos de divulgação da ontologia, que podem estar em formato de artigos publicados em periódicos, em anais de eventos científicos ou em relatórios técnicos. Caso as informações coletadas sejam insuficientes, aconselha-se entrar em contato com os desenvolvedores da ontologia para acesso às informações necessárias. Além disso, é essencial acessar o arquivo com o código da ontologia, que contém toda a sua estrutura.

Outra alternativa é realizar a busca da ontologia em repositórios desses artefatos ou de Sistemas de Organização do Conhecimento (SOC) em geral, visto que esses repositórios apresentam alguns metadados referentes às ontologias, facilitando, dessa forma, a descoberta de informações. Entre os repositórios conhecidos, estão: BARTOC²⁷ (*Basel Register of Thesauri, Ontologies &*

²⁷ <https://bartoc.org/>

Classifications), OBO (*Open Biological and Biomedical Ontology*)²⁸, BioPortal²⁹, *Protégé Ontology Library*³⁰, AgroPortal³¹, entre outros.

Como resultado dessa etapa, gera-se um documento com as principais informações da ontologia (doc. 1 de acordo com o fluxograma da metodologia).

Responsável pela etapa: ontologista / especialista do domínio (opcional)

Etapa 1.B – Avaliação da necessidade de enriquecimento

Descrição:

Nesta etapa, analisa-se a capacidade de representação da ontologia quanto ao seu objetivo e questões de competência, tendo como base os métodos da área de avaliação de ontologias. Essa análise também envolve a completude e a atualização da ontologia de acordo com o domínio que modela. Por meio dela, decide-se sobre a necessidade de enriquecimento da ontologia.

Diretrizes:

Os insumos para a realização desta etapa referem-se ao documento com o código da ontologia e ao documento 1 (apresentação da ontologia).

A primeira diretriz para realizar esta etapa refere-se à verificação da ontologia para saber se ela representa adequadamente os conceitos presentes nas questões de competência e nos objetivos de construção. Novas questões de competência também podem ser formuladas de acordo com o escopo de modelagem da ontologia. Se a ontologia não consegue responder a todas as questões de competência, provavelmente será necessário passar pelo processo de enriquecimento. Essa análise da estrutura pode ser realizada com o uso de softwares editores de ontologias: Protégé³², Fluent Editor³³, Hozo³⁴, Onto4all³⁵, entre vários outros.

²⁸ <http://www.obofoundry.org/>

²⁹ <https://bioportal.bioontology.org/>

³⁰ https://protegewiki.stanford.edu/wiki/Protege_Ontology_Library#OWL_ontologies

³¹ <http://agroportal.lirmm.fr/>

³² <https://protege.stanford.edu/>

³³ <https://www.cognitum.eu/semantics/FluentEditor/>

³⁴ <http://www.hozo.jp/>

³⁵ <https://onto4allditor.com/en>

Outra forma de verificar se é necessário aplicar o processo de enriquecimento é comparar a ontologia com a literatura do domínio. Para isso, podem-se aplicar métodos da área de avaliação de ontologia (MAEDCHE e STAAB, 2002; GANGEMI *et al.*, 2006; BRANK, GROBELNIK e MLADENIC, 2005; DELLSCHAFT e STAAB, 2006; ZAVITANOS, PALIOURAS e VOUIROS, 2011; BREWSTER *et al.*, 2004; ELHADAD, GABAY e NETZER, 2010; OUYANG e QU, 2011; HLOMANI e STACEY, 2013; HLOMANI e STACEY, 2014). Como o conhecimento é dinâmico, a ontologia precisa ser atualizada para acompanhar essa dinamicidade.

Ainda como forma de verificar a necessidade de enriquecimento, pode-se solicitar o auxílio de especialistas do domínio para avaliar o conteúdo da ontologia e analisar se é necessária ou não a realização do enriquecimento dela. Entretanto, para que isso seja realizado, muitas vezes é preciso um esforço do ontologista para transformar a representação existente na estrutura da ontologia em uma forma mais intuitiva para os especialistas do domínio. Isso pode ser feito traduzindo-se o conhecimento representado na ontologia para linguagem natural, em forma de frases. O uso de recursos como o mapa conceitual também pode ser um facilitador. Existem vários métodos de avaliação de ontologias com base na análise humana, entre eles: Lozano-Tello e Gómez-Pérez (2004), Almeida (2009) e Orme, Yao e Etkorn (2007).

Em síntese, se a ontologia apresenta problemas de incompletude ou não representa conceitos do domínio, então há indícios de que é necessário enriquecê-la. Esses indícios são explorados na próxima etapa (diagnóstico da ontologia).

Após a decisão sobre a necessidade ou não da realização do processo de enriquecimento, deve-se gerar um documento com o resultado da avaliação (Doc. 2). O Quadro 7, a seguir, apresenta um modelo desse documento.

Quadro 7 – Modelo de documento referente à avaliação da necessidade de enriquecimento

É necessário realizar o processo de enriquecimento?	Método de avaliação utilizado	Justificativa
Sim	Análise das questões de competência. Análise da estrutura taxonômica da ontologia.	Justificativa 1: Algumas questões de competência não podem ser respondidas pela ontologia, por falta de representação de conceitos fundamentais dessas QCs. Justificativa 2: Após análise da estrutura taxonômica da ontologia, constatou-se que algumas classes da ontologia não estão especializadas.

Fonte: elaborado pelo autor.

Responsável pela etapa: ontologista / especialista do domínio (opcional)

Etapa 1.C – Diagnóstico da ontologia

Descrição:

Esta etapa refere-se a um exame minucioso na estrutura da ontologia para determinar em quais partes o enriquecimento é necessário e quais componentes (classes, subclasses, relações) devem ser contemplados. Logo o resultado da etapa progressa é um ponto de partida para a realização do diagnóstico.

Diretrizes:

Para realização do diagnóstico da ontologia, primeiramente, deve-se acessar o arquivo que contém o código dela. Em seguida, deve-se acessar o documento gerado na Etapa 1.B, pois ele contém informações sobre os problemas identificados na ontologia, sendo um ponto de partida para o diagnóstico.

O diagnóstico pode ser realizado por meio de diferentes estratégias. Uma delas é checar toda a estrutura da ontologia, cada classe, propriedade, relacionamento e

axioma com o apoio de um especialista do domínio, com a finalidade de identificar partes nas quais melhorias sejam necessárias.

Outra possibilidade é usar os *pitfalls* (erros comuns) para diagnóstico de ontologias (apresentado no [Capítulo 3](#)), desenvolvido por Poveda-Villalón (2016). Esses *pitfalls* lidam com erros comuns feitos por ontologistas no momento de desenvolvimento da ontologia. Sendo assim, alguns desses erros são indicações para a realização do processo de enriquecimento. O método de diagnóstico da autora possui um software, batizado de OOPS (*Ontology Pitfall Scanner*)³⁶, que visa facilitar a análise da ontologia. O software possui uma versão Web, na qual é possível colar o código RDF da ontologia para realizar uma análise automática dos erros existentes. Dentre os 41 *pitfalls* apresentados pela autora, Poveda-Villalón, oito deles possuem relação com o enriquecimento de ontologias. Eles estão descritos no Quadro 8 a seguir.

Quadro 8 – Pitfalls para diagnóstico de ontologia aplicáveis ao processo de enriquecimento

			Continua...
Pitfall	Descrição do <i>pitfall</i>	Detalhes do <i>pitfall</i>	
P08. <i>Missing annotations</i> (Falta de anotações)	Um elemento é criado na ontologia, porém não se proveem anotações anexadas a ele. Rótulos como <code>rdfs:label</code> , <code>rdfs:comment</code> , <code>skos:prefLabel</code> , etc.	ver Anexo A	
P09. <i>Missing domain information</i> (Falta de informação do domínio)	Parte da informação necessária para a modelagem do domínio não está incluída na ontologia.	ver Anexo B	
P10. <i>Missing disjointness</i> (Falta do axioma de disjunção)	A ontologia não apresenta o axioma de disjunção entre classes ou entre relações que deveriam ser definidas como disjuntas.	ver Anexo C	
P11. <i>Missing domain or range in properties</i> (Falta de domínio e imagem nas relações)	As relações entre dois componentes não possuem domínio e imagem, ou seja, não é possível saber a direção de determinada relação.	ver Anexo D	
P12. <i>Equivalent properties not explicitly declared</i> (Propriedades equivalentes não estão declaradas)	A ontologia não informa a equivalência entre relacionamentos e/ou propriedades (atributos).	ver Anexo E	

³⁶ <http://oops.linkeddata.es/>

Conclusão.

Pitfall	Descrição do pitfall	Detalhes do pitfall
P13. <i>Inverse relationships not explicitly declared</i> (Relacionamentos inversos não estão declarados)	A ontologia não explicita as relações inversas. Ex: a relação temLínguaOficial possui como relação inversa éLínguaOficialde .	ver Anexo F
P14. <i>Missing "owl:allValuesFrom"</i> (Falta da restrição owl:allValuesFrom)	Uso inadequado de restrição universal.	ver Anexo G
P30. <i>Equivalent classes not explicitly declared</i> (Classes equivalentes não estão declaradas)	A ontologia não explicita a equivalência entre classes.	ver Anexo H

Fonte: elaborado pelo autor.

Nesta etapa, gera-se um relatório detalhado com o resultado do diagnóstico, identificando principalmente a localização dos problemas na estrutura da ontologia. A seguir, um exemplo de como esse relatório pode ser apresentado.

Quadro 9 – Exemplo de relatório com resultado do diagnóstico (documento 3)

Classes / subclasses	Anotações	Relações	Axiomas
Classe X precisa ser especializada (<i>pitfall</i> 09)	Faltam anotações para a Classe Y (<i>pitfall</i> 08)	Relação A que associa a Classe Z e W não define o que é <i>domain</i> e o que é <i>range</i> (<i>pitfall</i> 11)	Definir disjunção entre Classe F e G (<i>pitfall</i> 10)
Classe H deve ser especializada para representar informações das questões de competência (<i>pitfall</i> 09)	Falta anotação "definição" para a relação R (<i>pitfall</i> 08)	Falta definir relação entre Classe H e K.	Falta definir propriedade lógica para a relação P.

Fonte: elaborado pelo autor.

Responsável pela etapa: ontologista / especialista do domínio (opcional)

Etapa 1.D – Definição do escopo do enriquecimento

Descrição:

Nesta etapa, define-se qual será o escopo do enriquecimento, isto é, o objetivo e os limites do processo de enriquecimento. Desse modo, devem ser indicados os componentes da ontologia que serão enriquecidos (tipos de enriquecimento), bem como as partes da ontologia em que o enriquecimento deverá acontecer.

O escopo do enriquecimento está intrinsecamente conectado com o documento de identificação da ontologia (doc. 1), que especifica o objetivo da ontologia, as questões de competência e o domínio de cobertura. Assim, será possível determinar o nível de granularidade do conhecimento do domínio que a ontologia deve abranger. Além disso, o escopo do enriquecimento também está associado ao documento gerado na Etapa 1.C – Diagnóstico da ontologia, visto que é a partir desse documento que será possível determinar os pontos exatos nos quais o enriquecimento deverá acontecer.

Diretrizes:

O primeiro passo para a definição do escopo é acessar e analisar o documento com as informações da ontologia (doc. 1) e o documento com o resultado do diagnóstico da ontologia (doc. 3).

A partir dos resultados gerados na etapa de Diagnóstico da Ontologia e considerando sempre o objetivo de representação da ontologia, preenche-se um quadro com as informações de quais componentes serão enriquecidos. Esse quadro pode ser dividido em quatro tipos de enriquecimento, a saber: (1) enriquecimento lexical, (2) enriquecimento conceitual, (3) enriquecimento de relações e (4) enriquecimento de axiomas. Nesse documento, deve estar claro em quais partes da ontologia será aplicado o processo de enriquecimento, bem como quais componentes serão enriquecidos.

Ao final, esta etapa deve gerar um documento com informações bem detalhadas para a condução do processo de enriquecimento, posto que funcionará como apoio para tomada de decisões nas fases seguintes. Em síntese, este documento funcionará como um guia para o processo de enriquecimento, pois contém

informações sobre os limites do enriquecimento, grau de granularidade e tipos de componentes que deverão ser enriquecidos.

Quadro 10 – Exemplo do documento de escopo do enriquecimento

Enriquecimento conceitual	Enriquecimento lexical	Enriquecimento de relações	Enriquecimento de axiomas	Observações
Classe X precisa ser especializada.	Anotações (sinônimos) para a Classe X.	Relação da Classe X com outras Classes da ontologia.	Definição de equivalência da Classe X com Classe Y. Definição de disjunção entre Classe X e Classe Z.	Especialização da Classe X considerando o grau de granularidade da ontologia.

Fonte: elaborado pelo autor.

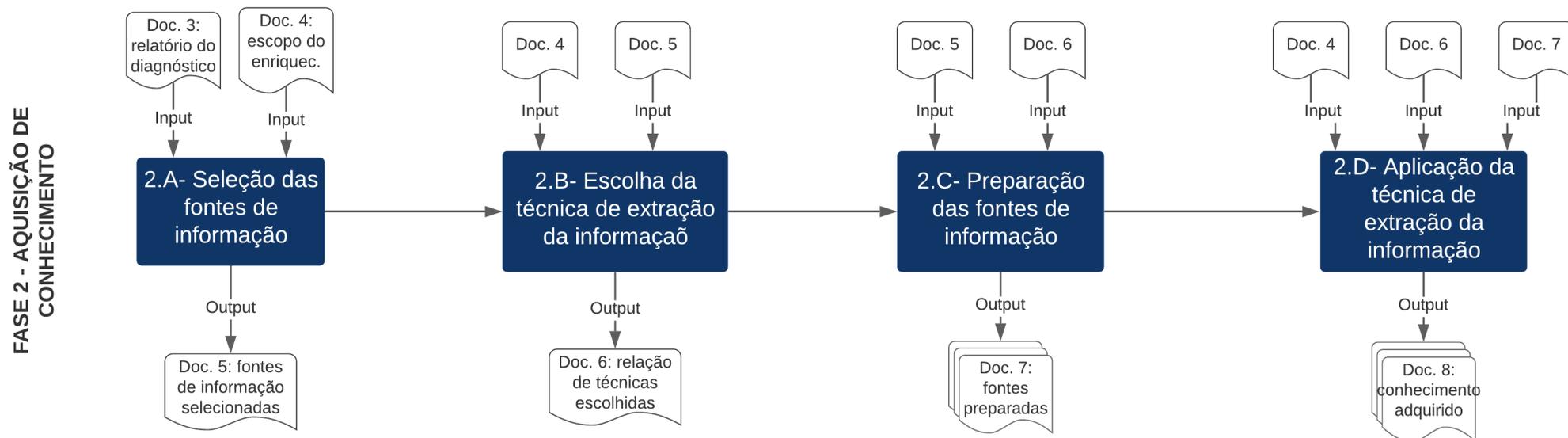
Responsável pela etapa: ontologista

6.2 Aquisição de Conhecimento (Fase 2)

Após a fase de planejamento, na qual se estabelece a direção que o processo de enriquecimento irá seguir, passa-se à fase de Aquisição de Conhecimento. Nessa fase, o objetivo é identificar insumos (matéria-prima) para realizar o processo de enriquecimento. Nessa perspectiva, a fase de Aquisição de Conhecimento conta com as seguintes etapas: **2.A – Seleção das fontes de informação**, na qual o propósito é identificar fontes passíveis de insumos para realização do enriquecimento da ontologia; **2.B – Escolha da técnica de extração da informação**, que se refere à seleção de qual estratégia será utilizada para extrair o conhecimento das fontes selecionadas; **2.C – Preparação da fonte de informação**, na qual o intuito é preparar as fontes de informação de acordo com a técnica de extração escolhida; **2.D – Aplicação da técnica de extração da informação**, a qual trata da extração de conteúdo nas fontes preparadas.

Esta fase pode ser considerada transversal às demais, apesar de ser apontada na METHODOE como fase 2, visto que, durante as outras etapas do enriquecimento, podem surgir dúvidas sobre um conceito, relação, definição; e seja necessário consultar o conhecimento adquirido nesta fase. A seguir, ela está resumida em um fluxograma.

Figura 19 – Fluxograma da Fase 2 – Aquisição de Conhecimento



Fonte: elaborada pelo autor.

Etapa 2.A – Seleção das fontes de informação

Descrição:

Nesta etapa, o objetivo é identificar fontes de informação passíveis de insumos para realização do processo de enriquecimento. As fontes de informação podem ser documentos textuais, como artigos científicos, livros, relatórios técnicos. Todavia podem-se extrair insumos a partir do reuso de Sistemas de Organização do Conhecimento (SOC), como também pode-se entrevistar especialistas de domínio para a obtenção de conhecimento referente ao domínio.

Diretrizes:

O primeiro passo para realização desta etapa é verificar os documentos oriundos da Etapa 1.C (Diagnóstico da ontologia) e 1.D (Definição do escopo do enriquecimento). Esses documentos contêm informações que auxiliam na seleção das fontes de informação, dado que os conceitos das questões de competência bem como os rótulos das classes (aquelas identificadas no diagnóstico) podem ser utilizados para pesquisa das fontes de informação.

Em seguida, deve-se definir o tipo de fonte de informação que se deseja utilizar. As fontes textuais são as mais utilizadas (como nos estudos de PETERS, 2009; ASTRAKHANTSEV, FEDORENKO e TURDAKOV, 2014; GUERRAM e MELLAL, 2018; ALI *et al.*, 2019), devido à sua disponibilidade em vários domínios do conhecimento. Entre essas fontes, podem ser utilizados os manuais, normas, relatórios técnicos, documentos científicos, entre outros. O levantamento (pesquisa bibliográfica) dessas fontes pode ser realizado em bases de dados, repositórios e bibliotecas digitais do domínio que a ontologia representa. Nesta etapa, as experiências na recuperação da informação utilizadas na BCI podem ser empregadas para identificação das fontes mais adequadas de acordo com a necessidade de informação, ou seja, extrair conhecimento que tenha relação com o escopo do enriquecimento. Para verificar se determinado documento tem relação com a temática com a qual se tem interesse, pode-se utilizar o processo de Análise de Assunto para auxiliar no exame e apreensão do conteúdo de documentos, por exemplo, selecionar documentos que tratam apenas de agricultura familiar.

A qualidade das fontes de informação também é um aspecto importante a ser considerado. Portanto deve-se optar por fontes que possuam credibilidade no

domínio. Isso pode ser feito com a utilização de periódicos científicos com bons indicadores para o caso de busca de artigos científicos. Fontes de referência, manuais e livros técnicos consolidados na área também são fontes confiáveis. Nesse ponto, a consulta a especialistas de domínio pode poupar tempo e garantir o acesso a fontes de informação pertinentes para o processo de enriquecimento.

O conhecimento do domínio também pode ser apreendido com o reúso de SOCs que representam domínios próximos da ontologia. Isso pode ser feito consultando-se repositórios de SOCs (citados na Etapa 1.A). Alguns desses repositórios disponibilizam a consulta por temáticas ou por áreas de conhecimento.

Os especialistas de domínio também podem contribuir para a captura de conhecimento para a realização do enriquecimento, sendo, portanto, uma espécie de fonte também.

É importante que, ao selecionar as fontes, tenha-se noção das técnicas de extração que serão utilizadas nessas fontes. Há uma ligação intrínseca entre as fontes de informação e as técnicas de extração da informação (abordadas na próxima etapa), visto que determinadas técnicas só podem ser aplicadas em tipos específicos de fontes. Por exemplo, se uma das fontes selecionadas são especialistas do domínio, as técnicas que poderão ser utilizadas estão relacionadas a entrevistas, técnica de *brainstorming* e aplicação de questionários.

Ao final desses procedimentos, organizam-se as fontes de informação de acordo com suas tipologias, com seus dados de identificação e a justificativa para seleção, conforme exemplificado no Quadro 11.

Quadro 11 - Exemplo de documento com o resultado da seleção das fontes de informação

Fonte	Referência ou endereço de acesso	Motivo para escolha
AGROVOC	http://aims.fao.org/vest-registry/vocabularies/agrovoc	SOC relacionado ao domínio da ontologia.
Artigo	https://bit.ly/3by5LgY	Relacionado ao tema geral da ontologia.

Fonte: elaborado pelo autor.

Responsável pela etapa: ontologista / especialista do domínio (opcional)

Etapa 2.B – Escolha da técnica de extração da informação

Descrição:

Após a seleção das fontes, esta etapa preocupa-se com a escolha da técnica que será empregada para extrair a informação das fontes. As técnicas de extração da informação podem ser classificadas em manuais, semiautomáticas e automáticas. As manuais são aquelas em que toda a análise é realizada por humanos. As técnicas automáticas e semiautomáticas envolvem sistemas ou algoritmos computacionais e podem ser classificadas de acordo com o método que empregam.

Diretrizes:

Inicialmente, para realização desta etapa, consulta-se novamente o documento resultante da Etapa 1.D – Escopo do enriquecimento, para saber quais tipos de componentes serão enriquecidos. Logo depois, analisa-se o documento gerado na Etapa 2.A, que contém a lista de fontes de informação selecionadas.

Em seguida, deve-se escolher a técnica de extração da informação que será utilizada nas fontes selecionadas na etapa prévia. Salienta-se que cada tipo de fonte de informação pode exigir uma técnica específica. Para fontes textuais, como artigos científicos, podem-se empregar técnicas de Processamento de Linguagem Natural (PLN), de Aprendizagem de Máquina ou análise manual, por exemplo.

Na METHODOE, advoga-se que não existe um método perfeito ou uma técnica perfeita, visto que existem muitas variáveis, entre elas o tipo de fonte de informação acessível no domínio do conhecimento, a disponibilidade de um profissional com expertise necessária para aplicar determinado método, o idioma da ontologia, o tipo de enriquecimento que será aplicado. Portanto, recomenda-se analisar o [Apêndice B – Dados sobre documentos da Revisão de Literatura](#), que apresenta um quadro com as referências de estudos que aplicaram determinados métodos de acordo com o tipo de enriquecimento que estava sendo realizado. Ressalta-se que tal quadro visa apenas exemplificar esta etapa com alguns métodos que foram aplicados em estudos de caso, não tendo a intenção de apresentar uma lista cabal com as técnicas de extração da informação. O intuito é que o ontologista tenha conhecimento das técnicas e faça uma análise (considerando as particularidades do seu contexto) da técnica que mais se adéqua à ontologia de domínio que ele está enriquecendo. A partir dos dados

apresentados no [Apêndice B](#), podem-se empregar técnicas que apresentam similaridades com a ontologia que se pretende enriquecer e com o escopo de enriquecimento delineado. A seguir é exibido um recorte dos dados apresentados no [Apêndice B](#).

Quadro 12 – Recorte dos dados apresentados no Apêndice B

Referência	Tipo de fonte de informação	Tipo de enriquecimento					Técnica aplicada
		Lexical	Conceitual	Relação		Axioma	
				Taxonômica	Não taxonômica		
KUNTARTO, ISYHRANI e GUNAWAN (2019)	Páginas Web		x				Algoritmo de Clustering K-Means; TF-IDF
IDOUDI et al (2018)	ontologia		x				Hierarchical Fuzzy Clustering
KOKLA e TOMAI (2018)	ontologia; Páginas Web		x	x	x		<u>análises</u> das definições dos conceitos; análise sintática (tokenização, part-of-speech (POS) tagging, divisão de sentenças e lematização)

Fonte: elaborado pelo autor.

Percebe-se, por meio do recorte apresentado, que o estudo de Kokla e Tomai (2018) tinha como fonte de informação ontologias e páginas Web, e os tipos de enriquecimentos são: conceitual e de relações (definido no escopo do enriquecimento). Para tanto, empregaram-se técnicas de análise manual das definições e dos conceitos da ontologia e técnicas ligadas à área de PLN, tais como tokenização, POS *tagging* e lematização. Portanto, se algum ontologista deseja enriquecer uma ontologia em circunstâncias similares (fontes de informação do mesmo tipo, objetivo de realizar enriquecimento do mesmo tipo) às apresentadas por Kokla e Tomai (2018), a utilização das técnicas empregadas pelos autores pode ser um caminho viável.

Por fim, ressalta-se que o uso de mais de uma técnica para extração da informação também é válido e pode até ser mais eficaz do que o uso de apenas uma técnica. Assim pode-se combinar a análise manual de fontes de informação com a extração automática, por exemplo.

Quadro 13 - Exemplo de documento resultante da etapa de escolha da técnica de extração

Técnica de extração escolhida	Aplicável a qual tipo de fonte de informação	Motivo da escolha
Análise manual	Corpus textual	Foram selecionadas apenas fontes de informação do tipo artigo científico. Além disso, o número de fontes selecionadas é pequeno, permitindo assim que seja possível a análise manual de cada fonte.

Fonte: elaborado pelo autor.

Responsável pela etapa: ontologista / profissional da Ciência da Computação (opcional)

Etapa 2.C – Preparação da fonte de informação

Descrição:

Esta etapa lida com a adequação da fonte de informação ao método ou técnica de extração da informação que será empregada, dado que cada tipo de técnica exigirá que a fonte seja preparada de uma forma específica. Portanto, a depender do tipo de técnica selecionada, será necessário o apoio técnico de um profissional que possua familiaridade com ela, posto que a preparação da fonte de informação pode ser complexa e envolver a configuração de alguma ferramenta.

Diretrizes:

Os documentos de entrada (*input*) desta etapa são aqueles gerados nas Etapas 2.A (seleção das fontes de informação) e 2.B (escolha da técnica de extração da informação).

Após acessar os documentos, a próxima atividade é conhecer os procedimentos da técnica de extração da informação. Em seguida, prepara-se a fonte de informação de acordo com as exigências da técnica selecionada. Por exemplo, se a fonte de conhecimento selecionada são artigos científicos e optou-se por uma técnica estatística que calcula a frequência absoluta dos termos, então deve-se preparar a fonte de acordo com os requisitos da ferramenta que executa a técnica. Geralmente, o corpus textual precisa estar em um formato básico, como o .txt, em que não há formatação.

Dessa forma, para preparar a fonte de informação de forma adequada, deve-se conhecer bem a técnica de extração que se estará aplicando. Nesse sentido, o acesso a outros estudos que tenham aplicado tal técnica também pode oferecer pistas relevantes para a preparação da fonte.

O resultado desta etapa são as fontes de informação selecionadas na Etapa 2.A e preparadas de acordo com as exigências da técnica de extração da informação.

Responsável pela etapa: ontologista; profissional da área de computação (de acordo com a demanda, ou seja, se a técnica de extração exigir uma configuração mais avançada).

Etapa 2.D – Aplicação da técnica de extração da informação

Descrição:

Esta etapa trata da aplicação da técnica de extração da informação nas fontes selecionadas e preparadas. É a etapa principal da fase de Aquisição de Conhecimento, visto que irá gerar como resultado elementos passíveis de serem utilizados no enriquecimento da ontologia.

Nesta etapa, também devem ser selecionadas ferramentas que facilitem a aplicação da técnica e definidas as regras para extração da informação. Existem muitas ferramentas (softwares e repositórios de ferramentas) que auxiliam as técnicas de Processamento de Linguagem Natural e de análise estatística, tais como: *Sketch Engine*³⁷, *OpenNLP*³⁸, *Natural Language Toolkit*³⁹, *GATE*⁴⁰, *Stanford CoreNLP*⁴¹, etc. Também é possível utilizar linguagens como o Python⁴² e R⁴³ para desenvolver um algoritmo próprio, visto que existem muitas bibliotecas com códigos que podem ser adaptados. As regras para extração referem-se aos critérios que serão observados para se examinar aquela informação extraída. Por exemplo, pode-se considerar apenas os substantivos extraídos de determinado *corpus*, posto que eles têm possibilidades de representar conceitos para a ontologia; ou podem-se considerar apenas os verbos de um dado *corpus*, já que podem ser indicativos de relações para a ontologia.

Diretrizes:

A implementação desta etapa se respalda nos documentos gerados nas etapas 1.D – Escopo do enriquecimento, que contribuirá para saber o que se deseja extrair das fontes; 2.B – Escolha da técnica de extração da informação, que aponta a técnica escolhida acompanhada do motivo de escolha; 2.C – Preparação da fonte de informação, que contém as fontes de informação tratadas de acordo com as diretrizes

³⁷ <https://www.sketchengine.eu/>

³⁸ <https://opennlp.apache.org/>

³⁹ <https://www.nltk.org/>

⁴⁰ <https://gate.ac.uk/>

⁴¹ <https://stanfordnlp.github.io/CoreNLP/>

⁴² <https://www.python.org/>

⁴³ <https://www.r-project.org/>

da técnica. Esses documentos servirão de guia para a aplicação da técnica de extração.

A partir das fontes de informação preparadas e da técnica de extração selecionada, inicia-se a extração da informação. Se a técnica for totalmente manual, deve-se analisar diretamente as fontes com o respaldo das diretrizes da técnica. Caso seja um método automático ou semiautomático, devem-se pesquisar e configurar as ferramentas que servirão de suporte para a implementação da técnica. A pesquisa e configuração da ferramenta podem ser auxiliadas por um especialista da computação, caso necessário. Também devem ser determinadas as regras para extração da informação.

Antes da execução definitiva da extração do conhecimento, recomenda-se a realização de um pré-teste para verificar se a fonte de informação foi preparada corretamente e se a ferramenta que dá suporte à técnica de extração foi configurada adequadamente (quando se trata de um método automático ou semiautomático); assim como se as regras para extração estão adequadas. A partir do pré-teste, será possível corrigir as eventuais falhas e então executar a extração da informação.

Os documentos resultantes da aplicação da técnica de extração não seguem um padrão predefinido e vão depender muito do escopo do enriquecimento, do tipo de fonte de informação explorada e principalmente da técnica de extração aplicada. Portanto recomenda-se a organização da informação extraída em tabelas ou quadros de acordo com os tipos de enriquecimento previstos no documento de escopo.

A título exemplificativo, numa situação A em que o escopo do enriquecimento aponta para o enriquecimento conceitual, a fonte de informação é um *corpus* textual de artigos científicos, e a técnica de extração da informação aplicada refere-se à análise estatística do tipo frequência do termo com a regra de extração dos termos que aparecem no mínimo 50 vezes. Como resultado, gera-se uma tabela com a listagem de termos na primeira coluna e sua respectiva frequência na segunda coluna. Em outra situação, B, poderia-se ter como escopo o enriquecimento conceitual e de relações; a fonte de informação é a mesma da situação A, e as técnicas de extração aplicadas referem-se à análise morfológica e à frequência do termo com a mesma regra de extração da situação A. Então, como resultado, ter-se-ia uma tabela com a lista de termos. Na primeira coluna, ter-se-ia o termo extraído; na segunda coluna, o grupo linguístico ao qual o termo pertence (substantivo, adjetivo, verbo) e, na terceira coluna, poderia-se ter a frequência do termo no *corpus* analisado. Dessa forma, os

substantivos poderiam ser indicativos de conceito para a ontologia, bem como os adjetivos (pois podem ser substantivados) e os verbos poderiam ser indicativos de relações.

A Tabela 3 apresenta um exemplo de documento que considera as variáveis apresentadas da situação B. Enfatiza-se que, a depender do tipo de fonte de informação e da técnica empregada para captura do conhecimento, podem-se gerar diversificados formatos de documentos. Para ilustrar isso, o Quadro 4 apresenta outro exemplo de documento.

Caso o conhecimento extraído após a aplicação das técnicas de extração se mostre insuficiente (escasso), recomenda-se a realização de uma nova iteração a partir da Etapa 2.A – Seleção da fonte de informação. Assim, escolhem-se novas fontes e aplicam-se novamente as técnicas de extração da informação ou escolhem-se outras técnicas.

Tabela 3 – Exemplo 1 de documento resultante desta etapa

Termo extraído⁴⁴	Grupo linguístico (morfologia)	Frequência
Erosão	substantivo	1500
Poluição	substantivo	1000
Influencia	verbo	500

Fonte: elaborado pelo autor.

No Quadro 4 apresenta-se um exemplo de documento em que a técnica de extração é a análise manual de textos, com a extração de trechos relevantes.

Quadro 14 – Exemplo 2 de documento resultante desta etapa

Continua...

Documento analisado	Trecho extraído	Conceitos relevantes
Potencial de impacto da agricultura sobre os recursos hídricos na Região do Cerrado	Eutrofização causada pelo enriquecimento dos corpos hídricos com fósforo e nitrogênio.	Eutrofização; corpos hídricos; fósforo; nitrogênio.

⁴⁴ Os dados apresentados no quadro são hipotéticos.

Potencial de impacto da agricultura sobre os recursos hídricos na Região do Cerrado	Deposição de sedimentos carreados diminuindo a capacidade dos reservatórios (assoreamento).	Deposição; sedimentos carreados; reservatórios; assoreamento.
---	---	---

Conclusão.

Documento analisado	Trecho extraído	Conceitos relevantes
Potencial de impacto da agricultura sobre os recursos hídricos na Região do Cerrado	Erosão pode ajudar a levar substâncias químicas para águas superficiais.	Erosão; substâncias químicas; águas superficiais.

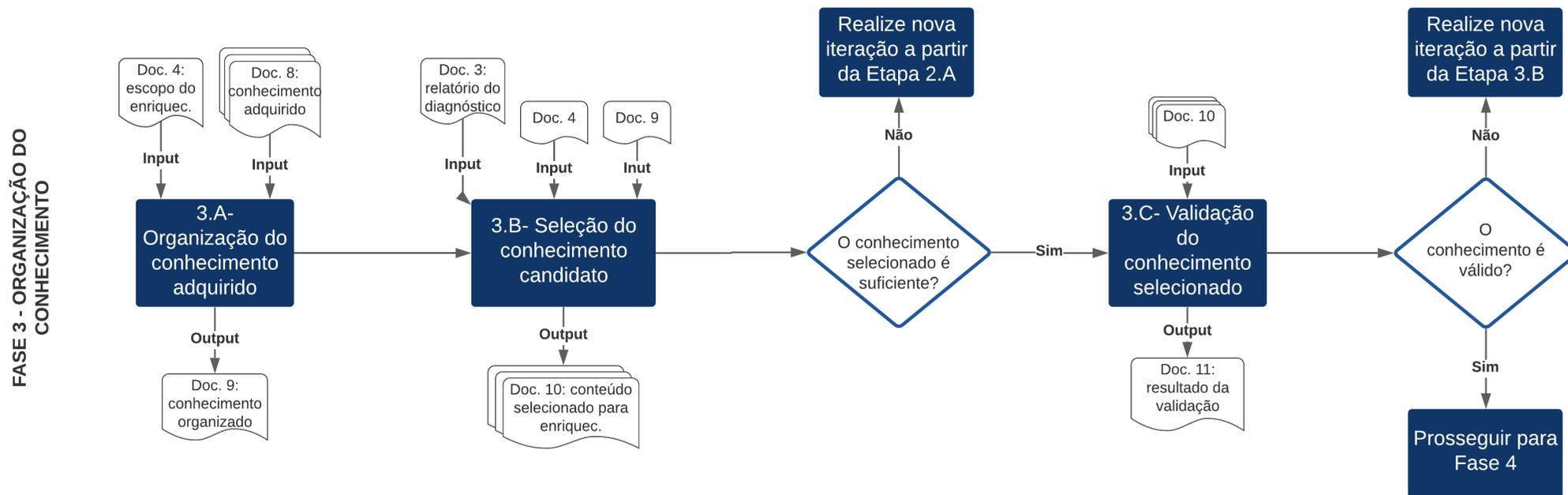
Fonte: elaborado pelo autor.

Responsável pela etapa: ontologista; profissional da área de computação (de acordo com a demanda, ou seja, se a técnica de extração exigir uma configuração mais avançada).

6.3 Organização do conhecimento (Fase 3)

Nesta fase, estrutura-se o conhecimento adquirido na fase anterior, com o propósito de identificar elementos para enriquecimento da ontologia. Esta fase é composta pelas seguintes etapas: **3.A – Organização do conhecimento adquirido**, que processa o conteúdo extraído do domínio com o objetivo de facilitar a inserção na ontologia; **3.B – Seleção do conhecimento candidato**, que aborda a escolha do conteúdo mais relevante para enriquecimento da ontologia; **3.C – Validação do conhecimento selecionado**, que visa garantir que esse conhecimento esteja de acordo com os propósitos da ontologia e com o escopo do enriquecimento. A seguir, as etapas da Fase 3 estão sintetizadas em um fluxograma.

Figura 20 – Fluxograma da Fase 3 – Organização do Conhecimento



Fonte: elaborada pelo autor.

Etapa 3.A – Organização do conhecimento adquirido

Descrição:

Nesta etapa, o objetivo é organizar o conhecimento adquirido na fase anterior. Essa organização visa facilitar a inserção do conhecimento na estrutura da ontologia, visto que, uma vez que o conhecimento está organizado, será mais fácil indicar o que vai ser inserido na ontologia, como: classe, relação, propriedade, axioma, ou seja, o conhecimento do domínio representado por conceitos será transformado em componentes ontológicos (classes, subclasses, relações).

Nesta etapa, recomenda-se o uso de alguns construtos da Biblioteconomia e Ciência da Informação (BCI), que se mostram relevantes para a organização do conhecimento. Entre esses construtos, estão os princípios da Teoria do Conceito de Dahlberg (1978) e da Teoria da Análise Facetada de Ranganathan (1967). De forma geral, os princípios da Teoria do Conceito podem ser utilizados na definição dos conceitos, visto que a definição é primordial para a posterior categorização do conceito, bem como para identificação de sinônimos, polissemia, correlação entre conceitos, etc. Os princípios da Teoria da Análise Facetada podem ser utilizados principalmente para categorização dos conceitos, contribuindo, desse modo, para o estabelecimento de classes e subclasses, para a identificação de relações entre conceitos no âmbito das ontologias.

Definir os conceitos é uma das principais ações da etapa de organização do conhecimento, pois facilitará as outras atividades, uma vez que a compreensão do domínio fica mais evidente. Vale destacar que o conceito precisa ser definido de preferência de acordo com o contexto da ontologia, ou seja, quanto mais específica e próxima do domínio representado pela ontologia a definição for, melhor. Em Campos (2010, 2017), é possível encontrar fundamentos teóricos e metodológicos sobre a função da definição no contexto das ontologias.

Diretrizes:

Para organizar o conhecimento adquirido, recomenda-se a análise dos documentos gerados na Etapa 2.D – Aplicação da técnica de extração da informação e do documento de escopo do enriquecimento (produto da Etapa 1.D). Se tiver

ocorrido mais de uma iteração, os documentos de todas as iterações devem ser considerados.

Após a análise dos documentos resultantes da aplicação da técnica de extração da informação (gerados na Etapa 2.D), deve-se:

- I. unificar os termos presentes em tais documentos em uma tabela;
- II. definir os termos de acordo com os princípios da Teoria do Conceito de Dahlberg (1978) (apresentada no [Capítulo 3](#)) e criar um glossário: as definições para os termos devem ser contextualizadas de acordo com o domínio coberto pela ontologia e, de preferência, deve-se definir uns termos em relação aos outros. Nesse sentido, ao definir, por exemplo, o termo “influencia”⁴⁵, que foi abordado na etapa prévia a título exemplificativo, deve-se fazê-lo de acordo com o domínio e de preferência mencionando outros elementos da ontologia que possuam relação com ele. Desse modo, a definição para “influencia” no domínio da transformação ambiental seria: tipo de relação entre um agente causador e um agente afetado (desertificação influencia a erosão, por exemplo);
- III. identificar termos sinônimos e relacionados; tratar a polissemia: com o suporte das definições, será possível fazer esse tratamento a partir da categorização dos termos, caso se julgue necessário. Nesse ponto, os princípios da Teoria do Conceito também podem ser empregados, principalmente com relação às categorias e subcategorias propostas por Dahlberg (1978), posto que isso contribuirá para entendimento da natureza (o que é entidade, processo, qualidade) do conceito e facilitará o estabelecimento de classes, relações entre os conceitos e axiomas;
- IV. classificar os termos de acordo com o tipo de componente ontológico (classe, relação, axioma) ou tipo de enriquecimento (lexical, conceitual, de relações, de axiomas): a execução dessa atividade pretende auxiliar a inserção do conteúdo na ontologia, pois facilitará o estabelecimento do que constituirá uma nova classe na ontologia (enriquecimento conceitual), o que serão variações terminológicas (enriquecimento lexical), quais serão as associações entre os conceitos (enriquecimento de relações) e quais regras podem ser incluídas na ontologia (enriquecimento de axiomas). Recomenda-se a organização do conhecimento em um quadro, conforme Quadro 15, a seguir.

⁴⁵ Verbo “influenciar” no presente do indicativo.

Quadro 15 – Modelo de quadro para organização do conhecimento adquirido

Conceito (enriquecimento conceitual)	Definição	Termos sinônimos (enriquecimento lexical)	Relações com outros conceitos (enriquecimento de relações)	Axioma (enriquecimento de axiomas)
Ciclo hidrológico	É um tipo de fenômeno natural, é componente da água, usa o processo 'circulação da água'.	Ciclo da água	É desenvolvido por: evaporação, condensação, precipitação, infiltração. É tipo de: ciclo biogeoquímico.	A relação É tipo de é transitiva. A relação É desenvolvido por é inversa da relação Desenvolve .

Fonte: elaborado pelo autor.

As orientações no exemplo do Quadro 15 se baseiam em um cenário em que os quatro tipos de enriquecimento são necessários. Porém podem existir situações em que o objetivo é enriquecer apenas dois tipos de componentes (isso está definido no documento de escopo). Assim uma das possibilidades é a organização do conteúdo extraído por temática ou por classe da ontologia que será enriquecida. Dessa forma, recomenda-se que cada ontologista analise o seu cenário e escolha o método de categorização mais adequado a ele.

Vale ressaltar que algumas diretrizes desta etapa se assemelham ao processo de construção de ontologias, principalmente com a fase de conceitualização ou modelagem conceitual das metodologias de construção. Portanto pode-se reutilizar algumas das estratégias empregadas nas metodologias de construção de ontologias, mas deve-se levar em conta sempre o objetivo do processo de enriquecimento, que difere da construção. No enriquecimento de ontologias, lida-se com uma estrutura construída previamente e qualquer modelagem deve levar em conta a estrutura atual da ontologia. A estrutura existente é um dos principais pontos de uma eventual modelagem conceitual. Ademais, no processo de enriquecimento, não se objetiva refazer a estrutura da ontologia, visa-se tão somente à expansão da atual. Qualquer alteração/reestruturação envolve outro processo da engenharia ontológica, tais como reengenharia, reestruturação, mudança (abordadas no [capítulo de fundamentação teórica](#)).

Responsável pela etapa: ontologista; especialista do domínio (opcional)

Etapa 3.B – Seleção do conhecimento candidato

Descrição:

Nesta etapa, o objetivo é identificar, dentre o conhecimento adquirido e organizado, aquele que é relevante para enriquecer a ontologia. Portanto precisa existir uma conexão entre o resultado da etapa de diagnóstico (Etapa 1.C) e esta etapa, visto que as lacunas existentes na ontologia que foram evidenciadas no diagnóstico precisam ser preenchidas com o conhecimento extraído das fontes de informação do domínio, além do documento de escopo do enriquecimento que contribui para o direcionamento da seleção.

Diretrizes:

Os documentos de entrada desta etapa são o relatório do diagnóstico (produto da Etapa 1.C), o documento de escopo do enriquecimento (Etapa 1.D) e os documentos gerados da etapa de organização do conhecimento adquirido (Etapa 3.A).

Primeiramente, analisa-se o documento de diagnóstico da ontologia e verificam-se quais conteúdos do conhecimento adquirido e organizado são úteis para preenchimento das lacunas identificadas, ou seja, faz-se uma correlação entre os problemas identificados no diagnóstico e as potenciais soluções identificadas na organização do conhecimento. A partir dessa convergência, é possível indicar qual fração do conhecimento extraído será selecionada para enriquecer a ontologia.

Em seguida, analisa-se o documento de escopo do enriquecimento para saber quais os limites do processo de enriquecimento, ou seja, até que nível deve-se especializar uma classe, por exemplo. Isso contribuirá diretamente na seleção do conhecimento considerado pertinente para a ontologia.

Outra recomendação que poderá auxiliar na seleção do conhecimento é acessar novamente cada uma das Questões de Competência e suas respectivas respostas. Os termos presentes nas questões de competência podem apresentar indícios importantes, e assim elas poderão guiar a atividade de seleção do conhecimento candidato para enriquecer a ontologia.

Caso o conhecimento selecionado seja considerado incompleto para preenchimento das lacunas identificadas no diagnóstico, recomenda-se uma nova

iteração a partir da Etapa 2.A – Seleção das fontes de informação. Os documentos gerados nessa nova iteração devem ser identificados com um marcador para diferenciar dos documentos gerados na primeira iteração.

Após a seleção do conhecimento candidato, deve-se organizar o conteúdo em formato de quadros ou tabelas para que seja de fácil acesso e compreensão. O mesmo modelo de quadro apresentado na etapa prévia pode ser utilizado nesta etapa. O único diferencial é que agora o quadro só conterá conhecimento que tem relação direta com o escopo de enriquecimento da ontologia.

Responsável pela etapa: ontologista e especialista de domínio.

Etapa 3.C – Validação do conhecimento selecionado

Descrição:

Esta etapa visa garantir que o conhecimento selecionado na etapa anterior para enriquecer a ontologia esteja de acordo com o conteúdo que a ontologia representa. Desse modo, recomenda-se que um ou mais especialistas de domínio analisem o conhecimento selecionado para enriquecer a ontologia. Os especialistas poderão indicar os pontos em que o conhecimento selecionado se mostra adequado e aqueles que precisarão ser revistos.

A análise dos especialistas deverá abranger os termos selecionados para representar os conceitos, as definições conceituais, as eventuais relações entre os conceitos e os axiomas (quando se aplica). Portanto todo o conhecimento deverá estar representado em um formato acessível e de simples assimilação pelos especialistas.

Diretrizes:

Os documentos de entrada desta etapa são aqueles gerados na Etapa 3.B – Seleção do conhecimento candidato. Esses documentos deverão ser analisados para verificar se estão claros para serem apresentados para o especialista de domínio ou se é necessária alguma intervenção do ontologista para facilitar a compreensão do especialista.

O próximo passo é convidar um ou mais especialistas do domínio e explicar o objetivo daquele processo de enriquecimento especificamente. Em seguida, deve-se apresentar o conteúdo do conhecimento selecionado para o especialista e solicitar que ele analise a adequabilidade do conteúdo à ontologia de domínio. Esse conteúdo deve ser apresentado da forma mais didática possível, por meio da linguagem natural.

Para cada conteúdo, o especialista deve dar o seu parecer (se o conhecimento é válido ou não) e a justificativa. O especialista deve ser indagado sobre como o conhecimento pode ser corrigido quando for identificado algum erro, posto que a validação do especialista, de certa forma, já é uma fonte de onde se pode capturar conhecimento. Quanto mais informação for registrada, mais fácil será para o ontologista identificar possível conteúdo faltante.

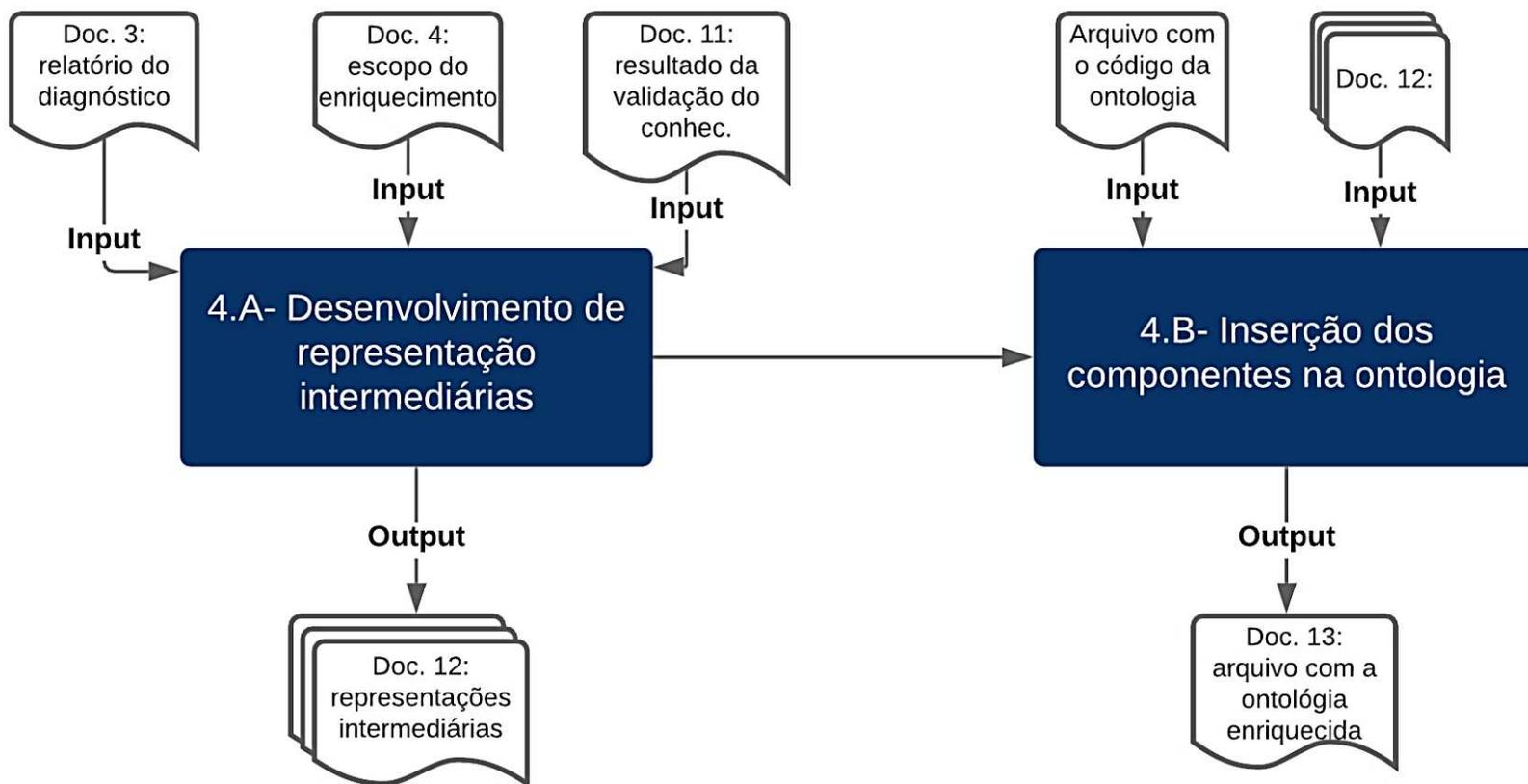
Caso parte do conhecimento ou sua totalidade não seja aprovada pelos especialistas de domínio, deve-se realizar uma nova iteração a partir da Etapa 3.B – Seleção do conhecimento candidato.

Responsável pela etapa: ontologista e especialista do domínio.

6.4 Implementação do enriquecimento (Fase 4)

A fase de implementação do enriquecimento visa à própria extensão da ontologia com o conhecimento apreendido e estruturado do domínio. Nessa fase, o propósito é inserir novos componentes na ontologia com o intuito de enriquecê-la. Para tal, respalda-se nas seguintes etapas: **4.A – Desenvolvimento de representações intermediárias**, que lida com a concepção de esquemas e modelos do conhecimento para a posterior inserção do conteúdo como recursos ontológicos na estrutura da ontologia, **4.B – Inserção dos componentes nos locais corretos da ontologia** – essa etapa lida com a decisão sobre onde o conhecimento apreendido e organizado será inserido na ontologia. O fluxograma representado na Figura 21 esquematiza as etapas da Fase 4.

Figura 21 – Fluxograma da Fase 4 – Implementação do Enriquecimento

FASE 4 - IMPLEMENTAÇÃO DO ENRIQUECIMENTO

Fonte: elaborada pelo autor.

Etapa 4. A – Desenvolvimento de representações intermediárias

Descrição:

Esta etapa visa à concepção de representações do conhecimento adquirido, selecionado e validado na Fase 3. Essas representações envolvem a disponibilização do conhecimento em linguagem natural (LN) e a elaboração de esquemas (por exemplo, mapas conceituais) para facilitar a posterior inclusão do conhecimento na ontologia. As relações entre os conceitos podem ser representadas graficamente. Posto que essa representação gráfica pode facilitar a observação de quaisquer equívocos na modelagem. Pressupõe-se que quando se inserem os recursos na estrutura da ontologia, essa observação pode ser mais complexa, principalmente para os especialistas de domínio, que geralmente não possuem familiaridade com os editores de ontologias.

Diretrizes:

Os documentos de entrada desta etapa são: o relatório de diagnóstico (produto gerado na Etapa 1.C), o documento de escopo do enriquecimento (produto gerado na Etapa 1.D) e o resultado da validação (produto gerado na Etapa 3.C).

A primeira atividade é partir das lacunas identificadas no documento de diagnóstico e rememorar a direção do processo de enriquecimento por meio do documento de escopo do enriquecimento; em seguida, inicia-se a modelagem de soluções a partir dos documentos gerados na Fase 3. Assim, caso o escopo de enriquecimento preveja os quatro tipos de enriquecimento, deve-se iniciar primeiramente pela modelagem das hierarquias de classes (considerando as classes já existentes na ontologia). A partir das hierarquias já desenvolvidas, deve-se representar as relações entre os conceitos, de preferência em esquemas gráficos (tais como mapas conceituais), e posteriormente os sinônimos (realizações alternativas dos conceitos). Com relação aos axiomas, recomenda-se representar em linguagem natural quais serão as restrições. Por exemplo, pode-se dizer em LN que a Classe 1 é disjunta da Classe 2.

No desenvolvimento das representações intermediárias, já se deve apontar quais conceitos serão transformados em classes ou subclasses da ontologia, quais serão relacionados e por quais relações. O ideal é que especialistas de domínio

contribuam para esta fase, porém, a depender do escopo do enriquecimento, ela pode ser conduzida apenas pelo ontologista.

No contexto da METHODOE, não serão apontados quais tipos de representações intermediárias podem ser utilizadas. Podem-se utilizar os diversos recursos e ferramentas disponíveis. O intuito desta etapa é apenas contribuir para que se raciocine e se observem as representações antes de inseri-las como recursos ontológicos na estrutura da ontologia. Essas representações devem envolver os componentes da estrutura existente da ontologia e não apenas os elementos que serão usados para enriquecer. A ideia é integrar o conhecimento novo com a estrutura existente da ontologia. Portanto, o produto dessa etapa são as representações intermediárias geradas.

Responsável pela etapa: ontologista e especialista do domínio (opcional).

Etapa 4.B – Inserção dos componentes na ontologia

Descrição:

Esta etapa trata da expansão da ontologia, seja em especialização seja em extensão com o conteúdo do conhecimento adquirido e selecionado, ou seja, o próprio enriquecimento da ontologia. Logo todas as etapas anteriores contribuem para esta etapa, visto que, nesse ponto da metodologia, já é possível saber em quais partes da ontologia o enriquecimento deve acontecer (ou seja, quais são as suas lacunas), bem como quais componentes são necessários.

Diretrizes:

Os documentos de entrada desta etapa são o arquivo com o código da ontologia e as representações intermediárias desenvolvidas na Etapa 4.A. Analisados os documentos de entrada, deve-se determinar o responsável ou responsáveis pelo enriquecimento. As diretrizes da metodologia sugerem duas formas de realizar a inclusão do conhecimento na ontologia: (i) pelo ontologista; (ii) pelo ontologista e especialista de domínio juntos.

O segundo passo é a escolha de um software editor de ontologias para facilitar a inserção dos componentes na estrutura da ontologia. Conforme já mencionado, existem várias opções de softwares.

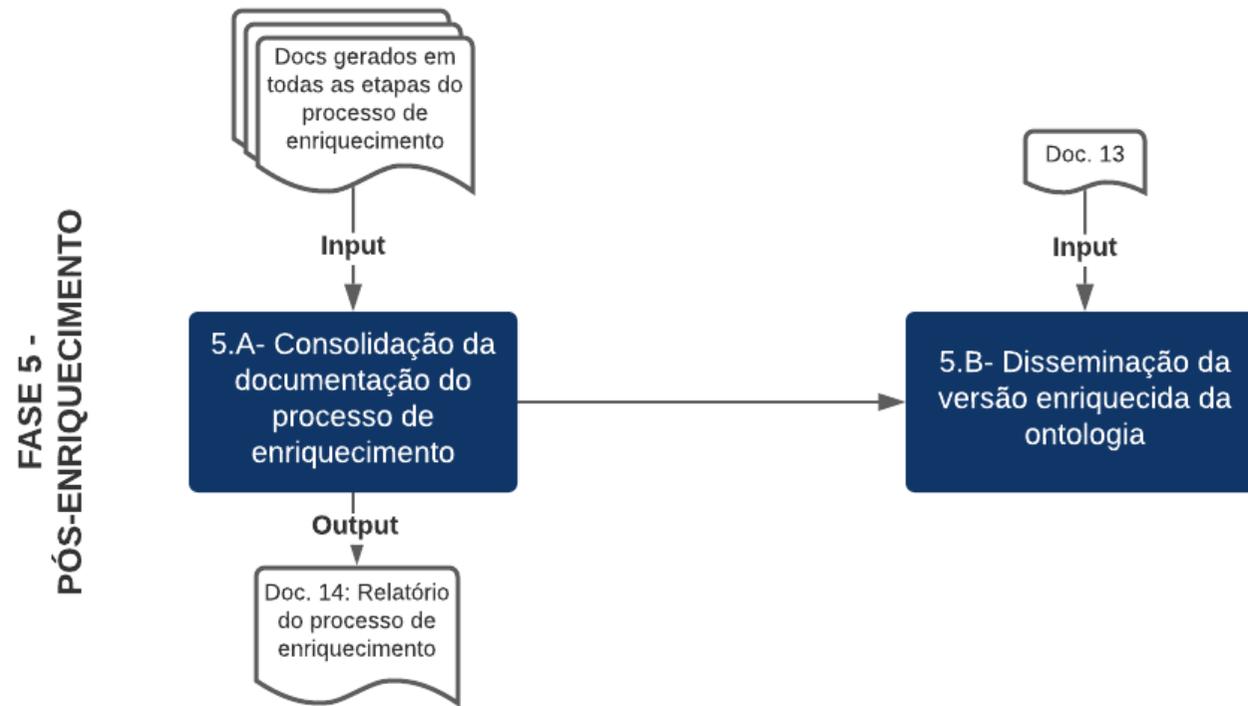
O terceiro passo é a inclusão dos componentes na estrutura da ontologia, em que devem ser utilizados os documentos gerados nas representações intermediárias da etapa precedente. Primeiramente, os conceitos devem ser incluídos como classes e subclasses. Em seguida, incluem-se as anotações da classe, como definições e sinônimos. A terceira tarefa trata da criação dos relacionamentos e suas propriedades lógicas (por exemplo, transitivos, reflexivos, simétricos, assimétricos) e da conexão das classes por meio dos relacionamentos. Finalmente, devem-se estabelecer os axiomas entre as classes enriquecidas. Ressalta-se que, nessa diretriz, estão previstos os quatro tipos de enriquecimento. Portanto, se o escopo do enriquecimento em determinado contexto prevê a realização de enriquecimento de menos componentes, deve-se seguir a diretriz relacionada àquele tipo de enriquecimento. Como resultado desta etapa, gera-se um arquivo com a ontologia enriquecida, por meio do uso de um editor de ontologias.

Responsável pela etapa: ontologista / especialista de domínio (opcional)

6.5 Pós-Enriquecimento (Fase 5)

A fase de pós-enriquecimento é a última fase da METHODOE e trata das tarefas que devem ser executadas posteriormente ao enriquecimento da ontologia. Essa fase é composta das seguintes etapas: **5.A – Consolidação da documentação do processo de enriquecimento**, que visa deixar explícitas todas as estratégias (registro das atividades realizadas e como foram realizadas) que foram adotadas durante o processo de enriquecimento; **5.B – Disseminação da versão enriquecida da ontologia**, que objetiva difundir (comunicar) a nova versão da ontologia para o público interessado.

Figura 22 – Fluxograma da fase de Pós-Enriquecimento



Fonte: elaborada pelo autor.

Etapa 5.A – Consolidação da documentação do processo de enriquecimento

Descrição:

Esta etapa trata do registro de todas as atividades executadas nas etapas anteriores do processo de enriquecimento. A etapa de documentação acontece durante todo o processo de enriquecimento, visto que cada fase deve gerar um ou mais documentos com o resultado (ou seja, o produto daquela etapa). Porém somente no final do processo de enriquecimento é possível reunir toda essa documentação.

Diretrizes:

Os documentos de entrada (*input*) para esta etapa são os documentos gerados em todas as etapas prévias. Conforme foi apontado em cada uma das etapas, deve-se gerar um documento para cada etapa do processo de enriquecimento.

No próximo passo, realiza-se o registro da documentação do processo de enriquecimento, gerando um relatório com as estratégias e escolhas adotadas para realização de cada etapa e o documento resultante daquela etapa. Desse modo, este relatório se configura como a documentação do processo de enriquecimento.

Responsável pela etapa: ontologista

Etapa 5.B – Disseminação da versão enriquecida da ontologia

Descrição:

Esta é a última etapa do processo de enriquecimento e refere-se à divulgação na versão enriquecida da ontologia para a comunidade de usuários interessados.

Diretrizes:

A partir do documento gerado na Etapa 4.B – Inserção dos componentes na ontologia, que se refere ao código da ontologia enriquecida, deve-se realizar a divulgação para a comunidade de usuários interessados. Isso pode ser feito por meio da disponibilização da ontologia em repositórios de Sistemas de Organização do Conhecimento (alguns deles foram citados na primeira etapa desta metodologia), caso

seja de interesse dos desenvolvedores tornar a ontologia acessível a qualquer pessoa.

Uma forma complementar para disseminação da ontologia é apresentá-la por meio de artigos científicos, em eventos técnico-científicos, entre outras publicações e comunicações.

Responsável pela etapa: ontologista e especialista do domínio.

7 APLICAÇÃO DA METHODOE E DISCUSSÃO DOS RESULTADOS

Este capítulo discorre sobre o exercício empírico de aplicação da metodologia de enriquecimento (METHODOE) em uma ontologia de domínio com o intuito de testar e validar a metodologia. Para isso, escolheu-se uma ontologia do domínio agrícola, a OntoAgroHidro.

A OntoAgroHidro é uma ontologia desenvolvida no âmbito da Embrapa Informática Agropecuária, sediada em Campinas, São Paulo. A motivação para a escolha dessa ontologia como universo para aplicação da metodologia ora desenvolvida refere-se ao fato de o proponente já ter trabalhado em uma pesquisa de mestrado com essa ontologia e com membros da equipe de desenvolvimento desse artefato. Portanto presumiu-se que essa experiência e conhecimento facilitariam o exercício empírico.

Para apresentação dos resultados referentes à aplicação da METHODOE na ontologia OntoAgroHidro (OAH), decidiu-se separar as seções de acordo com as cinco fases principais da metodologia de enriquecimento. Nesse sentido, na [seção 7.1](#), aborda-se o planejamento do enriquecimento; a [seção 7.2](#) discorre sobre a aquisição de conhecimento para enriquecimento da ontologia; a [seção 7.3](#) versa sobre a organização do conhecimento; já a [seção 7.4](#) trata da implementação do enriquecimento na OntoAgroHidro; e a [seção 7.5](#) aborda o pós-enriquecimento. A única seção divergente das fases da metodologia de enriquecimento é a [seção 7.6](#), na qual discutem-se as dificuldades encontradas na aplicação da METHODOE, bem como os aspectos positivos da metodologia e suas limitações.

7.1 Planejamento do enriquecimento

Seguindo as diretrizes da METHODOE, a fase de planejamento possui quatro etapas: (1) identificação da ontologia alvo, (2) avaliação da necessidade de enriquecimento, (3) diagnóstico da ontologia e (4) escopo do enriquecimento.

Na etapa de identificação da ontologia alvo, realizou-se o levantamento de informações sobre a OAH a partir de buscas de publicações existentes sobre ela, resultando em três artigos científicos recuperados. No entanto não foi possível realizar o levantamento de informações em repositório de ontologias, pois a OAH não está catalogada em nenhum repositório. A seguir, encontra-se o quadro com as informações referentes à OAH.

Quadro 16 – Identificação da ontologia OntoAgroHidro

Nome da ontologia: OntoAgroHidro

Idiomas: Português e inglês (devido ao reúso de outras ontologias no processo de desenvolvimento, algumas classes herdadas estão em inglês).

Domínio coberto: Impacto das **mudanças climáticas** na **agricultura** e nos **recursos hídricos dos biomas brasileiros**.

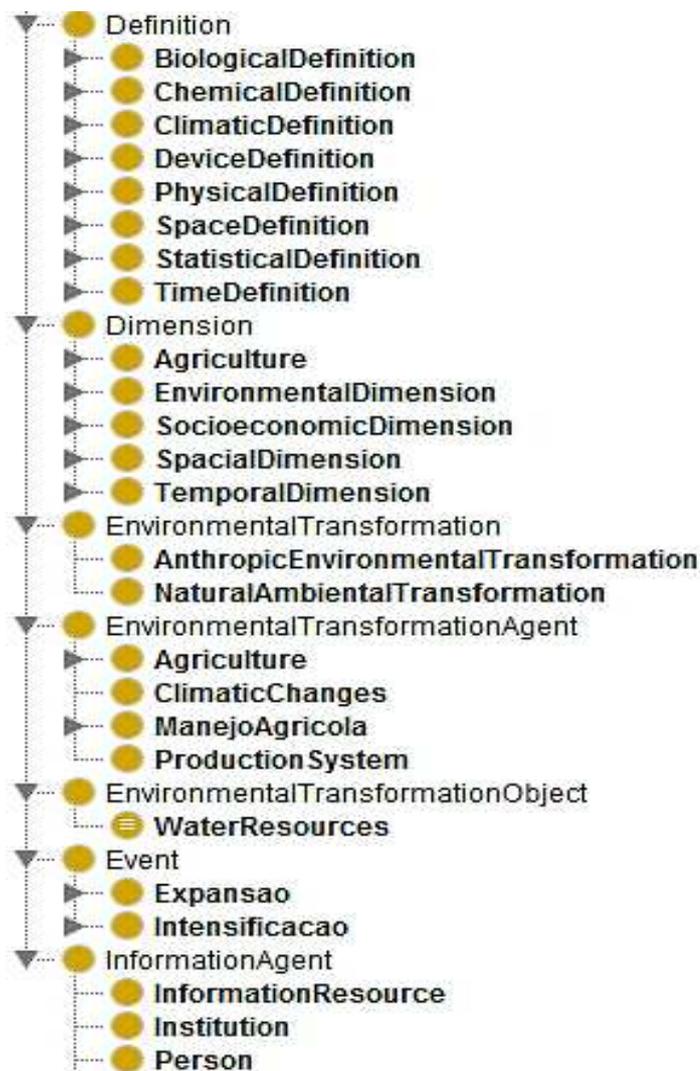
Descrição: OntoAgroHidro é uma ontologia de domínio criada por especialistas da Embrapa Informática Agropecuária. O objetivo principal dessa ontologia é representar o conhecimento sobre os impactos das mudanças climáticas e da agricultura nos recursos hídricos dos biomas brasileiros, funcionando como um vocabulário padrão para integração e compartilhamento de informações entre pesquisadores e outras instituições que atuam em parceria com a Embrapa. A versão atual do OntoAgroHidro possui aproximadamente 8500 elementos, divididos entre classes, subclasses, propriedades, relações e instâncias. A Figura 1 mostra a estrutura hierárquica principal da OAH, com as sete classes principais e suas respectivas subclasses. As principais classes são: *Definition* (**Definição**) – que se refere às definições de conceitos necessários para caracterizar outros conceitos no domínio; *Environmental Transformation* (**Transformação Ambiental**) – que se refere ao fenômeno de transformação ambiental que pode ter causas antrópicas e/ou naturais; é causada por um agente; é resultado de um evento e muda um objeto; *Environmental Transformation Object* (**Objeto de Transformação Ambiental**) – refere-se ao objeto no qual ocorre a transformação ambiental; na OntoAgroHidro, o foco está nos recursos hídricos, ou seja, o objeto alterado pela transformação ambiental. Outra classe principal é *Environmental Transformation Agent* (**Agente de Transformação Ambiental**) – que se refere aos agentes que causam a transformação ambiental; *Dimension* (**Dimensão**) – que versa sobre as dimensões (no sentido de múltiplas variáveis/aspectos) de um problema específico; as dimensões são especificadas e relacionadas à classe Definição; *Event* (**Evento**) – que se refere aos eventos que geram mudanças ambientais; um evento possui múltiplas dimensões, por exemplo, ocorre em um período e local, está associado a uma cultura de cultivo; um evento é descrito por um agente de informação; *Information Agent* (**Agente de Informação**) – refere-se aos agentes e recursos que contêm ou fornecem informações na rede. Na Figura 23, essas classes principais estão representadas.

Questões de competência: Quais os fatores que influenciam na qualidade da água? Quais são os impactos das mudanças climáticas nos recursos hídricos e na agricultura? Quais os impactos da agricultura nos recursos hídricos?

Fonte: Bonacin, Nabuco e Pierozzi Jr. (2014); Bonacin, Nabuco e Pierozzi Jr. (2015); Bonacin, Nabuco e Pierozzi Jr. (2016).

Fonte: elaborado pelo autor.

Figura 23 – Estrutura hierárquica da OntoAgroHidro com as classes e subclasses principais



Fonte: Captura de tela no software Protégé.

Após as informações da ontologia alvo do processo de enriquecimento terem sido apresentadas, passa-se para a etapa de avaliação da necessidade de enriquecimento, que foi realizada com base na análise da estrutura da ontologia. Consideraram-se, também, o objetivo principal da OAH e suas questões de competência.

Para esse exercício empírico, optou-se por enriquecer apenas uma amostra da OntoAgroHidro, com o uso do software Protégé versão 5.5, o qual permitiu verificar toda a estrutura hierárquica da ontologia. Além disso, pode-se visualizar a estrutura da OAH graficamente usando o *plugin* Ontograf do Protégé, que fornece uma visão mais abrangente da ontologia e a visualização das relações entre as classes.

A análise da OAH revelou que, dentre as sete classes principais, aquelas que estão menos desenvolvidas (falta de especialização) são: *Environmental Transformation* e *Information Agent*. Essas classes se desdobram apenas em mais um nível hierárquico (como mostra a Figura 23). A classe *Environmental Transformation* foi escolhida arbitrariamente para enriquecimento por tratar do conhecimento mais crítico para a ontologia, já que lida com fenômenos que causam mudanças no meio ambiente. Portanto, é essencial que os conceitos relacionados a esses fenômenos estejam presentes na estrutura da ontologia.

A classe *Environmental Transformation* é uma classe central da ontologia que indica um fenômeno que altera algo no meio ambiente. Esse conceito é fundamental para entender como a agricultura transforma o meio ambiente (em particular, os recursos hídricos) e como as transformações nos recursos hídricos (aquelas causadas por mudanças climáticas, por exemplo) podem afetar a agricultura e a produção de alimentos em geral, e vice-versa. A classe *Environmental Transformation* possui duas subclasses: *Anthropic Environmental Transformation* e *Natural Environmental Transformation*. A primeira cobre as mudanças que são causadas por influência do ser humano; e a segunda cobre as mudanças em decorrência de fenômenos naturais (BONACIN, NABUCO e PIEROZZI JR. 2015).

Em seguida, adotou-se a diretriz da METHODOE referente ao uso de Questões de Competência para identificar a necessidade de enriquecimento. Portanto, desenvolveram-se duas questões de competência que fazem parte do escopo de modelagem da ontologia para avaliar a necessidade de enriquecimento: QC1: quais são as transformações ambientais antrópicas? QC2: quais são as mudanças ambientais naturais? Percebeu-se que a OAH não possui, em sua estrutura, conteúdo para responder a tais questões, o que resultou em um apontamento para a necessidade de seu enriquecimento.

Quadro 17 – Resultado da avaliação da necessidade de enriquecimento na classe *Environmental Transformation*

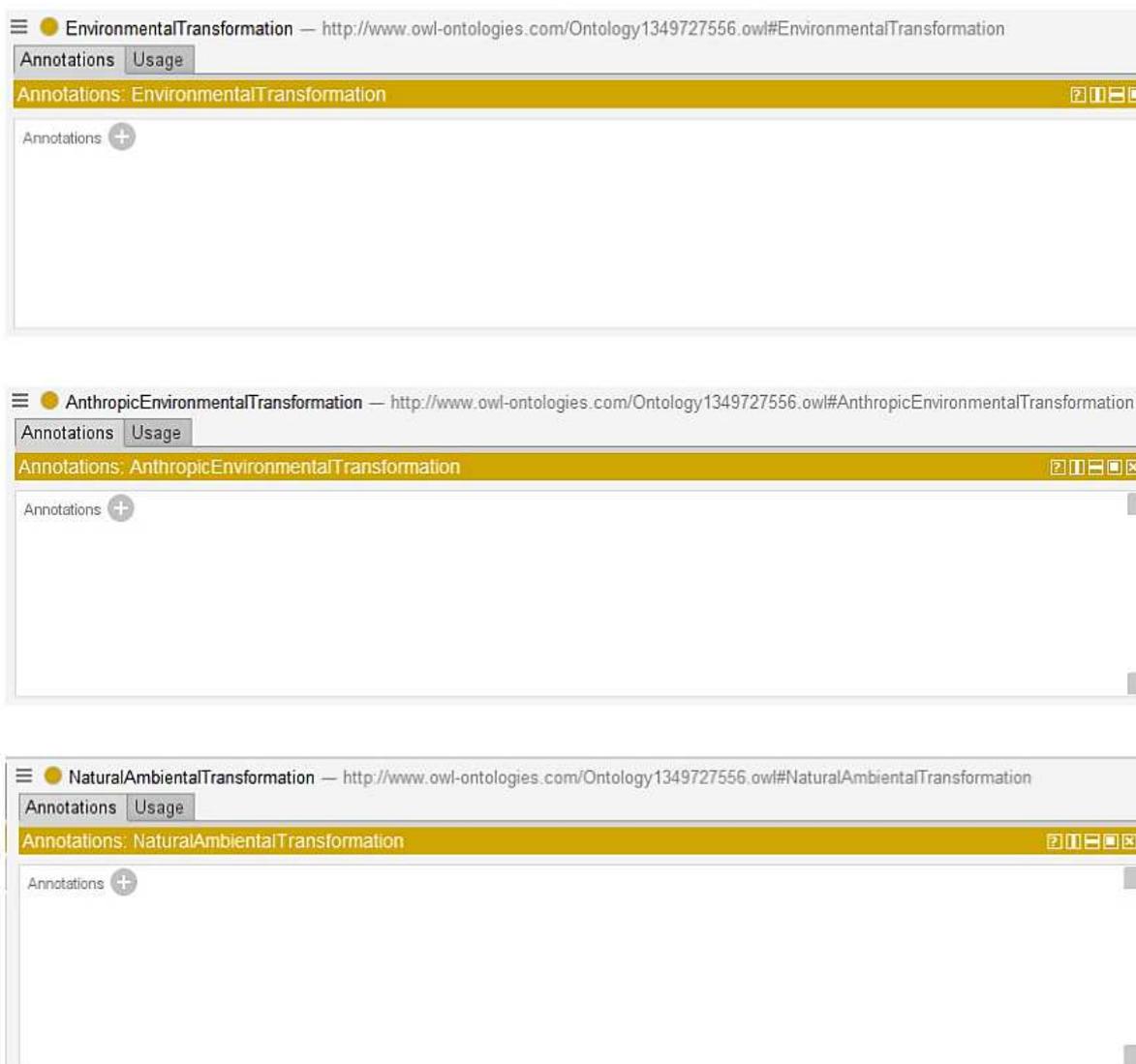
O enriquecimento é necessário?	Justificativa
Sim	Identificou-se a necessidade de realização do enriquecimento, porque a OAH não apresentava qualquer tipo de transformação ambiental ou antrópica. Dessa forma, não é possível responder às QCs 1 e 2.

Fonte: elaborado pelo autor.

Na etapa seguinte, realizou-se o Diagnóstico da Ontologia na classe *Environmental Transformation* e nas classes relacionadas a ela. Para identificar os relacionamentos dessa classe com as outras classes principais, realizou-se uma análise da estrutura da ontologia com o software Protégé.

O resultado dessa etapa indica que a classe *Environmental Transformation* precisa ser enriquecida (mais especificamente as subclasses *Anthropic Environmental Transformation* e *Natural Environmental Transformation*), porque ficou evidenciada a falta de conceitos que especializam as duas subclasses; nesse caso, os conceitos sobre as transformações ambientais. Além disso, notou-se a ausência de anotações (definição e rótulo em português) para a classe *Environmental Transformation* e suas subclasses, como pode ser constatado na Figura 24.

Figura 24 – Ausência de anotações para a classe *Environmental Transformation* e suas subclasses



Fonte: Captura de tela no software Protégé.

O diagnóstico da OAH gerou o relatório a seguir, apresentado no Quadro 18.

Quadro 18 – Relatório do diagnóstico da OntoAgroHidro

Classe / subclasse	Anotações	Relações	Axiomas
Classe <i>Environmental Transformation</i>	Falta definição e rótulo em português.	-	-
Especializar subclasse <i>Anthropic Environmental Transformation</i>	Falta definição e rótulo em português.	-	-
Especializar Subclasse <i>Natural Environmental Transformation</i>	Falta definição e rótulo em português.	-	-

Fonte: elaborado pelo autor.

Por último, na etapa de escopo do enriquecimento, conforme delineado pela METHODOE, acessou-se o documento com as informações da ontologia (apresentado no Quadro 16) e o documento com o relatório do diagnóstico (demonstrado no Quadro 18). A partir desses documentos, preencheu-se o documento de escopo do enriquecimento, apresentado no Quadro 19.

Em consonância com o objetivo principal da OntoAgroHidro e com as suas questões de competência, definiu-se que o limite para especialização da classe *Environmental Transformation* e as subclasses *Anthropic Environmental Transformation* e *Natural Environmental Transformation* seria todas aquelas transformações que possuem relação com agricultura, com recursos hídricos e mudanças climáticas no contexto de ecossistemas brasileiros. Nesse sentido, deve descartar-se todos os conceitos que não estejam associados ao contexto apresentado.

Quadro 19 – Documento de escopo do enriquecimento da OntoAgroHidro

Enriquecimento conceitual	Enriquecimento lexical	Enriquecimento de relações	Enriquecimento de axiomas
Classe <i>Environmental Transformation</i>	Definir sinônimos para a classe <i>Environmental Transformation</i> . Incluir definição para a classe <i>Environmental Transformation</i> .	-	-
Especializar a subclasse <i>Anthropic Environmental Transformation</i> com conceitos que envolvam transformações no âmbito da agricultura, recursos hídricos e mudanças climáticas nos ecossistemas brasileiros.	Definir sinônimos para a subclasse <i>Anthropic Environmental Transformation</i> e para eventuais subclasses específicas. Incluir definição para a subclasse <i>Anthropic Environmental Transformation</i> e para eventuais subclasses específicas.	Verificar relações que possam ser geradas a partir dos novos componentes.	Verificar propriedades das eventuais relações. Verificar se “ <i>domain</i> ” e “ <i>range</i> ” das relações estão definidos.
Especializar a subclasse <i>Natural Environmental Transformation</i> com conceitos que envolvam transformações no âmbito da agricultura, recursos hídricos e mudanças climáticas nos ecossistemas brasileiros.	Definir sinônimos para a subclasse <i>Natural Environmental Transformation</i> e para eventuais subclasses específicas. Incluir definição para a subclasse <i>Natural Environmental Transformation</i> e para eventuais subclasses específicas.	Verificar relações que possam ser geradas a partir dos novos componentes.	Verificar propriedades das eventuais relações. Verificar se “ <i>domain</i> ” e “ <i>range</i> ” das relações estão definidos.

Fonte: elaborada pelo autor.

Na próxima seção (Aquisição de conhecimento), apresentam-se as estratégias adotadas para captura do conhecimento e levantamento de insumos para o enriquecimento da OntoAgroHidro.

7.2 Aquisição de conhecimento

A METHODOE prevê quatro etapas na fase de Aquisição de conhecimento, a saber: (1) seleção das fontes de informação, (2) escolha da técnica de extração da informação, (3) preparação da fonte de informação e (4) aplicação da técnica de extração da informação.

Para selecionar as fontes de informação, procedeu-se da seguinte forma: (i) acesso ao relatório de diagnóstico da ontologia e ao documento de escopo para manter o enfoque nas lacunas identificadas e no objetivo do processo de enriquecimento; (ii) a partir dos conceitos da classe Transformação Ambiental, foram feitas consultas (utilizando-se as expressões "ambiente" e "antropogênico") no repositório Agroportal para identificar SOC (principalmente tesouros e ontologias) relacionados ao tema "transformação ambiental natural e antrópica"; (iii) busca no Google Scholar pelas expressões: impactos de atividades antrópicas, ecossistema de atividades antrópicas, impacto ambiental, transformação ambiental natural; (iv) busca no Google Scholar pelas seguintes questões de competência: QC1 – quais são as mudanças ambientais antrópicas? e QC2 – quais são as mudanças ambientais naturais? Vale ressaltar que se optou pela busca em uma base de dados genérica, pois as buscas em bases de dados especializadas não retornaram resultados considerados relevantes.

Os resultados das consultas foram selecionados a partir da análise do título e leitura técnica, no caso de textos. Em relação aos SOCs, foram analisadas as estruturas hierárquicas e as definições dos conceitos. O número de fontes foi considerado suficiente quando as informações se tornaram repetitivas (os mesmos conceitos foram encontrados). O Quadro 20 mostra as fontes selecionadas e o motivo para escolha delas.

Quadro 20 – Fontes de informação selecionadas

N.º	Fonte	Endereço de acesso	Motivo para escolha
1	AGROVOC	http://aims.fao.org/vest-registry/vocabularies/agrovoc	SOC relacionado ao domínio da ontologia.
2	Artigo	https://bit.ly/3by5LgY	Relacionado ao tema geral da ontologia.
3	Artigo	https://www.sciencedirect.com/science/article/pii/S0045653599001150	Relacionado ao tema geral da ontologia.
4	Artigo	https://www.sciencedirect.com/science/article/pii/S0168192314000033	Relacionado ao tema geral da ontologia.
5	Artigo	https://core.ac.uk/download/pdf/15436376.pdf	Relacionado ao tema geral da ontologia.
6	Artigo	https://bit.ly/2wAqRgA	Relacionado ao tema geral da ontologia.
7	Artigo	http://www.seer.ufu.br/index.php/caminhosdegeografia/article/view/16613	Relacionado à questão de competência n.º 1.
8	Livro	https://smastr16.blob.core.windows.net/igeo/2019/05/mudancas_climaticas_suguio_2008.pdf	Relacionado à questão de competência n.º 2.
9	<i>Global Agricultural Concept Scheme</i>	https://bit.ly/3fQUDz9	SOC relacionado ao domínio da ontologia.
10	<i>Environment Ontology</i>	http://agroportal.lirmm.fr/ontologies/ENVO	SOC relacionado ao domínio da ontologia.
11	<i>Biodiversity Thesaurus</i>	http://agroportal.lirmm.fr/ontologies/BIODIVTHES	SOC relacionado ao domínio da ontologia.
12	<i>Sustainable Development Goals Interface Ontology</i>	http://agroportal.lirmm.fr/ontologies/SDGIO	SOC relacionado ao domínio da ontologia.

Fonte: elaborado pelo autor.

Após a seleção das fontes de informação, realizou-se a escolha da técnica de extração da informação. Desse modo, por ser recomendação da METHODOE que a técnica escolhida esteja relacionada às fontes de informação selecionadas, este estudo concentrou-se principalmente na abordagem manual em vez da automática, pois o enriquecimento está concentrado em uma classe específica da ontologia. Assim foi possível realizar análises manuais/intelectuais. Nesta etapa, usou-se a análise

manual de documentos textuais e a análise manual de SOC como técnicas para extrair conhecimento. O Quadro 21 apresenta as técnicas escolhidas seguidas do motivo para escolha.

Quadro 21 – Técnicas de extração da informação

Técnica de extração	Grau de automação	Motivo para escolha
Análise de documentos	Manual	A leitura técnica pode trazer pistas para enriquecimento da ontologia.
Análise de SOC	Manual	A análise manual permite uma exploração de toda a estrutura do SOC.

Fonte: elaborado pelo autor.

A etapa de preparação das fontes de informação envolveu o *download* dos textos completos dos documentos selecionados e a organização deles em um diretório para posterior análise. No caso dos SOCs, optou-se por analisá-los em seus endereços online e não fazer o *download* do arquivo para posterior análise. Isso porque o acesso a todos os recursos da ferramenta de gestão e disponibilização do SOC facilita a navegação, a busca e a identificação das relações entre os termos.

Por último, na etapa de aplicação da técnica de extração da informação, iniciou-se a extração dos termos nos SOCs selecionados: AGROVOC, GACS (*Global Agricultural Concept Scheme*), ENVO (*Environment Ontology*), *Biodiversity Thesaurus* e SDGIO (*Sustainable Development Goals Interface Ontology*). A extração da terminologia envolveu a análise da hierarquia dos conceitos, suas definições (quando disponíveis) e a relação com outros conceitos. Foram selecionados o máximo de conceitos que pudessem ter relação com a classe *Environmental Transformation*. As relações (principalmente as hierárquicas) entre alguns conceitos também foram consideradas para enriquecimento. Como exemplo, apresenta-se a classe "*Deforestation*" que é *subclasse* de da classe "*Environment degradation*". O Quadro 22 traz um exemplo da terminologia extraída dos SOCs.

Quadro 22 – Resultado da análise dos SOCs

Fonte	Termos extraídos
AGROVOC	<p><i>Anthropogenic changes</i> (sinônimos: <i>anthropogenic factors, anthropic factors, anthropic influence</i>);</p> <p><i>Environmental degradation</i> (descritores específicos: <i>forest degradation, land degradation, soil degradation, deforestation, desertificação, poluição, forest fragmentation</i>);</p> <p>Erosion;</p> <p>Ozone depletion;</p> <p><i>Natural phenomenon</i> (descritores específicos: <i>atmospheric circulation, atmospheric formations, climate, climate change, cycling, water circulation, greenhouse effect</i>)</p>
GACS	<p><i>anthropic pressure</i></p> <p><i>anthropogenic impacts, human activity (environmental impact), anthropic factors, anthropogenic factors,</i></p> <p><i>anthropogenic effects, anthropogenic disturbances, human impact, anthropogenic influence, anthropogenic influences, anthropic influence</i></p> <p><i>Environment degradation</i> (descritores: <i>Deforestation; Forest fragmentation; Fouling; Habitat destruction ; Land degradation; Desertification; Pollution; Water pollution</i>)</p>
ENVO	<i>Anthropogenic environmental process</i>
Biodiversity Thesaurus	<p><i>Anthropogenic factor</i></p> <p><i>Environmental degradation</i> (descritores específicos: <i>Biodiversity erosion; Deforestation; Desertification; Ecocide; Eutrophication; Pollution; Trampling; Habitat fragmentation; Human activity</i>);</p> <p><i>Climate-smart agriculture</i></p> <p><i>Conventional agriculture</i></p>
SDGIO	<p><i>Anthropogenic environmental process</i> (subclasses: <i>Climate change; Desertification; Anthropisation</i>)</p> <p><i>Anthrophisation</i> (subclasses: <i>Aforestation; Deforestation; Revegetation</i>)</p>

Fonte: elaborado pelo autor.

A análise dos SOCs trouxe uma quantidade relevante de sinônimos e conceitos relacionados ao conceito de Transformação Ambiental. Além disso, identificaram-se as relações hierárquicas entre esses conceitos.

No que tange à análise dos documentos, realizou-se uma leitura exploratória de cada documento selecionado, com o intuito de identificar elementos relacionados ao escopo do enriquecimento (ao objetivo da ontologia e às questões de competência). Foram extraídos trechos que continham algum conceito relacionado à classe Transformação Ambiental. A ideia de extrair trechos se mostrou relevante, uma

vez que apresenta o contexto de uso do conceito no domínio e, nesse sentido, pode contribuir para a identificação de relações entre os conceitos. A seguir, no Quadro 23, apresenta-se uma amostra do resultado da análise dos documentos. O resultado completo encontra-se no [Apêndice E](#).

Quadro 23 – Amostra do resultado da análise dos documentos

Título do documento	Trecho extraído
Potencial de impacto da agricultura sobre os recursos hídricos na Região do Cerrado	Eutrofização causada pelo enriquecimento dos corpos hídricos com fósforo e nitrogênio. Deposição de sedimentos carregados diminuindo a capacidade dos reservatórios (assoreamento). Erosão pode ajudar a levar substâncias químicas para águas superficiais.
Um novo olhar na Geografia para os conceitos e aplicações de geossistemas, sistemas antrópicos e sistemas ambientais	Um Sistema Ambiental pode ser caracterizado como entidade organizada na superfície terrestre formada pelos subsistemas físico/natural (Geossistema) e antrópico, bem como por suas interações. O subsistema físico-natural (Geossistema) é composto por elementos e processos relacionados ao clima, solo, relevo, águas e seres vivos, enquanto os componentes e processos do subsistema Antrópico são aqueles ligados a: população, urbanização, industrialização, agricultura e mineração, entre outras atividades e manifestações humanas. Ambos mantêm um funcionamento parcialmente independente, e também um funcionamento dependente um do outro, ou seja, mesmo a natureza apresentando suas leis e dinâmica própria, esta pode sofrer alterações em decorrência da ação antrópica, por exemplo, alterações nos níveis pluviométricos e alterações na temperatura, como consequência do desmatamento. Do mesmo modo, os Sistemas Antrópicos sofrem interferência das leis da natureza, por exemplo, a restrição na produção de energia hidráulica no Brasil, entre os anos de 2001 e 2002, em decorrência das estiagens que diminuíram o nível de armazenamento de água nas barragens.

Fonte: elaborado pelo autor.

A exploração do conteúdo extraído dos textos contribuiu para a identificação de relações semânticas entre conceitos do domínio, o que se mostrou pertinente, posto que contribuiu para a complementação do conteúdo extraído dos SOCs, já que, nesses instrumentos, identificaram-se, principalmente, relações taxonômicas (hierárquicas).

A seção seguinte (Organização do conhecimento) versa sobre a ordenação e o tratamento dos insumos identificados nesta seção para que sejam descobertos elementos passíveis de serem utilizados no enriquecimento da OntoAgroHidro.

7.3 Organização do conhecimento

Na fase de Organização do Conhecimento, realiza-se a estruturação do conhecimento adquirido na fase anterior, segundo as orientações da METHODOE. Para isso, essa fase é composta de três etapas: (1) organização do conhecimento adquirido, (2) seleção do conhecimento candidato e (3) validação do conhecimento selecionado.

Na etapa de organização do conhecimento adquirido, os elementos extraídos na etapa precedente foram organizados em um quadro. Em seguida, os conceitos considerados mais relevantes para a OntoAgroHidro foram definidos, resultando em um glossário ([Apêndice F](#)). Ressalta-se que, conforme as diretrizes da METHODOE, os conceitos devem ser definidos uns em relação aos outros. Ademais devem-se seguir os princípios da Teoria do Conceito. Portanto respaldou-se nessas diretrizes para a definição dos conceitos presentes nos glossários.

A partir do documento resultante da aplicação da técnica de extração, no qual é possível verificar o contexto de extração dos termos, e com base nas definições dos conceitos, foi possível identificar os termos sinônimos e os termos relacionados. Portanto, após a execução desses passos, desenvolveu-se um quadro com cinco colunas: (i) numeração conceito; (ii) enriquecimento conceitual; (iii) enriquecimento lexical; (iv) enriquecimento de relações; (v) enriquecimento de axiomas. Observou-se que, para realizar a classificação dos conceitos, foi necessário voltar algumas vezes às fontes selecionadas e extrair conteúdo que contribuísse para a compreensão do conceito. A seguir, apresenta-se, no Quadro 24⁴⁶, o resultado dessa organização.

Quadro 24 – Organização do conhecimento adquirido

Continua...				
#	Enriquecimento conceitual	Enriquecimento lexical	Enriquecimento de relações	Enriquecimento de axiomas
1	Erosion	-	Is influenced by desertification; type of soil degradation.	A relação is influenced by é inversa da relação influences . A relação type é transitiva.

⁴⁶ O conteúdo está organizado em inglês, pois será inserido na ontologia neste idioma.

Continua...

#	Enriquecimento conceitual	Enriquecimento lexical	Enriquecimento de relações	de Enriquecimento de axiomas
2	Greenhouse effect	-	Affects atmosphere	-
3	Climate change	Climatic change	influences climate; is influenced by Anthropoic Transformation.	As relações influence e is influenced by são inversas.
4	Hydrological cycle	Water cycle	It develops by evaporation, condensation, precipitation, infiltration, and transpiration processes. Type of cycling	A relação type é transitiva.
5	Resource depletion	-	is caused by desertification, environmental degradation, and soil degradation	A relação is caused by é inversa da relação causes .
6	Hydrological evaporation	-	Type of Physical Process	A relação type é transitiva.
7	Cycling	Biogeochemical cycle	has subclass hydrological cycle	A relação has subclass é inversa da relação subclass of .
8	Precipitation	-	Type of Physical Process	A relação type é transitiva.
9	Anthropic Environmental Transformation	Anthropogenic changes, Anthropogenic impacts, Anthropogenic factors, Anthropoic factors, Anthropoic influence, Anthropoic impact	is caused by deforestation, anthropization, eutrophication, environmental degradation, forest degradation, land degradation, desertification, soil degradation, deforestation, afforestation, revegetation, pollution	A relação is caused by é inversa da relação causes .

Continua...

#	Enriquecimento conceitual	Enriquecimento lexical	Enriquecimento de relações	Enriquecimento de axiomas
10	Environmental degradation	Environmental damage, Ecological degradation, Environmental deterioration, Ecological damage	Is caused by deforestation, desertification, pollution; causes erosion and resource depletion	A relação causes e is caused by são inversas.
11	Forest degradation	Forest damage	Type environmental degradation	A relação type é transitiva.
12	Fatores meteorológicos	-	-	-
13	Land degradation	-	Type environmental degradation; caused by degradation; subclass desertification	A relação type é transitiva. is soil A relação is caused by é inversa da relação causes . has A relação has subclass é inversa da relação subclass of .
14	Desertification	Desertization	Subclass of land degradation; influences erosion; is caused by soil degradation; causes environmental degradation	A relação subclass of é inversa da relação has subclasse . A relação influences é inversa da relação is influenced by . A relação causes é inversa da relação is caused by .
15	Soil degradation	Soil deterioration	causes land degradation and desertification; type of environmental degradation	A relação causes é inversa da relação is caused by . A relação type é transitiva.
16	Deforestation	-	causes environmental degradation	A relação causes é inversa da relação is caused by .

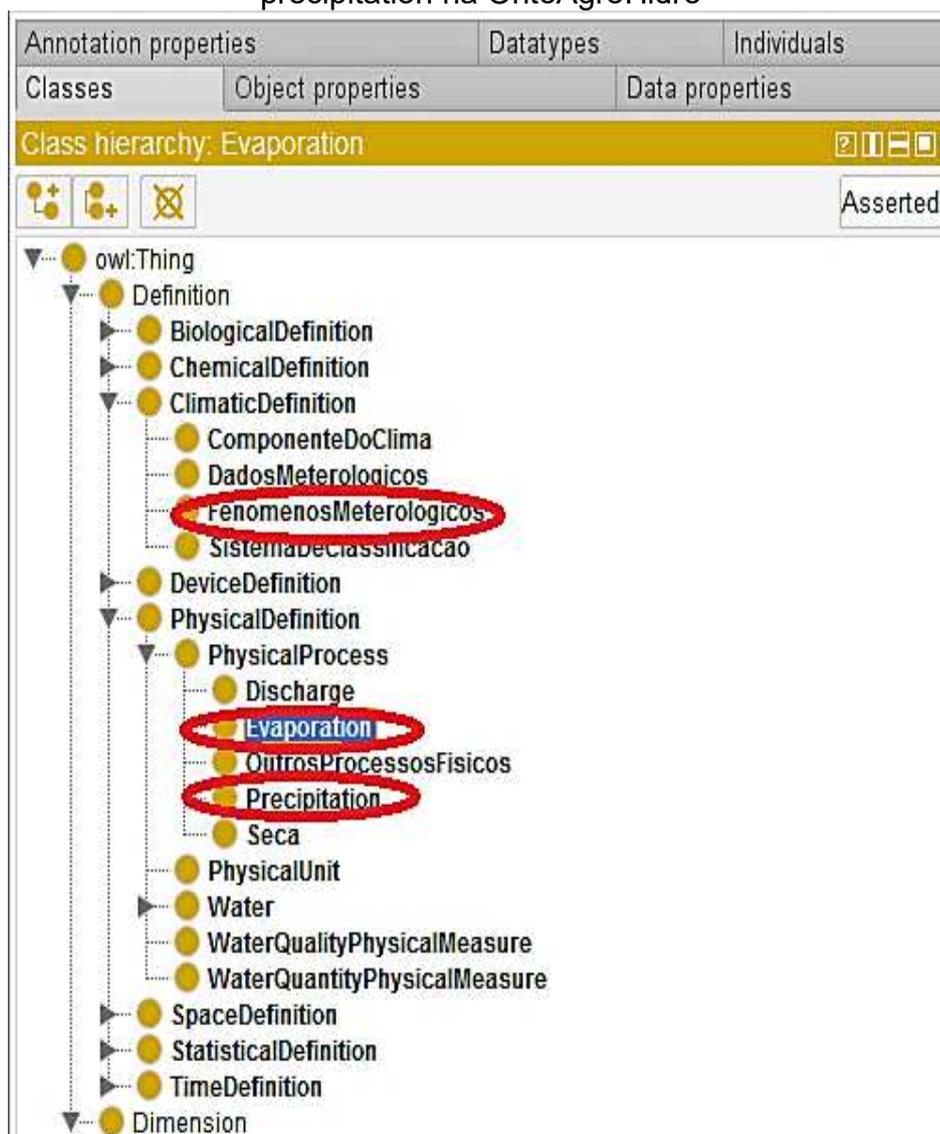
				Conclusão
#	Enriquecimento conceitual	Enriquecimento lexical	Enriquecimento de relações	Enriquecimento de axiomas
17	Pollution	-	causes environmental degradation; has subclass water pollution and soil pollution	A relação causes é inversa da relação is caused by . A relação has subclass é inversa da relação subclass of .
18	Water pollution	-	Type of pollution; affects water quality; has subclass eutrophication	A relação type é transitiva. A relação has subclass é inversa da relação subclass of .
19	Soil pollution	-	causes soil degradation; is caused by fertilizers; type of pollution	A relação causes é inversa da relação is caused by . A relação type é transitiva.
20	Anthropisation	Anthropization	-	-
21	Afforestation	-	uses process of revegetation	-
22	Revegetation	-	Helps to control erosion; is process for afforestation	-
23	Eutrophication	-	Type of water pollution	A relação type é transitiva.

Fonte: elaborado pelo autor.

Em seguida, na etapa de seleção do conhecimento adquirido, analisaram-se quais dos elementos organizados no Quadro 24 são relevantes para enriquecimento da ontologia. Portanto foram novamente avaliados o relatório do diagnóstico da ontologia e o documento de escopo do enriquecimento para garantir que todos os elementos tivessem relação com a ontologia. Nesse ponto, foi preciso ir além do que recomenda a METHODOE, pois foi indispensável analisar novamente a estrutura da OntoAgroHidro e verificar se cada conceito estava presente na ontologia, seja em forma de classe/subclasse, seja como relação, anotações, etc. A METHODOE não prevê a análise da estrutura da ontologia nesta etapa.

Dentre os elementos apresentados no Quadro 24, apenas os conceitos “evaporação”, “precipitação” e “fatores meteorológicos” foram desconsiderados, pois já estavam presentes na OAH. O conceito de evaporação (*evaporation*) e precipitação (*precipitation*) constam na ontologia como subclasses de processo físico. Já o conceito “fatores meteorológicos” (ou fenômeno meteorológico) consta na OAH como subclasse de definição climática, conforme ilustra a Figura 25, a seguir.

Figura 25 – Presença dos conceitos fenômenos meteorológicos, evaporation e precipitation na OntoAgroHidro



Fonte: Captura de tela no software Protégé.

Para finalizar, a fase de Organização do Conhecimento contou com a contribuição de um especialista do domínio para validação do conhecimento adquirido. Conforme diretriz da METHODOE, a análise do especialista deve abranger os termos selecionados para representar os conceitos, os sinônimos, as definições do

glossário, as relações entre os conceitos e os axiomas. Além disso, o conhecimento precisa ser apresentado da forma mais didática possível.

Portanto criou-se um documento em formato textual para facilitar a análise do especialista do domínio. Julgou-se importante representar o conhecimento em linguagem natural para auxiliar na compreensão do especialista e assim garantir que a qualidade da validação não fosse afetada. Além do envio do documento, foi realizada uma conversa por videoconferência para discutir detalhes que não estavam claros no documento.

No documento textual enviado para o especialista, primeiro, foram definidas três questões que ele deveria levar em consideração ao realizar a validação. São elas: 1) o conhecimento representado tem relação com os objetivos da OntoAgroHidro?; 2) o conhecimento está modelado adequadamente (hierarquia, relações com outras classes, definições, sinônimos)?; 3) estão faltando conceitos e/ou relações importantes dentro da Classe *Environmental Transformation*?

Em seguida, a validação foi segmentada em três partes: (i) hierarquia das classes e subclasses; (ii) definição, sinônimos; (iii) relações entre as classes. Além disso, inicialmente o especialista fez a validação apenas do conteúdo para enriquecimento da subclasse *Natural Environmental Transformation* e, em seguida, da subclasse *Anthropic Environmental Transformation*.

O especialista aprovou a maioria do conteúdo inserido na ontologia, porém fez algumas anotações sobre as subclasses presentes na classe *Natural Environmental Transformation*. Segundo ele, as subclasses *Erosion*, *Greenhouse Effect*, *Climate Change* e *Resource Depletion* devem possuir relação com a classe *Anthropic Environmental Transformation*, pois, no contexto de OntoAgroHidro, são todas as transformações causadas pela agricultura. Assim a relação "é causada por" foi utilizada para relacionar as classes.

O especialista também sugeriu mudar a classe "*Cycling*", para a classe "*Nutrient Cycling*". Ele recomendou a exclusão dos sinônimos "dano ambiental" e "dano ecológico" associados à subclasse "Degradação Ambiental", bem como a exclusão do sinônimo "dano florestal" que estava associado à subclasse "degradação florestal". Por fim, sugeriu a inclusão das seguintes associações entre as classes: (i) Desmatamento is_caused_by LandDegradation; (ii) O desmatamento is_caused_by degradação ambiental.

A seção subsequente versa sobre o próprio enriquecimento da OntoAgroHidro. Portanto os elementos organizados e tratados nesta fase serão utilizados para expansão da ontologia.

7.4 Implementação do enriquecimento

A fase de implementação do enriquecimento da METHODOE estabelece duas etapas: (1) desenvolvimento de representações intermediárias e (2) inserção dos componentes na ontologia.

As representações intermediárias envolveram a representação do conhecimento por meio de textos em linguagem natural e a criação da hierarquia de classes e subclasses, conforme pode ser observado no Quadro 25, a seguir. Nesse caso, optou-se por inserir os rótulos dos conceitos em inglês e, como anotação da classe, inserir a tradução do termo para o português. Isso porque a disponibilização da ontologia em língua inglesa facilita o compartilhamento com a comunidade internacional.

Quadro 25 – Representação intermediária em formato hierárquico

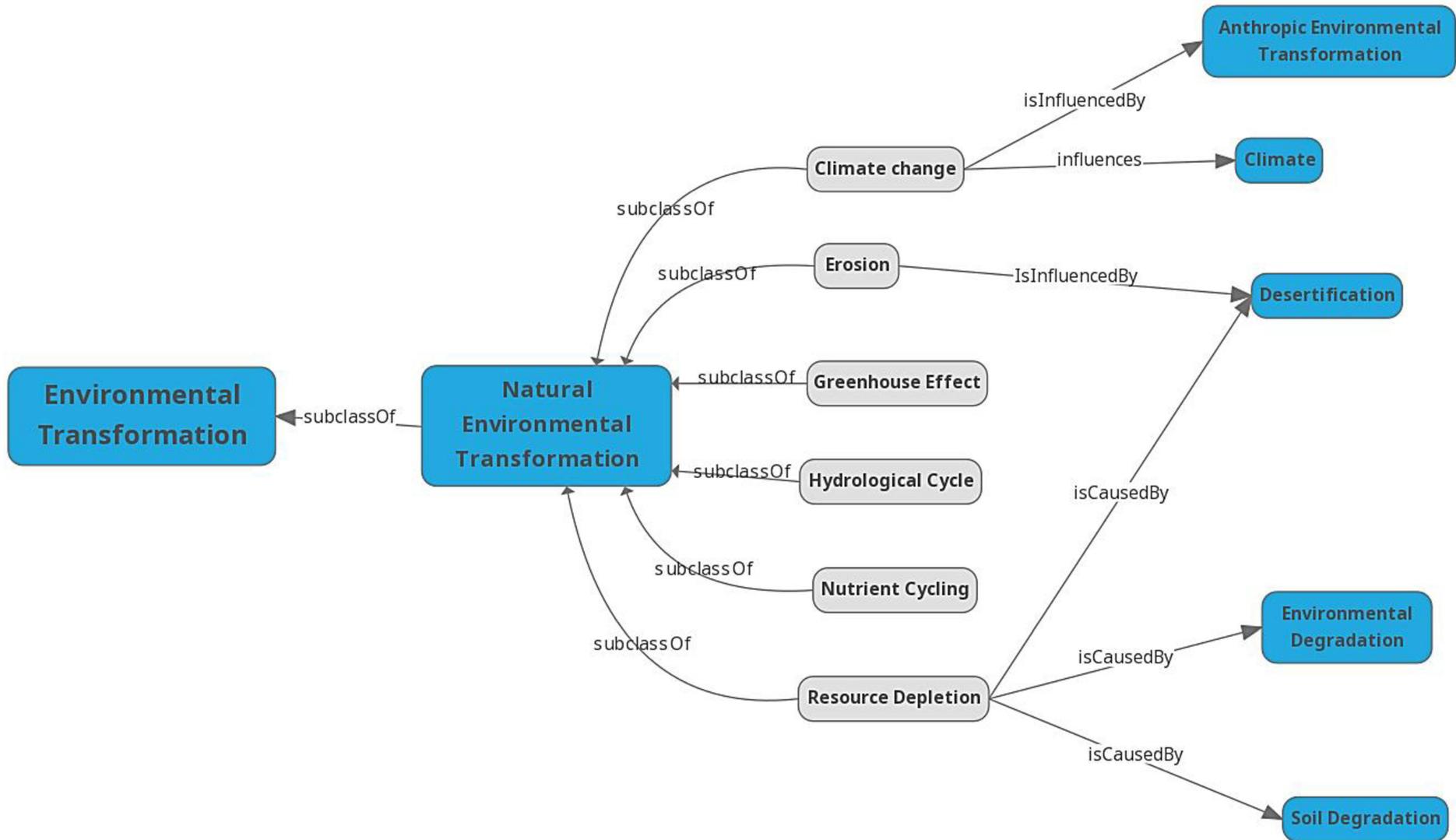
Environmental Transformation
<i>Natural Environmental Transformation</i>
Erosion
Greenhouse Effect
Climate Change
Hydrological Cycle
Resource Depletion
Nutrient Cycling
<i>Anthropic Environmental Transformation</i>
Antropization
Eutrophication
Environmental Degradation
Forest Degradation
Land Degradation
Desertification
Soil Degradation
Deforestation
Afforestation
Revegetation
Pollution
Water Pollution
Soil Pollution

Fonte: elaborado pelo autor.

Após a elaboração da hierarquia das classes (que basicamente é uma taxonomia), desenvolveram-se alguns mapas conceituais com as relações taxonômicas de cada conceito. Isso facilitou a visualização de como os conceitos que serão utilizados para enriquecimento impactam em conceitos relacionados a outras partes da OntoAgroHidro (ou seja, partes que não estavam previstas no enriquecimento). A seguir, apresenta-se, na Figura 26, o mapa conceitual com as relações dos conceitos pertencentes à classe *Natural Environmental Transformation* e, na Figura 27, o mapa conceitual referente à classe *Anthropic Environmental Transformation*. Os mapas conceituais foram elaborados utilizando-se o software Mindmup⁴⁷.

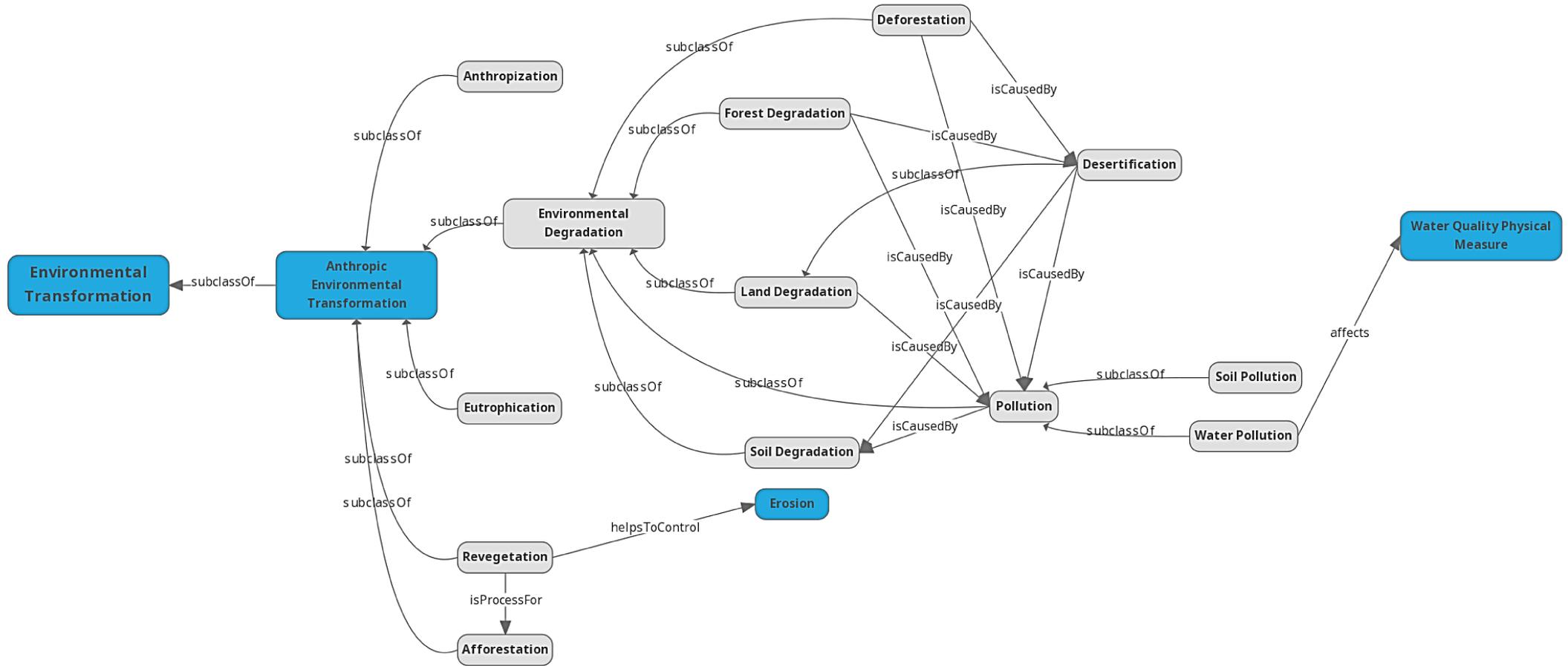
⁴⁷ <https://www.mindmup.com/>

Figura 26 – Mapa conceitual da classe Natural Environmental Transformation



Fonte: elaborada pelo autor.

Figura 27 – Mapa conceitual da classe Anthropoc Environmental Transformation



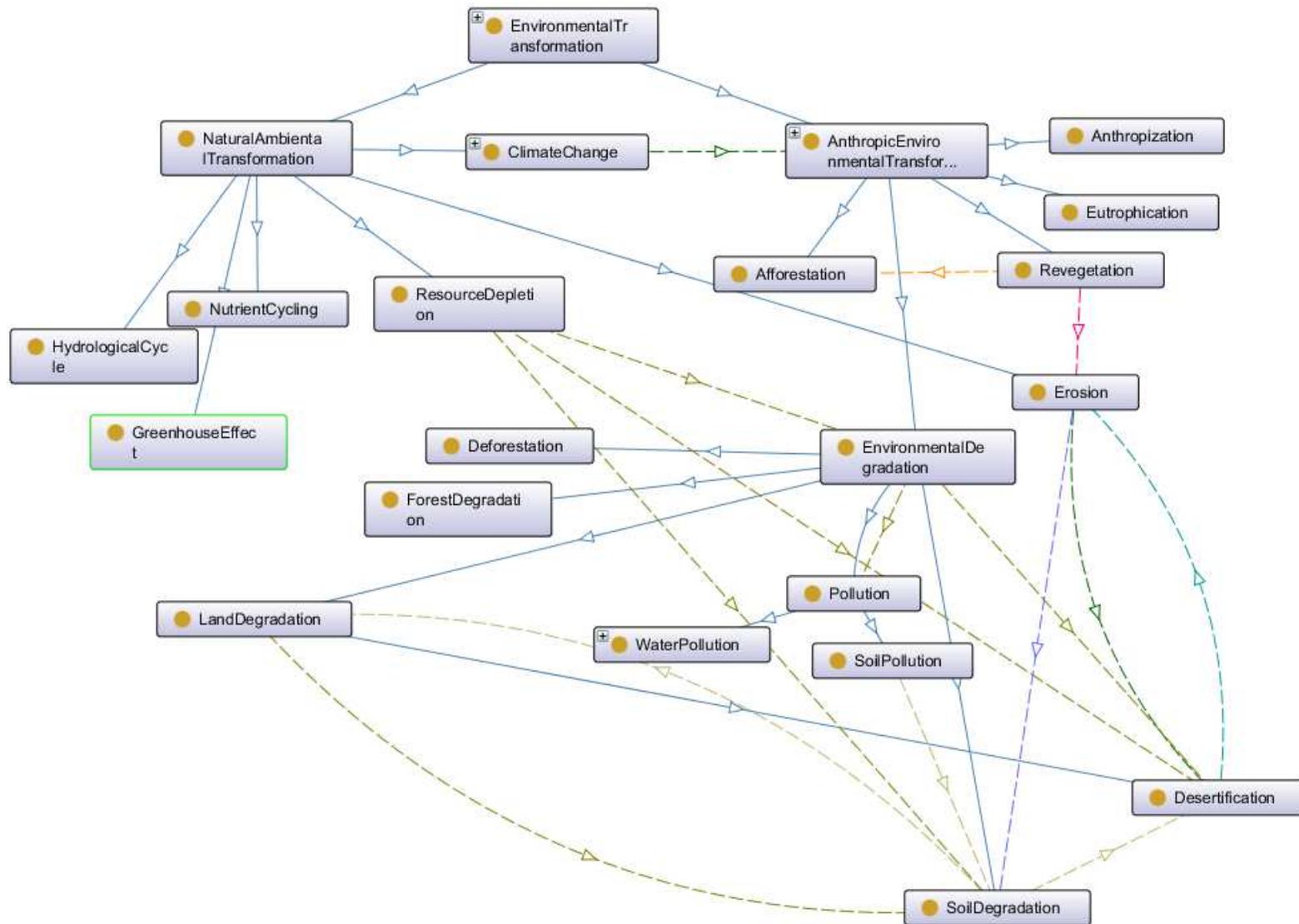
Fonte: elaborada pelo autor.

A partir dos documentos gerados na fase de organização do conhecimento e com base nas representações intermediárias, iniciou-se o enriquecimento da OntoAgroHidro. Para enriquecimento da ontologia, o arquivo OWL da OntoAgroHidro foi carregado no software Protégé versão 5.5.

O enriquecimento começou por meio da criação da hierarquia de classes a partir da classe *Environmental Transformation*. Em seguida, foram atribuídas as anotações (com definição e sinônimos) a partir do vocabulário (ou *namespace*) SKOS (*Simple Knowledge Organization System*). Depois das anotações, foram criadas as relações, definidas suas propriedades e associação entre as classes a partir das definições criadas. Por último, configuraram-se os axiomas.

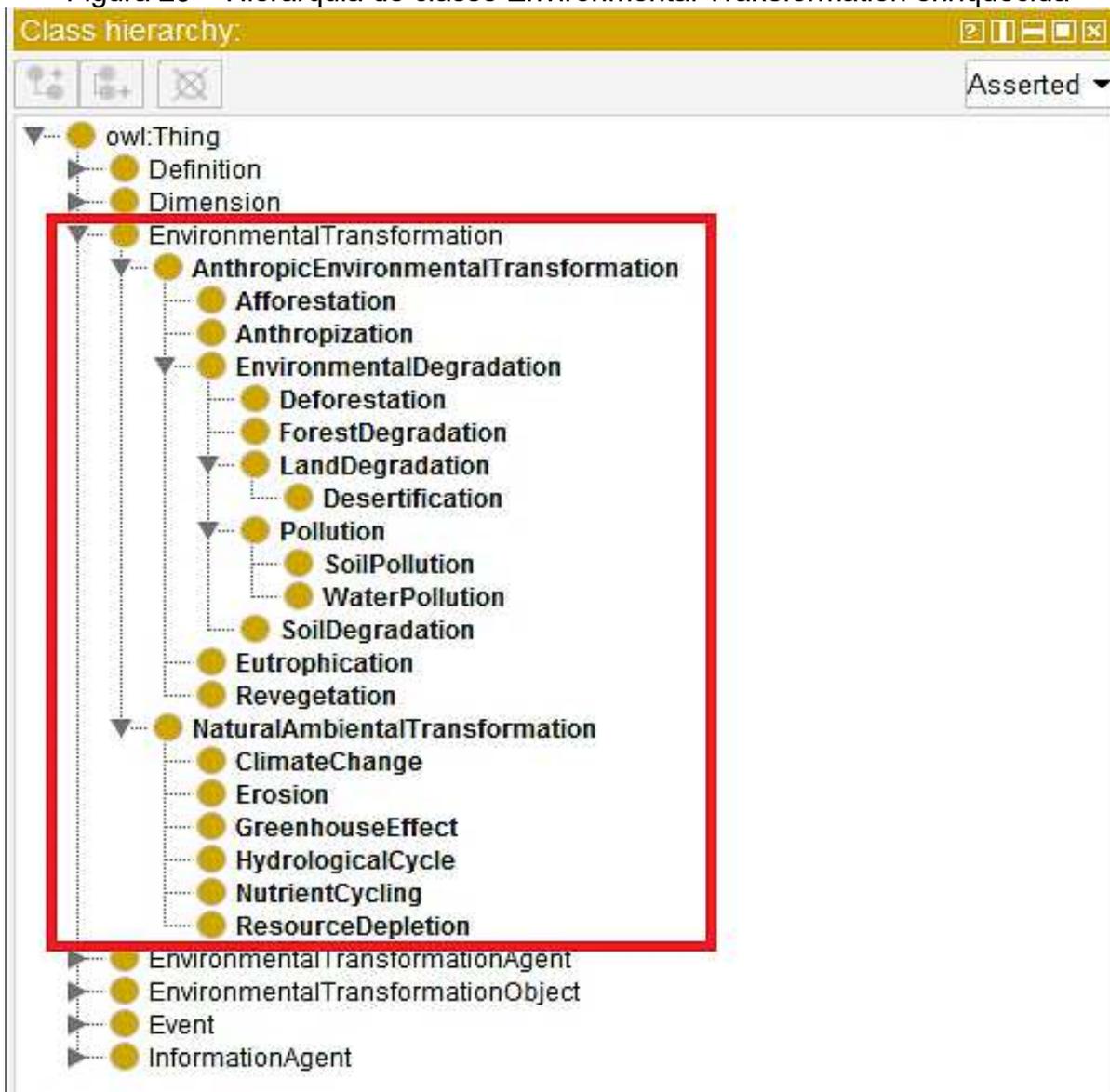
Como resultado do enriquecimento, apresenta-se, na Figura 28, as relações entre os conceitos enriquecidos, bem como a taxonomia da classe *Environmental Transformation* (na Figura 29).

Figura 28 – Classe Environmental Transformation enriquecida



Fonte: Captura de tela no software Protégé.

Figura 29 – Hierarquia de classe Environmental Transformation enriquecida



Fonte: Captura de tela no software Protégé.

Na seção seguinte, discorre-se sobre as etapas que ocorrem posteriormente ao processo de enriquecimento.

7.5 Pós-enriquecimento

A METHODOE propõe duas etapas pertencentes à fase de pós-enriquecimento: (1) consolidação da documentação do processo de enriquecimento e (2) disseminação da versão enriquecida da ontologia.

Com relação à documentação da OntoAgroHidro, foram gerados documentos em cada etapa, conforme as recomendações da METHODOE. Além disso,

explicaram-se em cada etapa as estratégias adotadas. Sendo assim, todo o processo de enriquecimento da ontologia foi registrado. Portanto, acredita-se que, se, futuramente, a ontologia precisar passar por um novo processo de enriquecimento, a equipe responsável estará apoiada em informações para conduzi-lo.

No que tange à disseminação da versão enriquecida da OntoAgroHidro, comunica-se que o arquivo com a extensão OWL da ontologia foi encaminhado para um dos especialistas responsáveis pelo desenvolvimento da OAH. Além disso, pretende-se divulgar a experiência de enriquecimento da ontologia em eventos e periódicos científicos. Desse modo, contribui-se para que a OAH seja reconhecida por outras comunidades.

7.6 Considerações gerais sobre a METHODOE e sobre a aplicação na OntoAgroHidro

Nesta seção, apresentam-se as considerações sobre a aplicação da Metodologia de Enriquecimento de Ontologias de Domínio, bem como as características e limitações da metodologia.

Conforme pode ser observado neste capítulo e no capítulo anterior, a METHODOE foi elaborada considerando-se a intervenção humana em praticamente todas as etapas, pois acredita-se que a capacidade humana de inferência e raciocínio ainda são primordiais para o enriquecimento de modelos complexos, como as ontologias de domínio. Os métodos automáticos de enriquecimento ainda não constroem estruturas conceituais complexas. Todavia sabe-se que, cada vez mais, os métodos automáticos para aquisição de conhecimento estão sendo aperfeiçoados. Dessa forma, defende-se a combinação entre métodos automáticos e manuais para se extrair a terminologia com mais rapidez; e a intervenção humana para manipular essa terminologia ou validar os modelos primários construídos de forma automática.

No desenvolvimento e aplicação da METHODOE, tentou-se superar as lacunas identificadas nos estudos descritos na revisão de literatura. Entre as lacunas, destacam-se: (i) técnicas que exigem a supervisão de um especialista da computação; (ii) ausência de detalhamento dos passos para execução do método; (iii) validação em um cenário limitado; (iv) métodos muito específicos para o domínio biomédico; (v) situações em que o enriquecimento com base no método totalmente automático não

produziu bons resultados; (vi) falta de orientações para enriquecimento de todos os componentes; (vii) dificuldade de acesso às fontes de informação por serem muito específicas.

A METHODOE supera parcialmente a **lacuna (i) – técnicas que exigem a supervisão de um especialista da computação**. Isso porque, na metodologia, não se impõe qual técnica de extração da informação o ontologista precisa utilizar, deixando a critério dele e de sua equipe a decisão sobre qual técnica empregar. Porém o ontologista pode optar por uma técnica complexa que exige conhecimentos técnicos sobre alguma linguagem de programação ou sobre o conhecimento de alguma ferramenta específica. Sendo assim, será necessária a supervisão de um especialista da computação.

No tocante à **lacuna (ii) – ausência de detalhamento dos passos para execução do método**, esclarece-se que a METHODOE atende parcialmente a essa lacuna. Grande parte das etapas apresenta orientações claras sobre como executá-la, porém, em algumas partes, em que são oferecidas muitas opções para o ontologista, há necessidade de detalhar melhor cada uma das opções.

Já no que concerne à **lacuna (iii) – validação em um cenário limitado**, esclarece-se que a validação da METHODOE não conseguiu superar essa lacuna, posto que a validação aconteceu em uma parte específica da OntoAgroHidro, e, portanto, em um cenário restringido. Apesar de aplicado em um escopo considerado pequeno se comparado com a quantidade de classes existentes na ontologia, o exercício de validação cumpriu o seu propósito.

Com relação à **lacuna (iv) – métodos muito específicos para o domínio biomédico**, não há uma dependência de domínio nas diretrizes apontadas em cada etapa da METHODOE. Embora a validação da metodologia tenha se dado em um campo específico, todo o seu desenvolvimento visou ultrapassar as particularidades de domínios, propondo orientações gerais e possibilitando escolhas ao ontologista.

No que tange à **lacuna (v) – situações em que o enriquecimento com base no método totalmente automático não produziu bons resultados**, destaca-se que a METHODOE foi integralmente idealizada pensando em uma sequência lógica de atividades dentro do processo de enriquecimento, sendo que se considera a intervenção humana em todas as cinco grandes fases. Porém isso não impede que algumas etapas ou até mesmo alguma fase por inteira seja automatizada, desde que os resultados sejam relevantes para o processo de enriquecimento.

A **lacuna (vi) – falta de orientações para enriquecimento de todos os componentes** também foi superada pela metodologia de enriquecimento, visto que um dos preceitos básicos para desenvolvimento da METHODOE é que ela oferecesse suporte para enriquecimento de todos os componentes da ontologia.

Em relação à **lacuna (vii) – dificuldade de acesso às fontes de informação por serem muito específicas**, salienta-se que a METHODOE não possui dependência de um tipo de fonte de informação específica. A metodologia oferece várias opções de fontes de informação para o ontologista e, assim, ele e sua equipe estão livres para escolher aquelas que mais se adequam ao seu domínio, ou aquelas mais acessíveis.

Além de superar grande parte das lacunas dos métodos abordados na Revisão de Literatura (RL), a METHODOE apresenta alguns dos aspectos favoráveis identificados nesses mesmos estudos da RL, como:

- **Independência de ferramenta específica:** as etapas da metodologia de enriquecimento não estão vinculadas a nenhuma ferramenta específica. Essa foi uma das funcionalidades nas quais se respaldou para desenvolvimento da METHODOE. Contudo, isso não impede o uso de softwares nas etapas da metodologia, pelo contrário, em algumas etapas (por exemplo no diagnóstico na ontologia, aplicação da técnica de extração da informação) da METHODOE, indica-se a utilização de softwares para facilitar a implementação da etapa. O diferencial é que, se o software for descontinuado, é possível utilizar outros do mesmo tipo, ou seja, que executem aquela tarefa de forma similar. Portanto a metodologia não é dependente de nenhuma ferramenta específica.
- **Possibilidade da realização do enriquecimento de axiomas:** as diretrizes da metodologia estão focadas no enriquecimento de todos os componentes das ontologias de domínio. Portanto, a partir das definições dos conceitos que são realizadas na fase 3 (organização do conhecimento) da METHODOE, é possível identificar axiomas e configurá-los na fase 4 (implementação do enriquecimento).
- **Detalhamento das etapas:** a maioria das etapas estão bem explicadas e facilitam a sua implementação por ontologistas.

- **Independência de domínio:** posto que na METHODOE apresentam-se diretrizes genéricas que em hipótese⁴⁸ podem ser aplicadas em variadas áreas do conhecimento.

A Metodologia para Enriquecimento de Ontologias de Domínio (METHODOE) fornece uma forma de enriquecimento de ontologias que visa superar as idiossincrasias de cada campo. Assim apresenta orientações gerais de como cada atividade pode ser realizada. Porém, nessa tentativa de ser um metamodelo para o processo de enriquecimento, a metodologia apresenta algumas falhas, que foram percebidas durante o exercício de enriquecimento da OntoAgroHidro. Nos parágrafos a seguir, elas serão discutidas.

Na etapa de avaliação da necessidade de enriquecimento (Etapa 1.B) da fase de planejamento do enriquecimento, a metodologia deveria apresentar exemplos de incompletudes para facilitar a identificação de problemas em sua estrutura pelo ontologista.

Na fase de aquisição de conhecimento, percebe-se que é necessário deixar mais claro na etapa de seleção das fontes de informação (Etapa 2.A) como cada Sistema de Organização do Conhecimento pode ser analisado, isto é, por onde começar a análise para averiguar se o SOC contém insumos para a realização do enriquecimento. Faltou estabelecer também critérios para auxiliar o ontologista a definir um limite de fontes, isto é, como saber se já se selecionou a quantidade suficiente. Também, na etapa de aplicação da técnica de extração da informação (Etapa 2.D), teve-se dificuldade na organização do conteúdo extraído dos SOCs, visto que a METHODOE não aponta quais elementos precisam ser extraídos, se são apenas aqueles conceitos considerados relevantes ou se a estrutura hierárquica (ou *cluster*, no caso dos tesouros) ao qual o conceito pertence também, assim como seus sinônimos.

Na fase de organização do conhecimento, é preciso deixar explícito nas diretrizes da etapa organização do conhecimento adquirido (Etapa 3.A) como fazer a definição do conceito de acordo com os princípios da Teoria do Conceito de Dahlberg. E, na etapa seleção do conhecimento adquirido (3.B), teve-se dificuldade para identificar o que seria considerado relevante para o enriquecimento, já que, nessa

⁴⁸ Em hipótese porque a metodologia não foi validada em mais de uma área do conhecimento.

etapa, não está prevista a análise da estrutura da ontologia, e, durante o exercício de aplicação, percebeu-se que tal análise é fundamental.

Ao longo do exercício de aplicação da METHODOE na ontologia OntoAgroHidro, percebeu-se que a metodologia possibilita a realização do processo de enriquecimento sem grandes problemas. A sistematização do processo de enriquecimento por meio das relações e interdependência entre as etapas permite toda a operacionalização do enriquecimento. Além disso, os documentos gerados em cada etapa facilitam a implementação da etapa seguinte.

Em síntese, o exercício de enriquecimento da OntoAgroHidro teve como principal foco funcionar como um teste de aplicação da METHODOE. Nessa direção, contribui-se para aperfeiçoamento da metodologia de enriquecimento em suas próximas versões. Com relação à contribuição para a ontologia OntoAgroHidro em si, acredita-se que esse exercício de enriquecimento pode servir como parâmetro para que os responsáveis pela ontologia apliquem a METHODOE em outras partes da OAH.

8 CONSIDERAÇÕES FINAIS

As ontologias de domínio são Sistemas de Organização do Conhecimento que possibilitam a estruturação de modelos de mundo por meio de seus componentes e de linguagens computacionais. Devido ao seu potencial na representação do conhecimento, elas podem ser utilizadas em diversas atividades humanas, como: interoperabilidade entre sistemas e base de dados heterogêneos, recuperação da informação, compartilhamento e reúso de conhecimento em uma área específica. É com razão que variadas áreas do conhecimento (Turismo, Agricultura, Medicina, Biologia, Direito) encontram nas ontologias a solução para seus problemas de representação e recuperação da informação.

Por serem construídas com base em conceitualizações ou abstrações que determinado grupo de especialistas faz de fenômenos do mundo físico ou social, as ontologias de domínio estão suscetíveis a limitações. Primeiro, porque a forma de conceber o mundo está sujeita a alterações. Segundo, porque o próprio mundo físico e social se transforma e evolui, caracterizando-se como um sistema dinâmico. Diante disso, as ontologias de domínio precisam acompanhar a dinamicidade do fenômeno que representa. Caso contrário, está sujeita a obsolescência e incompletudes.

A presente pesquisa partiu das questões com relação às ontologias de domínio serem desenvolvidas em várias áreas do conhecimento, e nem sempre são completas e atualizadas. Além disso, constatou-se na revisão de literatura a incipiência de uma metodologia abrangente para o processo de enriquecimento de ontologias de domínio, haja vista que se identificaram vários estudos que abordavam métodos específicos e aplicáveis a domínios e casos particulares. Sendo assim, traduziram-se tais problemas em duas questões de pesquisa: 1) como desenvolver uma metodologia para enriquecimento de ontologias de domínio que supere as lacunas dos atuais métodos? 2) O desenvolvimento de uma metodologia pode simplificar o processo de enriquecimento de ontologias de domínio? Frente a essa problemática, o **objetivo geral** desta pesquisa foi propor uma metodologia para o processo de enriquecimento de ontologias de domínio com base em construtos da engenharia ontológica e da Biblioteconomia e Ciência da Informação e que superasse as lacunas dos atuais métodos.

As duas questões de pesquisa foram respondidas durante a investigação. Com relação à primeira delas, a forma identificada para desenvolver a metodologia de

enriquecimento foi por meio da extração de insumos de estudos existentes na literatura das áreas de Engenharia Ontológica e Biblioteconomia e Ciência da Informação com o contributo do método de Análise de Conteúdo. No que tange à segunda questão, chegou-se à conclusão que uma metodologia pode sim simplificar o processo de enriquecimento, pois esclarece a relação entre as etapas do processo, bem como como realizá-las, além de orientar sobre o produto gerado e o tipo de profissional responsável por cada etapa.

O objetivo principal desta pesquisa foi alcançado mediante o desenvolvimento da *Methodology for Domain Ontology Enrichment* (METHODOE), demonstrada no [capítulo 6](#). A METHODOE possui cinco fases que possibilitam a realização de todas as atividades relacionadas ao processo de enriquecimento, desde o seu planejamento até o pós-enriquecimento. Entre as funcionalidades e características dessa metodologia, destacam-se: (1) apresentação de uma visão sistemática do processo de enriquecimento; (2) independência de domínio e de ferramentas específicas, possibilitando a aplicação em ontologias de domínios diversos; (3) independência de fontes de informação específicas e de difícil acesso; (4) apresentação de forma detalhada da maioria das etapas e com algumas exemplificações; (5) possibilidade da realização de todos os tipos de enriquecimento (lexical, conceitual, de relações e de axiomas).

Os procedimentos metodológicos adotados nesta investigação se demonstraram adequados para alcance dos objetivos específicos, assim como as temáticas discutidas no referencial teórico metodológico. O primeiro objetivo específico (caracterizar o processo de enriquecimento e desenvolver um fluxo para ele) foi alcançado parcialmente no [capítulo 3 – fundamentação teórica e metodológica](#) – e complementado no [capítulo 6](#), no qual se apresenta a METHODOE. Tal metodologia se configura como o fluxo para o processo de enriquecimento. O segundo objetivo específico (identificar insumos na área de engenharia ontológica e na Biblioteconomia e Ciência da Informação para o desenvolvimento da metodologia de enriquecimento) foi alcançado no [capítulo 5](#), no qual se explorou o material empírico com o respaldo do método de análise de conteúdo. O terceiro objetivo específico (demonstrar as principais lacunas nos métodos existentes para enriquecimento de ontologias de domínio) foi atingido no [capítulo 2](#), em que se apresenta a revisão sistemática da literatura. Nessa revisão, foi estabelecido o estado da arte da área de enriquecimento de ontologias de domínio

e apontadas as lacunas dos estudos. Já o quarto objetivo (dar subsídios aos desenvolvedores de ontologias para melhoria do processo de enriquecimento) foi contemplado no [capítulo 6](#), por meio do desenvolvimento da metodologia de enriquecimento de ontologias de domínio. Por último, o quinto objetivo específico (contribuir com os estudos sobre atualização de ontologias de domínio) foi atingido nos [capítulos 6 e 7](#), visto que a própria metodologia contribui para atualização por meio do processo de enriquecimento; e, no [capítulo 7](#), por meio de exercício de aplicação da metodologia, apresenta-se um estudo de caso.

Esta pesquisa partiu de alguns [pressupostos](#) descritos no capítulo inicial. No primeiro, argumentava-se que o enriquecimento de ontologias pode contribuir para que as ontologias representem o domínio com mais completude. Considerando a seção do referencial teórico que aborda o enriquecimento de ontologias, percebe-se que tal processo visa expandir uma ontologia, seja por meio de novas classes, seja por subclasses, relações, axiomas. Portanto constata-se que, nesse sentido, o enriquecimento pode contribuir para a completude das ontologias. Ademais, isso ficou evidenciado por meio da metodologia de enriquecimento desenvolvida (METHODOE), visto que, em uma das etapas, é possível identificar os problemas na estrutura da ontologia, principalmente problemas de incompletude. A partir desses problemas, é que se realiza o enriquecimento.

O segundo pressuposto, “metodologias de enriquecimento facilitam a atualização de conhecimento em ontologias de domínio”, confirmou-se, posto que a metodologia de enriquecimento desenvolvida nesta pesquisa possui diretrizes que contribuem para atualização de conhecimento em ontologias de domínio, por meio da inclusão de novos componentes. Esses componentes devem ser extraídos de fontes atuais do domínio, conforme recomendações da METHODOE.

No terceiro pressuposto, afirmava-se que “a Biblioteconomia e Ciência da Informação possui arcabouço teórico-metodológico que pode auxiliar no enriquecimento de ontologias de domínio”, o que se comprovou, já que na METHODOE constam recomendações de utilização do processo de análise de assunto para seleção das fontes de informação, além dos princípios da Teoria do Conceito (DAHLBERG, 1978) para criação das definições. Fez-se uso também de insumos identificados em documentos sobre manutenção e atualização de SOCs, que se mostram úteis, principalmente no que concerne à gestão e organização do processo de enriquecimento. Além disso, percebe-se que outros aportes da BCI

também encontram aplicação no enriquecimento de ontologias, como a Teoria da Análise Facetada de Ranganathan (1967), embora não se tenha explorado o potencial desses outros aportes na METHODOE.

Os resultados demonstraram que a metodologia de enriquecimento desenvolvida apresenta diretrizes pertinentes que possibilitam a aplicação em ontologias de domínio. Isso ficou evidenciado por meio da aplicação da METHODOE na ontologia OntoAgroHidro do domínio agrícola. Seguindo as diretrizes apresentadas em cada etapa da metodologia, foi possível realizar o enriquecimento de todos os componentes de uma parte específica da ontologia mencionada. O fluxo descrito na metodologia também se mostrou adequado, visto que apresenta um encadeamento das etapas, em que uma etapa é dependente da atividade realizada na etapa prévia. Além disso, a metodologia apresenta as principais coordenadas de cada etapa, com os documentos que serviam de insumos, os documentos que a etapa deveria gerar e o tipo de profissional responsável pela execução da etapa. Entretanto também foram percebidas falhas no que concerne ao detalhamento de determinadas etapas: a generalidade de algumas etapas pode deixar o ontologista confuso na aplicação do enriquecimento.

A principal contribuição desta pesquisa é a proposição de uma metodologia inédita que possibilita que o processo de enriquecimento de ontologias de domínio se torne claro e operacionalizável. Nessa perspectiva, a METHODOE contribui para que ontologias representem o conhecimento atualizado do domínio e de forma mais completa. Conforme já foi discutido aqui, são diversos os campos que fazem uso das ontologias. Entre esses campos está a agricultura, que é uma atividade essencial para a subsistência humana. Assim, o enriquecimento de uma ontologia desse campo, tal como o do exercício de validação realizado nesta tese, contribui para que essa ontologia continue cumprindo o seu papel, seja no auxílio à comunicação e compartilhamento do conhecimento entre especialistas, como artefato que possibilita a interoperabilidade entre diferentes bases de dados ou em uma aplicação no contexto da recuperação da informação. É primordial que esses instrumentos continuem sendo úteis nas diversas áreas onde são aplicados para solução de problemas de ordem prática, social ou teórica.

Ademais, as ontologias têm-se provado artefatos relevantes para que a Ciência da Informação lide com os desafios da representação e recuperação da informação no contexto digital. Portanto iniciativas que visem contribuir para o aprimoramento

desses instrumentos são oportunas para a CI. Outra contribuição desta investigação refere-se ao mapeamento de vários métodos para enriquecimento de ontologias, que possuem potencial para serem aplicados em outros SOCs e assim possibilitar a atualização desses instrumentos de forma semi ou automática.

Ainda como contribuição da pesquisa, aponta-se o percurso metodológico utilizado para desenvolvimento da metodologia de enriquecimento, ou seja, o respaldo em estudos empíricos para extração de insumos que podem ser generalizados em uma metodologia. Tal percurso se mostrou pertinente para o desenvolvimento da METHODOE e pode ser aplicável no desenvolvimento de metodologias para outros processos e atividades da engenharia ontológica.

8.1 Limitações da pesquisa

As limitações desta pesquisa podem ser divididas em dois grupos: aquelas que possuem relação com a abordagem metodológica adotada e aquelas relacionadas ao desenvolvimento e aplicação da metodologia de enriquecimento. Com relação às limitações do primeiro grupo, destacam-se: (1) a utilização apenas da análise de conteúdo como instrumento de coleta dos dados – já que se pressupõe que a aplicação de questionários com ontologistas e especialistas de domínio poderia trazer elementos pertinentes sobre a prática do processo de enriquecimento; (2) a restrição do material empírico a estudos sobre enriquecimento de ontologias e documentos sobre manutenção e atualização de SOCs – acredita-se que a exploração de estudos que aplicaram outros processos da engenharia ontológica, tais como evolução de ontologia, mudança ontológica e reengenharia ontológica, poderia apresentar elementos concernentes ao processo de enriquecimento; (3) a exploração de um maior número de estudos sobre manutenção e atualização de SOCs, especialmente sobre atualização de tesouros. Com relação ao segundo grupo de limitações: (1) a aplicação da metodologia em um cenário restrito, posto que o enriquecimento aconteceu em uma parte específica da OntoAgroHidro e não na ontologia como um todo; (2) os vieses com relação ao fato de a aplicação ter sido desenvolvida pelo proponente da pesquisa, que também foi o responsável pelo desenvolvimento da METHODOE.

8.2 Trabalhos futuros

No decorrer desta pesquisa, vislumbraram-se algumas direções que podem culminar em trabalhos futuros. A seguir apresentam-se alguns deles:

1. Aprimorar a METHODOE com base nas falhas identificadas durante o enriquecimento da OntoAgroHidro.
2. Desenvolvimento de *framework*, *software* ou *plugin* para editores de ontologias que facilitem a implementação da METHODOE por ontologistas.
3. Aplicação da METHODOE em escopos mais amplos, ou seja, na estrutura completa de ontologias de domínio e não apenas em um extrato, bem como aplicação em ontologias de outras áreas do conhecimento.
4. Aplicação de outro método de enriquecimento na OntoAgroHidro para comparar os resultados com o enriquecimento conduzido de acordo com as diretrizes da METHODOE.
5. Avaliação entre a versão original da OntoAgroHidro e a versão enriquecida para identificação das potenciais melhorias na ontologia.
6. Análise da possibilidade de integração da METHODOE com outros projetos da engenharia ontológica, evidenciando, assim, a incorporação do processo de enriquecimento com outros processos de manutenção e atualização ontológica.
7. Exploração com mais afinco de outros construtos teórico-metodológicos da BCI que possam ser úteis para enriquecimento de ontologias.
8. Exploração das técnicas empregadas na extração de informação para enriquecimento de ontologias em outros tipos de SOCs que são estudados pela BCI, tais como tesouros e taxonomias.

REFERÊNCIAS

- AGIRRE, E.; et al. **Enriching very large ontologies using the WWW**. 2000. Disponível em: <http://arxiv.org/abs/cs/0010026>. Acesso em: 25 jun. 2021.
- ALI, M.; et al. CLOE: a cross-lingual ontology enrichment using multi-agent architecture. **Enterprise Information Systems**, [S.l.], v. 13, n. 7-8, p. 1002-1022. 2019. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/17517575.2019.1592232>. Acesso em: 14 abr. 2021.
- ALMEIDA, M. B. Revisiting ontologies: a necessary clarification. **Journal of the American Society for Information Science and Technology**, [S.l.], v. 64, n. 8, p. 1682-1693, 2013. Disponível em: 10.1002/asi.22861. Acesso em: 20 jun. 2020
- ALMEIDA, M. B. A proposal to evaluate ontology content. **Applied Ontology**, Birmingham, v. 4, n. 3-4, p. 245-265, 2009. Disponível em: 10.3233/AO-2009-0070. Acesso em: 20 jun. 2020.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília-DF, v. 32, n. 3, p. 7-20, set./dez. 2003.
- ARP, R.; SMITH, B.; SPEAR, A. D. **Building Ontologies with basic formal ontology**. Cambridge, Massachusetts: The Mit Press, 2015.
- ASTRAKHANTSEV N. A.; FEDORENKO D. G.; TURDAKOV D. Y. Automatic enrichment of informal ontology by analyzing a domain-specific text collection. **Komp'juternaja Lingvistika i Intellektual'nye Tehnologii**. [S.l.], 2014 p. 29-42. Disponível em: https://www.ispras.ru/en/publications/automatic_enrichment_of_informal_ontology_by_analyzing_a_domain_specific_text_collection.pdf. Acesso em: 13 ago. 2021.
- BADA, M.; HUNTER, L. Enrichment of OBO ontologies. **Journal of Biomedical Informatics**. Colorado, v. 40, n. 3, jun. 2007, p. 300-315, 2006. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1532046406000827>. Acesso em: 10 maio 2021.
- BARBUR, G.; BLAGA, B.; GROZA, A. OntoRich - A support tool for semi-automatic ontology enrichment and evaluation. *In*: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTER COMMUNICATION AND PROCESSING, 7., 2011, ClujNapoca, Romani. **Processing [...]** Cluj-Napoca, Romania: IEEE, 2011. p. 129-132. Disponível em: 10.1109/ICCP.2011.6047855
- BARDIN, L. **Análise de conteúdo**. 4. ed. Lisboa: Edições, 2011.
- BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2016.
- BEN AMAR, F. B.; GARGOURI, B.; BEN HAMADOU, A. Domain ontology enrichment based on the semantic component of LMF-standardized dictionaries. *In*: WANG, M. (Org.). **Lecture notes in computer science (including subseries**

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Berlin: SpringerVerlag, 2013. p. 404–419.

BENDAOU, R.; TOUSSAINT, Y.; NAPOLI, A. Pactole: a methodology and a system for semi-automatically enriching an ontology from a collection of texts. In: INTERNATIONAL CONFERENCE ON CONCEPTUAL STRUCTURES, 16., 2008, Toulouse, France. **Anais [...]** Toulouse, France: [s.n.], 2008. p. 203–216. Disponível em: <https://hal.inria.fr/inria-00315530>. Acesso em: 21 fev. 2019.

BENZ, D. **Collaborative ontology learning**. 2007. Dissertação (Mestrado em Ciência da Computação) – Departamento de Ciência da Computação – Universidade de Freiburg, Freiburg, 2007.

BFO. Basic Formal Ontology website. 2019. Disponível em: <https://basic-formal-ontology.org/>. Acesso em: 20 jan. 2019.

BONACIN, R.; NABUCO, O. F.; PIEROZZI JUNIOR, I. Modeling the impacts of agriculture on water resources: semantic interoperability issues. In: International WETICE Conference. 23., 2014, [S.l.]. **Proceedings [...]**, [S.l.]: [s.n.], 2014, p. 447-452. Disponível em: <https://ieeexplore.ieee.org/document/6927099>. Acesso em: 04 ago. 2021.

BONACIN, R.; NABUCO, O. F.; PIEROZZI JUNIOR, I. Ontology models of the impacts of agriculture and climate changes on water resources: scenarios on interoperability and information recovery. **Future Generation Computer Systems**. v. 54, p. 423-434, 2016. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/137810/1/Ontology-Bonacin.pdf>. Acesso em: 04 ago. 2021.

BONACIN, R.; NABUCO, O. F.; PIEROZZI JUNIOR, I. Representação do conhecimento gerado pelo projeto Agrohidro: estrutura conceitual e aplicação computacional. In: SEMINÁRIO DA REDE AGROHIDRO. 3., 2015, Corumbá. **Anais [...]** Corumbá: Rede Agrohidro, 2015. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1030587/representacao-do-conhecimento-gerado-pelo-projeto-agrohidro-estrutura-conceitual-e-aplicacao-computacional>. Acesso em: 14 ago. 2021.

BOOSHEHRI, M.; et al. Ontology enrichment by extracting hidden assertional knowledge from text. **International Journal of Computer Science and Information Security**, [S.l.], v. 11, n. 5, p. 64-72, 2013. Disponível em: <http://arxiv.org/abs/1308.0701>. Acesso em 22 jun. 2020

BOOSHEHRI, M.; LUKSCH, P. An Ontology Enrichment Approach by Using DBpedia. In: INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE, MINING AND SEMANTICS, 5., 2015, Larnaca, Cyprus: Association for Computing Machinery, 2015. p. 1-11.

BRATKOVÁ, E.; KUČEROVÁ, H. Knowledge organization systems and their typology. **Revue of Librarianship**, [S.l.], v. 25, p. 1-11, 2014.

BREWSTER, C. et al. Data driven ontology evaluation. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 2004, Lisboa.

Anais [...] Lisboa: [s.n], 2004. Disponível em: <https://aclanthology.org/L04-1476/>. Acesso em: 22 fev. 2019.

BUITELAAR, P.; CIMIANO, P.; MAGNINI, B. **Ontology learning from text: methods, evaluation and applications**. Amsterdam: IOS Press. 2005.

CAMPOS, M. L. A. **A organização de unidades do conhecimento em hiperdocumentos**: modelo conceitual como espaço comunicacional para realização da autoria. 2001. 186 f. Tese (Doutorado em Ciência da Informação) – Conselho Nacional de Desenvolvimento Científico e Tecnológico/Instituto Brasileiro de Informação em Ciência e Tecnologia - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.

CAMPOS, M. L. A. C. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 22-32, 2004.

CAMPOS, M. L. A. O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, v. 15, n. 1, p. 220-238, 2010. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/36983>. Acesso em: 17 ago. 2020.

CAMPOS, M. L. A. Elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição. **Ciência da Informação**, Brasília, DF, v. 46, n. 1, p. 23-33, jan./abr. 2017. Disponível em: <http://revista.ibict.br/ciinf/article/view/4016/3455>. Acesso em: 17 ago. 2020.

CIMIANO, P.; HOTH, A.; STAAB, S. Learning concept hierarchies from text corpora using formal concept analysis. **JAIR - Journal of AI Research**, [S.l.], v. 24, p. 305-339. 2005.

CIMIANO, P.; VAOLKER, J. Text2Onto - a framework for ontology learning and data-driven change discovery. In: INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING AND INFORMATION SYSTEMS, 10., 2005, Alicante, **Proceedings [...]** Alicante: Springer, 2005.

DAHLBERG, I. A referent-oriented analytical concept theory of interconcept. **International Classification**. Frankfurt, v. 5, n. 3, p. 142-150, 1978.

DE NICOLA, A.; MIKSIKOFF, M.; NAVIGLI, R. A proposal for a unified process for ontology building: UPON. In: K.V., A.; J., D. and R., W., **Database and expert systems applications**. DEXA 2005, 2005, Springer, Berlin, Heidelberg. p.655-664.

DELLSCHAFT, K.; STAAB, S. On How to Perform a Gold Standard Based Evaluation of Ontology Learning. In: PROC. FIFTH INT'L CONF. SEMANTIC WEB, 2006, **Anais [...]** Heidelberg. Heidelberg: [s.n], 2006. p. 228-241. Disponível em: <http://gesec.deinf.ufma.br>. Acesso em: 23 fev. 2019.

EJEL, F.; et al. Enriching semantic relations of basic sciences ontology. **Knowl. Org.** v. 44, n. 5, p. 318-325, 2017. Disponível em: <https://irandoc.ac.ir/sites/fa/files/attach/article/ko4420175b.pdf>. Acesso em: 11 abr. 2021.

ELHADAD, D.; GABAY, M.; NETZER, Y. Automatic evaluation of search ontologies in the entertainment domain using text classification. In: **Applied semantic technologies: using semantics in intelligent information processing**. [S.l.]: Taylor and Francis, 2010. Disponível em: <https://www.cs.bgu.ac.il/~elhadad/papers/OntologyEvaluation.pdf>. Acesso em: 10 jun. 2019.

FAATZ, A.; STEINMETZ, R. **Ontology enrichment with texts from the WWW** [...] Darmstadt, Germany: [s.n.], 2002.

FARINELLI, F. **Realismo ontológico aplicado à interoperabilidade semântica entre sistemas de informação**: um estudo de caso no domínio obstétrico e neonatal. Tese (Doutorado em Gestão e Organização do Conhecimento), Belo Horizonte: Escola de Ciência da Informação - Universidade Federal de Minas Gerais, 2017.

FERNANDEZ-BREIS, J. T.; et al. Enriching the gene ontology via the dissection of labels using the ontology pre-processor language. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT. 17., 2010, Lisboa. **Anais [...]** Lisboa: Universidad de Murcia, 2010. p. 59-73. Disponível em: <http://www.cs.man.ac.uk/~stevensr/papers/ekaw2010.pdf>. Acesso em: 04 ago. 2021.

FERNÁNDEZ-LÓPEZ, M. Overview of methodologies for building ontologies. In: IJCAI- 99 WORKSHOP ON ONTOLOGIES AND PROBLEM-SOLVING METHODS, 1999, Stockholm, Sweden. **Proceedings [...]**, Stockholm, Sweden: [s.n.], 1999. p.1-13.

FERRÉ, S.; RUDOLPH, S. Advocatus diaboli: exploratory enrichment of ontologies with negative constraints. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT, 13, 2012, Galway. **Anais [...]** Springer: Heidelberg, 2012. p. 42-56. Disponível em: https://link.springer.com/chapter/10.1007%2F978-3-642-33876-2_7. Acesso: 12 abr. 2021.

FERREIRA, A. C. **Metodologia de revisão e atualização de tesouros aplicada ao Tesouro de Contas de Minas Gerais**: abordagem de pesquisa-ação. 2020. 433 f. Tese (Doutorado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2020.

FONSECA, F. The double role of ontologies in information science research. **Journal of the American Society for Information Science and Technology**, [S.l.] v. 58, n. 6, p. 786-793, 2007.

GANGEMI, A. et al. Qood grid: a metaontology-based framework for ontology evaluation and selection. In: INTERNATIONAL WORKSHOP ON EVALUATION OF ONTOLOGIES FOR THE WEB, 4., Edinburgh, UK, 2006, Edinburgh, UK. **Proceedings [...]** Edinburgh, UK, 2006. Acesso em: 13 abr. 2019.

GHARIB, T. F. et al. Enriching ontology concepts based on texts from WWW and corpus. **Journal of Universal Computer Science**, v. 18, n. 16, p. 2234-2251, 2012.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GILCHRIST, A. Thesauri, taxonomies and ontologies: an etymological note. **Journal of Documentation**, v. 59, n. 1, p. 7-18, 2003.

GILLANI, S.; KO, A. Incremental ontology population and enrichment through semantic-based text mining: an application for IT audit domain. **International Journal on Semantic Web & Information Systems**, [S.l.], v. 11, n. 3, jul. 2015. Disponível em: <https://dl.acm.org/doi/10.4018/IJSWIS.2015070103>. Acesso em: 4 abr. 2021.

GOMEZ-MORENO, P. U.; MESTRE-MESTRE, E. M. Automatic domain specific learning: towards a methodology for ontology enrichment. **Revista de Linguas para Fines Específicos**, v. 23, n.2, p. 63-85. 2017.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. J. Towards a method to conceptualize domain ontologies. In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE (ECAI'96), 12., 1996, Budapest, Romênia. **Anais [...]** Budapest, Romênia: Facultad de Informática (UPM), 1996. Disponível em: . Acesso em: 12 ago. 2018.

GÓMEZ-PÉREZ, A.; MANZANO-MACHO, D. An overview of methods and tools for ontology learning from texts. **The Knowledge Engineering Review**, [S.l.],v. 19, n.3, p. 187-212. 2004.

GOMÉZ-PÉREZ, A.; ROJAS-AMAYA, D. Ontological Reengineering for Reuse. In: EUROPEAN WORKSHOP (EKAW), 11, 1999, Dagstuhl Castle, Alemanha. **Anais [...]** Dagstuhl Castle, Alemanha. 1999.

GRUBER, T. R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In: INTERNATIONAL WORKSHOP ON FORMAL ONTOLOGY, 1993, Padova, Italy. **Proceedings [...]** Palo Alto, CA: Stanford Knowledge Systems Laboratory, 1993.

GRUBER, T. Toward principles for the design of ontologies used for knowledge sharing. **International journal of human-computer studies**, v. 43, n. 5–6, p. 907-928, 1995. Disponível em: <http://dx.doi.org/10.1006/ijhc.1995.1081>. Acesso em: 25 jun 2021.

GRUNINGER, M.; FOX, M. S. Methodology for the Design and Evaluation of Ontologies. In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING (IJCAI). 1995, **Anais [...]** [S.l.]; [s.n.], 1995.

GUARINO, N. Formal ontology in information systems In: _____. **First formal ontology in information systems**, Trento, Italy: IOS press, 1998 p. 81-97.

GUARINO, N. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In: Paziienza, M. T. (Ed.). **Information extraction a multidisciplinary approach to an emerging information technology**: Lecture Notes in Computer Science 1299, Springer Verlag. 1997. p. 139-170.

GUARINO, N.; GIARETTA, P. Ontologies and knowledge bases: towards a terminological clarification. In: _____ (Ed.). **Towards very large knowledge bases: knowledge building & knowledge sharing**, 1995. p.25-32.

GUARINO, N.; OBERLE, D.; STAAB, S. What is an Ontology? In: STAAB, S.; STUDER, R. (Ed.) **Handbook on Ontologies**. International Handbooks on Information Systems. Heidelberg: Springer Berlin, 2009. p.1-17.

GUARINO, N.; WELTY, C. A formal ontology of properties. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT, 12., 2000, Juan-les-Pins, France. **Anais [...]** 2000, Juan-les-Pins, France: Springer Berlin Heidelberg, October 2–6. p.97-112.

GUERRAM, T.; MELLAL, N. Domain independent approach for ontology semantic enrichment. In: INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING, 7., 2018, Stroudsburg. **Conference [...]** Stroudsburg: Association for Computational Linguistics, 2018.

GUIZZARDI, GIANCARLO; WAGNER, GERD. Towards ontological foundations for agent modelling concepts using the unified foundational ontology (UFO). In: INTERNATIONAL CONFERENCE ON AGENT-ORIENTED INFORMATION SYSTEMS II, 6. 2004, Berlin: Springer, 2004. p.110-124.

HARB, Ali.; HAJLAOUI, Kafil. Enhanced semantic automatic ontology enrichment. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS DESIGN AND APPLICATIONS, 10., 2010, Cairo. **Conference [...]** Cairo, 2010. p. 1113-1118. Disponível em: <http://doi.org/10.1109/ISDA.2010.5687038>. Acesso em: 20 abr. 2021

HASHIMY, A. S; KULATHURAMAIYER, N. Ontology enrichment with causation relations. In: IEEE CONFERENCE ON SYSTEMS, PROCESS & CONTROL (TCSPC 2013), 2013, Kuala Lumpur, Malaysia. **Anais [...]** Kuala Lumpur, Malaysia: Institute of Electrical and Electronics Engineers 2013. p.186-192.

HAZMAN, M.; EL-BELTAGY, S. R.; RAFEA, A. Ontology Learning from Domain Specific Web Documents. **International Journal of Metadata, Semantics and Ontologies**, [S.l.], v. 4, n. 1-2, pp: 24 – 33. 2009.

HJØRLAND, B. **Concept in knowledge organization (KO)**. The epistemological lifeboat, 2008. Disponível em: http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/concept_in_knowledge_organization.htm. Acesso em: 3 set. 2018.

HJØRLAND, B. Theories are knowledge organizing systems (KOS). **Knowledge Organization**, [S.l.], v. 42, n. 2, p. 113-128, 2015.

HJØRLAND, B. Theories of knowledge organization-theories of knowledge. **Knowledge Organization**, [S.l.], v. 40, n. 3, p. 169-181, 2013. Disponível em: <http://www.isko-de.org/data/uploads/KOAndTheoriesOfKnowledge.pdf>.

HLOMANI, H.; STACEY, D. A. An Extension to the Data-driven ontology evaluation. In: INTERNATIONAL CONFERENCE ON INFORMATION REUSE AND

INTEGRATION (IRI), 15., 2014, Ontário, Canadá. **Proceedings [...]** Ontário, Canadá: IEEE, 2014.

HLOMANI, H.; STACEY, D. A. Contributing evidence to data-driven ontology evaluation: Workflow ontologies perspective. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND ONTOLOGY DEVELOPMENT, 5., Vilamoura, Portugal. **Proceedings [...]** Vilamoura, Portugal: SCITEPRESS, 2013. p. 207-213.

HJØRLAND, B. Domain analysis in information science: eleven approaches-traditional as well as innovative. **Journal of Documentation**, v. 58, n. 4, p. 422-462, 2002.

HODGE, G. Systems of knowledge organization for digital libraries: beyond traditional. Washington, DC: The Digital Library Federation, 2000.

IDOUDI, R.; et al. Ontology knowledge mining for ontology conceptual enrichment. **Knowledge Management Research & Practice**, [S.l.], v. 17, n. 2, p. 151- 160, nov. 2018. Disponível em: <http://doi.org/10.1080/14778238.2018.1538599>. Acesso em: 23 abr. 2021.

INTERNATIONAL STANDARD ORGANIZATION. ISO 25.964: thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval. Geneve: International Standard Organization, 2011.

KATSUMI, M.; GRUNINGER, M. Chosing ontologies for reuse. **Applied Ontology**, v.12, n. 3, p. 195-221. 2017.

KOKLA, M.; PAPADIAS, V.; TOMAI, E. Enrichment and population of a geospatial ontology for semantic information extraction. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, [S.l.], v. 42, n. 14, p. 309-314, 2018. Disponível em: <http://doi.org/10.5194/isprs-archives-XLII-4-309-2018>. Acesso em: 14 abr. 2021.

KUNTARTO, G. P.; et al. Dwipa ontology III: implementation of ontology method enrichment on tourism domain. **International Journal on Smart Sensing and Intelligent Systems**, Jakarta, v. 10, n. 4, p. 903-919, dez. 2017. Disponível em: https://www.researchgate.net/publication/322294468_Dwipa_Ontology_III_Implementation_of_Ontology_Method_Enrichment_on_Tourism_Domain. Acesso em: 14 abr. 2021.

KUNTARTO, G. P.; ISYHRANI, S.; GUNAWAN, I. P. Performance of k-means clustering algorithm in enriching a new concept of amenities into dwipa ontology III within the Indonesia tourism domain. In: INTERNATIONAL SEMINAR ON RESEARCH OF INFORMATION TECHNOLOGY AND INTELLIGENT SYSTEMS (ISRITI), [s.n.], 2019. Yogyakarta. **Anais [...]** Yogyakarta: IEEE, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9034563/>. Acesso em: 12 abr. 2021.

LIMA, G. A. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 25, n. Especial, p. 57-97, 2020.

LIMA, G. A.; MACULAN, B. C. M. S. Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. **Ciência da Informação**, Brasília, v. 46, n. 1, p. 60-72, jan./abr. 2017.

LIU, K. **Ontology enrichment from free-text clinical documents: a comparison of alternative approaches**. 2012. 168 f. Tese (Doutorado em Filosofia) - School of Medicine, University of Pittsburgh, Pittsburgh, 2012. Disponível em: <http://d-scholarship.pitt.edu/10843/>. Acesso em: 10 abr. 2021.

LONSDALE, D. W.; et al. Reusing ontologies and language components for ontology generation. **Data Knowl. Eng.**, [S.l.] v. 69, p. 318-330. 2010.

LOZANO-TELLO, A.; GÓMEZ-PÉREZ, A. Ontometric: A method to choose the appropriate ontology. **Journal of Database Management, Hershey, Pensilvânia**, [S.l.] v. 15, n. 2, 18 p. 2004. Disponível em: . Acesso em: 10 set. 2014.

MAEDCHE, A.; STAAB, S. Measuring similarity between ontologies. *In: EUROPEAN CONFERENCE ON KNOWLEDGE ACQUISITION AND MANAGEMENT, 2002*, Madri, Espanha. **Proceedings [...]** Madri, Espanha: [s.n], 2002. Disponível em: . Acesso em: 20 set. 2019.

LUKSOVÁ, I. **Ontology enrichment based on unstructured text data**. 2013. 83 f. Dissertação (Mestrado em Software systems) - Faculty of Mathematics and Physics, Charles University, Praga, 2013. Disponível em: <https://is.cuni.cz/webapps/zzp/download/120119086>. Acesso em: 11 abr. 2021.

MAEDCHE, A. **Ontology learning for the semantic web**. [S.l.]: Springer. 2002.

MAEDCHE, A.; STAAB, S. Measuring similarity between ontologies. *In: EUROPEAN CONFERENCE ON KNOWLEDGE ACQUISITION AND MANAGEMENT, 2002*, Madri, Espanha. **Proceedings [...]** Madri, Espanha: [s.n], 2002. Disponível em: . Acesso em: 20 set. 2019.

MARCONDES, C. H.; CAMPOS, M. L. A. Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. **Ponto de Acesso**, Salvador, v. 2, n. 1, p. 107-136, jun./jul. 2008.

MARCONI, M.; LAKATOS, E. M. **Fundamentos de Metodologia científica**. 7. ed. São Paulo: Atlas, 2016.

MARTÍNEZ-ÁVILA, D.; FOX, M. J. The construction of ontology: a discourse analysis. *In: SMIRAGLIA, R. P.; LEE, Hur-Li. **Ontology for Knowledge Organization***. Würzburg: Ergon Verlag, 2015, p. 13-37.

MENDONÇA, F. M. **Ontoforinfoscience**: metodologia para construção de ontologias pelos cientistas da informação - Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA). 2015. Tese de Doutorado (Doutorado em Ciência da Informação). Universidade Federal de Minas Gerais. 2015.

MOREIRO GONZÁLEZ, J. A. **Linguagens documentárias e vocabulários semânticos para a web**: elementos conceituais. Salvador: EDUFBA, 2011.

MUSTAPHA, N. B., AUFAURE, M., BAAZHAOUI-ZGHAL, H. Towards an Architecture of Ontological Components for the Semantic Web. In: WEB INFORMATION SYSTEMS MODELING WORKSHOP (WISM), Wuhan, China. **Proceedings [...]** Wuhan, China. 2006.

NATIONAL INFORMATION STANDARDS ORGANIZATION. ANSI/NISO Z39.19-2005: guidelines for the construction, format, and management of monolingual thesauri. 2005.

NAVES, M. M. L. Análise de assunto: concepções. **Revista de Biblioteconomia de Brasília**, v. 20, n.2, p. 215-226. 1996.

NAVIGLI, R.; VELARDI, P. Ontology Enrichment Through Automatic Semantic Annotation of On-Line Glossaries. *In*: INTERNATIONAL CONFERENCE ON KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT, 2006, Stanford, CA. **Proceedings [...]**. Berlin: Springer Verlag, 2006. p. 126–140.

NOY, N. F.; McGUINNESS, D. L. **Ontology Development 101**: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

ORME, A.; YAO, H.; ETZKORN, L. Indicating ontology data quality, stability, and completeness throughout ontology evolution. **Journal of Software Maintenance and Evolution: Research and Practice**, Orme, TN, v. 19, n. 1, p. 49-55, jan./fev. 2007. Disponível em: Acesso em: 20 jan. 2020.

OTLET, P. **Traité de documentation: le livre sur le livre**. Théorie et pratique. Liège: Centre de Lecture Publique de la Communauté Française de Belgique, 1989. 432p.

OUYANG, L. et al. A method of ontology evaluation based on coverage, cohesion and coupling. In: INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY. 8., 2011, Shanghai. **Proceedings [...]** Shanghai: IEEE, 2011. p. 2451-2455.

PETASIS, G.; et al. Ontology population and enrichment: State of the art. In: Knowledge-driven multimedia information extraction and ontology evolution. [S.l.]: Springer-Verlag, 2011, p. 134-166.

PETERS, W. Text-based Legal Ontology Enrichment. In: CEUR WORKSHOP PROCEEDINGS. 465., 2009, Barcelona. **Anais [...]** Barcelona: Institute of Law and Technology, 2009. Disponível em: <http://ceur-ws.org/Vol-465/paper7.pdf>. Acesso em: 04 ago. 2021.

POVEDA-VILLALÓN, M. **Ontology Evaluation**: a pitfall-based approach to ontology diagnosis. 2016. 236 f. Tese (Doutorado em Inteligência Artificial). Departamento de Inteligência Artificial Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, 2016.

QUESADA-MARTÍNEZ, M.; et al. Approaching the axiomatic enrichment of the Gene Ontology from a lexical perspective. **Artificial Intelligence in Medicine**, [S.l.] v. 65, n.

1, p. 35-48, set. 2015. Disponível em: <http://doi.org/10.1016/j.artmed.2014.09.003>. Acesso em: 11 abr. 2021.

RANGANATHAN, S. R. Prolegomena to library classification. New York: Asia Publishing House, 1967.

REITER, Nils; BUITELAAR, Paul. Lexical enrichment of a human anatomy ontology using wordnet. 2008. Disponível em: <https://www.cl.uni-heidelberg.de/~reiter/publications/Reiter2008aa.pdf>. Acesso em: 04 abr. 2021.

SCHIESSL, M.; BRÄSCHER, M. Do texto às ontologias: uma perspectiva para a ciência da informação. **Ciência da Informação**, Brasília, v. 40, n. 2, maio/ago. 2011.

SHAMSFARD, M; BARFOROUSH, A. A. Learning Ontologies from Natural Language Texts. **Human-Computer Studies**, v. 60, n. 1, p. 17-63, 2004.

SILVA, D. L. **Uma proposta metodológica para construção de ontologias**: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação. Dissertação (Mestrado em Ciência da Informação). Universidade Federal de Minas Gerais. 2008.

SMITH, B. Basic Concepts of Formal Ontology. In: Guarino, N. (Ed.). **Formal Ontology in Information Systems**: IOS Press, 1998a. p.19-28.

SMITH, B. Ontology. In: Floridi, L. (Ed.). **The Blackwell Guide to the Philosophy of Computing and Information**. Oxford: Blackwell, 2003. p.155-166.

SMITH, B.; WELTY, C. Ontology: Towards a new synthesis. In: International Conference on Formal Ontology in Information Systems (FOIS 2001), 2., 2001, Ogunquit, Maine, USA. ACM Press, p.3-9.

SOERGEL, D. Knowledge Organization Systems. Overview Dagobert Soergel 504. [S.l.: s.n.], 2009. Disponível em: <http://www.dsoergel.com/UBLIS514DS-08.2a-1Reading4SoergelKOSOverview.pdf>.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the Association for Information Science and Technology**, [S.l.]: Syracuse University, v. 50, n. 12, p. 1119, 1999.

SOERGEL, D. **Indexing languages and thesauri**: Construction and maintenance. Los Angeles, CA: Melville. 1974. 632p.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. **Knowledge Organization**, [S.l.], v. 39, n. 3, p. 179–192, 2012.

STANFORD. **Grupo de Processamento de Linguagem Natural de Stanford**. 2019. Disponível em: <https://nlp.stanford.edu/software/tokenizer.html>. Acesso em: 20 jul. 2021. Conferir data de acesso

STOJANOVIC, L. **Methods and tools for ontology evolution**. 2004. 249 f. Tese (Doutorado em Economia). Universidade de Fridericana para Karlsruhe, Alemanha. 2004.

STUDER, R.; BENJAMINS, V.; FENSEL, D. Knowledge Engineering: Principles and Methods. **Data and Knowledge Engineering**, [S.l.] v. 25, n. 1-2, p. 161-197, 1998.

SUÁREZ-FIGUEROA, M. C. **NeOn methodology for building ontology networks: specification, scheduling and reuse**. 2010. Tese (Doutorado), Facultad de Informatica da Universidad Politécnica de Madrid, Madrid, 2010.

SURE, Y.; STAAB, S.; STUDER, R. On-to-knowledge methodology (OTKM). In: STAAB, S.; STUDER, R. (Ed.). *Handbook on Ontologies*. Berlin: Springer Berlin Heidelberg, 2004. Cap. 6, p.117-132.

TABOADA, M.; et al. A new synonym-substitution method to enrich the human phenotype ontology. **BMC Bioinformatics**. [S.l.], v. 18, n. 446, out. 2017. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-18587>. Acesso em: 04 abr. 2021.

TENNIS, J. T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? **BJIS**, Marília (SP), v. 6, n. 1, p.3-15, jan./jun. 2012. Disponível em: Acesso em: 4 jan. 2020.

TÖPPER, G.; KNUTH, M.; SACK, H. DBpedia ontology enrichment for inconsistency detection. In: I-SEMANTICS Int. Conf. on Semantic Systems Sept., 8., **Proceedings [...]** Austria: Graz, 2012.

USCHOLD, M.; KING, M. Towards a methodology for building ontologies. In: Workshop on Basic Ontological Issues in Knowledge Sharing, **Proceedings [...]** United Kingdom: University of Edinburgh, 1995.

VALARAKOS, A. G. *et al.* A name-matching algorithm for supporting ontology enrichment. In: HELLENIC CONFERENCE ON AI, 3., 2004, Samos, Greece. **Proceedings [...]**. VOUIROS, G.A.; PANAYIOTOPOULOS, T. (Org.). Berlin: Springer-Verlag, 2004. p. 381– 389.

VALARAKOS, A. G.; et al. **A Methodology for enriching a multi-lingual domain ontology using machine learning**. [S.l.: s.n.], 2001. Disponível em: <http://www.iit.demokrtios.gr/skel/crossmarc>. Acesso em: 15 maio 2020

VELARDI, P.; FABRIANI, P.; MISSIKOFF, M. Using text processing techniques to automatically enrich a domain ontology. In: Proceedings of the international conference on Formal Ontology in Information Systems. 1., 2001, Trento, Italy. **Anais [...]** Trento, Italy: [s.n.], 2001. p. 270-284. Disponível em: <http://doi.org/10.1145/505168.505194>. Acesso em 10 abr. 2021.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277-286, 1997.

W3C. **Semantic Web Activity**. [S.l.: s.n.]. 2020. Disponível em: <https://www.w3.org/2001/sw/>. Acesso em: 15 maio 2020.

ZAHRA, F. M.; et al. Ferramentas para aprendizagem de ontologias a partir de textos. **Perspectivas em Ciência da Informação**, v. 19, n. 1, p. 3-21, jan./mar. 2014.

ZAVITSANOS, E.; PALIOURAS, G.; VOUIROS, G. A. Gold standard evaluation of ontology learning methods through ontology transformation and alignment. **Transactions on Knowledge and Data Engineering**, [S.l: s.n.], v. 23, n. 11, p. 1635-1648, 2011.

ZENG, M. L. Knowledge organization systems (KOS). **Knowledge Organization**, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

APÊNDICE A – PROTOCOLO DE REVISÃO SISTEMÁTICA DE LITERATURA

Objetivo

Este Protocolo de Revisão Sistemática da Literatura (PRSL) apresenta a estrutura metodológica para a execução da etapa de revisão da literatura sobre métodos para o processo de enriquecimento de ontologias de domínio. Para construção desse protocolo respaldou-se no método proposto por Kitchenham (2004) e no software Start para gestão dos documentos. Sendo assim, os tópicos deste protocolo estão relacionados às fases do *Start*.

Equipe

Nome	Papel	Afiliação
Gercina Angela de Lima	Orientadora	PPGGOC/UFMG
Webert Júnio Araújo	Aluno de doutorado	PPGGOC/UFMG

Pergunta de Pesquisa

Quais são os trabalhos que aplicaram métodos para enriquecimento de ontologias de domínio publicados entre 2000 e 2019?

O quadro apresenta as perguntas de pesquisa secundárias desta RSL:

Pergunta	Descrição da Pergunta
P1	Quais as características dos métodos para enriquecimento de ontologias de domínio?
P2	Quais as lacunas existentes nesses métodos?
P3	Quais são os aspectos positivos dos métodos existentes?

Bases de Dados

Para definição sobre quais bases dados seriam consideradas, realizou-se uma busca exploratória no Google Acadêmico e a partir dos resultados das primeiras páginas, identificou-se aquelas bases que possuíam mais estudos referentes à temática indexados.

As seguintes bases de dados são utilizadas nesta RSL:

- *ACM Digital Library*;
- *Google Acadêmico*;
- *IEEE*;
- *Scopus*;
- *Web of Science*.

Termos de Busca

Termos	Sinônimos/variações	Tradução
Ontology	ontologies	ontologia
Enrichment	enriching, enrich	enriquecimento

Idiomas: Inglês e Português.

Período temporal: Após o ano 2000.

Strings de Busca

Base de dados	String utilizada
ACM Digital Library	Publication title (ontolog*) AND publication title (enrich*)
Google Acadêmico	Allintitle "ontology enrichment" - período temporal: 2000 a 2019
IEEE	Document title "ontolog*" AND document title "enrich"
Scopus	title(ontolog* AND enrich*)
Web of Science	TI=(ontolog* AND enrich*)

Critérios de Inclusão

Os Critérios de Inclusão dos estudos recuperados são apresentados no quadro

Critério	Descrição do Critério de Inclusão
C11	Título possui relação com a temática enriquecimento de ontologia de domínio;
C12	Resumo possui relação com a temática enriquecimento de ontologia de domínio;
C13	Serão incluídos trabalhos publicados e disponíveis integralmente nas bases de dados científicas buscadas;
C14	Serão incluídos trabalhos que tenham aplicado um método de enriquecimento em ontologia de domínio;

Critérios de Exclusão

Os Critérios de Exclusão dos estudos recuperados são apresentados no quadro

Critério	Descrição do Critério de Exclusão
CE1	Título não possui relação com a temática enriquecimento de ontologias de domínio;
CE2	Resumo não possui relação com a temática enriquecimento de ontologia de domínio;
CE3	Serão excluídos trabalhos que não apresentem resumo/abstract;
CE4	Serão excluídos <i>short papers</i> e trabalhos disponibilizados em bases <i>pre-print</i> ;
CE5	Serão excluídos documentos que não estejam em português ou inglês;
CE6	Serão excluídos trabalhos que não estejam disponíveis em texto completo;
CE7	Serão excluídos os trabalhos que não apliquem o processo de enriquecimento em ontologias de domínio.

Processo de Seleção das Publicações

1. Inicialmente se executou as Strings de Busca adaptadas em cada uma das bases de dados. A partir da lista de estudos retornados, exportaram-se os resultados em formato RIS, para importação na ferramenta StArt.
2. Na etapa de seleção, executou-se a análise do título de cada um dos estudos, descartando aqueles que claramente não possuíam relação com a temática da revisão ou que estivessem relacionados com os critérios de exclusão. Em seguida realizou-se a análise do resumo dos estudos restantes.
3. Ainda na etapa de seleção, realizou-se a leitura do texto completo dos estudos restantes, aplicando-se os critérios de inclusão e exclusão.
4. Na etapa de extração, realiza-se uma leitura atenta dos estudos para extração das informações pertinentes às perguntas de pesquisa definidas.
5. Finalizada a etapa de extração dos dados, realiza-se a análise, interpretação e documentação dos resultados encontrados nos estudos, com a redação do capítulo de revisão.

Extração de informações

Para extração da informação definiu-se os campos apresentados no Quadro abaixo.

Campo	Tipo	Conteúdo
Tipo de enriquecimento	PickonMany	{léxico, conceitual, relações, axiomas}
Técnica aplicada	PickonMany	{Estatística, Machine Learning, PLN }
Detalhamento da técnica	Text	
Fonte de informação / conhecimetro	Text	
Realizou algum tipo de validação	Pickonlist	{Sim,Não}
Descrição das etapas	Pickonlist	{Sim,Não}
Detalhamento das etapas	PickonMany	{Sim,Não, Parcialmente}
Principais lacunas	Text	
Principais aspectos favoráveis	Text	

Legenda: Tipo: Pickonlist (Escolha um na lista); PickonMany (Escolha em vários); Text (Campo texto aberto).

APÊNDICE B – DADOS SOBRE DOCUMENTOS DA REVISÃO DE LITERATURA

Referência	Tipo de fonte de informação	Tipo de enriquecimento					Técnica aplicada	Detalhamento das etapas			Principais lacunas	Aspectos favoráveis
		Lexical	Conceitual	Relação		Axioma		sim	não	parcial		
				Taxonômica	Não taxonômica							
FABRIANI e MISSIKOFF (2001)	corpus textual		x	x	x		Named Entity Recognition (NER); Parsing; POS tagging; Chunking (sentence segmentation); análise gramática		x		As técnicas exigem a supervisão de um especialista (linguista computacional)	Uso de fonte de informação de fácil acesso
FAATZ e STEINMETZ (2002)	Corpus textual		x				Análise estatística; medidas de similaridade		x		Enriquecimento de apenas um componente; falta de detalhamento das etapas	Uso de fonte de informação de fácil acesso;
VALARAKOS et al (2004)	Corpus anotado semanticamente	x					Machine Learning		x		Fontes de dados específicas; falta de detalhamento das etapas; enriquecimento de apenas um componente	Não identificado
NAVIGLI e VELARDI (2006)	Tesouro (Art e Architecture Thesaurus)	x	x	x	x		Análise sintática (POS tagging, NER); Machine Learning		x		Fonte de informação não aplicável para todos os domínios; as técnicas exigem a supervisão de um especialista (linguista computacional)	Não identificado
BADA e HUNTER (2007)	ontologias	x	x	x	x		tokenização	x			Método muito específico para o domínio biomédico; fonte de informação específica	As etapas do método estão bem descritas

BENDAOU, TOUSSAINT E NAPOLI (2008)	corpus textual;própria ontologia	x	x	x	x		Parsing; Formal Concept Analysis (FCA)			x	Falta de detalhamento de algumas etapas;	Uso de fonte de informação de fácil acesso; Independente de domínio
PETERS (2009)	corpus textual		x	x	x		Análise sintática (tokenização, sentence splitting, POS tagging, lematização); Análise estatística (TF-IDF, Mutual Information); Padrão Léxico sintático (Headword matching, Hearst)			x	As técnicas exigem a supervisão de um especialista (linguista computacional)	Uso de fonte de informação de fácil acesso
REITER e BUITELAAR (2008)	Base léxica (WordNet); Website (Wikipedia); Corpus especializado (British National Corpus)	x					Word Sense Disambiguation (WSD)			x	Enriquecimento de apenas um componente; fonte de informação não aplicável para todos os domínios ou idiomas	Não identificado
FERNANDEZ-BREIS et al (2010)	ontologia					x	Extração de padrões léxicos			x	Método muito específico para domínio biomédico; enriquecimento de apenas um componente	Não identificado
HARB e HAJLAOUI (2010)	própria ontologia; corpus textual		x				Análise morfosintática; Técnica de Data Mining (Association rules); Medida estatística (Mutual Information)		x		Falta de detalhamento das etapas; enriquecimento de apenas um componente	Uso de fonte de informação de fácil acesso
MONACHESI e MARKUS (2010)	Mídias sociais; Dataset (DBpedia)	x	x	x	x		Medida de similaridade (co-ocorrência,			x	Fonte de informação não aplicável para todos os domínios; as técnicas exigem a	Não identificado

							similaridade de Cosine)				supervisão de um especialista	
LIU (2011)	Relatórios clínicos;	x	x	x	x		Hearst Lexico Sintatic Pattern (LSP); POS tagging; NER; Análise estatística; Medidas de similaridade	x			As técnicas exigem a supervisão de um especialista (linguista computacional)	As etapas do método estão bem descritas
FERRÉ e RUDOLPH (2012)	ontologia					x	Navegação manual na estrutura da ontologia			x	Enriquecimento de apenas um componente	Proposta de enriquecimento raro na literatura
BEM AMAR, GARGOURI e BEM HAMADOU (2013)	Dicionários Léxicos		x	x	x		Análise sintática (POS tagging, tokenização, parsing)			x	Fonte de informação específica;	Não identificado
GROZA, BARBUR e BLAGA (2013)	Páginas Web	x	x	x	x		Análises estatísticas; Análise sintática (Named Entity Recognition, POS tagging)		x		Falta detalhamento das etapas	Software guia o processo de enriquecimento
LUKSOVÁ (2013)	Corpus textual		x	x	x		Machine Learning; Análise sintática (lematização, análise morfológica, parsing)	x			As técnicas exigem a supervisão de um especialista (linguista computacional)	Uso de fonte de informação de fácil acesso;
ASTRAKHANTSEV, FEDORENKO E TURDAKOV (2014)	corpus textual	x	x		x		Detecção de sentença; Tokenização; POS Tagging; Lematização; Word Sense Disambiguation;		x		Falta detalhamento das etapas; enriquecimento totalmente automático não produziu bons	Uso de fonte de informação de fácil acesso

							Medidas estatísticas (TF-IDF, Contagem de palavras, Frequência, etc); Algoritmo de Machine Learning				resultados; aplicação de várias técnicas que exigem a supervisão de um especialista	
BOOSHEHRI e LUKSCH (2015)	Dados estruturados em Linked Data				x		Análise sintática (tokenização, remoção de stop words, POS tagging, lemmatization)			x	Enriquecimento de apenas um componente	Não identificado
GILLANI e KO (2015)	Corpus textual; recursos léxicos (Wiktionary, WordNet, glossários)				x		Análise sintática (tokenização, POS tagging, stemming, remoção de stop words, lematização); análise estatística (cálculo de frequência); expansão de termos			x	Enriquecimento de apenas um componente; uso de fonte de informação não aplicável a todos os domínios.	Software guia parte do processo de enriquecimento.
QUESADA-MARTÍNEZ et al (2015)	ontologias					x	Análise léxica			x	Enriquecimento de apenas um componente; método muito específico para o domínio biomédico; fonte de informação específica	Proposta de enriquecimento raro na literatura
EJEI et al (2017)	ontologia; tesouro				x	x	Extração de padrões léxicos			x	Fonte de informação específica; enriquecimento de apenas um componente; falta de detalhamento das etapas.	Não identificado

GÓMEZ-MORENO E MESTRE-MESTRE (2017)	corpus especializado (Corpus of Contemporary American English)	x	x	x			Análise estatística (high frequency, Normalized Pointwise Mutual Information, peso estatístico); Análise sintática (tokenização, filtragem léxica, n-grams)			x	Validação do método em cenário muito limitado;	Uso de fonte de informação de fácil acesso;
KUNTARTO et al (2017)	Páginas Web		x	x	x		Análise estatística (TF-IDF); Análise sintática (POS tagging)			x	Fonte de informação não aplicável para todos os domínios	Não identificado
TABOADA et al (2017)	ontologia; base de dados	x					Named Entity Recognition (NER); lexical similarity			x	Enriquecimento apenas lexical; Fonte de informação específica	Enriquecimento da ontologia por completo
GUERRAM e MELLAL (2018)	Corpus textual		x	x	x		Análise sintática (POS tagging, Parsing, análise morfosintática, NER, word stemming); Medidas de similaridade		x		Falta de detalhamento de algumas das etapas;	Independente de domínio; Uso de fonte de informação de fácil acesso
IDOUDI et al (2018)	ontologia		x				Hierarchical Fuzzy Clustering	x			Enriquecimento de apenas um componente; fonte de informação muito específica; a técnica exige a supervisão de um especialista	Não identificado
KOKLA e TOMAI (2018)	ontologia; Páginas Web	x	x	x	x		análises das definições dos conceitos; análise sintática (tokenização, part-of-speech (POS) tagging, divisão de sentenças e lematização)		x		Fontes de dados específicas; falta de detalhamento das etapas	Enriquecimento de vários componentes

ALI et al (2019)	corpus textual multilíngue; ontologias		x	x	x		POS tagging, sentence splitter, tokenização, stop words,; machine learning			x	As técnicas exigem a supervisão de um especialista (linguista computacional)	Independente de domínio; independente de idioma
KUNTARTO, ISYHRANI e GUNAWAN (2019)	Páginas Web		x				Algoritmo de Clustering K-Means; TF-IDF			x	Enriquecimento de apenas um componente, fonte de informação muito específica, aplicação do método em uma amostra pequena e falta de detalhamento das etapas do método.	As etapas do método estão bem descritas

APÊNDICE C – COLETA DE DADOS REFERENTES À APLICAÇÃO DA ANÁLISE DE CONTEÚDO NOS 29 DOCUMENTOS DA REVISÃO DE LITERATURA

			Continua...
Autores	Ano	Trecho Extraído	Procedimento Generalizado
Velardi, Fabriani e Missikoff	2001	Identificação de conceitos do domínio	Aplicação da técnica de extração da informação
		Detecção de Named Entities (nomes próprios)	Aplicação da técnica de extração da informação
		Extração de terminologia	Aplicação da técnica de extração da informação
		Detecção de relações verticais (taxonômicas)	Aplicação da técnica de extração da informação
		Análise experimental	Validação da ontologia enriquecida
Faatz e Steinmetz	2002	Input ontologia que será enriquecida	Identificação da ontologia alvo
		Input corpus textual	Seleção da fonte de informação
		Uso de algoritmo que mede distancia semântica	Aplicação da técnica de extração da informação
Valakaros <i>et al</i>	2004	Anotação semântica do corpus textual com base na ontologia inicial	Preparação da fonte de informação
		Descoberta de conhecimento usando algoritmo Machine Learning	Aplicação da técnica de extração da informação
		Refinamento do conhecimento usando algoritmo de clustering	Tratamento do conhecimento adquirido
		Validação e inserção de elementos na ontologia com o auxílio de um especialista	Validação do conhecimento adquirido / Enriquecimento da ontologia

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Navigli e Velardi	2006	Cada vocabulário é processado	Seleção da fonte de informação
		Cada vocabulário é processado com part-of-speech tagger	Escolha da técnica de extração da informação
		Para cada gloss uma string POS tags é produzida	Definir as regras para seleção do conhecimento extraído
		Capturar nome de entidades relacionadas a localização, organização, pessoa, numeros e expressões de tempo	Definir as regras para seleção do conhecimento extraído
		Anotação de segmentos de sentenças com propriedades da CIDOC	Aplicação da técnica de extração da informação
		Restrição semântica sobre o domain e range da relação	Enriquecimento da ontologia
Bada e Hunter	2007	Representação das ontologias GO, CHEBI, CTO	Identificação da ontologia alvo
		Recuperação de padrões (expressões comuns) em termos OBO	Seleção da fonte de informação / Aplicação da técnica de extração da informação
		Uso de padrões para criar afirmações	Aplicação da técnica de extração da informação
		Representação primária das afirmações para enriquecimento	Tratamento do conhecimento adquirido / Desenvolvimento de representação intermediária
		Integração das afirmações na ontologia	Enriquecimento da ontologia
		Validação da integração das afirmações inseridas na ontologia	Validação do conhecimento adquirido

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Reiter e Buitelaar	2007	Extração de todos os termos do campo descrição dos termos de todas as classes da ontologia	Seleção da fonte de informação / Aplicação da técnica de extração da informação
		Busca dos termos na base WorNet	Aplicação da técnica de extração da informação
		Aplicação da técnica Word Sense Disambiguation	Aplicação da técnica de extração da informação
		Selecionar o conjunto de termos mais relevantes	Tratamento do conhecimento adquirido
		Adicionar sinônimos para o conjunto de termos selecionado	Tratamento do conhecimento adquirido
Bendadoud, Toussaint e Napoli	2008	Extração da informação/análise de textos	Escolha da técnica de extração da informação
		Classificação de objetos celestiais dos textos usando Formal Concept Analysis (FCA)	Aplicação da técnica de extração da informação
		Manter especialista no ciclo para validar operações	Validação do conhecimento adquirido
		Classificação de objetos celestiais da base de dados Simbad usando FCA	Aplicação da técnica de extração da informação
		Fusão e enriquecimento das hierarquias	Tratamento do conhecimento adquirido/ Enriquecimento da ontologia
Representação dos conceitos em Lógica Descritiva	Representação do conhecimento usando linguagem formal		

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Peters	2009	A plataforma GATE forma a base metodológica do nosso trabalho	Identificar recursos que suportem o enriquecimento
		Pré-processamento dos textos com as técnicas de extração da informação	Preparação da fonte de informação
		Extração de termos	Aplicação da técnica de extração da informação
		Identificação de padrões léxico-sintáticos para descoberta de relações semânticas	Aplicação da técnica de extração da informação
		Aplicação da técnica Headword matching para identificação de relações hiponômicas	Aplicação da técnica de extração da informação
		Recuperação de 10 head matching relações na ontologia	Tratamento do conhecimento adquirido
		Aplicação da técnica Hearst patterns para descoberta de relações hiponímicas	Aplicação da técnica de extração da informação
		Aplicação da técnica para medir Mutual Information	Aplicação da técnica de extração da informação
		Recuperação de 40 Mutual Information relations	Tratamento do conhecimento adquirido
		Tranformação dos elementos linguísticos em modelo semântico	Desenvolvimento de representação intermediária
		Tranformação do modelo semântico em linguagem de representação ontológica	Representação do conhecimento usando linguagem formal
		Decisão sobre o status ontológico do conhecimento semântico (se deveria se encaixar como classe, atributo, relação)	Enriquecimento da ontologia

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Fernandez-Breis et al	2010	Inspeção dos rótulos, definições textuais da Gene Ontology para diagnosticar o que pode ser transformar em axiomas	Diagnóstico da ontologia
		Desenvolver padrões de axiomas que capturem o conhecimento sobre os conceitos, relacionando a necessidade de axiomas a elementos dentro do termo rótulo para o conceito	Aplicação da técnica de extração da informação
		Identificar ontologias suportes ou módulo que capturam entidades fundamentadas nos padrões desenvolvidos	Seleção da fonte de informação
		Aplicar padrões entre ontologias fontes usando OPPL2 (The Ontology Pre-Processing Language's)	Tratamento do conhecimento adquirido
		Executar motor de inferência e inspecionar a ontologia resultante	Enriquecimento da ontologia / Revisão da ontologia enriquecida
Harb e Hajlaoui	2010	Pré-processamento dos textos com as técnicas de extração da informação	Preparação da fonte de informação
		Eliminação de stop words, definição das regras para seleção	Definir as regras para seleção do conhecimento extraído
		Buscar no corpus palavras relacionadas aos conceitos da ontologia	Aplicação da técnica de extração da informação
		Extração de conceitos com o uso de regras de associação	Aplicação da técnica de extração da informação
		Filtragem dos conceitos recuperados	Tratamento do conhecimento adquirido
Validação e enriquecimento da ontologia	Enriquecimento da ontologia / Validação do conhecimento adquirido		

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Monachesi e Markus	2010	Ontologia inicial	Identificação da ontologia alvo
		Extração de dados de aplicações de mídias sociais tais como Delicious, Youtube e Slideshare	Seleção da fonte de informação / Aplicação da técnica de extração da informação
		Aplicação de medidas de similaridade para identificar etiquetas que possuem relação com os conceitos existentes na ontologia	Tratamento do conhecimento adquirido
		Identificação de relações entre conceitos existentes na ontologia e novos conceitos derivados das etiquetas	Tratamento do conhecimento adquirido
Ferré e Rudolph	2012	Existência de uma ontologia como ponto de partida	Identificação da ontologia alvo
		Especialista analisa a estrutura da ontologia e tenta incluir conhecimentos absurdos sobre o domínio, respeitando o que a ontologia permite	Aplicação da técnica de extração da informação
		Especialista constrói descrições que a ontologia não restringe, mas que são absurdas de acordo com o conhecimento aceito no domínio	Aplicação da técnica de extração da informação
		Uma vez criada as situações problemáticas, desenvolve-se axiomas, restrições que impossibilitam a criação desse tipo de descrição	Enriquecimento da ontologia
		Exclui-se todas as descrições que não estão de acordo com os axiomas	Enriquecimento da ontologia

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Kaihong	2012	Textos de relatórios clínicos	Seleção da fonte de informação
		Pré-processamento usando técnicas PLN	Preparação da fonte de informação
		Output: documento etiquetado com POS tagging	Organização dos documentos de entrada e saída de cada etapa
		Output Documentos anotados com entidades por meio do uso da técnica NER	Organização dos documentos de entrada e saída de cada etapa
		Combinação de padrão léxico sintático	Aplicação da técnica de extração da informação
		Cálculo de similaridade	Aplicação da técnica de extração da informação
		Output: conjunto de dados de LSP contendo sentenças	Organização dos documentos de entrada e saída de cada etapa
		Output: lista de novas entidades candidatas	Organização dos documentos de entrada e saída de cada etapa
		Análise e avaliação	Tratamento do conhecimento adquirido
		Avaliação por especialista do domínio	Validação do conhecimento adquirido
		Avaliação por ontologistas que julgam a significância ontológica	Validação da ontologia enriquecida

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Ben Amar, Gargouri e Hamadou	2013	Anotação conceitual de campos textuais	Seleção da fonte de informação
		A anotação pode ser realizada com plataformas linguísticas	Identificar recursos que suportem o enriquecimento
		Identificar candidatas a triplas (Relação, Conceito 1, Conceito 2)	Aplicação da técnica de extração da informação
		Checar triplas duplicadas	Tratamento do conhecimento adquirido
		Validação das triplas extraídas	Validação do conhecimento adquirido
Groza, Barbur e Blaga	2013	Extração automática de conteúdo da Web usando a tecnologia RSS Feed	Seleção da fonte de informação
		Extração automática de conteúdo da Web usando a tecnologia RSS Feed	Aplicação da técnica de extração da informação
		Extração de termos do documento RSS	Aplicação da técnica de extração da informação
		Aplicação de métodos relacionados à base léxica WordNet	Tratamento do conhecimento adquirido
		Cada palavra exibida na árvore de hipônimo do WordNet pode ser adicionada à ontologia	Enriquecimento da ontologia
Luksová	2013	Input documentos textuais do domínio descrição de empregos	Seleção da fonte de informação
		Input ontologia que será enriquecida	Identificação da ontologia alvo
		Escolha de quais classes da ontologia serão analisados	Definição do escopo do enriquecimento
		Extração de realizações léxicas que representam os conceitos	Aplicação da técnica de extração da informação
		Extração de relações (taxonômicas e não taxonômicas)	Aplicação da técnica de extração da informação
		Uso de machine learning para extração dos elementos	Aplicação da técnica de extração da informação
		Avaliação da performance do algoritmo de ML	Tratamento do conhecimento adquirido
		Pré-anotação da ontologia	Enriquecimento da ontologia

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Astrakhantsev, Fedorenko e Turdakov	2014	Utiliza textos específicos do domínio e uma ontologia como input	Seleção da fonte de informação
		Pré-processamento dos textos com as técnicas de extração da informação	Preparação da fonte de informação / Aplicação da técnica de extração da informação
		Uso da ferramenta OpenNLP	Identificar recursos que suportem o enriquecimento
		Pega todos os textos pré-processados e retorna um conjunto de termos específicos de domínio que não estão contidos na ontologia de entrada	Tratamento do conhecimento adquirido
		Como termos extraem-se todas unigrama, bigramas e trigramas que ocorrem pelo menos 2 vezes e satisfaçam o padrão determinado	Definir as regras para seleção do conhecimento extraído
		Formação de conceitos	Aplicação da técnica de extração da informação
Booshehri e Luksch	2015	O sistema fornece uma ontologia primária	Identificação da ontologia alvo
		O sistema fornece um banco de dados de fragmentos de texto indicando o contexto de cada termo da ontologia	Seleção da fonte de informação / Preparação da fonte de informação
		O banco de dados de fragmentos é passado para o Módulo Extrator de Contexto para serem convertido para um banco de dados de vetores dos termos	Aplicação da técnica de extração da informação
		Os vetores são passados para o Módulo de Cálculo de Similaridade para identificar a relação entre termos	Aplicação da técnica de extração da informação

		da ontologia e seus recursos correspondentes em banco de dados em Linked Open Data (LOD)	
		O Módulo de Descoberta de Homônimos prepara um banco de dados de triplas relacionadas a alguns recursos que possuem o mesmo nome dos termos da ontologia	Tratamento do conhecimento adquirido
		O banco de dados com as triplas é processada pelo Módulo Extrator de Relações, que encontra relações adequadas para os termos da ontologia	Tratamento do conhecimento adquirido / Seleção do conhecimento candidato para enriquecer a ontologia
		Gera-se um banco de dados de relações que é usado no Módulo de Enriquecimento da Ontologia e validado pelo ontologista	Validação do conhecimento adquirido / Enriquecimento da ontologia
Gillani e Ko	2015	Input: arquivos de texto e XML	Seleção da fonte de informação
		Fase preliminar: Pré-processamento dos dados	Preparação da fonte de informação
		Aplicação das técnicas tokenização, Filtragem de Stop Words, POS tagging, Key Term Extraction	Aplicação da técnica de extração da informação
		Output da fase 1: arquivo com Key Terms	Organização dos documentos de entrada e saída de cada etapa
		Enriquecimento de conceitos (input Key Terms)	Aplicação da técnica de extração da informação
		Ligação dos Termos Chave (Key Terms) com recursos externos (WordNet e Wiktionary)	Identificar recursos que suportem o enriquecimento
		Extração de palavras similares	Aplicação da técnica de extração da informação

		Expansão de vocabulário sobre determinado conceito para extração de elementos de conhecimento úteis	Aplicação da técnica de extração da informação
		Filtragem de palavras candidatas	Tratamento do conhecimento adquirido
		Output da fase 2: Conceitos chave	Organização dos documentos de entrada e saída de cada etapa
		Ranqueamento de conceitos e seleção	Seleção do conhecimento candidato
		Categorização semântica do conceito	Seleção do conhecimento candidato
		Output: Enriquecimento da ontologia	Enriquecimento da ontologia
Quesada-Martínez	2015	O usuário especifica uma ontologia fonte	Identificação da ontologia alvo
		Análise léxica da ontologia fonte	Aplicação da técnica de extração da informação
		Output: arquivo XML com informações da análise léxica	Organização dos documentos de entrada e saída de cada etapa
		Input: arquivo da fase anterior é usado para cálculo de métricas	Organização dos documentos de entrada e saída de cada etapa
		Redução léxica	Aplicação da técnica de extração da informação
		Output: arquivo com redução léxica	Organização dos documentos de entrada e saída de cada etapa
		Validação através da comparação a um modelo padrão	Validação do conhecimento adquirido

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Ejei et al	2017	Ontologia implementada	Identificação da ontologia alvo
		Refinamento das relações hierárquicas para diferenciar relações genéricas de partitivas	Seleção da fonte de informação / Preparação da fonte de informação
		Alinhamento dos conceitos da ontologia de domínio com ontologia de alto-nível	Aplicação da técnica de extração da informação
		Escolha de uma ontologia de alto nível para alinhar os conceitos da ontologia e facilitar a identificação do domain e range da relação	Identificar recursos que suportem o enriquecimento
		Converter relacionamentos em relações semânticas	Tratamento do conhecimento adquirido
		Definição de regras para refinamento das relações	Definir as regras para seleção do conhecimento extraído
		Especialista do domínio seleciona os padrões de relações apropriados	Validação do conhecimento adquirido
Gómez-Moreno e Mestre-Mestre	2017	Identificação da ontologia que será enriquecida	Identificação da ontologia alvo
		Escolha do conceito "dengue virus"	Definição do escopo do enriquecimento
		Seleção de lista de atributos da base de dados ExpASy	Seleção da fonte de informação
		Compilação do corpus	Preparação da fonte de informação
		Aplicação de filtros no corpus para exclusão de palavras sem importância	Preparação da fonte de informação / Aplicação da técnica de extração da informação
		Classificação dos dados usando a medida NPMI	Tratamento do conhecimento adquirido
		Atribuição das palavras a um conceito relacionado na ontologia	Enriquecimento da ontologia

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Kuntarto	2017	Documentos Web (dados em vários formatos)	Seleção da fonte de informação
		Preparar uma coleção de documentos Web	Preparação da fonte de informação
		Identificação de conceitos	Aplicação da técnica de extração da informação
		Construção da taxonomia	Aplicação da técnica de extração da informação
		Extração de relações semânticas	Aplicação da técnica de extração da informação
		Avaliação da ontologia	Validação da ontologia enriquecida
Taboada et al	2017	Sinônimos da Human Phenotype Ontology (HPO)	Seleção da fonte de informação
		Descarte de sinônimos redundantes	Preparação da fonte de informação
		Identificação de sobreposições léxicas	Aplicação da técnica de extração da informação
		Geração de sinônimos candidatos com base nos relacionamentos hierárquicos da HPO	Aplicação da técnica de extração da informação
		Descarte de candidatos irrelevantes com base na MEDLINE	Tratamento do conhecimento adquirido
		Inferir escopo ou tipo dos sinônimos	Tratamento do conhecimento adquirido
		Geração de novos sinônimos para a ontologia HPO	Enriquecimento da ontologia
Guerram e Mellal	2018	Conhecimento básico do domínio representado em uma ontologia inicial	Identificação da ontologia alvo
		Anotação dos textos usando análise morfossintática	Preparação da fonte de informação
		Simplificação de sentenças complexas	Aplicação da técnica de extração da informação
		Comparação de cada relacionamento extraído com o conteúdo da ontologia inicial usando medida de similaridade	Validação do conhecimento adquirido

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Idoudi et al	2018	Definir qual será a ontologia fonte (de onde os dados serão extraídos)	Seleção da fonte de informação
		Definir a ontologia alvo, ou seja, ontologia que será enriquecida	Identificação da ontologia alvo
		Hierarchical fuzzy clustering	Escolha da técnica de extração da informação
		Criar estruturas hierárquicas de cluster fuzzy da ontologia fonte	Aplicação da técnica de extração da informação
		Criar estruturas hierárquicas de cluster fuzzy da ontologia alvo	Aplicação da técnica de extração da informação
		Alinhamento entre os cluster extraídos das duas ontologias	Tratamento do conhecimento adquirido
		Enriquecimento da ontologia alvo	Enriquecimento da ontologia
Kokla, Papadias e Tomai	2018	Processo de enriquecimento foi empreendido com base em outros recursos	Seleção da fonte de informação
		Extração de hiperônimos com base em análise das definições dos conceitos da ontologia e conceitos encontrados na hierarquia de base léxica (WordNet)	Aplicação da técnica de extração da informação
		Extração de hipônimos com base no recurso Geonames	Aplicação da técnica de extração da informação
		Extração de relação meronômica com base em recursos léxicos (WordNet, Babel Net)	Aplicação da técnica de extração da informação
		Extração de outras relações semânticas que não sejam hiperônimo, hipônimo ou meronômica	Aplicação da técnica de extração da informação
		Extração de informação semântica de textos para identificação de conceitos adicionais	Aplicação da técnica de extração da informação
		A tarefa é suportada por uma ferramenta implementada em R que	Identificar recursos que suportem o enriquecimento

		<p>guia o especialista na identificação de conceitos adicionais</p> <p>O processo é baseado na análise semântica do corpus textual para a identificação de frases chave que podem se consituir como conceito candidato para enriquecer a ontologia</p> <p>Frases chave para a tarefa específica são noun phrases que consistem na combinação de adjetivos, substantivos, preposições e determiners.</p> <p>Noun phrases deve conter um ou mais termos que referem a um conceito da ontologia e que seja encontrado no corpo textual pelo menos duas vezes</p> <p>Entre as noun phrases identificadas, as mais proeminentes para se consituir como candidatas a conceitos nas ontologias são XYZ</p>	<p>Escolha da técnica de extração da informação</p> <p>Definir as regras para seleção do conhecimento extraído</p> <p>Definir as regras para seleção do conhecimento extraído</p> <p>Tratamento do conhecimento adquirido</p>
Ali et al	2019	<p>A fonte de entrada (input) são textos em diferentes idiomas</p> <p>A ontologia a ser enriquecida é a entrada (input)</p> <p>Pré-processamento dos textos</p> <p>Extração de conceitos usando técnicas de PLN</p> <p>Identificação do domínio do texto de entrada</p> <p>Seleção de sentença candidata</p> <p>Aplicação de algoritmo para enriquecimento da ontologia em vários idiomas</p> <p>Verificação e validação</p>	<p>Seleção da fonte de informação</p> <p>Identificação da ontologia alvo</p> <p>Preparação da fonte de informação</p> <p>Aplicação da técnica de extração da informação</p> <p>Tratamento do conhecimento adquirido</p> <p>Seleção do conhecimento candidato para enriquecer a ontologia</p> <p>Enriquecimento da ontologia</p> <p>Validação da ontologia enriquecida</p>

Continua...

Autores	Ano	Trecho Extraído	Procedimento Generalizado
Kuntarto, Isyahrani e Gunawan	2019	Coleta dos dados no site TripAdvisor	Seleção da fonte de informação
		Classificação dos termos extraídos usando TF-IDF	Aplicação da técnica de extração da informação
		Classificação usando algoritmo K-Means	Aplicação da técnica de extração da informação
		Verificar correspondência entre conceitos extraídos e conceitos existentes na ontologia	Avaliação da necessidade de inserção de determinado componente
		Identificação de conceito candidatos para enriquecer a ontologia	Seleção do conhecimento candidato para enriquecer a ontologia
		Construção da taxonomia	Enriquecimento da ontologia
		Extração de relação semântica	Aplicação da técnica de extração da informação

APÊNDICE D – COLETA DE DADOS REFERENTES À APLICAÇÃO DA ANÁLISE DE CONTEÚDO NOS DOCUMENTOS SOBRE MANUTENÇÃO E ATUALIZAÇÃO DE SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

Documento	Trecho extraído	Procedimento equivalente para o contexto de enriquecimento
SOERGEL (1974)	Introdução de um novo descritor	Enriquecimento da ontologia
	Eliminação de um descritor	Alteração no conteúdo da ontologia
	Subdivisão de um descritor existente em um número de descritores específicos	Alteração no conteúdo da ontologia
	Mudança na definição	Alteração no conteúdo da ontologia
	Adição ou eliminação de um relacionamento hierárquico	Alteração no conteúdo da ontologia
	Adição ou eliminação de um relacionamento associativo	Alteração no conteúdo da ontologia
	Fontes para novos termos, conceitos e relacionamentos	Seleção da fonte de informação
	Fontes de informação sobre novos desenvolvimentos no campo científico	Seleção da fonte de informação
	Uso de formulários de tesouros na atualização	Identificar recursos que suportem o enriquecimento
Para a incorporação de informação no tesouro, os procedimentos descritos para construção de tesouros se aplicam com algumas modificações apropriadas	Considerar o uso de procedimentos do desenvolvimento de ontologia para o processo de enriquecimento	

	É especialmente importante manter em mente que o tesauro é uma estrutura altamente interconecta, então alterações em um local leva a alterações em outro local.	Verificar o impacto da alteração realizada no conteúdo da ontologia
	Todo elemento de dado é acompanhado pela data em que entrou no arquivo em desenvolvimento	Controlar a data em que o componente foi inserido
	[Atualização] deveria ser feita em intervalos regulares, o tamanho dos intervalos depende de algumas circunstâncias	Planejar a frequência do enriquecimento
	Listagem de mudanças realizadas	Controlar as alterações realizadas no conteúdo da ontologia
	A data em que um descritor foi introduzido deve ser descrita na nota de escopo	Controlar a data em que o componente foi inserido
	Organização para tomada de decisão na atualização do tesauro	Validação do conhecimento adquirido
	Problemas de reindexação devido aos novos descritores	Verificar o impacto da alteração realizada no conteúdo da ontologia
ANSI/NISO Z39.19-2005 (2010)	Os vocabulários controlados são reflexos da linguagem e, portanto, instrumentos dinâmicos	Avaliação da necessidade de enriquecimento
	Políticas e procedimentos devem ser estabelecidos para revisão periódica da terminologia, estabelecimento de novos termos e substituição de termos obsoletos,	Planejar a frequência do enriquecimento

especialmente em campos onde a terminologia muda rapidamente

Editores de vocabulário controlado devem atualizar o vocabulário controlado em intervalos que serão determinados pela frequência e volume das mudanças feitas, e pelo método de distribuição

Planejar a frequência do enriquecimento

Sempre que um termo apropriado ou combinação de termos não podem ser encontrados em um vocabulário controlado, o indexador ou pesquisador deve nomear um novo termo como um termo candidato

Avaliação da necessidade de enriquecimento

Os novos termos e aqueles indicados para reconsideração devem ser revisados pelo editor do vocabulário controlado e, preferencialmente, por um conselho editorial que funcione da mesma forma que o grupo responsável pela edição original.

Validação do conhecimento adquirido

Este termo deve ser adicionado ao vocabulário controlado ou já está coberto por um termo existente ou combinação de termos? Se for coberto, uma referência de USE adicional deve ser adicionada?

Diagnóstico da ontologia

Como o novo termo deve ser inter-relacionado com os termos existentes em uma base

Verificar o impacto da alteração realizada no conteúdo da ontologia

	hierárquica e de termos relacionados?	
	Os indexadores e as buscas devem ser capazes de propor modificações aos termos existentes ou seus relacionamentos, explicando a razão e fornecendo documentação de apoio para as mudanças propostas.	Possibilitar que especialistas proponham alterações na ontologia
	Se um termo for modificado, a data da mudança deve ser registrada na nota de histórico	Controlar a data em que o componente foi inserido
ISO 25964-1 (2011)	Antes que tempo e dinheiro significativos sejam gastos, os objetivos devem ser esclarecidos	Definir o escopo do enriquecimento
	Determinando as características do tesouro	Identificação da ontologia alvo
	Qual o formato previsto para atualizações e com que frequência elas serão necessárias.	Planejar a frequência do enriquecimento
	Durante as fases de planejamento de um projeto de tesouro, os principais recursos, como pessoas, financiamento, ferramentas de software e recursos de vocabulário devem ser determinados.	Definir os recursos necessários (humano, ferramentas, financiamento)
	Tesauros ou esquemas de classificação existentes com um escopo que atenda parcial ou totalmente ao requisito. Em casos extremos, pode ser	Seleção da fonte de informação

econômico simplesmente adotar um dicionário de sinônimos existente sem alterações. Mais comumente, partes de um vocabulário existente podem ser obtidas e / ou adaptadas, sujeitas a direitos autorais. Os trabalhos existentes são uma fonte de ideias para termos, ou estrutura, ou ambos.

Obras de referência padrão como dicionários, terminologias, nomenclaturas e enciclopédias são úteis para verificação do significado de um termo e validação de estruturas hierárquicas, ao invés de um guia para a seleção de termos.

Identificar recursos que suportem o enriquecimento

Estabelecendo responsabilidades

Definir responsabilidades

Embora seja possível compilar e produzir um tesouro totalmente sem suporte de software, esse procedimento não é recomendado. O software de gerenciamento de dicionário de sinônimos desenvolvido para esse fim deve ser usado para fins de eficiência e para evitar erros de escrita.

Identificar recursos que suportem o enriquecimento

Durante a construção do tesouro, o objetivo geral é coletar e organizar conceitos, mas, na prática, isso é conseguido coletando termos e organizando-os de forma que representem adequadamente os conceitos necessários.

Seleção da fonte de informação

O texto completo da literatura central para o escopo exigido também é útil

Conforme descrito em 13.2.1, técnicas automáticas podem ser usadas para extrair palavras-chave e frases da literatura

Os termos coletados devem ser classificados em ordem sistemática antes que possam ser estudados adequadamente e inseridos no tesouro.

Se o software de gerenciamento de tesouro não oferecer suporte a essa tarefa, ferramentas de desktop padrão, como planilhas, podem ser úteis para essa classificação preliminar.

Uma vez que os termos coletados estão em uma planilha, junto com sua fonte e frequência, eles devem ser classificados em áreas temáticas ou facetas (ver Cláusula 11) para que termos semelhantes com todas as suas variantes e sinônimos sejam reunidos

O principal produto dessa etapa analítica deve ser uma lista ou conjunto de pequenas listas de termos, agrupadas por assunto e / ou faceta, com indicação da fonte e frequência de uso de cada termo.

Os termos devem ser tratados em grupos de assunto / faceta ou hierarquias. Grupos de sinônimos e quase-sinônimos devem ser considerados juntos e as preferências e relações de equivalência estabelecidas

Tratamento do conhecimento adquirido

Nesta fase, os termos e suas relações hierárquicas devem ser selecionados cuidadosamente, levando em consideração as potenciais duplicações, sobreposições, omissões e o grau de especificidade exigido

Tratamento do conhecimento adquirido

Se desejado, a fonte e a definição também podem ser inseridas

Tratamento do conhecimento adquirido

É melhor que a data de entrada e o identificador do termo sejam adicionados automaticamente pelo software, em vez de digitação manual.

Controlar a data em que o componente foi inserido

Se o exercício inicial de avaliação tiver revelado um tesouro existente, do qual uma parte substancial deve ser adotada, pode ser possível importá-lo diretamente para o software de gerenciamento de tesouros escolhido.

Seleção da fonte de informação

É eficiente trabalhar com grupos de termos agrupados hierarquicamente, inserindo-os junto com equivalência e relações hierárquicas.

Enriquecimento da ontologia

O controle de qualidade deve ser realizado para cobrir aspectos como integridade, ausência de duplicação, precisão de grafia, escolha correta de relações, etc.	Validação do conhecimento adquirido
Envolvendo especialistas e partes interessadas	Validação do conhecimento adquirido
A manutenção do dicionário de sinônimos deve estar ativa durante toda a vida útil do produto.	Planejar a frequência do enriquecimento
Embora a mudança seja essencial para lidar com os requisitos emergentes, é importante reconhecer que cada mudança potencialmente prejudica o desempenho de recuperação para o "backfile"	Verificar o impacto da alteração realizada no conteúdo da ontologia
Um mecanismo fácil de sugerir alterações deve ser fornecido a todos os usuários do dicionário de sinônimos, em particular aos indexadores. O mecanismo fornecido pode ser um formulário, impresso ou eletrônico	Identificar recursos que suportem o enriquecimento
O formulário deve convidar as pessoas a sugerir um novo termo ou outra alteração necessária e fornecer o máximo possível de informações básicas. As informações de base úteis incluem uma definição do conceito necessário, quaisquer referências à sua ocorrência na literatura, quaisquer sinônimos	Identificar recursos que suportem o enriquecimento

ou formas alternativas de expressão, quaisquer relações com termos já presentes no tesouro e uma justificativa de por que a mudança é necessária.

Um estudo cuidadoso pode mostrar que o conceito sugerido já estava presente no tesouro com outro nome.

Diagnóstico da ontologia

Embora o editor deva criar uma nota de histórico apropriada quando os termos são alterados significativamente, a data de entrada e a data da última modificação devem ser preenchidas automaticamente pelo software. Da mesma forma, o identificador para novos termos deve ser adicionado automaticamente

Controlar a data em que o componente foi inserido

Disseminação das atualizações

Disseminar a versão da ontologia enriquecida

A forma mais simples de divulgar as mudanças é distribuir uma nova edição do tesouro

Disseminar a versão da ontologia enriquecida

APÊNDICE E – RESULTADO DA ANÁLISE DOS DOCUMENTOS

Título	Informação extraída
<p>Anthropic Impacts On The Environmental Factors And Its Reduction Measures Within The Oltet Mining Perimeter</p>	<p>O impacto ambiental produzido pela exploração de carvão é significativo, inevitável e irreversível, afetando a sistemas aquáticos, ar, recursos minerais naturais, ecossistemas, clima, geomorfologia e paisagem, uso da terra e comunidade humana.</p> <p>Exploração de carvão impacta na qualidade do ar devido à emissão de gases poluentes</p>
<p>Trace Element Inputs Into Soils By Anthropogenic Activities And Implications For Human Health</p>	<p>Solos de práticas agrícolas que usa produtos químicos can be dispersed over large areas from the point of release, and are difficult to control and limit. (ou seja, pode afetar as águas subterrâneas).</p> <p>Degradação de solo pode ser causada por práticas agrícolas</p> <p>Degradação do solo pode ser causado por produtos químicos que por sua vez podem contaminar águas subterrâneas</p> <p>Although some trace elements are essential plant micronutrients, excessive concentrations of them in agricultural soils are a general matter of concern to society as they may adversely affect the quality of agricultural products and enter the food chain, or leach down to groundwater and contaminate drinking water resources, and may cause, in both cases, hazards to the health of humans and animals.</p> <p>Agrochemicals, especially pesticides, including herbicides, fungicides, insecticides, etc., are widely used in intensive agriculture for the control of weeds, insects and diseases in various crops.</p> <p>The most important anthropogenic sources of trace elements for soils include: commercial fertilizers, liming materials and agrochemicals, sewage sludges and other wastes used as soil amendments, irrigation waters, and atmospheric deposition from industrial, urban and road emissions [5].</p> <p>Águas residuais podem ser usadas para irrigação. Essas águas podem ter concentrações de elementos que impactam no solo</p>

Título	Informação extraída
<p>The impact of climate change and anthropogenic activities on alpine grassland over the Qinghai-Tibet Plateau</p>	<p>livestock and implementation of many large scale ecological projects, such as fencing degrading grassland, ecological compensation and restoring the land cover vegetation.</p> <p>The major anthropogenic activities over the QTP that can influence alpine grassland NPP are grazing and human-induced land cover change; the former is the dominant one and can lead to grassland degradation, even desertification in the plateau. However, it is not clear whether the number of livestock directly affects NPP or whether there are other influencing factors, so we recorded the number of total livestock over grassland restoration</p>
<p>Potencial de impacto da agricultura sobre os recursos hídricos na Região do Cerrado</p>	<p>Conversão de áreas naturais em agroecossistemas</p> <p>Eutrofização causada pelo enriquecimento dos corpos hídricos com fósforo e nitrogênio</p> <p>Deposição de sedimentos carregados diminuindo a capacidade dos reservatórios (assoreamento)</p> <p>Erosão pode ajudar a levar substâncias químicas para águas superficiais</p>
<p>Um Novo Olhar Na Geografia Para Os Conceitos E Aplicações De Geossistemas, Sistemas Antrópicos E Sistemas Ambientais</p>	<p>Um Sistema Ambiental pode ser caracterizado como entidade organizada na superfície terrestre formada pelos subsistemas físico/natural (Geossistema) e antrópico, bem como por suas interações. O subsistema físico-natural (Geossistema) é composto por elementos e processos relacionados ao clima, solo, relevo, águas e seres vivos, enquanto os componentes e processos do subsistema Antrópico são aqueles ligados a população, urbanização, industrialização, agricultura e mineração, entre outras atividades e manifestações humanas.</p> <p>Ambos mantêm um funcionamento parcialmente independente, e também um funcionamento dependente um do outro, ou seja, mesmo a natureza apresentando suas leis e dinâmica própria, esta pode sofrer alterações em decorrência da ação antrópica, como, por exemplo, alterações nos níveis pluviométricos e alterações na temperatura, como consequência do desmatamento. Do mesmo modo os Sistemas Antrópicos sofrem interferência das leis da natureza, como, por exemplo, a restrição na produção de energia hidráulica no Brasil, no entre os anos de 2001 e 2002,</p>

Título	Informação extraída
	em decorrência das estiagens que diminuíram o nível de armazenamento de água nas barragens.
Mudanças ambientais da terra	<p>Na realidade o efeito-estufa é um fenômeno natural existente na superfície terrestre.</p> <p>o ser humano passou a exalar gases estufa (principalmente dióxido de carbono) em grande quantidade, acompanhando a crescente industrialização e passou a intensificar o fenômeno natural</p> <p>O efeito-estufa é um fenômeno natural mas pela influência de gases-estufa exalados por atividades antrópicas está ocorrendo recrudescimento artificial do processo</p> <p>Acompanhando o aumento de temperatura deverá ocorrer incremento de pluviosidade provocando numerosas tempestades.</p> <p>Regionalmente poderá ocorrer mudança na umidade de solo e com isso áreas hoje agricultáveis podem ser transformadas em semi-desertos ou verdadeiros desertos.</p> <p>circulação da atmosfera / condição atmosférica</p> <p>vento/tornado....</p> <p>desastre natural</p> <p>intemperismo, erosão e sedimentação</p> <p>Os cultivos agrícolas, as terraplanagens para construção e expansão urbanas, as atividades de escavação para mineração e as obras de construção de usinas elétricas ou de rodovias e portos</p> <p>Em futuro próximo, como consequência do desflorestamento, deverá ocorrer mudança na precipitação pluvial das regiões sul e sudeste do Brasil, pois haverá diminuição na umidade suprida pelo Oceano Atlântico à Floresta Pluvial Tropical da Amazônia.</p> <p>O incessante incremento da destruição dos ambientes naturais, que não se limita à Amazônia, suscita a possibilidade de deterioração do clima em todo o planeta Terra</p> <p>Chuva ácida – Designação empregada para chuva com pH inferior a 5,6, devida à contaminação por ácidos sulfúricos (H₂ SO₄) e nítrico (HNO₃), produzidos pela queima de combustíveis fósseis (carvão e petróleo). A chuva ácida tem danificado coberturas vegetais, lagos e solos pela</p>

Título	Informação extraída
	<p>acidificação, além de causar estragos em objetos de arte pela dissolução do mármore e oxidação do bronze.</p> <p>Chuva polínica – Quantidade total de pólen e esporos que caem sobre uma determinada área em um certo intervalo de tempo. Esta estimativa é executada pelo estudo de amostras de sedimentos e solos ou usando-se armadilhas para retenção de palinomorfos</p> <p>Desertificação – Processo de deterioração ambiental resultante do rompimento da estabilidade de uma paisagem, dando origem a eventos catastróficos de erosão de vertentes, denudação de solos cultiváveis, formação de campos de dunas, etc., que são fenômenos mais comuns em um deserto. A desertificação é, hoje em dia, freqüentemente induzida pelo homem (ação antrópica), mesmo em regiões de clima não-desértico, através do uso inadequado do espaço físico, como parece ter acontecido na Antiga Grécia.</p> <p>Desflorestamento – Ação de desflorestar através da derrubada sistemática de árvores, em grande escala, para se desfazer de florestas transformando-as em áreas de pastagem ou de cultivo.</p> <p>Efeito-estufa – Fenômeno de aquecimento da superfície da Terra pela radiação terrestre de grande comprimento de onda, que é absorvida e re-emitida. O gás carbônico (CO₂) e outros gases-estufa (CH₄, CFCs, etc.) retêm essas radiações, que causam o aquecimento global.</p> <p>Poluição de água – Estado da água em que suas propriedades físico-químicas e químicas acham-se deterioradas a ponto de prejudicar o seu uso. Entre os poluentes mais comuns tem-se produtos químicos, radioativos e microrganismos, além de temperatura, introduzidos comumente por ação antrópica direta ou indireta.</p> <p>Savanização – Refere-se à transformação de uma área em termos climáticos (estações bem demarcadas em termos de pluviosidade, mas pouco distintas em amplitudes térmicas anuais) com fisionomia vegetacional característica de uma savana (ou cerrado).</p> <p>Sedimentação – Deposição de partículas inorgânicas (minerais) e orgânicas (de origem vegetal) em meio subaquoso ou subaéreo sob condições físico-químicas próximas às da superfície terrestre.</p>
Influence of anthropogenic activities on water quality of	Furthermore, the need to produce more food for a growing population has a link to the quality aspects of the aquatic environment as it results in increased soil erosion, chemical pollution by fertilizers and pesticides, and pollution from animal operations.

Título	Informação extraída
a tropical stream ecosystem	The need to produce more food for a growing population has a link to the quality aspects of the aquatic environment as it results in increased soil erosion, chemical pollution by fertilizers and pesticides, and pollution from animal operations. evaporação

APÊNDICE F – GLOSSÁRIO COM OS CONCEITOS PARA ENRIQUECIMENTO DA ONTOAGROHIDRO

Afforestation: O estabelecimento da floresta por meio de plantio e / ou semeadura deliberada em terras que, até então, estavam sob outro uso do solo, implica na transformação do uso do solo de não florestal em floresta.

Anthropization (antropização): um processo durante o qual um sistema ambiental natural é alterado pela ação humana.

Circulação atmosférica: fenômeno natural. influencia no clima. é influenciada por formações atmosféricas. Entende-se por Circulação Atmosférica a movimentação das massas de ar. Essa movimentação ocorre na Troposfera, a camada da atmosfera mais próxima da Terra.

Clima (climate): Fenômeno natural. Pode ser influenciado por circulação atmosférica, mudança climática. A disposição de um sistema ambiental, particularmente aqueles que influenciam as atmosferas, para manifestar um padrão cíclico de clima ao longo do tempo.

Climate change: As alterações climáticas são mudanças do clima que ocorrem a nível global. Relação com outras classes: **influencia** clima; **influenciado** por AnthropicTransformation.

Cycling: Ciclos biogeoquímicos são processos que reciclam constantemente elementos químicos na natureza.

Deforestation: A desflorestação é a remoção de uma floresta ou grupo de árvores onde a terra é posteriormente convertida para um uso não-florestal. **causa degradação ambiental.**

Erosion: Erosão é o processo de desgaste, transporte e sedimentação do solo, dos subsolos e das rochas como efeito da ação dos agentes erosivos, tais como a água, os ventos e os seres vivos. Relação com outras classes: **é influenciado por** desertification; **é tipo de** degradação do solo.

Eutrophication (eutroficação ou eutrofização): A eutrofização ou eutroficação é o crescimento excessivo de plantas aquáticas, para níveis que afetem a utilização normal e desejável da água. O fator substancial para este aumento é a maior

concentração de nutrientes, essencialmente o nitrogênio e fósforo. A eutrofização (ou eutroficação) é um processo normalmente de origem antrópica (provocado pelo homem), ou raramente de ordem natural, tendo como princípio básico a gradativa concentração de matéria orgânica acumulada nos ambientes aquáticos. Relações com outras classes: é tipo de water pollution

Evaporação (hydrological evaporation): fenômeno físico. A evaporação hidrológica é a evaporação da água, geralmente como parte de um ciclo planetário da água.

Greenhouse effect: O efeito estufa ou efeito de estufa é um processo físico que ocorre quando uma parte da radiação infravermelha é emitida pela superfície terrestre e absorvida por determinados gases presentes na atmosfera, os chamados gases do efeito estufa ou gases estufa.

Hydrological cycle: O ciclo da água é o permanente processo de transformação da água na natureza, passando de um estado para outro (líquido, sólido ou gasoso). A essa transformação e circulação da água dá-se o nome de ciclo da água ou ciclo hidrológico, que se desenvolve através dos processos de evaporação, condensação, precipitação, infiltração e transpiração.

Precipitação: é componente de ciclo hidrológico.

Resource depletion: Relação com outras classes é **causado por** desertificação, degradação do meio ambiente e degradação do solo.

Water pollution: um processo de poluição durante o qual os contaminantes são lançados na água. Relação com outras classes: **afeta** qualidade da água.

ANEXO A – PITFALL 8 DE POVEDA-VILLALÓN (2016)

Title	P08. Missing annotations	Importance level	Minor
Aspects	Ontology understanding Ontology clarity	Affects to	Classes Object properties Datatype properties
Description			
This pitfall consists in creating an ontology element and failing to provide human readable annotations attached to it. Consequently, ontology elements lack annotation properties that label them (e.g. <code>rdfs:label</code> , <code>lemon:LexicalEntry</code> , <code>skos:prefLabel</code> or <code>skos:altLabel</code>) or that define them (e.g. <code>rdfs:comment</code> or <code>dc:description</code>). This pitfall is related to the guidelines provided in [3].			
Examples			
Graphical representation and/or OWL code		Natural language description	
<div style="border: 1px solid black; display: inline-block; padding: 2px;">Person</div> No explicit evidence of: <code>AnnotationAssertion(rdfs:label :Person "name")</code> <code>AnnotationAssertion(rdfs:comment :Person "description")</code>		The graphical example shows the class “Person” without any annotation that labels or describes it.	
How to solve it			
<p>(a) Include label annotation properties (e.g. <code>rdfs:label</code>, <code>lemon:LexicalEntry</code>, <code>skos:prefLabel</code> or <code>skos:altLabel</code>) to provide terms that identify ontology elements.</p> <p>(b) Include description annotation properties (e.g. <code>rdfs:comment</code> or <code>dc:description</code>) to provide natural language definitions of ontology elements.</p> <p>See [1] and [2] for more detailed guidelines.</p>			
References	<p>[1] Aguado-De Cea, G., Montiel-Ponsoda, E., Poveda-Villalón, M., and Giraldo-Pasmin, O. X. (2015). Lexicalizing Ontologies: The issues behind the labels. In <i>Multimodal communication in the 21st century: Professional and academic challenges</i>. 33rd Conference of the Spanish Association of Applied Linguistics (AESLA), XXXIII AESLA.</p> <p>[2] Montiel-Ponsoda, E., Vila Suero, D., Villazón-Terrazas, B., Dunsire, G., Escolano Rodríguez, E., Gómez-Pérez, A. (2011). <i>Style guidelines for naming and labeling ontologies in the multilingual web</i>.</p> <p>[3] Vrandečić, D. (2010). <i>Ontology Evaluation</i>. PhD thesis.</p>		

Table 4.8: P08. Missing annotations

ANEXO B – PITFALL 9 DE POVEDA-VILLALÓN (2016)

Title	P09. Missing domain information	Importance level	Minor
Aspects	No inference Real world modelling Requirement completeness	Affects to	Ontology
Description			
Part of the information needed for modelling the intended domain is not included in the ontology. This pitfall may be related to (a) the requirements included in the Ontology Requirement Specification Document (ORSD) that are not covered by the ontology, or (b) to the lack of knowledge that can be added to the ontology to make it more complete.			
Examples			
Graphical representation and/or OWL code		Natural language description	
 <pre> graph LR Route[Route] -- startsInPoint --> Point[Point] </pre> <p><i>No explicit evidence of:</i> Declaration(ObjectProperty(:endsInPoint))</p>		The graphical example shows that the object property “startsInPoint” has been defined to indicate in which point a route starts. The lack of a property representing the end of the route is also indicated. However, that information should also be included.	
How to solve it			
It is advisable to check the ontology against the ORSD in order to discover missing classes, relationships and attributes and to include them in the ontology being built. In addition, one could check for new information in the following ways, among others: (a) Check whether any new class can be added and whether it belongs to a hierarchy. (b) Check whether any new property might have a complementary action that indicates an inverse property. Then one should include the new property.			

Table 4.9: P09. Missing domain information

ANEXO C – PITFALL 10 DE POVEDA-VILLALÓN (2016)

Title	P10. Missing disjointness		Importance level	Important
Aspects	Real world modelling		Affects to	Classes Object properties Datatype properties
Description				
The ontology lacks disjoint axioms between classes or between properties that should be defined as disjoint. This pitfall is related with the guidelines provided in [1], [2] and [3].				
Examples				
Graphical representation and/or OWL code			Natural language description	
 <p><i>No explicit evidence of:</i> DisjointClasses(:Odd:Even)</p>			The graphical example shows that the classes “Odd” and “Even” are included in the ontology. However, the disjoint axiom <code>owl:disjointWith</code> among them is not declared. In this situation there could be individuals belonging to “Odd” and “Even” numbers at the same time which is conceptually wrong.	
How to solve it				
For classes:				
(a) Look for pairs of concepts that can not share instances and define them as disjoint classes using the primitive <code>owl:disjointWith</code> . If a set of classes from the ontology are pairwise disjoint, then use <code>owl:AllDisjointClasses</code> from OWL 2.				
For properties (only for OWL 2):				
(a) Look for pairs of properties for which there are no two individuals that are interlinked by both properties and define them as disjoint properties using the primitive <code>owl:propertyDisjointWith</code> . If a set of properties from the ontology are pairwise disjoint, then use <code>owl:AllDisjointProperties</code> .				
References	<p>[1] Gómez-Pérez, A. (2004). Ontology evaluation. In <i>Handbook on ontologies</i>, pages 251-273. Springer.</p> <p>[2] Noy, N. F., McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology.</p> <p>[3] Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., and Wroe, C. (2004). OWL pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns. In <i>Engineering Knowledge in the Age of the Semantic Web</i>, pages 63-81. Springer.</p>			

Table 4.10: P10. Missing disjointness

ANEXO D – PITFALL 11 DE POVEDA-VILLALÓN (2016)

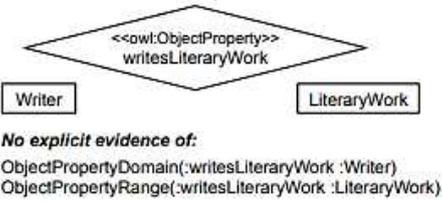
Title	P11. Missing domain or range in properties	Importance level	Important
Aspects	No inference Ontology understanding	Affects to	Object properties Datatype properties
Description			
Object and/or datatype properties without domain or range (or none of them) are included in the ontology.			
Examples			
Graphical representation and/or OWL code		Natural language description	
 <p><i>No explicit evidence of:</i> ObjectPropertyDomain(:writesLiteraryWork :Writer) ObjectPropertyRange(:writesLiteraryWork :LiteraryWork)</p>		<p>The graphical example shows that both the object property “writesLiteraryWork” and the classes “Writer” and “LiteraryWork” are included in the ontology. However, no connection between the classes and the object property is declared.</p>	
How to solve it			
<p>Domains and ranges should be declared in order to provide object and datatype properties with a more complete definition.</p> <p>For domains:</p> <p>(a) For each object or datatype property in the ontology without a domain defined, we recommend answering the following question “What is the most general class in the ontology whose instances could serve as subject of the property?”. The class that represents the answer will be the domain of the property. If the answer is <code>owl:Thing</code>, we recommend looking for several classes whose instances could serve as subjects of the property. Then, join these classes by the “or” operator (i.e., <code>owl:unionOf</code>) and set this union of classes as the domain of the property.</p> <p>For ranges:</p> <p>(a) For each object property in the ontology without a range defined, we recommend answering the following question “What is the most general class in the ontology whose instances could serve as object of the property?”. The class that represents the answer will be the range of the property. If the answer is <code>owl:Thing</code>, we recommend looking for several classes whose instances could serve as subjects of the property. Then, join these classes by the “or” operator (i.e., <code>owl:unionOf</code>) and set this union of classes as the range of the object property.</p> <p>(b) For each datatype property in the ontology without a range defined, we recommend answering the following question “What would be the format of the data (strings of characters, positive numbers, dates, floats, etc.) used to fill in this information?”. Choose the built-in datatype from the hierarchy in [1] that is closest to the answer to the previous question and set it as the range of the datatype property. If there is more than one plausible candidate, find their lowest common ancestor and set it as the range of the datatype property.</p>			
References	[1] Built-in datatype hierarchy http://www.w3.org/TR/xmlschema-2/built-in-datatypes		

Table 4.11: P11. Missing domain or range in properties

ANEXO E – PITFALL 12 DE POVEDA-VILALLÓN (2016)

Title	P12. Equivalent properties not explicitly declared	Importance level	Important
Aspects	No inference Ontology understanding	Affects to	Object properties Datatype properties
Description			
The ontology lacks information about equivalent properties (<code>owl:equivalentProperty</code>) in the cases of duplicated relationships and/or attributes.			
Examples			
Graphical representation and/or OWL code		Natural language description	
<pre> graph TD ns3Org[ns3:Organization] -- ns1:has-member --> ns3Mem[ns3:Member] ns3Org -- ns2:hasMember --> ns3Mem </pre> <p>No explicit evidence of: EquivalentObjectProperties(ns1:has-member ns2:hasMember)</p>		<p>The graphical example shows two properties defined in different namespaces (ns1 and ns2) that refer to the same relationship between concepts, more precisely, both indicate the members of an organization. However, such properties are not defined as equivalent of each other, even though one could say that both properties have the same meaning and every time one property holds, the other property holds as well.</p>	
How to solve it			
<p>In order to check whether two properties (object properties or datatype properties) defined in two different namespaces could be defined as equivalent properties one should:</p> <ol style="list-style-type: none"> Check that both properties hold for the same set of individuals, verifying that their domains are the same class or denote the same set of individuals, even though they might involve different class restrictions. Check ranges: <ol style="list-style-type: none"> For object properties: verify that the ranges of both object properties are the same class or denote the same set of individuals, even though they might involve different class restrictions. For datatypes properties: verify that the ranges of both datatype properties are the same datatype. Check that both properties represent the same real-world relation between classes or the same attributes in case of datatype properties. <p>If (a), (b) and (c) hold, the properties should be declared equivalent using <code>owl:equivalentProperty</code>.</p> <p>Finally, check if both properties are really needed: if they are defined in the same namespace, they could be either redundant or they could have different intended meanings. In both cases, a disambiguation step would be needed. If they are redundant and they are defined in a namespace under your control, remove one of them. If they actually have the same meaning and are defined in namespaces you do not control, keep them, along with the equivalent property statement.</p>			

Table 4.12: P12. Equivalent properties not explicitly declared

ANEXO F – PITFALL 13 DE POVEDA-VILLALÓN (2016)

Title	P13. Inverse relationships not explicitly declared	Importance level	Minor
Aspects	No inference Ontology understanding	Affects to	Object properties
Description			
This pitfall appears when any relationship (except for those that are defined as symmetric properties using <code>owl:SymmetricProperty</code>) does not have an inverse relationship (<code>owl:inverseOf</code>) defined within the ontology.			
Examples			
Graphical representation and/or OWL code		Natural language description	
<pre> graph TD AA[AdministrativeArea] -- hasOfficialLanguage --> L[Language] L -- isOfficialLanguageOf --> AA </pre> <p><i>No explicit evidence of:</i> <code>InverseObjectProperties(hasOfficialLanguage :isOfficialLanguageOf)</code></p>		<p>The graphical example shows two object properties, namely “hasOfficialLanguage” and “isOfficialLanguageOf”. The former describes that an administrative area has an official language, and the latter describes that a language is official language of an administrative area. In this case, both properties are defined in the ontology but no <code>owl:inverseOf</code> statement is declared between them.</p>	
How to solve it			
<p>In principle, all relationships, apart from the symmetric ones, could have an inverse relationship. An exception for this rule might be the properties created for an n-ary relationship, but not necessarily.</p> <p>For each object property without an inverse property (<code>owl:inverseOf</code>) in the ontology, follow one of the alternatives:</p> <ol style="list-style-type: none"> Check whether there is already an object property in the ontology that represents its inverse relationship. If so, verify that the domain of each relationship matches the range of the other. Make sure that the expressions used to define the domain of one relationship represent the same set of individuals denoted by the range of the other relationship, even though they might involve different class restrictions. If this is the case, define the two object properties as inverse using <code>owl:inverseOf</code>. Verify whether it is reasonable to create an inverse object property by altering the verb from active to passive voice or by using an adequate converse term*. In case this new object property is created, define it as inverse of the one being analysed using <code>owl:inverseOf</code>. Include, if possible, the domain and range of the new object property. Make sure that the expressions used to define the domain of each relationship denote the same set of individuals denoted by the range of the other property, even though they might involve different class restrictions. <p>*Converse terms are pairs of terms that refer to a relationship from opposite points of view, such as buy/sell, parent/child, among others. Note that a converse term does not always represent an inverse relationship and it might have to be adapted. For example, “buy” would not be the inverse of “sell” according to the semantics of <code>owl:inverseOf</code>. However, the relationship “hasParent” would be inverse of “hasChild”. See pitfall “P05. Defining wrong inverse relationships” for further details.</p>			

Table 4.13: P13. Inverse relationships not explicitly declared

ANEXO G – PITFALL 14 DE POVEDA-VILALLÓN (2016)

Title	P14. Misusing “owl:allValuesFrom”	Importance level	Critical
Aspects	Modelling decisions	Affects to	Classes
Description			
This pitfall consists in using the universal restriction (owl:allValuesFrom) as the default qualifier instead of the existential restriction (owl:someValuesFrom). Additional information about this pitfall is provided in [1].			
Examples			
Graphical representation and/or OWL code		Natural language description	
		<p>In the graphical example a definition of the class “Book” is provided by means of an owl:equivalentClass axiom in the following way: $Book \equiv \exists producedBy.Writer \sqcap \forall uses.Paper$. While the owl:someValuesFrom axiom is properly used for stating that the book has to be produced by at least one writer, it is not correct to say that all the materials used in the book have to belong to the class “Paper”, as for example, another material used during the production of a book might be “Ink”, among others.</p>	
How to solve it			
<p>An universal restriction (owl:allValuesFrom) can be used to restrict the range of a relationship, i.e., to state that, in the context of an axiom, only individuals belonging to the specified class can act as objects of that relationship. Considering a class “ClassA” used as target of an universal restriction for a given relationship, one should:</p> <ol style="list-style-type: none"> Answer the question “Is it possible to have individuals that do not belong to “ClassA” acting as object of such property?” If the answer is “yes”, the universal restriction (owl:allValuesFrom) should be deleted. To analyse whether the universal restriction should be replaced by an existential one check the point (b). Check whether the intended meaning of the restriction is to state that at least one individual belonging to “ClassA” should appear as object in an instantiation of such property. In this case, an existential restriction (owl:someValuesFrom) should be used instead of the universal one (owl:allValuesFrom). 			
References	<p>[1] Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H., and Wroe, C. (2004). OWL pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns. In <i>Engineering Knowledge in the Age of the Semantic Web</i>, pages 63-81. Springer.</p>		

Table 4.14: P14. Misusing “owl:allValuesFrom”

ANEXO H – PITFALL 30 DE POVEDA-VILLALÓN (2016)

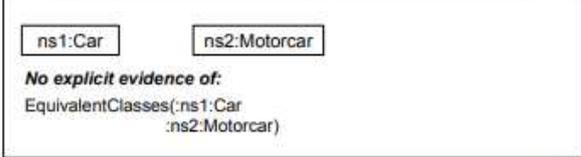
Title	P30. Equivalent classes not explicitly declared	Importance level	Important
Aspects	No inference	Affects to	Classes
Description			
This pitfall consists in missing the definition of equivalent classes (<code>owl:equivalentClass</code>) in case of duplicated concepts. When an ontology reuses terms from other ontologies, classes that have the same meaning should be defined as equivalent in order to benefit the interoperability between both ontologies.			
Examples			
Graphical representation and/or OWL code		Natural language description	
		The graphical example shows two classes, namely “Car” and “Motorcar”, defined in different namespaces and that refer to the same real world concept. However, no <code>owl:equivalentClass</code> axiom between them has been stated.	
How to solve it			
If the classes involved in the pitfall are defined in different namespaces and both are required, define them as equivalent by means of the <code>owl:equivalentClass</code> primitive. Consider also using other primitives such as <code>rdfs:subclassOf</code> in case they are not completely equivalent but one concept is more specific than the other.			
If both classes are defined in the same namespace under the developer’s control, we recommend removing one of them and attach its label (via <code>rdfs:label</code> or other label annotation property) to the remaining class. See “P02. Creating synonyms as classes” for further details.			

Table 4.30: P30. Equivalent classes not explicitly declared