

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

RAFAEL DENERSON RAMOS DE SOUSA

**UMA ONTOLOGIA PARA O DOMÍNIO DE TRABALHOS DE CONCLUSÃO DE
CURSO DE PÓS-GRADUAÇÃO STRICTO SENSU NO BRASIL:
IMPLEMENTAÇÃO EM DADOS LIGADOS INTEROPERÁVEIS NA
WEB SEMÂNTICA**

Vitória

2023

RAFAEL DENERSON RAMOS DE SOUSA

**UMA ONTOLOGIA PARA O DOMÍNIO DE TRABALHOS DE CONCLUSÃO DE
CURSO DE PÓS-GRADUAÇÃO STRICTO SENSU NO BRASIL:
IMPLEMENTAÇÃO EM DADOS LIGADOS INTEROPERÁVEIS NA
WEB SEMÂNTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Ciência da Informação.

Linha de pesquisa: Linha 2 - Memória, Representação e Informação.

Orientador: Prof. Dr. Henrique Monteiro Cristovão.
Coorientadora: Prof^a. Dra. Maria das Graças da Silva Teixeira.

Vitória

2023

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

S725o Sousa, Rafael Denerson Ramos de, 1985-
Uma ontologia para o domínio de trabalhos de conclusão de
curso de pós-graduação stricto sensu no Brasil : implementação em
dados ligados interoperáveis na Web semântica / Rafael
Denerson Ramos de Sousa. - 2023.
211 f. : il.

Orientador: Henrique Monteiro Cristovão.
Coorientadora: Maria das Graças da Silva Teixeira.
Dissertação (Mestrado em Ciência da Informação) -
Universidade Federal do Espírito Santo, Centro de Ciências
Jurídicas e Econômicas.

1. Ciência da informação. 2. Recuperação da informação. 3.
Ontologia. 4. Web semântica. 5. Metadados. 6. Educação. I.
Cristovão, Henrique Monteiro. II. Teixeira, Maria das Graças da
Silva. III. Universidade Federal do Espírito Santo. Centro de
Ciências Jurídicas e Econômicas. IV. Título.

CDU: 001

RAFAEL DENERSON RAMOS DE SOUSA

UMA ONTOLOGIA PARA O DOMÍNIO DE TRABALHOS DE CONCLUSÃO DE CURSO DE PÓS-GRADUAÇÃO STRICTO SENSU NO BRASIL: IMPLEMENTAÇÃO EM DADOS LIGADOS INTEROPERÁVEIS NA WEB SEMÂNTICA

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo, como requisito para a obtenção do título de Mestre em Ciência da Informação.

Linha de pesquisa: Linha 2 - Memória, Representação e Informação.

Aprovado em 4 de julho de 2023.

COMISSÃO EXAMINADORA

[assinatura digital]

Prof. Dr. Henrique Monteiro Cristovão
Universidade Federal do Espírito Santo
Membro Titular Interno
Orientador

[assinatura digital]

**Prof^a. Dr^a. Maria das Graças da Silva
Teixeira**
Universidade Federal do Espírito Santo
Membro Titular Interno
Co Orientadora

[assinatura digital]

Prof^a. Dr^a. Daniela Lucas da Silva Lemos
Universidade Federal do Espírito Santo
Membro Titular Interno

[assinatura digital]

Prof. Dr. Dalton Lopes Martins
Universidade de Brasília
Membro Titular Externo





Ata de defesa - Rafael

Data e Hora de Criação: 05/07/2023 às 07:50:36

Documentos que originaram esse envelope:

- Ata de defesa - Rafael.pdf (Arquivo PDF) - 1 página(s)



Hashs únicas referente à esse envelope de documentos

[SHA256]: 9e1bd850765925124314e2eb9e003656c8d43e62643c2c90442e5f6b04dd755

[SHA512]: 8d5f94f9772a735e4c54b07d2e5d10e4e795fedc32a214a4db691bf13b07d65c753ed06b56034721281fad9dc0e378b872bfc7bd25bf8f1a73c16a3c872a152a

Lista de assinaturas solicitadas e associadas à esse envelope



ASSINADO - Dalton Lopes Martins (daltonmartins@unb.br)

Data/Hora: 05/07/2023 - 09:36:20, IP: 179.255.8.133

[SHA256]: f58c6466b3d6b8e345cff925573aa86bd0f881ce7ca6b9f9a43e51bd1acfb762



ASSINADO - Daniela Lucas da Silva Lemos (daniela.l.silva@ufes.br)

Data/Hora: 05/07/2023 - 08:16:56, IP: 187.36.170.78, Geolocalização: [-20.296757, -40.292876]

[SHA256]: b0451d4d590cb3a0613f7b5ed1ea5d54165051f23058cc93400a837a406e70ca



ASSINADO - Henrique Monteiro Cristovão (henrique.cristovao@ufes.br)

Data/Hora: 05/07/2023 - 11:05:29, IP: 177.104.235.45, Geolocalização: [-20.258265, -40.264288]

[SHA256]: 5de25c7cc348853af356d007ec5a90941cdabbc03a4839f69d7f4a73c2a55256



ASSINADO - Maria das Graças da Silva Teixeira (maria.teixeira@ufes.br)

Data/Hora: 05/07/2023 - 09:17:35, IP: 189.91.159.34, Geolocalização: [-18.716087, -39.842722]

[SHA256]: 0c0b6754b76cf52d23f36c17d89e36a93c8294b8390417ef80aba431ff59d13f



ASSINADO - Rafael Denerson Ramos de Sousa (rafael.d.sousa@edu.ufes.br)

Data/Hora: 07/07/2023 - 11:43:18, IP: 201.54.47.20

[SHA256]: 098cec2b733b27a36343c9a88a862af3a448b6b5f4bd57285d9e4535b98a3ee0

Histórico de eventos registrados neste envelope

07/07/2023 11:43:18 - Envelope finalizado por rafael.d.sousa@edu.ufes.br, IP 201.54.47.20
07/07/2023 11:43:18 - Assinatura realizada por rafael.d.sousa@edu.ufes.br, IP 201.54.47.20
07/07/2023 11:41:34 - Envelope visualizado por rafael.d.sousa@edu.ufes.br, IP 201.54.47.20
05/07/2023 11:05:29 - Assinatura realizada por henrique.cristovao@ufes.br, IP 177.104.235.45
05/07/2023 11:05:08 - Envelope visualizado por henrique.cristovao@ufes.br, IP 177.104.235.45
05/07/2023 09:36:21 - Assinatura realizada por daltonmartins@unb.br, IP 179.255.8.133
05/07/2023 09:36:13 - Envelope visualizado por daltonmartins@unb.br, IP 179.255.8.133
05/07/2023 09:17:35 - Assinatura realizada por maria.teixeira@ufes.br, IP 189.91.159.34
05/07/2023 09:17:24 - Envelope visualizado por maria.teixeira@ufes.br, IP 189.91.159.34
05/07/2023 08:16:56 - Assinatura realizada por daniela.l.silva@ufes.br, IP 187.36.170.78
05/07/2023 08:16:16 - Envelope visualizado por daniela.l.silva@ufes.br, IP 187.36.170.78
05/07/2023 07:53:52 - Envelope registrado na Blockchain por edma.jantorno@ufes.br, IP 200.137.65.106
05/07/2023 07:53:51 - Envelope encaminhado para assinaturas por edma.jantorno@ufes.br, IP 200.137.65.106
05/07/2023 07:50:37 - Envelope criado por edma.jantorno@ufes.br, IP 200.137.65.106

Esta dissertação é dedicada à minha verdadeira família.

AGRADECIMENTOS

Aos meus pais, por me apoiarem durante todo esse tempo de estudos, mesmo nos momentos mais turbulentos que fizeram parte da minha vida.

Ao Renato de Lima Stelzer e à Merilin Soraia Ramos de Sousa, por terem me concedido um lar em vários momentos, e assim eu pude realizar minhas pesquisas. E inclusive por me emprestarem também um computador.

Ao Luciano Marques da SUPG/CCJE, por ter me atendido com presteza em todas as vezes que solicitei informações e precisei resolver ou encaminhar assuntos do Programa de Pós-Graduação em Ciência da Informação (PPGCI).

Ao meu orientador Henrique Monteiro Cristovão, por todo conhecimento compartilhado, por toda dedicação, serenidade, acompanhamento pedagógico implacável, e todo cuidado em fornecer um conhecimento de excelência, cada incentivo e cobrança, e principalmente, pelo exemplo ético, profissional e humano.

A minha coorientadora Maria das Graças da Silva Teixeira, por ter enriquecido esta pesquisa com sua capacidade intelectual especializada, pelas sugestões pontuais no decorrer desta pesquisa, por ter me conduzido durante esse período que pesquisamos, e por toda paciência em cada reunião.

Aos membros da banca examinadora Dalton Lopes Martins e Daniela Lucas da Silva Lemos, por cada recomendação, comentário, indicação de materiais e por fornecerem uma avaliação cuidadosa e criteriosa.

As professoras Lucileide Andrade de Lima do Nascimento, Rosa da Penha Ferreira da Costa e Maria Lucia Teixeira Garcia, por contribuírem de maneira imensurável com a pesquisa e aceitaram ser entrevistadas sobre o âmbito do domínio pesquisado.

Aos professores do PPGCI, pelo conhecimento de excelência que foi transmitido ao longo do curso.

A cada colega do curso: pela convivência e por cada experiência compartilhada.

A Universidade Federal do Espírito Santo, por me proporcionar a oportunidade de tentar avançar mais um degrau em minha carreira acadêmica.

"Cause when you trust someone, illusion has begun."

Type O Negative

RESUMO

O aumento, tanto em quantidade quanto em variedade, de conteúdos na Web tem provocado dificuldades quanto à recuperação de informação (RI) que atenda às necessidades de interessados em tais informações. A organização do conhecimento é fragilizada pela ausência de ontologias de domínio que representem de forma adequada a realidade e a subutilização de boas práticas de publicação na Web, como os princípios FAIR, em especial quanto à interoperabilidade, e isso têm também provocado problemas na RI. Nesse contexto, a Web semântica pode oferecer recursos de publicação para permitir RI por meio de consultas integradas com outras bases e repositórios de dados. Ilustrando tal situação, de organização e publicação de dados na Web, observam-se deficiências no conjunto de trabalhos de conclusão de curso de pós-graduação no Brasil. Diante do exposto, o problema de pesquisa deste trabalho é como organizar os metadados de trabalhos de conclusão de curso utilizando-se de ontologias para que seja possível recuperá-los de forma interoperável? Acredita-se que economizar tempo, minimizar gastos, expandir a transparência e aprimorar a qualidade dos dados e dos serviços prestados pelos setores públicos fazem parte da publicidade e da eficiência. Alinhado ao problema de pesquisa, o objetivo geral é criar uma ontologia de referência de domínio e mapeá-la para uma ontologia operacional para representar trabalhos de conclusão de curso de pós-graduação no contexto da Web semântica e fundamentados em boas práticas de publicação. O trabalho possui abordagem de pesquisa qualitativa, natureza aplicada e procedimentos de estudo de caso. Utilizou a metodologia de construção de ontologias denominada SABiO, suportada pela ontologia de fundamentação UFO e a linguagem de modelagem orientada a ontologia OntoUML. Como resultados, destacam-se o modelo conceitual do estudo de caso por meio de uma ontologia de referência de domínio, o modelo de *design* para a ontologia operacional, a ontologia operacional implementada em *Turtle*, uma base de dados ligados RDF situada em um repositório de um banco de dados orientado a grafos, e algumas consultas em SPARQL como prova de conceito. Observou-se que os dados ligados serializados em RDF atenderam boa parte dos preceitos das boas práticas de publicação na Web. Como continuidade da pesquisa, indica-se a validação dos artefatos desenvolvidos em uma base de dados substancialmente maior e o desenvolvimento de interfaces para oferecer aos usuários finais possibilidades de consultas com boa usabilidade.

Palavras-chave: ontologia de domínio; Web semântica; dados ligados; interoperabilidade; Educação.

ABSTRACT

The increase, in quantity and variety, of content on the Web has caused difficulties in terms of information retrieval (IR) that meets the needs of those interested in such information. The organization of knowledge is weakened by the absence of domain ontologies that adequately represent reality and the underuse of data on the Web best practices, such as the FAIR principles, especially with regard to interoperability, and this has also caused problems in IR. In this context, the semantic Web can offer publishing resources to allow IR through integrated queries with other databases and data repositories. Illustrating this situation, of organization and publication of data on the Web, deficiencies are observed in the set of postgraduate course conclusion works in Brazil. Given the above, the research problem of this work is how to organize the metadata of course completion papers using ontologies so that it is possible to retrieve them in an interoperable way? It is believed that saving time, minimizing expenses, expanding transparency and improving the quality of data and services provided by public sectors are part of publicity and efficiency. Aligned with the research problem, the general objective is to create a domain reference ontology and map it to an operational ontology to represent postgraduate course conclusion works in the context of the semantic Web and based on good publishing practices. The work has a qualitative research approach, applied nature and case study procedures. It used the ontology construction methodology called SABiO, supported by the UFO foundation ontology and the ontology-oriented modeling language OntoUML. As results, the conceptual model of the case study through a domain reference ontology, the design model for the operational ontology, the operational ontology implemented in Turtle, a database linked RDF located in a repository of a database graph oriented, and some SPARQL queries as a proof of concept. It was observed that the linked data serialized in RDF met a good part of the precepts of good practices of Web publishing. As a continuation of the research, it is indicated the validation of the artifacts developed in a substantially larger database and the development of interfaces to offer end users possibilities of queries with good usability.

Keywords: domain ontology; semantic Web; linked data; interoperability; Education.

LISTA DE FIGURAS

Figura 1 - Mapa conceitual de ontologia na Filosofia (link para acesso em tamanho maior)	49
Figura 2 - Mapa conceitual de ontologia na Ciência da Computação (link para acesso em tamanho maior)	50
Figura 3 - Mapa conceitual de ontologia na Ciência da Informação (link para acesso em tamanho maior)	52
Figura 4 - Arquitetura da Web semântica conhecida como "bolo de noiva" ou "pirâmide da Web semântica" (link para acesso em tamanho maior)	57
Figura 5 - As quatro camadas da interoperabilidade (link para acesso em tamanho maior)	73
Figura 6 - Mapa Conceitual contextualizando alguns elementos da Fundamentação Teórica (link para acesso em tamanho maior)	85
Figura 7- Processo de metodologia SABiO (link para acesso em tamanho maior) ...	94
Figura 8 - Diagrama de atividades das etapas metodológicas do desenvolvimento (link para acesso em tamanho maior).....	116
Figura 9 - <i>Knowledge graph</i> do modelo conceitual preliminar do domínio (link para acesso em tamanho maior).....	128
Figura 10 - Ontologia de referência de domínio - diagrama Defesa (link para acesso em tamanho maior)	130
Figura 11 - Ontologia de referência de domínio - diagrama Colaborador (link para acesso em tamanho maior).....	131
Figura 12 - Ontologia de referência de domínio - diagrama Bolsa (link para acesso em tamanho maior)	132
Figura 13 - Ontologia de referência de domínio - diagrama Área (link para acesso em tamanho maior)	133
Figura 14 - Modelo de <i>design</i> - diagrama Defesa (link para acesso em tamanho maior)	135
Figura 15 - Especificação das entidades - anotações de implementação do modelo de <i>design</i> da ontologia operacional de domínio (link para acesso em tamanho maior)	136
Figura 16 - Especificação das relações - anotações de implementação do modelo de <i>design</i> da ontologia operacional de domínio (link para acesso em tamanho maior)	138

Figura 17 - Modelo de <i>design</i> - diagrama Colaborador (link para acesso em tamanho maior)	139
Figura 18 - Modelo de <i>design</i> - diagrama Bolsa (link para acesso em tamanho maior)	140
Figura 19 - Modelo de <i>design</i> - diagrama Área (link para acesso em tamanho maior)	142
Figura 20 - Arquivo em <i>Turtle</i> no VocBench (link para acesso em tamanho maior)	154
Figura 21 - Arquivo em <i>Turtle</i> no Notepad++ (link para acesso em tamanho maior)	154
Figura 22 - Reconciliação de dados com bases externas (link para acesso em tamanho maior)	156
Figura 23 - Esquema de mapeamento para dados ligados (link para acesso em tamanho maior)	158
Figura 24 - Repositório de dados ligados em RDF (link para acesso em tamanho maior)	158
Figura 25 - Consulta SPARQL para as três palavras-chave de maior ocorrência (link para acesso em tamanho maior).....	160

LISTA DE QUADROS

Quadro 1 - Levantamento bibliográfico do Estado da Arte	30
Quadro 2 - Exposição das fases do levantamento bibliográfico	113
Quadro 3 - Atividades das etapas procedimentais do desenvolvimento	117
Quadro 4 - Objetos das etapas procedimentais do desenvolvimento	118
Quadro 5 - Vocabulários de termos usados na interoperabilidade do modelo de <i>design</i>	142
Quadro 6 - Equivalência de classe usada no modelo de <i>design</i> com <i>equivalentClass</i>	145
Quadro 7 - Equivalência de propriedade de dados adotada no modelo de <i>design</i> com <i>DatatypeProperty</i>	148
Quadro 8 - Equivalência de propriedade de objetos adotada no modelo de <i>design</i> com <i>ObjectProperty</i>	150
Quadro 9 - Semelhança entre esta dissertação e a pesquisa científica de Silva, Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b).....	164
Quadro 10 - Divergência entre esta dissertação e a pesquisa científica de Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b).....	165
Quadro 11 - Glossário de termos da ontologia de domínio	200

LISTA DE ABREVIATURAS E SIGLAS

ACM	<i>Association for Computing Machinery</i>
ADP	Aliança de Dados de Pesquisa
API	<i>Application Programming Interface</i>
AVA	Ambiente Virtual de Aprendizagem
BDA	Barômetro de Dados Abertos
BDTD	Biblioteca Brasileira de Teses e Dissertações
BFO	<i>Basic Formal Ontology</i>
BIBFRAME	<i>Bibliographic Framework Initiative</i>
BNF	<i>Bibliothèque Nationale de France</i>
BRAPCI	Base de Dados em Ciência da Informação
CAFe	Comunidade Acadêmica Federada
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBIR	<i>Stanford Center for Biomedical Informatics Research</i>
CC	<i>Creative Commons Rights Expression Language</i>
CCJE	Centro de Ciências Jurídicas e Econômicas
CI	Ciência da Informação
CSV	<i>Comma Separated Values</i>
DC	Dublin Core
DCAT	<i>Data Catalog Vocabulary</i>
DL	<i>Description Logics</i>
DLAN	Dados Ligados e Abertos em Nuvem
DLP	<i>Description Logic Programs</i>
DQV	<i>Data Quality Vocabulary</i>
DOLCE	<i>Descriptive Ontology for Linguistic and Cognitive Engineering</i>
DWBP	<i>Data on the Web Best Practices</i>
ENANCIB	Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação
FAIR	<i>Findable, Accessible, Interoperable and Reusable</i>
FAPESP	Fundação de Apoio à Pesquisa do Estado de São Paulo
FDP	<i>FAIR Data Point</i>
FOAF	<i>Friend of a Friend</i>
GODI	<i>Global Open Data Index</i>

GO-FAIR	<i>Global Open-Findable, Accessible, Interoperable and Reusable</i>
HTTP	<i>Hypertext Transfer Protocol</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IA	Inteligência Artificial
IBICT	Instituto Brasileiro de Informações em Ciência e Tecnológica
IDE	<i>Integrated Development Environment</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
IFLA	<i>International Federation of Library Associations and Institutions</i>
IGDA	Índice Global de Dados Abertos
IHMC	<i>Institute for Human and Machine Cognition</i>
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IODE	<i>International Oceanographic Data and Information Exchange</i>
IoT	<i>Internet of Things</i>
ISKO	<i>International Society for Knowledge Organization</i>
JSON-LD	<i>JavaScript Object Notation for Linked Data</i>
KG	<i>knowledge graph</i>
KO	<i>Knowledge Organization</i>
KOS	<i>Knowledge Organization Systems</i>
KR	<i>Knowledge Representation</i>
LODC	<i>Linked Open Data Cloud</i>
LOD	<i>Linked Open Data</i>
LOV	<i>Linked Open Vocabularies</i>
MEC	Ministério da Educação e Cultura
MPDW	Melhores Práticas para os Dados na Web
NeOn	<i>Network Ontology</i>
OC	Organização do Conhecimento
OD	<i>Open Data</i>
ODB	<i>Open Data Barometer</i>
OntoUML	<i>Ontological Unified Modelling Language</i>
OO	Orientado a Objeto
OpenAIRE	<i>Open Access Infrastructure for Research in Europe</i>
OTK	<i>On-To-Knowledge</i>
OWL	<i>Web Ontology Language</i>

PDA	Plano de Dados Abertos
PDAА	Princípios de Dados Abertos Anticorrupção
PPGCI	Programa de Pós-Graduação em Ciência da Informação
RI	Recuperação da Informação
RDA	<i>Research Data Alliance</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>RDF Schema</i>
SABiO	<i>Systematic Approach for Building Ontologies</i>
SBC	Sociedade Brasileira de Computação
SciELO	<i>Scientific Electronic Library Online</i>
SI	Sistemas de Informação
SKOS	<i>Simple Knowledge Organization System</i>
SMPO	Sistema de Melhores Práticas Oceânicas
SOC	Sistema de Organização do Conhecimento
SOL	SBC-Open Lib
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
SWRL	<i>A Semantic Web Rule Language</i>
TI	Tecnologia da Informação
TIC	Tecnologia da Informação e Comunicação
TTL	<i>Terse RDF Triple Language</i>
EU	União Européia
UFES	Universidade Federal do Espírito Santo
UFO	<i>Unified Foundational Ontology</i>
UML	<i>Unified Modeling Language</i>
URI	<i>Uniform Resource Identifier</i>
VCD	Vocabulário de Catálogo de Dados
VQD	Vocabulário de Qualidade de Dados
W3C	<i>World Wide Web Consortium</i>
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	19
1.1	PROBLEMA DE PESQUISA	20
1.2	OBJETIVO GERAL	23
1.3	OBJETIVOS ESPECÍFICOS	23
1.4	JUSTIFICATIVA	24
1.5	ORGANIZAÇÃO DO TRABALHO	27
2	ESTADO DA ARTE: BOAS PRÁTICAS DE PUBLICAÇÃO DE DADOS NA WEB NO CONTEXTO DE TRABALHOS DE CONCLUSÃO DE CURSO DE PÓS-GRADUAÇÃO	28
3	FUNDAMENTAÇÃO TEÓRICA	36
3.1	ORGANIZAÇÃO DO CONHECIMENTO NA CIÊNCIA DA INFORMAÇÃO	36
3.1.1	Organização do Conhecimento	37
3.1.2	Sistemas de Organização do Conhecimento	40
3.1.3	Representação da Informação	41
3.1.4	Metadados	42
3.1.5	Recuperação da Informação	43
3.1.6	Comunicação Científica	46
3.2	ONTOLOGIA.....	47
3.3	WEB SEMÂNTICA	54
3.3.1	Organização da Web semântica	56
3.3.2	Vocabulários Controlados	61
3.3.3	Knowledge Graph (Grafo de Conhecimento)	65
3.3.4	Dados Ligados	66
3.3.5	Interoperabilidade	70
3.4	BOAS PRÁTICAS NA PUBLICAÇÃO DE DADOS NA WEB	74
3.4.1	Princípios FAIR	74

3.4.2	Boas Práticas na Publicação de Dados na Web	76
3.4.3	Boas Práticas na Publicação de Dados Ligados	83
3.5	CONSIDERAÇÕES FINAIS DA FUNDAMENTAÇÃO TEÓRICA.....	84
4	FUNDAMENTAÇÃO METODOLÓGICA	87
4.1	ORGANIZAÇÃO GERAL DA PESQUISA.....	87
4.1.1	Mapas Conceituais	87
4.1.2	Gerenciador Bibliográfico	89
4.2	DESENVOLVIMENTO.....	91
4.2.1	Modelagem conceitual baseada em ontologia	91
4.2.1.1	Metodologias para construção de ontologias de domínio.....	91
4.2.1.2	Ontologias de fundamentação.....	99
4.2.1.3	Linguagem de modelagem orientada a ontologia.....	102
4.2.1.4	Ferramenta gráfica para modelagem de ontologias.....	103
4.2.2	Design e Implementação da ontologia operacional de domínio	104
4.2.2.1	Ferramentas para construção e implementação de ontologias.....	104
4.2.2.2	Linguagens para representação e serialização de ontologias operacionais.....	106
4.2.3	Preparação de Dados	107
4.2.4	Mapeamento para base de dados ligados RDF	108
4.2.5	Banco de dados orientados a grafos	109
4.2.6	Consultas em RDF	109
4.2.7	Visualização em RDF	110
5	PROCEDIMENTOS METODOLÓGICOS	111
5.1	CLASSIFICAÇÃO DA PESQUISA.....	111
5.2	LEVANTAMENTO BIBLIOGRÁFICO.....	112
5.3	CENÁRIO DO ESTUDO DE CASO.....	114
5.4	ETAPAS PROCEDIMENTAIS DO DESENVOLVIMENTO.....	115
5.4.1	Etapa 1: Elaboração do modelo conceitual	119

5.4.2	Etapa 2: Construção do modelo de <i>design</i>	121
5.4.3	Etapa 3: Implementação da ontologia operacional	124
5.4.4	Etapa 4: Preparação e reconciliação de dados	125
5.4.5	Etapa 5: Mapeamento para base de dados ligados RDF	126
5.4.6	Etapa 6: Recuperação de Informação.....	127
6	RESULTADOS E DISCUSSÃO	128
6.1	ONTOLOGIA DE REFERÊNCIA DO DOMÍNIO	128
6.2	MODELO DE <i>DESIGN</i> PARA A ONTOLOGIA OPERACIONAL.....	134
6.3	ONTOLOGIA OPERACIONAL	152
6.4	DADOS PRÉ-PROCESSADOS E RECONCILIADOS.....	155
6.5	MAPEAMENTO DOS DADOS PARA RDF	157
6.6	BASE DE DADOS LIGADOS RDF	158
6.7	CONSULTAS E INFORMAÇÃO RECUPERADA	159
6.8	DISCUSSÕES GERAIS	160
7	CONSIDERAÇÕES FINAIS	167
8	REFERÊNCIAS.....	170
	APÊNDICES	194
	APÊNDICE A – ENTREVISTA EM 13/12/2022 USADA NA ELABORAÇÃO DA ONTOLOGIA DE REFERÊNCIA DE DOMÍNIO	195
	APÊNDICE B – ENTREVISTA EM 20/12/2022 USADA NA ELABORAÇÃO DA ONTOLOGIA DE REFERÊNCIA DE DOMÍNIO	197
	APÊNDICE C – GLOSSÁRIO DE TERMOS DA ONTOLOGIA DE DOMÍNIO	200

1 INTRODUÇÃO

Atualmente, informações compartilhadas por meio de tecnologias têm um papel relevante na sociedade, seja para divulgação de decisões políticas, compartilhamento e assimilação de conhecimento científico ou de senso comum, de intenções ou emoções pessoais por meio das tecnologias multimidiáticas nas comunidades virtuais. Nesse contexto, a Ciência da Informação (CI) tem discutido o impacto das tecnologias na comunicação contemporânea, bem como analisado e investigado as dimensões teóricas e epistemológicas que articulam cultura, preservação, memória, patrimônio e documentos (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2023a). A CI também aborda as bases ontológicas da representação e recuperação do conhecimento, conhecidas como Sistemas de Organização do Conhecimento (SOC), e, dessa maneira, apresenta possibilidades e mecanismos para organizar, publicar e recuperar informações que possam ser processadas em repositórios, bases de dados ou acervos.

Os mecanismos e as estratégias para representar e recuperar informação são diversos, considerando um ou mais domínios. Tanto que a Web semântica como padronização da Web quanto aos dados ligados (W3C, 2011) se consolidou como uma rede de informações e tornou-se um desses mecanismos para representar e recuperar informação na Web (LAUFER, 2015). Para isso, ela conta com a organização e a representação por meio de ontologias, visto como um conceito interdisciplinar definido como uma especificação explícita e formal de uma conceituação compartilhada (GRUBER, 1994), capaz de proporcionar organização e representação do conhecimento.

Ontologias podem ser categorizadas como ontologia de referência de domínio e como ontologia operacional de domínio, sendo a primeira responsável por tratar de especificações em um nível conceitual, enquanto a segunda refere-se a especificação sistemática para estabelecer uma implementação possível para um determinado domínio (FALBO, 2014). O uso de ontologias na Web semântica possibilita a interoperabilidade¹ das camadas sintática, estrutural e semântica da rede Web, pois na camada sintática trata-se das diferenças de codificação, decodificação e

¹ Interoperabilidade é a capacidade de um sistema se relacionar e se comunicar com outro. É a faculdade de dados ou dispositivos de recursos não cooperativos serem capazes de incorporar ou funcionar em conjunto com o mínimo de esforço possível (WILKINSON *et al.*, 2016).

representação dos dados (ZENG, 2019), na camada estrutural busca-se alteração nas representações e arquitetura da informação contida nas estruturas de dados usando representações pré-definidas por comunidades e bibliotecas reconhecidas (ZENG, 2019), e na camada semântica almeja-se uma comunicação por meio de dados, informação e conhecimento entre diferentes agentes, serviços e aplicativos tanto dentro como fora da Web (USHOLD; MENZEL, 2005). Atualmente, a importância desses elementos interoperáveis está em possibilitar mecanismos de buscas que atendam aos princípios de Encontrabilidade, Acessibilidade, Interoperabilidade e Reusabilidade (*Findable, Accessible, Interoperable e Reusable* (FAIR)), que é o ideal de uma ciência aberta, guiando os pesquisadores nos problemas de administração e acessibilidade de dados de maneira mais ampla, buscando a reutilização de dados de pesquisas (BOECKHOUT; ZIELHUIS; BREDENOORD, 2018).

1.1 PROBLEMA DE PESQUISA

Existem diversos problemas relacionados à recuperação, organização e publicação de dados e informações que requerem Organização do Conhecimento (OC) de maneira adequada aos desafios da Web semântica, por exemplo, a existência de uma linguagem que expresse dados e regras para raciocinar sobre metadados², bem como permitir que regras de qualquer Sistema de Representação do Conhecimento existente possam representar esses metadados com uma linguagem padronizada em termos sintáticos e semânticos. Outro problema está na falta de uma linguagem coletiva com mecanismos de interoperabilidade e a prosperidade dessa linguagem para proporcionar escala de produção do conhecimento científico local e global (REVEZ; SILVA, 2021). Pearlman *et al.* (2021) exemplificam dizendo que a comunidade científica oceânica precisa de metodologias acessíveis em diferentes idiomas e modalidades, isto é, não diz respeito apenas aos documentos físicos, mas também aqueles que são oriundos de qualquer região, incluindo comunidades sub-representadas, pois isso estimula a propagação de recursos tecnológicos utilizados para descoberta aperfeiçoada de novos métodos científicos.

Esses problemas de recuperação, organização e publicação de dados e informações são decorrentes do aumento, tanto em quantidade como em variedade,

² Metadados são dados estruturados que possibilitam classificar, descrever e gerenciar documentos (BRASIL, 2020a).

de conteúdos na Web contemporânea. Isso provocou certas limitações quanto à exploração eficiente das consultas, bem como dificuldade em procurar e selecionar conteúdos especializados, como, por exemplo, buscar por pessoas com nomes iguais ou comuns, ou mesmo indivíduos famosos com nomes equivalentes (INIAKPOKEIKIYE; FAWEI, 2022).

Devido aos múltiplos contextos mencionados, o baixo uso de boas práticas tanto na organização como na publicação de dados provoca problemas na Recuperação da Informação (RI). A ausência de interoperabilidade em bases de dados acarreta uma RI reduzida, isto é, apenas na base pesquisada, além de dificultar que os dados da base sejam encontrados, acessados e reusados por outros sistemas de busca. A inexistência de práticas recomendadas para publicação e interrelação de dados estruturados para serem acessados por humanos e máquinas como dados ligados, dificulta consultas integradas com outras bases e repositórios de dados, como em acervos institucionais que utilizam Banco de Dados Relacionais em sua maior parte.

A CI é uma área que pode colaborar bastante com estudos da Web semântica (CONEGLIAN; SOUZA; SEGUNDO, 2016) e há baixa produção de estudos. Problemas encontrados nesse âmbito estão relacionados às pesquisas acadêmicas com nível baixo de interoperabilidade para publicar dados e informação na Web (GONÇALVES; JACYNTHO, 2020; INIAKPOKEIKIYE; FAWEI 2022), bem como a adoção de banco de dados relacional existente para publicar dados e informação na Web (GONÇALVES; JACYNTHO, 2020), o que acarreta em custo acentuado para adaptar ontologias que gerenciem e permitam consultas de forma simples e eficiente de dados universitários (INIAKPOKEIKIYE; FAWEI 2022).

O portal de periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) permite recuperar informações rápidas por assunto, periódicos, livros, bases e websites por meio da Comunidade Acadêmica Federada (CAFe) (FUNDAÇÃO COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2021). Embora a CAPES tenha a maior base virtual de apoio à pesquisa do Brasil, dispõe de conteúdo científico nacional e internacional composto por teses, dissertações, normas técnicas, patentes e material audiovisual (FUNDAÇÃO COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR, 2022) e possibilite um bom cruzamento de informações para o

pesquisador recuperar informação, há limitações como baixa adesão de interoperabilidade na RI do próprio acervo, não disponibiliza a recuperação de metadados e as consultas gratuitas são permitidas somente para instituições cadastradas segundo a Portaria n° 122/2017³ (BRASIL, 2017).

No âmbito da Organização do Conhecimento de trabalhos de conclusão de curso no nível de mestrado e doutorado existem diversas deficiências na literatura, tais como falta de interoperabilidade da camada sintática (SILVA; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B; ZHENG; TÖRMÄ; SEPPÄNEN, 2021; HAMIM; BENABBOU; SAEL, 2021) e da camada semântica (GOMES *et al.*, 2020; ASHOUR *et al.*, 2020), falta de conceitos relevantes em ontologias, como, por exemplo, instituições educacionais e no contexto de defesa e bolsa de estudos (ANDRADE; NARDI; RUY, 2023; SILVA; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B; SILVEIRA; ARAGÓN; CURY; MENEZES, 2021; HAMIM; BENABBOU; SAEL, 2021), ausência de implementação de consultas para recuperar informação e metadados de maneira interoperável (SILVA; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B; HAMIM; BENABBOU; SAEL, 2021; GOMES *et al.*, 2020; ASHOUR *et al.*, 2020), restrições na gestão de dados universitários devido ao uso reduzido de ontologia (ANDRADE; NARDI; RUY, 2023; INIAKPOKEIKIYE; FAWEI, 2022), falta de mapeamento automático para o esquema da base de dados sem etapas metodológicas (ANDRADE; NARDI; RUY, 2023; INIAKPOKEIKIYE; FAWEI, 2022), baixa adoção da utilização de dados ligados por meio de repositórios orientado a grafos (SILVA; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B; GONÇALVES; JACYNTHO, 2020; HAMIM; BENABBOU; SAEL, 2021), baixa adoção quanto a recuperação de informação e metadados de teses e dissertações (SILVA; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B; GONÇALVES; JACYNTHO, 2020; HAMIM; BENABBOU; SAEL, 2021; GOMES *et al.*, 2020; ASHOUR *et al.*, 2020), dificuldade considerável para adaptar ontologias que administrem e permitam consultas detalhadas e eficientes (INIAKPOKEIKIYE; FAWEI 2022; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B;

³ Portaria n° 122, de 19 de julho de 2017 estabelece os critérios de acesso ao Portal de Periódicos por parte das Instituições que são: Instituições federais de ensino superior; Instituições de pesquisa que possuam pós-graduação com pelo menos um programa com nota superior a 4 avaliada pela CAPES; Instituições públicas de ensino superior estaduais e municipais com pós-graduação (*stricto sensu*) com pelo menos um programa que tenha obtido nota 4 avaliada pela CAPES ou superior; Instituições privadas de ensino superior com pelo menos um doutorado avaliado pela CAPES nota 5 ou superior; Instituições com programas de pós-graduação recomendados pela CAPES e que atendam critérios de excelência estabelecidos pelo Ministério da Educação (BRASIL, 2017).

ZHENG; TÖRMÄ; SEPPÄNEN, 2021; HAMIM; BENABBOU; SAEL, 2021; GOMES *et al.*, 2020).

Diante dessa situação problemática contextualizada a partir de revisão de literatura, acredita-se que seja possível organizar metadados de trabalhos de conclusão de curso para recuperá-los de forma integrada (interoperável), utilizando-se de boas práticas para organizar informações, visando publicá-las e recuperá-las, baseando-se nos princípios FAIR para permitir que elas sejam recuperadas de maneira encontrável, acessível, interoperável e reutilizável. Além disso, a adoção de dados ligados interoperáveis na Web semântica possibilita a RI integrada entre bases por intermédio do uso de bases de dados orientadas a grafos, permitindo mecanismos de buscas em dados ligados em rede como grafos do conhecimento (YANG *et al.*, 2020).

Assim, a questão de pesquisa elaborada a partir deste contexto seria: como organizar metadados de trabalhos de conclusão de curso utilizando-se de ontologias e outros tipos de vocabulários para que seja possível recuperá-los de forma interoperável? Tal questão foi a base para definir as etapas e os produtos da presente pesquisa.

1.2 OBJETIVO GERAL

Criar uma ontologia de referência para o domínio de trabalhos de conclusão de curso de pós-graduação *stricto sensu* e mapeá-la em uma ontologia operacional no contexto de dados ligados na Web semântica.

1.3 OBJETIVOS ESPECÍFICOS

I. Investigar elementos da organização do conhecimento com suporte de dados ligados interoperáveis na Web semântica e fundamentados em boas práticas de publicação;

II. Elaborar um modelo conceitual baseado em ontologia de referência do domínio de trabalhos de conclusão de curso de pós-graduação *stricto sensu* no Brasil;

III. Criar um modelo de *design* para a ontologia operacional mapeada da ontologia de referência do objetivo específico (II), munida de elementos interoperáveis da camada estrutural e da camada semântica;

IV. Implementar uma ontologia operacional a partir do modelo de *design*

criado no objetivo específico (III);

V. Gerar uma base de dados ligados na Web semântica a partir de um mapeamento da ontologia operacional de domínio implementada no objetivo específico (IV) e de uma base de dados experimental, e previamente preparada, de dissertações com vínculos interoperáveis para elementos externos;

VI. Realizar consultas, para fins de validação do artefato ontológico implementado, na base de dados ligados gerada no objetivo específico (V) como prova de conceito.

1.4 JUSTIFICATIVA

A União Europeia (UE) tem discutido a interoperabilidade em dados ligados como questão pertinente à cidadania digital e a portabilidade dos direitos que fazem parte dos países que a compõem. Ela alerta que a digitalização da administração pública objetiva economizar tempo, minimizar gastos, expandir a transparência e aprimorar a qualidade dos dados e dos serviços prestados pelos setores públicos (FERREIRA *et al.*, 2019). Entretanto, Ferreira *et al.*, (2019) alerta que há certa dificuldade relacionada à inclusão na cidadania digital, devido a desterritorialização e a desmaterialização que impulsionaram a criação de novos direitos fundamentais, minimizando certas barreiras existentes nos territórios físicos. O Portal Oficial de Publicações da União Europeia⁴ (*The official portal for European data*) permite o acesso aos serviços de publicação de todas as instituições, organismos e agências da UE como publicações, legislação da UE, dados abertos, resultados de pesquisas, avisos de licitação, dentre outras informações oficiais (EUROPEAN UNION, 2023). Essa mobilização de abertura dos dados dos governos permite à população ter acesso aos indicadores de contribuição tributária, licitações orçamentárias, saúde, educação e também dos resultados de pesquisas institucionais (EUROPEAN UNION, 2023).

Macedo (2021) levantou estudos no âmbito da CI acerca das iniciativas e fases de evolução de portais de dados abertos dos EUA, Reino Unido, Irlanda e no Brasil, e apresentou algumas iniciativas de projetos voltados à disseminação de dados abertos do governo nas esferas públicas como Publicação de dados abertos e semânticos no Brasil, Portais de transparência da administração pública, Sistema Fiscalize, Portais de dados abertos estaduais, Portal Brasileiro de Dados Abertos, dentre outros. O

⁴ Disponível em: <https://data.europa.eu/en>

acesso e uso de dados e informações públicas disponibilizados pelo governo do Brasil é o Portal Brasileiro de Dados Abertos (BRASIL, 2023a) e há um projeto bem definido para acessibilidade e (re)uso de dados e informações públicas que é Plano de Dados Abertos (PDA) (BRASIL, 2021).

Isso se reflete aqui no Brasil, especialmente no âmbito da educação, devido ao Ministério da Educação e Cultura (MEC)⁵ ter como objetivo disponibilizar para toda comunidade brasileira dados e informações educacionais para serem usadas no desenvolvimento de aplicativos e ações com planos bienais bem definidos (BRASIL, 2023b), além de haver uma variedade considerável de pesquisas em desenvolvimento neste âmbito (ANDRADE; NARDI; RUY, 2023; INIAKPOKEIKIYE; FAWEI, 2022; SILVA; MUTZ; RUY, 2022A; SILVA; MUTZ; RUY, 2022B; SILVEIRA; ARAGÓN; CURY; MENEZES, 2021; ZHENG; TÖRMÄ; SEPPÄNEN, 2021; HAMIM; BENABBOU; SAEL, 2021; GONÇALVES; JACYNTHO, 2020; GOMES *et al.*, 2020; ASHOUR *et al.*, 2020; ASHOUR *et al.*, 2020).

É lenta a comunicação das pesquisas científicas (NOGUEIRA, 2022; FONSECA, 2020; JOHANSSON, *et al.*, 2018; POWELL, 2016). Estudos mostram que demora-se para publicar as produções da ciência (NOGUEIRA, 2022; FONSECA, 2020; POWELL, 2016), pois em um recorte de 30 anos, as publicações científicas chegam a demorar cerca de 100 dias para serem publicadas, e apenas uma parcela pequena de áreas do conhecimento consegue publicar entre 70 e 80 dias (POWELL, 2016). Além da demora na publicação, o cientista não deve se apropriar da própria descoberta, impedindo que ela seja disponibilizada para a academia e para a sociedade, pois as descobertas da ciência são resultados da colaboração social e por essa razão devem ser destinadas à comunidade e serem publicadas (MERTON, 1968, p.190 *apud* FONSECA, 2020, p. 4-5).

Economizar tempo, minimizar gastos, expandir a transparência e aprimorar a qualidade dos dados e dos serviços prestados pelos setores públicos fazem parte da publicidade e da eficiência. Seus fundamentos de responsabilidade social presumem a prestação de contas à população e a boa gestão dos recursos e serviços públicos, dentre outras razões, uma vez que são dados oriundos de interesses sociais e da legislação vigente. A publicação de maneira transparente e eficiente é basilar para uma sociedade informada como fundamentado no art. 37 da Constituição da

⁵ Disponível em: <https://www.gov.br/mec/pt-br>

República Federativa do Brasil de 1988 que estabelece a “[...] administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência [...]” (BRASIL, 1988).

Para fundamentar e contextualizar a necessidade de permitir à sociedade a publicação de maneira eficiente e transparente, conforme o art. 37 da CF/88, a **Lei de Acesso à Informação** nº 12.527 de 2011 regulamenta o direito de acesso à informação de todos os cidadãos brasileiros (BRASIL, 2011). Na direção contrária, em alguns aspectos, a **Lei de Proteção de Dados** n.º 67/98, de 26 de outubro (BRASIL, 1998), alterada pela Lei n.º 13.709, de 14 de agosto de 2018 (BRASIL, 2018a), tutela no sentido de aplicar normas que visem o sigilo quanto ao tratamento dos dados pessoais.

Conforme a Lei de Acesso à Informação, o acesso à informação é um direito adquirido pelos cidadãos brasileiros e existem mecanismos por intermédio da Web semântica, em conjunto com os conceitos de dados ligados, que permitem acessibilidade em diversos idiomas e modalidades, e também de qualquer região e comunidades diversas. Assim, a interoperabilidade proporciona vários benefícios à sociedade contemporânea no contexto da publicidade e eficiência. Organizações diversas, como a própria academia, necessitam aprimorar suas infraestruturas de suporte e reutilização de dados com a finalidade de se tornar uma ciência mais aberta. Assim, vencer problemas relacionados à encontrabilidade, acessibilidade, interoperabilidade e reusabilidade na recuperação de dados e informação proporciona vantagem considerável aos pesquisadores que buscam disseminar, ter credibilidade e reutilizar dados uns dos outros, conforme orientam as boas práticas, incluindo os princípios FAIR.

A organização e publicação de maneira interoperável, tanto de dados como de metadados conceituados, estruturados e notacionados que utilizem padrões de publicação comum entre eles (BOECKHOUT; ZIELHUIS; BREDENOODB, 2018) no contexto de dados ligados, estão em evidência no poder público, pois esses mecanismos são utilizados pelos governos de todo planeta como o principal instrumento de busca tanto por organizações internacionais como startups ágeis (HYLAND, ATEMEZING; VILLAZÓN, 2014).

Após os aspectos apresentados, é evidente a necessidade e utilidade de um

trabalho como o aqui descrito para apresentar alternativas às deficiências diversas encontradas na literatura. Esses elementos são de grande valia para a sociedade devido ao conjunto de boas práticas que são ideais para retribuir uma ciência mais aberta, tornando a acessibilidade aos dados das pesquisas mais amplos e buscando a reutilização delas.

1.5 ORGANIZAÇÃO DO TRABALHO

O presente trabalho está organizado em oito capítulos. A Introdução, que aborda a problemática, os objetivos e a justificativa. O segundo capítulo, Estado da Arte, apresenta elementos novos discutidos na literatura por intermédio de pesquisas sistematizadas relativas à interoperabilidade em dados ligados e as boas práticas de publicação na Web buscando o contexto de trabalhos de conclusão de curso. O terceiro capítulo, Fundamentação Teórica, aborda os fundamentos da OC, comunicação científica, ontologia, Web semântica e boas práticas na publicação de dados na Web. O quarto capítulo, Fundamentação Metodológica, trata dos fundamentos conceituais e metodológicos para preparar o leitor ao entendimento dos elementos apresentados nos próximos capítulos. O quinto capítulo, Procedimentos Metodológicos, apresenta os procedimentos utilizados na pesquisa. O sexto capítulo, Resultados e Discussão, mostra os resultados da pesquisa e os discute à luz da fundamentação teórica e metodológica. Por fim, há as considerações finais, referências e os apêndices.

2 ESTADO DA ARTE: BOAS PRÁTICAS DE PUBLICAÇÃO DE DADOS NA WEB NO CONTEXTO DE TRABALHOS DE CONCLUSÃO DE CURSO DE PÓS-GRADUAÇÃO

Na linha de padrões internacionais de transferência de dados na Web, Gonçalves e Jacyntho (2020) elaboraram um método almejando a publicação semântica com dados ligados⁶ em um estudo de caso para artigos acadêmicos. Eles aplicaram um método objetivando mapear e publicar uma base relacional existente, baseando-se nos fundamentos de dados ligados para os artigos acadêmicos da conferência interna da Semana de Integração Acadêmica de uma universidade pública federal brasileira. Embora o trabalho apresentado por Gonçalves e Jacyntho (2020) seja correlato ao presente projeto, a principal diferença é que, enquanto eles utilizaram um banco de dados relacional existente para mapear e publicar dados e informação, a presente pesquisa busca organizar, publicar e recuperar dados e informação por intermédio de um banco de dados orientado a grafos a partir de uma ontologia de domínio.

Outro trabalho correlato à presente pesquisa foram os estudos realizados por Iniakpokeikiye e Fawei (2022), que adaptaram uma ontologia para gerenciar de forma simples e eficiente dados universitários da Niger Delta⁷. Os autores aplicaram a tecnologia da Web semântica almejando a unificação do gerenciamento de dados universitários, usando SPARQL⁸ como linguagem de consulta, com a finalidade de testar a ontologia adotada. A principal diferença com relação a esta proposta é que os autores sugerem uma ontologia para entender apenas os dados relacionados aos alunos e a Instituição, excluindo as produções acadêmicas como Trabalhos de Conclusão de Curso (TCC)⁹ como teses, dissertações, dentre outros. Ainda, eles não

⁶ Dados ligados trata-se de uma coleção de conjuntos de dados inter-relacionados na Web. Ele compõe a parte central da Web Semântica que integra em larga escala e raciocina sobre dados na Web. Disponível em: <https://www.w3.org/standards/semanticweb/data>

⁷ Niger Delta é uma Universidade localizada na Ilha Wilberforce, no estado de Bayelsa, no sul da Nigéria. Disponível em: <https://www.ndu.edu.ng/>

⁸ SPARQL é o acrônimo de SPARQL *Protocol and RDF Query Language* (Protocolo SPARQL e linguagem de consulta RDF). Disponível em: <https://www.w3.org/TR/rdf-sparql-protocol/>

⁹ TCC é o acrônimo de Trabalho de Conclusão de Curso. Trata-se da realização de uma pesquisa monográfica ou artigo acadêmico a ser desenvolvido na disciplina de mesmo nome, após um projeto elaborado e aprovado na disciplina de Projeto de Pesquisa. Ele pode corresponder a uma monografia, um artigo científico ou outra produção acadêmica de acordo com o Regulamento do Trabalho de Conclusão de Curso. Disponível em: https://arquivologia.ufes.br/sites/arquivologia.ufes.br/files/field/anexo/regulamento_de_tcc.pdf

usam uma metodologia para construir ontologias e mapeiam os dados de um banco de dados relacional sem almejar buscas interoperáveis e integradas com outras bases. Esta dissertação difere-se, pois propõem-se um modelo conceitual visando compreender o âmbito de TCC de pós-graduações *stricto sensu*, utiliza-se de metodologia para construir ontologia e realiza-se o mapeamento para banco de dados orientado a grafos almejando alcançar a interoperabilidade das camadas estruturais e semântica.

Os padrões internacionais de transferência de dados na Web são uma boa prática, pois almejam a publicação e o mapeamento semântico com dados ligados em base de dados orientada a grafos. Para proporcionar tais mecanismos de maneira eficiente, surgiram as boas práticas de publicação de dados na Web que estão sendo aplicadas na atualidade. Pesquisas atuais apontam a importância do uso da Web semântica para organizar e publicar dados de maneira interoperável (INIAKPOKEIKIYE; FAWEI, 2022; SILVA; MUTZ; RUY, 2022a; SILVA; MUTZ; RUY, 2022b; ZHENG; TÖRMÄ; SEPPÄNEN, 2021; ASHOUR *et al.*, 2020, GOMES *et al.*, 2020; GONÇALVES; JACYNTHO, 2020; RAUTENBERG, 2019; ROCHA; RATHKE, 2019). Assim, levantou-se um estudo de pesquisas que adotam boas práticas de publicação de dados na Web e seus componentes, ressaltando a necessidade de interoperabilidade em dados ligados, bem como a relevância da produção desses estudos em busca das melhores práticas relacionadas à publicação e uso de dados na Web. Esse levantamento está exposto no Quadro 1 e uma análise comparativa dos resultados da presente pesquisa com alguns dos trabalhos citados neste capítulo é apresentada na subseção 6.8, no contexto das discussões sobre os resultados.

Quadro 1 - Levantamento bibliográfico do Estado da Arte

Autor e Base	Título	Escopo	Estratégias tecnológicas utilizadas
Andrade, Nardi e Ruy (2023): Univali.	Integração de dados de publicações científicas usando uma abordagem baseada em ontologias.	Apresenta a iniciativa de integração de dados de publicações científicas baseada em ontologia para três bases em um repositório compartilhado.	Ontologia, OntoUML, UFO, OWL, integração semântica, mapeamento semântico e RDF.
Iniakpokeikiye e Fawei (2022): <i>Asian Journal of Research in Computer Science.</i>	Uma ontologia adaptável para gerenciamento fácil e eficiente de dados universitários na <i>Niger Delta University</i> .	Adaptação de uma ontologia para gerenciar eficientemente os dados universitários na <i>Niger Delta University</i> .	Web semântica, ontologia, OWL, OWL2, dados ligados, RDF, SPARQL e <i>knowledge graph</i> ¹⁰ .
Silva, Mutz e Ruy (2022a): Ontobras.	Uso de ontologias no suporte a aplicação de <i>machine learning</i> : um caso no domínio de evasão escolar.	Apresenta um caso no campo da evasão escolar combinando a utilização de ontologias com o suporte a aplicação de <i>machine learning</i> , buscando prever a evasão de alunos em diferentes instituições da educação	Ontologia, vocabulário comum, vocabulário controlado ¹¹ , metadados, OntoUML, SABIO, UFO, interoperabilidade semântica e integração semântica.

¹⁰ *Knowledge graph* tem diversas definições (abordadas na subseção 3.3.3) e sua tradução para o português refere-se a grafo de conhecimento. É conhecido também como rede semântica e retrata uma rede de elementos do mundo real como objetos, eventos, situações ou conceitos, representando ainda a relação entre eles. Disponível em: <https://www.ibm.com/topics/knowledge-graph#:~:text=A%20knowledge%20graph%2C%20also%20known,the%20term%20knowledge%20%E2%80%9Cgraph.%E2%80%9D>

¹¹ Vocabulário controlado refere-se à combinação organizada de frases e palavras usadas para indexar conteúdos e recuperá-los (discutido na subseção 3.3.2) por intermédio pesquisa ou navegação (HARPRING et al., 2006). Disponível em: https://edisciplinas.usp.br/pluginfile.php/5506147/mod_resource/content/1/Cataloging_Cultural_Objects.pdf

		superior.		
Silva; Mutz e Ruy (2022b): Ontobras.	Abordagem para análise de múltiplas fontes de dados de evasão escolar.	Apresenta como a combinação <i>machine learning</i> e ontologia podem se tornar recursos para prever a evasão escolar em diferentes instituições da educação superior.	Ontologia, vocabulário comum, vocabulário controlado, metadados, OntoUML, SABio, UFO, interoperabilidade semântica e integração semântica.	
Zheng, Törmä e Seppänen (2021): <i>Science Direct</i> .	Um conjunto de ontologia compartilhada para <i>workflow</i> de construção digital.	Suíte de ontologia compartilhada para <i>workflow</i> de construção digital.	Web semântica, ontologia, OWL, dados ligados, dados ligados e abertos, RDF, vocabulários, SPARQL e GraphDB.	
Hamim, Benabbou e Sael (2021): <i>The Science and Information Organization</i> .	Um sistema de apoio à decisão baseada em ontologia por meio de tarefas de predição multi-objetivas.	Sugere uma ontologia para representar perfil de alunos e criar um sistema de apoio à decisão por meio de tarefas de predição multi-objetivas.	Ontologia, OWL, RDF, metadado, Protégé e interoperabilidade semântica.	
Silveira, Aragón, Cury e Menezes (2021): SBC-OpenLib (SOL).	Uma ontologia de referência para arquiteturas pedagógicas.	Apresenta uma ontologia de referência para arquiteturas pedagógicas.	Ontologia, SABio, UFO, OWL, OntoUML,	

Ashour <i>et al.</i> (2020): Ieeexplore.	Um modelo ontológico para representação de cursos e perfis acadêmicos: um estudo de caso da <i>King Abdulaziz University</i> .	Um modelo ontológico de representação de cursos e perfis acadêmicos da <i>Universidade King Abdulaziz</i> objetivando mapear professores acadêmicos em novos cursos para dar aula.	Web semântica, ontologia, dados ligados, RDF, Protégé e SPARQL.
Gomes <i>et al.</i> (2020): Em Questão.	Proposta de uma ferramenta para classificação arquivística com base em ontologias.	Ferramenta para classificação arquivística com base em ontologias.	Web semântica, ontologia, SPARQL, OWL2, Python e RDF.
Gonçalves e Jacyntho (2020): SciELO.	Um método para publicação semântica <i>linked data</i> de bases de dados convencionais e um estudo de caso real de artigos acadêmicos.	Um método para publicação semântica de dados ligados com bases de dados convencionais em um estudo de caso de artigos acadêmicos.	Web semântica, metadados, ontologia, vocabulário, OWL, SKOS, dados ligados, dados ligados e abertos, RDF.
Rautenberg (2019): Em Questão.	Web semântica e bibliografia: um estudo para a publicação de registros bibliográficos como dados abertos conectados.	Uso de Web semântica para a publicação de registros bibliográficos como dados ligados e abertos.	Web semântica, metadados, ontologia, dados ligados e abertos, vocabulários ligados e abertos.
Rocha e Rathke (2019): Brajils.	Sistema de informação de pesquisa: uso da ontologia	Sistema de informação de pesquisa para o uso de ontologia no contexto	Web semântica, ontologia, ontologia VIVO, plataforma VIVO, dados

	de VIVO no contexto das instituições brasileiras.	das instituições brasileiras.	ligados e abertos.
Silveira, Cury e Menezes (2019): SBC-OpenLib (SOL).	Superando fronteiras da educação com ecossistemas de aprendizagem.	Discute como os Ecossistemas de Aprendizagem podem ajudar a superar ou atenuar as fronteiras da educação a partir de uma ontologia desenvolvida com a metodologia SABIO.	Ontologia, SABIO e OntoUML.
Carchedi, <i>et al.</i> (2018): Repositório Institucional UFC.	Onto4LA: uma ontologia para integração de dados educacionais.	Estabelece <i>Ontology For Learning Analytics</i> (Onto4LA) com a finalidade de facilitar a integração de dados educacionais e proporcionar interoperabilidade semântica.	Ontologia, metadados, SKOS e interoperabilidade semântica de dados e informação.
Obeid <i>et al.</i> (2018): ACM <i>Digital Library</i> .	Sistema de recomendação baseado em ontologia no ensino superior.	Sugere um sistema baseado em ontologia para orientar estudantes do ensino médio a escolherem cursos em Instituições de ensino superior.	Web semântica, Ontologia, Protégé.
Rozsa, Dutra, Moisés e Nhacuongue (2017b): Brajis.	O paradigma tecnológico da Internet das Coisas e sua relação com a Ciência da Informação.	Dados ligados e abertos no contexto acadêmico para identificar e analisar os vocabulários utilizados na academia e na pesquisa científica.	Ontologia, metadados, dados ligados, dados ligados e abertos, vocabulários abertos, vocabulários ligados e abertos e RDF.
Segundo,	Conceitos e tecnologias da	Conceitos e tecnologias da Web	Web semântica, ontologia, ontologia

Coneglian e Lucas (2017): SciELO.	Web semântica no contexto da colaboração acadêmico-científica: um estudo da plataforma Vivo.	semântica que colaboram como âmbito acadêmico- científico para a construção de redes semânticas colaborativas e para a organização das informações.	VIVO, plataforma VIVO, dados ligados, RDF, prova de conceito.
Subirats <i>et al.</i> (2017): EBSCOhost.	Alinhamento de metadados e vocabulários controlados no ecossistema persistente internacional PHAIDRA.	Propõe-se a implementação de uma ferramenta open source usando VocBench, objetivando apoiar o desenvolvimento colaborativo e alinhamento de conceito multilíngue de vocabulários controlados no ecossistema persistente internacional PHAIDRA.	Web semântica, ontologia, metadados, dados ligados, dados ligados e abertos, vocabulários controlados, RDF, SKOS e VocBench.
Coneglian, Sousa e Segundo (2016): Brajís.	Análise de publicações da temática “Web semântica” na América Latina: um olhar nas bases de dados da <i>Web of Science</i> .	Analisa a produção/ publicação científica na América Latina do tema Web semântica, observando como uma região está frente a outras, e as características que as produções de tal região apresentam.	Web semântica, ontologia, RDF, OWL, SPARQL e vocabulários.
Ambinder e Marcondes (2013): SciELO.	Leitura de artigos científicos digitais na Web.	Apresentação, acesso e leitura de artigos científicos digitais na Web.	Web semântica, metadados e ontologia.

<p>Ameen <i>et al.</i> (2012): Ieeexplore.</p>	<p>Criação de ontologia no domínio da educação.</p>	<p>Elaboração de ontologia no domínio da educação, contendo detalhes sobre vários cursos oferecidos pela universidade.</p>	<p>Web semântica, ontologia e Protégé.</p>
<p>Marcondes (2011): Liinc em Revista.</p>	<p>Um modelo semântico de publicações eletrônicas.</p>	<p>Modelo de conteúdo como uma ontologia computacional. Trata-se de um protótipo de interface Web de submissão de artigos a um sistema eletrônico de publicação de periódicos que implementa o modelo. Busca-se extrair e representar o conteúdo de artigos científicos biomédicos em formato “inteligível” por programas.</p>	<p>Web semântica, RDF, OWL e termos controlados do UMLS <i>Thesaurus</i>.</p>

Fonte: Elaborado pelo próprio autor

3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda de maneira concisa a contextualização da Organização do Conhecimento (OC) e da Recuperação da Informação (RI) no âmbito da Ciência da Informação (CI), seus componentes operacionais, dando ênfase a Web semântica, ontologias, interoperabilidade, e boas práticas na publicação de dados na Web.

3.1 ORGANIZAÇÃO DO CONHECIMENTO NA CIÊNCIA DA INFORMAÇÃO

A Organização do Conhecimento (OC) (*Knowledge Organization (KO)*), teve seu desenvolvimento na História como disciplina de pesquisa, ensino, acadêmica e profissional na Biblioteconomia e na Ciência da Informação (CI), entre outras áreas. Com o advento e propagação da Informática, o homem gerou conhecimentos especializados, e com isso originou grandes bancos de dados e informação (VALENTIM, 2010), “capazes de fazer automaticamente algumas conexões pertinentes entre as representações, mais ou menos como se compreendesse seu sentido” (LÉVY, 1993, p.108), mas que tem suas limitações, pela própria estrutura que os sistemas computadorizados apresentam.

Representantes da Inteligência Artificial e produtores de Sistemas Especialistas, profissionais em Hiperfídia e especialistas no ramo da Educação estão se interessando em aplicar metodologias de OC, pois o conhecimento por si só também tem as suas limitações, pois ele não pode ser apreendido ou representado a não ser que seja exibido por unidades de conhecimento e suas diversas combinações possíveis estejam dispostas em palavras, termos ou declarações (DAHLBERG, 1993).

Apesar das limitações mencionadas, os bancos de dados e informação fornecem base importante para aqueles que o acessam quando estruturados de maneira adequada e, conseqüentemente, apoiam na formação do conhecimento daqueles que os utilizam. Desse modo, é importante gerenciar a informação digital de maneira eficiente por intermédio de sistemas de informação corporativos, com a finalidade de proporcionar acesso e recuperação consistente de dados e informação, e que sua utilização seja transformada tanto em conhecimento como em vantagem competitiva (VALENTIM, 2010).

A seção seguinte discute de forma resumida o que se entende sobre OC para o interesse da presente pesquisa.

3.1.1 Organização do Conhecimento

Ao apresentar a Ciência da Informação (CI) em outros contextos, González (2000), Souza, Stumpf (2009) e Marteleto (2009 *apud* ARAÚJO 2018) esclarecem que tanto no Brasil como em Portugal, houve um processo maior de autonomização da CI norteada por temáticas como gestão da informação, transferência da informação e, inclusive, Organização do Conhecimento (OC) decorrente de perspectivas dialogadas nos Estados Unidos e na Europa.

Do ponto de vista histórico, os resultados das pesquisas de Araújo (2018) estabelecem que a OC surgiu da representação da informação no final da década de 1970. A propagação da influência das teorias cognitivistas na CI se deu no campo da representação, porque houve desenvolvimento na elaboração de linguagens de representação e sistemas de informação pensados para os usuários, ou para as estratégias cognitivas deles (VAKKARI, 1999 *apud* ARAÚJO, 2018). Isso complementa a noção de que em toda atividade de organizar, classificar, indexar e representar informação é indispensável considerar o escopo dos documentos existentes, das fontes informacionais disponíveis, bem como a área dos conhecimentos nos ramos aos quais pertencem as fontes mencionadas. Essas contribuições advêm da Teoria do Conceito de Dahlberg (1978) na elaboração de tesouros e outros recursos de linguagem documentária (SOERGEL, 1974 *apud* ARAÚJO, 2018) e das execuções dos fundamentos da Teoria da Classificação Facetada proposta por Ranganathan (2009) em sistemas automatizados de RI (FUGMANN, 1993, ARAÚJO, 2018). Essencial para o estabelecimento desta subárea foi a origem da *International Society for Knowledge Organization* (ISKO¹²), pois reuniu pesquisadores dedicados ao tema e começou a organizar eventos e possibilitar publicações. Desde então, oficializou-se a área de OC (ARAÚJO, 2018, p.39).

Para Toutain (2007), a OC é uma área central tanto de ensino quanto de pesquisa em CI e Biblioteconomia. Isso ocorre porque os processos de produção, tratamento e disseminação de informação passaram por mudanças significativas na trajetória da comunicação do conhecimento. O autor ainda argumenta que pode-se identificar pelo menos quatro fases importantes que impulsionaram a criação de

¹² ISKO é o acrônimo de *International Society for Knowledge Organization*, e sua tradução do Inglês para o Português significa Sociedade Internacional para a Organização do Conhecimento. É a principal sociedade científica da área de Organização do Conhecimento. Disponível em: <https://isko.org.br/isko-internacional/>

instrumentos de classificação, e o desenvolvimento de métodos e técnicas de indexação relacionados a recuperação de documentos e informação no contexto de bibliotecas, e outros sistemas e redes de informação: (1) “Caos” Documentário, (2) “Explosão” da Informação, (3) “Avalanche” de Conhecimento, e (4) “Revolução” Tecnológica.

Nessa perspectiva, Hjørland (2008) estabelece que OC, no sentido restrito do termo, diz respeito às atividades conhecidas como descrição, indexação e classificação de documentos em bibliotecas, bases de dados bibliográficas, arquivos e outras classificações de instituições de memória por bibliotecários, arquivistas, especialistas em informação, especialistas no assunto, por algoritmos de computador e mesmo por leigos. O campo de estudo da OC está preocupado com a natureza e a qualidade dos processos de OC, assim como os Sistemas de Organização do Conhecimento (SOC) são usados para organizar documentos, documentar representações, obras e conceitos. Hjørland defende que a OC é um campo de pesquisa, ensino e prática, essencialmente vinculado à Biblioteconomia e a CI, isto é, a OC está especialmente institucionalizada em cátedras, universidades ao redor do mundo, programas de ensino e pesquisa, instituições de pesquisa, escolas de ensino superior, revistas acadêmicas (por exemplo, a *Knowledge Organization* de 1993), conferências nacionais e internacionais, organizações nacionais e internacionais, como por exemplo, a *International Society for Knowledge Organization* (ISKO) (HJØRLAND, 2016).

Em contrapartida, Brascher e Café (2008) discordam dos argumentos anteriores, pois alegam que Hjørland (2008) estabelece o processo de OC usado no sentido restrito em CI, que compreende a elaboração de resumos, a catalogação, a classificação, a indexação, o estabelecimento de elos, etc. Mas que na visão dos autores, os processos mencionados aplicam-se apenas a objetos físicos, pois eles acreditam e defendem que ao se tratar de objetos informacionais, esses são processos de organização da informação e não do conhecimento.

Os argumentos de Hjørland (2003) convergem com aqueles apresentados por Jaenecke (1994 *apud* RABELLO, 2017) sobre a OC, sendo que suas alegações se complementam na seguinte contextualização:

A organização do conhecimento é um campo de investigação no âmbito da ciência da informação (HJØRLAND, 2003) que busca “ordenar e suprir o conhecimento” (JAENECKE, 1994) previamente

selecionado, trazendo subsídios teóricos e metodológicos para a elaboração de produtos documentais mediante a produção e a utilização de linguagens documentais, ou seja, constitui um trabalho intelectual que tem a organização do documento e da informação nele contida (conteúdo) como objeto de estudo. Portanto, o campo objetiva a produção e a utilização da metainformação para fins de organização, visando recuperação, acesso e uso da informação num sistema de armazenamento e recuperação (RABELLO, 2017, p. 147).

A OC corresponde ao campo de investigação na CI e almeja estruturar e complementar o conhecimento. Na concepção de Dahlberg (2006 *apud* CARLAN, 2010), a OC é definida como a ciência que ordena os conceitos de maneira estruturada e sistematizada mediante seus atributos, e que podem ser estabelecidas como elementos externados da herança de objeto, sendo que a aplicação dos conceitos e classes desses conceitos são ordenados por valores, conteúdos dos objetos ou assuntos. Por intermédio dessa OC é que as ferramentas são criadas e apresentam uma interpretação baseada na organização e estrutura do objeto, e que são chamados de SOC.

Dahlberg (1993) estabelece que o item mais essencial da fundamentação teórica da OC está no fato de que qualquer OC deve ser baseada em unidades de conhecimento. Ela explica que a OC como domínio específico de bibliotecários e da CI pode-se tornar a metodologia necessária para os seguintes três grupos principais de usuários: (1) alunos que estejam iniciando seus estudos universitários, (2) professores, em especial, para ensinar aos alunos da Educação, e (3) todos aqueles que ajudam os líderes políticos, industriais e sociais. A autora acredita que as informações precisam ser organizadas e processadas com certa prioridade para se tornarem conhecimento disponível, e não de forma apenas pessoal, mas também interpessoal, objetiva e pública. Ela esclarece que para o conhecimento se tornar disponível, muito precisa ser feito até que alguns desses resultados se tornem visíveis.

Em conformidade com Dahlberg (2006), Hjørland (2008) argumenta que a OC pode ser realizada de diversas maneiras, entretanto, todas elas têm em comum a tipificação de um domínio que pode ser atingido por meio de um *software* fundamentado em um algoritmo. Nesse sentido, Coelho, Lima e Borges (2012) explicam que uma das convergências em que a OC pode ser realizada está na era pré-Inteligência Artificial relacionada a interpretação humano-cognitiva e integrada ao contexto social, mesmo quando realizada por meio de um computador, e que o produto desse processo acarreta a origem do SOC, tratado na próxima subseção.

3.1.2 Sistemas de Organização do Conhecimento

O termo Sistemas de Organização do Conhecimento (SOC) (*Knowledge Organization Systems* (KOS)) foi proposto pela primeira vez no ano de 1998, em Pittsburgh, *Pennsylvania* pelo *Networked Knowledge Organization Systems Group*, na primeira Conferência da ACM¹³ *Digital Libraries* (BRASCHER; CAFÉ, 2008). Para Hjørland (2021), SOC abrange qualquer tipo de esquema de organização da informação, bem como a promoção do ato ou o efeito de gerir o conhecimento. Ele contém classificação de esquemas e categorias que organizam materiais em um nível amplo, mas também pode ser entendido como mecanismos para organizar a informação, ou seja, é uma espécie de aparato central na organização de biblioteca, museu e arquivo. O autor argumenta que o SOC corresponde a seleção de conceitos com indicação de relações semânticas escolhidas, explicando que seus exemplos são os sistemas de classificação, listas de cabeçalhos de assuntos, tesouros, ontologias e demais sistemas de metadados (HJØRLAND, 2016).

Na concepção de Zeng e Chan (2004), os SOC correspondem às ferramentas de interpretação de estruturas do conhecimento. Esses autores acreditam e defendem que o conhecimento é inseparável do indivíduo, mas que ele pode ser exteriorizado do organismo humano mediante um sistema estruturado que possua relações semânticas, e que permita também a recuperação e o armazenamento da informação, bem como possibilita a criação de um novo conhecimento em um ciclo contínuo.

A expansão da influência provocada pelas teorias cognitivistas na CI colaborou com o campo da representação da informação. Na medida em que houve desenvolvimento de linguagens de representação e sistemas de informação pensados para os usuários ou para as estratégias cognitivas deles, completou-se a ideia de organizar, classificar e indexar informação, e conseqüentemente representá-la. Pode-se destacar as contribuições da teoria do conceito no desenvolvimento de tesouros e instrumentos de linguagem documentária, bem como das aplicações e dos fundamentos da teoria da classificação facetada criados por Ranganathan (2009) em sistemas automatizados de RI (VAKKARI, 1999; DAHLBERG, 1978; SOERGEL, 1974; RANGANATHAN, 2009 *apud* ARAÚJO, 2018).

¹³ ACM é o acrônimo de *Association for Computing Machinery*, e sua tradução do Inglês para o Português significa Associação de Máquinas de Computação. É uma sociedade internacional de computação científica e educacional. Disponível em: <https://dl.acm.org/>

De maneira complementar, a informação representada e recuperada nesse contexto está amparada nos fundamentos definidos por Le Coadic (1994), por argumentar e defender a informação como um conhecimento gravado sob a forma impressa ou numérica, oral ou mesmo audiovisual. O autor defende que a informação comporta um componente de sentido, ou seja, é um significado emitido a ser consciente através de uma mensagem inscrita por meio de um aparato espacial-temporal, seja ele de maneira impressa, sinal elétrico, onda sonora, etc.

Para possibilitar tais representações e o gerenciamento dos dados, é importante que ambos estejam em consonância com certos princípios para a RI, que será tratada na próxima subseção.

3.1.3 Representação da Informação

Na CI a organização e representação da informação e do conhecimento tem a ênfase dada aos processos de organização, usando-se de representações de informações e de conhecimentos, e os componentes para serem compreendidos, precisam conceitos antecedentes (LIMA; ALVARES, 2012).

A representação da informação presume o uso de uma linguagem específica para proporcionar a comunicação entre documentos e usuários dos sistemas de informação (ALBUQUERQUE, 2013). Ela tem como objetivo analisar a modelagem conceitual em diferentes tipos de Sistema de Organização do Conhecimento (SOC), investigando a representação das estruturas conceituais e as relações semânticas que são definidas, pretendendo reconhecer elementos comparativos entre eles (ALMEIDA, 2017).

Sobre a recuperação da informação, transferência de informações e a representação da informação no âmbito da CI, tratam-se de abordagens em que a importância está na organização do conhecimento (ALBUQUERQUE, 2013; NOVELLINO, 1996) com o objetivo de produzir instrumentos e métodos (estratégias) para representar informações nas áreas abrangentes do assunto (ALBUQUERQUE, 2013). Nesse sentido, Novellino (1996) explica que a característica mais importante do processo de representação da informação está na substituição de uma entidade linguística longa e complexa para o texto do documento com uma descrição abreviada (NOVELLINO, 1996).

3.1.4 Metadados

Metadados são dados sobre dados (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2023; GILLILAND, 2016; LAUFER, 2015). É um termo que se refere a quaisquer dados utilizados com a finalidade de apoiar na identificação, descrição e localização de recursos eletrônicos disponibilizados em rede (INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS, 2023), e são importantes para fornecer informações adicionais sobre os dados que são publicados como o seu conteúdo e a sua estrutura.

Como exemplo de metadados temos o nome da coluna de uma tabela, o nome da coluna colocada na primeira linha de um arquivo no formato CSV (LAUFER, 2015), e eles podem ser aplicados para auxiliar nas tarefas como descoberta, reutilização do conjunto de dados, granularidade que é a capacidade de dividir algo em partes menores e promover formas de acessar uma única propriedade de um recurso (de uma coluna de uma tabela) (LAUFER, 2015). Para Gilliland (2016) os metadados são classificados administrativo, descritivo, preservação, técnico e uso, e cada um deles é correspondente ao seu objetivo ou utilização:

- **Metadados administrativos:** são aqueles utilizados para gerenciar e administrar grupos ou recursos informacionais.
- **Metadados descritivos:** usados para diferenciar, validar e caracterizar grupos ou recursos informacionais.
- **Metadados de preservação:** trata-se da aplicação para preservação de grupos ou recursos informacionais.
- **Metadados técnicos:** são aqueles que correspondem a maneira de um sistema funcionar ou se comportar.
- **Metadados de uso:** são relativos ao grau e tipo de uso de grupos ou mesmo recursos informacionais.

É importante a classificação dos metadados como instrumento de busca em meios eletrônicos disponibilizados em rede, pois eles fornecem informações adicionais sobre os dados para ajudar desenvolvedores de aplicações e usuários finais a entender melhor o significado dos dados (LAUFER, 2015; LÓSCIO, BURLE E CALEGARI, 2017). Gilliland (2016) argumenta que uma estrutura de metadados

mínimos pode-se considerar um padrão de estrutura de dados. E nesse sentido, ele corresponde a um conjunto básico de categorias utilizadas para representar informações sobre documentos (SILVA; LARA, 2023), e auxiliam a esclarecer questões quanto a licença de uso, empresa (organização) que gerou dados, a qualidade dos dados, a frequência de atualização de um conjunto de informações, a proveniência, como realizar o acesso, dentre outros (LAUFER, 2015; LÓSCIO, BURLE E CALEGARI, 2017). Além disso, como os metadados correspondem aos dados estruturados que proporcionam classificar, especificar e gerenciar documentos (BRASIL, 2020a), eles ajudam no processo de comunicação entre publicadores e consumidores de dados (LAUFER, 2015).

3.1.5 Recuperação da Informação

A informação no âmbito da CI pode ser organizada e recuperada em áreas diversas como produção e comunicação científicas, nos estudos sobre os sujeitos, na gestão da informação, em âmbitos como economia política da informação, nas áreas de estudos métricos da informação, memória, patrimônio e documento (ARAÚJO; VALENTIM, 2019). Todavia, as maneiras de organizar e recuperar dados e informações que estão disponíveis tanto no mercado como na comunidade acadêmica e que são aplicados no ramo da CI, possuem diversas limitações quanto a publicação e recuperação em bases de dados relacionais e não orientadas a grafos (GONÇALVES; JACYBTHO, 2020), e principalmente a escassez de RI em pesquisas acadêmicas mais completas, ou seja, realizar RI Trabalhos de Conclusão de Curso (TCCs) como dissertações e teses de maneira interoperável e integrada com bases externas. Tanto que, sobre a *Ocean Best Practices System (OBPS)*¹⁴, uma equipe internacional de cientistas defendeu que a comunidade pesquisadora necessita de uma linguagem comum para que os seus resultados possam ser utilizados e reutilizados continuamente e com o mínimo possível de imprevisibilidade (REVEZ; SILVA, 2021).

De modo geral, a RI no domínio da CI é focada em proporcionar facilidade ao usuário no acesso às informações de seu interesse. Essa área contempla aplicações

¹⁴ OBPS corresponde a *Ocean Best Practices System* (Sistema de Melhores Práticas Oceânicas (SMPO)). É um repositório do *International Oceanographic Data and Information Exchange (IODE)*, que dispõe descobertas da Web semântica de última geração e indexação de metadados atendendo às melhores práticas. Disponível em: <https://www.oceanbestpractices.org/about/>

quanto a representação, armazenamento, organização e acesso aos itens de informação tais como documentos, páginas da Web, catálogos online, objetos multimídia, registros estruturados e semiestruturados (BAEZA *et al.*, 1999).

A área de RI tem se propagado mais que seus objetivos iniciais de indexar texto e pesquisar documentos em uma coleção. É importante destacar que os seus estudos incluem modelagem, pesquisa na Web, classificação de texto, arquitetura de sistemas, interfaces de usuário, dados de visualização, filtragem e linguagens. Quanto à pesquisa, a área pode ser estudada em duas perspectivas complementares: uma centrada no computador e outra centrada no ser humano. Sobre a perspectiva centrada no computador, RI corresponde a investigação de construção de índices eficientes, processamento de consultas do usuário em alta performance e implementação de algoritmos de classificação para alcançar resultados melhores. Já na perspectiva centrada no ser humano, a RI consiste particularmente em estudar o comportamento do usuário com a finalidade de compreender suas necessidades primordiais, e conseqüentemente apontar como essa compreensão afeta tanto a organização como a operação do sistema de recuperação (BAEZA *et al.*, 1999).

Assim como a RI relacionada a pesquisa pode ser estudada sob a perspectiva centrada no computador, a informação tornou-se conceito recorrente na obra de diversos pesquisadores relacionados com as tecnologias computacionais, tais como Claude Shannon, Bertrand Russell, Alan Turing, Kurt Gödel e Richard Dawkins (GLEICK, 2013). Em 1953, Mooers propôs a criação de uma disciplina científica para desenvolver a ambição de Bush¹⁵: a RI, devido à "explosão" informacional que foi resultante do aumento da quantidade de documentos e da dificuldade em recuperar informação diante disso (ARAÚJO, 2018). Os primeiros experimentos científicos decorrentes dessa proposta são os *Cranfield Projects*¹⁶ I e II desenvolvidos em 1957 e 1963. Nos anos seguintes, a proposta de recuperação automatizada da informação foi encampada dentro do projeto da Ciência da Informação, chegando mesmo a ser entendida como aquilo que deveria ser o núcleo da área por diferentes autores, entre

¹⁵ Vannevar Bush foi um respeitado cientista do Instituto de Tecnologia de Massachusetts (MIT) e chefe do esforço científico dos EUA na Segunda Guerra Mundial (SILVA; FREIRE, 2012). Trabalhou para resolver problemas de organização da informação e RI (SILVA; FREIRE, 2012; ARAÚJO, 2014; ARAÚJO, 2018).

¹⁶ *Cranfield Projects* diz respeito aos projetos dos primeiros experimentos científicos empíricos entre 1957 e 1963, quando a CI se tornou uma disciplina específica e delimitada em função da RI (CAPURRO, R.; HJØRLAND, 2003; SILVA; FREIRE, 2012; ARAÚJO, 2014; ARAÚJO, 2018).

os quais Tefko Saracevic, em seu livro *Introduction to Information Science* (ARAÚJO, 2018).

O ser humano organiza informações para posteriormente recuperá-las e pesquisá-las por mais de 5.000 anos. Conforme a biblioteca ganhou notoriedade tanto no ambiente físico como no ambiente digital (ARAÚJO, 2018), suas áreas de representação e recuperação da informação são os campos da CI que se relacionam fortemente com a biblioteconomia (ARAÚJO, 2014). Essas relações se deram com a RI no campo dos sistemas de informação, ainda na década de 1960, na busca por encontrar a melhor maneira de representar a informação em termos formais e de conteúdo, e também na sua recuperação otimizada. Com relação a períodos anteriores, os esforços de pesquisas em RI foram iniciados na década de 50 por Calvin Moores, Eugene Garfield, Hans Peter Luhn e Philip Bagley, sendo que o primeiro deles supostamente deu nome ao termo RI (BAEZA *et al.*, 1999). Esses estudos colaboraram com o surgimento de diferentes instrumentos de linguagem controlada e de sistemas de classificação, todos com o objetivo de economizar custos, reduzir ruídos, diminuir redundância e aplicar princípios lógicos. Nesse contexto, os sistemas e linguagens criados se desenvolviam na busca da criação pela melhor linguagem e melhor representação, sendo estes os objetivos iniciais da RI na CI (ARAÚJO, 2014).

Como o volume de informações nas bibliotecas esteve crescendo continuamente, é necessário elaborar estruturas de dados especializadas para pesquisa performática, como por exemplo, os índices. De uma forma ou outra, os índices estão no centro dos sistemas modernos de RI. Eles proporcionam um acesso veloz aos dados e permite melhorar a performance do processamento da consulta. Durante séculos, os índices foram feitos de maneira manual como conjuntos de categorias. Cada categoria no índice é constituída de rótulos que identificam seus tópicos relacionados e de indicações para os documentos que trabalham esses tópicos. Embora esses índices sejam comumente elaborados por pesquisadores de Biblioteconomia e CI, a chegada dos computadores modernos tem permitido a construção automática de índices grandes, o que tornou mais rápido o desenvolvimento da área de RI (BAEZA *et al.*, 1999).

Atualmente, houve migração da RI na CI para a Ciência da Computação com o desenvolvimento de sistemas usados nos buscadores conhecidos como, por exemplo, o Google (HJØRLAND, 2021). Portanto, o processo de RI nesses sistemas de busca

conta com fundamentos organizacionais definidos em OC. A RI e a OC (mencionado na seção 3.1.1) são dois âmbitos de pesquisa de estudos específicos, entretanto, ambos compartilham do mesmo objetivo que está em facilitar a localização de documentos, informações e conhecimentos (HJØRLAND, 2021). A relação é restrita entre OC e RI, pois, RI dedica-se aos processos de busca e OC trata-se de planejar estruturas ótimas de RI (dentre outros propósitos) (ANDERSON; JOSÉ, 2005 *apud* HJØRLAND, 2021). Conforme a OC é uma área de estudos fundamental da CI na RI, logo, organizar adequadamente elementos como documentos, informações e conhecimentos favorece o processo de recuperá-los. E para organizar devidamente esses elementos, há um conjunto de princípios orientadores e práticos para isso.

3.1.6 Comunicação Científica

A comunicação científica é uma etapa imprescindível e inseparável do fazer ciência, porque sem ela não existe comunicação, elucidação e discussão quanto às descobertas da própria ciência (ERDMANN, 2011). A comunicação é importante para avanço científico, pois é por intermédio dela que se possibilita a troca de informações e ideias entre os pesquisadores para a retroalimentação do processo científico (DROESCHER; SILVA, 2014). Várias finalidades da comunicação científica foram sintetizadas por Figueiredo (1992) como: estimular o pensamento e a ação, inserir ou interagir com ideias de outras pessoas, conhecer, experimentar e realizar; promoção do conhecimento com o que outras pessoas estão fazendo, buscar conhecer o desenvolvimentos dos próprios campos especializados e em outras áreas; minimizar a possibilidade da redundância dos trabalhos, economizando esforço e tempo; incorporar conhecimento basilar para pesquisas em outras áreas; e captar informações suficientes para a execução de produções científicas.

O domínio das descobertas científicas está na comunicação dos seus resultados (FONSECA, 2020). O autor argumenta que o crescimento da ciência acontece pela cooperação entre os pesquisadores e a sociedade. As questões dos estudos de Ciência, Tecnologia e Sociedade, devem demonstrar mais que a cooperação entre os pesquisadores, a produção científica precisa levar em conta a demanda social (FONSECA, 2020).

3.2 ONTOLOGIA

A utilização de ontologias para representar o conhecimento de um domínio aumentou na Computação, principalmente com o advento da Web semântica. Todavia, o foco tem sido na eficiência computacional, o que pode causar prejuízo na qualidade da representação, não correspondendo a certos aspectos, tais como os temporais, que são características importantes do domínio (ZAMBORLINI, 2011).

A palavra ontologia é de origem latina e seu significado corresponde ao “estudo da existência” (GUIZZARDI, 2005). A etimologia do prefixo “ont” vem do particípio do presente do verbo grego “einaí” que quer dizer “ser”, sendo que “logia” em latim significa “estudo” (GUIZZARDI, 2005). O termo foi usado pela primeira vez no séc. XVII no âmbito da Filosofia, e é proveniente dos estudos de Aristóteles no século IV a.C. como “a ciência do ser enquanto ser” (GUIZZARDI, 2005, p.52). Depois disso, o termo foi utilizado para denominar uma disciplina da Filosofia, e quando utiliza-se a palavra “Ontologia” com “O” maiúsculo, trata-se da intitulação de um sistema de categorias autônomas de linguagem adequada para conceituar teorias científicas (GUIZZARDI, 2005; GUARINO; GIARETTA, 1995). O termo “ontologia” com “o” minúsculo diz respeito a um artefato concreto no nível simbólico a ser usado para um determinado fim (GUARINO; GIARETTA, 1995), e pode ter dois sentidos: uma teoria lógica que fornece uma explicação parcial e explícita de uma conceituação (sentido 1); ou trata-se do sinônimo de conceituação (sentido 2) (GUARINO; GIARETTA, 1995).

Há diferentes definições para o termo ontologia (GRUBER, 1994; GUARINO; GIARETTA, 1995), sendo que um dos mais divulgados é o que define ontologia como uma especificação explícita de uma conceituação (GRUBER, 1994). Terminologicamente, ela é definida como (1) uma área da metafísica que estuda a natureza e as relações do ser (lida com entidades abstratas) e (2) uma teoria específica sobre a natureza do ser ou os tipos de coisas que possuem existência (MERRIAM-WEBSTER DICTIONARY, 2022). É um conceito pesquisado em diversos ramos de estudo tais como Filosofia, Ciência da Computação e CI (ALMEIDA, 2013; FARINELLI, ALMEIDA, 2019), e também, em diversos domínios do conhecimento, tais como Medicina, Biologia, Direito e Geografia. Independente da propagação do termo, não é uma tarefa simples definir o seu significado (FARINELLI, ALMEIDA, 2019).

Argumentos similares são realizados por Carlan (2011) ao explicar que ontologia conceitua os termos usados para representar e escrever um domínio, sendo que ela se expandiu e ganhou notoriedade importante em diversas comunidades científicas como Ciência da Computação, CI, Medicina. Em essência, as ontologias fornecem contexto e significado a uma declaração (SILVA; MARTINS; SIQUEIRA, 2019).

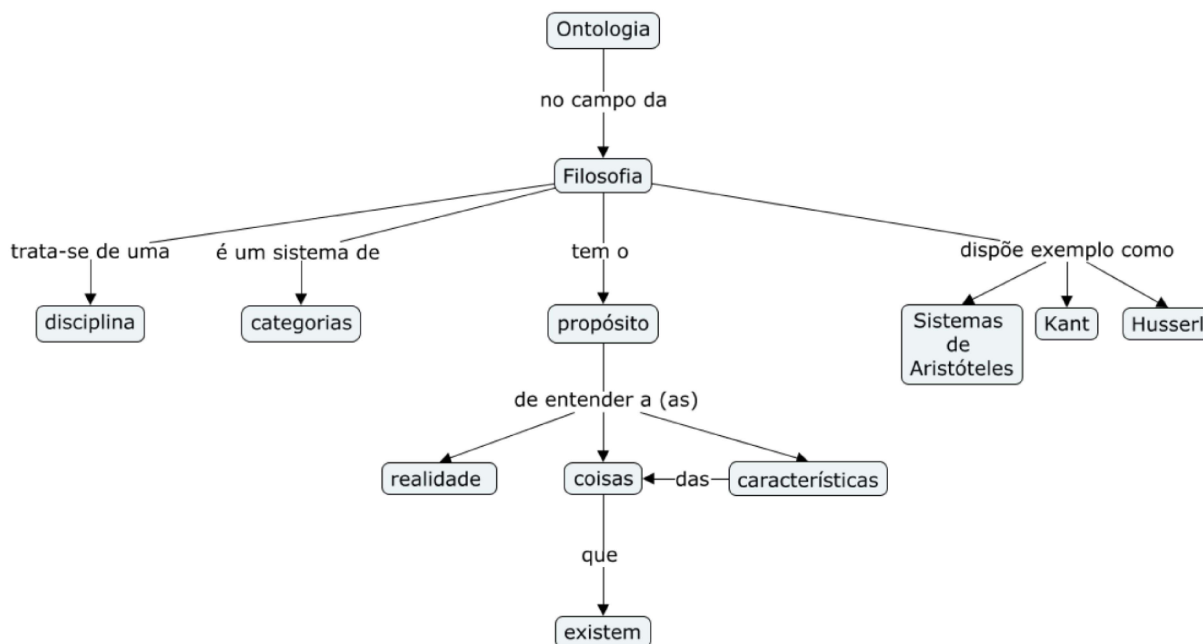
Guarino e Giaretta (1995) sugerem três possíveis interpretações técnicas da palavra “ontologia”: (1) “ontologia” é sinônimo de “teoria ontológica”, ou seja, é um conjunto de fórmulas com o objetivo de ser sempre verdadeiro com base em uma conceituação; (2) “ontologia” quer dizer “especificação de um compromisso ontológico¹⁷”; e (3) “ontologia” quer dizer “conceituação¹⁸”.

Em Filosofia, onde o termo se originou, ontologia é uma área da Metafísica que estuda os diversos tipos de categorias de entidades existentes e os seus relacionamentos (GONZÁLEZ, 2011). Esse termo é emprestado do âmbito filosófico para se referir a um relato sistemático da existência (GRUBER, 1994). Trata-se de uma teoria sobre a natureza da existência, de quais tipos de coisas existem, e como disciplina, estuda tais teorias (BERNERS-LEE; HENDLER; LASSILA, 2002). Neste ramo, refere-se ao conceito da Filosofia que pesquisa ou estuda o ser em geral, as suas propriedades transcendentais, bem como a organização da realidade (GONZÁLEZ, 2011). Almeida (2013) define o termo “ontologia” no ramo da Filosofia conforme apresentado na Figura 1.

¹⁷ Compromisso ontológico é uma explicação semântica parcial da conceituação que se pretende em uma teoria lógica (GUARINO; GIARETTA, 1995).

¹⁸ Conceituação é uma estrutura semântica com a intenção de codificar regras implícitas restringindo a estrutura de um fragmento da realidade (GUARINO; GIARETTA, 1995).

Figura 1 - Mapa conceitual de ontologia na Filosofia ([link para acesso em tamanho maior](#))



Fonte: autoria própria, a partir dos conceitos de Almeida (2013).

Conforme apresentado no mapa conceitual da Figura 1, ontologia no campo da Filosofia é uma disciplina, ou seja, um sistema de categorias, que tem o propósito de atender a realidade, as coisas existentes e suas características. E dispõe de exemplos como Sistemas de Aristóteles, Kant e Husserl.

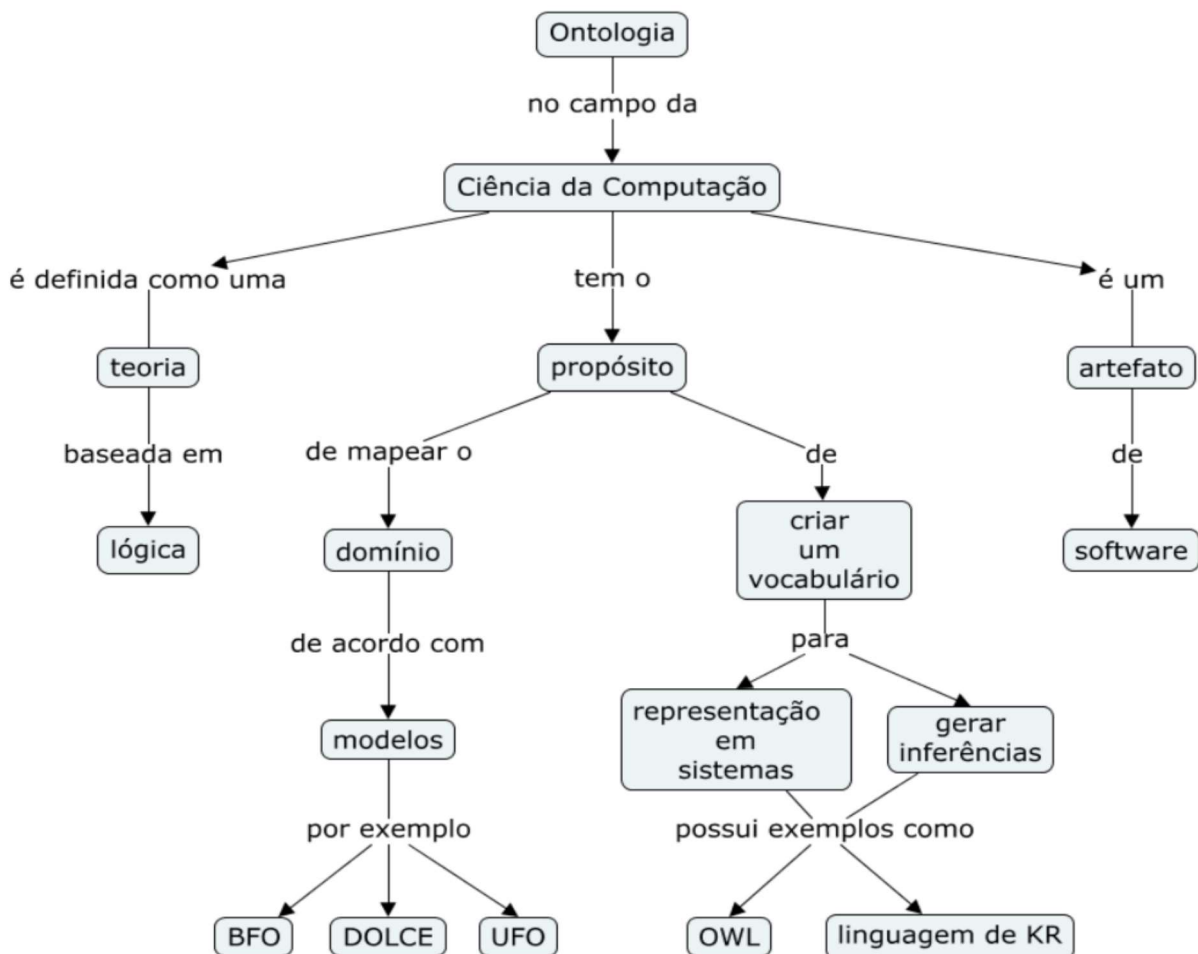
Na Ciência da Computação, o termo ontologia surgiu pela primeira vez em 1967, na obra *Another Look at Data* (Outro Olhar Sobre Dados) proposta por George Mealy (FARINELLI; ALMEIDA, 2019). Ela diz respeito à modelagem de Sistemas de Informação (SI) em campos distintos como os sistemas tradicionais baseados em Bancos de Dados, bem como os sistemas baseados em conhecimento, ambos relacionados ao ramo da Representação do Conhecimento (FARINELLI, ALMEIDA, 2019).

Pesquisadores de Inteligência Artificial (IA) e da Web associam o termo ontologia para seu próprio entendimento e, para eles, uma ontologia pode ser entendida como um documento ou um arquivo que define formalmente as relações entre os termos (BERNERS-LEE; HENDLER; LASSILA, 2002). No âmbito dos sistemas de IA, o que “existe” é aquilo que pode ser representado (GRUBER, 1994). Nesse ramo ela foi introduzida na década de 1970, e desde os anos 80 na Modelagem Conceitual de Sistema de Informação. Já nos anos 90, sua pesquisa expandiu-se com a Web semântica, aumentando a necessidade por ontologias em aplicativos de

diversos tipos para resolver problemas de interoperabilidade, e também para conceder uma representação estrutural unificada de comunicação (FARINELLI; ALMEIDA, 2019).

Complementando a noção da informação poder ser ordenada e estruturada por intermédio de ontologias, Borst (1997) agrega a definição de Gruber (1994) defendendo que ontologia é uma especificação formal e explícita de uma conceituação compartilhada, onde, “formal” por ser legível por computadores, “explícita” por considerar conceitos, propriedades, relações, funções, restrições e axiomas determinados e processados por computadores, “de uma conceituação” por relacionar modelos abstratos aos fenômenos do mundo real, e “compartilhada” devido ao conhecimento consensual reutilizado (BORST, 1997). Para Almeida (2013), esse termo no ramo da Ciência da Computação tem o significado conforme exposto na figura 2.

Figura 2 - Mapa conceitual de ontologia na Ciência da Computação ([link para acesso em tamanho maior](#))



Fonte: autoria própria, a partir dos conceitos de Almeida (2013).

O mapa conceitual da Figura 2 apresenta a ontologia no campo da Ciência da Computação como uma teoria baseada na lógica. Ela possui o propósito de mapear o domínio de acordo com modelos como BFO¹⁹, DOLCE²⁰ e UFO²¹, e tem o intuito de criar um vocabulário para representação em sistemas e gerar inferências como por exemplo OWL²² e linguagem de KR²³. Por fim, para a Computação, ontologias podem ser vistas também como artefatos de *software*. A ontologia de fundamentação UFO será abordada com mais detalhes na subseção 4.2.1.2.

Já em CI, ontologia também foi introduzida através de George Mealy. A CI trata de problemas relacionados à modelagem de informações, sendo que essas questões são estudadas pela comunidade de engenheiros do conhecimento (VICKERY, 1997). Nesse sentido, González (2011) argumenta que ontologia é a especificação explícita de uma conceituação ou a parcela da realidade com associações que estes conceitos determinam entre si, bem como axiomas que definem esses conceitos e relações.

Almeida (2013) argumenta que os princípios ontológicos são assuntos relevantes da pesquisa em CI. Como trata-se de uma terminologia estudada em diferentes campos e com diversos significados, acredita-se que uma explicação abrangente do termo permite relacioná-la à pesquisa interdisciplinar. A Figura 3 representa o termo ontologia segundo Almeida (2013).

¹⁹ BFO é o acrônimo de *Basic Formal Ontology* (Ontologia Formal Básica). É uma teoria das estruturas básicas da realidade desenvolvida pelo Instituto de Ontologia Formal e Ciência da Informação Médica. Disponível em: <http://ontology.buffalo.edu/medo/biodynamic.pdf>

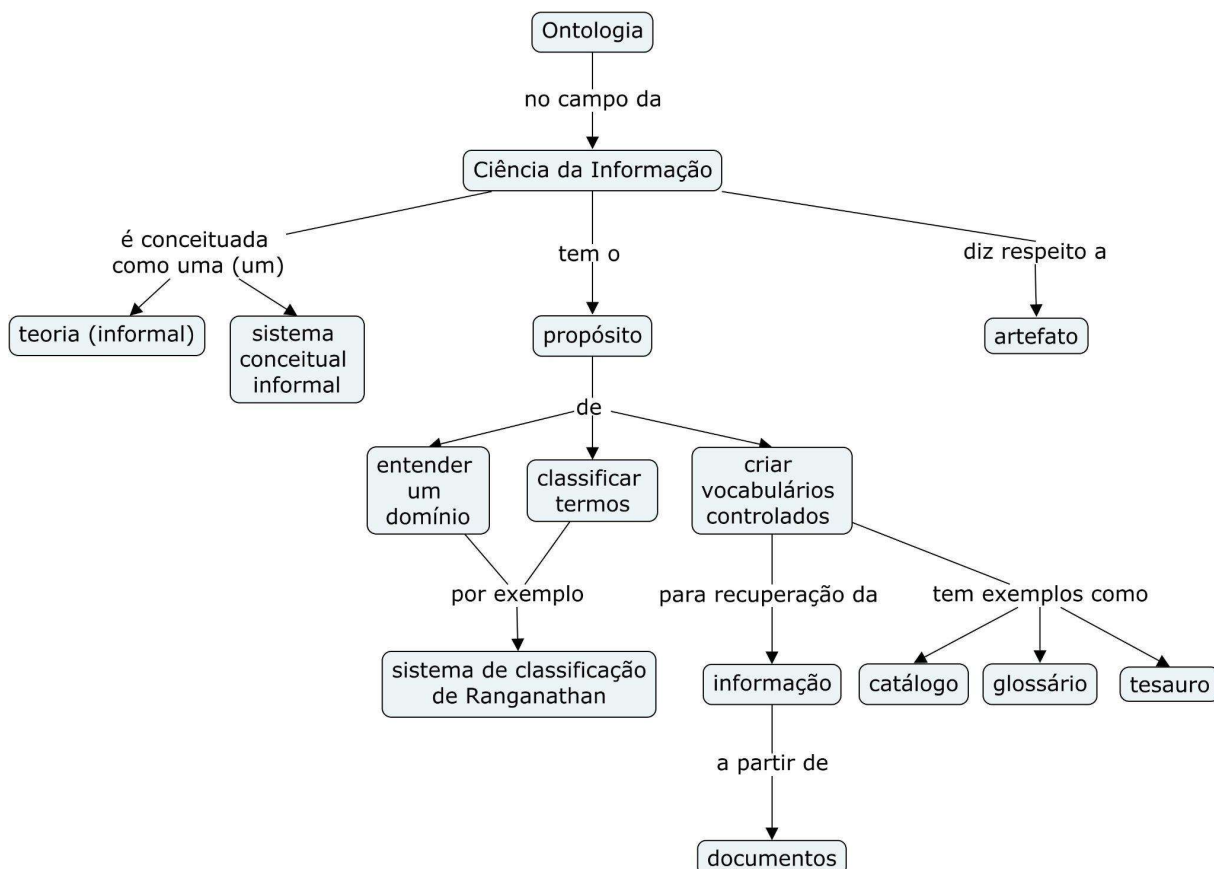
²⁰ DOLCE é o acrônimo de *Descriptive Ontology for Linguistic and Cognitive Engineering* (Ontologia Descritiva para Engenharia Linguística e Cognitiva). É uma ontologia de particulares, e seu domínio de discurso é restrito a eles. Disponível em: <http://www.loa.istc.cnr.it/old/Papers/DOLCE-EKAW.pdf>

²¹ UFO é uma ontologia de fundamentação que tem como objetivo principal fornecer uma base para linguagens gerais de modelagem conceitual (GUIZZARDI; WAGNER, 2005). Disponível em: https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/some_applications_of_a_unified_foundational_ontology_to_business_modeling_2005.pdf

²² OWL é o acrônimo de *Web Ontology Language* (Linguagem de Ontologia da Web). É uma linguagem que pode ser usada para descrever as classes e as relações entre elas que são inerentes a documentos e aplicativos da Web. Disponível em: <https://www.w3.org/TR/owl-guide/>

²³ KR é o acrônimo de *Knowledge Representation* (Representação do Conhecimento). É a ciência, a tecnologia e a aplicação dos métodos e ferramentas para representar o conhecimento objetivando a pesquisa e a recuperação otimizada em bancos de dados digitais. Disponível em: <https://www.isko.org/cyclo/blog>

Figura 3 - Mapa conceitual de ontologia na Ciência da Informação ([link para acesso em tamanho maior](#))



Fonte: autoria própria, a partir dos conceitos de Almeida (2013).

Conforme expresso no mapa conceitual da Figura 3, na CI ontologia é um artefato - uma teoria informal ou um sistema conceitual informal. Ela tem o propósito de atender a um domínio ou de classificar termos, como por exemplo o Sistema de Classificação de Ranganathan, ou de criar vocabulários controlados para a RI a partir de documentos como um catálogo, um glossário ou um tesauro.

Como a ontologia pode ser entendida como um sistema de categorias que lida com a organização da realidade e das entidades abstratas, um dos elementos disciplinares filosóficos dela trabalha com a área de Ontologia Formal, que vincula-se com representações formais de classes independentemente da sua própria natureza (ZAMBORLINI, 2011, p.26). No âmbito de CI, já que o termo informação é amplo e pode ser analisado por várias perspectivas, Neches *et al.* (1991 *apud* GONZÁLEZ, 2011) explicam que na representação da informação, a ontologia pode ser entendida como a soma do encadeamento de conceitos relevantes que retratam o conhecimento compartilhado pelos integrantes de um domínio particular. Nesse âmbito, os princípios ontológicos podem ser usados para auxiliar na construção de estruturas categóricas

para representar o conteúdo dos documentos (ALMEIDA, 2013). Isto é, não limita-se apenas à representação do conteúdo e dos recursos da temática, mas também da abrangência da representação do recurso como um todo (geralmente da perspectiva de uma determinada comunidade de usuários) (ALMEIDA, 2013). Nesse sentido, a ontologia possui o potencial de explicar todo o ambiente social que faz parte da análise realizada por cientistas da informação e possibilita novos olhares do ponto de vista da disciplina de Ontologia Social²⁴ (ALMEIDA, 2013).

No contexto de ontologia como coleções de informações, é esperado que a informação seja ordenada, estruturada ou contida de alguma maneira, caso contrário ela permanecerá isenta de forma e sem utilidade. A ontologia possibilita atribuir semântica aos metadados (CASTRO; SIMIONATO; ZAFALON, 2016), e isso permite certa reflexão sobre sua aplicação na Web semântica, pois é dependente de infraestrutura de metadados que devem ser interpretados e processados por máquinas que transmitem linguagem de indexação para processar e inferir sobre buscas (CARLAN, 2010). Estas linguagens usadas na descrição dos metadados devem ser formatadas e codificadas de forma padrão, para que se torne fácil de ser processada e compartilhada por máquinas (CARLAN, 2010). Caso contrário, a objetividade mecânica disso recai sobre nossa qualidade humana, ou seja, nas limitações do nosso aparelho sensorial (MCGARRY, 1999).

Ainda sobre a conceituação do termo, González (2011) esclarece que ontologia é a própria conceituação formal de uma área ou de uma fração da realidade, com a qual podem operar distintas aplicações de *software* em que: (1) conceitos ou termos usados para a definição tem como objetivo um vocabulário comum (sintático e semântico), favorecendo a troca de mensagem bem como a interoperabilidade de recursos; (2) fornecem sentido integral à informação, relacionando-a como parte de um contexto.

Com relação às tecnologias digitais, Taylor e Joudrey (2008 *apud* Araújo, 2018, p.50) abordam a ontologia como a padronização da codificação de vocabulários controlados. Elas apresentam importantes particularidades referentes aos modelos como tesouros e as taxonomias digitais (GUARINO, 1998; CAMPOS; GOMES, 2017 *apud* Araújo, 2018).

²⁴ Ontologia social é uma ontologia regional que representa o social e seus fenômenos (SELL, 2021). Para Schmitt (2013, p.923 *apud* SELL, 2021, p.1) “a ontologia social indaga em que sentido as entidades sociais existem e qual é a sua natureza básica, bem como das relações sociais”.

As ontologias proporcionam a interoperabilidade entre SI, promovendo e facilitando o compartilhamento e a reutilização do conhecimento entre sistemas mediante um processo inteligente. Na literatura, existem autores que definem ontologias como linguagens documentárias por meio de elementos de formação como termos, definições e relações (CARLAN, 2011).

González (2011) esclarece que para que a camada ontológica possa inferir conhecimento é preciso contextualizar a informação no âmbito de um cenário peculiar. É esclarecido que nesta camada são definidas as ontologias correspondendo a três aspectos distintos e relacionados, sendo eles: (1) a soma de conceitos importantes do conhecimento, distribuídos pelos membros de um determinado domínio, (2) as associações entre esses conceitos e, (3) os axiomas estabelecidos sobre os conceitos e relações mencionados (GONZÁLEZ, 2011).

Para González (2011), às ontologias contemplam um salto qualitativo que está no uso de motores de busca com palavras-chave, e usam-se também agentes de *software* inteligentes por intermédio de conceitos. Elas possibilitam a transposição da simples recuperação de informação para o arrojado alcance de respostas precisas a consultas concretas com a finalidade de recuperar o conhecimento.

O capítulo da Fundamentação Metodológica desta dissertação, mais especificamente na subseção 4.2.1, trata da modelagem conceitual baseada em ontologia, abordando metodologias para a construção de ontologias de domínio, ontologias de fundamentação, e linguagem de modelagem orientada a ontologia, e portanto, complementando as definições expostas no presente tópico.

3.3 WEB SEMÂNTICA

A Web semântica é definida como uma teia de informações elaborada com o propósito de ser facilmente processada por máquinas em uma escala global. Ela busca criar uma forma eficiente para representar dados na *World Wide Web*, buscando estabelecer um banco global de Dados Ligados (LAUFER, 2015).

Para Berners-Lee, Hendler e Lassila (2002), a Web semântica não é a própria Web, mas sim uma extensão dela em que as informações fornecidas são bem definidas. Isso significa que, com a Web semântica, os computadores e as pessoas podem trabalhar em melhor cooperação. Ela inaugura uma nova funcionalidade à medida que as máquinas se tornam capazes de processar e "[...] entender [...]" os

dados existentes na Web (BERNERS-LEE; HENDLER; LASSILA, 2002). González (2011) define Web semântica como uma extensão da Web ao acrescentar semântica no presente modelo de representação de dados. Ele explica que a Web semântica, por intermédio das suas linguagens e padrões, integra o alicerce para a concretização do tratamento acurado das informações no ambiente Web. Para alcançar tais mecanismos, diversas tecnologias foram elaboradas, tais como o uso de ontologias, pois ela objetiva incorporar significado e sentido ao conteúdo dos documentos, operando como instrumento de representação do conhecimento. No âmbito da Tecnologia da Informação (TI), ontologias são designadas as classificações de identidades da informação ou da matéria (objetos existentes no planeta). Assim como as ontologias são componentes básicos da Web semântica, Carlan (2010) argumenta que elas permitem refletir sobre como o desenvolvimento da Web semântica é dependente da infraestrutura de metadados, e que esses últimos podem ser entendidos e processados por máquinas que transmitem linguagens de indexação. Estas linguagens utilizadas na descrição de metadados precisam ser formatadas e codificadas em um padrão para que o processamento e compartilhamento por máquinas se torne mais fácil (CARLAN, 2010).

A intenção da Web semântica é que a Internet seja pesquisável não apenas por meio de palavras, mas também por meio do significado. Isso requer interoperabilidade semântica e sintática de um vocabulário de assunto, pois não é suficiente se a descrição do assunto for baseada em conceitos isolados, e muitas vezes também deve ser baseada na lógica proposicional (VELTMAN, 2001, 2002, 2004 *apud* SLAVIC, 2005). Soergel (1999) afirma que:

Os SOC são reconhecidos como fonte importante de vocabulários estruturados e formalizados que podem ser explorados para dar suporte ao desenvolvimento da web semântica, a qual contempla, ainda, dados estruturados ou semiestruturados, metadados como padrões de intercâmbio e os modelos de representação para controle da linguagem e organização na recuperação da informação (Soergel, 2006 *apud* CARLAN, 2010, p.32).

Assim, para Web semântica funcionar, os computadores devem ter acesso a coleções de informações e conjuntos de regras de inferência para conduzir o raciocínio automatizado. Cabe ressaltar que pesquisadores de Inteligência Artificial (IA) estudavam esses sistemas antes da Web ser desenvolvida e que a Representação do Conhecimento é comparada ao hipertexto antes do advento da

Web (BERNERS-LEE; HENDLER; LASSILA, 2002). Sobre o conhecimento, Gutierrez e Juan (2021) argumentam que em 1980 houve uma conquista importante para a compreensão entre o poder expressivo de uma linguagem lógica e a complexidade computacional das tarefas de raciocínio. Eles defendem que aumentando o poder expressivo em uma linguagem lógica, conseqüentemente aumenta a sua complexidade computacional, e isso norteou as pesquisas, originando o conceito de Lógica da Descrição. Os autores evidenciam *softwares* como KL-ONE²⁵, LOOM²⁶ e CLASSIC²⁷, e foi através de sistemas como esses que ocorreu o desenvolvimento do raciocínio sobre esquemas e estruturas, assim como a capacidade de rastreabilidade subsunçadora, pois possibilitou-se apresentar que a lógica do raciocínio pode ser implementada em um *software*.

As próximas subseções tratam de algumas possibilidades de organização da Web semântica, tais como vocabulários controlados, *knowledge graph*, dados ligados e interoperabilidade.

3.3.1 Organização da Web semântica

A importância da Web semântica está na possibilidade de pessoas criarem programas que coletam conteúdo da Web de fontes diversas, processam essas informações e trocam esses resultados com outros programas (BERNERS-LEE; HENDLER; LASSILA, 2002). Para isso ser possível, Laufer (2005) esclarece que a Web semântica define três blocos básicos: um modelo de dados padrão, um conjunto de vocabulários de referência e um protocolo padrão de consulta. Inclusive, ela dispõe de quatro regras fundamentais para a criação de dados ligados na Internet que são (BERNERS-LEE *et al.*, 2001): (1) usar URIs como nomes para coisas; (2) utilizar URIs HTTP para as pessoas pesquisarem esses nomes; (3) na pesquisa, a URI fornecer informações através dos padrões RDF e SPARQL; e (4) incluir links para outros URIs com a finalidade de se encontrar mais pesquisas.

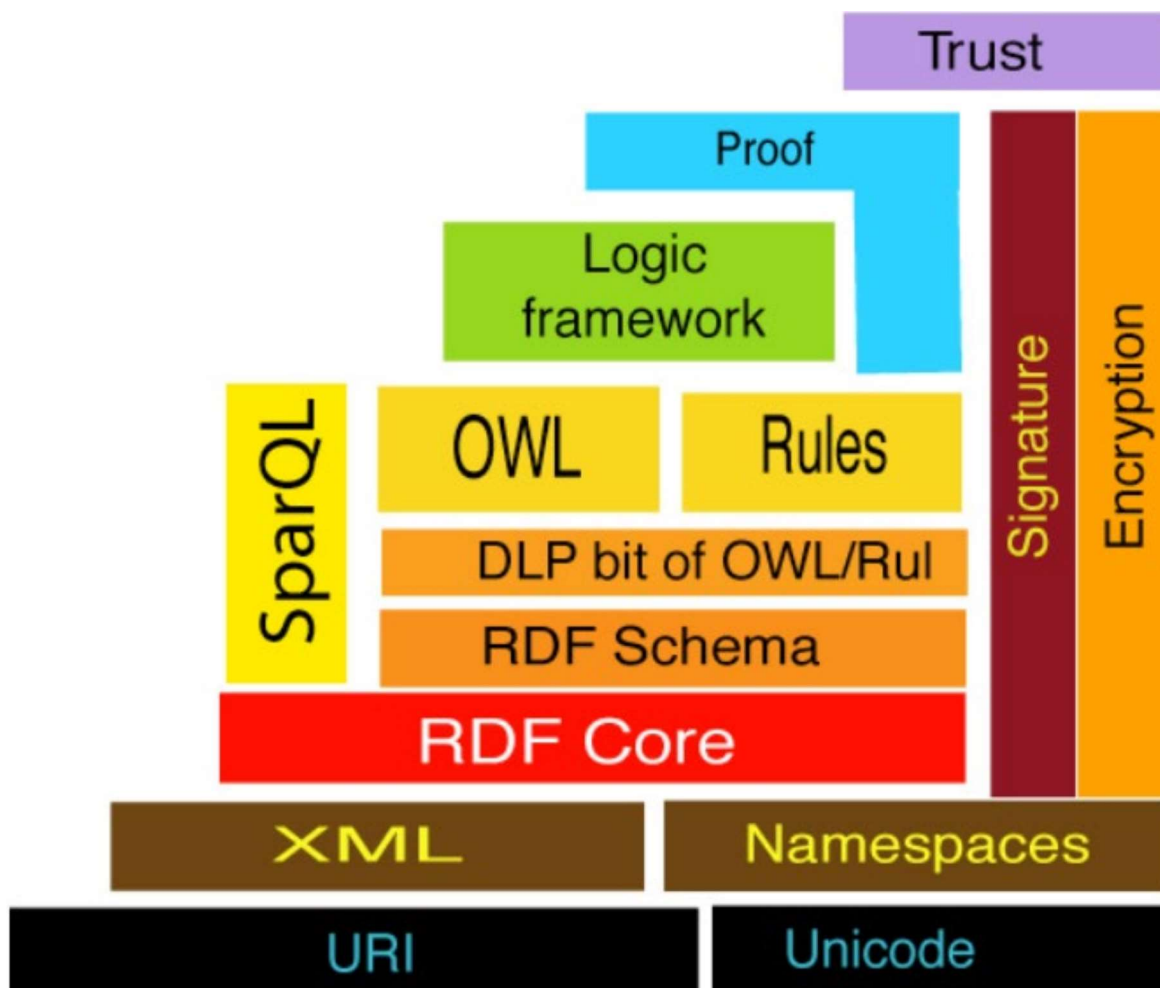
²⁵ KL-ONE é um sistema para representação de conhecimento para programas de Inteligência Artificial. Disponível em: https://onlinelibrary.wiley.com/doi/epdf/10.1207/s15516709cog0902_1

²⁶ LOOM é uma linguagem e um ambiente de programação que almeja a construção de sistemas especialistas e outras aplicações inteligentes. Disponível em: <https://www.ime.usp.br/~eudenia/iia/onto/loom.html>

²⁷ CLASSIC é um *software* que utiliza conceito de Lógica da Descrição para desenvolvimento do raciocínio tanto sobre esquemas como estruturas, bem como a capacidade de rastreabilidade subsunçadora (GUTIERREZ; JUAN, 2021, p.100).

Para estabelecer a organização da Web semântica, Berners-Lee (2005) apresentou a pilha da Web semântica em um *Workshop* de regras em 2005, conforme disposto na Figura 4.

Figura 4 - Arquitetura da Web semântica conhecida como "bolo de noiva" ou "pirâmide da Web semântica" ([link para acesso em tamanho maior](#))



Fonte: Berners-Lee (2005).

A Figura 4 apresenta a arquitetura da Web semântica com suas linguagens e outros componentes. Destaca-se abaixo cada camada representada, do nível de complexidade mais simples para o mais complexo (BERNERS-LEE, 2005; MONTEIRO, 2013; CRISTOVÃO, 2016):

- **URI e Unicode:** Constituem a camada que integra a Web com a Web semântica, usando sempre que possível um padrão internacional, como por exemplo, o Unicode; Os endereços URI são usados para denominar de maneira única elementos como classes, conceitos, relações, propriedades. O Unicode é um padrão universal que possibilita computadores representar

e manipular textos de qualquer sistema de escrita existente, e é adotado para endereçar e garantir a legibilidade;

- **XML²⁸ e Namespaces:** Interliga a camada Web com a Web semântica, dispondo de informações e documentos presentes, e convertendo da linguagem HTML para XML. O XML é uma evolução do HTML, uma metalinguagem que serve para descrever linguagens. Em outras palavras, ele define a organização dos dados em conjunto com os namespaces, separando o próprio dado (conteúdo) do seu nome (rótulo). Já os namespaces são os apelidos para URI (geralmente são apresentados por meio de um prefixo).
- **RDF Core:** É uma camada específica da Web semântica que representa uma rede de informações formada por triplas (sujeito, predicado e objeto). Ele apresenta o modelo e a sintaxe para descrever uma tripla. Trata-se de um tipo de estrutura de descrição de recursos criado para rearticular especificações adicionando novos recursos (incluindo valores tipificados de dados);
- **RDF Schema:** É um conjunto de classes e propriedades RDF que compõem a dimensão do vocabulário básico de RDF. Ele atribui propriedade taxonômica na camada RDF Core, objetivando melhorar o relacionamento entre as informações e sua facilidade de acesso, tratando-se de uma camada que possibilita realizar inferências;
- **DLP²⁹ bit of OWL/Rul:** É uma camada que permite o uso de ontologias da Web semântica para facilitar a comunicação entre RDF e OWL. Essa camada permite formalizar, padronizar e melhorar a expressividade semântica das informações por meio de linguagens de ontologia, proporcionando a interoperabilidade e o acesso delas;
- **OWL e Rules:** Camada que trabalha com informação em alto nível, onde as regras são usadas para representar o conhecimento. Essa camada permite compor a linguagem recomendada para definir ontologias que descrevem

²⁸ XML é o acrônimo de *eXtensible Markup Language* (Linguagem de Marcação Estendida). É uma tecnologia projetada para facilitar a implementação e proporcionar a interoperabilidade na Web. Disponível em: <https://www.w3.org/TR/REC-xml/>

²⁹ DLP corresponde a *Description Logic Programs* (Descrição Programas Lógicos). Disponível em: <http://www.cs.man.ac.uk/~horrocks/Publications/download/2003/p117-grosof.pdf>

semanticamente os dados. OWL trata-se de uma linguagem usada para descrever as classes e as relações entre elas, e que são inerentes aos documentos e aplicativos da Web, bem como permite implementar regras (rules) com esses mecanismos;

- **SparQL:** É a camada de definição da linguagem utilizada em query ou pesquisa, durante as inferências disponibilizadas pela Web semântica, ou seja, trata-se de uma linguagem de consulta e protocolo para RDF. Por essa razão ela é vertical quanto às camadas RDF Schema, DLP bit of OWL/Rul, OWL e Rules;
- **Logic framework:** Camada que abriga modelos lógicos e a semântica dos dados, proporcionando atividades de inferências, isto é, ela disponibiliza mecanismos de inferência que trabalham com lógica de primeira ordem para manipular conhecimento;
- **Proof:** Camada responsável pelas verificações necessárias objetivando comprovar e exibir inferências. Ela tem a capacidade de apresentar os procedimentos detalhados do resultado por intermédio de uma máquina de inferência de lógica de primeira ordem. São testes para verificar o quanto uma afirmação, ou um teste, poderia ser provado, objetivando alcançar a confiança (Trust);
- **Signature e Encryption:** Camada que corresponde a assinatura (signature) e criptografia (encryption), sendo perpendicular aos elementos RDF Core, RDF Schema, DLP bit of OWL/Rul, SPARQL, OWL, Rules, Logic Framework e Proof. Signature tem o objetivo de fornecer suporte à assinatura e identificação das descrições e definições que implicam diretamente na realização de inferências, enquanto encryption dá suporte à criptografia e identificação que são relacionadas às inferências;
- **Trust:** Essa camada representa a confiança da informação e do conhecimento proporcionado pela Web semântica, usando instrumentos como chaves públicas, assinatura digital e criptografia, conforme definido na camada Signature e Encryption.

A arquitetura para a Web semântica foi proposta para organizar de maneira simplificada, em camadas, as linguagens e elementos utilizados no cenário da Web semântica. Berners-Lee (2005) especifica os recursos e as linguagens para a Web

semântica e esclarece que todas as linguagens são mais visíveis do que linguagens de script, além de mencionar a existência da aplicação do princípio da menor potência.

Por outro lado, González (2011) apresenta as cinco camadas da Web semântica conhecidas como Camada Sintática, Camada Semântica, Camada Ontológica, Camada Lógica e Camadas de Prova e Confiança. As camadas Sintática, Semântica e Ontológica (três primeiras) são basilares para representar a informação de forma semanticamente acessível. As camadas Lógica e de Prova e Confiança determinam os elementos que possibilitam aos agentes de *software* reconhecer e comprovar de maneira confiável os diferentes elementos do modelo como recursos, agentes, inferências obtidas (GONZÁLEZ, 2011).

Essas camadas favorecem o princípio fundamental da Web semântica, que está em permitir humanos e máquinas entender e interpretar dados na web. Nesse contexto, as máquinas não servem apenas para exibir dados, mas também para usá-los com a finalidade de integração e reutilização em diversas aplicações (SURE; STUDER, 2005 *apud* RAZA *et al.*, 2019). Tim Berners-Lee *et al.* (2001, p.1) complementa argumentando que a Web semântica estabelece a estrutura do conteúdo significativo para as páginas Web, elaborando um espaço onde os agentes de *software* se propagam a cada página, realizando de maneira facilitada tarefas sofisticadas para os usuários.

González (2011) esclarece que a Web semântica tem a sua finalidade baseada em duas ideias distintas e principais que são: (1) rótulo semântico de recursos (uma separação formal entre conteúdo e estrutura de documentos), bem como (2) elaboração de aplicações de *software* “inteligentes” (conhecidos como agentes com capacidade de receber, processar e operar os recursos comentados em nível semântico).

Laufer (2015) explica que a Web semântica é um mecanismo que objetiva facilitar a comunicação entre os diversos componentes do ecossistema na Web, elaborando um modelo mental comum, reduzindo ambiguidades, e aprimorando o trabalho necessário para o desenvolvimento de aplicações que manipulem fontes de dados distintas.

E para sugerir os mecanismos mencionados, González (2011) indica que a Web semântica propõe ao menos quatro itens: (1) a sintaxe comum (como por exemplo, o XML como expressão do conhecimento em estruturas simples

predefinidas com o RDF), (2) o uso de ontologias e vocabulários de metadados (como por exemplo, o *Simple Knowledge Organization System - SKOS*³⁰), (3) o ato ou o efeito de referenciar termos por intermédio de recursos que esclareçam seu conteúdo por meio da URI e, (4) o uso por parte das máquinas e das pessoas (*RDF Site Summary*³¹).

3.3.2 Vocabulários Controlados

Harpring *et al.* (2006) definem vocabulário controlado como uma combinação organizada de palavras e frases que são usadas para indexar conteúdos com o propósito de recuperá-los por meio de navegação ou pesquisa. Os autores afirmam que normalmente o vocabulário controlado diz respeito aos termos preferenciais sobre um escopo limitado, ou representa um campo de domínio específico. Por sua vez, Zeng (2005) esclarece que o controle de vocabulário é usado com o objetivo de aprimorar a eficácia de sistemas de armazenamento e recuperação de informações, sistemas de navegação na Web e outros artefatos que almejam identificar e localizar conteúdos desejados por intermédio de certo tipo de descrição usando uma linguagem. Ela argumenta que o controle de vocabulário objetiva a obtenção consistente da descrição de objetos e também de conteúdo, para facilitar a recuperação.

Konstantinou e Spanos (2015) esclarecem que, além das bases sintáticas especificadas pelos padrões, os vocabulários precisam ser utilizados para que as informações retratadas sejam habituais e inequívocas quando interpretadas e compreendidas. Os autores alegam que esses vocabulários devem ser reutilizados por diversos produtores de dados ao descrever dados sobre um assunto, tornando semanticamente interoperável cada um desses dados. Eles orientam que é preciso observar os mínimos detalhes entre os vocabulários e os conjuntos de dados, pois são coleções de fatos que correspondem a uma ou mais entidades e ontologias individuais próprias, porque contribuem com descrição terminológica abstrata de alto

³⁰ SKOS é o acrônimo de *Simple Knowledge Organization System* (Sistema Simples de Organização do Conhecimento). Fornece um modelo para expressar a estrutura elementar e o conteúdo de esquemas conceituais como tesouros, taxonomias, e outros tipos parecidos de vocabulário controlado (MILES, A; BECHHOFFER, 2009). Disponível em: <https://www.w3.org/TR/skos-primer/>

³¹ Trata-se de um vocabulário RDF que proporciona descrição de metadados extensíveis, de multiuso, leve e em formato de distribuição. É uma forma produtiva de descrever, gerenciar e disponibilizar informações importantes para distribuição em massa. Disponível em: <https://www.w3.org/2001/09/rdfprimer/rss.html>

nível de um domínio, possuindo axiomas que colaboram com a descrição terminológica de cada elemento.

De maneira complementar, González (2011) argumenta que o principal objetivo dos vocabulários para a Web está em proporcionar buscas por conceitos. Ela explica que para alcançá-los, utiliza-se um mecanismo de mescla de vocabulários registrados (ontologias, tesouros e cabeçalhos de assunto), bem como a ordenação dos termos. Entretanto, a autora adverte que, na maioria das vezes, os vocabulários controlados não são aplicáveis nos locais que se pode usar metadados. Ela afirma que em se tratando da construção, da manutenção e da aplicação, eles têm um custo alto quando relacionados ao tempo de desenvolvimento e na preparação dos usuários. O controle de vocabulário oferece a vantagem de produzir confiança na busca de informações arquivísticas em um sistema de organização, evita a dispersão de informações e, conseqüentemente, gera precisão na RI (SMIT; KOBASHI, 2003).

Não há uma divisão clara entre vocabulários e ontologias. Há certa tendência no uso de ontologia para se referir a uma coleção de termos complexos e formal, enquanto o termo vocabulário geralmente é usado quando o formalismo não é necessário ou para ser aplicado em um sentido livre. Os blocos de construção básicos para técnicas de inferência na Web semântica são os vocabulários (W3C, 2022). No âmbito da Web semântica, os vocabulários definem os conceitos e relacionamentos, que também são denominados termos. Eles são usados para descrever e representar uma área de domínio, e servem também para classificar os termos utilizados em uma aplicação, caracterizar relacionamentos prováveis e definir possíveis restrições à aplicação desses termos (W3C, 2022).

Para especificar em SKOS, o termo ontologia é utilizado com graus distintos de estrutura, perpassando desde as simples taxonomias (hierarquia do Yahoo³²), esquemas de metadados (como por exemplo, o Dublin Core³³), até o que se conhece como teorias lógicas. Dessa maneira, a Web semântica necessita de ontologias com certa estrutura que seja significativa, e para isso é preciso particularizar descrições para os tipos de conceitos conhecidos como classes (aspectos gerais sobre muitos domínios de interesses), relações (ligação que pode existir entre elementos) e

³² Yahoo é um provedor de serviços de acesso à Internet que funciona por meio de um portal. Disponível em: <https://legal.yahoo.com/br/pt/yahoo/terms/otos/index.html>

³³ Dublin Core é uma organização de inovação no *design* de metadados que utiliza as melhores práticas em toda a ecologia de metadados. Disponível em: <https://www.dublincore.org/about/>

propriedades (atributos que esses elementos podem possuir) (GONZÁLEZ, 2011). O SKOS possui a vantagem de aproveitar o conhecimento do setor bibliotecário na OC e pode ser introduzido na Web semântica (GONZÁLEZ, 2011). Ainda, o SKOS oferece a vantagem de orientar conceitos e não termos, e isso melhora o valor dos modelos configurados pelos tesouros, aumentando a capacidade de estabelecer relações entre unidades léxicas ou etiquetas, e ampliando os relacionamentos básicos de um vocabulário controlado tradicional. SKOS também incorpora funções avançadas que permitem estabelecer relações entre esquemas de conceitos e agrupá-los em coleções.

São exemplos de vocabulários controlados, as listas simples de termos preferencialmente exclusivos, os conjuntos de termos correspondente ao mesmo conceito (sinônimos), termos preferenciais e não preferenciais, identificação de hierarquias de termos (taxonomias), bem como esses conjuntos com relações semânticas entre termos e outros conceitos (tesouros). Ou seja, os variados tipos de vocabulários controlados são conhecidos como listas controladas, sinônimo de arquivo anel, taxonomia, cabeçalhos de assunto e tesouros (HARPRING, *et al.*, 2006).

Zeng (2005), reforça Harpring *et al.* (2006), sobre os tipos de vocabulários conhecidos ao definir as estruturas de vocabulários controlados. Ela explica que os vocabulários controlados são estruturados e permitem a apresentação de diferentes tipos de relações entre os termos que eles integram, sendo que existem quatro tipos diferentes de vocabulários controlados, definidos por sua estrutura dá mais simples para a mais complexa: Lista, Anel de Sinônimos, Taxonomia e Tesouro.

Como metodologia para criar vocabulários controlados, Harpring *et al.* (2006) recomenda quais são os elementos de dados que precisam ser descritos e quais desses devem ser arquivos de autoridade na forma de tesouros. Eles esclarecem que os vocabulários controlados devem ser particularizados para que funcionem adequadamente com a situação ou com as coleções específicas da demanda, argumentando que cada organização deve elaborar uma estratégia específica para criar vocabulários controlados numa coleção personalizada. Eles advertem que, se a coleção é consultada em um ambiente federado, os vocabulários controlados devem ser personalizados para haver recuperação em coleções distintas.

Complementando os mecanismos propostos por Harpring *et al.* (2006) como metodologia para criar vocabulários controlados, González (2011) defende que para

criar um vocabulário básico pode-se ter como ponto de partida as frases de busca dos usuários ou de entrevistas com especialistas, buscando determinar a metodologia de transformação do conhecimento tácito do especialista em conhecimento explícito. Zeng (2005) apresenta a existência de quatro princípios de controle de vocabulário que orientam o projeto e desenvolvimento: eliminação de ambiguidade, controle de sinônimos, estabelecimento de relacionamentos entre termos quando apropriado, e teste com validação de termos. Ela elucida que um dos objetivos mais importantes do controle de vocabulário está em garantir que cada conceito diferente se refira a uma forma linguística única. Ela aponta que a eliminação da ambiguidade e a compensação da sinonímia por intermédio do controle de vocabulário garante que cada termo possua um significado único e que apenas um termo poderá ser utilizado para representar um determinado conceito ou entidade.

Harpring *et al.* (2006) recomenda que se um elemento é indicado como controlado, isso quer dizer que os valores deste elemento devem ser derivados do vocabulário controlado com a perspectiva de serem vinculados a uma autoridade competente. Os autores destacam que os padrões de dados não objetivam apenas o registro de informações precisas, mas são fundamentais para a recuperação de registro de informações com eficiência. González (2011) conclui que o termo vocabulário está ocupando o espaço do seu precedente das linguagens documentárias, pois está convertendo-se em uma designação genérica para tratar de estruturas ou de elementos normalizados que retratam e organizam a informação no âmbito da Internet. Isso significa que sua polissemia está limitando-se ao seu contexto de utilização. Na Web Social, adequa-se a concepção de folksonomia³⁴. Nos padrões de metadados, o vocabulário é definido como a coleção de propriedades e de elementos descritivos (GONZÁLEZ, 2011).

Em um contexto relacionado a RI no âmbito da CI, ao comentar sobre a geração de dados ligados, Konstantinou e Spanos (2015) esclarecem que reutilizar termos de vocabulários controlados da Web semântica na elaboração do mapeamento de banco de dados para ontologia não é suficiente para gerar gráficos RDF que interliguem

³⁴ Folksonomia - Trata-se de um termo introduzido em 2005 por Vander Val para denominar a etiquetagem dos recursos da web em ambiente social realizada pelos próprios usuários (ARAÚJO, 2018). É uma maneira de indexar informações (FOLKSONOMY, 2022), surgida com a Web 2.0 e, no campo da CI, articulada a descentralização das ações de representação da informação (ARAÚJO, 2018; FOLKSONOMY, 2022).

facilmente com outras bases de dados. Para ocorrer a verdadeira geração de dados ligados, as entidades do mundo real que dispõem os valores do banco de dados que representam devem ser identificadas e os links entre eles estabelecidos, o que não ocorre com a maioria dos métodos atuais, pois representam valores de banco de dados como literais em RDF.

3.3.3 *Knowledge Graph* (Grafo de Conhecimento)

Recuperação da Informação na era digital contemporânea ocorre principalmente por meio da Internet e um de seus veículos é o *knowledge graph* (KG). Para Yang *et al.* (2020), a incorporação de grafos em rede conquistou o interesse de pesquisadores. KG compreende a efetivação de nós em rede em um ambiente composto por codificação de estruturas de rede e propriedades de nós (HODGE, 2021). Essa técnica foi aplicada em vários ramos como recomendação, classificação de nós e previsão de ligação.

Gutierrez e Juan (2021) defendem que a definição de KG perpassa por progressos científicos em áreas distintas de pesquisa, como por exemplo: Web semântica, Bancos de Dados, representação e raciocínio do conhecimento, PLN³⁵ e aprendizagem de máquina. A fusão dessas ideias e técnicas apresenta um estímulo para conhecer como os novos avanços estão consolidados nessas técnicas (GUTIERREZ; JUAN, 2021).

O termo KG surgiu em 1972 (SCHNEIDER, 1973), sendo que Hogan *et al.* (2021) discutem que sua definição é controversa, e explicam que há uma série de definições conflitantes e variação de técnicas particulares que são propostas para métodos gerais e inclusivos para registrar e transferir conhecimento da realidade. A definição inclusiva estabelece que um KG é equivalente ao grafo de dados e tem como objetivo acumular e transmitir conhecimento do mundo real através de nós, sendo que esses representam entidades de interesse e a ligação entre eles por arestas representam as relações entre essas entidades. Por sua vez, Gutierrez e Juan (2021) definem KG como o avanço da vinculação de conhecimento e dados em larga escala integrados a formatos heterogêneos.

A academia aplica o conceito de KG para designar sistemas que integram dados com estrutura de grafos, ou uma organização proveniente da Web semântica e

³⁵ PLN corresponde a Processamento Linguagem Natural (*Natural Language Processing* (NLP)).

dados ligados (GUTIERREZ; JUAN, 2021), e isso significa que a sua concepção está em consonância com o modelo de dados que é baseado em grafo (HOGAN *et al.*, 2021).

Conforme a RI ocorre por meio da Internet, o KG corresponde a um modelo de conhecimento que abarca uma coleção de descrições interligadas de conceitos, entidades, relacionamentos e eventos. Eles incorporam dados em contextos por intermédio de metadados semânticos e links, e com isso fornecem certa estrutura para integração, centralização, análise e compartilhamento de dados (ONTOTEXT, 2022b).

Na prática, KG propõe-se a servir como um fundamento de compartilhamento do conhecimento em constante expansão dentro de uma comunidade ou mesmo organização (NATASHA *et al.*, 2019). Suas características mais importantes estão em combinar vários paradigmas de gerenciamento de dados como Banco de Dados (dados explorados por meio de consultas estruturadas), Grafo (podem ser analisados como qualquer estrutura de dados de rede) e Base de Conhecimento (carregam semântica formal e pode ser usado como fonte de interpretação dos dados e inferir novas informações) (ONTOTEXT, 2022b).

Sob a ótica do KG como um grafo de dados que objetiva acumular e transmitir conhecimento do mundo real, a aplicação de suas pesquisas pode se tornar um aglomerado de técnicas de áreas distintas com o objetivo comum de maximizar o conhecimento. Depois de discutir os conceitos apresentados acerca de *knowledge graph*, é notório que o seu uso favorece a recuperação da informação em rede, e isso pode ser aprimorado com a adoção de dados ligados.

3.3.4 Dados Ligados

Dados Ligados (*Linked Data*) são conceituados como um conjunto de boas práticas com a finalidade de publicar e conectar uma coleção de dados estruturados na Web, objetivando a elaboração de uma Web de Dados (BIZER *et al.*, 2006 *apud* ISOTANI, 2015). São um conjunto de práticas recomendadas para disseminação e inter-relação de dados estruturados para serem acessados tanto por humanos como máquinas por intermédio dos padrões conhecidos como Estrutura de Descrição de Recursos (*Resource Description Framework* (RDF)) para troca de dados e a linguagem SPARQL para a consulta deles (HYLAND; ATEMEZING; VILLAZÓN, 2014).

Hyland, Ateazing e Villazón (2014) afirmam que, ao utilizar os princípios de dados ligados, eles proporcionam aos desenvolvedores consultar dados vinculados de bases distintas simultaneamente, além de associá-los sem a exigência de um único esquema comum que esses dados compartilham. Eles argumentam que os dados ligados estão em evidência, e explicam que esses mecanismos são usados por governos como principal mecanismo de busca, por organizações internacionais, e também por startups ágeis.

Dados ligados correspondem a uma coleção de princípios de *design* para compartilhar dados interligados e legíveis por equipamentos computáveis na Web (ONTOTEXT GRAPHDB, 2022c; CONELEGIAN, 2016; BAUER; KALTENBÖCK, 2012), enquanto o Dado aberto corresponde aos dados que são capazes de serem usados e distribuídos livremente (ONTOTEXT GRAPHDB, 2022a; BAUER; CONELEGIAN, 2016; KALTENBÖCK, 2012). Essas afirmações consideram que a concepção de Dados Abertos é realizada sobre a ideia da Web Social (BAUER; KALTENBÖCK, 2012), e o conceito de dados ligados é originário da Web semântica (BAUER; KALTENBÖCK, 2012).

A fusão entre *Linked Data* e *Open Data* (Dado Aberto) chama-se *Linked Open Data* (LOD) (Dado Ligado e Aberto), sendo esse último um Dado Ligado disponibilizado sob licença aberta (BERNERS-LEE, 2010). Segundo Conelegian (2016), LOD foi introduzido em 2006 por Berners-Lee ao relatar que a Web semântica tem como principal fundamento promover a interligação entre os dados, permitindo tanto homens como agentes computacionais explorarem a Web de dados. Esses mecanismos de dados ligados proporcionam a descoberta dos dados por parte dos usuários (Berners-Lee, 2006), isto é, LOD corresponde ao conjunto das melhores práticas objetivando publicar e conectar dados estruturados na Web, para assim, proporcionar que sejam criadas ligações entre dados de fontes distintas espalhadas em âmbito global (HEATH, 2012 *apud* CONELEGIAN, 2016). O progresso dos dados ligados no contexto de Dados Abertos está em tornar esses dados disponíveis na Web de forma pública (CRISTOVÃO; FERNANDES, 2018), por meio do conceito de LOD (BIZER; HEATH; BERNERS-LEE, 2009; BERNERS-LEE, 2010). É um movimento em expansão para as organizações disponibilizarem seus dados, a fim de permitir que usuários criem e combinem conjuntos de dados objetivando realizar suas

interpretações sobre eles e disponibilizá-los em formatos para aplicações processá-los (BAUER; KALTENBÖCK, 2012).

Conforme os dados ligados, no contexto de Dados Abertos, objetiva tornar os dados disponíveis na Web de forma pública, a preocupação com a privacidade das organizações foi um obstáculo para a adoção de tecnologias de dados ligados em bibliotecas digitais (YADAGIRI; RAMESH, 2013 *apud* RAZA *et al.*, 2019). Assim, as leis de direitos autorais foram discutidas por Gonzales (2014 *apud* RAZA *et al.*, 2019) e estão entre os principais problemas para disseminar dados na Web. As bibliotecas empenharam-se para adquirir licença de diversos periódicos, bancos de dados e recursos online diferentes que objetivavam atender seus usuários sem impedimento. A utilização de dados comerciais e não comerciais deve ser decidida no âmbito dos dados ligados. Isso ficou evidente quando Charles *et al.* (2014 *apud* RAZA *et al.*, 2019) apontaram que entre os desafios para os dados ligados e Abertos na Web estão os tipos de dados heterogêneos e multilíngues. Embora haja certa preocupação com a privacidade por parte das organizações quanto ao modelo de dados ligados e Abertos, esses modelos e padrões de disponibilização dos dados ligados compõem a base de elaboração da Web de Dados, e conseqüentemente proporcionam uma das bases necessárias para o estabelecimento da Web semântica (ISOTANI, 2015).

Dados ligados são um conceito que se refere à maneira de vincular os dados para processá-los. Esses dados podem localizar-se em fontes de dados interconectados, distribuídos ou não. Assim, no sentido de organizá-los e processá-los, frequentemente é preciso preparar a interoperabilidade para os sistemas que os armazenam (KONSTANTINO; SPANOS, 2015). Dados ligados podem revolucionar a maneira dos bibliotecários catalogarem e organizarem itens, possibilitando a reutilização do trabalho de outros bibliotecários e transformando a catalogação centrada em descrições baseadas em entidades. Vários trabalhos ao redor do mundo adotaram o paradigma de dados ligados para publicar coleções e exemplos, isso inclui organizações como a *Library of Congress* e a *British National Bibliography*. Entretanto, ainda existem etapas que precisam ser aprimoradas para alcançar a integração plena de dados bibliográficos na nuvem de dados ligados e abertos, bem como a interligação dos conjuntos de dados com os conjuntos de dados de outras bases (KONSTANTINO; SPANOS, 2015).

Um acervo de dados ligados é definido por uma coleção de triplas no formato RDF. Uma tripla nesse formato determina um sujeito (*subject*) (recurso), um objeto (*object*) (valor), e uma relação entre o sujeito e o objeto. A relação entre eles é caracterizada por um predicado (*predicate*), sendo que, uma tripla é decodificada por meio de uma sentença definida por um sujeito, um predicado e um objeto. Considerando que um KG é um grafo de dados, então um conjunto de dados ligados é um KG.

A principal tecnologia utilizada para realizar o mapeamento de dados ligados na Web semântica é o RDF. Ele contém o significado das estruturas e o codifica em conjuntos de triplas, cada uma sendo semelhante a formação de sujeito, verbo e objeto de uma frase elementar. Essas triplas podem ser escritas utilizando tags XML. Em RDF, um documento faz afirmações de que coisas específicas (pessoas, páginas da Web, ou o que for) têm características como “é uma irmã de” ou “é o autor de”, com certos valores para “outra pessoa” ou “outra página da Web”. Esta estrutura é uma maneira natural de descrever a maioria dos dados processados por máquinas. Cada sujeito e objeto são identificados por uma URI, assim como é usado em um link de uma página Web (URLs são os mais comuns tipos de URI). Os verbos são identificados por URIs, que permitem a qualquer um definir um novo conceito (um novo verbo) apenas definindo uma URI para o verbo em algum lugar da Web (BERNERS-LEE; HENDLER; LASSILA, 2002). Essa tecnologia permite realizar o mapeamento de dados ligados na Web semântica, pois as triplas RDF compõem uma rede de informações de coisas e a relação entre elas. Como o RDF utiliza URIs para compilar essas informações em um documento, as URIs afirmam que conceitos não são apenas palavras em um documento, mas que estão vinculados a um significado único ao serem encontrados na Web (BERNERS-LEE *et al.*, 2001).

Para Konstantinou e Spanos (2015), a maior parte do processamento de dados ligados corresponde à criação da fonte de dados registrados semanticamente, e não é realizada de maneira online. Eles explicam que as aplicações tradicionais dispõem de conjuntos de dados armazenados e processados manipulando diferentes tecnologias como bancos de dados relacionais, planilhas, arquivos XML, dentre outros, que precisam ser transformados em conjuntos de dados ligados e Abertos para tornarem-se publicados na Web (KONSTANTINOU; SPANOS, 2015).

Konstantinou e Spanos (2015) explicam como os dados ligados estão relacionados aos Dados Abertos, definindo que Dados Abertos são aqueles publicamente acessíveis por meio da internet e sem qualquer barreira física ou virtual para acessá-los. Já os dados ligados são aqueles que permitem que os relacionamentos sejam expressos entre esses dados. O RDF é ideal para representar *Linked Data*, porém não a única forma, como normalmente é estabelecido.

Dados Ligados é o paradigma que se adequa ao contexto de sistemas IoT³⁶, pois são facilmente integrados à Web e seus nós são identificáveis de maneira única. Ao se manifestar sobre a implantação dos dados ligados e Abertos relacionados às metodologias e ferramentas de *software*, Konstantinou e Spanos (2015) argumentam que *software* de visualização em dados ligados é importante porque as ferramentas que fazem parte desta categoria não precisam realizar suposições de conhecimento técnico, mas sim, ter como meta o usuário comum, uma vez que este pode ter conhecimento restrito ou nulo sobre dados ligados e o ecossistema relacionado.

Konstantinou e Spanos (2015) esclarecem que os dados Ligados têm seu valor fundamentado no significado de que o resultado do mapeamento é publicado com base nos princípios dos dados ligados, que está em todas as URIs usarem o esquema HTTP e, quando desreferenciadas, fornecerem informações importantes para o recurso que as identificam.

O futuro dos repositórios digitais está no emprego dos dados ligados. Isso significa que um repositório digital compatível com dados ligados possibilita a reutilização de conteúdo, concedendo também atividades de coleções individuais na nuvem global de dados ligados em progresso, bem como fornecendo ao usuário a oportunidade de encontrar novas fontes de dados no período em que o programa está em execução, percorrendo links de nível de dados e desse modo, oferecendo retornos mais completos quando novas fontes de dados surgem na Web (STEVENSON, 2011 *apud* KONSTANTINO; SPANOS, 2015).

3.3.5 Interoperabilidade

À medida que a acessibilidade aumenta, eleva-se também a necessidade de realizar a interoperabilidade e usufruir dos benefícios dessa. A interoperabilidade

³⁶ IoT é o acrônimo de *Internet of Things* (Internet das Coisas). É um paradigma tecnológico que proporciona a interconexão entre os mundos físico e digital, oportunizando um meio informacional que amplia a percepção dos indivíduos quanto ao cenário ao seu redor (ROZSA *et al.*, 2017a, p.255).

proporciona vários benefícios à sociedade contemporânea, e isso fica evidente quando Paganine *et al.* (2020) fundamentam que os padrões de interoperabilidade visam guiar as instituições de pesquisa a elaborar ou mesmo aperfeiçoar seus repositórios de dados científicos (PAGANINE, *et al.*, 2020).

O dicionário Priberam (2022) define interoperabilidade como “1. Qualidade do que é interoperável. 2. Capacidade de um sistema para interagir e comunicar com outro. (PRIBERAM, 2022)”. Wilkinson *et al.* (2016) complementa que a interoperabilidade é a capacidade de dados ou ferramentas de recursos não cooperativos serem capazes de incorporar ou funcionar em conjunto, com esforço mínimo.

Para Boeckhout, Zielhuis e Bredenoord (2018), o princípio da interoperabilidade determina que tanto os dados como os metadados devem ser conceituados, estruturados e expressos utilizando padrões de publicação comum entre eles. Assim, torna-se necessária a exploração de formatos de dados técnicos e semânticos que sejam padrões, variáveis, ontologias e similares.

A interoperabilidade está intimamente ligada à possibilidade de dados ou ferramentas de recursos compartilharem dados de bases diferentes, proporcionando o funcionamento cooperado entre bases diversas. Embora o funcionamento cooperado seja importante, Wilkinson *et al.* (2016) explica que uma boa administração de dados deve priorizar a possibilidade de as máquinas descobrirem e explorarem dados por meio de aplicações tecnológicas e padrões de interoperabilidade mais generalizados no nível de dados ou repositórios.

É sugerido o uso de **vocabulários** para apoiar a interoperabilidade. Como exemplo, pesquisas acerca da interoperabilidade no nível semântico indicam que as extensões das diretrizes OpenAIRE³⁷ são imprescindíveis para elaborar repositórios que satisfaçam aos princípios FAIR³⁸. Tanto que foi realizada uma extensão da FAIR

³⁷ OpenAIRE é a fusão da palavra *Open* com o acrônimo de *Access Infrastructure for Research in Europe*, e sua tradução do Inglês para o Português significa Infraestrutura de Acesso Aberto para Pesquisa na Europa. São diretrizes para Gerentes de Repositório de Literatura 4.0 que fornecem orientação para definir e implementar políticas locais de gerenciamento de dados com base nos requisitos OpenAIRE. Disponível em: <https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/introduction.html>

³⁸ FAIR é o acrônimo de *Findable, Accessible, Interoperable e Reusable*, e sua tradução do Inglês para o Português significa respectivamente Encontrável, Acessível, Interoperável e Reusável. Disponível em: <https://www.go-fair.org/fair-principles/>

Data Point (FDP) para adaptar as diretrizes OpenAIRE (WILKINSON *et al.*, 2016; PAGANINE *et al.*, 2020).

Hyland, Ateazing e Villazón (2014) apresentam o Vocabulário de Catálogo de Dados (VCD) (*Data Catalog Vocabulary* (DCAT)), um vocabulário RDF elaborado com o objetivo de simplificar a interoperabilidade entre os catálogos de dados disseminados na Web. Assim, os editores aprimoram a capacidade de descoberta e permitem aos aplicativos executar metadados de catálogos diversos usando VCD para representar conjuntos de dados em catálogos de dados. Esse mecanismo possibilita a disseminação descentralizada de catálogos, bem como simplifica a busca por conjuntos de dados federados em sites.

Dois, ou mais, sistemas são considerados interoperáveis quando ocorre troca de informações com êxito entre eles (USCHOLD; GRUNINGER, 2002 *apud* KONSTANTINOU; SPANOS, 2015). As abordagens que alcançam a interoperabilidade entre clientes de *software* distribuídos são categorizadas levando em consideração três aspectos: (1) Mapeamento entre conceitos de cada fonte, pois a descrição de informações de um determinado esquema transpõe os conceitos de uma fonte para outra; (2) Intermediação para a transposição de consultas, pois nessa perspectiva, uma camada intermediária é introduzida objetivando transpor as consultas com base em cada dado relacionado ao esquema da fonte. Essa camada pode ser, por exemplo, *software* adicional, conjunto de regras, ontologia, agente de *software*; (3) Abordagem fundamentada em consultas, pois o usuário tem a capacidade de realizar consultas que poderão ser avaliadas em cada fonte de dados (PARK; RAM, 2004 *apud* KONSTANTINOU; SPANOS, 2015).

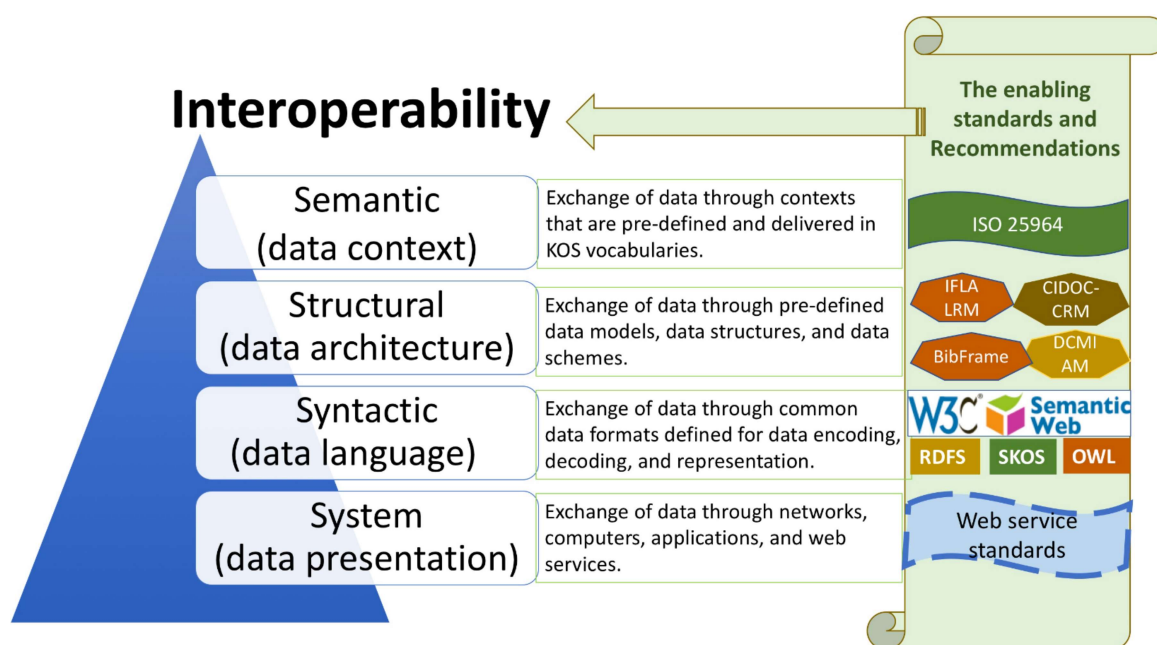
Konstantinou e Spanos (2015) argumentam que, embora a interoperabilidade seja utilizada como sinônimo de integração de informações, há diferença entre ambos. Eles explicam que interoperabilidade se refere à troca de informação bem-sucedida entre sistemas objetivando alcançar a integração entre eles, e não simplesmente a conformidade entre padrões para permitir a comunicação entre os sistemas. Já a integração de informação, no âmbito da Ciência da Computação, se refere ao processo pelo qual a informação, oriunda de diversas fontes e sistemas, é associada com o objetivo de permitir o seu processamento como um todo.

A interoperabilidade como troca de informação bem-sucedida entre dois ou mais sistemas para alcançar a integração entre eles pode ser obtida em níveis. Nesse

sentido, ela é composta por diferentes camadas, tais como: semântica (relacionada ao significado dos dados), estrutural (estrutura dos dados), sintática (linguagens e formatos) e sistêmica (redes, *softwares*, etc.) (Zeng, 2019). Embora existam diferentes camadas, ter interoperabilidade mais generalizada, no nível de dados ou repositórios, colabora diretamente com uma boa administração de dados. No entanto, pesquisas relacionadas à interoperabilidade de nível semântico indicam a necessidade de extensões orientadoras para a elaboração de repositórios que satisfaçam os princípios FAIR (WILKINSON *et al.*, 2016).

Nesse contexto, Zeng (2019) apresenta padrões e recomendações das melhores práticas que foram desenvolvidas no mundo, e para isso ela aborda 4 camadas da interoperabilidade, conforme identificado na Figura 5.

Figura 5 - As quatro camadas da interoperabilidade ([link para acesso em tamanho maior](#))



Fonte: Zeng (2019)

As camadas apresentadas por Zeng (2019) são:

- **Interoperabilidade Semântica**³⁹: Camada que trata da integração semântica, sendo norteadada pela comunicação de propósito coerente. Diversos contextos devem ser abordados, bem como as semelhanças, diferenças e os conceitos e relacionamentos devem ser compreendidos.

³⁹ A camada Interoperabilidade Semântica utiliza ontologias de domínio (por exemplo, owl:sameAs - equivalência de instâncias [valores de classes]) (CRISTOVÃO, 2021a).

Nessa camada são usados os KOS (*Knowledge Organization System*), que também são conhecidos por vocabulários KOS;

- **Interoperabilidade Estrutural**⁴⁰: Nessa camada os desafios de interoperabilidade estão nas alterações de arquitetura de informações em estruturas de dados, modelos de dados, estruturas de dados e esquemas. Adota-se estratégias para permitir a comunicação de dados por intermédio de estruturas pré-definidas, modelos conceituais estabelecidos por comunidades reconhecidas como biblioteca, arquivo e museu na era digital;
- **Interoperabilidade Sintática**⁴¹: Camada que trata dos problemas relacionados a qualquer esforço de interoperabilidade como as diferenças na codificação, decodificação e representação de dados. Os padrões de linguagem de dados fundamentais que permitem a comunicação de dados por meio de formatos de dados comuns dizem respeito às recomendações do W3C que foram elaboradas para a Web semântica;
- **Interoperabilidade Sistêmica**: Nessa camada, a interoperabilidade aborda questões quanto às incompatibilidades entre hardware e sistemas operacionais para a comunicação dos dados por intermédio de redes, computadores, aplicativos e serviços da web.

3.4 BOAS PRÁTICAS NA PUBLICAÇÃO DE DADOS NA WEB

Esta subseção aborda as boas práticas de publicação apresentando os princípios FAIR, as Boas Práticas na Publicação de Dados na Web e as Boas Práticas na Publicação de Dados Ligados.

3.4.1 Princípios FAIR

Para Wilkinson *et al.* (2016) há uma urgente necessidade para aprimorar a infraestrutura de auxílio à reutilização de dados, sendo que diversas organizações, tais como academia, indústria, agências de financiamento e editores acadêmicos, têm se reunido para projetar e apoiar uma coleção concisa e mensurável de fundamentos

⁴⁰ Interoperabilidade Estrutural utiliza vocabulários abarcando termos estruturais da base de dados, como vocabulários de metadados ou esquemas de metadados ou ontologias leves (por exemplo, Dublin Core (dc) - dc:title, dc:abstract, dc:created). Nessa categoria, os modelos conceituais são independentes de sintaxe de codificação e de aplicativos peculiares (CRISTOVÃO, 2021a).

⁴¹ Interoperabilidade Sintática usa vocabulários recomendados pelo W3C (como por exemplo: RDF - rdf:type, rdf:langString) (CRISTOVÃO, 2021a).

conhecidos como os princípios de dados FAIR. Os princípios FAIR podem ser uma forma poderosa de tornar verdadeiro o ideal de uma ciência mais aberta, o que estimula pesquisadores a abordar questões quanto ao gerenciamento de dados e acessibilidade mais ampla, com a finalidade de apresentar certa capacidade de reutilização (BOECKHOUT; ZIELHUIS; BREDENOORD, 2018). A ideia de utilização dos princípios FAIR é que possam servir de orientação para aprimorar a reutilização de acervos de dados (WILKINSON *et al.*, 2016).

O nome FAIR corresponde ao acrônimo dos termos *Findability* (facilidade de localização), *Accessibility* (acessibilidade), *Interoperability* (interoperabilidade), e *Reusability* (capacidade de reutilização).

FAIR foi introduzido no início de 2014 e é apresentada como um conjunto de princípios orientadores e de práticas mínimas para o gerenciamento de dados de pesquisa. Desde então, esses princípios impulsionaram rapidamente a pesquisa e a política de pesquisa (BOECKHOUT; ZIELHUIS; BREDENOORD, 2018).

Com finalidade de impulsionar a transposição linear de conhecimento para a transição mais dinâmica da circulação de conhecimento, especialistas acreditam que a criação e o apoio da inovação aberta de um ecossistema são essenciais para facilitar a tradução de conhecimento em valor socioeconômico. Além dos componentes formais do lado da oferta, tais como habilidades de pesquisa, ciência de excelência, financiamento e gestão de propriedade intelectual, existe ainda certa necessidade em se dedicar, no lado da demanda, quanto aos aspectos da circulação do conhecimento, garantindo que o trabalho científico cumpra às necessidades dos usuários e que esse conhecimento seja encontrável, acessível, interoperável e reutilizável (FAIR) (EUROPÄISCHE KOMMISSION, 2016). Wilkinson *et al.* (2016) argumentam que o ecossistema digital existente impede que a publicação de dados possa extrair o máximo de benefício na pesquisa, porém, como uma resposta parcial a essa questão, financiadores científicos, editores e agências governamentais investem no gerenciamento de dados e planos de administração de dados gerados em experimentos com financiamento público, por exemplo.

Há benefícios esperados com a superação de problemas como encontrabilidade, acessibilidade, interoperabilidade e reusabilidade na recuperação de dados e informação. Assim, a superação desses problemas pode beneficiar os pesquisadores que almejam compartilhar, obter credibilidade e reutilizar os dados e

interpretações uns dos outros; profissionais editores de dados que disponibilizam seus serviços; *software* e produtores de ferramentas que oferecem análise de dados e serviços de processamento nos fluxos de trabalho reutilizáveis; agências privadas e públicas preocupadas com o gerenciamento de dados a longo prazo e na comunidade de ciência de dados que integra e analisa dados novos e existentes para avançar em descobertas (WILKINSON *et al.*, 2016).

De acordo com Guizzardi (2020) os princípios orientadores do FAIR têm a interoperabilidade como atributo principal para aumentar o valor agregado dos artefatos de informação. Ele explica que a interoperabilidade em FAIR só é possível com o suporte de estruturas de informação que sejam ontologicamente consistentes e que estabeleçam compromissos ontológicos explícitos. O autor argumenta que para alcançar a interoperabilidade mencionada, precisa-se mais do que vocabulários, mas de uma ontologia de domínio.

No Brasil, um plano de gestão de dados fundamentado nos princípios FAIR para a liberação de recursos tem sido implementado por algumas instituições, tais como, a Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP). Devido ao potencial desses princípios, o Instituto Brasileiro de Informações em Ciência e Tecnológica (IBICT) já possui acordo com o GO-FAIR⁴², objetivando adotar esses princípios no Brasil (HENNING, 2019). Entretanto, para possibilitar a aplicação desses princípios no atual ecossistema digital é preciso que os dados estejam modelados.

Conforme visto, esse conjunto de princípios orientadores e práticos ajudam na encontrabilidade, acessibilidade, interoperabilidade e reusabilidade relacionados aos dados e informações. Para que isso ocorra por intermédio do processo de busca na Web, existem mecanismos para definir, organizar, disciplinar, categorizar, representar, modelar e qualificar tanto os dados como as informações.

3.4.2 Boas Práticas na Publicação de Dados na Web

Lóscio, Burle e Calegari (2017) documentaram sugestões de Melhores Práticas para Dados na Web (MPDW) (*Data on the Web Best Practices* (DWBP)). Essa

⁴² GO-FAIR é uma iniciativa *bottom-up* (de baixo para cima) para interessados e autogovernados que visam implementar os princípios de dados FAIR, isto é, tornar os dados Encontráveis, Acessíveis, Interoperáveis e Reutilizáveis (FAIR). Disponível em: <https://www.go-fair.org/>

proposta foi adotada como recomendação pela W3C⁴³ e oferece uma coleção de práticas sobre o uso e a publicação de dados na Web (LÓSCIO, BURLE E CALEGARI, 2017). Penteado (2020) apresenta que um dos princípios da MPDW está na forma de consumo de dados por humanos e máquinas, embora trate-se de um mecanismo difícil de implementar. Ele explica que é um desafio fazer com que dados e metadados estejam disponíveis nessas duas maneiras de consumo: por máquinas e por humanos. A forma de consumo por humanos está relacionada ao **catálogo de dados**, enquanto a maneira de consumo por máquinas está associada a triplestore⁴⁴. Segundo Bertolini *et al.* (2023), a Federação Internacional de Associações e Instituições Bibliotecárias (*International Federation of Library Associations and Institutions* (IFLA)) estabelece os princípios de catalogação de boas práticas voltadas ao usuário humano, levando em consideração aspectos como localização, identificação, seleção, aquisição, navegação (RIVA; BOEUF; ŽUMER, 2017; BERTOLINI *et al.*, 2023):

- **Localização:** Refere-se à capacidade de encontrar todas as manifestações possíveis de expressões de uma obra. É localizar todas as expressões de uma obra que estão escritas em um idioma.
- **Identificação:** Significa reconhecer um recurso entre os resultados de uma busca. Trata-se do recurso que possibilita materializar a pesquisa sobre uma obra, ainda que a pesquisa tenha sido realizada por um título diferente da obra buscada pelo usuário. É a capacidade de identificar um recurso de uma pesquisa levando em conta os meios destinados a um público ou um propósito específico.
- **Seleção:** Significa escolher os recursos identificados da obra. Tem o sentido de selecionar recursos de uma obra pesquisada, incluindo conteúdos adicionais relevantes, contribuição secundária por agentes específicos e disposto nos formatos mais convenientes ao propósito do usuário.

⁴³ W3C corresponde a *World Wide Web Consortium* (Consórcio da Rede de Alcance Mundial). É uma comunidade internacional de desenvolvimento de padrões para Web (W3C, 2021).

⁴⁴ Triplestore ou RDF refere-se a um banco de dados elaborado especificamente para armazenamento e recuperação em triplas (entidade composta por sujeito-predicado-objeto) por intermédio de consultas semânticas. Disponível em: <https://dbpedia.org/page/Triplestore>

- Aquisição: Significa obter um recurso ou informações como data, localização e país por meio de download (usando link online), empréstimo físico ou compra no catálogo da biblioteca;
- Navegação: significa explorar os relacionamentos para compreender a estrutura do domínio do assunto bem como a sua terminologia. É compreender os relacionamentos de diversos nomes para uma instância de entidade. Indica que um catálogo deve apresentar claramente as relações entre as entidades para além do catálogo, inclusive para contextos não bibliotecários.

Como o objetivo principal da MPDW é fazer com que os dados sejam compreendidos e encontrados por humanos e máquinas, isso auxilia a interatividade entre consumidores e produtores de dados, proporcionando dados precisos, atualizados e disponíveis a qualquer tempo. Isso colabora com a propagação contínua da Web como ambiente para a transferência de dados e promove benefícios tais como o reuso dos dados, o mecanismo de interligar um fato a outro, a descoberta de recursos e a elaboração de visualizações interativas (LÓSCIO, BURLE E CALEGARI, 2017).

A MPDW ocorre em um contexto de dados na Web, e para isso, ela conta com certos elementos como dados ligados, vocabulários e ontologias. Diante disso, Penteado (2020, p.189) argumenta que um dos pontos principais para a publicação de dados ligados está no reaproveitamento de vocabulários e ontologias já existentes para o domínio que se deseja representar.

Alguns dos exemplos de portais do crescimento e uso da Web para publicação de dados e do avanço no compartilhamento de dados abertos em todo planeta, segundo Lóscio, Burle e Calegari (2017) são:

- *Global Open Data Index (GODI)*⁴⁵: Corresponde ao portal Índice Global de Dados Abertos (IGDA). Proporciona o acompanhamento do estado dos dados abertos governamentais como medida mundial de como os governos estão usando e publicando os dados abertos para inovação, impacto social e responsabilidade (OPEN KNOWLEDGE FOUNDATION, 2016).

⁴⁵ Disponível em: <http://index.okfn.org/>

- *Open Data Barometer* (ODB)⁴⁶: Equivale ao portal Barômetro de Dados Abertos (BDA). Disponibiliza como os governos estão usando e publicando dados com responsabilidade, inovação e impacto social, e com base nos Princípios de Dados Abertos Anticorrupção (PDAA) do G20 (*OPEN DATA BAROMETER*, 2008).
- *Research Data Alliance* (RDA)⁴⁷: Condiz com o portal Aliança de Dados de Pesquisa (ADP). É um exemplo de progresso da publicação online de dados por meio de um portal de pesquisa promovido pela RDA (*RESEARCH DATA SHARING WITHOUT BARRIERS*, 2013), que tem como objetivo construir uma infraestrutura social e técnica para possibilitar o compartilhamento aberto e a reutilização de dados.
- *Bibliothèque Nationale de France* (Biblioteca Nacional da França (BNF))⁴⁸: É um portal com mecanismos de coleta, análise e publicação de dados online de mídias sociais, *Crowdsourcing* (contribuição colaborativa) de informações, dentre outros. É um portal que enfatiza a propagação na Web das principais coleções de patrimônio cultural (*ACCUEIL*, 2014).
- *Linked Open Data Cloud* (LODC)⁴⁹: Corresponde a Dados Ligados e Abertos em Nuvem (DLAN). É um portal exemplar quanto ao crescimento do uso da Web para a publicação de dados. Ele possui atualmente 1.255 conjuntos de dados com 16.174 links.

Para Lóscio, Burle e Calegari (2017), é importante que tanto humanos quanto agentes de *software* entendam e processem facilmente os dados que necessitam ser atualizados e descobertos na Web. Assim, de acordo com os autores, as melhores práticas para publicar e consumir dados na Web contém ao menos um dos seguintes requisitos e componentes:

- **Uso de Turtle ou JSON-LD**⁵⁰ são exemplos de algumas das melhores aplicações práticas de RDF⁵¹. Tanto *Turtle* como JSON-LD são exemplos

⁴⁶ Disponível em: https://opendatabarometer.org/?_year=2017&indicator=ODB#

⁴⁷ Disponível em: <https://rd-alliance.org/>

⁴⁸ Disponível em: <https://data.bnf.fr/>

⁴⁹ Disponível em: <https://lod-cloud.net/>

⁵⁰ JSON-LD é um formato para dados ligados de fácil compreensão para leitura e escrita de humanos. Disponível em: <https://json-ld.org/>

⁵¹ RDF corresponde a *Resource Description Framework* (Estrutura de Descrição de Recursos). É um *framework* para representar informações na Web (WOOD; LANTHALER; CYGANIAK, 2014). Disponível em: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

de boas práticas, pois permitem leitura e escrita fáceis por humanos, além de serem formatos de dados para diversos ambientes de programação em RDF para máquinas.

- **Metadados** são requisitos fundamentais em um espaço de informação aberto como a *Web*. Eles devem fornecer informações que auxiliem consumidores de dados a entender melhor o significado dos dados, sua estrutura, direitos e termos de licença, a organização que gerou os dados, a qualidade dos dados, os métodos de acesso e o cronograma de atualização dos dados. Em suma, os metadados devem ser utilizados com a finalidade de auxiliar tarefas que envolvem a descoberta e reutilização de conjuntos de dados, e eles podem ser de diferentes tipos, ou seja, podem ser classificados em diferentes taxonomias e com diferentes critérios de agrupamento;
- **Licença de Dados** diz respeito a uma informação que deve ser anexada a dados na *Web* quando se deseja evidenciar a existência de restrições tanto no compartilhamento como na reutilização de dados. Trata-se de uma licença que fornece permissão oficial para realizar algo com os dados aos quais está relacionada;
- **Proveniência dos Dados** o termo proveniência é originado do francês *provenir* que significa “vir de”, e é utilizado para descrever o processo de curadoria em obras de arte conforme a arte é passada de autor para autor. De maneira semelhante, a proveniência dos dados são metadados que permitem aos provedores de dados fornecerem detalhes quanto ao histórico dos dados. Determina uma maneira própria dos consumidores de dados confiarem na integridade e credibilidade dos dados compartilhados na *Web*;
- **Qualidade dos Dados** é a “adequação ao uso” para um aplicativo ou caso de uso peculiar. Diz respeito a sua avaliação e engloba diferentes dimensões de qualidade, sendo que cada uma delas representa grupos de características que são relevantes para editores e consumidores de dados. Para avaliar a qualidade dos dados em cada dimensão de qualidade, usa-

se o Vocabulário de Qualidade de Dados (VQD)⁵², pois estabelece conceitos tanto de medidas como de métricas (ALBERTONI; ISAAC, 2016);

- **Versão dos Dados** corresponde aos conjuntos de dados disseminados na Web que são atualizados de maneira programada, ou mesmo alterados quando melhorias fazem parte da coleta de dados;
- **Identificadores de Dados** são especificidades globalmente exclusivas que podem ser consultadas por intermédio de desreferenciação na Internet. Exemplos de Identificadores de dados são URI⁵³s, HTTP⁵⁴ (ou HTTPS⁵⁵), utilizados fundamentalmente na descoberta, uso ou citação de dados na Web. Ou seja, são especificidades de formas diversas que fazem parte de todos os sistemas de informação;
- **Formatos de Dados** é um padrão específico para representação de dados. Trata-se da forma como a informação é codificada e armazenada para ser utilizada em um sistema de computador, geralmente delimitada por um tipo de dado formal ou um conjunto de padrões. Tem o seu uso motivado pela W3C e utilizados pelo público em geral para que também possam ser processados por sistemas computacionais. Formatos de dados padronizados proporcionam a interoperabilidade e a reusabilidade, e quando são padronizados e legíveis por máquina, são neutros quanto a localidade;
- **Vocabulários de Dados** definem os conceitos e os relacionamentos (conhecidos como “termos” ou “atributos”). Eles são aplicados com a finalidade de descrever e representar uma área de domínio. Assim, eles são empregados na classificação dos termos que fazem parte de um aplicativo

⁵² VQD corresponde a Vocabulário de Qualidade de Dados (*Data Quality Vocabulary* (DQV)). É uma extensão de Vocabulário de Catálogo de Dados (*Data Catalog Vocabulary* (DCAT)) usada para melhorar a qualidade dos dados. Aceita-se correções de usuário, é utilizado por editores e trata-se de um vocabulário adotado com frequência por desenvolvedores. Disponível em:

<https://www.w3.org/TR/vocab-dqv/>

⁵³ URI é o acrônimo de *Uniform Resource Identifier*, e sua tradução do Inglês para o Português significa Identificador Uniforme de Recurso. Trata-se de uma sequência comprimida de caracteres que identifica um recurso físico ou abstrato. Disponível em:

<https://philarcher.org/diary/2013/uripersistence/>

⁵⁴ HTTP é o acrônimo de *Hypertext Transfer Protocol*, e sua tradução do Inglês para o Português significa Protocolo de Transferência de Hipertexto. Trata-se do protocolo de comunicação para acesso aos documentos. Disponível em: <https://datatracker.ietf.org/doc/html/rfc7230>

⁵⁵ HTTPS é o acrônimo de *Hypertext Transfer Protocol Secure*, e sua tradução do Inglês para o Português significa Protocolo de Transferência de Hipertexto Seguro. Disponível em:

<https://datatracker.ietf.org/doc/html/rfc7230>

específico, e representam possíveis relacionamentos ou definem possíveis restrições ao uso desses termos. Vários sinônimos para “vocabulário” foram definidos, como por exemplo: Ontologia, Vocabulário Controlado, Tesouro, Taxonomia, Lista de Códigos, Rede Semântica;

- **Acesso de Dados** trata da capacidade de permitir que humanos e máquinas possam aproveitar os benefícios do compartilhamento de dados usando a infraestrutura da Web. O ato ou o efeito de fornecer acesso fácil a dados na Web permite que humanos e máquinas aproveitem os benefícios do compartilhamento de dados usando a infraestrutura da Web. Por padrão, a Web oferece acesso usando métodos HTTP (*Hypertext Transfer Protocol*). Isso fornece acesso aos dados em um nível de transação atômica, e pode ocorrer por intermédio do download em massa simples de um arquivo ou, onde os dados são distribuídos em vários arquivos ou exigem métodos de recuperação mais sofisticados, por meio de uma API⁵⁶. Os dois métodos básicos, download em massa e API, não são mutuamente exclusivos;
- **Preservação de Dados** são processos e operações que visam garantir a permanência técnica e intelectual de objetos ao longo do tempo. Foi introduzido pela *Alliance for Permanent Access Network* (Aliança para Rede de Acesso Permanente) e trata-se de um plano de gerenciamento de dados com foco no planejamento de preservação e metadados (DPGLOSSARY, 2015). Indica se os dados foram removidos ou arquivados. Esse item é relevante, pois é complicado supor que os dados na Web estarão disponíveis permanentemente;
- **Feedback** (retroalimentação/comentário) favorece aos editores e consumidores, pois além de ajudar os editores a melhorar a integridade de seus dados publicados, incentivam a publicação de novos dados. A retroalimentação permite que os consumidores de dados descrevam suas experiências de uso (por exemplo, em aplicativos que usam dados), preferências e necessidades. É esperado que o feedback esteja disponível publicamente para outros consumidores de dados examinarem;

⁵⁶ API é o acrônimo de *Application Programming Interface* (Interface de Programação de Aplicação). É um conjunto de rotinas e padrões que favorecem a comunicação e troca de informações entre sistemas. Disponível em: <https://www.ibm.com/cloud/learn/api>

- **Enriquecimento de Dados** diz respeito a um conjunto de processos que tem a finalidade de aprimorar, refinar ou melhorar dados brutos ou processados anteriormente. Essa e outras ideias equivalentes colaboram para tornar os dados um ativo valioso para qualquer campo de domínio;
- **Republicação** trata do ato ou do efeito de reutilização dos dados. Ela geralmente combina dados existentes com outros conjuntos de dados, cria aplicativos ou visualização da Web ou reempacota os dados em um novo formato.

3.4.3 Boas Práticas na Publicação de Dados Ligados

Hyland, Ateazing e Villazón (2014) explicam que a modelagem de dados ligados envolve dados que são trocados entre modelos diferentes. Eles exemplificam que a modelagem de dados ligados pode tratar da transposição da representação tabular de dados em uma representação baseada em grafo. Modelos de Bancos de Dados Relacionais podem ser convertidos em dados ligados para integrar conjuntos de dados de diferentes bases ou com outros conjuntos de dados de código aberto. Nesse processo, os interessados são incentivados a definir como os objetos estão relacionados, sendo que o especialista no assunto registra como os vários objetos estão relacionados ao usar vocabulários padrão sempre que for possível. Em dados ligados, o esquema de dados é representado com os próprios dados. Esse mecanismo difere da abordagem relacional, no qual documentos externos e diagramas descrevem os dados. Isto é, enquanto nos dados ligados baseia-se no uso de padrões internacionais da Web aberta para a troca de dados, a abordagem relacional utiliza-se de dicionários de dados, diagramas de entidade-relacionamento, esquemas lógicos, sem a preocupação do uso de padrões internacionais (HYLAND; ATEMEZING; VILLAZÓN, 2014).

Antes do estabelecimento de padrões internacionais de transferência de dados na Web, era complicado e vagaroso elaborar aplicações utilizando as técnicas tradicionais de gerenciamento de dados. Assim, conforme os dados governamentais abertos estão sendo publicados na Web, as melhores práticas para os dados também estão avançando na Web (HYLAND; ATEMEZING; VILLAZÓN, 2014).

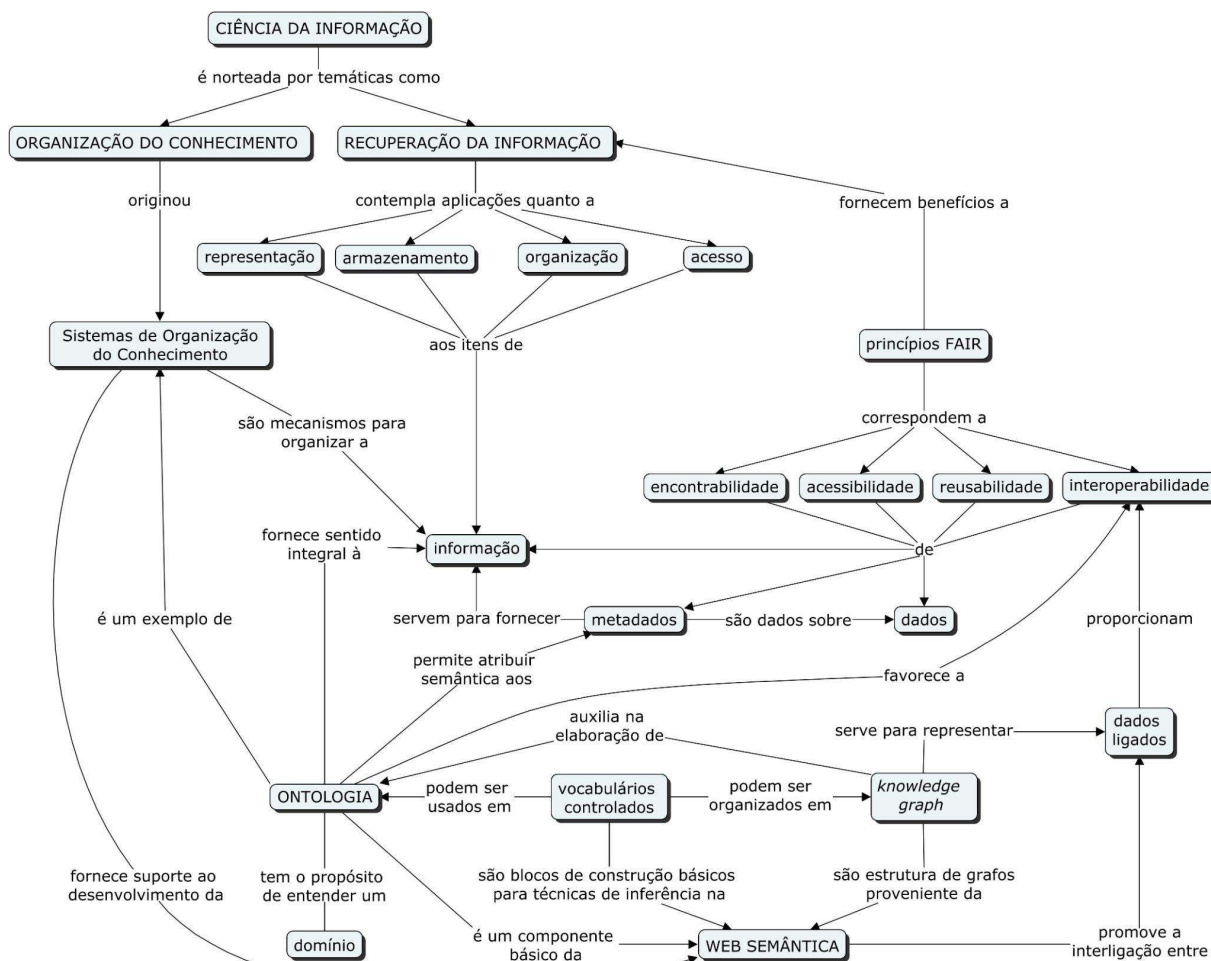
3.5 CONSIDERAÇÕES FINAIS DA FUNDAMENTAÇÃO TEÓRICA

Foram abordados neste capítulo os fundamentos relacionados à OC no âmbito da CI, destacando o conceito de OC e os desdobramentos da publicação dos dados relacionados aos SOC. Em seguida, foi trabalhada a contextualização da RI no âmbito da CI, apresentando a definição dos princípios norteadores FAIR e Web semântica, além de expor seus principais componentes operacionais que dizem respeito aos vocabulários controlados, ontologias, dados ligados e interoperabilidade para a publicação de dados na Web.

Pôde-se observar que a OC na CI aborda os SOC como mecanismos para organizar a informação, que também é um dos itens resultantes das aplicações de recursos como representação, armazenamento, organização e acesso da RI na CI. No contexto da CI, foi apresentado que uma ontologia tem o propósito de classificar termos, bem como criar vocabulários controlados que podem ser organizados em *knowledge graphs*. Para isso, utiliza-se mecanismos da Web semântica como dados ligados para proporcionar interoperabilidade de informação e que está entre os elementos mais importantes da RI e faz parte dos princípios norteadores FAIR.

Parte dos elementos que foram abordados neste capítulo são importantes para o cumprimento dos dois primeiros objetivos específicos da dissertação e estão apresentados na Figura 6.

Figura 6 - Mapa Conceitual contextualizando alguns elementos da Fundamentação Teórica ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

A Figura 6 apresenta um mapa conceitual que contextualiza diversos elementos abordados na fundamentação teórica com algumas relações entre os conceitos abordados. A Ciência da Informação é norteada pelas temáticas como Organização do Conhecimento (OC) e Recuperação da Informação (RI). Os Sistemas de Organização do Conhecimento (SOC) foram originários de OC, fornecem suporte ao desenvolvimento da Web semântica, e também são mecanismos para organizar a informação. A RI contempla aplicações quanto a representação, o armazenamento, a organização e o acesso aos itens de informação. Ontologia é um exemplo de SOC. Ela é um componente básico da Web semântica e tem o propósito de entender um domínio.

O *knowledge graph* serve para auxiliar na elaboração de ontologias. É ténue a divisão entre ontologia e *knowledge graph* do ponto de vista da implementação de ontologia de domínio (discutido nas subseções 3.2 e 3.3.3). O *knowledge graph* é um

grafo de dados, e um conjunto de dados ligados é considerado um grafo de conhecimento que ocorre por meio de triplas (representado em RDF). Já a ontologia não necessariamente é representada por intermédio de triplas, ou seja, ela é uma especificação formal, de conceituação explícita, disciplinada e objetiva compreender um domínio.

4 FUNDAMENTAÇÃO METODOLÓGICA

Este capítulo trata dos fundamentos conceituais e metodológicos para preparar o leitor ao entendimento dos elementos apresentados nos próximos capítulos que abordam os procedimentos metodológicos e os resultados com as discussões, explicando também o motivo da escolha das ferramentas computacionais adotadas. Ele é organizado em duas seções. A primeira trata de elementos do contexto da organização da pesquisa, e a segunda dos elementos do contexto do desenvolvimento.

4.1 ORGANIZAÇÃO GERAL DA PESQUISA

A fundamentação metodológica da organização geral da pesquisa apresenta os elementos efetivamente usados para auxiliar a organização da pesquisa, abordando e contextualizando elementos como tecnologias, ferramentas, dentre outros. A presente seção é dividida em duas subseções que abordam: os mapas conceituais e o gerenciador bibliográfico.

4.1.1 Mapas Conceituais

O conhecimento pode ser sistematizado por meio de mapas conceituais. No âmbito da CI, o ato ou o efeito de representar algo está diretamente relacionado com as diversas formas de simbolizar ou mesmo registrar símbolos que substituem coisas ou idéias (LIMA; ALVARES, 2012 *apud* RODRIGUES; CERVANTES, 2017, p. 104). Nesse sentido, Novo (2013, p.16 *apud* RODRIGUES; CERVANTES, 2017, p. 104) explica que a representação conceitual diz respeito à relação entre as questões, o que implica que os conceitos fazem parte de diferentes domínios do conhecimento e respondem a um determinado problema.

Pelizzari *et al.* (2002) conceitua que mapas conceituais são:

[...] instrumentos que permitem descobrir as concepções equivocadas ou interpretações não aceitas (podem não ser errôneas) de um conceito, ilustradas por uma frase que inclui no conceito. Devem ser hierárquicos, quer dizer, os conceitos mais gerais devem situar-se na parte superior, e os conceitos mais específicos e menos inclusivos na parte inferior. Também podem ser considerados instrumentos úteis para negociar significados, quer dizer, os alunos sempre trazem alguma coisa deles mesmos para a negociação. Não são como uma

tábua rasa ou um recipiente vazio que o professor deve preencher (PELIZARI *et al.*, 2002, p.41).

Para Novak e Cañas (2008), mapas conceituais são definidos como ferramentas gráficas que servem para organizar e representar o conhecimento. Eles possuem conceitos e relações entre os termos indicados. As linhas que ligam os conceitos são chamadas de frases ou palavras de ligação, sendo que elas determinam a relação entre os dois ou mais conceitos. Quanto ao conceito, este é estabelecido por intermédio de eventos, objetos definidos por um rótulo. Já o rótulo dos conceitos geralmente é representado por uma palavra.

O mapa conceitual é um instrumento proposto por Joseph Donald Novak em 1972 (NOVAK; CAÑAS, 2008), baseado na Teoria da Aprendizagem Significativa de David Ausubel (1968). Ele é uma ferramenta que serve para organizar, representar e compartilhar o conhecimento (NOVAK, 1977; SHERRATT; SCHLABACH, 1990; CAÑAS *et al.*, 2004 *apud* CRISTOVÃO, 2016).

Os mapas conceituais são ferramentas que auxiliam na representação do conhecimento e foram usados para representar e sistematizar alguns assuntos apresentados e discutidos na fundamentação teórica desta pesquisa. Eles foram aplicados para elaborar a representação dos conceitos, *knowledge graph*, organizar e adaptar figuras de ontologia nas áreas de conhecimento (subseção 3.3.2) e o fechamento de capítulo (subcapítulo 3.4). É importante ressaltar que esse instrumento foi utilizado também para auxiliar na representação da elaboração inicial de ontologias da fase de projeto, bem como esboçar os fechamentos dos capítulos. Este instrumento foi usado para realizar a Etapa 1 que corresponde a concepção da modelagem conceitual relacionada a ontologia de domínio.

O CmapTools⁵⁷ é um editor desenvolvido pelo *Institute for Human and Machine Cognition* - IHMC (Instituto de Cognição Humano-Máquina). Ele foi utilizado para elaborar os mapas conceituais utilizados nesta pesquisa. Para Novak e Cañas (2010, p.17), CmapTools é um *software* que:

[...] torna fácil para usuários de todas as idades elaborarem e modificarem mapas conceituais da mesma forma que um processador de texto facilita a produção textual. Ele permite aos usuários trabalharem juntos à distância na elaboração de seus mapas; publicarem seus mapas conceituais para o acesso de qualquer pessoa

⁵⁷ CmapTools é um *software* que serve para construir, navegar e compartilhar modelos de conhecimento representados como mapas conceituais, mapas mentais, dentre outros. Disponível em <https://cmap.ihmc.us/cmaptools/>

conectada à internet; fazerem links para fontes externas em seus mapas para melhor explicarem seus conteúdos; e buscarem informações relacionadas ao mapa na Rede Mundial de Computadores (NOVAK; CAÑAS, 2010, p.17).

O CmapTools incorpora um elenco de recursos que tem como finalidade auxiliar o desenvolvimento e publicação de mapas conceituais interligados e recursos relacionados. O *software* apoia na construção e navegação de certa publicação e isso significa possibilitar a realização de mapas por intermédio do navegador. O autor argumenta que a maior parte das pessoas têm maior interesse na busca por informações do que criá-las, e essa diferença impacta no ramo da Educação, pois há divergência considerável do ponto de vista pedagógico entre alunos construtores de conhecimento e alunos consumidores de informação na web.

Esta ferramenta foi escolhida para representar os artefatos da dissertação como novos conceitos da Fundamentação Teórica, ontologias da fase de projeto, o esboço dos fechamentos dos capítulos e as considerações finais porque é uma ferramenta gratuita, de manuseio simples, que permite publicação gratuitamente das produções de mapas conceituais nos servidores públicos Cmap, bem como o fato permitir criar arquivos colaborativos que podem ser editados simultaneamente. Os arquivos com os mapas conceituais elaborados estão disponíveis no servidor público do *Institute for Human and Machine Cognition* (IHMC)⁵⁸.

4.1.2 Gerenciador Bibliográfico

Foi realizada uma análise preliminar sobre os *softwares* gerenciadores bibliográficos EndNote⁵⁹, Mendeley⁶⁰, Zotero⁶¹ (YAMAKAWA, *et al.*, 2014) e RefWorks⁶² (KRATOCHVÍL, 2017; IVEY; CRUM, 2018).

EndNote é um *software* de gerenciamento de referências que auxilia pesquisadores a conduzir e compartilhar pesquisas (ENDNOTE, 2023). É um *software* usado para buscar e organizar referências bibliográficas de maneira sistemática

⁵⁸ Disponível em: <https://cmapspublic.ihmc.us/>

⁵⁹ Disponível em <https://endnote.com/product-details>

⁶⁰ Disponível em <https://www.elsevier.com/solutions/mendeley>

⁶¹ Disponível em <https://www.zotero.org/>

⁶² Disponível em <https://www.refworks.com/refworks2/default.aspx>

(YAMAKAWA, *et al.*, 2014), contudo, ele está disponível apenas para os Sistemas Operacionais macOS⁶³ e Windows⁶⁴ (IVEY; CRUM, 2018).

Assim como EndNote, o Mendeley é um gerenciador de referências que propicia hospedar materiais de pesquisa, proporciona compilar e organizar referências, bem como possibilita extrair metadados e disponibilizá-los em redes sociais (ELSTON, 2019, p.1071). Ele permite ser usado em conjunto com editor de texto, é gratuito (ELSTON, 2019; MENDELEY, 2022; YAMAKAWA, E. K. *et al.*, 2014), possui compatibilidade com Sistemas Operacionais diferentes como macOS, Linux⁶⁵, Windows (MENDELEY, 2022), possibilita a realização de backups com armazenamento em nuvem (cópia de segurança), facilidade de uso e possibilita certa melhora na performance do desenvolvimento da pesquisa no que tange às citações e a realização de referências (ELSTON, 2019; MENDELEY, 2022; YAMAKAWA, E. K. *et al.*, 2014).

Assim como os gerenciadores de referências mencionados, o RefWorks permite criar bancos de dados pessoais e usá-los com várias finalidades em atividades de pesquisa como: gerenciar, armazenar e compartilhar as informações, bem como formatar bibliografias e textos de maneira automatizada, aumentando assim a performance na elaboração de pesquisas e diminuindo o número de erros na formatação de referências bibliográficas (REFWORKS, 2009). Ele não é aplicável a um Sistema Operacional⁶⁶ pois está disponível somente na Web, e ele é compatível com os navegadores Internet Explorer, Microsoft Edge, Firefox, Safari e Chrome. Após estudos e testes entre todos os gerenciadores de referência mencionados, Kratochvíl (2017) sugere que é importante experimentá-los antes de se estabelecer uma escolha definitiva de apenas um deles.

O Zotero é um *software* livre de código aberto para organizar referências bibliográficas (YAMAKAWA, *et al.*, 2014; KRATOCHVÍL, 2017; IVEY; CRUM, 2018;

⁶³ Trata-se do Sistema Operacional da Apple. Disponível em:

https://support.apple.com/pt_BR/downloads/macOS

⁶⁴ É o Sistema Operacional da Microsoft. Disponível em: <https://www.microsoft.com/pt-br/software-download/windows11>

⁶⁵ É um Sistema Operacional de computador pertencente ao movimento de código aberto (*open-source*) (TANENBAUM; BOS, 2016; DEITEL; DEITEL, 2016). Disponível em: <https://www.linux.org/pages/download/>

⁶⁶ Sistema Operacional é um dispositivo de *software* que tem a função tornar a utilização de computadores mais simples para serem manuseados por parte dos usuários (TANENBAUM; BOS, 2016).

CORPORATION FOR DIGITAL SCHOLARSHIP, 2020; ZOTERO, 2023a). Ele está disponível para os Sistemas Operacionais macOS, Windows e Linux (YAMAKAWA, *et al.*, 2014; CORPORATION FOR DIGITAL SCHOLARSHIP, 2020), bem como permite utilização por meio do navegador com armazenamento em nuvem (IVEY; CRUM, 2018; ZOTERO, 2023a). Este *software* foi escolhido para organizar as referências bibliográficas da dissertação por possuir um repositório com mais de 10 mil estilos de citação (ZOTERO, 2023b) e ser uma recomendação de Kratochvíl (2017) para quem busca uma solução gratuita e consolidada (ZOTERO, 2023a). Ele permitiu organizar a literatura científica especializada da temática, funcionando como um gerenciador de referências bibliográficas e contribuindo significativamente para obter mais produtividade na elaboração dos textos escritos. Fácil de usar, ele disponibiliza mecanismos de armazenamento em nuvem que ajudaram na coleta, organização, citação, compartilhamento e backup das pesquisas consultadas (CORPORATION FOR DIGITAL SCHOLARSHIP, 2020).

4.2 DESENVOLVIMENTO

O desenvolvimento da pesquisa abarcou tecnologias, ferramentas, linguagens e ambientes computacionais que serão apresentados nas próximas subseções a saber: modelagem conceitual baseada em ontologia, *design* e implementação da ontologia operacional de domínio, mapeamento para base de dados ligados RDF, e consultas RDF.

4.2.1 Modelagem conceitual baseada em ontologia

Essa subseção apresenta as metodologias para elaboração de ontologias de domínio e suas ferramentas; discorre sobre a importância das ontologias de fundamentação; denota a linguagem de modelagem orientada a ontologia usada, bem como manifesta a adoção de uma ferramenta gráfica para modelagem de ontologias.

4.2.1.1 Metodologias para construção de ontologias de domínio

Existem diversas pesquisas que adotam estratégias, linguagens e ferramentas para elaborar ontologias ou processos baseados em ontologia para especificação de domínio como abordado por Silva, Mutz e Ruy (2022a), Silva, Mutz e Ruy (2022b), Farinelli e Elkin (2019), Detoni (2019), Campos (2019), Reginato *et al.* (2019),

Salamon (2018), Teixeira (2017), Smith (2015), Teixeira, Falbo e Guizzardi (2014), Falbo (2014), Guizzardi *et al.* (2011), Suárez-Figueroa (2010), Fernandez *et al.* (1997), Guizzardi (2006), Guizzardi e Wagner (2005), Guizzardi (2005), Sure *et al.* (2004) e Pinto *et al.* (2004). Assim, foi realizado um estudo sobre três metodologias para construir ontologias, sendo analisado os processos de desenvolvimento de Neon⁶⁷, OntoForInfoScience⁶⁸ e SABiO⁶⁹.

A metodologia NeOn (*Network Ontology*) é um processo para construir ontologias de forma colaborativa, e também para a reutilização e evolução dinâmica de ontologias para ambientes distribuídos em rede (SUÁREZ-FIGUEROA, 2010; FARINELLI; ELKIN, 2017). Ela foi estabelecida com base nas melhores práticas de outras três metodologias: Methontology⁷⁰, On-To-Knowledge⁷¹ e Diligent⁷² (FARINELLI; ELKIN, 2017). A fusão dessas estratégias fornece orientação para uma grande parte dos pontos-chave do processo de engenharia ontológica que diz respeito ao desenvolvimento colaborativo, na reutilização de engenhos ontológicos e não ontológicos, e também no progresso e manutenção de ontologias com desenvolvimento distribuído ou em rede. De acordo com Munn e Smith (2008 *apud* FARINELLI; ELKIN, 2017, p.134), a vantagem da metodologia NeOn está no seu processo de elaboração por caráter incremental, e isso é apropriado para desenvolver ontologias com base no falibilismo no qual o realismo ontológico diz respeito, pois o

⁶⁷ Disponível em: http://neon-project.org/nw/Welcome_to_the_NeOn_Project.html

⁶⁸ Disponível em: <http://hdl.handle.net/1843/BUBD-A35H3K>

⁶⁹ Disponível em: https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/sabio_systematic_approach_for_building_ontologies_2014.pdf

⁷⁰ Methontology é uma metodologia estruturada baseada na experiência de desenvolvimento de uma ontologia no domínio de química e que é usada para elaborar ontologias desde o início de sua elaboração (FERNANDEZ *et al.*, 1997). Ela é apresentada em várias publicações e é usada em diversos domínios como Saúde, Geociências, Citometria, Química, etc (BOER *et al.*, 2022). Methontology é usada por vários engenheiros de ontologia devido a sua documentação vasta e detalhada, além de ter sido usada em várias experiências científicas (BOER *et al.*, 2022).

⁷¹ On-To-Knowledge (OTK) é uma metodologia para criar e possibilitar soluções de gerenciamento de conhecimento baseada em ontologias para empresas com foco nos processos de conhecimento e meta-processos de conhecimento (SURE *et al.*, 2004). Ela serve para criar ontologias e pode ser usada em diferentes aplicações baseadas em ontologias (SURE *et al.*, 2004). OTK é centrada na identificação, ilustração e instanciação do processo de conhecimento e do meta-processo de conhecimento, e suas limitações estão na sua adoção em redes de ontologias devido a natureza dinâmica, pois o suporte metodológico de OTK e suas ferramentas não são adequados para a construção de redes (SURE *et al.*, 2004).

⁷² Diligent diz respeito a uma metodologia para construir outras ontologias e para apoiar especialistas de domínio (PINTO *et al.*, 2004).

falibilismo é impulsionado pela disposição evolutiva da ciência e está contida no quadro do realismo ontológico.

OntoForInfoScience é uma metodologia para construção de ontologias (MENDONÇA; ALMEIDA, 2015; MENDONÇA, 2015). Ela é uma alternativa para desenvolvimento de ontologias desenvolvida para atender aos profissionais que não são engenheiros do conhecimento (MENDONÇA; ALMEIDA, 2015). Para Mendonça e Castro (2021), a principal característica de OntoForInfoScience está em abordar de maneira detalhada as etapas e tarefas do ciclo de desenvolvimento ontológico, possibilitando uma compreensão mais clara das atividades que são imprescindíveis para construir ontologias para diversos tipos de desenvolvedores. Esta metodologia já foi usada para desenvolver ontologias no ramo da biomédica (MENDONÇA; ALMEIDA, 2015; MENDONÇA, 2021), mas não há uma avaliação de sua utilização para implementações ontológicas mais robustas (MENDONÇA; CASTRO, 2021).

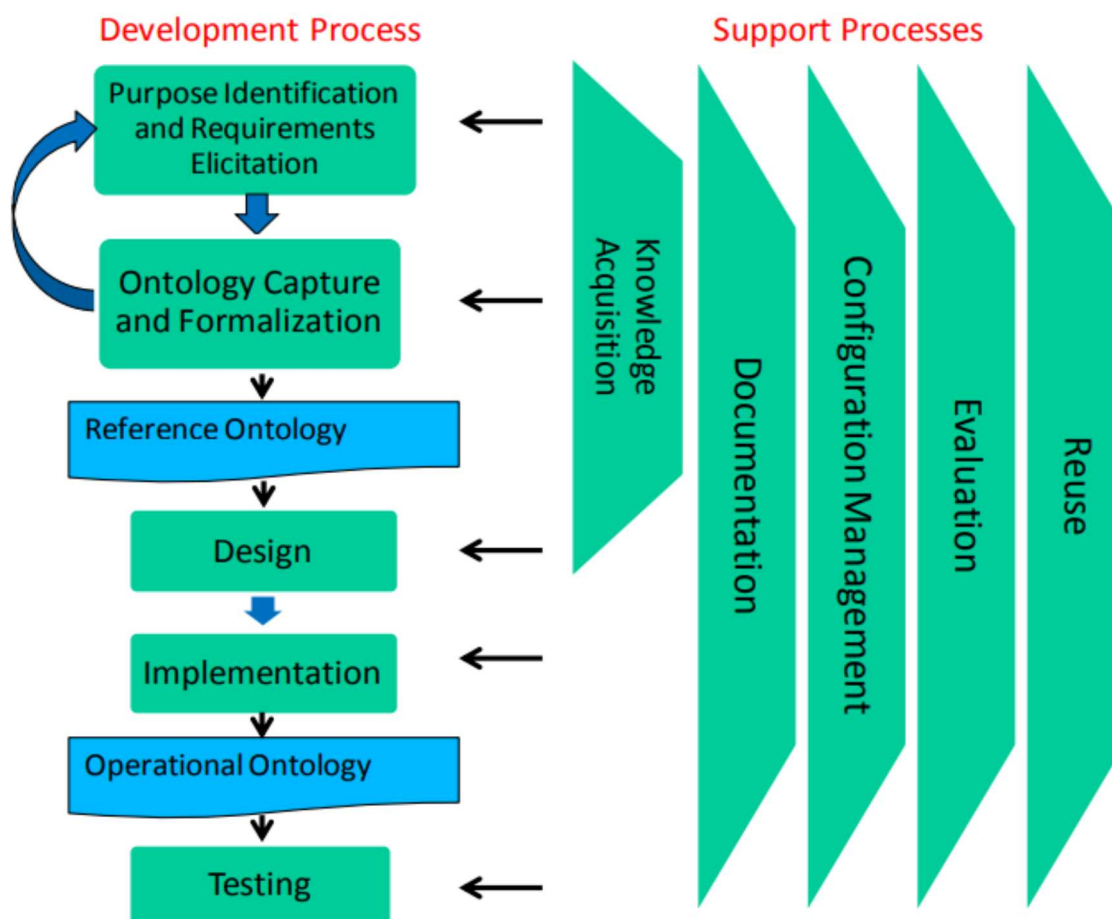
Segundo Falbo (2014), *Systematic Approach for Building Ontologies* (SABiO) é uma metodologia para elaborar ontologias de domínio. Ele explica que podem ser desenvolvidos dois tipos de ontologias de domínio: ontologias de referência e operacionais. Enquanto a primeira é um tipo especial de modelo conceitual (FALBO, 2014), a segunda é aquela projetada para assegurar as propriedades computacionais esperadas (GUIZZARDI, 2006). Para Falbo (2014), a construção de ontologias deve considerar o processo de desenvolvimento e também cinco processos auxiliares como aquisição de conhecimento, documentação, gerência de configuração, avaliação e reutilização. O autor argumenta que para construir uma ontologia de referência é suficiente realizar as três primeiras fases de SABiO, mas para elaborar uma ontologia operacional é necessário realizar todo processo da metodologia mencionada que é composta por 5 fases.

SABiO possui um processo de desenvolvimento organizado em cinco fases, conforme mostra a Figura 7, (FALBO, 2014):

- I. Identificação do propósito e elicitação de requisitos (*Purpose identification and requirements elicitation*);
- II. Captura e formalização de ontologias (*Ontology capture and formalization*);
- III. Projeto (*Design*);
- IV. Implementação (*Implementation*);

V. Teste (*Testing*).

Figura 7- Processo de metodologia SABiO ([link para acesso em tamanho maior](#))



Fonte: Falbo (2014)

A metodologia SABiO, Figura 7, tem como cerne descrever uma realidade por intermédio da elaboração da ontologia de domínio. Por essa razão, a ontologia de domínio está apresentada em cor azul e corresponde aos componentes Ontologia de Referência (*Reference Ontology*) e a Ontologia Operacional (*Operational Ontology*), onde o primeiro refere-se ao modelo conceitual e o segundo diz respeito ao projeto de especificações computacionais que são legíveis por máquinas.

Falbo (2014) esclarece que uma ontologia de domínio precisa ser elaborada com respaldo em uma ontologia de fundamentação e que pode ser estruturada em três níveis distintos e interrelacionados: ontologia de fundamentação (*foundational ontology*), ontologia de referência de domínio (*domain reference ontology*) e ontologia operacional de domínio (*domain operational ontology*). A ontologia de fundamentação

corresponde a disciplina filosófica para representar classes rigorosamente (SMITH, 2015) de todas as formas e maneiras de ser (GUARINO; GIARETTA, 1995) e é apresentada com mais detalhes na próxima subseção (4.2.1.2). A ontologia de referência de domínio realiza especificações em um nível mais conceitual e tem como objetivo auxiliar na comunicação (FALBO, 2014), ajudar no entendimento do domínio (ANDRADE; NARDI; RUY, 2023), aprendizagem (FALBO, 2014), resolver dúvidas durante o processo de alinhamento de significado entre conceitos e relações (ANDRADE; NARDI; RUY, 2023) e resolução de outros problemas (FALBO, 2014). Já a ontologia operacional de domínio trata da especificação sistêmica com o intuito de implementar um determinado domínio (FALBO, 2014).

A primeira fase da metodologia SABiO na Figura 7 corresponde a **Identificação do propósito e elicitação de requisitos (*Purpose identification and requirements elicitation*)** e diz respeito às quatro atividades que ocorrem de forma iterativa, sendo que todas envolve a colaboração de engenheiro de ontologia, especialista do domínio e do potencial usuário da ontologia. Falbo (2014) classifica as quatro atividades desta fase em modularização de ontologias, identificação do propósito da ontologia e a aplicação pretendida, elicitação de requisitos e identificação das questões que a ontologia deve ser capaz de responder.

Inicialmente busca-se identificar o propósito da ontologia e a aplicação pretendida. Em seguida os requisitos são elicitados e podem ser classificados em requisitos funcionais⁷³ e requisitos não funcionais⁷⁴ (FALBO, 2014). Os requisitos funcionais dizem respeito ao conteúdo (conhecimento) a ser estruturado pela ontologia e podem expressar as questões que a ontologia deve ser capaz de responder (GRÜNINGER, 1995). Ao definir as questões que a ontologia deve ser capaz de responder, consegue-se uma forma eficiente de indicar o que pode ser relevante para a ontologia e aquilo que não é (definição do seu escopo), bem como pode-se realizar uma estratégia para avaliá-la (FALBO). Sob outra perspectiva, os requisitos não funcionais são às características, qualidades e aspectos gerais não

⁷³ Requisitos funcionais são aqueles que descrevem a funcionalidade ou os serviços do sistema (detalham as funções que o sistema deve fazer) (SOMMERVILLE, 2011). Exemplos desses requisitos são: cadastro de clientes, emissão de nota fiscal, consulta ao estoque, dentre outros (OLIVEIRA; NUNO, 2020).

⁷⁴ Requisitos não funcionais são aqueles que não estão ligados diretamente à funcionalidade do sistema, mas manifestam propriedades do sistema ou restrições quanto às funções ou os serviços munidos por ele (SOMMERVILLE, 2011).

associados ao conteúdo da ontologia, como por exemplo, os requisitos de interoperabilidade⁷⁵ (SUÁREZ-FIGUEROA *et al.*, 2012).

Os requisitos funcionais precisam descrever as questões que a ontologia deve ser capaz de responder (FALBO, 2014). Essas questões podem ser identificadas e notacionadas utilizando métodos diversos e em diferentes níveis de granularidade como *Top-down*⁷⁶, *Bottom-up*⁷⁷ ou *Middle out*⁷⁸ (SUÁREZ-FIGUEROA *et al.*, 2012).

Falbo (2014) argumenta que não há uma forma universal de modularizar uma ontologia e esclarece que SABI_O sugere sua decomposição em sub-ontologias. Já SUÁREZ-FIGUEROA *et al.* (2012) recomenda que a escolha de uma determinada abordagem deve ser norteada pelos requisitos da ontologia e as questões que ela (ontologia) deve ser capaz de responder.

Falbo (2014) explica que SABI_O propõe a identificação de propósito e elicitación de requisitos considerando potenciais usuários, engenheiros e especialistas do domínio com ciclos de quatro atividades que devem ser executados de forma iterativa (modularização da ontologia; propósito e identificação pretendida pelos usuários; elicitación de requisitos; e identificação das questões que a ontologia deve responder). Ele exemplifica da seguinte maneira: baseando-se no propósito e no uso inicialmente levantados com os potenciais usuários da ontologia é que são levantados os requisitos e também são delineadas as questões que a ontologia deve ser capaz de responder. Assim, o engenheiro identifica sub-ontologias (modularização da ontologia) e separa essas questões que a ontologia deve ser capaz de responder em sub-ontologias identificadas.

⁷⁵ Requisitos de interoperabilidade estabelece que a ontologia deve ser fundamentada em uma ontologia de fundamentação particular para favorecer sua integração com ontologias externas e que já existem (FALBO, 2014).

⁷⁶ *Top-down* pode-se traduzir como “de cima para baixo”. Nesta estratégia tanto a ontologia como o engenheiro partem das questões mais complexas que são ramificadas questões mais simples (SUÁREZ-FIGUEROA *et al.*, 2012).

⁷⁷ *Bottom-up* pode-se traduzir como “de baixo para cima”. Nesta técnica o engenheiro de ontologia parte das questões mais simples que resultam na composição de outras questões que são mais complexas (SUÁREZ-FIGUEROA *et al.*, 2012).

⁷⁸ *Middle out* pode-se traduzir como “meio-termo”. Neste método o engenheiro inicia a elaboração da ontologia notacionando as questões que são compostas ou decompostas para posteriormente realizar resumos ou perguntas simples. Sua finalidade estará sempre focada em descrever as questões que a ontologia deve ser capaz de responder (conteúdo, conhecimento e definição de escopo), não importando a estratégia usada (questões compostas/abstratas ou decompostas/simples). Enquanto as questões abstratas são importantes para produzir os casos de teste, as questões simples são importantes para guiar a modularização da ontologia (SUÁREZ-FIGUEROA *et al.*, 2012).

A segunda fase da metodologia SABiO na Figura 7 corresponde a **Captura e formalização de ontologias (*Ontology capture and formalization*)**, e o objetivo principal desta fase está em capturar os conceitos do domínio com base nas questões que a ontologia deve responder. Como SABiO sugere que os conceitos e as relações de ontologia de domínio e ontologia de referência são analisados com base em uma ontologia de fundamentação, a análise ontológica pode ser aplicada em situações diversas como: devem ser analisados os módulos, os recursos ontológicos⁷⁹ e recursos não ontológicos⁸⁰ (como tesauros, esquemas de classificação, folksonomias) a serem reutilizados das ontologias que são desenvolvidas do zero; e deve-se escolher uma ontologia de fundamentação para realizar esta tarefa.

Tanto os conceitos como as relações relevantes devem ser identificados e organizados. Um modelo de representação gráfica é um instrumento importantíssimo para apoiar a comunicação, além de possibilitar a discussão, negociação e representação do consenso alcançado entre os especialistas do domínio (FALBO, 2014). Guarino (2009) argumenta que não há um consenso universal sobre as distinções ontológicas construídas na linguagem da representação. Ele explica que os pressupostos sobre as construções básicas das linguagens de representação localizam-se implícitas à mente do engenheiro do conhecimento, e que é complicado expressá-las ou compartilhá-las. Por outro lado, para elaborar ontologia de domínio e ontologia de referência é preciso usar linguagens altamente expressivas para criar ontologias fortemente axiomatizadas e que se aproximem ao máximo da ontologia ideal do domínio (FALBO, 2014). O objetivo dessas linguagens está na adaptação da representação porque o resultado das especificações é para o uso humano (GUIZZARDI, 2006).

É importante mencionar que a atividade de modelagem conceitual começa nesta segunda fase, e durante esse processo o engenheiro de ontologia deve identificar os principais conceitos e relações no domínio, bem como precisam ser organizados adequadamente (em taxonomia) (FALBO, 2014).

⁷⁹ Recursos ontológicos são as ontologias de domínio, ontologias de fundamentação, ontologias *core* e padrões ontológicos (*ontology patterns*) que pertencem a um campo de conhecimento e podem ser reusados para elaborar ontologia (FALBO, 2014).

⁸⁰ Recursos não ontológicos (*non-ontological resources* (NOR)) são recursos de conhecimento consolidado (FALBO, 2014; SUÁREZ-FIGUEROA *et al.*, 2012). Tratam-se de recursos de conhecimento aplicados nas atividades de obtenção de conhecimento que apoiam na construção de ontologias (SUÁREZ-FIGUEROA *et al.*, 2012). São exemplos de NORs elementos como tesauros, esquemas de classificação, léxicos, folksonomias, dentre outros (SUÁREZ-FIGUEROA *et al.*, 2012).

O conhecimento para a elaboração da ontologia pode ser obtido por meio dos especialistas de domínio, fontes de conhecimento consolidados como livros, padrões internacionais e modelos de referência (FALBO, 2014). Elementos como conceitos, relações e propriedades de recursos não ontológicos precisam ser analisados com base na ontologia de fundamentação escolhida.

A elaboração da ontologia deve ser guiada pelas questões que a ontologia deve ser capaz de responder. Se os elementos da ontologia (conceitos, relações, propriedades e axiomas) não forem suficientes para questões que a ontologia deve ser capaz de responder, então conceitos adicionais, relações, propriedades ou axiomas devem ser adicionados (FALBO, 2014). Assim, a elaboração da ontologia é um processo iterativo e que possui um processo de avaliação árdua. Nesse sentido, Falbo (2014) explica que durante a fase de elaboração e formalização da ontologia podem surgir outras questões que a ontologia deve responder (novos propósitos). Assim, a iteração com a fase anterior (Identificação de Propósito e Elicitação de Requisitos) é muito comum.

A terceira fase da metodologia SABiO na Figura 7 é o **Projeto (*Design*)**. Para Falbo (2014), esta é a fase em que se busca a elaboração de uma ontologia operacional que sirva para ser usada por aplicativos computacionais, isto é, trata-se de uma versão operacional almejando o projeto e implementação de uma ontologia que seja legível por máquinas em uma linguagem específica, como por exemplo OWL (FALBO, 2014). Nesse sentido, Guizzardi (2007) relata que a ontologia de referência pode ser usada para elaborar diversos projetos (*designs*) diferentes que sejam importantes para a implementação computacional, uma vez que nessa fase busca-se uma forma de facilitar e apoiar no desenvolvimento de uma ontologia que seja legível por máquinas computacionais.

Falbo (2014) argumenta que a ontologia operacional difere da ontologia de referência porque ela não é voltada para as adaptações da representação conceitual, mas sim para garantir um projeto com as propriedades computacionais que são importantes para a implementação computacional. O autor explica que na fase de projeto a especificação conceitual da ontologia de referência é transformada em especificação de projeto, considerando elementos que vão desde as questões arquitetônicas até os requisitos tecnológicos não funcionais com a finalidade de criar um ambiente que sirva para a implementação computacional.

Após a definição do ambiente de implementação, deve-se reavaliar a modularização da ontologia que foi estabelecida no início do projeto, levando em conta os requisitos tecnológicos não funcionais e as características do ambiente de implementação visando determinar a arquitetura da ontologia (FALBO, 2014).

A quarta fase de **Implementação (Implementation)** contida na Figura 7 refere-se à implementação da ontologia na linguagem operacional escolhida (FALBO, 2014), ou seja, nesta fase implementa-se em uma linguagem computacional a versão operacional do projeto de implementação, como por exemplo OWL ou SWRL⁸¹.

Já a quinta etapa **Teste (Testing)** da Figura 7 diz respeito à verificação e validação da ontologia operacional para um conjunto finito de casos de teste. Nesta fase é possível aplicar tanto o teste caixa-branca como o caixa-preta, desde que se verifique e valide-se as questões em que a ontologia deve ser capaz de responder (FALBO, 2014). Para testar, no ambiente de implementação escolhido e com instância de dados para o módulo da ontologia a ser testada, realiza-se a elaboração de uma consulta sobre uma questão que esteja contida no módulo em que a ontologia deve ser capaz de responder, e considera-se o resultado esperado com base nos dados de instâncias conhecidas.

Na presente pesquisa optou-se por adotar a metodologia SABiO, pois ela permite modelar com facilidade ontologias que representam elementos do mundo real utilizando-se de entidades, propriedades e relacionamentos. Além disso, ela apresentou excelentes resultados quanto ao processo de reuso de padrões relacionados a organização de domínios (FALBO 2014) e existe uma boa quantidade de pesquisas que usam ela no contexto da educação (SILVA; MUTZ; RUY, 2022a; SILVA; MUTZ; RUY, 2022b; SILVEIRA; ARAGÓN; CURY; MENEZES, 2021; SILVEIRA, CURY E MENEZES, 2019).

4.2.1.2 Ontologias de fundamentação

Ontologia de fundamentação é a disciplina filosófica que trabalha com representações formais de classes independente da sua natureza (SMITH, 2015). É o desenvolvimento sistemático, rigoroso e axiomático da lógica de todas as maneiras e formas de ser (GUARINO; GIARETTA, 1995). Guarino (2009) explica que o objetivo

⁸¹ SWRL é o acrônimo de *Semantic Web Rule Language*. Disponível em: <https://www.w3.org/Submission/SWRL/>

da formalidade na representação do conhecimento está em expressar de maneira rigorosa os compromissos da representação ontológica elaborada, e isso pode ser feito de duas formas: (1) criando ontologias gerais que representem uma construção neutra, ou (2) adotar uma construção que não seja neutra e que a semântica seja adequada e particular a fim de garantir modelos ontologicamente bem fundamentados. O autor recomenda a segunda opção quando o engenheiro do conhecimento busca elaborar modelos com transparência ontológica, boa fundamentação, reusabilidade, e argumenta que esse tipo raciocínio pode ser um pouco mais fácil por não usar linguagem-objeto, mas sim metalinguagem (linguagem usada para dar conta da semântica ontológica).

Nesse sentido, BFO é uma ontologia formal projetada para ser neutra quando aplicada aos domínios para o qual foi planejada, almejando auxiliar na interoperabilidade das ontologias de domínio definidas em sua base, e para assim sustentar um registro consistente de dados em diferentes domínios (SMITH, 2015). BFO é uma Ontologia Formal Básica desenvolvida pelo Instituto para Ontologia Formal Médica e Ciência da Informação (*Institute for Formal Ontology and Medical Information Science*) como uma teoria das estruturas básicas da realidade, e possui uma aceitação considerável nas áreas de medicina, biologia, bioinformática e áreas parecidas, direito, geografia (FARINELLI; ELKIN, 2017).

UFO é uma ontologia de fundamentação (GUIZZARDI; WAGNER, 2005) baseada em teorias como Ontologia Formal, Lógica Filosófica, Filosofia da Linguagem, Linguística e Psicologia Cognitiva, que tem como foco principal fornecer uma base para linguagens gerais de modelagem conceitual (ZAMBORLINI, 2011). A UFO simplifica resultados de diferentes ontologias de fundamentação, tais como GFO/GOL, OntoClean/DOLCE, que apesar de mostrarem propriedades importantes, possuem certas limitações tais como a capacidade de capturar conceitos elementares nas linguagens de modelagem conceitual (ZAMBORLINI, 2011). Então, a UFO tem como meta propor a fusão das ontologias mencionadas, apresentando características favoráveis e corrigindo as desvantagens identificadas (DEGEN *et al.*, 2001; WELTY; GUARINO, 2001; GUIZZARDI; WAGNER, 2005, 2010 *apud* ZAMBORLINI, 2011).

UFO corresponde ao acrônimo de *Unified Foundational Ontology* (Ontologia de Fundamentação Unificada), e é organizada em aspectos, tais como: UFO-A, UFO-B e UFO-C que determina os termos relacionados aos aspectos estruturais como

conceitos abrangentes de objetos, suas propriedades inerentes e relacionais, os tipos que eles instanciam, bem como os papéis que eles desempenham, dentre outros (GUIZZARDI; WAGNER, 2005). A UFO possui estereótipos de classes (conceitos) que servem para solucionar problemas clássicos de modelagem conceitual (GUIZZARDI, 2005), a saber:

- **Category:** É uma classe de estereótipos do tipo rígido que incorpora propriedades comuns, essenciais e que dispõe de diferentes princípios de identidade (por exemplo, Organização de Ensino).
- **Collective:** Trata-se da classe de estereótipos de tipo rígido que caracteriza coletivos. Instâncias desta classe representam coleções com características em comum.
- **Kind:** É uma classe de estereótipos de tipo rígido que fornece um princípio de identidade (por exemplo: Pessoa, Instituição de Ensino Superior, dentre outros).
- **Mixin:** É a classe de estereótipos que caracteriza propriedades que são indispensáveis para certas instâncias, mas que são dispensáveis para outras.
- **Phase:** Trata-se de uma classe de estereótipos de tipo anti-rígido que determina cada fase de classe que incorpora um princípio de entidade.
- **Quantity:** Classe de estereótipos que define a categoria das coisas que tem existência autônoma. Instâncias desta classe possuem características relacionadas a estrutura qualitativa.
- **Relator:** Diz respeito a classe de estereótipos que representa instância de propriedades relacionais. Ele ocorre no tempo e passa a existir no tempo, destacando próprio caráter estrutural.
- **Role:** É um tipo anti-rígido de estereótipos que define o papel de instâncias no escopo de uma relação como *Kind*, *Quantity* ou *Collective*.
- **Role Mixin:** Classe de estereótipos anti-rígidos que possui diferentes critérios para princípio de identidade. Instâncias desta classe não dispõem de princípios de identidade iguais.
- **SubKind:** Trata-se da classe de estereótipos de tipo rígido que especializa instâncias do seu supertipo *Kind* (por exemplo: Local Virtual e Local Físico são subtipos [*Subkind*] do supertipo [*Kind*] Local).

Optou-se usar, na presente pesquisa, a ontologia de fundamentação UFO. Com base na literatura consultada, a estruturação da UFO é mais natural e mais simples de entender que as ontologias de fundamentação consultadas, em especial a BFO. Existe também a disponibilidade de ferramentas com boa usabilidade para trabalhar com a UFO, como o plugin do Visual-Paradigm que será visto na subseção 4.2.3. Há linguagem de modelagem ontológica, como a OntoUML, que será tratada na próxima subseção, que faz uma boa integração com a UFO. Outro item que pesou no momento de escolher UFO, é que ela está sendo utilizada de maneira ampla para elaborar outras ontologias (ANDRADE; NARDI; RUY, 2023) e existem pesquisas em curso no ramo da educação que a utilizam para modelar ontologias (SILVA; MUTZ; RUY, 2022a; SILVA; MUTZ; RUY, 2022b; SILVEIRA; ARAGÓN; CURY; MENEZES, 2021; SILVEIRA, CURY E MENEZES, 2019).

4.2.1.3 Linguagem de modelagem orientada a ontologia

OntoUML⁸² é uma extensão da linguagem UML⁸³ baseada na *Unified Foundational Ontology* (UFO) (ZAMBORLINI, 2011). Trata-se de uma linguagem ontologicamente fundamentada para modelagem conceitual orientada por ontologia (ONTOUML, 2017), cujos estereótipos como *Kind*, *Subkind*, *Relator*, *Role*, são advindos da UFO (ZAMBORLINI, 2011). Embora não haja consenso sobre ter distinções ontológicas embutidas na linguagem de representação, a OntoUML é uma recomendação concreta nesse caminho (ZAMBORLINI, 2011).

A metodologia SABiO (FALBO, 2014) recomenda o uso de OntoUML por ser ela uma linguagem de fácil modelagem e bastante integrada às ideias propostas na metodologia. Guizzardi (2005) sugere OntoUML como um exemplo de linguagem de representação apropriada para ontologias de referência, pois utiliza elementos importantes do diagrama de classe da UML e abarca distinções fundamentais que são definidas pela UFO.

⁸² OntoUML corresponde a *Ontological Unified Modelling Language* (LA-ONGSRI; RODDICK, 2015). É uma extensão UML baseada na *Unified Foundational Ontology* (UFO). Disponível em: <https://ontouml.org/ontouml/>.

⁸³ UML corresponde a sigla de *Unified Modeling Language* (Linguagem de Modelagem Unificada). Seus diagramas e modelos foram originalmente projetados para o paradigma orientado a objetos, mas também são utilizados em outras áreas como modelagem de processos, modelagem de serviços, modelagem de componentes e interface. Disponível em: <https://www.softwareideas.net/uml>

Como SABiO recomenda usar OntoUML para modelar ontologia, o engenheiro de ontologia deve respeitar as restrições para relacionar os tipos, bem como classificar cada conceito com base nos tipos definidos na OntoUML (*kind, subkind, phase, role, category, rolemixin, etc.*). As regras e os padrões ontológicos de modelagem da OntoUML como *subkinds, phases, roles* (GUIZZARDI *et al.*, 2011) e *rolemixins* (GUIZZARDI, 2005) precisam ser aplicados para obter-se modelos conceituais consistentes (FALBO, 2014).

Adotou-se OntoUML, pois ele conta com uma ferramenta de suporte à engenharia de ontologias que permite a verificação de modelos e possibilita a exportação para linguagens de implementação como *Ontology Web Language* (OWL) (ANDRADE; NARDI; RUY, 2023).

4.2.1.4 Ferramenta gráfica para modelagem de ontologias

Draw.io⁸⁴ é um aplicativo de diagramação gratuito e atualmente é o mais instalado no Atlassian Marketplace (DRAW.IO, 2023).

O Visual-paradigm⁸⁵ é uma suíte de desenvolvimento de *software* e gerenciamento empresarial que fornece recursos para elaborar arquiteturas corporativas, gerenciar projetos e desenvolver *softwares* em equipe (VISUAL-PARADIGM, 2020). Ele contém um conjunto de ferramentas de análise, *design*, gerenciamento e desenvolvimento de modelos para TI com diversos recursos visuais de modelagem gráfica, (VISUAL-PARADIGM, 2020), não é um *software* gratuito (VISUAL-PARADIGM, 2022), mas existe uma versão simplificada e sem custos (versão *Community*).

Optou-se por modelar a ontologia da presente pesquisa no Visual-paradigm, pois ele dispõe de uma quantidade maior de recursos de edição visual para modelar ontologias com relação ao Draw.io. Além disso, ele permite a diagramação de ontologias na linguagem OntoUML.

⁸⁴ Draw.io é um *software* de diagramação. Disponível em <https://marketplace.atlassian.com/apps/1210933/draw-io-diagrams-for-confluence?hosting=cloud&tab=overview>

⁸⁵ Visual-paradigm é um *software* que serve para fazer diagramas e modelagem para TI. Disponível em: <https://www.visual-paradigm.com/>.

O gUFO⁸⁶ é uma Implementação de ontologia leve da UFO e que serve para implementar ontologias em OWL. Ele é um plug-in usado no Visual-Paradigm que serve para modelar diagramas na linguagem OntoUML e serviços de inteligência de modelo com o uso de estereótipos (ALMEIDA *et al.*, 2023; GUIZZARDI, *et al.*, 2022). Ele foi adotado pois permite o reuso em ontologias computacionais em OWL 2 DL⁸⁷ e tem mecanismos para detectar erros e aplicação de *design patterns*⁸⁸.

4.2.2 Design e implementação da ontologia operacional de domínio

Esta subseção aborda a fase de projeto (*design*) da ontologia operacional.

4.2.2.1 Ferramentas para construção e implementação de ontologias

Protégé⁸⁹ é um *software* de código aberto, gratuito e que serve para editar ontologias e elaborar sistemas inteligentes (MUSEN, 2015a). Esse ambiente foi projetado pelo Centro de *Stanford* para Pesquisa de Informática Biomédica (*Stanford Center for Biomedical Informatics Research (CBIR)*) (MUSEN, 2015b). Esta aplicação é apoiada por uma comunidade forte de usuários acadêmicos, governamentais e corporativos com a finalidade de construir soluções baseadas em áreas como Biomedicina, Comércio Eletrônico e Modelagem Organizacional (MUSEN, 2015a). Embora ele tenha sido desenvolvido para modelagem de sistemas biomédicos objetivando beneficiar a pesquisa clínica e translacional, suas pesquisas avançam no estado da arte em tecnologia semântica e estatística. Resumindo, esse sistema é para o desenvolvimento e manutenção de ontologias, disponibiliza um repositório e serviços da Web para clínicos e cientistas biomédicos usarem ontologias e pode ser utilizado para projetar ontologias de áreas diversas (MUSEN, 2015b).

Onto4AllEditor⁹⁰ é um editor gráfico gratuito que serve para criar, editar e exportar ontologias (ALMEIDA, 2015). Ele conta com recursos visuais disponíveis por meio de uma interface gráfica (MENDONÇA *et al.* (2021), proporcionando uma melhor

⁸⁶ O gUFO é um *plug-in* que serve para modelar ontologias em OntoUML usando o Visual-Paradigm. Disponível em <https://github.com/nemo-ufes/gufo>

⁸⁷ DL corresponde a *Description Logics* (Lógicas de Descrição). Disponível em: <https://www.w3.org/TR/2004/REC-owl-guide-20040210/#DescriptionLogics>

⁸⁸ *Design Patterns* é um padrão de projeto de *software* (*software design pattern*) no âmbito da Engenharia de *Software*. Disponível em: http://staff.cs.utu.fi/staff/jouni.smed/doos_06/material/DesignPrinciplesAndPatterns.pdf

⁸⁹ Disponível em <https://protege.stanford.edu/>

⁹⁰ Disponível em: <https://onto4all.com/en>

visualização do modelo ontológico e permitindo uma interação dinâmica por parte do usuário com o modelo (MENDONÇA; CASTRO, 2021). Ele engloba meios de associação entre ontologias, console de avisos, reuso de ontologias, relatório de erros e editor de tesouros (MENDONÇA *et al.*, 2021). Sua limitação está na ausência de avaliação para uso nos casos mais robustos e que respalde a sua utilização (MENDONÇA; CASTRO, 2021).

VocBench⁹¹ é uma plataforma de desenvolvimento colaborativo, multilíngue e baseada na Web para gerenciar ontologias OWL, tesouros, vocabulários controlados como o SKOS, léxicos Ontolex-lemon e conjuntos de dados RDF genéricos. Ele tem como objetivo estabelecer novos padrões de flexibilidade, abertura e poder expressivo de modelagem RDF grátis e de código aberto, sendo que ele pode ser baixado gratuitamente na plataforma VochBench. Projetado para apoiar as necessidades de ambientes da Web semântica e dados ligados, a sua propagação foi impulsionada pela retroalimentação de comunidade de usuários em organizações públicas, empresas e usuários independentes que almejam soluções de código aberto. O VocBench se destaca pela sua organização e multiplicidade de mecanismos de edição, simplicidade para gerenciamento de ontologias OWL, tesouros SKOS e qualquer tipo de conjunto de dados RDF (VOCBENCH, 2022).

O vocBench foi escolhido em detrimento do Protégé, pois oferece um suporte robusto para representação de conhecimento apoiada pelo padrão de metadados SKOS, e também proporciona desenvolvimento, edição e personalização de ontologias com amplo suporte de pesquisa no âmbito da tecnologia semântica. Ele permite gerenciar ontologias OWL, tesouros, SKOS e conjuntos de dados RDF genéricos com grande poder expressivo em sua modelagem, além de se destacar por sua organização e multiplicidade de mecanismos de manutenção e simplicidade para gerenciar ontologias.

O Notepad++⁹² é um editor de texto de código-fonte gratuito que oferece suporte a diversos idiomas (NOTEPAD++, 2023). Ele foi usado como apoio para trabalhar diretamente com o código da ontologia completa em um único arquivo, ou seja, com a ontologia operacional serializada.

⁹¹ Disponível em <http://vocbench.uniroma2.it/>

⁹² Disponível em: <https://notepad-plus-plus.org/>

4.2.2.2 Linguagens para representação e serialização de ontologias operacionais

O desafio da Web semântica está em fornecer uma linguagem que expresse dados e regras para raciocinar sobre os dados (os metadados) e permitir regras de qualquer sistema de Representação do Conhecimento existente a ser exportado para a Web (BERNERS-LEE; HENDLER; LASSILA, 2002).

Duas tecnologias de destaque para o desenvolvimento da Web semântica são XML e RDF. A tecnologia XML permite criar tags próprias, rótulos ocultos em páginas da Web ou seções de texto em uma página. Resumidamente, XML permite aos usuários adicionar uma estrutura arbitrária aos seus documentos, mas não diz nada sobre o que as estruturas significam (BERNERS-LEE; HENDLER; LASSILA, 2002).

Para Laufer (2015), RDF é um arcabouço que serve para representar informações na Web. Ele permite fazer afirmações sobre recursos, e esses recursos podem ser qualquer coisa, ou seja, tanto os elementos concretos como os abstratos podem ser recursos. Tem-se como exemplo de recursos uma empresa, uma pessoa, uma página Web, dentre outros.

Objetivando desenvolver o mapeamento de dados ligados na Web semântica do estudo de caso, adotou-se a tecnologia RDF porque ele permite representar dados e informações na Web para armazenamento e recuperação triplas por intermédio de consultas semânticas (entidade composta por sujeito-predicado-objeto).

O RDF/XML⁹³ foi a primeira notação a ser utilizada que permite um grafo RDF ser escrito num formato de texto compacto. Essa linguagem carrega a importância de ser padronizada pelo W3C, tem no seu uso a vantagem das linguagens de programação possuírem mais suporte para XML, além de possibilitar o uso de namespaces XML e evitar o uso de URIs completas tornando as URIs menores (LAUFER, 2015).

O *Turtle* é uma linguagem ou sintaxe textual para RDF que permite um grafo RDF ser completamente escrito em formato de texto compacto, nativo e com abreviações para padrões de uso, bem como tipos de dados comuns (HOMMEAUX; CAROTHERS, 2014). Ele fornece níveis de compatibilidade com o formato N-Triples⁹⁴

⁹³ RDF/XML é uma sintaxe XML para RDF em termos de Namespaces em XML, XML *Information Set* e XML Base. Disponível em <https://www.w3.org/TR/rdf-syntax-grammar/>

⁹⁴ N-Triples é um formato de texto elementar organizado em linha para codificar um grafo RDF. Disponível em: <https://www.w3.org/TR/n-triples/>

(que corresponde a estrutura de sintaxe textual para RDF com sujeito, predicado e objeto), ou seja, sua sintaxe é de padrão triplo da recomendação SPARQL W3C.

A leitura da codificação RDF/XML é bastante complicada para seres humanos, e por essa razão optou-se por *Turtle*⁹⁵. Assim, utilizou-se *Turtle* para documentar as triplas RDF cuja extensão de arquivo é *ttl*⁹⁶, isto é, usou-se essa linguagem para implementar o RDF objetivando a realização do mapeamento para Base de Dados Ligados em RDF.

O *W3C Web Ontology Language* (Linguagem de Ontologia Web - OWL⁹⁷) é uma linguagem da Web Semântica projetada para representar conhecimento rico e complexo sobre coisas, grupos de coisas e relações entre coisas (OWL, 2012). Ela é baseada em lógica computacional a fim dele poder ser explorado por programas de computador. Um documento OWL também é conhecido como ontologia, pode ser publicado na *World Wide Web*, ou também pode se referir a outras ontologias OWL (OWL, 2012). Trata-se do fragmento da pilha de tecnologia da Web Semântica do W3C que abarca elementos como RDF, RDFS, SPARQL (OWL, 2012), dentre outros. Essa linguagem foi utilizada por ser uma recomendação e também um padrão da W3C, além de proporcionar um maior número de interpretação por máquinas, pois ela fornece vocabulário adicional por intermédio da semântica formal.

4.2.3 Preparação de dados

Esta subseção trata do *software* OpenRefine⁹⁸ e os procedimentos de limpeza e estruturação dos dados do estudo de caso pré-processados. A preparação dos dados é importante para construir ontologia de referência de domínio e ontologia de referência com dados mais precisos (limpando os dados). Ela também possibilita o mapeamento atendendo a interoperabilidade da camada semântica por intermédio da reconciliação.

OpenRefine, anteriormente chamado de “Google Refine”, é um programa bem difundido, de código aberto e gratuito, e que serve para trabalhar com dados não

⁹⁵ Disponível em: <https://www.w3.org/TR/turtle/#language-features>.

⁹⁶ Disponível em [https://www.oxfordsemantic.tech/fundamentals/what-is-a-ttl-file#:~:text=TTL%20\(pronounced%20'turtle'\),representing%20information%20in%20the%20web%E2%80%9D](https://www.oxfordsemantic.tech/fundamentals/what-is-a-ttl-file#:~:text=TTL%20(pronounced%20'turtle'),representing%20information%20in%20the%20web%E2%80%9D).

⁹⁷ Disponível em: <https://www.w3.org/TR/owl-guide/>

⁹⁸ OpenRefine é um *software* que serve para trabalhar com dados não estruturados. Disponível em <https://openrefine.org/>

estruturados. Trata-se de uma aplicação consolidada para depurar e analisar dados estruturados, porém confusos. Isso quer dizer que suas principais funções referentes aos dados estão em limpá-los ou transformá-los em outro formato, disponibilizando também funções na Internet e os integrando com dados externos (OPENREFINE, 2021).

O OpenRefine resolve problemas tais como coluna sem nome, conteúdos sem preenchimento, tipificação dupla de tipo de dado para uma mesma coluna. Para Hooland *et al.* (2013) não se deve tomar os dados pelo valor que se apresenta. Isso enfoca como os estudiosos podem diagnosticar e trabalhar com base na precisão deles.

4.2.4 Mapeamento para base de dados ligados RDF

Nesta subseção foi realizada a elaboração do mapeamento da base de dados do estudo de caso pré-processados em RDF, baseando-se na ontologia criada no VochBench.

O RMLMapper⁹⁹ é um aplicativo que serve para mapear dados ligados com o intuito de gerar gráficos de conhecimento. Ele está disponível para os Sistemas Operacionais Linux, Windows e macOS e é compatível com o RMLStreamer¹⁰⁰ (RML.IO, 2022).

Ontotext Refine (também conhecido como OntoRefine), é um aplicativo gratuito que serve para automatizar a conversão de dados em um grafo de conhecimento por intermédio de um mapeamento de dados (ONTOTEXT REFINE, 2022).

Entretanto, o mapeamento foi realizado no Ontotext Refine¹⁰¹ porque ele disponibiliza uma interface gráfica de fácil manuseio, diferentemente do RMLMapper que funciona por linhas de comando no prompt e com com arquivo de script. Ele foi escolhido pois é aceito para retornar buscas de informação com consultas SPARQL e é compatível com o GraphDB.

⁹⁹ RMLMapper é uma aplicação para mapear dados ligados em RDF (RML.IO, 2022).

¹⁰⁰ RMLStreamer é uma aplicação de visualizações diversas de dados ligados em RDF (RML.IO, 2022).

¹⁰¹ Ontotext Refine é um aplicativo que serve para realizar o mapeamento em dados ligados (ONTOTEXT REFINE, 2022).

4.2.5 Banco de dados orientados a grafos

Esta subseção apresenta o banco de dados orientado a grafo usado para consultas em RDF para recuperação da informação.

O Neo4j¹⁰² é um banco de dados orientado a grafo que se estabeleceu como ferramenta de produção confiável e escalável, seja como servidor autônomo ou em conjunto no armazenamento distribuído (DRAKOPOULOS; GOURGARIS; KANAVOS, 2020). Tem licença GPL3 "edição da comunidade" de código aberto e é acessível usando *software* escrito em outras linguagens por meio da linguagem de consulta particular Cypher (NEO4J GRAPH DATA PLATFORM, 2023; FRANCIS *et al.*, 2018).

O Ontotext GraphDB¹⁰³ foi a ferramenta escolhida para se utilizar com RDF e realizar consultas SPARQL e está disponível no portal GraphDB (ONTOTEXT GRAPHDB, 2022a). Trata-se de um banco de dados orientado a grafos que trabalha com o formato RDF completo para grande quantidade de dados e cargas de consultas razoáveis e se estabeleceu progressivamente para níveis de produção em sistemas distribuídos e para servidores autônomos (DRAKOPOULOS; GOURGARIS; KANAVOS, 2020). Esta ferramenta possibilita o trabalho em projetos com banco de dados em grafos semânticos, em ambiente local ou online, bem como proporciona as bases de desenvolvimento de sistemas de dados inteligentes (ONTOTEXT GRAPHDB, 2022a).

Foi adotado o uso do GraphDB para recuperar informação por meio de consultas, pois Neo4j não possui suporte para realização de consultas SPARQL.

4.2.6 Consultas em RDF

Esta subseção aborda a linguagem de consulta escolhida para a recuperação da informação.

Cypher¹⁰⁴ é uma linguagem de consulta para grafos (FRANCIS *et al.*, 2018). Trata-se da linguagem de consulta que é escrita para dados ligados em RDF e que é usada de maneira específica no *software* Neo4j.

¹⁰² Neo4J é um banco de dados orientado a gráfico (NEO4J GRAPH DATA PLATFORM, 2023)

¹⁰³ GraphDB é um *software* para realizar consultas SPARQL em RDF. Disponível em <https://www.ontotext.com/products/graphdb/graphdb-free/>.

¹⁰⁴ Disponível em: <https://neo4j.com/developer/cypher/>

SPARQL¹⁰⁵ é uma linguagem e também um protocolo de consulta para RDF que serve para recuperar informação (CLARK; FEIGENBAUM; TORRES, 2008).

No entanto, escolheu-se o SPARQL pois é a linguagem de consulta apropriada para realizar as consultas no GraphDB e que são compatíveis com o mapeamento em RDF que é realizado. Essa linguagem foi utilizada para recuperar informação por meio do retorno dos resultados nas consultas e para realizar a prova de conceito objetivando cumprir o último objetivo específico desta dissertação.

4.2.7 Visualização em RDF

Essa subseção aborda as tecnologias utilizadas para visualizar RDF.

O RMLStreamer¹⁰⁶ é uma aplicação capaz de gerar grafo do conhecimento por meio de um código-fonte que contém os dados ligados já mapeados (RML.IO, 2022). Ele é compatível com o RMLMapper e pode ser trabalhado com o código gerado dele.

Visual Graph¹⁰⁷ é uma tecnologia que serve para visualizar grafos do conhecimento. Adotou-se essa tecnologia, pois ele é compatível com os mapeamentos de dados ligados que são realizados no GraphDB. Ele foi usado para gerar visualizações diversas dos grafos de conhecimento em RDF sobre os repositórios elaborados no GraphDB, ou seja, ele permitiu criar visualizações personalizadas dos dados em RDF no formato de grafos, bem como possibilitou apresentar a recuperação de informação como resultados desta dissertação.

¹⁰⁵ SPARQL é uma linguagem de consulta e protocolo para RDF. Disponível em: <https://www.w3.org/TR/rd-sparql-protocol/>

¹⁰⁶ Disponível em: <https://github.com/RMLio/RMLStreamer>

¹⁰⁷ Visual Graph é um *software* para construção de visualizações diversas de RDF. Disponível em: <https://graphdb.ontotext.com/documentation/10.0/exploring-data.html#create-your-own-visual-graph>

5 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta os procedimentos metodológicos da pesquisa realizada. Descreve a classificação da pesquisa, o levantamento bibliográfico, o cenário do estudo de caso e as etapas procedimentais do desenvolvimento da pesquisa. Gil (2019) explica que é natural buscar a classificação das pesquisas devido a diversidade de objetos que elas se referem e por perseguirem objetivos distintos. Dessa maneira, a seção 5.1 Descreve a classificação da pesquisa, a seção 5.2 trata do levantamento bibliográfico, a seção 5.3 discorre sobre o cenário do estudo de caso e a seção 5.4 corresponde às etapas procedimentais do desenvolvimento da pesquisa.

5.1 CLASSIFICAÇÃO DA PESQUISA

Quanto à abordagem, trata-se de uma Pesquisa Qualitativa, pois tem o seu foco no processo, interpretando os fatos e dando-lhes significado. Quanto à natureza, trata-se de uma pesquisa aplicada, pois volta-se para a solução de problemas específicos vinculados à organização, disseminação e recuperação de informações com suporte de dados ligados interoperáveis na Web semântica e fundamentada em boas práticas. Quanto aos procedimentos, relaciona-se a pesquisa estudo de caso, pois na visão de Stake (1998), embora o estudo de caso não seja exatamente uma escolha metodológica, trata-se da seleção de um objeto a ser estudado, chamando a atenção para a questão daquilo que, especificamente possa ser aprendido a partir de um determinado caso.

O procedimento estudo de caso possibilita entender tanto as questões contextuais como as fenomenológicas da temática a ser estudada, Yin (2005, p.32) explica que o estudo de caso trata-se uma investigação empírica que “[...] investiga um fenômeno contemporâneo dentro de seu contexto da vida real, especialmente quando os limites entre o fenômeno e contexto não estão claramente definidos”.

No que tange às questões contextuais e fenomenológicas da temática, ressalta-se que o estudo de caso será realizado no contexto educacional, utilizando-se de procedimentos similares aos apresentados por Gonçalves e Jacyntho (2021). Eles realizaram o mapeamento do banco de dados para o padrão da Web semântica em RDF, resultando assim na aplicação de um conjunto de ontologias consagradas para

integrar os dados ligados com fonte de dados semânticos conhecidos e reusados, almejando também publicação desses dados na Web de Dados Ligados.

No que se refere ao estudo de caso, por se tratar de um acervo de dissertação de pós-graduação, o instrumento de validação correspondente foi a prova de conceito. Este é um procedimento relativamente comum em pesquisa. Por exemplo, para validar os dados de sua pesquisa, Coneglian (2020) aplicou a prova de conceito em sua tese de doutorado. Neto, Borges e Roque (2021), defendem a prova de conceito por identificarem como prática importante para validar ou convalidar o resultado dos dados no âmbito dos Sistemas de Informação e da Ciência da Informação (CI). Na visão de Neto, Borges e Roque, esse recurso de validação é uma prática de pesquisa que pode ser utilizada como instrumento de elaboração do conhecimento na submissão de objetos, artefatos e fenômenos de qualquer natureza quanto a um grupo de atividades. Isso contribui com o entendimento e conhecimento dos objetos, artefatos e fenômenos mencionados no âmbito do estudo, no desenvolvimento e no acolhimento de tecnologias ou produtos novos das organizações e seus atores (NETO; BORGES; ROQUE, 2021).

Como trata-se de uma temática que utiliza, como estudo de caso, um acervo de dissertações acadêmicas, a concepção metodológica para elaboração desta dissertação baseou-se na concepção pós-positivista definida por Creswell (2010, p.29), contendo características determinísticas, reducionistas, observações e mensurações empíricas, bem como a verificação da teoria.

5.2 LEVANTAMENTO BIBLIOGRÁFICO

Essa seção aborda como foi feito o levantamento bibliográfico referente ao estado da arte e à fundamentação teórica da pesquisa.

As consultas de publicações foram realizadas no portal CAPES-CAFe (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Comunidade Acadêmica Federada), Google Acadêmico, Ambiente Virtual de Aprendizagem (AVA) da disciplina Sistemas de Recuperação de Informação do ano de 2021, de livros e artigos especializados e que foram disponibilizados nas disciplinas Documento, Cultura e Estado, Informação e Cultura, Pesquisa em Ciência da Informação, Epistemologia da Ciência da Informação, Informação e Memória, todas cursadas no

Programa de Pós-Graduação em Ciência da Informação (PPGCI)¹⁰⁸ da Universidade Federal do Espírito Santo (UFES)¹⁰⁹, bem como pesquisa nos livros da Biblioteca Central da UFES.

O Quadro 2 especifica de maneira categorizada o levantamento bibliográfico executado durante a pesquisa.

Quadro 2 - Exposição das Fases do Levantamento Bibliográfico

Pesquisa Bibliográfica		
Fontes Gerais de Informação	Bases de Dados	<i>Asian Journal of Research in Computer Science</i> ; Base de Dados em Ciência da Informação (BRAPCI); Biblioteca Brasileira de Teses e Dissertações (BDTD); <i>Computer and Information Systems</i> ; EBSCOhost; Google Scholar; Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT); SBC ¹¹⁰ -OpenLib (SOL) ¹¹¹ ; SciELO – <i>Scientific Electronic Library Online</i> ; ScienceDirect; Scopus (Elsevier B.V); <i>Web of Science</i> .
	Periódicos Nacionais	Levantamento em periódicos nacionais em CI com Qualis “A” e “B”; Brajis; Liinc em Revista; Em Questão.
	Periódicos Estrangeiros	IEEE ¹¹² <i>Intelligent Systems</i> ; IEEEExplore; <i>International Journal on Semantic Web and Information Systems</i> .

¹⁰⁸ PPGCI possui Área de Concentração Informação, Sociedade e Cultura e duas Linhas de Pesquisa: Linha 1 - Cultura, Mediação e Uso da Informação, e Linha 2 - Memória, Representação e Informação. Disponível em:

https://cienciadainformacao.ufes.br/sites/cienciadainformacao.ufes.br/files/field/anexo/regimento_inter_no_ppgci_ufes.pdf

¹⁰⁹ UFES é uma instituição autárquica vinculada ao Ministério da Educação com autonomia didático-científica e vocação para atuar em todas as áreas do saber. Disponível em:

<https://www.ufes.br/institui%C3%A7%C3%A3o>

¹¹⁰ SBC é o acrônimo de Sociedade Brasileira de Computação. Disponível em:

<https://www.sbc.org.br/>

¹¹¹ SOL corresponde a **SBC-Open Lib**. Trata-se da biblioteca digital da SBC. Disponível em:

<https://sol.sbc.org.br/index.php/indice>

¹¹² IEEE corresponde a *Institute of Electrical and Electronics Engineers*. Disponível em:

<https://ieeexplore.ieee.org/Xplorehelp/overview-of-ieee-xplore/about-ieee-xplore>

	Evento Científico	Anais do Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (Enancib); <i>Computer on the Beach</i> ; Ontobras; Simpósio Brasileiro de Informática na Educação; <i>Workshop on Ontologies on Conceptual Modeling and Information Systems Engineering</i> .
Palavras-Chave Usadas	Português	“dados ligados”, “dados ligados e abertos”, “fair”, “interoperabilidade”, “ontologia”, “Ontologia”, “ontologia de fundamentação”, “ontologia na educação”, “ontologia educacional”, “organização do conhecimento”, “sistemas de organização do conhecimento”, “vocabulário controlado”, “web semântica”.
	Inglês	“controlled vocabulary”, “linked data”, “linked open data”, “LOD”, “educational ontology”, “fair”, “interoperability”, “knowledge organization”, “KO”, “knowledge organization systems”, “KOS”, “ontology”, “Ontology”, “foundational ontology”, “ontology in education”, “semantic web”.

Fonte: Realizado pelo próprio autor.

5.3 CENÁRIO DO ESTUDO DE CASO

Apesar da pesquisa criar uma ontologia de referência ampla para representar trabalhos de conclusão de curso de pós-graduação, a parte específica dos testes da ontologia operacional na recuperação de informação foi realizada sobre acervo¹¹³ de dissertações¹¹⁴ dos alunos egressos do PPGCI da UFES, sendo esses trabalhos acadêmicos de avaliação final obrigatórios para os alunos obterem o título de mestre no PPGCI da UFES e que estão disponíveis no repositório da Biblioteca Central da UFES. O curso de mestrado em CI da UFES vinculado ao PPGCI surgiu em 2019, possui uma carga horária de atividade pedagógica mínima de 480 (quatrocentos e oitenta horas), em no mínimo 24 (vinte e quatro) créditos distribuídos, sendo 16 (dezesseis) créditos em disciplinas obrigatórias e 8 (oito) créditos em disciplinas optativas para que o aluno obtenha o título de Mestre em CI. Ele é um curso em

¹¹³ Acervo de dissertações do PPGCI. Disponível em: <https://cienciainformacao.ufes.br/pt-br/pos-graduacao/PPGCI/disserta%C3%A7%C3%B5es-defendidas>

¹¹⁴ Dissertação é o resultado de um trabalho de pesquisa desenvolvido, documentado e defendido publicamente em cursos de pós-graduação de mestrado, destinado a obtenção do grau acadêmico de mestre. Disponível em: <https://repositorio.ufes.br/bitstream/10/1533/1/Normalizacao%20e%20apresentacao%20de%20trabalhos%20cientificos%20e%20academicos.pdf>

funcionamento, sendo que o programa já formou 21 (vinte e um) mestres e dispõe de 34 (trinta e quatro) alunos regularmente matriculados (UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, 2023b). Na data 25 de março de 2023, o acervo de dissertações do curso de mestrado do PPGCI apresentou 17 (dezessete) registros digitalizados e metadados cadastrados, sendo que o estudo de caso utilizará desses dados.

5.4 ETAPAS PROCEDIMENTAIS DO DESENVOLVIMENTO

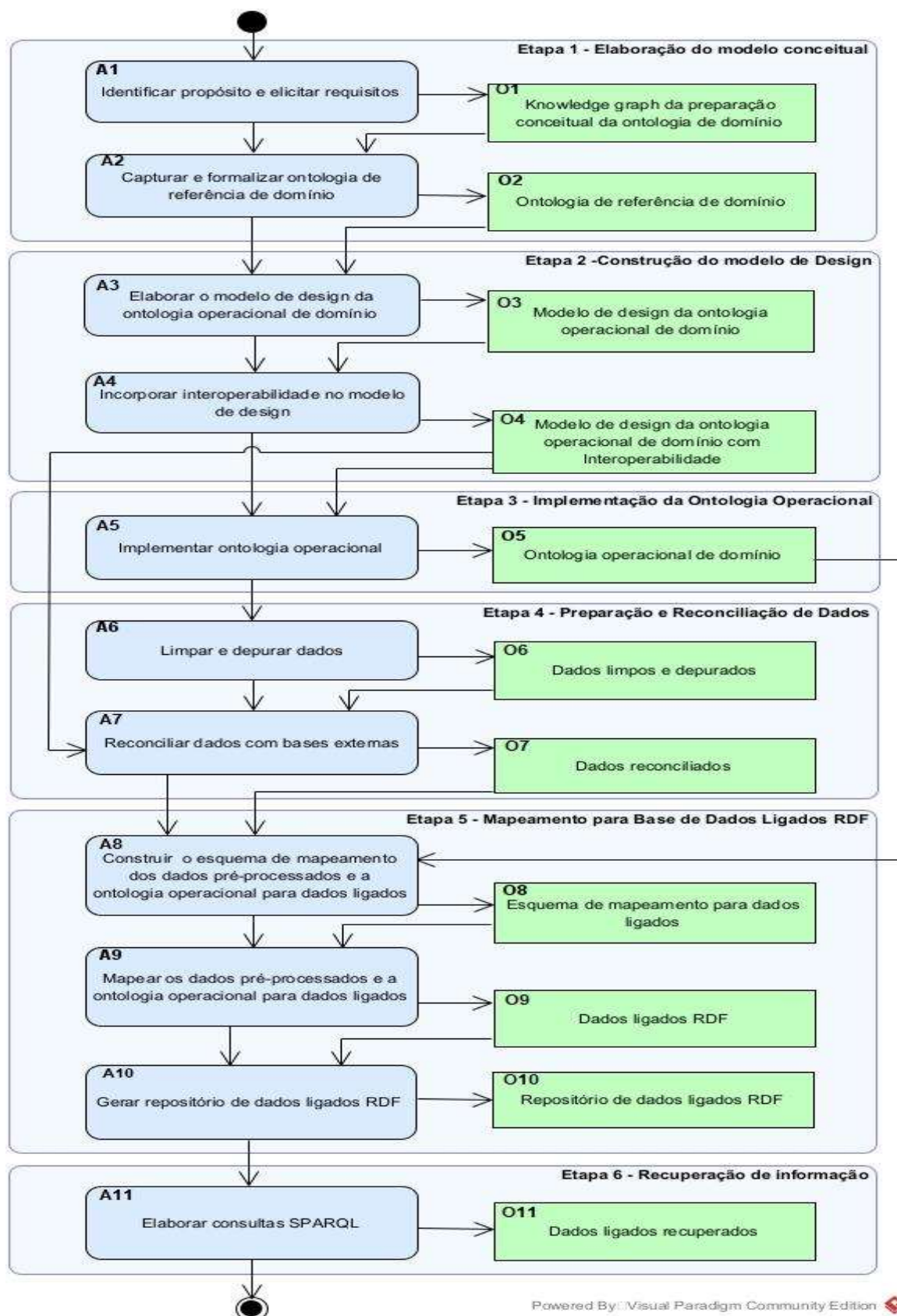
O procedimento de desenvolvimento da pesquisa foi organizado em seis etapas, com 11 ações (de A1 a A11) e 11 objetos resultantes dessas ações (de O1 a O11), apresentadas em formato de um diagrama de atividades¹¹⁵ UML, exposto na Figura 8.

Os Quadros 3 e 4 descrevem as atividades e os objetos gerados, respectivamente, utilizados no diagrama de atividades da Figura 8. As Etapas de 1 a 6 foram representadas pela legenda E1 a E6 em ambos Quadros.

As etapas identificadas são explicadas nas próximas subseções com maiores detalhes.

¹¹⁵ Diagrama de atividades da UML servem para exibir as atividades que integram um processo de sistema e o fluxo de controle que ocorre de uma atividade para outra (SOMMERVILLE, 2011). Ele pode ser elaborado por meio dos seguintes elementos de modelo: atividades, ações, nós de controle, nós do objeto e linhas de atividade. Disponível em: <https://www.ibm.com/docs/pt-br/rational-soft-arch/9.7.0?topic=diagrams-activity>

Figura 8 - Diagrama de atividades das etapas metodológicas do desenvolvimento ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

Quadro 3 - Atividades das etapas procedimentais do desenvolvimento

E1	A1	<p>Identificar propósito e elicitar requisitos</p> <p>Corresponde à primeira atividade da metodologia SABiO. Realizou-se entrevistas como elicitação de requisitos, estabeleceu-se o que a ontologia de domínio deveria responder e elaborou-se com o <i>software</i> um <i>knowledge graph</i> inicial com apoio do <i>software</i> CmapTools contendo entidades com suas respectivas propriedades do domínio da fase de análise conceitual.</p>
	A2	<p>Capturar e formalizar ontologia de referência de domínio</p> <p>Refere-se à segunda atividade da metodologia SABiO. Foi criada a ontologia de referência de domínio, baseada na ontologia de fundamentação UFO, utilizando-se o <i>knowledge graph</i> construído em A1, entrevistas com especialistas e consultas a documentos especializados. Ela foi elaborada com o suporte do <i>software</i> Visual-Paradigm, com a adoção da linguagem OntoUML e usando o plug-in gUFO.</p>
E2	A3	<p>Elaborar o modelo de <i>design</i> da ontologia operacional de domínio</p> <p>Refere-se à terceira atividade da metodologia SABiO. Elaborou-se com o <i>software</i> Visual-Paradigm o modelo de <i>design</i> da ontologia operacional de domínio tendo como base a ontologia de referência de domínio definida em A2. Esta etapa contou com o suporte da linguagem OntoUML embasada na ontologia de fundamentação UFO e usando o plug-in gUFO.</p>
	A4	<p>Incorporar interoperabilidade no modelo de <i>design</i></p> <p>Foram incorporados elementos de interoperabilidade da camada semântica e da camada estrutural no modelo de <i>design</i> da ontologia operacional de domínio construído em A3 por meio de anotações realizadas via o <i>software</i> Visual-Paradigm. Na interoperabilidade da camada estrutural associou-se elementos de vocabulários controlados, enquanto na interoperabilidade da camada semântica realizou-se associações com classes de outras ontologias, e indicando possíveis bases com indivíduos que pudessem ser equivalentes a indivíduos da base local do estudo de caso. Esses elementos foram relacionados com base nos portais especializados para consulta de termos e vocabulários classificados na subseção 5.4.2.</p>
E3	A5	<p>Implementar ontologia operacional</p> <p>Trata-se da quarta atividade da metodologia SABiO. Realizou-se a codificação do modelo da ontologia operacional na linguagem <i>Turtle</i>, utilizando-se dos <i>softwares</i> VocBench e Notepad++, a partir do modelo de <i>design</i> da ontologia operacional de domínio modificada em A4.</p>
E4	A6	<p>Limpar e depurar dados</p> <p>Realizou-se a limpeza e depuração dos dados do estudo de caso pré-processados com o <i>software</i> OpenRefine para eliminar espaços em branco, remover registros duplicados, eliminar tipificação dupla de dados, resolver conteúdo sem preenchimento e realizar o agrupamento de diferentes representações da mesma realidade.</p>
	A7	<p>Reconciliar dados com bases externas</p> <p>Realizou-se a interligação de dados do estudo de caso com dados de outras bases (repositórios externos) utilizando-se do <i>software</i> OpenRefine e por meio do modelo de <i>design</i>, elaborado em A4, com indicações de interoperabilidade da camada semântica.</p>

E5	A8	Construir o esquema de mapeamento dos dados pré-processados e a ontologia operacional para dados ligados Elaborou-se um esquema de mapeamento que recebe os dados pré-processados (O7) e a ontologia operacional de domínio (O5) para a geração dos dados ligados RDF correspondentes aos dados originais do estudo de caso. Este procedimento foi realizado com suporte da tecnologia RDF no <i>software</i> Ontotext Refine.
	A9	Mapear os dados pré-processados e a ontologia operacional para dados ligados O esquema de mapeamento elaborado em O8, foi executado por meio do <i>software</i> Ontotext Refine gerando um repositório de dados ligados em RDF no formato N-triples e correspondentes aos dados do estudo de caso.
	A10	Gerar repositório de dados ligados RDF Utilizou-se um banco de dados orientado a grafos, GrapDB, para criar um repositório capaz de armazenar o conjunto de dados ligados RDF gerados em A9.
E6	A11	Elaborar consultas SPARQL Foram escritas consultas na linguagem SPARQL para recuperação de informação em dados ligados no repositório criado (O10), como prova de conceito da pesquisa, e executados por meio do GrapDB.

Fonte: Elaborado pelo próprio autor.

Quadro 4 - Objetos das etapas procedimentais do desenvolvimento

E1	O1	Knowledge graph da preparação conceitual da ontologia de domínio É uma representação visual que indica, de forma preliminar, as principais entidades do estudo de caso com algumas conexões entre elas. Ele foi elaborado com o <i>software</i> CmapTools.
	O2	Ontologia de referência de domínio Modelo conceitual que usa como base a ontologia de fundamentação UFO. A representação gráfica foi feita com a linguagem OntoUML através da ferramenta Visual-Paradigm com o plugin gUFO.
E2	O3	Modelo de design da ontologia operacional de domínio É um modelo que representa elementos da ontologia de referência tendo como foco a indicação de elementos para a implementação da ontologia operacional. Esse modelo foi desenvolvido no <i>software</i> Visual-Paradigm com o plugin gUFO e com a linguagem OntoUML.
	O4	Modelo de design da ontologia operacional de domínio com Interoperabilidade É o modelo O3 com a incorporação de elementos de interoperabilidade das camadas semântica e estrutural. Desenvolvido com o <i>software</i> Visual-Paradigm e com o apoio dos portais especializados para consulta de termos e vocabulários classificados na subseção 5.4.2.
E3	O5	Ontologia operacional de domínio É uma ontologia de domínio codificada para a linguagem específica de representação dos dados que, nesse caso, foram os dados ligados RDF para a Web semântica e serializados no formato <i>Turtle</i> . O arquivo gerado possui

		extensão TTL e foi gerado com apoio dos <i>softwares</i> VocBench e Notepad++.
E4	O6	Dados limpos e depurados São os dados do estudo de caso pré-processados que passaram por correções de erros de conteúdo e formatação. Eles são gerados com apoio do <i>software</i> OpenRefine.
	O7	Dados reconciliados São os dados limpos e depurados do estudo de caso (O6) com a indicação de ligações para dados de outras bases de dados e bibliotecas externas. Adotou-se o OpenRefine.
E5	O8	Esquema de mapeamento para dados ligados É uma codificação preparada para gerar a base de dados ligados RDF do estudo de caso. O esquema é representado por um arquivo construído com o <i>software</i> Ontotext Refine e registrado no formato JSON.
	O9	Dados ligados RDF É o conjunto de dados ligados RDF em formato <i>N-triples</i> , que representa os dados do estudo de caso, criado a partir do esquema de mapeamento (O8) com o <i>software</i> Ontotext Refine.
	O10	Repositório de dados ligados RDF É o repositório que armazena o conjunto de dados ligados RDF (O9) em um banco de dados orientado a grafos, GraphDB.
E6	O11	Dados ligados recuperados É o resultado das consultas SPARQL realizadas sobre os dados ligados em RDF no repositório (O10), bem como a visualização dos grafos visuais como resultado da recuperação de informação. Adotou-se o GraphDB.

Fonte: Elaborado pelo próprio autor.

5.4.1 Etapa 1: Elaboração do modelo conceitual

A Etapa 1 correspondeu ao segundo objetivo específico desta dissertação - de elaborar um modelo conceitual baseado em ontologia de referência de domínio sobre trabalhos de conclusão de curso de pós-graduação. Assim, aplicou-se a etapa 1 proposta por Falbo (2014) (subseção 4.2.1.1), em SABiO (Identificação do propósito e elicitação de requisitos), construindo-se inicialmente um *knowledge graph* contendo as entidades e suas relações do estudo de caso (A1 e O1). O *knowledge graph* discutido na subseção 3.3.3 foi elaborado com base em documentos, artigos e livros especializados no ramo da educação, bem como reuniões informais com professores, orientadores e coorientadores especialistas do domínio educacional de graduação e pós-graduação *stricto sensu*. Para criar o *knowledge graph* usou-se o CmapTools.

Em seguida, elaborou-se uma ontologia de referência de domínio (A2 e O2), a partir do *knowledge graph* preliminar, aplicando-se a atividade 2 de SABiO (Captura e

formalização de ontologias). Utilizou-se a fundamentação UFO (ver subseção 4.2.1.2 da Fundamentação Metodológica), com apoio da linguagem OntoUML (subseção 4.2.1.3), e suporte do *software* Visual-Paradigm (ver subseção 4.2.1.4), com uso plugin gUFO (ver subseção 4.2.1.4), sendo elaborada por meio de documentos institucionais especializados e entrevistas com especialistas do negócio, tais como professores e orientadores de pós-graduação.

Convidou-se quatro professores especialistas no domínio acadêmico e administrativo de pós-graduação para realizar entrevistas com o objetivo de validar a ontologia de referência de domínio, e agendou-se com três: dois já foram coordenadores do Mestrado do PPGCI/UFES, e outro foi coordenador dos programas de mestrado e doutorado do PPGPS/UFES. Para eles, foi explicado que a ontologia de referência de domínio desenvolvida representa o contexto de defesa de dissertações e teses. Foram feitas perguntas, tais como se na visão dos especialistas a ontologia de referência de domínio estaria no caminho certo e se atendia ou respondia/correspondia às questões sobre o domínio de teses e dissertações. Essas entrevistas foram registradas e o resumo delas pode ser consultado nos Apêndices A e B.

Solicitou-se que os especialistas comentassem sobre o domínio de defesa de dissertações e teses antes de mostrar a primeira versão da ontologia de referência de domínio (construída com base em documentos, artigos e livros especializados no ramo da educação e reuniões informais com especialistas do domínio educacional). Em seguida, explicou-se as relações identificadas entre a primeira versão do *knowledge graph* e da ontologia de referência de domínio, discutindo-se e registrando as possíveis alterações da ontologia de referência de domínio.

A experiência de comunicação do projeto com a apresentação da primeira versão do *knowledge graph* e da ontologia de referência de domínio foi importante porque serviu para adequar relações, discutir e alinhar representações conceituais das entidades estabelecidas na fase de elaboração do modelo conceitual, assim como propriedades de entidades, restrições, estereótipos, dentre outros. Os entrevistados entenderam as representações do *knowledge graph* e da ontologia de referência de domínio, mas também mostraram curiosidade sobre a definição dos estereótipos que fazem parte da formalidade das ontologias, bem como levantaram questionamentos e

apresentaram discordâncias entre relações nas entidades, e esses elementos também estão registrados nos Apêndices A e B.

5.4.2 Etapa 2: Construção do modelo de *design*

A Etapa 2, Figura 8, corresponde a construção do modelo de *design* e diz respeito ao objetivo específico “Criar um modelo de *design* para a ontologia operacional de domínio baseada na ontologia de referência do objetivo específico (II) e com elementos interoperáveis da camada estrutural e da camada semântica”. Nesse estágio, adotou-se a fase 3 indicada em SABiO (Projeto - *Design*), construindo um modelo de *design* para a ontologia operacional de domínio (A3 e O3), e em seguida, associou-se elementos de vocabulários controlados (seção 3.3.2) e esquemas de metadados externos para atendimento à interoperabilidade da camada estrutural (subseção 3.3.5). Assim, criou-se também as conexões de instâncias e entidades com outras bases objetivando a interoperabilidade da camada semântica por intermédio de vocabulários controlados (subseção 3.3.2) definidos em bibliotecas especializadas para a ontologia de domínio (A4 e O4). Para isso, usou-se entidades do modelo conceitual, os relacionamentos entre elas, e também as suas propriedades, bem como suas nomenclaturas por meio de bases externas compatíveis com a ontologia operacional de domínio (ação de reuso da engenharia de ontologias). Lemos, Martins e Carmo (2022) argumentam que as fases de criação de metadados, assim como o processo de descrição, indexação, catalogação e classificação no âmbito da CI favorecem a qualidade de dados quanto ao uso e reuso em diferentes aplicações.

Usou-se os padrões de notações para classes, relações e propriedades que são propostas pela UML, adotando-se a técnica camelCase¹¹⁶ que é uma convenção de nomenclatura adotada para escrever um código mais legível.

Decidiu-se elaborar o modelo de *design* com nomes mais descritivos nos “Datatype Properties”, pois (1) uma propriedade pode ser considerada um “DatatypeProperty” se o seu range é um literal ou (2) um “ObjectProperty” se o seu range não for um literal (por exemplo: uma classe ou enumeração). Os resultados provenientes desta decisão são encontrados na seção 6.2.

¹¹⁶ CamelCase é um tipo de padrão para definir nomes de variáveis, tipos e funções em diversas linguagens de programação para computador. Disponível em: <https://everything2.com/title/CamelCase>

Para “ObjectProperty” é consenso usar-se descrição verbal, isto é, a terminologia de cada nome das propriedades iniciada por um verbo seguido do substantivo do range (classe destino). Quanto ao uso de terminologias verbais para não literais (classe, enumeração, etc), a W3C recomenda que todas as propriedades sejam nomeadas para que as triplas possam ser lidas de maneira clara e objetiva, como por exemplo, “hasProperty” (temPropriedade) (HYLAND; ATEMEZING; VILLAZÓN-TERRAZAS, 2014).

Referente aos Datatype Properties, eles foram notacionados com o nome que identifica a propriedade (substantivo) e o domínio ao qual ele pertence (classe). Para esse elemento não usou-se de terminologia verbal para distingui-los de ObjectProperty, e por outros dois motivos: (1) Na tripla RDF não pode haver identificadores repetidos para uma mesma base URI, ou seja, se o “Datatype Properties” “nome” é propriedade das classes “Pessoa” e “Instituicao”, ocorre o erro de identificador repetido ao processar essa tripla. Dessa maneira, adotou-se uma solução notacionando “nomePessoa” para a entidade “Pessoa” e “nomeInstituicao” para a entidade “Insituicao”. Já para ontologias diferentes com base URI distintas, os nomes das propriedades podem se repetir. (2) Nas ferramentas de construção e implementação de ontologias (como por exemplo, VocBench ou Protégé discutido na subseção 4.2.3.1), as propriedades são exibidas de maneira isolada, dificultando assim que o usuário entenda o que elas representam. Por exemplo, a propriedade “dataInicio” aparece como todas as outras propriedades da classe e sem uma identificação clara do que ela representa, ou seja, só é possível entender do que se trata “dataInicio” ao selecioná-la para poder então verificar o seu domínio.

Utilizou-se de mapeamento em OWL para conectar vocabulários de ontologias diversas. O critério usado na seleção para o reuso desses vocabulários (relacionar ontologias) se deu pela descrição semântica da definição de cada um deles (classes, propriedades, relações e instâncias). Os vocabulários utilizados nas equivalências de classe (equivalentClass)¹¹⁷, de propriedades (DatatypeProperty)¹¹⁸, de relações

¹¹⁷ Mapeamento usado em OWL para equivalência de classes (entidades) (CRISTOVÃO, 2023). Serve para indicar que duas classes podem ser consideradas equivalentes, isto é, as entidades equivalentes têm as mesmas instâncias e usa-se “equivalentClass” para relacionar classes sinônimas (por exemplo, a classe Carro pode ser declarada como equivalente a classe Automóvel. Disponível em: <https://www.w3.org/TR/owl-features/>

¹¹⁸ Mapeamento em OWL usado para equivalência de propriedades (atributos) (CRISTOVÃO, 2023). Permite indicar que duas propriedades podem ser classificadas como equivalentes. Propriedades

(ObjectProperty)¹¹⁹ e de instâncias (sameAs)¹²⁰ apresentados a seguir foram consultados nos portais especializados para consulta de termos e vocabulários para as classificações:

- Vocabulários para interoperabilidade da camada semântica: DBpedia *Resource*¹²¹, GeoNames¹²² e Wikidata *Entity*¹²³.
- Vocabulários para interoperabilidade da camada sintática: OWL¹²⁴, RDF¹²⁵, RDFS¹²⁶ e SKOS¹²⁷.
- Vocabulários para interoperabilidade da camada estrutural: *Bibliographic Framework Initiative* (BIBFRAME)¹²⁸, *Creative Commons Rights Expression Language* (CC)¹²⁹, *Dublin Core* (DC)¹³⁰, *Friend of a Friend* (FOAF)¹³¹ e *Schema*¹³².

equivalentes podem ser usadas para relacionar propriedades sinônimas. Disponível em: <https://www.w3.org/TR/owl-features/>

¹¹⁹ Trata-se do mapeamento em OWL usado para equivalência de relações. São relações entre instâncias de duas classes (entidades). Disponível em: <https://www.w3.org/TR/owl-features/>

¹²⁰ Mapeamento em OWL usado para equivalência de instâncias (indivíduos de classes) (CRISTOVÃO, 2023). Indica que dois indivíduos podem ser ditos iguais. Isso permite a elaboração de vários nomes distintos que se dizem respeito ao mesmo indivíduo. Disponível em: <https://www.w3.org/TR/owl-features/>

¹²¹ Representa recursos informacionais nativos da Wikipedia (CRISTOVÃO, 2021b). Disponível em: <https://www.wikipedia.org/>

¹²² Representa a localização geográfica de um local (CRISTOVÃO, 2021b). Disponível em: <http://www.geonames.org/>

¹²³ Retrata entidades da base Wikidata (CRISTOVÃO, 2021b). Disponível em: https://www.wikidata.org/wiki/Wikidata:Main_Page

¹²⁴ Criado para elaborar ontologias no âmbito da Web semântica (CRISTOVÃO, 2021b). Disponível em: <https://www.w3.org/TR/2012/REC-owl2-quick-reference-20121211/>

¹²⁵ Disponibiliza vocabulários básicos para a modelagem de dados em RDF (CRISTOVÃO, 2021b). Disponível em: <https://www.w3.org/TR/rdf-schema/>

¹²⁶ Estende o vocabulário de RDF representando relacionamentos em um *knowledge graph* (CRISTOVÃO, 2021b). Disponível em: <https://www.w3.org/TR/rdf-schema/>

¹²⁷ Representa sistemas de organização de conhecimento como tesouros, esquemas de classificação, taxonomias, dentre outros. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>

¹²⁸ Utilizado para a descrição bibliografias (CRISTOVÃO, 2021b). Disponível em: <https://id.loc.gov/ontologies/bibframe.html>

¹²⁹ Trata da licença do vocabulário ou ontologia (CRISTOVÃO, 2021b). Disponível em: <https://creativecommons.org/ns>

¹³⁰ Usado na descrição de objetos digitais como, vídeos, sons, imagens, sites na web, dentre outros (CRISTOVÃO, 2021b). Disponível em: <http://dublincore.org/documents/dcmi-terms>

¹³¹ Utilizado para a descrição de dados de pessoas (CRISTOVÃO, 2021b). Disponível em: <http://xmlns.com/foaf/spec/>

¹³² Adotado para representar aplicações como páginas da Web e mensagens de e-mail (CRISTOVÃO, 2021b). Disponível em: <https://schema.org/docs/full.html>

Além dos repositórios de busca acima, foram utilizados os portais *Linked Open Vocabularies* (LOV)¹³³, BioPortal¹³⁴ e o JoinUp¹³⁵, pois contém uma grande quantidade de ontologias, termos e vocabulários de busca global, e também o Prefix.cc¹³⁶, que corresponde a um portal para buscar prefixos de namespaces (CRISTOVÃO, 2021b).

Cabe mencionar que as restrições¹³⁷ apresentadas no modelo de *design*, assim como na ontologia de referência de domínio (subseção 5.4.2) não foram implementadas, porque muitas delas não são possíveis de serem implementadas nesta fase devido aos limites de implementação da própria ontologia operacional. Entretanto, essas restrições podem ser implementadas enquanto sistemas de informação em uma fase mais adiante e que está fora da proposta desta dissertação (em uma linguagem de programação). Cada restrição é importante para orientar no desenvolvimento do sistema de informação, pois o programador poderá se basear nelas para codificar as regras que são fundamentais e que devem ser respeitadas para o sistema funcionar adequadamente (segundo as regras estabelecidas na ontologia de domínio).

5.4.3 Etapa 3: Implementação da ontologia operacional

A Etapa 3 condiz com a implementação da ontologia operacional e trata do objetivo específico “Implementar uma ontologia operacional de domínio baseada no modelo de *design* criado no objetivo específico (III)”. Neste momento usou-se a atividade de Implementação, fase 4 da metodologia SABiO (FALBO, 2014), e a partir do modelo de *design* da ontologia operacional de domínio, utilizou-se o gUFO (subseção 4.2.1.4) para gerar automaticamente o arquivo de implementação da ontologia operacional de domínio no formato *Turtle* (extensão TTL). E em seguida foi realizada manualmente uma limpeza da implementação com o Notepad++ (subseção 4.2.3.2), devido a geração automática de diversos elementos que atrapalham a leitura

¹³³ Disponível em: <https://lov.linkddata.es/dataset/lov/>

¹³⁴ Disponível em: <https://bioportal.bioontology.org/>

¹³⁵ Disponível em: <https://joinup.ec.europa.eu/>

¹³⁶ Disponível em: <http://prefix.cc/>

¹³⁷ Restrições especificadas em linguagem natural e que não foram implementadas nas seguintes entidades da ontologia de referência de domínio e ontologia operacional de domínio: diagrama Defesa: Discente Egresso (DiscenteEgresso), Membro (Membro), Orientador (Orientador), Organização de Ensino (OrganizaçãoEnsino) e Produto (Produto); diagrama Colaborador: Acervo (Acervo) e Documento Impresso (DocumentoImpresso).

e as fases posteriores do desenvolvimento, como entidades (classes) e propriedades replicadas e problemas de formatação no uso de comentários com caracteres especiais. Posteriormente, usou-se o VocBench (subseção 4.2.3.2) para a implementação da ontologia operacional de domínio no formato *Turtle*, realizando-se a depuração do arquivo em formato *Turtle* e o reaproveitamento de elementos definidos e que estavam planejados na fase anterior, mas que não foram gerados automaticamente como “*rdfs:label*” e “*rdfs:comment*” de todas as entidades (classes e propriedades). Assim, codificou-se as classes, propriedades e relacionamentos (A5 e O5) com a inserção de dados da própria base e de ontologias externas e, então, criou-se um documento escrito em linguagem *Turtle*. Os arquivos resultantes da implementação e documentos de dados do acervo de dissertações do estudo de caso (arquivos, código fonte, dados, dentre outros) estão disponíveis em um repositório na plataforma Github¹³⁸.

Ainda, usou-se o SKOS (discutido na subseção 4.2.3.2) para representar as listas de termos para algumas propriedades de classes da ontologia, e também para especificar algumas instâncias (elementos da ontologia domínio), uma vez que o SKOS é indicado para representar sistemas de organização de conhecimento.

5.4.4 Etapa 4: Preparação e reconciliação de dados

Na Etapa 4, Figura 8, aplicou-se os procedimentos de limpeza dos dados do estudo de caso em acervo experimental de dissertações (que representa uma pequena parte do domínio) com o *software* OpenRefine (subseção 4.2.4), como a eliminação de espaços em branco na base de dados, remoção de registros duplicados e agrupamento de diferentes representações da mesma realidade, obtendo-se assim dados limpos e depurados (A6 e O6). Nesse sentido, utilizou-se essa ferramenta para analisar e efetuar as devidas correções no banco de dados de dissertações do PPGCI, e que foi usado para mostrar os resultados por intermédio de consultas SPARQL. Adotou-se o OpenRefine também para realizar tarefas essenciais como separar vários valores contidos no mesmo campo, analisar a distribuição de valores em um conjunto de dados e executar uma das etapas da interoperabilidade de maneira manual.

¹³⁸ GitHub é um repositório de hospedagem de arquivos e código-fonte. Disponível em: <https://github.com/>

Executou-se ainda o recurso de reconciliação¹³⁹ para interligar os dados do estudo de caso com dados de outras bases (repositórios externos). Esse processo foi semiautomático, porque após a combinação dos valores foi necessário apreciar e confirmar os resultados gerados. Essa interligação foi realizada por intermédio do *software* OpenRefine viabilizando a interoperabilidade da camada semântica discutida na subseção 3.3.5 e aplicada no mapeamento da próxima Etapa, resultando assim na reconciliação dos dados com bases externas (A7 e O7).

5.4.5 Etapa 5: Mapeamento para base de dados ligados RDF

A Etapa 5, Figura 8, contemplou o quinto objetivo específico “Gerar uma base de dados ligados na Web semântica a partir de um mapeamento da ontologia operacional implementada no objetivo específico (IV) e de uma base de dados experimental de dissertações com vínculos interoperáveis para elementos externos”. Nesse estágio, construiu-se o esquema de mapeamento dos dados pré-processados com a ontologia operacional de domínio e com os dados reconciliados (A8 e O8) com a aplicação da fase 4 (quatro) de SABiO (subseção 4.2.1.4), ou seja, incorporou-se ao repositório do GraphDB a base em CSV¹⁴⁰ exportada do projeto realizado no OpenRefine.

Em seguida, realizou-se o mapeamento para os dados pré-processados (base experimental de dados ligados) e a ontologia operacional para dados ligados para obter os dados ligados em RDF (A9 e O9) utilizando-se do Ontotext Refine (subseção 4.2.5) e associando elementos de vocabulários controlados e esquemas de metadados externos para atendimento à interoperabilidade da camada estrutural (subseção 3.3.5). Esse procedimento estabeleceu o mapeamento do modelo de classes e suas propriedades (entidades, atributos e relacionamentos) com outras classes e propriedades de ontologias externas, bem como bibliotecas especializadas selecionadas. Os processos de mapeamento foram realizados usando o banco de

¹³⁹ Reconciliação trata-se do processo de combinar o conjunto de dados da própria base com os dados de um base externa, isto é, dados gerados por bibliotecas, instituições científicas, organizações acadêmicas, organizações diversas, arquivos, museus, com um conjunto de dados local (GLASS, 2021).

¹⁴⁰ CSV é a sigla de *comma-separated values* (valores separados por vírgula), ou também *character separated values* (caracteres separados por valores) para enfatizar que outros caracteres, além da vírgula podem ser usados como ponto-e-vírgula, tabulação, dentre outros. É um dos tipos de formatos de bases de dados estruturados mais populares para publicar dados na Web, pois é conciso, fácil de se entender por computadores e humanos, e também se adere à natureza tabular da maioria dos dados. Disponível em: <https://www.w3.org/TR/tabular-data-primer/>

dados orientado a grafos GraphDB (subseção 4.2.6), e contou com um procedimento manual para relacionar uma instância dos dados pré-processados (dado de uma coluna) com a ontologia operacional para dados ligados em RDF, sendo que todas as demais instâncias (outros instâncias dos dados pré-processados) foram mapeadas por meio de uma função automática disposta no GraphDB. Esse processo foi custoso, e à medida que aumentava-se a quantidade de elementos mapeados, o funcionamento do GraphDB tornou-se mais lento, ou apresentava-se erros. Para aumentar a performance no mapeamento e também vencer os desafios encontrados pelos erros ocasionados devido à quantidade de mapeamento realizado, adotou-se a estratégia de dividir o mapeamento em mais de um arquivo (arquivos menores funcionam mais rápidos e evitaram-se os erros promovidos pela quantidade de instâncias mapeadas).

Assim, criou-se um repositório de dados ligados RDF (A10 e O10) no GraphDB com a importação da ontologia operacional de domínio (Etapa 3) criada no formato *Turtle* (extensão TTL) no VocBench (subseção 4.2.3.1).

5.4.6 Etapa 6: Recuperação de Informação

A Etapa 6, Figura 8, corresponde ao último objetivo específico que é elaborar consultas SPARQL (subseção 4.2.7) no repositório de dados ligados em RDF como prova de conceito para validar a RI. Nesse estágio utilizou-se a fase 5 da estratégia SABiO que é conhecida como Teste. Ela trata da validação da ontologia operacional por meio de prova de conceito ao retornar dados ligados recuperados (A11 e O11). Assim, aplicou-se a prova de conceito no repositório de dados ligados em RDF, executando consultas SPARQL e apresentando resultados para o acervo do estudo de caso.

As consultas foram realizadas na base de dados ligados das dissertações dos alunos egressos do PPGCI/UFES, selecionado alguns indivíduos (alunos, professores, dissertações, etc) por meio do GraphDB (subseção 4.2.6). Esse procedimento foi validado com base nos resultados obtidos, mesmo considerando-se uma base de dados pequena do estudo de caso. Os resultados foram apresentados com o Visual Graph (subseção 4.2.8) e, assim, foi possível validar a ontologia operacional.

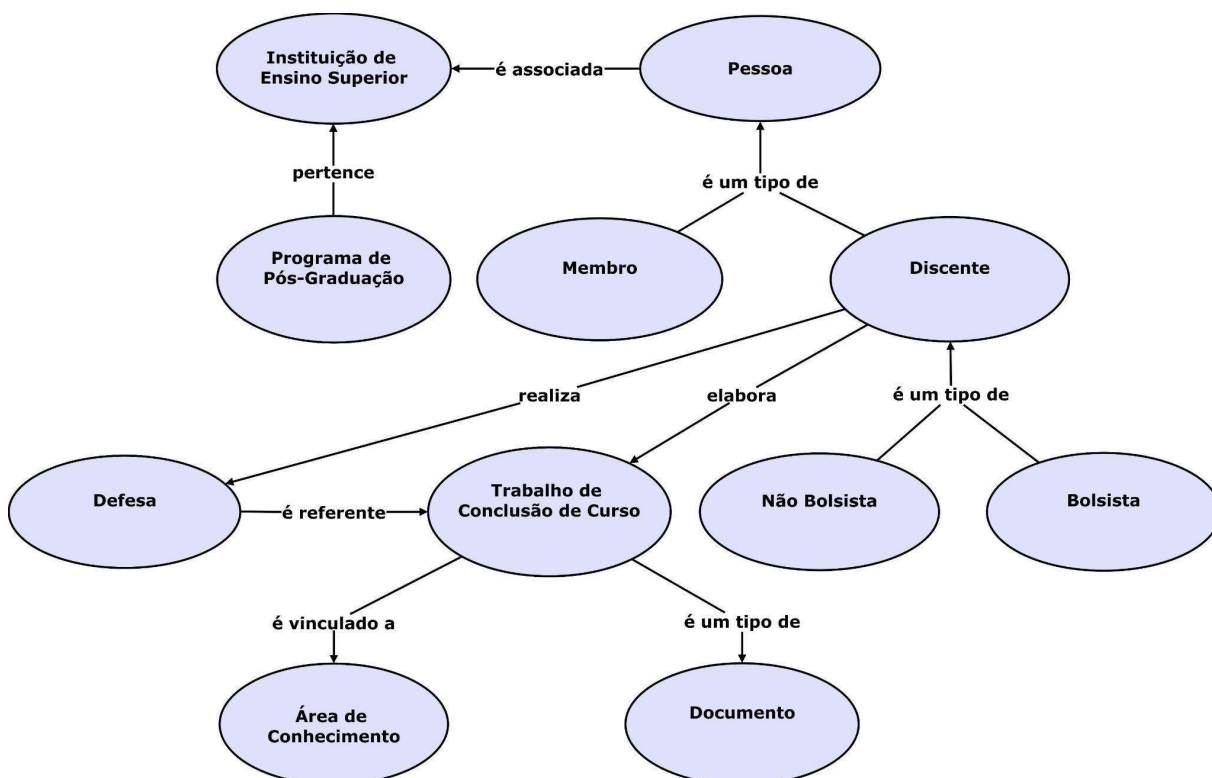
6 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados e discussão sobre os resultados da pesquisa. Esses resultados são oriundos da execução das seis etapas procedimentais do desenvolvimento (subseção 5.4) e são descritos detalhadamente nas próximas subseções.

6.1 ONTOLOGIA DE REFERÊNCIA DO DOMÍNIO

Inicialmente foi criado um modelo conceitual preliminar contendo as principais entidades e relações do acervo de dissertações do estudo de caso. A Figura 9 apresenta o *knowledge graph* (O1, Quadro 4) resultante.

Figura 9 - *Knowledge graph* do modelo conceitual preliminar do domínio ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

O Apêndice C, Quadro 11, explica de maneira sucinta os termos utilizados no *knowledge graph* desenvolvido.

A Figura 9 apresenta os tipos de entidades Instituição de Ensino Superior, Programa de Pós-Graduação, Pessoa, Discente, dentre outras, e os relacionamentos entre elas. Observa-se que um Programa de Pós-Graduação pertence a uma Instituição de Ensino Superior, assim como a entidade Pessoa é associada a alguma

Instituição de Ensino Superior, podendo desempenhar diferentes papéis nessa relação. Durante o levantamento de requisitos¹⁴¹ detalhado foi possível perceber que existem mais entidades envolvidas e relacionamentos, tais como o projeto que uma bolsa pode estar vinculada, a publicação em edital, o caso de edital e o projeto correspondente a uma bolsa de estudos.

A partir da compreensão fornecida pelo *knowledge graph* (O1, Quadro 4), que possui maior flexibilidade de regras de modelagem, construiu-se uma ontologia de referência de domínio¹⁴² (O2, Quadro 4) cuja representação está descrita nas Figuras 10, 11, 12 e 13. Trata-se de uma ontologia de referência de domínio (O2, Quadro 4) elaborada com o uso da linguagem OntoUML (ver subseção 4.2.1.3), e embasada com diretrizes de UFO (discutida na subseção 4.2.1.2), e que busca representar com maior detalhamento e expressividade a realidade do estudo de caso - trabalhos de conclusão de curso de pós-graduação stricto sensu.

Nas Figuras de 10 a 13 é possível perceber maior formalidade com relação ao modelo anterior (o modelo conceitual preliminar), pois além de conter entidades e relacionamentos, ela identifica estereótipos UFO (tais como *kind*, *subkind*, *role*, *relator*), cardinalidades e atributos (por exemplo, a sigla de uma Instituição de Ensino Superior).

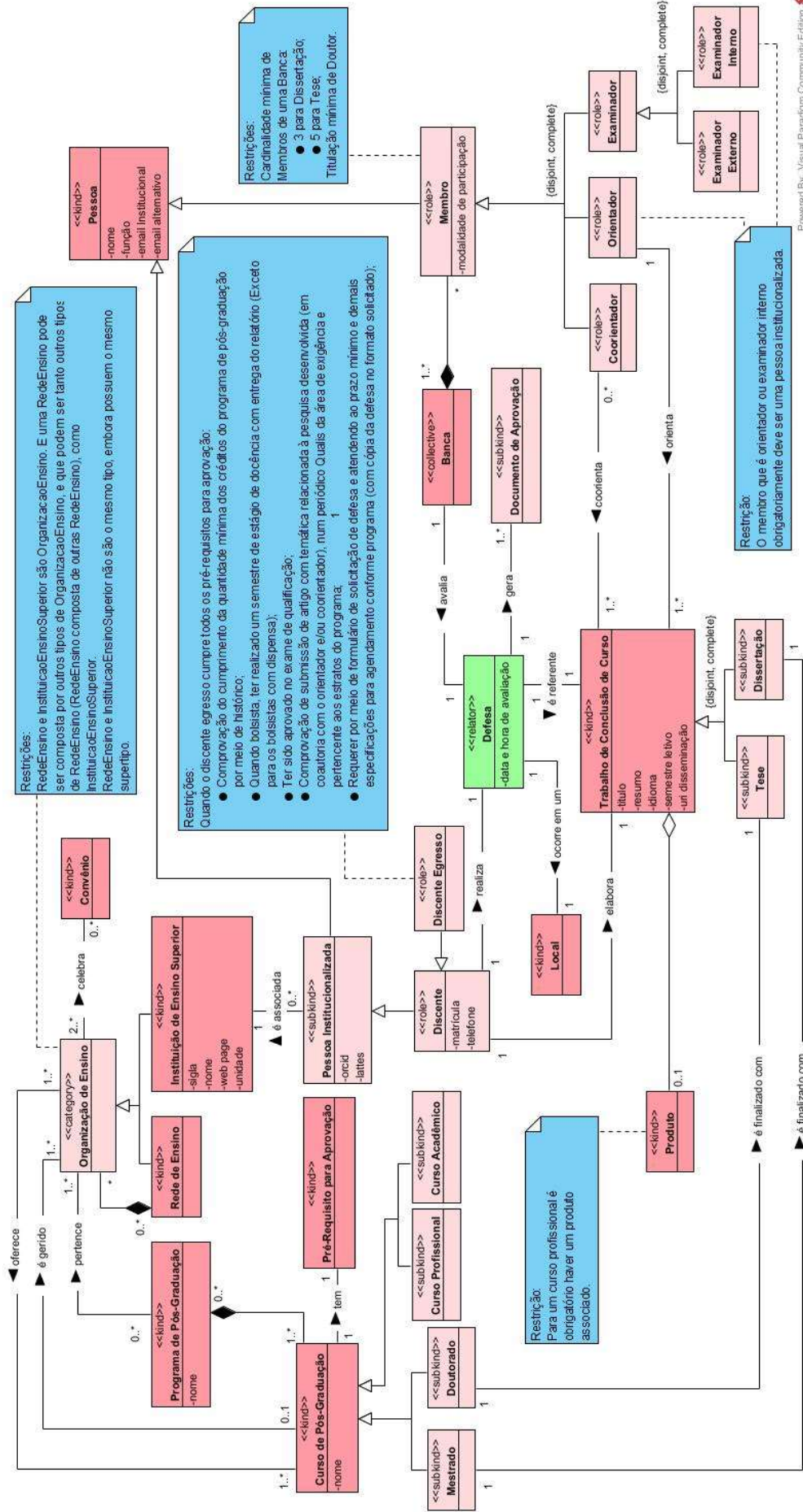
Na Figura 10 destacam-se entidades como Discente, Defesa, Trabalho de Conclusão de Curso. Bem como, notam-se atributos em tais entidades. Discente tem os atributos matrícula e telefone, enquanto Defesa possui a característica data e hora de avaliação.

¹⁴¹ Levantamento de requisitos é o procedimento de descoberta das funções que o sistema deve realizar (requisitos funcionais), bem como as restrições dessas funções (requisitos não funcionais) (WAZLAWICK, 2011).

¹⁴² A ontologia de referência de domínio completa foi desenvolvida no Visual-Paradigm e está disponível em:

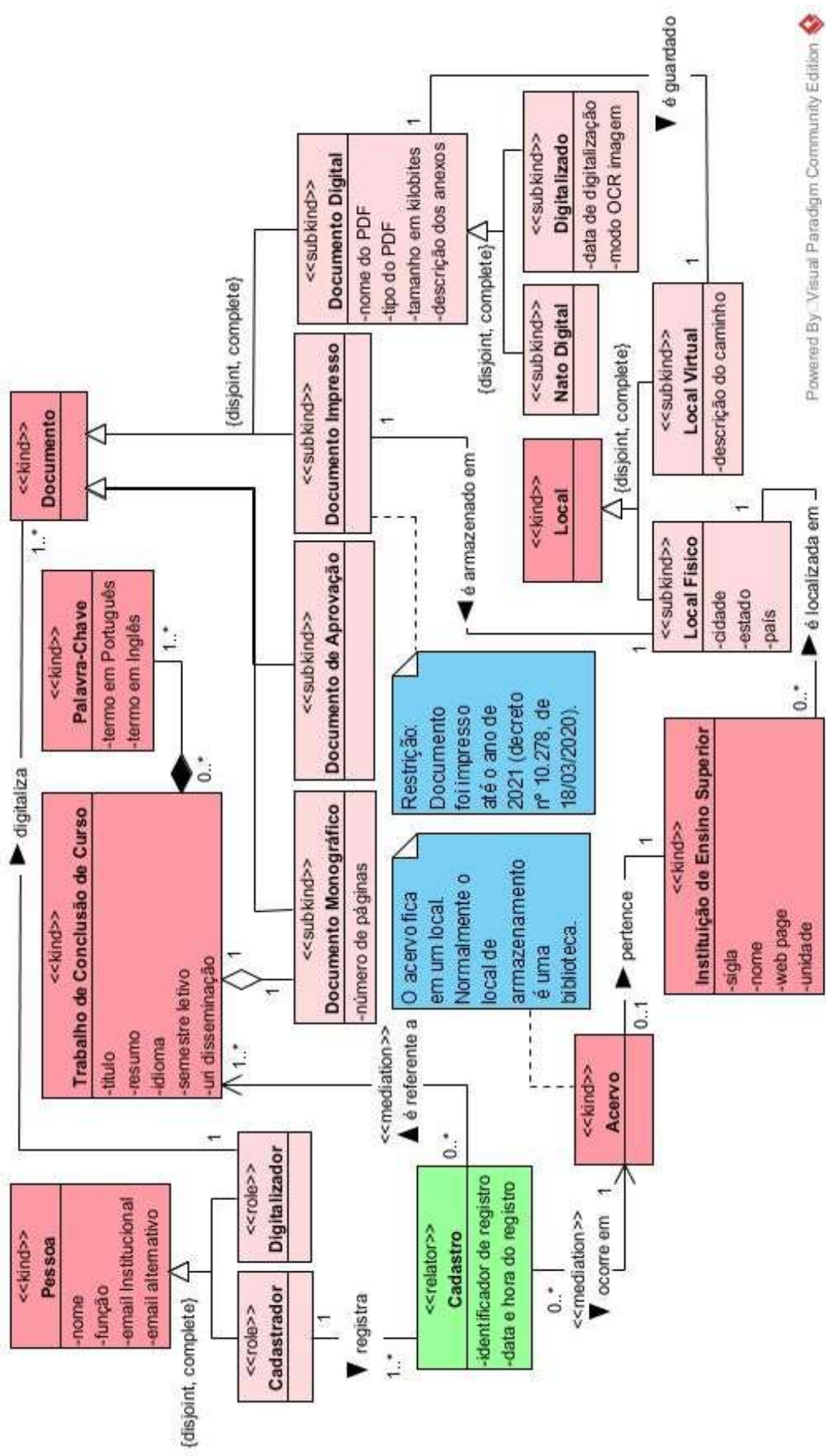
<https://github.com/rafaeldenersons/ontologiaReferenciaDominioCompleta/blob/main/2.1ontologiaReferencia.vpp>

Figura 10 - Ontologia de referência de domínio - diagrama Defesa (link para acesso em tamanho maior)



Fonte: Elaborado pelo próprio autor.

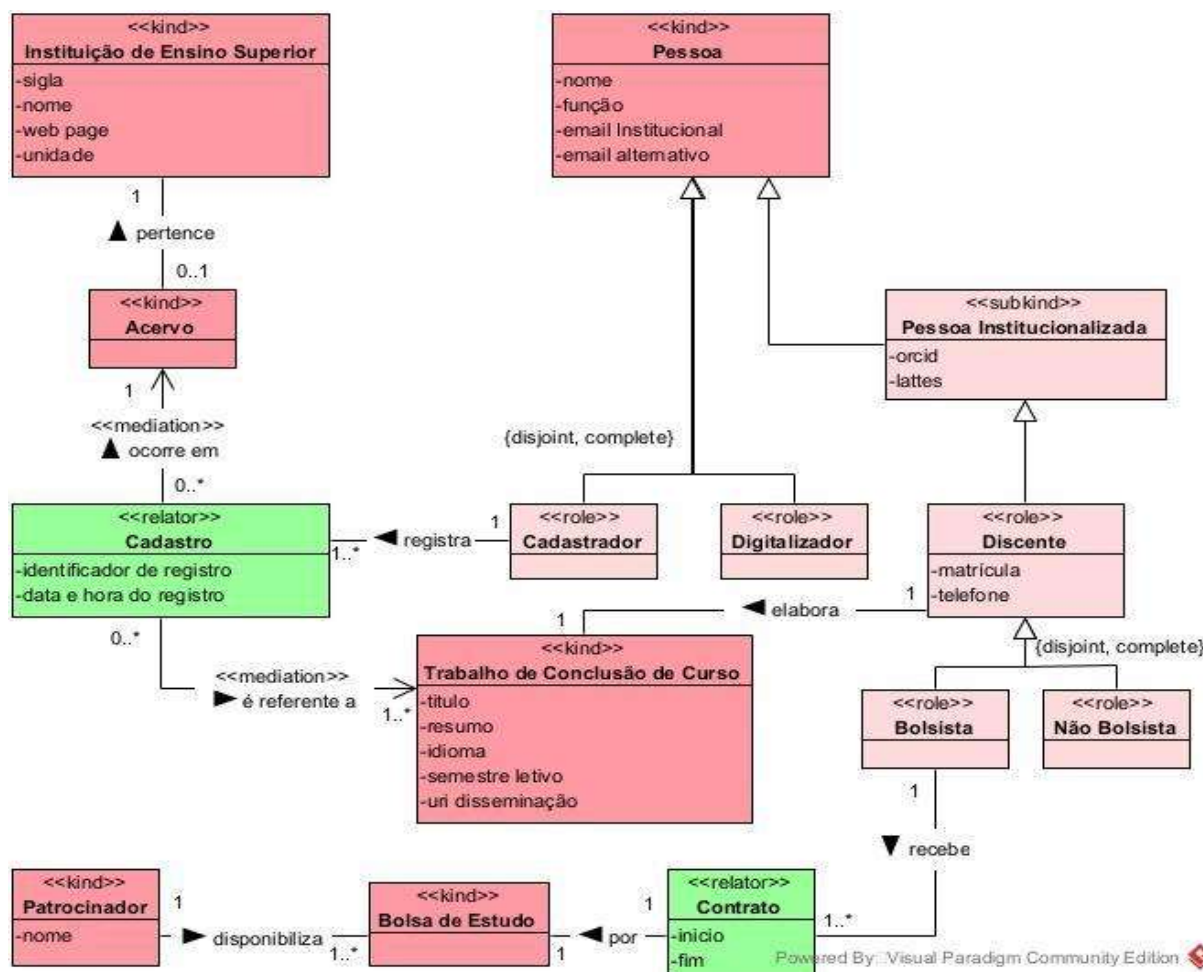
Figura 11 - Ontologia de referência de domínio - diagrama Colaborador (link para acesso em tamanho maior)



Fonte: Elaborado pelo próprio autor.

Na Figura 11 é possível perceber entidades como Cadastrador, Cadastro - que complementam a atividade central da defesa, e novamente Trabalho de Conclusão de Curso, fazendo uma ligação com o diagrama da Figura 10. Ressalta-se que Trabalho de Conclusão de Curso tem os atributos título, resumo, idioma, semestre letivo e iri disseminação, enquanto Cadastro tem as características identificador de registro e data e hora de registro, sendo que cada Trabalho de Conclusão de Curso é referente a um Cadastro, possui um Documento Monográfico e várias Palavras-Chave associadas.

Figura 12 - Ontologia de referência de domínio - diagrama Bolsa ([link para acesso em tamanho maior](#))

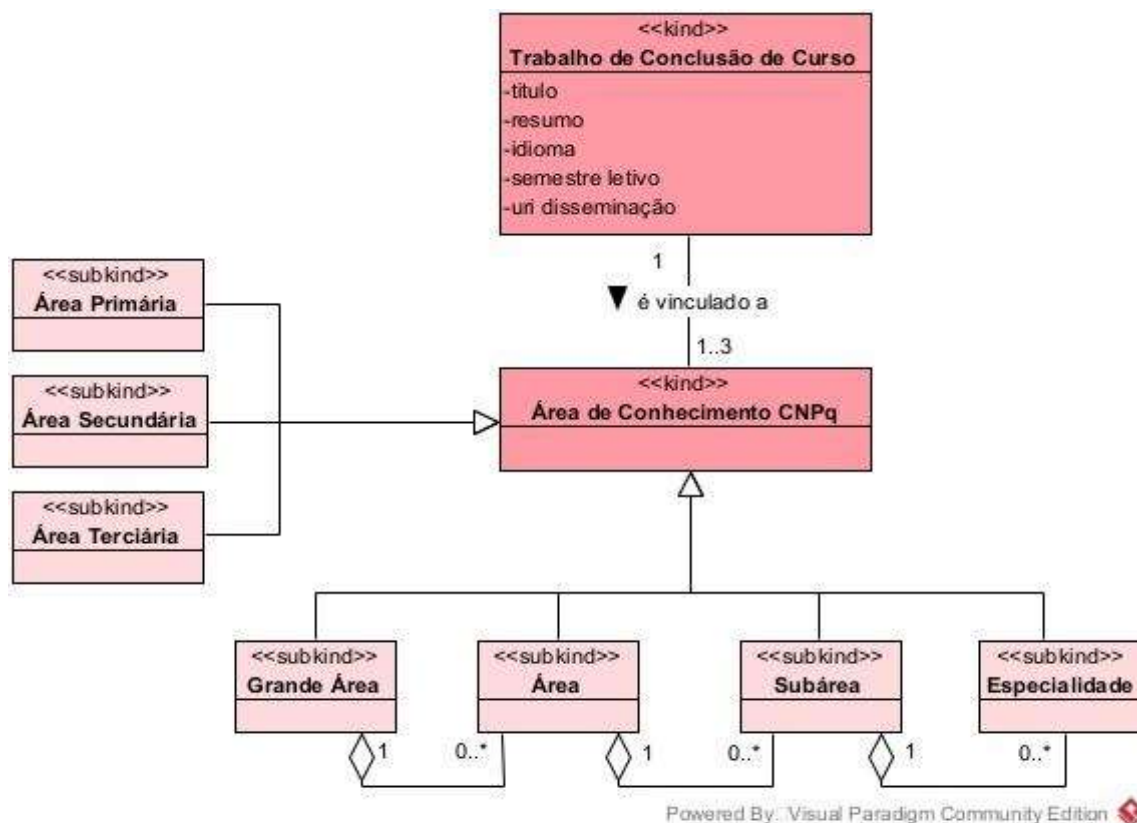


Fonte: Elaborado pelo próprio autor.

A Figura 12 apresenta entidades como Discente, Não Bolsista, Bolsista, Contrato, Bolsa de Estudo, também conceitos complementares ao Trabalho de Conclusão de Curso (que atua como elemento central de ligação entre os demais conceitos). A princípio não foram identificados atributos para o Bolsista, o mesmo

ocorrendo com Bolsa de Estudo (o que não significa que não há atributos em tais tipos, mas apenas que eles não foram julgados pertinentes ao contexto), enquanto Contrato detém propriedades como início e fim da vigência do contrato, visto que um Bolsista recebe um ou mais Contrato(s) por Bolsa de Estudo.

Figura 13 - Ontologia de referência de domínio - diagrama Área ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

A ontologia da Figura 13 representa entidades como Trabalho de Conclusão de Curso (novamente o elo de ligação entre diagramas) e Área de Conhecimento CNPq. Nota-se que um Trabalho de Conclusão de Curso deve ser vinculado a pelo menos uma Área de Conhecimento CNPq (podendo ser até três), denominadas de Área Primária, Área Secundária e Área Terciária. Cada uma dessas áreas é representada por uma Grande Área, que pode ser especializada em Área, Subárea e Especialidade.

O uso da metodologia SABiO proposta por Falbo (2014) para elaborar a ontologia de referência do estudo de caso e com respaldo em uma ontologia de fundamentação, segue em linha similar com outras pesquisas em andamento no contexto educacional (SILVA; MUTZ; RUY, 2022a; SILVA; MUTZ; RUY, 2022b; SILVEIRA; ARAGÓN; CURY; MENEZES, 2021; SILVEIRA; CURY; MENEZES, 2019).

A ontologia desenvolvida atende ao o primeiro objetivo específico desta pesquisa, e foi utilizada como entrada para artefatos trabalhados em atividades posteriores, como identificado nas seções seguintes.

6.2 MODELO DE *DESIGN* PARA A ONTOLOGIA OPERACIONAL

A partir do modelo conceitual fornecido pela ontologia de referência de domínio (O3, Quadro 4), desenvolvida de maneira formal e com regras de modelagem, iniciou-se a construção do modelo de *design* da ontologia operacional de domínio¹⁴³ (O3, Quadro 4) discutido na subseção 5.4.2, tendo como base a fase 3 da metodologia SABiO do Projeto (*Design*) e incorporando elementos de interoperabilidade (O4, Quadro 4), discutidos na subseção 4.2.3.

Esse modelo de *design* usa a linguagem OntoUML, com as diretrizes UFO, conforme já discutido na seção 4.2.1. Ele possui associação com itens de vocabulários controlados, esquemas de metadados de outras bases para atender à interoperabilidade da camada estrutural, bem como a ligação de instâncias e entidades de bases externas (ação de reuso da engenharia de ontologias) para atender a interoperabilidade da camada semântica (discutida na subseção 5.4.2 e fundamentada na subseção 3.3.5).

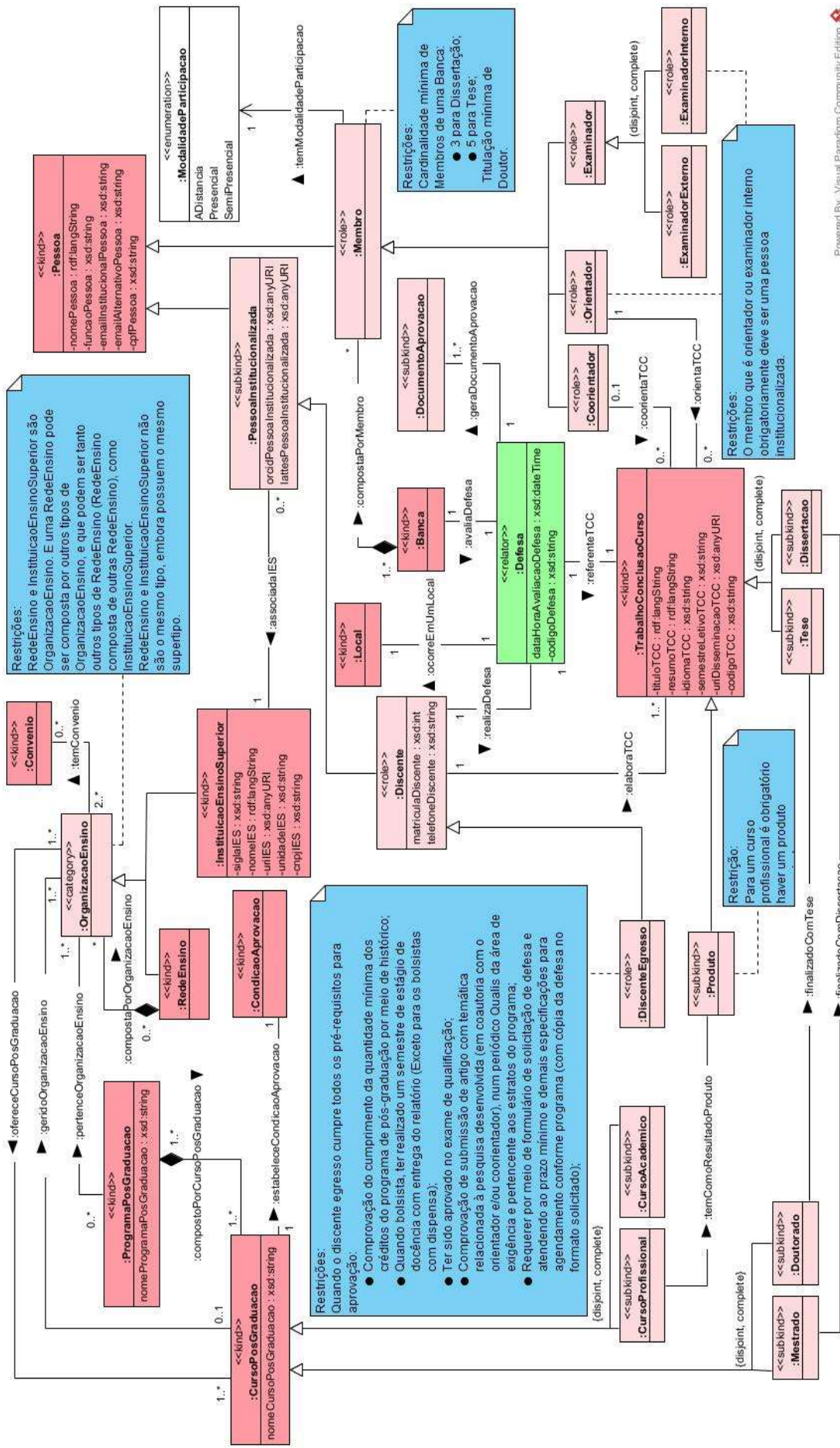
Cada um dos quatro diagramas da ontologia de referência de domínio (O2, Quadro 4), Figuras 10 a 13, possui um modelo de *design* correspondente (O4, Quadro 4), Figuras 14 a 17, respectivamente.

A Figura 14 apresenta o diagrama Defesa do modelo de *design*, equivalente ao modelo de referência de domínio da Figura 10 (O2, Quadro 4). Nota-se as entidades com a definição em que foram implementadas como :Discente, :Defesa, :Trabalho de Conclusão de Curso, etc. É possível perceber também que os atributos das entidades estão tipificados em cada entidade. Como por exemplo, a entidade :Trabalho de Conclusão de Curso possui o atributo codigoTCC : xsd:string, sendo codigoTCC o próprio nome a ser codificado na implementação, enquanto xsd:string¹⁴⁴ representando o tipo de dado que essa propriedade representa.

¹⁴³ Modelo de *design* da ontologia operacional de domínio completo e desenvolvido no Visual-Paradigm. Disponível em: https://github.com/rafaeldenersons/modeloDesignCompleto/blob/main/2.2modeloDeDesign_ontologiaOperacional.vpp

¹⁴⁴ xsd:string é o tipo de dados para representar caracteres. Disponível em: <https://www.w3.org/TR/xmlschema-2/>

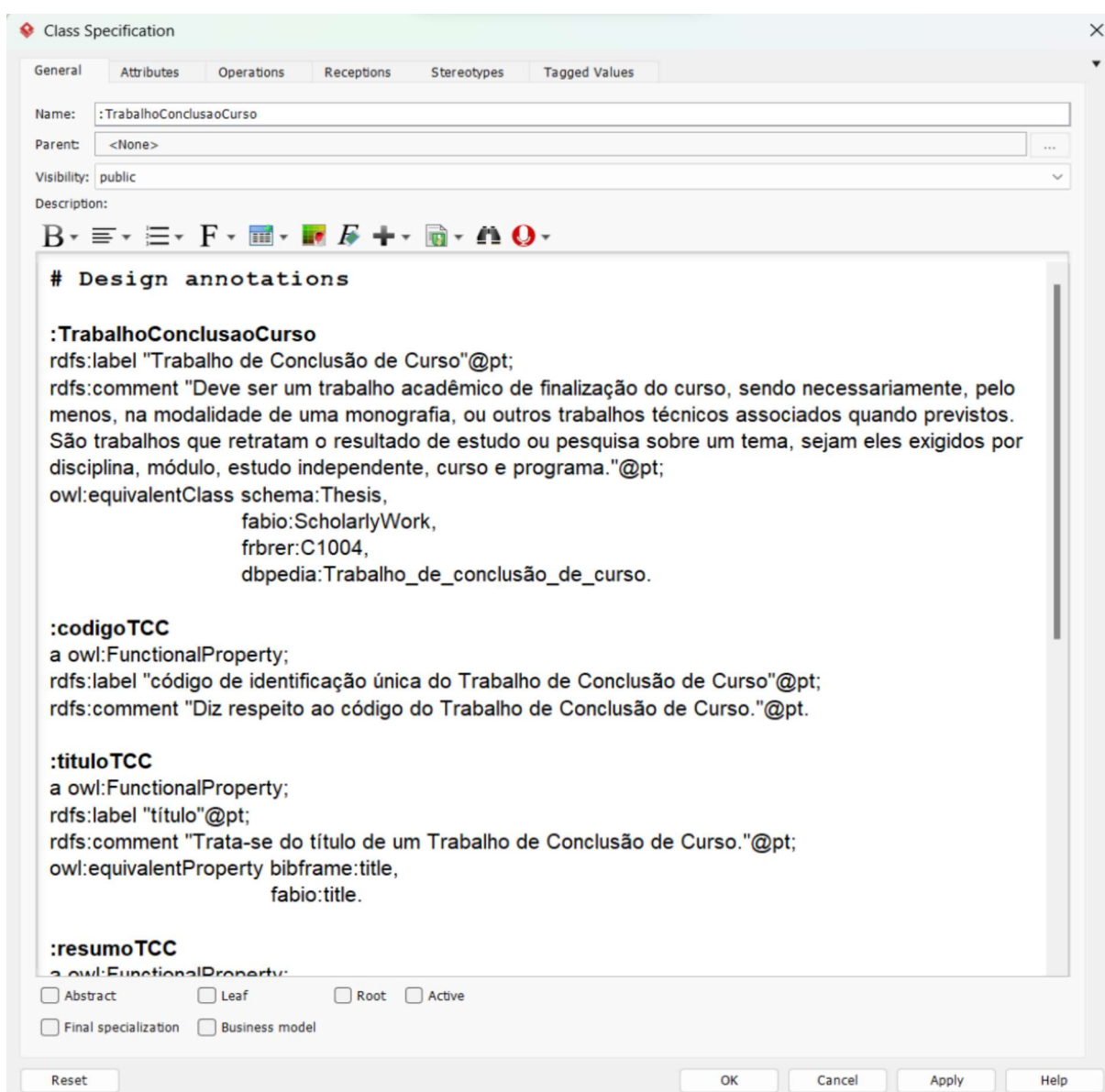
Figura 14 - Modelo de design - diagrama Defesa (link para acesso em tamanho maior)



Fonte: Elaborado pelo próprio autor.

Os procedimentos executados para pesquisar, selecionar e relacionar os vocabulários controlados junto ao modelo de *design* desta pesquisa são confluentes com fundamentos e discussão apresentados na subseção 3.3.2. Como trata-se do modelo de *design* da ontologia operacional de domínio (O4, Quadro 4) e que foi usado na implementação, todas as anotações de implementação foram especificadas para cada entidade (*Class Specification* do diagrama Defesa)¹⁴⁵.

Figura 15 - Especificação das entidades - anotações de implementação do modelo de *design* da ontologia operacional de domínio ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

¹⁴⁵ *Class Specification* do diagrama Defesa. Disponível em:
https://github.com/rafaeldenersons/3_ontologiaOperacionalDomnio/blob/main/1designAnnotationsDefesa.txt

A Figura 15 apresenta as anotações do modelo de *design* da ontologia operacional de domínio específicas para a entidade :TrabalhoConclusaoCurso, tendo o elemento rdfs:label para definir o label da entidade como “Trabalho de Conclusão de Curso” em português (pt). A propriedade rdfs:comment conecta a entidade a um comentário explicativo. A propriedade owl:equivalentClass para relacionar as classes externas (entidades definidas em bibliotecas especializadas).

É possível perceber que os relacionamentos também estão notacionados conforme a definição em que foram implementados. No modelo da Figura 15 o relacionamento “:Discente :elaboraTCC :TrabalhoConclusaoCurso” quer dizer “Discente elabora TrabalhoConclusaoCurso”.

As especificidades da propriedade :elaboraTCC (*Association Specification*) estão na Figura 16. Cada anotação de implementação das relações do tipo *Object Property* foi especificada conforme o exemplo :elaboraTCC mostrado na Figura 16.

O exemplo da Figura 16 também mostra as características das propriedades com axiomas para raciocinar (*Reasoner*) e que são usadas em *objectProperties*. owl:AsymmetricProperty: refere-se a uma propriedade assimétrica, isso corresponde a aplicação em apenas uma direção. Por exemplo, se :Discente elabora :TrabalhoConclusãoCurso, então não pode ocorrer de :TrabalhoConclusãoCurso elaborar :Discente.

Ainda, existe anotações quanto a propriedade rdfs:label para indicar o nome da relação (relacionamento) que é :elaboraTCC, contém o elemento rdfs:comment para indicar um comentário explicativo sobre a relação (relacionamento), e também o componente owl:equivalentPropert para ligar a relacionamentos de bibliotecas especializadas externas.

Figura 16 - Especificação das relações – anotações de implementação do modelo de *design* da ontologia operacional de domínio ([link para acesso em tamanho maior](#))

Fonte: Elaborado pelo próprio autor.

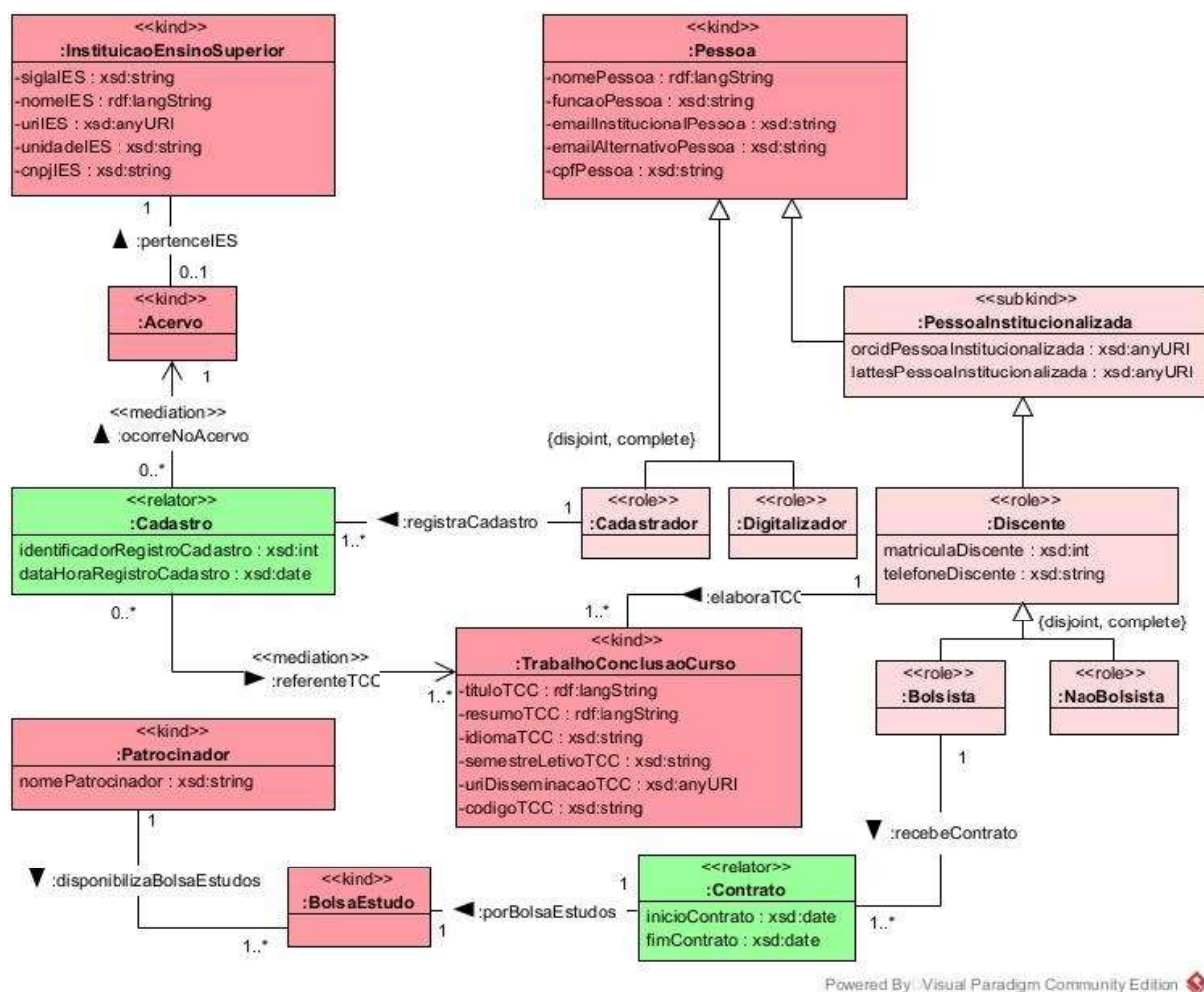
O diagrama colaborador, Figura 17, foi baseado na ontologia de referência de domínio (O2, Quadro 4), Figura 11, e possui um *Class Specification* do diagrama Colaborador¹⁴⁶.

¹⁴⁶ *Class Specification* do diagrama Colaborador. Disponível em: https://github.com/rafaeldenersons/3_ontologiaOperacionalDomnio/blob/main/2designAnnotationsColaborador.txt

No diagrama da Figura 17 nota-se a adoção de entidades enumeradas para representar listas SKOS (discutido na subseção 4.2.3.2) como `:OrganizacaoAcademica`, `:CategoriaAdministrativa` e `:TipoPDF`. Percebe-se que `:OrganizacaoAcademica` representa uma lista enumerada com os respectivos elementos `CentroFederalEducacaoTecnologica`, `CentroUniversitario`, `Faculdade`, `InstitutoFederalEducacaoCienciaTecnologia` e `Universidade`. E `:InstituiçãoEnsinoSuperior` está relacionada a `:OrganizacaoAcademica` por meio da propriedade `:temOrganizacaoAcademica`.

A Figura 18 apresenta o diagrama Bolsa do modelo de *design*. Da mesma forma, existe a *Class Specification* da bolsa¹⁴⁷ para detalhar elementos do modelo.

Figura 18 - Modelo de *design* - diagrama Bolsa ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

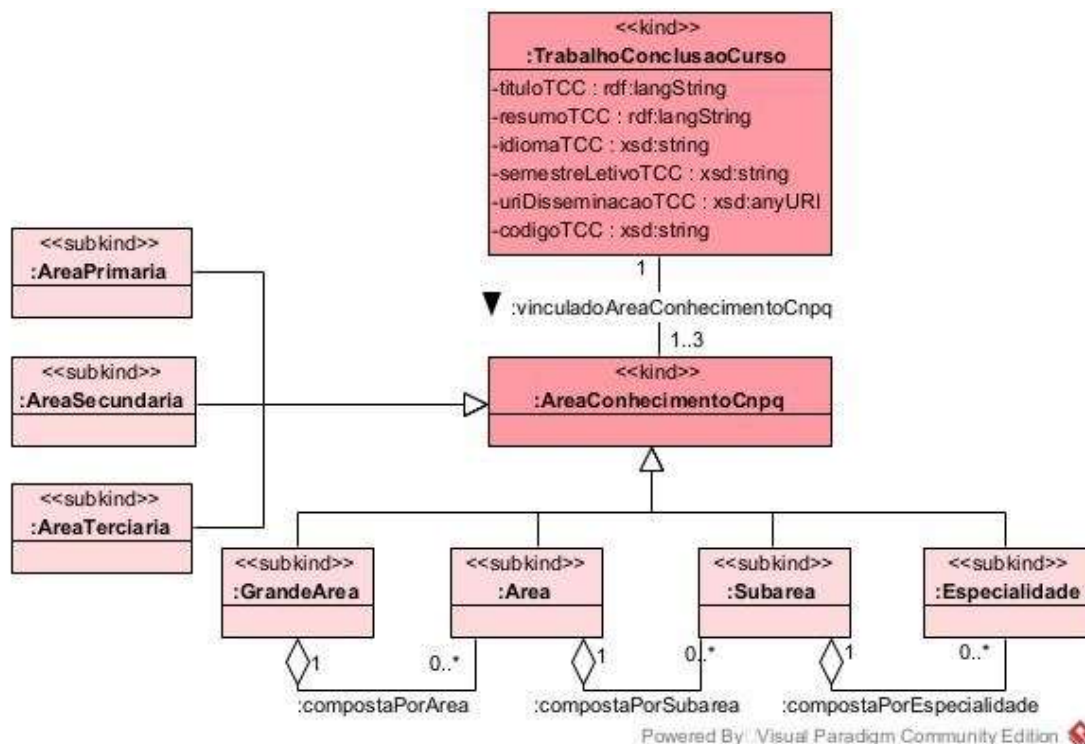
¹⁴⁷ *Class Specification* do diagrama Bolsa. Disponível em: https://github.com/rafaeldenersons/3_ontologiaOperacionalDomnio/blob/main/3designAnnotationsBolsa.txt

Nota-se na Figura 18 entidades como :Bolsista, :Contrato, :BolsaEstudos, dentre outras. Exibe-se que as entidades :Bolsista e :BolsaEstudos, não detém propriedades, enquanto a entidade :Contrato apresenta as propriedades “inícioContrato : xsd:date” e “fimContrato : xsd:date”, sendo “inícioContrato” e “fimContrato” indicando o nome de cada propriedade e “xsd:date” mostrando que as propriedades mencionadas são do tipo data.

Enfim, a Figura 19 representa o complemento do modelo de *design* do diagrama Área (O4, Quadro 4). Percebe-se que há maior representatividade para o relacionamento entre as entidades :GrandeArea, :Area, :Subarea e :Especialidade, pois as composições compostaPorArea, compostaPorSubarea e compostaPorEspecialidade são representadas por uma escrita única, e isso é preciso para realizar a implementação. Nela observa-se que apenas :TrabalhoConclusaoCurso tem atributo, que é codigoTCC : xsd:string, sendo que ela está relacionada com no mínimo uma, e no máximo três :AreaConhecimentoCnpq por meio da relação :vinculada. Esse diagrama detém o *Class Specification* do diagrama Área ¹⁴⁸.

¹⁴⁸ *Class Specification* do diagrama Área. Disponível em:
https://github.com/rafaeldenersons/3_ontologiaOperacionalDomnio/blob/main/4designAnnotationsAreaDeConhecimentoDeTrabalhoAcademico.txt

Figura 19 - Modelo de *design* - diagrama Área ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

O resultado da investigação e seleção das equivalências de vocabulários para o modelo de *design* estão expressos em cada entidade e relação dos diagramas Defesa, Colaborador, Bolsa e Área, e correspondem às figuras 14, 17, 18 e 19 respectivamente. Os Quadros 5, 6, 7 e 8 elucidam o uso dos vocabulários de bibliotecas e ontologias externas, bem como as equivalências de classes e propriedades de dados e objetos (equivalentClass, DatatypeProperty, ObjectProperty), e que foram utilizados no modelo de *design*. O Quadro 5 apresenta os vocabulários que são referenciados nas equivalências dos Quadros 6, 7 e 8.

Quadro 5 - Vocabulários de termos usados na interoperabilidade do modelo de *design*

Namespace do Vocabulário ¹⁴⁹	Descrição
itsmo	Assuntos de gerenciamento de serviços de TI.
grddl	Técnica para coletar descrições de recursos de dialetos de idiomas (linguagens).

¹⁴⁹ Os endereços Web dos referidos vocabulários estão disponíveis nas primeiras 52 linhas do arquivo que contém a ontologia operacional finalizada (abordada na subseção 6.3), disponível em: https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/export.ttl

rdf	Esquema RDF para os termos do vocabulário RDF no RDF Namespace.
owl	Descreve classes internas e propriedades que formam a base da sintaxe RDF/XML de OWL 2.
xsd	Especifica a linguagem de definição de esquema XML (<i>XML Schema Definition Language</i>).
skos	Modelo de dados para compartilhar e interligar sistemas de organização de conhecimento por intermédio da Web semântica.
rdfs	RDF Schema (RDFS) disponibiliza um vocabulário de modelagem de dados para dados em RDF.
dc	Esquema de metadados que descreve diversos tipos de objetos digitais.
schema	Esquema de metadados que tem como objetivo criar, manter e promover esquemas para dados estruturados na Internet, páginas da Web, mensagens de e-mail, dentre outros.
dqm	Descreve vocabulário de gerenciamento de qualidade de dados.
caresses	Ontologia que objetiva descrever recursos específicos para metas, operadores, ações e parâmetros para lidar com a interação verbal entre o robô e o usuário.
data	Trata-se do prefixo do endereço de URI usado para disponibilizar os dados da ontologia operacional de domínio de dissertações do estudo de caso.
pext	Esquema de metadados usados em relatório técnico gerado para serviço de documentação RIF e OWL.
frapo	Ontologia de financiamento, administração de pesquisa e projetos (<i>Funding, Research Administration and Projects Ontology - FRAPO</i>).
gufo	É uma implementação leve da UFO.
iol	Ontologia para objetos de informação, codificações e realizações.
frbrer	Esquema para um conjunto de elementos de classes e propriedades RDF nativas descritas do modelo de entidade-relacionamento de requisitos funcionais para registros bibliográficos (<i>Functional Requirements for Bibliographic Records - FRBR</i>).
meb	Ontologia que fornece elementos básicos para o gerenciamento de conteúdo.
cwrc	Ontologia de colaboração canadense de pesquisa escrita (<i>Canadian Writing Research Collaboratory - CWRC</i>).

sur	Esquema usado para instrumentos de pesquisas inovadores. Descreve abordagens de conversação, questionários de entrevistas e bate-papos.
ptop	Esquema de relatório técnico gerado para serviço de documentação RIF e OWL.
essglobal	Esquema de vocabulários usados na economia social e solidária.
whois	Esquema de vocabulários experimentais para descrever perfil ou histórico de uma pessoa. Define propriedades que descrevem o estágio da vida de alguém.
rico	Esquema de metadados do conselho internacional de arquivos e registros em ontologia de contextos (<i>Records in Contexts Ontology</i> - RICO) ICA RiC-O.
pproc	Ontologia de finanças públicas. Define os conceitos necessários para descrever o processo de compras e contratos do setor público (<i>public e-procurement</i>).
xapi	Ontologia criada para representar a descrição de dados estruturados publicados na Web como base de conhecimento.
vivo	Ontologia para descrever descoberta de pesquisadores.
seo	Ontologia de eventos científicos (<i>Scientific Events Ontology</i> - SEO).
foaf	Esquema baseado no uso de <i>homepages</i> da Web legíveis por máquina para pessoas, grupos, empresas e outros elementos.
scoro	Esquema que define tipo, categoria ou a natureza de uma contribuição feita com relação à autoria de uma publicação.
km4c	Modelo de conhecimento para descrever cidades inteligentes quanto a interconexão, armazenamento e pesquisa de dados de diversas fontes, portais de região, dentre outros.
fabio	FRBR-aligned Bibliographic Ontology (FaBiO) é uma ontologia para definir entidades que são publicadas ou potencialmente publicáveis no âmbito educacional, como artigos, livros, produções acadêmicas, etc.
ov	Modelo para especificar URI aberta para vocabulários que atendam a esquemas RDF ou documentos XML Namespace.
teach	Esquema de vocabulário básico para ensino com a tecnologia RDF do W3C.
gn	Ontologia que possibilita adicionar informações semânticas geoespaciais.

voag	Esquema que destina-se especificar licenciamento, atribuição, proveniência e governança de ontologia.
dbpedia	Plataforma de catalogação de dados para desenvolvedores e consumidores de dados.
poder	Descreve vocabulários para expressar informações sobre pessoas e suas relações com organizações, empresas, editora, ontologias e artefatos.
sro	SALT <i>Rhetorical Ontology</i> (SRO) descreve a retórica e a argumentação.
base-gleif	Ontologia que define conceitos genéricos para reaproveitamento de outras ontologias da <i>Global Legal Entity Identifier Foundation</i> (GLEIF).
cwmo	<i>Creative Workshop Management Ontology</i> (CWMO) é uma ontologia elaborada para descrever o domínio da produção criativa.
dcndl	Declaração de esquema RDF para termos de metadados da <i>National Diet Library</i> (NDL) no Japão.
wde	Banco de dados gratuito, colaborativo, multilíngue e com dados estruturados para fornecer suporte para a Wikipedia, Wikimedia Commons e outros movimentos Wikimedia.
bibframe	Esquema de vocabulários para RDF no contexto de bibliotecas, bibliotecários e catalogação com lista ordenada de classes e propriedades.
crsw	Esquema da ReSIST <i>Courseware Ontology</i> que representa os diversos cursos e recursos educacionais.
mrel	Esquemas de termos de relatoria e especificações que estabelecem a relação entre um agente e um recurso bibliográfico.
dbo	Ontologia gerada com base nas especificações criadas na <i>DBpedia Mappings Wiki</i> . Contém dados de instância de diferentes versões de idioma da Wikipedia.

Fonte: Elaborado pelo próprio autor.

Quadro 6 - Equivalência de classe usada no modelo de *design* com *equivalentClass*

Classe do modelo de <i>design</i>	Classes equivalentes
:Acervo	fabio:WorkCollection, fabio:DocumentRepository.
:AreaConhecimentoCnpq	bibframe:Topic, sio:SIO_001340, dbo:Area.
:Banca	fabio:ExaminationPaper, scoro:ContributionSituation.

:BolsaEstudo	frapo:scholarship, frapo:Studentship.
:Bolsista	scoro:FinancialRole, sio:SIO_000406.
:Cadastrador	caresses:User.
:Cadastro	fabio:ArchivalRecordSet.
:CondicaoAprovacao	schema:coursePrerequisites, crsw:Pre-requisite.
:Contrato	wdo:Q1221208.
:Convenio	owl:equivalentClass itsmo:Agreement.
:Coorientador	mrel:ths, sio:SIO_000878, mrel:Thesisadvisor.
:CursoPosGraduacao	wdo:Q600134, schema:Course, wdo:Q2607211.
:Defesa	schema:EducationEvent, sro:Evaluation, sro:Discussion, sro:Argument, sro:Claim.
:Discente	sio:SIO_000405, seo:Student.
:DiscenteEgresso	sio:SIO_000399.
:Dissertação	bibframe:Dissertation, fabio: MastersThesis, sio:SIO_000167.
:Documento	foaf:Document, dbpedia:Documento.
:DocumentoAprovacao	schema:EducationalOccupationalCredential, fabio:Diary.
:DocumentoDigital	fabio:DigitalItem.
:DocumentoMonografico	wdo:Q193495.
:Doutorado	wdo:Q849697.
:Especialidade	wdo:Q1047113
:Examinador	sur:Participant, cwmo:Evaluator, wdo:Q95674799.
:ExaminadorExterno	sur:Participant, cwmo:Evaluator.

:ExaminadorInterno	sur:Participant, cwmo:Evaluator.
:InstituicaoEnsinoSuperior	do:Q2385804, wdo:Q5341295, wdo:Q1211427, schema:EducationalOrganization, cwrc:EducationalOrganization, sio:SIO_000290, dbo:EducationalInstitution, pext:EducationalOrganization, pext:University.
:Local	bibframe:Place, schema:Place, dc:Location.
:LocalFisico	bibframe:Place, schema:Place, dc:Location.
:LocalVirtual	wdo:Q107307154.
:Membro	foaf:Group, scoro:EducationalRole, vivo:MemberRole.
:Mestrado	wdo:Q4218455, wdo:Q18189, wdo:Q189533, wdo:Q26269989, wdo:Q183816.
:NaoBolsista	wdo:Q24716636.
:NatoDigital	wdo:Q14083178.
:OrganizacaoEnsino	pext:Educational Organization, cwrc:EducationalOrganization, km4c:Higher_education.
:Orientador	mrel:ths, sio:SIO_000878, mrel:Thesisadvisor, wdo:Q104472852.
:PalavraChave	sio:SIO_000147
:Patrocinador	frapo:FinancialEntity, frapo:FundingProgramme.
:Pessoa	schema:Person.
:PessoaInstitucionalizada	scoro:EducationalRole, whois:academicCareer.
:Produto	schema:IndividualProduct.

:ProgramaPosGraduacao	fabio:AcademicProceedings, schema:EducationalOccupationalProgram, teach:StudyProgram, xapi:Grouping.
:RedeEnsino	sio:SIO_000620.
:Tese	fabio:DoctoralThesis, fabio:Thesis, fabio:Proposition.
:TrabalhoConclusaoCurso	schema:Thesis, fabio:ScholarlyWork, frbrer:C1004, dbpedia:Trabalho_de_conclusão_de_curso.

Fonte: Elaborado pelo próprio autor.

Quadro 7 - Equivalência de propriedade de dados adotada no modelo de *design* com DatatypeProperty

Classe do modelo de <i>design</i>	Data property do modelo de <i>design</i>	Propriedades equivalentes
:Cadastro	:identificadorRegistroCadastro	frapo:identifier.
	:dataHoraRegistroCadastro	fabio:hasDateCollected, bibframe:date, dc:date, dc:dateSubmitted.
:CursoPosGraduacao	:nomeCursoPosGraduacao	dc:title, teach:courseTitle.
:Defesa	:dataHoraAvaliacaoDefesa	fabio:hasDecisionDate, dcat:record, ptop:registeredIn.
	:codigoDefesa	dc:identifier.
:Digitalizado	dataDigitalizacaoDigitalizado	fabio:hasAccessDate, bibframe:date, dc:date, dcndl:dateDigitized.
	modoOcrImagemDigitalizado	bibframe:media.
:Discente	:matriculaDiscente	dc:identifier.
	:telefoneDiscente	foaf:phone.
:Defesa	:dataHoraAvaliacaoDefesa	fabio:hasDecisionDate,

		dcat:record, ptop:registeredIn.
	:codigoDefesa	dc:identifier.
:Documento	identificacaoUnicaDocumento	dc:identifier, fabio:identifier.
:DocumentoDigital	:nomePdfDocumentoDigital	foaf:name.
	:tamanhoKilobitesDocumentoDigital	bibframe:FileSize.
	:descricaoAnexosDocumentoDigital	bibframe:supplementaryContent.
:DocumentoMonografico	:numeroPaginasDM	bibframe:count, fabio:hasPageCount.
:InstituicaoEnsinoSuperior	:siglaIES	vivo:abbreviation
	:nomeIES	bibframe:grantingInstitution wdo:PP1813
	:uriIES	dqm:hasURI foaf:homepage, xapi:hasHomePage.
:LocalFisico	:cidadeLocalFisico	wdo:Q82799, #(name) schema:name, gn:name.
	:estadoLocalFisico	wdo:Q82799, #(name) schema:name, gn:name.
	:paisLocalFisico	wdo:Q82799, #(name) schema:name, gn:name.
:PalavraChave	:termoPortuguesPalavraChave	fabio:keyword, schema:keywords.
	:termoInglesPalavraChave	fabio:keyword, schema:keywords.
:Pessoa	:nomePessoa	foaf:name, dc:name.
	:funcaoPessoa	bibframe:role.
	:emailInstitucionalPessoa	foaf:mbox, pext:hasEMail.

	:emailAlternativoPessoa	foaf:mbox, pext:hasEMail, km4c:alternativeEmail.
:PessoaInstitucionalizada	:orcidPessoaInstitucionalizada	scoro:hasORCID.
	:lattesPessoaInstitucionalizada	frapo:hasCurriculumVitae.
:ProgramaPosGraduacao	:nomeProgramaPosGraduacao	xapi:hasName.
:TrabalhoConclusaoCurso	:tituloTCC	bibframe:title, fabio:title.
	:resumoTCC	dc:abstract, bibframe:summary, fabio:abstract.
	:idiomaTCC	bibframe:language.
	:semestreLetivoTCC	ov:semester.
	:uriDisseminacaoTCC	dc:publisher, meb:uri.

Fonte: Elaborado pelo próprio autor.

Quadro 8 - Equivalência de propriedade de objetos adotada no modelo de *design* com ObjectProperty

Object property do modelo de <i>design</i>	Propriedades equivalentes
:armazenadoNoLocalFisico	dcat:record, ptop:registeredIn.
:associadaIES	rico:teachingRelationHasSource, bibframe:grantingInstitution.
:avaliaDefesa	schema: assesses, pproc:criterionEvaluationMode.
:compostaPorArea	dc:title.
:compostaPorEspecialidade	dc:title.
:compostaPorMembro	ov:composedBy, essglobal:isComposedBy, xapi:hasMember, skos:member, bibframe:agentOf.
:compostaPorSubarea	dc:title.
:compostoPorCursoPosGraduacao	ov:composedBy, essglobal:isComposedBy, schema:hasCourseInstance,

	schema:hasCourse, schema:course, ov:deliveredBy, ov:offersCourse.
:compostoPorOrganizacaoEnsino	ov:composedBy, essglobal:isComposedBy.
:compostoPorPalavraChave	schema:keywords, fabio:keyword.
:digitalizaDocumento	iol:isEncodedBy.
:disponibilizaBolsaEstudos	frapo:contributions-in-kind, scoro:controls-project-finances, voag:approves.
:elaboraTCC	scoro:research-student.
:estabeleceCondicaoAprovacao	schema:coursePrerequisites, schema:programPrerequisites, voag:hasApprovalProcess.
:geraDocumentoAprovacao	schema:occupationalCredentialAwarded, bibiframe:degree, schema:coursePrerequisites, schema:programPrerequisites, voag:hasApprovalProcess.
:geridoOrganizacaoEnsino	base-gleif:isManagedBy, poder:hasAcademicOrganization, poder:organizationParticipant.
:guardadoLocalVirtual	dcat:record, ptop:registeredIn.
:localizadaEmLocalFisico	wdo:Q82799, schema:name, gn:name.
:ocorreEmUmLocal	scoro:research-student.
:ofereceCursoPosGreaduacao	schema:hasCourseInstance, schema:hasCourse, schema:course, ov:deliveredBy, ov:offersCourse.
:pertenceOrganizacaoEnsino	poder:hasAcademicOrganization, poder:organizationParticipant.
:porBolsaEstudos	frapo:contributions-in-kind, scoro:controls-project-finances, voag:approves.
:realizaDefesa	fabio:hasDecisionDate.
:recebeContrato	frapo:supports, vivo:teachingOverview.

:registraCadastro	dcat:record, ptop:registeredIn.
:temConvenio	itsmo:hasAgreement.
:temTipoPDF	bibframe:digitalCharacteristic.
:vinculadoAreaConhecimentoCnpq	dc:title.

Fonte: Elaborado pelo próprio autor.

A aplicação das técnicas de equivalências de classes e propriedades de dados e objetos contidas nos Quadros 5, 6, 7 e 8 foram importantes para alcançar a interoperabilidade (subseção 3.3.5) e a integração com bases externas de dissertações conforme apresentado por Revez e Silva (2021). Dessa maneira, é possível colaborar com uma RI interoperável objetivando o reuso contínuo das pesquisas científicas que adotam as técnicas propostas pela Web semântica. Como a quantidade de dissertações e teses está em expansão, proporcioná-las com técnicas de dados ligados e por intermédio de bibliotecas especializadas externas favorece o reaproveitamento das pesquisas como apresentado por Konstantinou e Spanos (2015) na subseção 3.3.2.

A realização do modelo de *design* para a ontologia operacional com interoperabilidade (O4, Quadro 4) corresponde à elaboração da terceira fase do processo de desenvolvimento SABiO (subseção 4.2.1) proposta por Falbo (2014). Esta etapa ajudou a resolver um dos grandes problemas enfrentados na Recuperação da Informação (subseção 3.1.6) que está na possibilidade de organizar dados e informações para disseminá-los e recuperá-los dados na Web semântica (subseção 3.3) de maneira interoperável. Dessa maneira, o modelo de *design* e a ontologia de referência de domínio cumprem o objetivo específico que busca resolver parte do problema na RI Interoperável, criando-se um modelo de *design* para a ontologia operacional de domínio com base na ontologia de referência e com elementos interoperáveis das camadas estrutural e semântica (subseção 3.3.5). Dessa forma, o terceiro objetivo específico foi cumprido.

6.3 ONTOLOGIA OPERACIONAL

A implementação da ontologia operacional foi realizada seguindo as diretrizes da fase 4 da metodologia SABiO que é conhecida como Implementação (subseção 5.4.3). Levando em conta o modelo de *Design* da ontologia operacional de domínio

(O4, Quadro 4) elaborado no Visual-Paradigm (subseção 4.2.3.1), realizou-se a exportação automática da ontologia operacional¹⁵⁰ em um arquivo no formato *Turtle* (extensão TTL) usando o plug-in gUFO, limpou-se esse arquivo com Notepad++ gerando-se uma nova versão da ontologia operacional¹⁵¹ (O5, Quadro 4). Foi realizada a cópia das anotações da fase de *Design* definidas no modelo de design da ontologia operacional¹⁵² (O4, Quadro 4) de domínio contida no Visual-Paradigm e a depuração com suporte do VocBench (discutido na subseção 5.4.3 e apresentada na subseção 6.2) e que permitiu a codificação de arquivo no formato *Turtle* por meio de elementos gráficos. Em seguida, inseriu-se alguns indivíduos na ontologia operacional¹⁵³ para poder analisar os resultados.

A Figura 20 mostra uma tela do VocBench com a ontologia operacional de domínio (O5, Quadro 4) carregada e disponível para edição. No lado esquerdo está disponível para edição as entidades :Local, :LocalFisico, :LocalVirtual, :OrganizacaoEnsino, dentre outras. Como a entidade :Local está selecionada, exibem-se elementos dela para edição como rdf:type, owl:equivalentClass, rdfs:subclass, dentre outros. Isso possibilitou a implementação codificada em *Turtle*, gerando-se ao final da atualização das entidades um novo arquivo da ontologia operacional de domínio¹⁵⁴ (O5, Quadro 4), contendo a representação das classes, propriedades e relacionamentos codificados em *Turtle*, e dispondo dos dados da própria base e das ontologias externas.

¹⁵⁰ Ontologia operacional implementada e gerada com a exportação em gUFO. Disponível em: https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/2.3.1ontologiaOperacional_implementada_geradaPorExport_gUFO..ttl

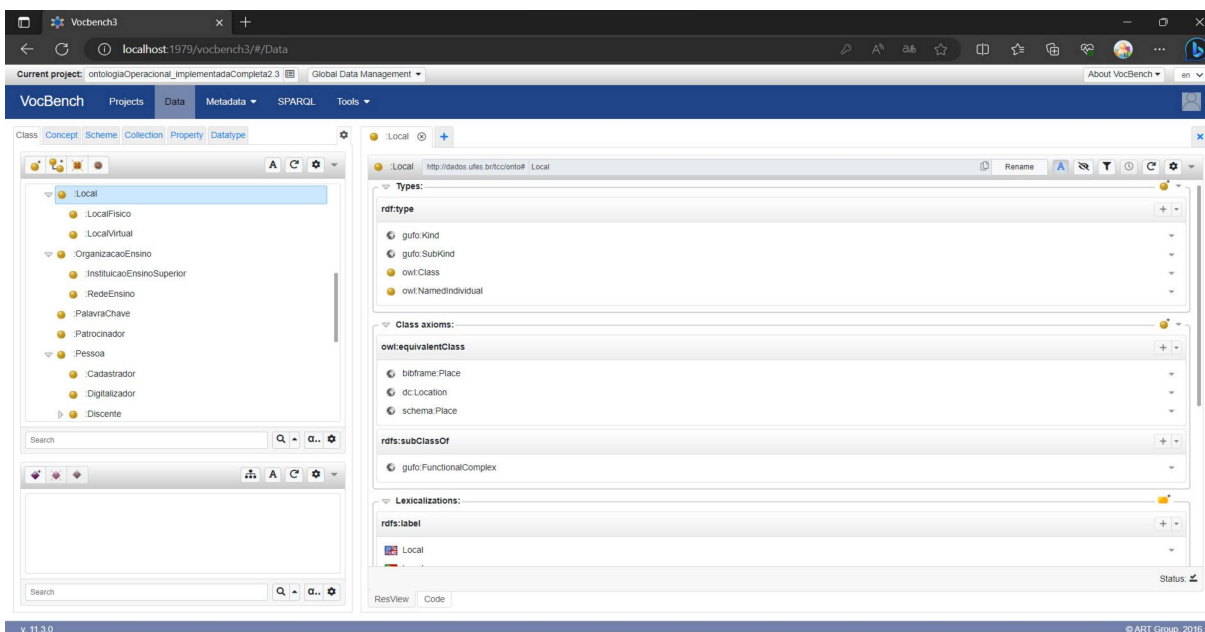
¹⁵¹ Ontologia operacional implementada com aplicação da limpeza manual. Disponível em: https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/2.3.2ontologiaOperacional_implementada_aposLimpezaManual.ttl

¹⁵² Ontologia operacional implementada com as anotações da fase de *Design* contidas no Visual-Paradigm. Disponível em: https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/2.3.3ontologiaOperacional_implementada_designAnnotationsCopiadosDoVP.ttl

¹⁵³ Ontologia operacional com alguns indivíduos. Disponível em: https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/2.3.4ontologia%20operacional_individuos.ttl

¹⁵⁴ export.ttl é o arquivo de implementação da ontologia operacional. Disponível em: https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/export.ttl

Figura 20 - Arquivo em *Turtle* no VocBench ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor, tela capturada do OpenRefine.

A Figura 21 apresenta um recorte do resultado da ontologia operacional de domínio (O5, Quadro 4) implementada, codificada e serializada em *Turtle*. A entidade :Local está selecionada na cor cinza, abarcando elementos como rdfs:subClassOf, rdfs:label e rdfs:comment, correspondendo respectivamente ao nome (rdfs:label), um comentário a seu respeito (rdfs:comment) e que está disponível no arquivo *Turtle* “export.ttl”.

Figura 21 - Arquivo em *Turtle* no Notepad++ ([link para acesso em tamanho maior](#))

The screenshot shows a Notepad++ window displaying the Turtle export file content. The text is as follows:

```

3597 :paisLocalFisico "Brasil"@pt-BR, "Brazil"@en, "http://www.wikidata.org/entity/Q155"^^xsd:anyURI,
3598 "http://dbpedia.org/resource/Brazil"^^xsd:anyURI, "http://www.geonames.org/3469034"^^xsd:anyURI .
3599
3600 :cidadeLocalFisico a owl:DatatypeProperty;
3601 rdfs:label "cidadeLocalFisico"@en;
3602 rdfs:range xsd:string;
3603 rdfs:domain :LocalFisico;
3604 rdfs:subPropertyOf gufo:hasQualityValue .
3605
3606 :estadoLocalFisico a owl:DatatypeProperty;
3607 rdfs:label "estadoLocalFisico"@en;
3608 rdfs:range xsd:string;
3609 rdfs:domain :LocalFisico;
3610 rdfs:subPropertyOf gufo:hasQualityValue .
3611
3612 :paisLocalFisico a owl:DatatypeProperty;
3613 rdfs:label "paisLocalFisico"@en;
3614 rdfs:range xsd:string;
3615 rdfs:domain :LocalFisico;
3616 rdfs:subPropertyOf gufo:hasQualityValue .
3617
3618 :Local a owl:Class, owl:NamedIndividual, gufo:Kind, gufo:SubKind;
3619 rdfs:subClassOf gufo:FunctionalComplex;
3620 rdfs:label "Local"@pt, "Local"@en;
3621 rdfs:comment "Trata-se do lugar fisico ou virtual de registro do Trabalho de Conclusão de Curso. Quanto ao local fis
3622 ""# Design annotations
3623
3624 :Local
  
```

The status bar at the bottom shows: User Defined language file - Turtle, length: 593.307, lines: 6.040, Ln: 3.625, Col: 25, Sel: 631 | 8, Windows (CR LF), UTF-8, INS.

Fonte: Elaborado pelo próprio autor.

A implementação da ontologia operacional correspondeu a codificar o modelo da ontologia operacional de domínio na linguagem *Turtle* (O5, Quadro 4) discutido nas subseções 5.4.3 e 4.2.3.2, com o auxílio do *software* VocBench (subseção 4.2.3.1) com base no modelo de *design* da ontologia operacional de domínio (O4, Quadro 4) apresentado na subseção 6.2. O cumprimento do quarto objetivo específico que esteve em realizar a implementação da ontologia operacional de domínio (O5, Quadro 4) por intermédio do modelo de design (O4, Quadro 4). Estes mecanismos proporcionam o resultado de representar os dados ligados (subseção 3.3.4) em RDF (subseção 4.2.3.2) para a Web Semântica e que proporcionam a interoperabilidade das camadas sintática e estrutural (subseção 3.3.5).

6.4 DADOS PRÉ-PROCESSADOS E RECONCILIADOS

Inicialmente limpou-se e depurou-se (O6, Quadro 4) os dados brutos do acervo de dissertações apresentados na seção 5.4.4 que resultaram em três arquivos do acervo experimental (bases de dados brutos das dissertações¹⁵⁵, pessoas¹⁵⁶ e Instituições¹⁵⁷). Com o apoio do *software* OpenRefine, removeu-se os espaços em branco, excluiu-se registros duplicados e foi realizado o agrupamento de representações da mesma realidade, como por exemplo, as palavras-chave definidas para mais de um TCC como memória, Memória e MEMÓRIA foram convertidas para apenas uma notação (Memória).

Em seguida foi realizada a reconciliação dos dados (O7, Quadro 4) com bases externas de maneira semi-automático, criando colunas adicionais contendo conexões com a propriedade owl:sameAs nas bases externas Wikidata¹⁵⁸, DBpedia¹⁵⁹ e

¹⁵⁵ Dissertações em dados brutos. Disponível em:

https://github.com/rafaeldenersons/0_dadosBrutos/blob/main/1Dados%20de%20egressos%20do%20PPGCI-UFES%20-%20TrabalhoConclusaoCurso.csv

¹⁵⁶ Pessoas em dados brutos. Disponível em:

https://github.com/rafaeldenersons/0_dadosBrutos/blob/main/2Dados%20de%20egressos%20do%20PPGCI-UFES%20-%20Pessoa.csv

¹⁵⁷ Instituições em dados brutos. Disponível em:

https://github.com/rafaeldenersons/0_dadosBrutos/blob/main/3Dados-de-egressos-do-PPGCI-UFES-Instituicao.csv

¹⁵⁸ Wikidata refere-se a base de dados externa de reconciliação dos dados. Disponível em:

https://www.wikidata.org/wiki/Wikidata:Main_Page

¹⁵⁹ Dbpedia diz respeito à base de dados externa de reconciliação dos dados. Disponível em:

<https://www.wikipedia.org/>

GeoNames¹⁶⁰. A Figura 22 apresenta um pequeno recorte da ação de reconciliação de dados com bases externas realizada com o OpenRefine. Ela apresenta as colunas título, resumo, dentre outras da base de dados de dissertação. As colunas cursoReconciliado, cursoSameAsWikidata e cursoSameAsDbPedia (dissertacoesCSV¹⁶¹, pessoasCSV¹⁶², instituicaoCSV¹⁶³).

Figura 22 - Reconciliação de dados com bases externas ([link para acesso em tamanho maior](#))

curso	título	resumo	abstract	cursoReconciliado	cursoSameAsWikidata	cursoSameAsDbPedia
Ciência da Informação	O papel da biblioteca no processo de obtenção da remição de pena pelo trabalho, estudo e pela leitura no Complexo Penitenciário de Xuri no Estado do Espírito Santo (ES)	O Estado do Espírito Santo possui trinta e seis estabelecimentos prisionais, destes, trinta possuem bibliotecas. Diante desta informação, este estudo investiga o papel da biblioteca no Complexo Penitenciário de Xuri nos processos de acesso à informação e da remição de pena pelo trabalho, estudo e leitura. Em específico, relata a conformação da pena e da prisão no cenário da execução penal no Brasil. Identifica o funcionamento e as práticas de remição de pena desenvolvidas pelas unidades prisionais a partir de suas vinculações com as bibliotecas do Complexo Penitenciário de Xuri em Vila Velha no Espírito Santo; caracteriza os usos das bibliotecas prisionais como espaço para práticas socioeducativas e de acesso à informação para os internos do Complexo Penitenciário de Xuri	The State of Espírito Santo has thirty-six prisons, of which thirty have libraries. Given this information, this study investigates the role of the library in the Xuri Penitentiary Complex in the processes of access to information and the remission of punishment through work, study and reading. Specifically, it reports the conformation of punishment and imprisonment in the context of criminal execution in Brazil. Identifies the functioning and practices of penalty redemption developed by the prison units from their links with the libraries of the Xuri Penitentiary Complex in Vila Velha in Espírito Santo, characterizes the uses of prison libraries as a space for socio-educational practices and access to information for inmates at the Xuri	information science	http://www.wikidata.org/entity/Q16387	http://dbpedia.org/resource/CI&I

Fonte: Elaborado pelo próprio autor, tela capturada do OpenRefine.

A realização da limpeza e depuração dos dados (O6, Quadro 4) no OpenRefine (subseção subseções 4.2.4) resultou nos dados pré-processados e que passaram por correções de conteúdo (O7, Quadro 4), ajuste de e formatações (subseções 5.4.4). Em seguida, a ligação dos dados pré-processados com outras bases e bibliotecas externas de maneira manual correspondeu a interoperabilidade da camada semântica (subseção 3.3.5). O cumprimento desta etapa permitiu a preparação dos dados para a realização do mapeamento para a base de dados ligados.

¹⁶⁰ GeoNames trata-se da base de dados externa de reconciliação de dados. Disponível em: <http://www.geonames.org/>

¹⁶¹ https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso.csv

¹⁶² https://github.com/rafaeldenersons/5_mapeamento/blob/main/2pessoa.csv

¹⁶³ https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.2.csv

6.5 MAPEAMENTO DOS DADOS PARA RDF

Esta subseção apresenta os esquemas de mapeamento (O8, Quadro 4) que recebem como entrada os dados da base experimental, já devidamente pre-processados e reconciliados (O7, Quadro 4), e a ontologia operacional de domínio escrita em *Turtle* (O5, Quadro 4) e fornece como saída os dados ligados em RDF (O9, Quadro 4). Os esquemas de mapeamento (O8, Quadro 4) foram construídos no Ontotext Refine.

A Figura 23 mostra a interface gráfica do Ontotext Refine para essa ação de mapeamento. A interface deste *software* é intuitiva e funciona pelo estabelecimento dos tipos de triplas que irão formar a base de dados ligados. Apesar dos esquemas de mapeamento poderem ser armazenados no formato interno do *software*, eles foram exportados para o formato JSON (O9, Quadro 4), por segurança e para que fossem permitidos pequenos ajustes, em caso de necessidade, diretamente em um editor de textos simples e de forma independente do Ontotexto Refine.

Os arquivos no formato JSON contendo os esquemas de mapeamento foram: trabalhoConclusaoCurso1¹⁶⁴, trabalhoConclusaoCurso2¹⁶⁵, trabalhoConclusaoCurso3¹⁶⁶, pessoa¹⁶⁷ e instituicao1¹⁶⁸, instituicao2¹⁶⁹, instituicao3¹⁷⁰, instituicao4¹⁷¹.

¹⁶⁴ Disponível em:

https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso1.1.json

¹⁶⁵ Disponível em:

https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso1.2.json

¹⁶⁶ Disponível em:

https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso1.3.json

¹⁶⁷ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/2pessoa.1.1.json

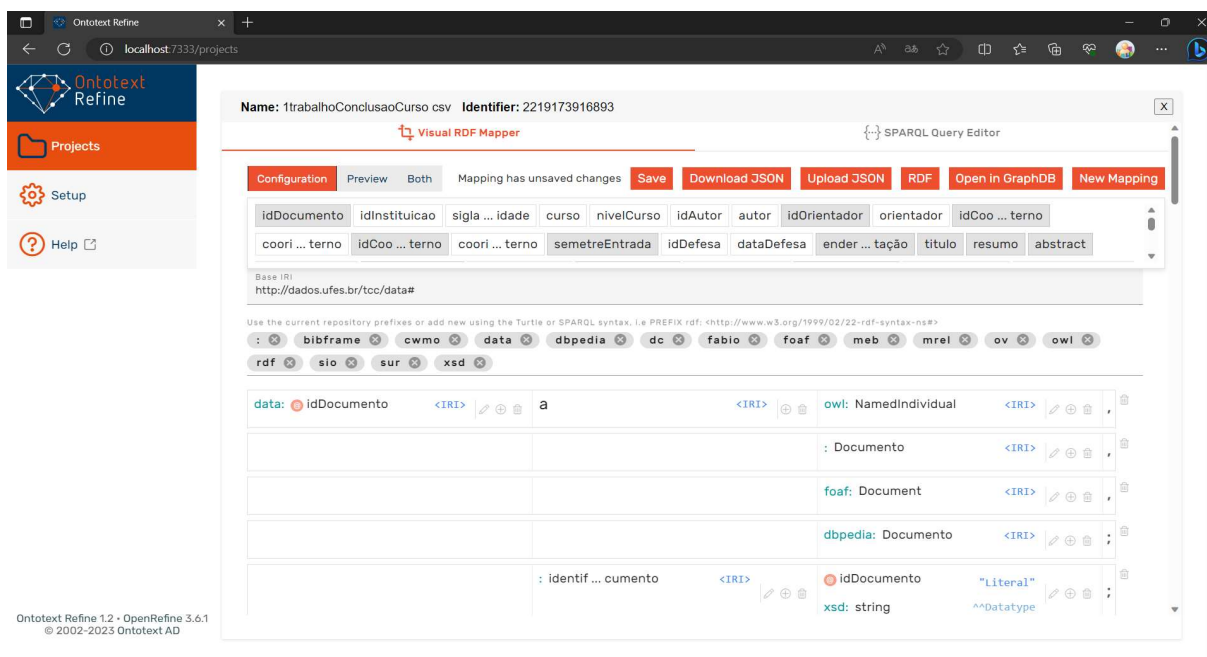
¹⁶⁸ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.1.1.json

¹⁶⁹ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.2.json

¹⁷⁰ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.3.json

¹⁷¹ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.4.json

Figura 23 - Esquema de mapeamento para dados ligados ([link para acesso em tamanho maior](#))

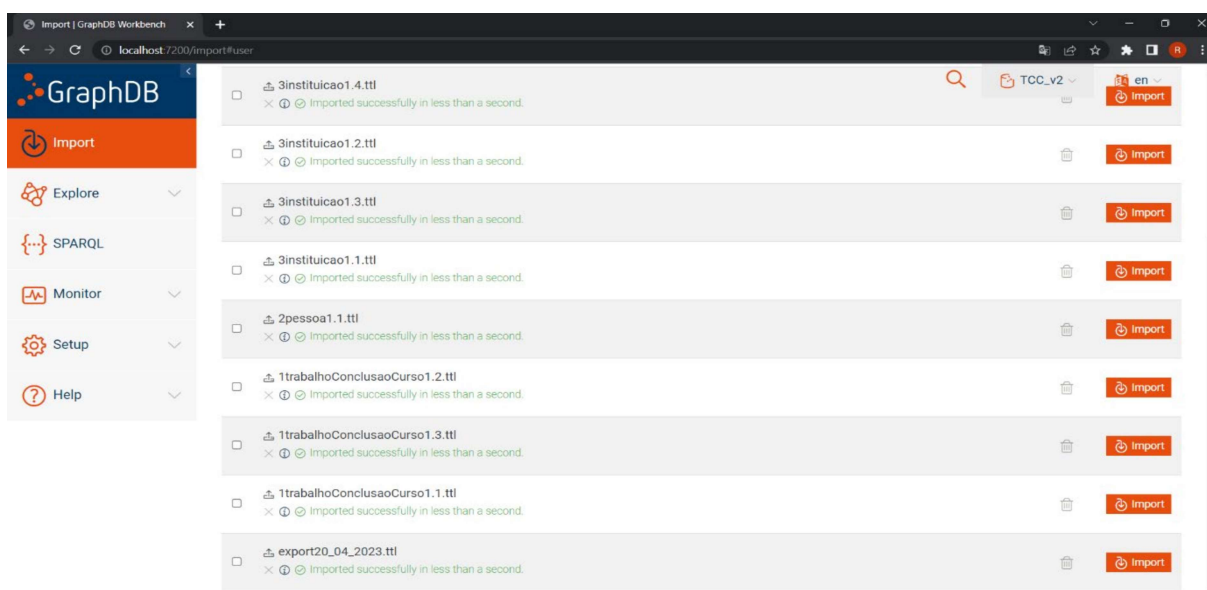


Fonte: Elaborado pelo próprio autor.

6.6 BASE DE DADOS LIGADOS RDF

A partir dos esquemas de mapeamento construídos (O8, Quadro 4), apresentados na subseção 6.5, foram gerados os dados ligados RDF, também utilizando-se do *software* Ontotext Refine. A Figura 24 apresenta essas ações de geração das bases de dados ligados (O9, Quadro 4).

Figura 24 - Repositório de dados ligados em RDF ([link para acesso em tamanho maior](#))



Fonte: Elaborado pelo próprio autor.

Os arquivos gerados em formato *Turtle* foram: trabalhoConclusaoCursoTtl1¹⁷², trabalhoConclusaoCursoTtl2¹⁷³, trabalhoConclusaoCursoTtl3¹⁷⁴, pessoaTtl¹⁷⁵, instituicaoTtl1¹⁷⁶, instituicaoTtl2¹⁷⁷, instituicaoTtl3¹⁷⁸, instituicao4Ttl¹⁷⁹.

Este resultado refere-se ao quinto objetivo específico que propôs gerar uma base de dados ligados na Web semântica (O10, Quadro 4) a partir de um mapeamento da ontologia operacional de domínio implementada (A9, Quadro 4) e de uma base de dados experimental, e previamente preparada (O9, Quadro 4), de dissertações com vínculos interoperáveis para elementos externos.

Em seguida todos os dados ligados RDF foram inseridos em um repositório no GraphDB (O10, Quadro 4) junto ao arquivo em *Turtle* completo¹⁸⁰ (com todos os dados juntos).

6.7 CONSULTAS E INFORMAÇÃO RECUPERADA

Utilizando-se do repositório de dados ligados criado no GraphDB, foram realizadas consultas SPARQL, no próprio GraphDB (O10, Quadro 4), e análises provenientes das informações recuperadas, como prova de conceito do trabalho realizado (O11, Quadro 4).

A primeira consulta foi elaborada com o objetivo de encontrar as três palavras-chave que mais apareceram nos TCCs da base experimental.

A Figura 25 apresenta, na própria interface do GraphDB, a consulta escrita em SPARQL e as informações recuperadas.

¹⁷² Disponível em:

https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso1.1.ttl

¹⁷³ Disponível em:

https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso1.2.ttl

¹⁷⁴ Disponível em:

https://github.com/rafaeldenersons/5_mapeamento/blob/main/1trabalhoConclusaoCurso1.3.ttl

¹⁷⁵ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/2pessoa1.1.ttl

¹⁷⁶ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.1.ttl

¹⁷⁷ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.2.ttl

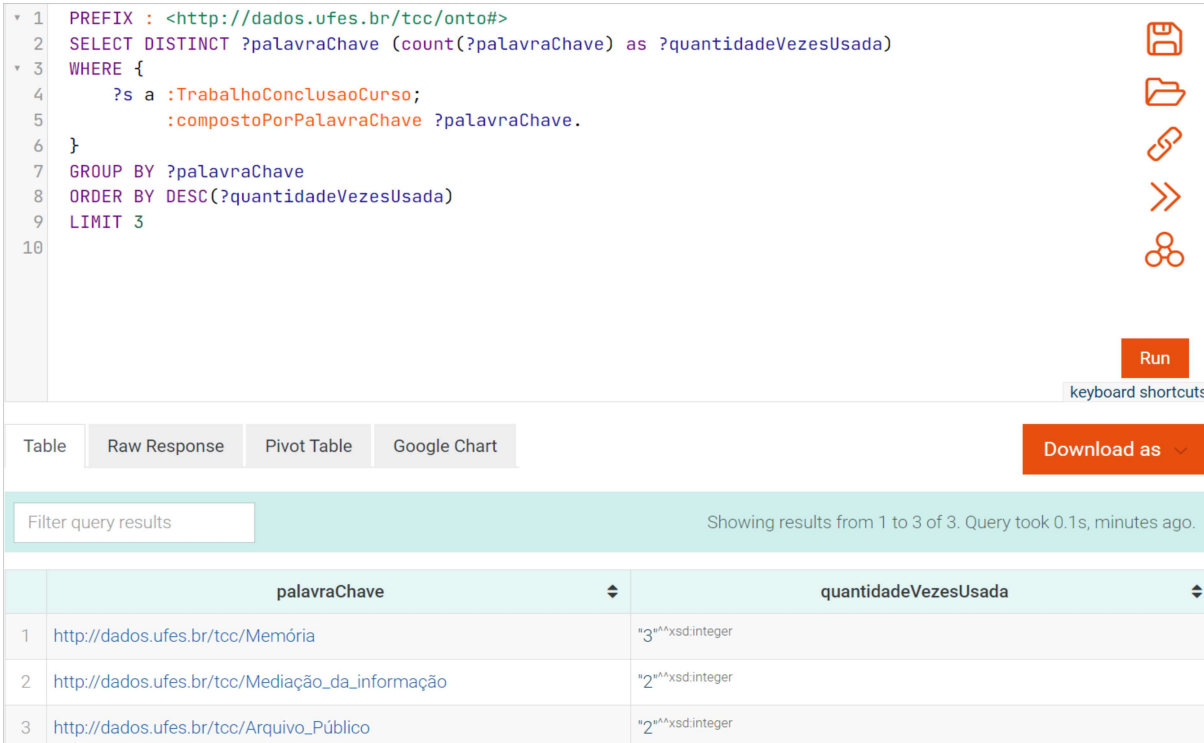
¹⁷⁸ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.3.ttl

¹⁷⁹ Disponível em: https://github.com/rafaeldenersons/5_mapeamento/blob/main/3instituicao1.4.ttl

¹⁸⁰ Disponível em:

https://github.com/rafaeldenersons/4_implementacaoOntologiaOperacional/blob/main/export.ttl .

Figura 25 - Consulta SPARQL para as três palavras-chave de maior ocorrência
([link para acesso em tamanho maior](#))



```

1 PREFIX : <http://dados.ufes.br/tcc/onto#>
2 SELECT DISTINCT ?palavraChave (count(?palavraChave) as ?quantidadeVezezUsada)
3 WHERE {
4   ?s a :TrabalhoConclusaoCurso;
5       :compostoPorPalavraChave ?palavraChave.
6 }
7 GROUP BY ?palavraChave
8 ORDER BY DESC(?quantidadeVezezUsada)
9 LIMIT 3
10

```

Showing results from 1 to 3 of 3. Query took 0.1s, minutes ago.

	palavraChave	quantidadeVezezUsada
1	http://dados.ufes.br/tcc/Memória	"3^^xsd:integer
2	http://dados.ufes.br/tcc/Mediação da Informação	"2^^xsd:integer
3	http://dados.ufes.br/tcc/Arquivo Público	"2^^xsd:integer

Fonte: Elaborado pelo próprio autor.

A Figura 25 apresenta no canto superior esquerdo a consulta escrita em SPARQL. Na parte superior da figura, a consulta em SPARQL seleciona (O11, Quadro 4) as palavras-chave que mais aparecem no repositório de dados de TCC (O10, Quadro 4), de maneira ordenada com base no maior número de ocorrência. Na parte inferior da figura é exibido o resultado em três linhas. Na primeira linha a palavra Memória (coluna palavraChave) aparece três (3) vezes, seguida das palavras-chave Mediação da Informação e Arquivo Público, ambas com duas (2) ocorrências.

Esse resultado representou a quinta fase da metodologia SABI^o correspondente (subseção 4.2.1.1) que é Teste (O11, Quadro 4). Observa-se, assim, o teste a validação da ontologia operacional (subseção da 6.1 a 6.6) para o fragmento correspondente a entidade Palavra-Chave e as suas propriedades por meio de prova de conceito.

6.8 DISCUSSÕES GERAIS

Articulou-se o conceito de ontologia no âmbito da Ciência da Informação (CI), pois ela tem o propósito de entender um domínio (ALMEIDA, 2013), e aprofundou-se na investigação desses estudos, pois estudos apontam que uma boa ontologia de

domínio serve para alcançar interoperabilidade existente nos princípios norteadores FAIR (*Findability, Accessibility, Interoperability e Reusability*) (GUIZZARDI, 2020; WILKINSON *et al.*, 2016).

Como uma ontologia de domínio (O4 e O11, Quadro 4) é decomposta em ontologia de referência de domínio (modelo conceitual) e ontologia operacional de domínio (projeto de detalhamentos computacionais legíveis por máquinas) (FALBO, 2014), realizou-se a primeira para representar os TCCs de pós-graduação adotando as duas primeiras etapas da metodologia SABiO (FALBO, 2014). Esses mecanismos permitiram representar informação por meio do resultado de um *knowledge graph*, que resultou em quatro diagramas de ontologia de referência de domínio (Defesa, Colaborador, Bolsa e Área de conhecimento) embasados na ontologia de fundamentação UFO e produzidos na linguagem OntoUML com suporte do Visual-Paradigm e gUFO. O benefício de usar UFO para elaborar a ontologia de referência domínio e o modelo de *design*, esteve em representar uma modelagem ontológica mais natural e simples de entender que as demais ontologias de fundamentação pesquisadas e que serviu para alinhar conceitos nas entrevistas disponíveis nos apêndices A e B. Além de UFO está sendo usada para criar ontologias em diversas áreas do conhecimento (ANDRADE; NARDI; RUY, 2023) e da educação (SILVA; MUTZ; RUY, 2022a; SILVA; MUTZ; RUY, 2022b; SILVEIRA; ARAGÓN; CURY; MENEZES, 2021; SILVEIRA, CURY E MENEZES, 2019), ela também possibilitou uma integração considerável para modelar ontologia em OntoUML e permitiu a adoção do plugin gUFO do Visual-Paradigm que é bem difundido e tem excelente usabilidade para modelar ontologias. Esses resultados foram importantes, pois permitiram alcançar uma descrição clara e precisa das entidades no âmbito educacional da defesa, do colaborador, da bolsa de estudos e da área de conhecimento de trabalhos acadêmicos, e isso possibilita que outras pesquisas possam (re)aproveitar os diagramas de ontologia de referência de domínio elaborados para apoiar na comunicação, na aprendizagem, e também na resolução de outros problemas que são defendidos por Falbo (2014).

O modelo de *design* foi projetado aplicando a terceira etapa da metodologia SABiO a partir da ontologia de referência de domínio (O2, Quadro 4), e assim conseguiu-se alcançar como resultado a ontologia de referência (FALBO, 2014) no domínio de TCCs (O4, Quadro 4). Isso correspondeu a elaboração de um modelo

conceitual que serviu de apoio a comunicação com especialistas, auxiliou no entendimento do domínio e ajudou a solucionar dúvidas durante o processo de negociação de significado entre conceitos e relações (FALBO, 2014; ANDRADE; NARDI; RUY, 2023). Além disso, proporcionou as equivalências de classe, de propriedades e de relações.

Adotar o uso dos vocabulários controlados disponíveis nos portais especializados permite a apresentação de diferentes tipos de relações entre os termos que eles integram, conforme defendido por Harpring *et al.* (2006) e Zeng (2005). Esses mecanismos foram implementados com a perspectiva de reaproveitar os vocabulários controlados e ontologias já existentes para o domínio que se deseja representar (PENTEADO, 2020) e estão em consonância com as Boas Práticas na Publicação de Dados na Web (*Data on the Web Best Practices*) propostas por Lóscio, Burle e Calegari (2017).

As soluções provenientes do modelo de *design* alcançado promovem benefícios inversamente proporcionais aos problemas encontrados na exploração das consultas. Iniakpokeikiye e Fawei (2022) apresentaram as limitações provocadas pela ambiguidade relacionada à recuperação da informação como a dificuldade em pesquisar por pessoas com nomes iguais, comuns, ou indivíduos famosos com nomes idênticos, e essas ambiguidades são eliminadas com o uso de ontologia no contexto da Web semântica (LAUFER, 2015) adotando procedimentos interoperáveis (ZENG, 2005). Esses resultados só foram alcançados com a adoção de ontologia no contexto da Web semântica associadas à implementação da multiplicidade de equivalência de classes, propriedades e relações.

A implementação da ontologia operacional (O5, Quadro 4) foi derivada de toda preparação proposta no modelo de *design* e com a aplicação da fase 4 de SABiO (O4, Quadro 4) segundo as recomendações de Falbo (2014). Ela resultou em arquivos no formato *Turtle* (extensão TTL) advindos do uso do plug-in gUFO e com o apoio dos *softwares* VocBench e Notepad++. Esses arquivos foram implementados buscando assegurar as propriedades computacionais esperadas explicado por Guizzardi (2006), pois esta fase correspondeu a especificação sistemática para realizar as implementações possíveis para o domínio de dissertações do PPGCI obedecendo as diretrizes de Falbo (2014). Esses resultados permitiram representar os dados ligados em RDF de maneira codificada e não visual para a Web semântica com os elementos

de interoperabilidade das camadas sintática, estrutural e semântica conforme fundamentados por Zeng (2019).

Essas soluções de dados ligados em RDF correspondem aos princípios de *design* que tem a finalidade de compartilhar dados interligados e legíveis computacionalmente na Web segundo argumentado por Ontotext Refine (2022), Conelegian (2016) e Bauer e Kaltenböck (2012). Servem também como mecanismos da Web semântica, proporcionando a extensão da própria Web conforme fundamentada por Berners-Lee, Hendler e Lassila (2002) ao permitir a interoperabilidade de elementos da ontologia operacional com bases externas.

A implementação de dados ligados em RDF corresponderam aos preceitos de Boas Práticas na Publicação de Dados Ligados, uma vez que eles refletem a modelagem de dados ligados para representar a tabulação de dados por representação baseada em grafo, correspondendo assim às exemplificações explanadas por Hyland, Ateazing e Villazón (2014).

A preparação e reconciliação (O6, Quadro 4) dos dados correspondeu ao resultado de obter os dados do acervo de dissertações do PPGCI/UFES depurados e reconciliados com o apoio do *software* OpenRefine, permitindo trabalhar com uma base precisa como defendida por Hooland *et al.* (2013). O procedimento de reconciliar os dados proporcionou a implementação de mecanismos da interoperabilidade da camada semântica que é disponibilizado por meio de funções na Internet para integrar os dados da base em questão com dados de bases externas (O7, Quadro 4). Ele foi atingido por intermédio de um método semi-automático do OpenRefine, e não só correspondeu a interoperabilidade da camada semântica, como também indicou as extensões das diretrizes propostas pela OpenAIRE que são fundamentais para produzir um repositório que atenda os princípios FAIR segundo a visão dos especialistas *Wilkinson et al.* (2016) e *Paganine et al.* (2020).

A implementação dos mecanismos de mapeamento para a base de dados ligados em RDF para a Web semântica (O10, Quadro 4) adotou a quarta fase da metodologia SABiO proposta por Falbo (2014). Esse resultado permitiu incorporar a base de dados de dissertações do PPGCI/UFES depurado (O6, Quadro 4), estruturado e reconciliado (O7, Quadro 4) ao repositório elaborado com apoio do GraphDB. Essa solução contou com o uso do Ontotext Refine para estabelecer o esquema de mapeamento e a implementação em JSON, e também com a elaboração

do repositório de dados ligados em RDF com a adoção do GraphDB. Essa etapa constitui nos dados ligados em RDF no formato N-triples (O10, Quadro 4) estabelecido a partir do esquema de mapeamento (O9, Quadro 4), e significa que trata-se de dados ligados com a adoção de Boas Práticas na Publicação de Dados na Web propostas por Lóscio, Burle e Calegari (2017).

O resultado da recuperação de informação foi alcançado aplicando-se a quinta fase da metodologia SABiO proposta por Falbo (2014), e isso correspondeu a realização de consultas SPARQL no repositório de dados ligados RDF (O10, Quadro 4) como prova de conceito, ao analisar o retorno dos dados ligados recuperados (O11, Quadro 4). Esse instrumento permitiu validar a ontologia operacional através da prática experimental, além de ser procedimento bem difundido para esse fim em pesquisas no âmbito dos Sistemas de Informação (SI) e da Ciência da Informação (CI) segundo as abordagens realizadas por Coneglian (2020) Neto, Borges e Roque (2021).

De todas as produções científicas consultadas e que estão expressas no Capítulo 2 (Estado da arte das boas práticas de publicação de dados na web no contexto de trabalhos de conclusão de curso de pós-graduação), aquelas que estiveram mais próximas do que foi desenvolvido nesta dissertação em termos de ontologia de referência de domínio (modelo conceitual) estiveram nas produções de Silva, Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b). Esses autores também utilizaram UFO, OntoUML e a metodologia SABiO para elaborar a ontologia de referência de domínio e alcançar a interoperabilidade semântica. A principal diferença com relação a presente pesquisa é que enquanto aqui a ontologia como modelo conceitual buscou entender o contexto de TCC de pós-graduação, as pesquisas dos autores mencionados esteve em entender o domínio da evasão escolar.

Ademais, foi observado as seguintes semelhanças e divergências que estão dispostas nos quadros 9 e 10.

Quadro 9 - Semelhança entre esta dissertação e a pesquisa científica de Silva, Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b)

Semelhança entre entidade	
A presente pesquisa	Silva, Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b)
<u>Área de Conhecimento</u>	<u>Área de Conhecimento</u>

<u>Pessoa</u>	<u>Pessoa</u>
<u>Curso de Pós-Graduação</u>	<u>Curso</u>
<u>Discente</u>	<u>Aluno</u>
<u>Grande Área</u>	<u>Grande Área de Conhecimento</u>
<u>Instituição de Ensino Superior</u>	<u>Instituição Educacional</u>
<u>Local</u>	<u>Tipo de Localização</u>
<u>Organização de Ensino</u>	<u>Organização</u>

Fonte: Elaborado pelo próprio autor.

Quadro 10 - Divergência entre esta dissertação e a pesquisa científica de Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b)

Divergência entre entidade e propriedade	
A presente pesquisa	Silva, Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b)
<u>modalidade de ensino</u> é um atributo da entidade <u>Membro</u>	<u>Modalidade de Ensino</u> é uma entidade
<u>matrícula</u> é um atributo da entidade <u>Discente</u>	<u>Matrícula</u> é uma entidade
: <u>CategoriaAdministrativa</u> de estereótipo «enumeration»	<u>Categoria Administrativa</u> de estereótipo «kind»
: <u>OrganizacaoAcademica</u> de estereótipo «enumeration» contém as organizações por atributos: : <u>CentroFederalEducacaoTecnologica</u> ; : <u>CentroUniversitario</u> ; : <u>Faculdade</u> ; : <u>InstitutoFederalEducacaoCienciaTecnologia</u> ; : <u>Universidade</u> ;	Classes de estereótipo «SubKind»: <u>Escola</u> ; <u>Colégio</u> ; <u>Faculdade</u> ; <u>Instituto</u> ; <u>Universidade</u> ;

Fonte: Elaborado pelo próprio autor.

De acordo com o quadro 9, é possível observar que as classes Área de Conhecimento e Pessoa são equivalentes, assim como Discente e Aluno, e também as demais classes, pois embora a nomenclatura seja diferente, a semântica é a mesma. Já no Quadro 10, possuem divergências de diversas naturezas. Por exemplo, nesta dissertação, modalidade de ensino é um atributo da entidade Membro, enquanto na pesquisa científica de Silva Mutz e Ruy (2022a) e Silva, Mutz e Ruy (2022b), Modalidade de Ensino é classificada como uma entidade. A pesquisa difere também

na maneira de possibilitar interoperabilidade semântica, porque enquanto eles combinaram ontologias com técnicas de *Machine Learning*, esta dissertação usou consultas SPARQL para recuperar informação de dados ligados em RDF.

A pesquisa científica de Andrade, Nardi e Ruy (2023) também houve semelhança considerável com o que foi desenvolvido na presente pesquisa. Eles adotaram UFO e OntoUML para realizar tanto a ontologia de referência de domínio como a ontologia operacional. O que se percebe de diferente, está no fato da ontologia de referência de domínio desta dissertação sugerir um modelo conceitual para entender o contexto de TCC de pós-graduação, enquanto a ontologia de referência de domínio deles propõe uma integração dos dados de publicações científicas como resumo, artigo, capítulo de livro, dentre outros, excluindo as teses e dissertações acadêmicas. Outra diferença esteve no fato da pesquisa deles propor o mapeamento semântico por meio das bibliotecas *scholarly*¹⁸¹, *semanticscholar*¹⁸², *fuzzywuzzy*¹⁸³, *pandas*¹⁸⁴ e *RDFlib*¹⁸⁵ da linguagem Python e o Protégé para gerar o repositório da ontologia operacional em um arquivo OWL, enquanto nesta pesquisa adotou-se o Ontotext Refine para realizar o mapeamento dos dados ligados em RDF para alcançar a interoperabilidade semântica com o apoio de consultas em portais especializados de termos e vocabulários, e a ontologia operacional de domínio foi implementada no VochBench para gerar o arquivo OWL.

Enfim, realizar as cinco fases da metodologia SABiO resultou na elaboração de uma ontologia segundo os preceitos referendados por Falbo (2014), abarcando a ontologia de referência de domínio e a ontologia operacional de domínio. A importância de elaborar e validar uma ontologia de domínio está em possibilitar a recuperação da informação no âmbito de pós-graduação stricto sensu no Brasil, usando-se um repositório experimental do acervo de dissertações dos alunos egressos do PPGCI/UFES, bem como fundamentado nos princípios FAIR defendidos por Guizzardi (2020) e argumentados por Wilkinson *et al.* (2016).

¹⁸¹ *Scholarly* possibilita resgatar informações de autores e publicações no Google Scholar (ANDRADE; NARDI; RUY, 2023)

¹⁸² *Semanticscholar* permite resgatar dados da API *Semantic Scholar* (ANDRADE; NARDI; RUY, 2023).

¹⁸³ *Fuzzywuzzy* concede o índice de semelhança entre duas strings (ANDRADE; NARDI; RUY, 2023).

¹⁸⁴ *Pandas* possibilita a leitura de arquivos contendo as informações, oferecendo estrutura e operações para manipular tabelas (ANDRADE; NARDI; RUY, 2023).

¹⁸⁵ *RDFlib* permite serializar dados em OWL (ANDRADE; NARDI; RUY, 2023).

7 CONSIDERAÇÕES FINAIS

Apresentou-se análise do estado da arte e considerável discussão sobre elementos da Organização do Conhecimento no domínio da CI e da Web semântica. Esses elementos permitiram elucidar os fundamentos e entender o processo da proposta desta dissertação que foi elaborar uma ontologia de referência para o âmbito de Trabalho de Conclusão de Curso de pós-graduação e implementá-la em dados ligados no contexto da Web semântica.

O texto discorreu acerca da organização do conhecimento, apresentando alguns SOCs como mecanismos para organizar a informação e a seleção de conceitos com relações semânticas. O assunto de ontologia enquanto SOC foi explorado com mais profundidade, por ser este um componente básico para a Web semântica, e que incorpora significado a conteúdos publicados na Web. Acredita-se assim, ter-se cumprido o primeiro objetivo específico estabelecido no trabalho.

O segundo objetivo específico foi atingido através da elaboração de um modelo conceitual como uma ontologia de referência para o domínio de TCCs, com a criação de um *knowledge graph* preliminar, em conformidade com as duas primeiras atividades da metodologia SABiO. A ontologia de referência foi baseada na ontologia de fundamentação UFO e adotou a linguagem de modelagem orientada a ontologia OntoUML. ênfase foi direcionada a entidades mais próximas da defesa de TCCs e metadados promissores para futuras recuperações de informação. No entanto, observa-se que outras entidades e relacionamentos poderiam fazer parte deste modelo conceitual, como por exemplo, o projeto vinculado à bolsa de estudos, e a publicação do edital de bolsas.

Em seguida, foi desenvolvido um modelo de *design* com a adoção da terceira atividade da metodologia SABiO, usando vocabulários para equivalências de classes, de propriedades de dados e propriedades de objetos. As equivalências foram o resultado da busca em portais especializados de termos e vocabulários atendendo à interoperabilidade da camada semântica e estrutural. Foram identificadas 44 entidades com equivalências (*owl:equivalentClass*), 36 propriedades de dados com equivalência (*owl:DatatypeProperty*) e 28 equivalências (*ObjectProperty*). Dessa forma foi cumprido o terceiro objetivo específico.

Como próximo passo, implementou-se a ontologia operacional de domínio em atendimento à quarta atividade da metodologia SABiO, a partir do modelo de *design*.

Essa implementação foi codificada como dados ligados em RDF para a Web Semântica, possibilitando a interoperabilidade das camadas sintática e estrutural, planejadas no modelo de *design*. Este resultado correspondeu ao cumprimento do quarto objetivo específico estabelecido.

A partir da ontologia operacional executou-se a depuração e reconciliação dos dados, com a aplicação de mecanismos semi-automáticos do OpenRefine, viabilizando parte da interoperabilidade da camada semântica. Em seguida, realizou-se o mapeamento dos dados pré-processados com a ontologia operacional para obter os dados ligados RDF de uma base de dados experimental, cumprindo o quinto objetivo específico.

A recuperação de informação foi alcançada com a adoção de algumas consultas SPARQL no repositório experimental de dados ligados. Os resultados das consultas permitiram validar parcialmente a ontologia (uma prova de conceito), gerando indícios da qualidade dos artefatos propostos, o que correspondeu ao cumprimento do sexto objetivo específico.

O objetivo geral da pesquisa foi cumprido por meio dos passos realizados nos objetivos específicos, com a elaboração de uma ontologia de referência de domínio para representar trabalhos de conclusão de curso de pós-graduação e o seu mapeamento em uma ontologia operacional no contexto de dados ligados na Web semântica.

A adoção de ontologias no âmbito da Web semântica proporciona um aumento na capacidade interoperável do mecanismo de recuperação de informação. No caso específico desse trabalho, o atendimento da interoperabilidade das camadas sintática, estrutural e semântica, associada a bases reconhecidas como GeoNames, Wikidata e DBpedia, colaboraram para que a organização da base ficasse sintonizada com a interoperabilidade prevista nos princípios FAIR, em sintonia com ideais para uma ciência mais aberta.

Os resultados desta pesquisa tem relevância científica e servem para balizar novas pesquisas no âmbito de organização do conhecimento, como também no domínio de TCC de pós-graduação *stricto sensu*, além de em atividades específicas para as quais pode servir de exemplo, tais como a construção de ontologia de referência de domínio e o mapeamento desta para uma ontologia operacional correspondente, e também encorajando novas pesquisas a ampliar o nível de

descobertas sobre esse âmbito ou com objetivos similares ao aqui proposto. E ainda, obteve-se amadurecimento e vivência científica por parte do autor desta pesquisa, quanto ao conhecimento explorado por intermédio da análise e levantamento da OC e RI na CI com suporte de dados ligados interoperáveis na Web semântica, com o uso de boas práticas de publicação, e devido aos processos inerentes às práticas metodológicas adotadas como a metodologia SABiO e a linguagem OntoUML.

Embora por si só as informações sobre disseminação de pesquisas acadêmicas ou a própria transparência dos dados nos portais de serviço público já sejam de importância considerável em termos de organização, publicação e recuperação para a sociedade, é preciso possibilitar que a sociedade encontre essas informações (encontrabilidade). Ademais, observa-se importância considerável na multiplicidade de maneiras da população acessar tais informações (acessibilidade), possuir capacidade de transferir, bem como utilizar dados de forma uniforme e eficiente entre várias organizações, instituições, multiplicidade de línguas, diversidade de plataformas ou sistemas e tecnologias institucionais (interoperabilidade), e ainda, reutilizar as informações para novas pesquisas sem custos para coletá-los novamente (reusabilidade).

Como recomendação de perspectiva futura, cita-se a procura em portais especializados de termos e vocabulários para a adoção de equivalências para os elementos que não foram tratados nessa pesquisa (sendo os números: 17 entidades, 8 propriedades e 8 relações sem equivalências). Aponta-se, inclusive, a elaboração de um conjunto de questões de competência na elicitação de requisitos para, então, aplicá-las de acordo com um estudo de usuários. Indica-se também a validação dos artefatos desenvolvidos em uma base de dados substancialmente maior e o desenvolvimento de interfaces para oferecer aos usuários finais possibilidades de consultas com boa usabilidade, uma vez que o uso da linguagem SPARQL para elaboração de consultas não é simples para o usuário final. Por fim, é pertinente a ideia de realizar um processo similar em um domínio diferente, aumentando assim a quantidade de domínios mapeados de forma similar e que poderiam aos poucos se interligarem.

8 REFERÊNCIAS

- ACADEMIA BRASILEIRA. **Nato-digital**. Disponível em: <https://www.academia.org.br/nossa-lingua/nova-palavra/nato-digital>. Acesso em: 8 nov. 2022.
- ACCUEIL. **Reference information about authors, works and topics**. 2014. Disponível em: <https://data.bnf.fr/>. Acesso em: 31 mar. 2022.
- ALBERTONI, R.; ISAAC, A. **Data on the Web Best Practices: data quality vocabulary**. 2016. Disponível em: <https://www.w3.org/TR/vocab-dqv/>. Acesso em: 30 mar. 2022.
- ALBUQUERQUE, M. E. B. C. et al. **Representação da informação: um universo multifacetado**. João Pessoa: Editora da UFPB, 2013. Disponível em: file:///C:/Users/rafae/Downloads/45-8.%20EDU_DIAG%20-%20PDF%20completo-10446-1-10-20220829.pdf. Acesso em: 01 jun. 2023.
- ALMEIDA, M. B. **Revisiting ontologies: a necessary clarification**. Journal of the American Society for Information Science and Technology. 2019. v. 17, p. e019037. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8653319>. Acesso em: 11 nov. 2022.
- ALMEIDA, M. B. **Ontologias na Ciência da Informação: estado da arte no Brasil. ciência da informação**. 2017. Disponível em: <https://revista.ibict.br/ciinf/issue/view/237>. Acesso em: 5 jun. 2023.
- ALMEIDA, J. P. A.; FALBO, R. A.; GUIZZARDI, G. S.; TIAGO P. Gufo: **A lightweight implementation of the Unified Foundational Ontology (UFO)**. Disponível em: <https://nemo-ufes.github.io/gufo/>. Acesso em: 26 abr. 2023.
- AMBINDER, D. M; MARCONDES, C. H. **Leitura de artigos científicos digitais na Web**. p. 8, 2013. Disponível em: <https://www.scielo.br/j/tinf/a/BFHGZPhVKGX8qLGdDBmsXCp/?lang=pt#>. Acesso em: 25 ago. 2022.
- AMEEN, A.; KHAN, K. U. R.; RANI, B. P. **Creation of ontology in education domain**. 2012. IEEE Fourth International Conference on Technology for Education. Hyderabad, India: IEEE, 2012. p. 237–238. Disponível em: <http://ieeexplore.ieee.org/document/6305981/>. Acesso em: 20 set. 2022.
- ANDRADE P. J.; NARDI C. J.; RUY B. F. **Integração de dados de publicações científicas usando uma abordagem baseada em ontologias**. In: Computer on the Beach, XIV, 2023, Santa Catarina, SC. Anais [...]. Itajaí, SC: Universidade do Vale do Itajaí/SC, 2023. p.158–165. Disponível em: <https://periodicos.univali.br/index.php/acotb/article/view/19428>. Acesso em: 8 maio 2023.
- ARAÚJO, C. A. Á. **O que é Ciência da Informação**. Belo Horizonte: KMA, 2018.

ARAÚJO, C. A. Á. **Arquivologia, Biblioteconomia, Museologia e Ciência da Informação: o diálogo possível.** 2014. Disponível em: <http://labds.eci.ufmg.br:8080/bitstream/123456789/54/1/ARA%C3%A9AJO.%20Arquivologia%2C%20Biblioteconomia%2C%20museologia%20e%20ciencia%20da%20informa%C3%A7%C3%A3o.pdf>. Acesso em: 30 mar. 2022.

ARAÚJO, C. A. Á.; VALENTIM, M. L. P. **A Ciência da Informação no Brasil: mapeamento da pesquisa e cenário institucional.** Bibliotecas. Anales de Investigación. 2019. v. 15, n. 2, p. 28. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=7871021>. Acesso em: 19 mar. 2023.

ASHOUR, G.; ALJOHANI, N.; AI-DUBAI, A.; ROMDHANI, I. **An ontological model for courses and academic profiles representation: a case study of King Abdulaziz University.** International Conference Engineering Technologies and Computer Science (EnT). Moscow, Russia: IEEE, 2020. p. 123–126. Disponível em: <https://ieeexplore.ieee.org/document/9140525/>. Acesso em: 31 ago. 2022.

BAEZA-YATES, R. et al. **Modern Information Retrieval.** New York: ACM press, 1999.

BAUER, F.; KALTENBÖCK, M. **Linked Open Data: the essentials: a quick start guide for decision makers.** Wien: Ed. mono/monochrom, 2012. Disponível em: https://www.researchgate.net/publication/280113736_Linked_Open_Data_The_Essentials_A_quick_start_guide_for_decision_makers. Acesso em: 19 jun. 2022.

BENÍCIO, C. D.; SILVA, A. K. A. **Do livro impresso ao e-book: o paradigma do suporte na biblioteca eletrônica,** 2005. Disponível em: https://www.brapci.inf.br/_repositorio/2010/11/pdf_83aedd01d5_0013473.pdf. Acesso em: 11 fev. 2023.

BERNERS-LEE, T. **Linked Data: up to design issues.** 2010. Disponível em: <https://www.w3.org/DesignIssues/LinkedData>. Acesso em: 10 jun. 2022.

BERNERS-LEE, T. **Web for real people.** W3C - World Wide Web Consortium, 2005. Disponível em: <https://www.w3.org/2005/Talks/0511-keynote-tbl/>. Acesso em: 19 jun. 2022.

BERNERS-LEE, T. ; HENDLER, J. ; LASSILA, O. **The semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities.** Scientific American. 2002. Disponível em: http://csis.pace.edu/~marchese/CS835/Lec9/112_SemWeb.pdf. Acesso em: 01 ago. 2021.

BERNERS-LEE, T; HENDLER, J; LASSILA, O. **The semantic Web.** Scientific american, v. 284, n. 5, p. 28–37, 2001.

BERTOLINI V. B. et al. **Statement of International Cataloguing Principles (ICP).** 2016. Disponível em: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-en.pdf. Acesso em: 10 maio 2023.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked Data**: the story so far. *International Journal on Semantic Web and Information Systems*, IGI global, v. 5, n. 3, p. 1–22, 2009. Disponível em: https://www.researchgate.net/publication/225070216_Linked_Data_The_Story_so_Far. Acesso em: 01 jun. 2022.

BOECKHOUT, M.; ZIELHUIS, G. A.; BREDENOORD, A. L. **The FAIR guiding principles for data stewardship**: FAIR enough? *European Journal of Human Genetics*, 2018, v. 26, n. 7, p. 931–936. Disponível em: <https://www.nature.com/articles/s41431-018-0160-0>. Acesso em: 18 set. 2022.

BOER, V.; RAVEN, D.; ESMEIJER, E; OOMEN, J. **Modeling ontologies for individual artists**: a case study of a dutch ceramic glass sculptor. *DH Benelux Journal*. 2022. Disponível em: <https://journal.dhbenelux.org/journal/issues/004/article-4-DeBoer.pdf>. Acesso em: 28 maio 2023.

BRASCHER, M.; CAFÉ, L. **Organização da informação ou Organização do Conhecimento?** In: Encontro Nacional de Pesquisa em Ciência da Informação, 9, 2008. São Paulo, Anais[...]. São Paulo, USP, 2008.

BRASIL. **Portal Brasileiro de Dados Abertos**. 2023a. Disponível em: <https://dados.gov.br/home>. Acesso em: 21 maio 2023.

BRASIL. **Dados Abertos do Ministerio da Educação**. 2023b. Disponível em: <https://dadosabertos.mec.gov.br/>. Acesso em: 14 jun. 2023.

BRASIL. **Resolução nº 3, de 28 de janeiro de 2022**. Trata-se do Regulamento Geral da Pós-Graduação da Universidade Federal do Espírito Santo e revoga a Resolução nº 11, de 13 de abril de 2010, com redação alterada pelas resoluções subsequentes. 2022a. Disponível em: https://cienciadainformacao.ufes.br/sites/cienciadainformacao.ufes.br/files/field/anexo/resolucao_no_3.2022_-_regulamento_geral_da_pos-graduacao.pdf. Acesso em: 29 jan. 2023.

BRASIL. **Regimento do PPGI, de 16 de setembro de 2022**. Constitui, em conjunto com o Estatuto e o Regimento Geral da Universidade Federal do Espírito Santo (UFES), o Regulamento Geral da Pós-Graduação da UFES e os demais dispositivos legais, o documento regulador e disciplinador das atividades de ensino e pesquisa do Programa de Pós- Graduação em Informática (PPGI) stricto sensu, nos cursos de Mestrado em Informática e de Doutorado em Ciência da Computação. UFES, Vitória. 2022b. Disponível em: https://informatica.ufes.br/sites/informatica.ufes.br/files/field/anexo/regimento_4.0_0.pdf. Acesso em: 02 nov. 2022.

BRASIL. **Plano de Dados Abertos**. 2021. Disponível em: <https://www.gov.br/cgu/pt-br/aceso-a-informacao/dados-abertos/arquivos-eleicoes/plano-de-dados-abertos.pdf>. Acesso em: 21 maio 2023.

BRASIL. **Decreto nº 10.278/2020, de 18 de março de 2020**. Regulamenta o disposto no, inciso X do caput do art. 3º da Lei nº 13.874, de 20 de setembro de 2019, e no art. 2º-A da Lei nº 12.682, de 9 de julho de 2012, para estabelecer a

técnica e os requisitos para a digitalização de documentos públicos ou privados, a fim de que os documentos digitalizados produzam os mesmos efeitos legais dos documentos originais. 2020a. Disponível em:
https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/d10278.htm. Acesso em: 11 fev. 2023.

BRASIL. Resolução nº 01/2020, de 14 de agosto de 2020. Estabelece normas e critérios para a Defesa de Dissertação no âmbito do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo (PPGCI/UFES). UFES, Vitória. 2020b. Disponível em:
https://cienciainformacao.ufes.br/sites/cienciainformacao.ufes.br/files/field/anexo/resolucao_01-2020_ppgci_.pdf. Acesso em: 03 nov. 2022.

BRASIL. Resolução nº 04/2019, de 09 de dezembro de 2019. Estabelece normas e critérios para o Exame de Qualificação no âmbito do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo (PPGCI/UFES). UFES, Vitória. 2019. Disponível em:
https://cienciainformacao.ufes.br/sites/cienciainformacao.ufes.br/files/field/anexo/resolucao_04-2019_ppgci_qualificacao_com_anexos.pdf. Acesso em: 06 nov. 2022.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018, que estabelece e regula a Lei de Proteção de Dados. 2018a. Disponível em:
http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm#:~:text=1%C2%BA%20Esta%20Lei%20disp%C3%B5e%20sobre,da%20personalidade%20da%20pessoa%20natural. Acesso em: 07 jun. 2022.

BRASIL. Resolução nº 4, de 11 de dezembro de 2018. Altera o inciso I do artigo 2º da Resolução CNE/CES nº 1, de 6 de abril de 2018, que estabelece diretrizes e normas para a oferta dos cursos de pósgraduação lato sensu denominados cursos de especialização, no âmbito do Sistema Federal de Educação Superior, conforme prevê o Art. 39, § 3º, da Lei nº 9394/1996, e dá outras providências. 2018b. Disponível em:
http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=103631-rces004-18&category_slug=dezembro-2018-pdf&Itemid=30192. Acesso em: 12 fev. 2023.

BRASIL. Portaria nº 122, de 19 de junho de 2017 (DOU de 21/06/2017, seção 2, p. 12). O presidente da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, no uso das atribuições que lhe são conferidas pelo Estatuto aprovado pelo Decreto nº 8.977, de 30 de janeiro de 2017. Disponível em:
https://www.periodicos.capes.gov.br/images/documents/Portaria_n.122_PAAP_Alterar%20o%20par%C3%A1grafo%20%C3%BAnico%20do%20Art.%2019%C2%BA%20da%20Portaria%20n%C2%BA%2074_19%20junho%202017.pdf. Acesso em: 21 maio 2023.

BRASIL. Resolução nº 36/2014, de 07 de agosto de 2014. Normas e procedimentos para reprodução de acervo no âmbito do SIB/UFES e dá outras providências. UFES, Vitória. 2014. Disponível em:

https://daocs.ufes.br/sites/daocs.ufes.br/files/field/anexo/resolucao_no._36.2014.pdf. Acesso em: 03 nov. 2022.

BRASIL. **Lei nº 12.682/2012, de 9 de julho de 2012**. Dispõe sobre a elaboração e o arquivamento de documentos em meios eletromagnéticos. 2012. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/Lei/L12682.htm. Acesso em: 11 fev. 2023.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011, que estabelece e regula o acesso à informação**. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 07 jun. 2022.

BRASIL. **Lei nº 10.973, de 2 de dezembro de 2004, que dispõe sobre incentivos à inovação e à pesquisa científica e tecnológica no ambiente produtivo e dá outras providências**. 2004a. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10.973.htm. Acesso em: 1 nov. 2022.

BRASIL. **Portaria nº 4.059, de 10 de dezembro de 2004 (DOU de 13/12/2004, seção 1, p. 34)**. Introduz na organização pedagógica e curricular de seus cursos superiores reconhecidos, a oferta de disciplinas integrantes do currículo que utilizem modalidade semi-presencial, com base no art. 81 da Lei n. 9.394, de 1.996, e no disposto nesta Portaria. 2004b. Disponível em: http://portal.mec.gov.br/sesu/arquivos/pdf/nova/acs_portaria4059.pdf. Acesso em: 12 fev. 2023.

BRASIL. **Resolução CNE/CES Nº 24, de 18 de dezembro de 2001, que estabelece normas para o funcionamento de cursos de pós-graduação**. 2002. Disponível em: <http://portal.mec.gov.br/docman/dezembro-2017-pdf/78331-rces024-02-pdf/file#:~:text=Altera%20a%20reda%C3%A7%C3%A3o%20do%20par%C3%A1grafo,de%20cursos%20de%20p%C3%B3s%2Dgradua%C3%A7%C3%A3o>. Acesso em: 14 fev. 2023.

BRASIL. **Resolução CNE/CES Nº 1, de 3 de abril de 2001, que altera a redação do parágrafo 4º do artigo 1º e o artigo 2º, da Resolução CNE/CES 1/2001, que estabelece normas para o funcionamento de cursos de pós-graduação**. 2001. Disponível em: <http://portal.mec.gov.br/seed/arquivos/pdf/tvescola/leis/CES0101.pdf>. Acesso em: 14 fev. 2023.

BRASIL. **Lei nº 67, de 26 de outubro de 1998, que estabelece e regula a Lei de Proteção de Dados**. 1998. Disponível em: <https://dre.pt/dre/detalhe/lei/67-239857>. Acesso em: 07 jun. 2022.

BRASIL. **Lei nº 9.394, de 20 de dezembro de 1996**. Estabelece as diretrizes e bases da educação nacional. 1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9394.htm. Acesso em: 12 fev. 2023.

BRASIL. **Lei nº 8.666, de 21 de junho de 1993, que regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da administração pública e dá outras providências**. 1993. Disponível em:

http://www.planalto.gov.br/ccivil_03/LEIS/L8666cons.htm#art24xiii. Acesso em: 1 nov. 2022.

BRASIL. **Constituição Federal de 1988**. 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 03 mar. 2022.

CAMPOS, P. M. C. **Designing a network of reference ontologies for the integration of water quality data**. 2019. Dissertação (mestrado) - Universidade Federal do Espírito Santo, Vitória, ES. Disponível em: https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/designing_a_network_of_reference_ontologies_for_the_integration_of_water_quality_data_2019.pdf. Acesso em: 30 abr. 2023.

CAPURRO, R.; HJØRLAND, B. **The concept of information**. 2003, Hochschule der Medien. Disponível em: <http://www.capurro.de/infoconcept.html>. Acesso em: 3 jun. 2022.

CARCHEDI, L. C. et al. **Onto4LA**: uma ontologia para integração de dados educacionais. em: VII Congresso Brasileiro de Informática na Educação, 2018, Fortaleza, Ceará, Brasil. Anais [...]. Fortaleza, Ceará, Brasil. p. 439. Disponível em: <https://repositorio.ufc.br/handle/riufc/44031>. Acesso em: 1 mar. 2023.

CARLAN, E.; MEDEIROS, M. B. B. **Sistemas de Organização do Conhecimento na visão da Ciência da Informação**. 2011. Revista Ibero-Americana de Ciência da Informação, Brasília, v. 4, n. 2, p. 53–73. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/1675/1474>. Acesso em: 07 jun. 2021.

CARLAN, E. **Sistemas de Organização do Conhecimento**: uma reflexão no contexto da Ciência da Informação. 2010. Dissertação (mestrado) - Universidade de Brasília, Brasília, DF. Disponível em: <https://repositorio.unb.br/handle/10482/7465>. Acesso em: 18 jun. 2022.

CASTRO, F. F.; SIMIONATO, A. C.; ZAFALON, Z. R. **Aspectos relacionais entre ontologia e metadados**: considerações interdisciplinares. 2016, XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB). Disponível em: <https://brapci.inf.br/index.php/res/v/191839>. Acesso em: 15 junho 2023.

CLARK, K. G.; FEIGENBAUM, L; TORRES, E. **SPARQL Protocol for RDF**. 2008. Disponível em: <https://www.w3.org/TR/rdf-sparql-protocol/>. Acesso em: 21 jun. 2022.

CONARQ. **Glossário: documentos arquivísticos digitais**. Rio de Janeiro: Conselho Nacional de Arquivos - CONARQ/Câmara Técnica de Documentos Eletrônicos - CTDE, 2020. Disponível em: https://www.gov.br/conarq/pt-br/assuntos/camaras-tecnicas-setoriais-inativas/camara-tecnica-de-documentos-eletronicos-ctde/glosctde_2020_08_07.pdf. Acesso em: 05 jun. 2023.

CONEGLIAN, C. S. **Recuperação da Informação com abordagem semântica utilizando linguagem natural**: a Inteligência Artificial na Ciência da Informação.

Tese (doutorado) - Universidade Estadual Paulista, Marília, SP, 2020. Disponível em: <https://repositorio.unesp.br/handle/11449/193051>. Acesso em: 01 mar. 2022.

CONEGLIAN, C. S.; SEGUNDO, J. E. S. **Europeana no Linked Open Data: conceitos de Web semântica na dimensão aplicada das humanidades digitais.** *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 22, n. 48, p. 88, 2016. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n48p88>. Acesso em: 01 jun. 2022.

CONEGLIAN, C. S.; SOUZA, J. O.; SEGUNDO, J. E. S. **Análise de publicações da temática “Web semântica” na América Latina: um olhar nas bases de dados da Web of Science.** *Brazilian Journal of Information Science: research trends*. 2016. v. 10, n. 2. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/6002>. Acesso em: 31 ago. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Estatísticas e indicadores da pesquisa no Brasil.** 2023. Disponível em: <http://memoria2.cnpq.br/documents/10157/f653f40d-8a19-468a-aa8f-bb3ac557e32c>. Acesso em: 09 fev. 2023.

CORPORATION FOR DIGITAL SCHOLARSHIP. **Zotero, your personal research assistant.** Zotero, 2020, Disponível em: <https://www.zotero.org/> Acesso em: 2 jun. 2022.

CRESWELL, J. W. **Projeto de pesquisa.** Porto Alegre: Sage, 2010.

CRISTOVÃO, H. M. **Recuperação de informação: aplicações na Web semântica.** 2023. ISBN 978-65-00-63129-6. Notas de aula da disciplina Sistemas de Recuperação da Informação. Programa de Pós-Graduação em Ciência da Informação, PPGCI/UFES. [web page]. Disponível em: <http://bit.ly/cristovaohm-ri>. Acesso em: 16 jun. 2023.

CRISTOVÃO, H. M. **Sistemas de Recuperação da Informação.** Apresentação da Unidade 3: Recuperação de Informação em Redes Semânticas, do curso de Mestrado Acadêmico da Universidade Federal do Espírito Santo, 2021a.

CRISTOVÃO, H. M. **Ontologias e vocabulários de metadados: portais de busca, coletânea e bibliografia.** Documento de apresentação de portais de busca de termos e vocabulários selecionados com foco em Ciência da Informação, do curso de Mestrado Acadêmico da Universidade Federal do Espírito Santo, 2021b.

CRISTOVÃO, H. M. **Um modelo híbrido de Recuperação de Informação e conhecimento baseado na síntese de mapas conceituais obtidos por operações de transformação de redes complexas orientadas por busca de relacionamentos entre termos de consulta em bases de dados ligados.** 2016. Doutorado - Universidade de Brasília. Disponível em: <https://repositorio.unb.br/handle/10482/22284>. Acesso em: 19 jun. 2022.

COELHO, A. G.; LIMA, G. Â.; BORGES, M. M. **Ciência da Informação: tendências temática e diacrónica** (2011-2020). p. 11, 2012.

DAHLBERG, I. **Knowledge Organization**: its scope and possibilities. *KO Knowledge Organization*, vol. 20, no 4, dezembro de 1993, p. 211–22, <https://doi.org/10.5771/0943-7444-1993-4-211>. Disponível em: <https://www.nomos-elibrary.de/index.php?doi=10.5771/0943-7444-1993-4-211>. Acesso em: 18 jun. 2022.

DEITEL, P.; DEITEL, H. **Java**: como programar. 10 ed. São Paulo: Pearson, 2016. Disponível em: <https://www.bvirtual.com.br/NossoAcervo/Publicacao/39590>. Acesso em: 11 jun. 2023.

DETONI, A. A. **EARly-OE**: atividades iniciais de engenharia de ontologias apoiadas em modelos de arquitetura organizacional. 2019. Tese (doutorado) - Universidade Federal do Espírito Santo, Vitória, ES. Disponível em: <https://informatica.ufes.br/pt-br/pos-graduacao/PPGI/detalhes-da-tese?id=14064>. Acesso em: 30 abr. 2023.

DPGLOSSARY. **Alliance for permanent access network**. 29 jun. 2015. Disponível em: <http://www.alliancepermanentaccess.org/index.php/consultancy/dpglossary/>. Acesso em: 21 jun. 2022.

DRAKOPOULOS, G.; GOURGARIS, P.; KANAVOS, A. **Graph communities in Neo4j**: four algorithms at work. *Evolving Systems*, v. 11, n. 3, p. 397–407, 2020. Disponível em: <https://link.springer.com.ez120.periodicos.capes.gov.br/article/10.1007/s12530-018-9244-x>. Acesso em: 27 abr. 2023.

DRAW.IO. **Supercharge your diagramming in confluence**. Disponível em: <https://drawio-app.com/product/>. Acesso em: 26 abr. 2023.

DROESCHER, F. D; SILVA, E. L. D. **O pesquisador e a produção científica**. *Perspectivas em Ciência da Informação*, v. 19, p. 170-189, 2014. Disponível em: <https://www.scielo.br/j/pci/a/ww5zR3KhYCK65bPkWJyTQtf/?lang=pt>. Acesso em: 26 abr. 2023.

ENDNOTE. **Clarivate**. 2023. Disponível em: <https://clarivate.com/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/endnote-site-license/>. Acesso em: 26 abr. 2023.

ELSTON, D. M. **Mendeley**. *Journal of the American Academy of Dermatology*, v. 81, n. 5, p. 1071, 2019. Disponível em: <https://doi.org/10.1016/j.jaad.2019.06.1291>. Acesso em: 24 abr. 2023.

ETHNOLOGUE LANGUAGES OF THE WORLD. **Browse by language name**. 2022. Disponível em: <https://www.ethnologue.com/browse/names>. Acesso em: 12 fev. 2023.

ERDMANN, A. L. **A importância da publicação científica**. *Revista de Enfermagem da UFSM*, v. 1, n. 2, 2011. Disponível em: <https://periodicos.ufsm.br/reufsm/article/view/2995>. Acesso em: 8 maio. 2023.

EUROPEAN UNION. **The official portal for European data**. Disponível em: <https://data.europa.eu/en>. Acesso em: 20 maio 2023.

FALBO, R. A. SABiO: **Systematic approach for building ontologies**. In: 1st Joint Workshop onto.com, 8th., 2014, Rio de Janeiro, RJ. Anais [...]. Rio de Janeiro, RJ: CEUR/RJ, 2014. p.14. Disponível em: http://ceur-ws.org/Vol-1301/ontocomodise2014_2.pdf. Acesso em: 30 jan. 2023.

FARINELLI, F.; ALMEIDA, M. B. **Ontologias biomédicas: teoria e prática**. Em: Minicursos do Simpósio Brasileiro de Computação Aplicada à Saúde. Sociedade Brasileira de Computação, 2019. p. 48. E-book. Disponível em: https://www.researchgate.net/publication/333812081_Ontologias_biomedicas_teorica_e_pratica. Acesso em: 06 mar. 2022.

FARINELLI, F.; ELKIN, P. L. **Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico**. Ciência da Informação, v. 46, n. 1, 2017. Disponível em: <https://revista.ibict.br/ciinf/article/view/4018>. Acesso em: 26 abr. 2023.

FERNANDEZ, M.; GOMEZ-PEAREZ, A.; JURISTO, N. **Methontology: from ontological art towards ontological engineering**. 1997. Disponível em: <https://aaai.org/papers/0005-methontology-from-ontological-art-towards-ontological-engineering/>. Acesso em: 27 abr. 2023.

FERREIRA, J.; COELHO, L.; RODRIGUES, L. G.; DIAS, L.; CORTIZO, L. S.; PORRUA, M. MACHADO, N. C.; LORENZ, N. A.; COSTA, T. B.; LEAL, V. M. M. **UNIO Ebook Interop 2019: o mercado único digital da União Europeia como desígnio político: a interoperabilidade como o caminho a seguir**. 1ª ed. Pensamento Sábio - Associação para o conhecimento e inovação, 2019. E-book. Disponível em: <http://hdl.handle.net/1822/61446>. Acesso em: 18 set. 2022.

FIGUEIREDO, N. M. **Serviços de referência e informação**. São Paulo: Polis, 1992. Disponível em: <https://abecin.org.br/wp-content/uploads/2021/03/Servicos-de-referencia.pdf>. Acesso em: 25 maio 2023.

FOLKSONOMY. **Corante: folksonomy**. Disponível em: <https://corante.com/many/folksonomy/>. Acesso em: 3 jul. 2022.

FONSECA, S. F. **A comunicabilidade da ciência e o ethos científico: a divulgação da ciência em períodos de pandemia, uma contribuição de Robert K. Merton**. Revista de Estudos Interdisciplinares, v. 2, n. 6, 2020. Disponível em: <https://revistas.ceeinter.com.br/index.php/revistadeestudosinterdisciplinar/article/view/85>. Acesso em: 25 maio 2023.

FRANCIS, N. et al. **Cypher: an evolving query language for property graphs**. Sigmod/Pods '18: International Conference on Management of Data. Proceedings of the 2018 International Conference on Management of Data. Houston TX USA: ACM, 2018. p. 1433–1445. Disponível em: <https://dl.acm.org/doi/10.1145/3183713.3190657>. Acesso em: 27 abr. 2023.

FUNDAÇÃO COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Tabela de áreas do conhecimento**. Disponível em:

https://sapx.ufrb.edu.br/2019/documentos/areas_de_conhecimento_capes.pdf. Acesso em: 28 out. 2023.

FUNDAÇÃO COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Aplicativos CAPES?**. 2022. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/aplicativos-capes>. Acesso em: 6 jun. 2023.

FUNDAÇÃO COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **O que é CAPES?** 2021. Disponível em: <https://web.archive.org/web/20220322094220/https://www.gov.br/mec/pt-br/assuntos/noticias/o-que-e-a-capes>. Acesso em: 20 fev. 2023.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2019. Disponível em: https://www.academia.edu/48899027/Como_Elaborar_Projetos_De_Pesquisa_6a_Ed_GIL. Acesso em: 12 mar. 2023.

GILLILAND, A. J. **Introduction to metadata**. 2016. Interactive Resource. Disponível em: <http://www.getty.edu/publications/intrometadata>. Acesso em: 14 maio 2023.

GLASS, J. **OpenRefine user manual**. OpenRefine Documentation, janeiro de 2021. Disponível em: <https://docs.openrefine.org/>. Acesso em: 06 mar. 2022.

GLEICK, J. **A informação: uma história, uma teoria, uma enxurrada**. São Paulo: Companhia das Letras, 2013. Disponível em: https://edisciplinas.usp.br/pluginfile.php/340345/mod_resource/content/1/A-informacao-Uma-historia-uma-teoria-uma-enxurrada-James-Gleick-pdf.pdf. Acesso em: 06 mar. 2022.

GLOSSÁRIO DE TERMOS REFERENTES AO ENSINO DE GRADUAÇÃO DO CEFET-MG. **Guia de Gestão Acadêmica da Graduação**. 2015. Disponível em: <https://www.dcsa.cefetmg.br/wp-content/uploads/sites/35/2017/08/Glossario-Academico.pdf>. Acesso em: 30 out. 2022.

GOMES, D. L.; BARROS, T. H. B.; SOUSA, R. T. B.; JUNIOR, R. L. **Proposta de uma ferramenta para classificação arquivística com base em ontologias**. Em *Questão*, v. 26, n. 1, p. 351–374, 2020. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/91853>. Acesso em: 31 ago. 2022.

GONÇALVES, A. O.; JACYNTHO, M. D. de A. **Um método para publicação semântica Linked Data de bases de dados convencionais e um estudo de caso real de artigos acadêmicos**. *Transinformação*, v. 32, p. e180051, 2020. Disponível em: <https://www.scielo.br/j/tinf/a/Kv7kymrYXsZtTyn8QvtYV9Q/?lang=pt>. Acesso em: 31 ago. 2022.

GONZÁLEZ, J. A. M. **Linguagens documentárias e vocabulários semânticos para a Web: elementos conceituais**. EDUFBA. 2011. Disponível em: <https://books.scielo.org/id/y89xh>. Acesso em: 11 nov. 2022.

GRUBER, T. R. **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:
<https://www.sciencedirect.com/science/article/pii/S1071581985710816>. Acesso em:
 18 set. 2022.

GRÜNINGER, M., FOX, M. S. **Methodology for the design and evaluation of ontologies**. Workshop on Basic Ontological Issues in Knowledge Sharing, 1995. Disponível em: <http://www.eil.utoronto.ca/wp-content/uploads/enterprise-modelling/papers/gruninger-ijcai95.pdf>. Acesso em: 29 abr. 2023.

GUARINO, N. **The ontological level: revisiting 30 years of Knowledge Representation**. In: Borgida, Alexander T. et al. (org.). *Conceptual Modeling: Foundations and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. (Lecture Notes in Computer Science). v. 5600, p. 52–67. E-book. Disponível em: http://link.springer.com/10.1007/978-3-642-02463-4_4. Acesso em: 22 maio 2023.

GUARINO, N.; GIARETTA, P. **Ontologies and knowledge bases: towards a terminological clarification**. In: *Towards Very Large Knowledge Bases*. Amsterdam: IOS Press, 1995. Disponível em: <http://www.loa.istc.cnr.it/old/Papers/KBKS95.pdf>. Acesso em: 25 maio 2023.

GUIZZARDI, G. et al. **Automated conceptual model clustering: a relator-centric approach**. *Software and Systems Modeling*, v. 21, n. 4, p. 1363–1387, 2022. Disponível em: <https://link.springer.com/article/10.1007/s10270-021-00919-5>. Acesso em: 29 maio 2023.

GUIZZARDI, G. **Ontology, Ontologies and the “I” of FAIR**. *Data Intelligence*, v. 2, n. 1–2, p. 181–191, 2020. Disponível em: https://doi.org/10.1162/dint_a_00040. Acesso em: 06 maio 2022.

GUIZZARDI, G.; GRAÇAS, A. P.; GUIZZARDI, R. S. S. **Design patterns and inductive modeling rules to support the construction of ontologically well-founded conceptual models in OntoUML**. Bayro-Corrochano, Eduardo; Hancock, Edwin (org.). *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Cham: Springer International Publishing, 2011. (Lecture Notes in Computer Science). v. 8827, p. 402–413. E-book. Disponível em: http://link.springer.com/10.1007/978-3-642-22056-2_44. Acesso em: 1 maio 2023.

GUIZZARDI, G. **On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models**. *Databases and Information Systems IV: Selected Papers from the Seventh International Baltic Conference, DB&IS'2006*. IOS Press, 2006. p. 18. Disponível em: https://www.researchgate.net/publication/221278057_On_Ontology_ontologies_Conceptualizations_Modeling_Languages_and_MetaModels. Acesso em: 1 maio 2023.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. 2005. PhD Tese (doutorado) - Centre for Telematics and Information Technology, Enschede. Disponível em: https://ris.utwente.nl/ws/portalfiles/portal/6042428/thesis_Guizzardi.pdf. Acesso em: 06 maio 2023.

GUIZZARDI, G.; WAGNER, G. **Some applications of a Unified Foundational Ontology in business modeling**. Eindhoven University of Technology. The Netherlands, p. 23, 2005.

GUTIERREZ, C.; JUAN F. S. **Knowledge Graphs**. Communications of the ACM, vol. 64, no. 3, 2021, pp. 96–104, doi:10.1145/3418294.

HAMIM, T.; BENABBOU, F.; SAEL, N. **An Ontology-based decision support system for multi-objective prediction tasks**. International Journal of Advanced Computer Science and Applications, v. 12, n. 12, 2021. Disponível em: <http://thesai.org/Publications/ViewPaper?Volume=12&Issue=12&Code=IJACSA&SerialNo=24>. Acesso em: 1 jun. 2023.

HARPRING, P. et al. **Cataloging Cultural Objects: A guide to describing cultural works and their images**. American Library Association, 2006. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=om3Q5QER-UYC&oi=fnd&pg=PR11&dq=BACA,+M.%3B+HARPRING,+P.%3B+LANZI,+E.%3B+MCRAE,+L.%3B+WHITESIDE,+A.+Cataloging+Cultural+Objects:+A+Guide+to+Describing+Cultural+Works+and+Their+Images.+&ots=nQlbDRJKuy&sig=ftiWKdapjMUzmM7upt0Di0k2DDY#v=onepage&q&f=false>. Acesso em: 19 jun. 2022.

HENNING, P. C; RIBEIRO , C. J. S; SALES, L, F; MOREIRA, J. L. R; SANTOS, L. B. S. **Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR**. Pesquisa Brasileira em Ciência da Informação e Biblioteconomia, Rio de Janeiro, 2019. Disponível em: <https://periodicos.ufpb.br/index.php/pbcib/article/view/46969>. Acesso em: 9 jun. 2022.

HJØRLAND, B. **Information Retrieval and Knowledge Organization: a perspective from the philosophy of science**. Information, vol. 12, no 3, p. 135, 2021. Disponível em: <https://doi.org/10.3390/info12030135>. Acesso em: 20 jun. 2022.

HJØRLAND, B. **Knowledge Organization**. 2016. Disponível em: https://www.isko.org/cyclo/knowledge_organization. Acesso em: 30 jun. 2022.

HJØRLAND, B. **What Is Knowledge Organization (KO)?** Knowledge Organization, v. 35, n. 2–3, p. 86–101, 2008.

HODGE, G. **Systems of Knowledge Organization for digital libraries**. Washington, DC: Digital Library Federation, Council on Library and Information Resources, 2000. Disponível em: <https://www.clir.org/pubs/reports/pub91/>. Acesso em: 30 jul. 2021.

HOGAN, A. et al. **Knowledge Graphs**. ACM Computing Surveys. v. 54, n. 4, p. 1–37, 2021. Disponível em: <http://arxiv.org/abs/2003.02320>. Acesso em: 22 jul. 2022.

HOMMEAUX, E. P. CAROTHERS, G. **RDF Turtle**. 2014. Disponível em: <https://www.w3.org/TR/turtle/#language-features>. Acesso em: 21 jun. 2022.

HOOLAND, S. V.; VERBORGH, R.; WILDE, M. **Cleaning data with OpenRefine, programming historian 2**. 2013, <https://doi.org/10.46430/phen0023>. Disponível

em: <https://cacm.acm.org/magazines/2021/3/250711-knowledge-graphs/fulltext#PageTop>. Acesso em: 21 de dez. de 2021.

HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. **Best Practices for Publishing Linked Data**. 2014. Disponível em: <https://www.w3.org/TR/2014/NOTE-ld-bp-20140109/>. Acesso em: 26 mar. 2023.

INIAKPOKEIKIYE, T.; FAWEL, B. **An adaptable Ontology for easy and efficient University data management in Niger Delta University**. *Asian Journal of Research in Computer Science*, p. 19–27, 2022. Disponível em: <https://journalajrcos.com/index.php/AJRCOS/article/view/30315>. Acesso em: 26 jun. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do censo da educação superior 2021**. 2021. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>. Acesso em: 12 fev. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Censo da Educação Superior 2013**. 2013. Disponível em: https://download.inep.gov.br/educacao_superior/censo_superior/questionarios_e_ma_nuais/2013/glossario_consolidado_censup_2013.pdf. Acesso em: 30 out. 2022.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. **Digital libraries: metadata resources**. Disponível em: <https://www.ifla.org/g/cataloguing/digital-libraries-metadata-resources/>. Acesso em: 22 maio 2023.

IVEY, C.; CRUM, J. **Choosing the right citation management tool**: EndNote, Mendeley, RefWorks, or Zotero. *Journal of the Medical Library Association*, v. 106, n. 3, 2018. Disponível em: <http://jmla.pitt.edu/ojs/jmla/article/view/468>. Acesso em: 24 abr. 2023.

JOHANSSON, M. A. et al. **Preprints: an underutilized mechanism to accelerate outbreak science**. *PLOS Medicine*, v. 15, n. 4, p. e1002549, 2018. Disponível em: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002549>. Acesso em: 24 maio 2023.

KONSTANTINOU, N.; SPANOS, D. E. **Materializing the Web of Linked Data**. Cham: Springer International Publishing, 2015. E-book. Disponível em: <http://link.springer.com/10.1007/978-3-319-16074-0>. Acesso em: 2 maio 2022.

KRATOCHVÍL, J. **Comparison of the accuracy of bibliographical references generated for medical citation styles by EndNote, Mendeley, RefWorks and Zotero**. *The Journal of Academic Librarianship*, v. 43, n. 1, p. 57–66, 2017. Disponível em: <https://www-sciencedirect.ez120.periodicos.capes.gov.br/science/article/pii/S0099133316302294?via%3Dihub>. Acesso em: 22 abr. 2023.

LA-ONGSRI, S.; RODDICK, J. F. **Incorporating ontology-based semantics into conceptual modelling**. *Information Systems*, v. 52, p. 1–20, 2015. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0306437915000265?via%3Dihub>. Acesso em: 22 mar. 2023.

LAUFER, C. **Guia de Web semântica**. São Paulo, SP: CeWeb.br, Embaixada Britânica, iGovSP, 2015. Disponível em: https://www.cgi.br/media/docs/publicacoes/13/Guia_Web_Semantica.pdf. Acesso em: 30 julho 2021.

LE COADIC, Y. F. **A ciência da informação**. Briquet de Lemos, 1994. Disponível em: <https://pt.scribd.com/document/655208024/LE-COADIC-Y-F-a-Ciencia-Da-Informacao>. Acesso em: 11 jun. 2023.

LEMOS, D. L. S.; MARTINS, D. L. CARMO, D. **Padrões de qualidade para dados e metadados endereçados a aplicações em ciência de dados**. *Advanced Notes in Information Science*, Tallinn, Estonia, v. 2, p. 161–170, 2022. DOI: 10.47909/anis.978-9916-9760-3-6.116. Disponível em: <https://pub.colnes.org/index.php/anis/article/view/116/127>. Acesso em: 05 jun. 2022.

LÉVY, P. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. Rio de Janeiro: Editora 34, 1993.

LIMA, J. L. O.; ALVARES, L. **Organização e representação da informação e do conhecimento. Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4 Editores, v. 248, p. 21-48, 2012. Disponível em: https://www.researchgate.net/profile/Jose-Leonardo-Lima-2/publication/281969932_Organizacao_e_representacao_da_informacao_e_do_conhecimento/links/5600067308ae07629e522ad1/Organizacao-e-representacao-da-informacao-e-do-conhecimento.pdf. Acesso em: 30 maio 2023.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. **Data on the Web Best Practices**. 2017. Disponível em: <https://www.w3.org/TR/dwbp/>. Acesso em: 30 mar. 2022.

MACEDO, D. F. **Dados abertos governamentais: modelo de governança voltado a qualidade de dados para publicação em rede**. 2021. Dissertação (mestrado) - Universidade Federal do Espírito Santo, Vitória, ES. Disponível em: <https://cienciainformacao.ufes.br/pt-br/pos-graduacao/PPGCI/detalhes-da-tese?id=15610>. Acesso em: 20 mar. 2023.

MARCONDES, C. H. **Um modelo semântico de publicações eletrônicas**. p. 22, 2011. Disponível em: <https://revista.ibict.br/liinc/article/view/3290>. Acesso em: 25 ago. 2022.

MILES, A.; BECHHOFER, S. **SKOS: Simple Knowledge Organization System reference**. 2009. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>. Acesso em: 30 jun. 2022.

MOROSINI, M. C. et al. **Enciclopédia de pedagogia universitária**. Inep/MEC – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. 2006. Disponível em: https://download.inep.gov.br/publicacoes/diversas/temas_da_educacao_superior/enc

iclopedia_de_pedagogia_universitaria_glossario_vol_2.pdf. Acesso em: 29 out. 2022.

MCGARRY, K. **O contexto dinâmico da informação**: Uma análise introdutória. 1999. p. vi, 206-vi, 206.

MENDELEY. **Mendeley reference manager for desktop**. 2022. Disponível em: <https://www.mendeley.com/download-reference-manager/windows>. Acesso em: 24 abr. 2023.

MENDONÇA, F. M.; CASTRO, L. P. **OntoForInfoScience e Onto4ALLEditor**: metodologia e editor de ontologias como facilitadores na construção de ontologias por especialistas do domínio e cientistas da informação. FRC: Belo Horizonte, v.1, n1, 2021. Disponível em: <https://periodicos.ufmg.br/index.php/fronteiras-rc/article/view/35446>. Acesso em: 20 maio 2023.

MENDONÇA, F. M. et al. **Onto4AllEditor**: um editor web gráfico para construção de ontologias por todos os tipos de usuários da informação. Em *Questão*, p. 401–430, 2021. Disponível em: <https://doi.org/10.19132/1808-5245273.401-430>. Acesso em: 23 maio 2023.

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 (HEMONTO). Tese (doutorado) – Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, Belo Horizonte, 2015. Disponível em: <http://hdl.handle.net/1843/BUBD-A35H3K>. Acesso em: 21 maio 2023.

MENDONÇA, F. M.; ALMEIDA, M. B. **OntoForInfoScience**: uma metodologia detalhada para construção de ontologias e sua aplicação no domínio da biomedicina. 2015, XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB). Disponível em: <https://brapci.inf.br/index.php/res/v/187104>. Acesso em: 19 maio 2023.

MERRIAM-WEBSTER DICTIONARY. **Definition of Ontology**. Disponível em: <https://www.merriam-webster.com/dictionary/ontology>. Acesso em: 20 jun. 2022.

MONTEIRO, F. S. **Web semântica e repositórios digitais educacionais na área de saúde**: uma modelagem com foco no objetivo de aprendizagem para refinar resultados de busca. 2013. Doutorado - Universidade de Brasília. Disponível em: <https://repositorio.unb.br/handle/10482/13461>. Acesso em: 19 jun. 2022.

MUSEN, M. A. **The Protégé project**: a look back and a look forward. *AI Matters*. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015a. DOI: 10.1145/2557001.25757003. Disponível em: <https://protege.stanford.edu/>. Acesso em: 27 de nov. de 2021a.

MUSEN, M. A. **The Protégé project**: a look back and a look forward. *AI Matters*. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015b. DOI: 10.1145/2557001.25757003. Disponível em: <https://protege.stanford.edu/about.php>. Acesso em: 27 nov. 2021b.

NATASHA, F. N.; YUQING, G.; ANSHU, J.; ANANT, N.; ALAN, P.; JAMIE T. **Industry-scale Knowledge Graphs: lessons and challenges**. 2019. ACM Queue 17, 2, p.20. Disponível em: <https://queue.acm.org/detail.cfm?id=3332266>. Acesso em: 23 jun. 2022.

NEO4J GRAPH DATA PLATFORM. **Graph database management system**. 2023. Disponível em: <https://neo4j.com/>. Acesso em: 27 abr. 2023.

NETO, A. J. R.; BORGES, M. M.; ROQUE, L. **A preliminary study of proof of concept practices and their connection with Information Systems and Information Science**. New York, NY, USA: Association for Computing Machinery. Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality. 2018. p. 270–275. Disponível em: <https://doi.org/10.1145/3284179.3284226>. Acesso em: 31 jul. 2021.

NOGUEIRA, M. **Quando publicar é também resistir**. Revista de Estudos em Relações Interétnicas| Interethnica, v. 23, n. 1, p. 1-6, 2022. Disponível em: <https://periodicos.unb.br/index.php/interethnica/article/download/42952/32984>. Acesso em: 25 maio 2023.

NOTEPAD++. **What is Notepad++**. Disponível em: <https://notepad-plus-plus.org/>. Acesso em: 27 abr. 2023.

NOVAK, J. D.; CAÑAS, A. J. **A teoria subjacente aos mapas conceituais e como elaborá-los e usá-los**. Tradução de Luis Fernando Cerri (PPGE/UEPG), com revisão técnica de Fabiano Morais. Práxis Educativa, Ponta Grossa, v.5, n.1, p. 9-29, jan. -jun. 2010. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/3251296.pdf>. Acesso em: 23 dez. 2021.

NOVAK, J. D; CAÑAS, A. J, **The theory underlying concept maps and how to construct and use them, technical report IHMC CmapTools**. 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008. Disponível em: <http://cmap.ihmc.us/docs/pdf/TheoryUnderlyingConceptMaps.pdf>. Acesso em: 31 mar. 2022.

NOVELLINO, M. S. F. **Instrumentos e metodologias de representação da informação**. Informação & Informação, v. 1, n. 2, p. 37, 1996. Disponível em https://www.researchgate.net/profile/Maria-Salet-Novellino/publication/276231427_Instrumentos_e_metodologias_de_representacao_da_informacao/links/564a39c408ae127ff9868b0a/Instrumentos-e-metodologias-de-representacao-da-informacao.pdf. Acesso em: 23 maio. 2023.

OBEID, C. et al. **Ontology-based recommender system in Higher Education**. In: Companion of the Web Conference, Lyon, France: ACM Press, 2018. p. 1031–1034. Disponível em: <http://dl.acm.org/citation.cfm?doid=3184558.3191533>. Acesso em: 1 abr. 2023.

OLIVEIRA, E.; NUNO, C. D. **Engenharia de Software**. Apresentação da Unidade 1: Engenharia de Software, do curso de Pós-Graduação Lato Sensu da Universidade Cândido Mendes, 2020. Disponível em

https://drive.google.com/file/d/1XeIjBlzrV63VrpvQHGNEjLoOAcsQhVJm/view?usp=share_link. Acesso em: 29 abr. 2023.

OPEN DATA BAROMETER. **The Open Data Barometer**. 2008. Disponível em: https://opendatabarometer.org/?_year=2017&indicator=ODB. Acesso em: 31 mar. 2022.

OPEN KNOWLEDGE FOUNDATION. **Global Open Data Index**. 2016. Disponível em: <http://index.okfn.org/>. Acesso em: 31 mar. 2022.

OPENREFINE. **A free, open source, powerful tool for working with messy data**. Disponível em: <https://openrefine.org/>. Acesso em: 27 de nov. de 2021.

ONTOTEXT GRAPHDB. **A fully featured RDF database for massive data and moderate query loads**. Disponível em: <https://www.ontotext.com/products/graphdb/graphdb-free/>. 2022a. Acesso em: 27 de maio de 2022.

ONTOTEXT GRAPHDB. **What is a Knowledge Graph?** 2022b. Disponível em: <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>. Acesso em: 27 de maio de 2022.

ONTOTEXT GRAPHDB. **What are Linked Data and Linked Open Data?** 2022c. Disponível em: <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>. Acesso em: 27 de maio de 2022.

ONTOTEXT REFINE. **Ontotext Refine: turn strings into things**. 2022. Disponível em: <https://www.ontotext.com/products/ontotext-refine/>. Acesso em: 27 abr. 2023.

ONTOUML. **OntoUML**. 2017. Disponível em: <https://ontouml.org/ontouml/>. Acesso em: 3 maio 2023.

OWL. **Web Ontology Language (OWL)**. 2012. Disponível em: <https://www.w3.org/OWL/>. Acesso em: 27 abr. 2023.

PAGANINE, L. N. et al. **Padrões de Interoperabilidade para repositórios de dados de pesquisa**. Rio de Janeiro, RJ: Ibict, 2020. E-book. Disponível em: <http://livroaberto.ibict.br/handle/123456789/1085>. Acesso em: 23 mar. 2022.

PEARLMAN, J. et al. **Evolving and sustaining Ocean Best Practices to enable interoperability in the UN decade of Ocean Science for sustainable development**. *Frontiers in Marine Science*, v. 8, p. 619685, 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fmars.2021.619685/full>. Acesso em: 29 mar. 2022.

PELIZZARI, A.; KRIEGL, M. L.; BARON, M. P.; FINCK, N. T. L.; DOROCINSKI, S. I. **Teoria da Aprendizagem Significativa segundo Ausubel**. *Revista PEC*, Curitiba, v. 2, n. 1, p. 37-42, 2002. Disponível em: https://rfp.sesc.com.br/moodle/pluginfile.php/2423/mod_resource/content/1/Teoria_a_aprendizagem_significativa.pdf. Acesso em: 30 ago. 2022.

PENTEADO, B. E. **Modelo de infraestrutura para publicação de dados abertos governamentais conectados de qualidade**. Tese (doutorado) - Universidade de São Paulo, São Carlos. Disponível em: <https://www.teses.usp.br/teses/disponiveis/55/55134/tde-14092020-175138/>. Acesso em: 20 mar. 2022.

PINTO, H S.; STAAB, S.; TEMPICH, C. **DILIGENT: Towards a fine-grained methodology for distributed, loosely-controlled and evolving engineering of ontologies**. European Conference on Artificial Intelligence, 16., 2004, Valencia, Spain. Proceedings. Amsterdam: IOS Press, 2004. P. 393-397. Disponível em: https://aifb.kit.edu/images/3/30/2004_690_Pinto_DILIGENT_Towar_1.pdf. Acesso em: 27 abr. 2023.

PORTABLE DOCUMENT FORMAT. **Entender os diferentes tipos de PDF forms e documentos/ adobe experience manager**. Disponível em: <https://experienceleague.adobe.com/docs/experience-manager-learn/forms/document-services/pdf-forms-and-documents.html?lang=pt-BR>. Acesso em: 12 fev. 2023.

POWELL, K. **Scientists are becoming increasingly frustrated by how long it seems to take to publish papers. But is it really getting worse? The waiting game**. In: Nature. V. 530, fev. 2016. Disponível em: <https://www.nature.com/articles/530148a>. Acesso em: 25 maio 2023.

PRIBERAM. **Interoperabilidade**. 2008 Disponível em: <https://dicionario.priberam.org/interoperabilidade>. Acesso em: 22 mar. 2022.

RABELLO, R. **Documento e institucionalidades**: dimensões epistemológica e política. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 23, n. 51, p. 138–156, 1 jan. 2018.

ROCHA, R. P.; RATHKE, S. B. **Sistema de informação de pesquisa**: uso da ontologia de VIVO no contexto das instituições brasileiras. Brazilian Journal of Information Science: research trends, v. 13, n. 4, p. 132–151, 2019. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/9275>. Acesso em: 31 ago. 2022.

RAUTENBERG, S. **Web semântica e bibliografia**: um estudo para a publicação de registros bibliográficos como dados abertos conectados. Em Questão, v. 25, p. 313–330, 2019. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/92147>. Acesso em: 26 ago. 2022.

RAZA, Z.; MAHMOOD, K.; WARRAICH, N. F. **Application of Linked Data technologies in digital libraries**: a review of literature. Library Hi Tech News, v. 36, n. 3, p. 9–12, 2019.

REFWORKS. **RefWorks**. 2009. Disponível em: <https://www.refworks.com/individual/index.html>. Acesso em: 26 abr. 2023.

REGINATO, C. C; SALAMON, J. S; NOGUEIRA, G. G; BARCELOS, M. P; SOUZA, VÉS; MONTEIRO, M. **GO-FOR**: uma estrutura orientada a objetivos para reutilização

de ontologias. Proceedings of the 20th IEEE International Conference on Information Reuse and Integration for Data Science Los Angeles, CA, USA: IEEE, 2019. Disponível em: https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/go_for__a_goal_oriented_framework_for_ontology_reuse_2019.pdf. Acesso em: 30 abr. 2023.

RESEARCH DATA SHARING WITHOUT BARRIERS. **Research data alliance**. 2013. Disponível em: <https://rd-alliance.org/>. Acesso em: 31 mar. 2022.

REVEZ, J. M. R.; SILVA, C. G. **Organização da Informação e a agenda 2030: a interoperabilidade e o desenvolvimento sustentável**. Ponto de Acesso, v. 15, n. 3, 2021. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/47474>. Acesso em: 8 jun. 2022.

RIVA P; , BOEUF, P. L.; ŽUMER M. et al. **IFLA library reference model: um modelo conceitual para a informação bibliográfica**. 2017. Disponível em: https://repository.ifla.org/bitstream/123456789/47/1/ifla-lrm-august-2017_rev201712-por.pdf. Acesso em: 11 maio 2023.

RML.IO. **Easily generate high-quality knowledge graphs with RML.io**. 2022. Disponível em: <https://rml.io/>. Acesso em: 27 abr. 2023.

ROZSA, V. DUTRA, M. L.; PINTO, A. L.; TORRADO, E. M. **O paradigma tecnológico da internet das coisas e sua relação com a Ciência da Informação**. Informação e Sociedade: Estudos, João Pessoa, v. 27, n. 3, 2017a. Disponível em: <https://periodicos.ufpb.br/index.php/ies/article/view/36975>. Acesso em: 17 jun. 2022.

ROZSA, V.; DUTRA, M. L.; NHACUONGUE, J. A. **Linked Open Data no contexto acadêmico: identificação e análise de vocabulários utilizados na academia e na pesquisa científica**. Brazilian Journal of Information Science: research trends, v. 11, n. 3, 2017b. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/6780>. Acesso em: 31 ago. 2022.

RODRIGUES, M. R.; CERVANTES B. M. N. **Os mapas conceituais e as múltiplas aplicações para a organização e representação do conhecimento**. Informatio. Revista del Instituto de Información de la Facultad de Información y Comunicación, vol. 22, n. 2, 2017, p. 101–21, Disponível em: <https://informatio.fic.edu.uy/index.php/informatio/article/view/199>. Acesso em: 31 mar. 2022.

SALAMON, J. S. **Uma abordagem orientada a objetivos para desenvolvimento de ontologias baseado em Integração**. 2018. Dissertação (mestrado) - Universidade Federal do Espírito Santo, Vitória, ES. Disponível em: <http://www.inf.ufes.br/~jssalomon/wp-content/uploads/DissertacaoJordanaversaoFinal.pdf>. Acesso em: 30 abr. 2023.

SAWAYA, M. R. **Dicionário de informática e Internet: inglês-português**. São Paulo, SP: Nobel : Centro Estadual de Educação Tecnológica Paula Souza, 1999.

Disponível em: <https://nosda18.files.wordpress.com/2009/04/dicionariode-informatica-e-internet.pdf>. Acesso em: 11 fev. 2023.

SCHNEIDER, E. W. **Course modularization applied**: the interface system and its implications for sequence control and data analysis. (436252004-001). American Psychological Association, , 1973. Disponível em: <http://doi.apa.org/get-pe-doi.cfm?doi=10.1037/e436252004-001>. Acesso em: 23 jun. 2022.

SEGUNDO, J. E.; S.; CONEGLIAN, C. S.; LUCAS, E. R. O. **Conceitos e tecnologias da Web semântica no contexto da colaboração acadêmico-científica**: um estudo da plataforma Vivo. *Transinformação*, v. 29, n. 3, p. 297–309, 2017. Disponível em: <https://www.scielo.br/j/tinf/a/VpcTfXFWvN5bsQ9LqvK3mYw/?lang=pt>. Acesso em: 01 jun. 2022.

SILVA, C. J. L.; FREIRE, G. H. A. **Um olhar sobre a origem da Ciência da Informação**: indícios embrionários para sua caracterização identitária. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Paraíba, v. 17, n. 33, p. 1–29, 2012.

SILVA, C. A. ; LARA, M. L. G. **Esquema básico de metadados para representação descritiva de obras de arte em museus brasileiros**. *Transinformação*, v. 33, 2021. Disponível em: <https://www.scielo.br/j/tinf/a/DTLyDN7trqnwFchLcLcBmQg/>. Acesso em: 9 maio 2023.

SILVA E. M.; MUTZ F. W.; RUY F. B. **Uso de ontologias no suporte a aplicação de machine learning**: um caso no domínio de evasão escolar. *Ontobras*, 2022a, p. 14. Disponível em: <https://www.inf.ufrgs.br/ontobras/trabalhos-aceitos-2022/>. Acesso em: 29 jun. 2022.

SILVA E. M.; MUTZ F. W.; RUY F. B. **Abordagem para análise de múltiplas fontes de dados de evasão escolar**. *Anais do XIII Computer on the Beach - COTB'22*. Itajaí - Santa Catarina - Brasil: Universidade do Vale do Itajaí, 2022. p. 149–156. Disponível em: <https://periodicos.univali.br/index.php/acotb/article/view/18791>. Acesso em: 30 maio 2023.

SILVA, M. F.; MARTINS, D. L.; SIQUEIRA, J. **Web semântica em repositórios**: ontologia para representação de bibliotecas digitais. *Ciência da Informação em Revista*, v. 6, n. 1, p. 99, 2019. Disponível em: <https://brapci.inf.br/index.php/res/v/113732>. Acesso em: 28 abr. 2023.

SILVA, J. L. C. **Múltiplas interlocuções da informação no campo da Ciência da Informação no âmbito dos fundamentos técnico-pragmáticos, humanos e científicos**. Universidade Federal da Bahia Instituto de Ciência da Informação. Programa de Pós-Graduação em Ciência da Informação. p. 491, 2014.

SILVEIRA, P. D. N.; ARAGÓN, R.; CURY, D.; MENEZES, C. S. **Uma ontologia de referência para arquiteturas pedagógicas**. Em: XXXII Simpósio Brasileiro de Informática na Educação, 32. 2019, Brasília, DF. *Anais [...]*. Brasília, DF:

Universidade de Brasília/DF, 2021. p.24. Disponível em:
<https://sol.sbc.org.br/index.php/sbie/article/view/18040>. Acesso em: 7 abr. 2023.

SILVEIRA, P. N.; CURY, D.; MENEZES, C. D. **Superando fronteiras da educação com ecossistemas de aprendizagem**. Em: XXX Simpósio Brasileiro de Informática na Educação, 30. 2019, Brasília, DF. Anais [...]. Brasília, DF: Universidade de Brasília/DF, 2019. p.209. Disponível em:
https://www.researchgate.net/profile/Credine-Menezes/publication/337412090_Superando_frenteiras_da_educacao_com_ecossistemas_de_aprendizagem/links/61f006568d338833e39455be/Superando-fronteiras-da-educacao-com-ecossistemas-de-aprendizagem.pdf. Acesso em: 5 abr. 2023.

SLAVIC, A. **Knowledge Organization Systems, network standards and semantic Web**. Informacijske Znanosti u Procesu Promjena. Zagreb: Zavod za informacijske studije, 2005. p. 5-22. Disponível em:
<https://repository.arizona.edu/handle/10150/106470?show=full>. Acesso em: 31 jul. 2021.

SMIT, J. W.; KOBASHI, N. Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Arquivo do Estado, Imprensa Oficial, 2003. E-book. Disponível em:
http://www.arquivoestado.sp.gov.br/site/assets/publicacao/anexo/como_elaborar_vocabulario_controlado_para_aplicacao_em_arquivos.pdf. Acesso em: 1 nov. 2020.

SMITH, B. **Ontology and Information Systems**. 2004, p. 93. Disponível em:
[http://ontology.buffalo.edu/ontology\(PIC\).pdf](http://ontology.buffalo.edu/ontology(PIC).pdf). Acesso em: 31 jun. 2022.

SMITH, B. et al. **Basic Formal Ontology 2.0**: specification and user's Guide. 2015. Disponível em: <https://raw.githubusercontent.com/BFO-ontology/BFO/master/docs/bfo2-reference/BFO2-Reference.pdf>. Acesso em: 27 abr. 2023.

SOMMERVILLE, I. **Engenharia de software**. 9. ed. São Paulo: Pearson Prentice Hall, 2011. Disponível em: <https://www.facom.ufu.br/~william/Disciplinas%202018-2/BSI-GSI030-EngenhariaSoftware/Livro/engenhariaSoftwareSommerville.pdf>. Acesso em: 29 abr. 2023.

STAKE, R.E. (1998). **Case Studies**. In N.K. Denzin e Y.S. Lincoln (eds.) Handbook of qualitative research, Thousands Oaks, Sage.

SUÁREZ-FIGUEROA, M. C. *et al.* **Ontology engineering in a networked world**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. E-book. Disponível em:
<https://link.springer.com/10.1007/978-3-642-24794-1>. Acesso em: 29 abr. 2023.

SUÁREZ-FIGUEROA, M. C. **NeOn methodology for building ontology networks: specification, scheduling and reuse**. 2010. PhD Tese (doutorado) - Universidad Politécnica de Madrid, Madrid. Disponível em: <http://oa.upm.es/3879/>. Acesso em: 1 maio 2023.

SUBIRATS, I.; SOLODOVNIK, I.; BUDRONI, P.; GANGULY, R.; HUDAK, R. **Metadata and controlled vocabularies alignment in PHAIDRA international**

persistent ecosystem. Information Services and Use, v. 37, n. 1, p. 25–32, 2017. Disponível em: <https://web-s-ebscohost.ez43.periodicos.capes.gov.br/ehost/pdfviewer/pdfviewer?vid=0&sid=075c0bf5-a25a-4913-af57-47b44308cadc%40redis>. Acesso em: 30 ago. 2022.

SURE, Y.; STAAB, S.; STUDER, R. **On-to-knowledge methodology (OTKM).** Handbook on ontologies. Berlin: Springer Berlin Heidelberg, 2004. Cap. 6. P. 117-132. Disponível em: https://link.springer.com/chapter/10.1007/978-3-540-24750-0_6. Acesso em: 27 abr. 2023.

TANENBAUM, A. S; BOS, H. **Sistemas Operacionais Modernos.** 4. ed. São Paulo: Pearson, 2016. Disponível em: http://ldemetrio.com.br/Livros/Livros_TI/segunda_unid/Sistemas%20Operacionais%20Modernos%20-%20Tanenbaum%20-%204%20Edi%C3%A7%C3%A3o.pdf. Acesso em: 11 jun. 2023.

TEIXEIRA, M. G. S. **An ontology-based process for domain-specific visual language design.** 2017. Tese (doutorado) - Universidade Federal do Espírito Santo e na Ghent University em Ciências Econômicas Aplicadas, Vitória, ES. Disponível em: <https://repositorio.ufes.br/handle/10/9873>. Acesso em: 20 jan. 2023.

TEIXEIRA, M. G. S. FALBO, R. A, GUIZZARDI, G. **Analyzing the behavior of modelers in interpreting relationships in conceptual models:** an empirical study. Onto.com / ODISE 2014 - 1st Joint Workshop on Ontologies on Conceptual Modeling and Information Systems Engineering, 2014, Rio de Janeiro - RJ, Brasil. Disponível em: https://ceur-ws.org/Vol-1301/ontocomodise2014_1.pdf. Acesso em: 30 abr. 2023.

TOUTAIN, L. M. B. B. **Para entender a ciência da informação.** Salvador: EDUFBA, 2007. Disponível em: <https://repositorio.ufba.br/handle/ufba/145>. Acesso em: 17 maio 2022.

UNIVERSIDADE FEDERAL DA FRONTEIRA DO SUL. **Documento digitalizado X nato digital.** Disponível em: <https://portalsei.uffs.edu.br/apresentacao/o-que-e-um-documento-nato-digital>. Acesso em: 11 fev. 2023.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, PPGCI. **Regimento do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo.** 2023a. Disponível em: https://cienciainformacao.ufes.br/sites/cienciainformacao.ufes.br/files/field/anexo/regimento_interno_ppgci_ufes.pdf. Acesso em: 28 abr. 2023.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, HISTÓRICO. **Regimento do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Espírito Santo.** 2023b. Disponível em: <https://cienciainformacao.ufes.br/pt-br/pos-graduacao/PPGCI/hist%C3%B3rico>. Acesso em: 4 maio 2023.

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO, NORMAS ABNT. **Normalização e apresentação de trabalhos científicos e acadêmicos.** 2015. Disponível em: <https://repositorio.ufes.br/bitstream/10/1533/1/Normalizacao%20e%20apresentacao>

%20de%20trabalhos%20cientificos%20e%20academicos.pdf. Acesso em: 27 out. 2022.

USHOLD M.; MENZEL. C. **Semantic integration and interoperability using RDF and OWL**. 2005. Disponível em: <https://www.w3.org/2001/sw/BestPractices/OEP/SemInt/>. Acesso em: 22 jul. 2023.

VALENTIM, M. **Gestão, mediação e uso da informação**. São Paulo: Cultura Acadêmica, 2010. Disponível em: <https://books.scielo.org/id/j4gkh>. Acesso em: 8 maio 2022.

VICKERY, B. C. **Ontologies**. *Journal of Information Science*, v. 23, n. 4, p. 277–286, 1997. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/016555159702300402?journalCode=jisb>. Acesso em: 8 mar. 2022.

VISUAL-PARADIGM. **Ideal Modeling and Diagramming Tool for Agile Team Collaboration**. Disponível em: <https://www.visual-paradigm.com/>. Acesso em: 21 jun. 2022.

VISUAL-PARADIGM. **Introduction of visual paradigm**. em: visual paradigm community circle. 3 mar. 2020. Disponível em: <https://circle.visual-paradigm.com/docs/introduction-and-system-requirements/introduction-of-visual-paradigm/>. Acesso em: 6 jun. 2023.

VOCBENCH. **VocBench**. Disponível em: <http://vocbench.uniroma2.it/>. Acesso em: 28 maio 2022.

W3C. **Semantic Web**: vocabularies. 2022. Disponível em: <https://www.w3.org/standards/semanticweb/ontology>. Acesso em: 25 jun. 2022.

W3C. **World Wide Web consortium escritório Brasil**. 2011. Disponível em: <https://www.w3c.br/Padroes/WebSemantica>. Acesso em: 22 maio 2023.

WAZLAWICK, R. S. **Análise e projeto de sistemas de informação orientados a objetos**. Rio de Janeiro: Elsevier Editora Ltda, 2011. Disponível em: https://kupdf.net/download/raul-wazlawick-analise-e-projeto-de-sistemas-de-informacao-orientados-campus-2011_598a1806dc0d602e32300d17_pdf. Acesso em: 28 maio 2023.

WILKINSON, M. D. et al. **The FAIR guiding principles for scientific data management and stewardship**. *Scientific Data*, 2016, v. 3, n. 1, p. 160018.

WOOD, D; LANTHALER, M.; CYGANIAK, R. **RDF: concepts and abstract syntax**. 2014. Disponível em: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>. Acesso em: 10 jun. 2022.

YANG, L. et al. **Dynamic heterogeneous graph embedding using hierarchical attentions**. Em: JOSE, J. M. et al. (org.). *Advances in Information Retrieval*. Cham: Springer International Publishing, 2020. (Lecture Notes in Computer Science). v.

12036, p. 425–432. E-book. Disponível em: http://link.springer.com/10.1007/978-3-030-45442-5_53. Acesso em: 22 jun. 2022.

YAMAKAWA, E. K. et al. **Comparativo dos softwares de gerenciamento de referências bibliográficas: Mendeley, EndNote e Zotero**. *Transinformação*, v. 26, n. 2, p. 167–176, 2014. Disponível em: <https://www.scielo.br/j/tinf/a/YCXRcdqjP6gGccddwgzwnCK/?lang=pt>. Acesso em: 24 abr. 2023.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 3. ed., Porto Alegre: Bookman, 2005. Disponível em: <https://ria.ufrn.br/jspui/handle/123456789/943>. Acesso em: 06 maio 2023.

ZAMBORLINI, V. C. **Estudo de alternativas de mapeamento de ontologias da linguagem OntoUML para OWL: abordagens para Representação de informação temporal**. 2011. Dissertação (Mestrado) - Universidade Federal do Espírito Santo, Vitória, ES, 2011. Disponível em: https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/estudo_de_alternativas_de_mapeamento_de_ontologias_da_linguagem_ontouml_para_owl__abordagens_para_representacao_de_informacao_temporal_2011.pdf. Acesso em: 22 maio 2022.

ZENG, M. L. **Construction of controlled vocabularies, a primer**. 2005. Disponível em: <https://marciazeng.slis.kent.edu/Z3919/index.htm>. Acesso em: 22 maio 2022.

ZENG, M. **Interoperability**. ISKO Encyclopedia of Knowledge Organization. 2019. Disponível em: <https://www.isko.org/cyclo/interoperability>. Acesso em: 08 jun. 2022.

ZENG, M. L., & CHAN, L. M. **Trends and issues in establishing interoperability among Knowledge Organization Systems**. 2004. *Journal of the American Society for Information Science and Technology*, 55(5), 377–395. Disponível em: <https://dl.acm.org/doi/10.1002/asi.10387>. Acesso em: 22 maio 2023.

ZHENG, Y.; TÖRMÄ, S.; SEPPÄNEN, O. **A shared ontology suite for digital construction workflow**. *Automation in Construction*. 2021. v. 132, p. 103930. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0926580521003812>. Acesso em: 12 set. 2022.

ZOTERO. **Why Zotero?** Disponível em: <https://www.zotero.org/why>. 2023a. Acesso em: 30 abr. 2023.

ZOTERO. **Style repository**. 2023b. Disponível em: <https://www.zotero.org/styles/>. Acesso em: 30 abr. 2023.

ZOTFILE. **Advanced PDF management for Zotero**. Disponível em: <http://zotfile.com/>. Acesso em: 30 abr. 2023.

APÊNDICES

APÊNDICE A – ENTREVISTA EM 13/12/2022 USADA NA ELABORAÇÃO DA ONTOLOGIA DE REFERÊNCIA DE DOMÍNIO



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Entrevista: validação da ontologia de referência de domínio da Defesa, da Colaboração, Bolsa de Estudos e Área de Conhecimento do Trabalho Acadêmico.


No dia 13 do mês de dezembro do ano de 2022, às 17h00min, via plataforma Google Meet, foi realizada uma entrevista visando validar a ontologia desenvolvida pelo mestrando **Rafael Denerson Ramos de Sousa** com a orientação dos professores Henrique Monteiro Cristovão e Maria das Graças Silva Teixeira, matriculado sob o número 2021131508, nos contextos educacionais de defesa, colaboração, bolsa de estudos e área de conhecimento do trabalho acadêmico. As entrevistadoras são as professoras doutoras Lucileide Andrade de Lima do Nascimento e Rosa da Penha Ferreira da Costa, consideraram que a ontologia de referência de domínio seja alterada conforme a seguinte descrição dos conceitos nos seguintes contextos:

Defesa: verificar a possibilidade de inserir a propriedade matrícula ou siape para o docente (representado pela entidade Pessoa Institucionalizada, Membro Interno, Orientador e/ou Coorientador da ontologia de referência de domínio). Na entidade Membro, verificar a possibilidade de inserir uma titulação mínima de doutor (conforme definição contida no programa stricto sensu). Inserir a entidade Rede de Ensino (porque um curso pode ser oferecido por Instituição de Ensino Superior ou por Rede de Ensino) e Pré-Requisito para Aprovação (alegando que existe pré-requisitos para a defesa acontecer). Eliminar a entidade Pós-Graduação vinculada ao curso, pois está redundante uma vez que existe a entidade Programa de Pós-Graduação pertencente a Organização de Ensino. Verificar a representação de curso de mestrado e doutorado na modalidade profissional. Sugeriram que uma defesa gera mais de um Documento de Aprovação (como ata, por exemplo).

Colaboração: Sugeriu-se que a cardinalidade não tem um máximo para a quantidade de Palavra-Chave, pois cada programa de mestrado ou doutorado. Recomendou informar na entidade Documento Impresso que a partir de 2020 os documentos passaram a não ser mais impressos de acordo com o decreto nº 10.278 (regulamento geral da pós-graduação da UFES).

Bolsa de estudos: Sugeriu-se alterar a nomenclatura da entidade Voluntário para aluno Não Bolsista quando relacionada ao aluno.

A entrevista foi encerrada às 19h36min.

Documento assinado digitalmente
 RAFAEL DENERSON RAMOS DE SOUSA
Data: 15/05/2023 19:07:27-0300
Verifique em <https://validar.iti.gov.br>

Rafael Denerson Ramos de Sousa

Mestrando

APÊNDICE B – ENTREVISTA EM 20/12/2022 USADA NA ELABORAÇÃO DA ONTOLOGIA DE REFERÊNCIA DE DOMÍNIO



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO CENTRO DE CIÊNCIAS JURÍDICAS E ECONÔMICAS PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Entrevista: validação da ontologia de referência de domínio da Defesa, da Colaboração, Bolsa de Estudos e Área de Conhecimento do Trabalho Acadêmico.

No dia 20 do mês de dezembro do ano de 2022, às 14h30min, presencialmente na Universidade Federal do Espírito Santo (UFES), sala 807 da coordenação do Programa de Pós-Graduação em Política Social (PPGPS), foi realizada uma entrevista visando validar a ontologia desenvolvida pelo mestrando **Rafael Denerson Ramos de Sousa** com a orientação dos professores Henrique Monteiro Cristovão e Maria das Graças Silva Teixeira, matriculado sob o número 2021131508, nos contextos educacionais de defesa, colaboração, bolsa de estudos e área de conhecimento do trabalho acadêmico. A entrevistada foi a professora doutora Maria Lucia Teixeira Garcia, considerou que a ontologia de referência de domínio pudesse ser alterada conforme a seguinte descrição dos conceitos nos seguintes contextos:

Defesa: Contextualizou dizendo que no âmbito da defesa de dissertação ou tese, existe um processo que perpassa pela aprovação do orientador assinalando se a dissertação ou tese há condições de defesa, em seguida escolhe-se a banca, faz-se a defesa na data agendada. Na defesa existe um ritual que varia conforme o regimento do programa. O discente egresso (candidato) tem em torno de trinta a cinquenta minutos para fazer uma síntese (conforme regimento de cada programa de pós-graduação stricto sensu), faz-se a arguição por cada membro da banca, e ao final a banca conclui-se pela aprovação ou não aprovação da dissertação ou tese. Esta aprovação pode estar condicionada a correções ou indicações, pois a defesa pode não ser o ponto final (isso depende se o regimento de programa de pós-graduação stricto sensu admite a aprovação condicionada à correção ou indicações). Assim


como existem programas que permitem a aprovação condicionada a correções, existem também programas que só admitem a conclusão da dissertação ou tese como aprovada ou não aprovada.

Há diferença na composição dos membros da banca de dissertação e tese. Na dissertação são três: o orientador e mais dois membros (um interno e um externo que realizam arguição). Na tese existem cinco membros: o orientador e mais quatro membros (dois externos e dois internos que fazem arguição). Na arguição o nível de complexidade é correspondente ao nível de titulação que o discente egresso (candidato) está se titulando (mestre ou doutor). Nesse sentido, foi recomendado que a entidade Membro pudesse expressar a cardinalidade de maneira correta para tese e dissertação, correspondendo no mínimo 3 para dissertação e no mínimo 5 para tese. Sugeriu-se também substituir o termo Aluno Acadêmico ou (Profissional Acadêmico) por Discente Egresso (ou discente concluinte conforme termos padronizados pela Capes), pois acadêmico é um termo ambíguo e os documentos da Capes abordam o termo discente para se referir ao aluno, bem como a terminologia egresso para relacionar-se ao aluno que está de saída da Instituição (estado de transição entre aluno regular e aluno egresso). Aconselhou usar termos padronizados e legitimados nacionalmente pela Capes, pois ela é quem regulamenta a pós-graduação no Brasil desde 1951 quando foi criada, e a portaria 81 define os tipos de docente, dentre outros termos. A defesa pode ter dois momentos: apresentação e arguição (é preciso que o aluno demonstre consistência e coerência nas ideias), embora existam programas que possuem apenas a arguição (isso é definido no regimento do programa). Argumentou que Membro(s) faz(em) arguição da Defesa, enquanto o aluno egresso apresenta a defesa respondendo às arguições dos membros da banca.

Quanto à defesa no contexto de Colaboração e Bolsa de Estudos recomendou-se usar os documentos normativos e portarias da Capes.

Área de conhecimento do trabalho acadêmico: O documento das áreas de conhecimento do CNPq não sofreu certas adequações de forma a serem aproveitadas no regimento dos programas de pós-graduação mostrando-se muitas vezes obsoletas, enquanto a classificação das áreas de conhecimento da Capes aborda as definições que fazem parte do regimento dos programas de pós-graduação que são definidos no Brasil.

A entrevista foi encerrada às 15h50min.

Documento assinado digitalmente
 **RAFAEL DENERSON RAMOS DE SOUSA**
Data: 15/05/2023 19:09:21-0300
Verifique em <https://validar.it.gov.br>

Rafael Denerson Ramos de Sousa

Mestrando

APÊNDICE C – GLOSSÁRIO DE TERMOS DA ONTOLOGIA DE DOMÍNIO

O Quadro 11 contém termos com sua descrição, usados no *knowledge graph* da preparação conceitual da ontologia de domínio, na ontologia de referência de domínio, e na ontologia operacional de domínio.

Quadro 11 - Glossário de termos da ontologia de domínio

Termo	Descrição	Referência
Acadêmico	Refere-se ao membro de uma academia ou um aluno (estudante).	Priberam (2022)
Acervo	Local que contém todos os documentos depositados em suas bibliotecas e que estão à disposição do público para estudo e pesquisa, na dependência específica de reprodução.	Brasil (2014)
Agência de Fomento	Diz respeito à instituição ou órgão de natureza pública ou privada que possua entre as suas finalidades o financiamento de ações que visem a promover e estimular o desenvolvimento da ciência, tecnologia e inovação.	Brasil (2004a)
Área	Trata-se do 2º nível da área de conhecimento do CNPq, ou seja, é o conjunto de conhecimentos interrelacionados, construídos coletivamente, reunidos segundo a natureza do tema de investigação com objetivos de ensino, pesquisa e aplicações experimentais.	Fundação (2022)
Área do Conhecimento	Grupo de informações e referenciais metódicos, agrupados segundo um ponto de vista, que sustenta relações de complementaridade recíproca. No Brasil, uma classificação difundida e que tem sido utilizada para a pesquisa e a pós-graduação, é a classificação de Áreas do Conhecimento do CNPq. Sua classificação original apresentou uma hierarquização em	Glossário de Termos Referentes ao Ensino de Graduação do CEFET-MG (2015), Fundação Coordenação de

	quatro níveis. Trata-se das áreas do conhecimento indicadas pelo CNPq.	Aperfeiçoamento de Pessoal de Nível Superior (2023), Conselho Nacional de Desenvolvimento Científico e Tecnológico (2023) e especialista do domínio
Área Primária	Refere-se à primeira área de concentração de uma linha de pesquisa.	Especialista do domínio
Área Secundária	Refere-se à segunda área de concentração de uma linha de pesquisa.	Especialista do domínio
Área Terciária	Refere-se à terceira área de concentração de uma linha de pesquisa.	Especialista do domínio
Ata	Trata-se do registo documentado de sessão de colectividades deliberativas.	Priberam (2022)
Banca	Trata-se da composição de membros (no mínimo três titulares e um suplente) para avaliar a defesa de um estudante de pós-graduação em uma Instituição de Ensino Superior.	Universidade (2023)
Bolsa de Estudos	Pagamento mensal concedido ao aluno de curso superior para que ele atue em função remunerada.	Glossário de Termos Referentes ao Ensino de Graduação do CEFET-MG (2015)
Bolsista	É todo aluno que recebe uma Bolsa de Estudos.	Glossário de Termos Referentes ao Ensino de Graduação do CEFET-MG (2015)

Cadastrador	Pessoa pertencente a comunidade acadêmica que cadastra o trabalho acadêmico no acervo.	Glossário de Termos Referentes ao Ensino de Graduação do CEFET-MG (2015)
Cadastro	Trata-se do registro da pessoa pertencente à comunidade acadêmica, da data e hora em que o trabalho acadêmico foi armazenado no acervo.	Glossário de Termos Referentes ao Ensino de Graduação do CEFET-MG (2015)
Classificação da Categoria Acadêmica	Especial, Privada comunitária, Privada confessional, Privada com fins lucrativos, Privada - Particular em sentido estrito, Privada sem fins lucrativos, Pública Estadual, Pública Federal, Pública Municipal,	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2021)
Classificação de Idioma	Diz respeito aos tipos de idioma (linguagem ou língua falada por um povo/nação). A capes usa como requisito de proficiência os idiomas na língua inglesa, francesa, alemã, espanhola e italiana. Entretanto, existem diversos outros tipos de idioma conforme Ethnologue Languages of the World.	Ethnologue Languages of the World (2022) e Fundação (2023)
Classificação da Modalidade de Participação	Diz respeito a modalidade de participação definida como: Presencial, a distância (BRASIL, 2018b) e semi-presencial.	Brasil (2004b) e Brasil (2018b)
Classificação da Organização Acadêmica	Centro Federal de Educação Tecnológica, Centro Universitário, Faculdade, Instituto Federal de Educação, Ciência e Tecnologia, Universidade.	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2021)

Classificação de PDF	Diz respeito a classificação do tipo de arquivo PDF: PDF tradicional, PDF/A para arquivamento, PDF XFA estático e PDF XFA dinâmico (PDF/A-1a, PDF/A-1b, PDF/A-21, PDF/A-2b, PDF/A-2u, PDF/A-3a, PDF/A-3b, PDF/A-3u, PDF/A-4f, PDF/A-4e).	Portable Document Format (2023)
Condição para Aprovação	Trata-se de cada pré-requisito para aprovação no curso stricto sensu.	Especialista de domínio
Contrato	É qualquer acordo entre órgãos ou entidades da Administração Pública e particulares, em que haja um ajuste de vontades para a formação de vínculo e definição de obrigações entre as partes.	Brasil (1993)
Convênio	Acordo firmado por organizações educacionais (instituições educacionais de ensino superior) para ofertar programas ou cursos de pós-graduação. Ele pode ofertar certos produtos (como por exemplo, formação: neste caso, não gere o curso) ou quando conta com professores de outras Instituições.	Especialista de domínio
Convidado	Trata-se do membro que recebe o convite para participar de uma banca de defesa de qualificação, dissertação ou tese.	Brasil (2022b) e Morosini <i>et al.</i> (2006)
Coorientador	É membro da Banca Examinadora que é caracterizado por acompanhar o processo de tutoria, direcionar e apoiar ao aluno orientando em uma série de aspectos pesquisados, que são decisivos na formação dos futuros orientadores e pesquisadores.	Brasil (2022) e Morosini <i>et al.</i> (2006)
Curso Acadêmico	Trata-se de um curso de pós-graduação stricto sensu relacionado a programas de mestrado ou doutorado.	Brasil (2019), Especialista do domínio
Curso de Pós-Graduação	Arranjo de disciplinas e atividades estruturadas em assuntos amplos ou particulares do conhecimento, visando auxiliar aos objetivos educacionais	Brasil (1996), Brasil (2001), Brasil (2002),

	<p>determinados pela instituição e de acordo com as diretrizes curriculares. São aplicados ao programa de pós-graduação, seja gerido por uma única Instituição de Ensino Superior ou por Rede de Ensino. Curso de pós-graduação stricto sensu, compreende programas de mestrado e doutorado que atendem às exigências de autorização, reconhecimento e renovação de reconhecimento previstas na legislação. Os cursos de pós-graduação de mestrado e doutorado profissionais visam contribuir para o incremento da qualificação da prática e experiência profissional, concedendo competências para avaliação crítica, intervenção e resposta a problemas a ela relacionados, bem como a realização de tecnologias aplicadas ao trabalho.</p>	<p>Especialista de domínio, Glossário de Termos Referentes ao Ensino de Graduação do CEFET-MG (2015) e Brasil (2022a)</p>
Curso Profissional	<p>Pretende colaborar para o acréscimo da qualificação da prática profissional, fornecendo competências para avaliação crítica, interferência e resolução de problemas, e também objetivando o desenvolvimento de tecnologias aplicadas ao trabalho.</p>	<p>Brasil (2022a) e Especialista do domínio</p>
Defesa	<p>É o evento de qualificação ou de defesa de dissertação que trata da apresentação de um trabalho acadêmico formal a uma banca examinadora.</p>	<p>Universidade Federal do Espírito Santo (2023a) e Brasil (2022b)</p>
Digitalizado	<p>É a representação digital do processo de digitalização do documento físico e seus metadados, isto é, trata-se de um documento digital no estado digitalizado. Um documento deste tipo tem o mesmo efeito jurídico conferido aos documentos microfilmados, nos termos da Lei nº 5.433, de 8 de maio de 1968, e de regulamentação posterior. São documentos no formato digital e que foram digitalizados (eram</p>	<p>Brasil (2012), Especialista do negócio e Priberam (2022)</p>

	físicos mas passaram a ser documentos digitais).	
Digitalizador	É uma pessoa institucionalizada que digitaliza as teses e as dissertações. Trata-se daquela pessoa que transcreve os dados para a tela do computador.	Especialista do domínio e Sawaya (1999)
Discente	Na Educação Superior o verbete aluno refere-se ao estudante. Trata-se do enfoque no processo de aprendizagem, dos saberes discentes e dos diferentes tipos de estudantes (considerados em suas distinções pessoais e nas circunstâncias da academia).	Morosini <i>et al.</i> (2006) e Especialista do domínio
Discente Egresso	Trata-se da transição do aluno regular para aluno concluinte.	Especialista do domínio.
Dissertação	Dissertação é o resultado de um trabalho de pesquisa desenvolvido, documentado e defendido publicamente em cursos de pós-graduação de mestrado, destinado a obtenção do grau acadêmico de mestre.	Universidade Federal do Espírito Santo (2015)
Documento	Trata-se de um documento digital ou impresso para o contexto aplicado. É a informação registrada, codificada em dígitos binários, acessível e interpretável com suporte de sistema computacional.	Priberam (2022) e Conarq (2020)
Documento de Aprovação	Entidade que representa os requisitos de aprovação que são definidos por um determinado programa de pós-graduação stricto sensu.	Especialista do domínio
Documento Digital	São artefatos digitais. Entende-se por documento digital a conversão da fiel imagem de um documento para o código digital. Trata-se de um documento que pode ser digitalizado (o ato ou o efeito da conversão de um documento que contém informação ou dados analógicos para um formato digital) ou nato digital (documento que já nasce digitalmente).	Brasil (2012), Especialista do domínio e Priberam (2022)

Documento Impresso	Qualquer escrito ou impresso que fornece uma informação ou prova, usado para esclarecimento de algo. Neste caso, trata-se de Trabalhos de Conclusão de Curso como teses e dissertações.	Benício e Silva (2005) e Especialista do domínio
Documento Monográfico	Trata-se de um documento especializado de escrita sobre um único assunto ou aspecto de assunto no âmbito acadêmico. Geralmente é um Trabalho de Conclusão de Curso.	Especialista do domínio
Doutorado	Diz respeito a um curso de pós-graduação <i>stricto sensu</i> visa propiciar formação científica e cultural aprofundada, formando profissionais com a finalidade de desenvolver formas independentes de atividades de ensino, pesquisa e inovação, e também para atuar na formação de profissionais de alta qualificação científica e técnico-profissional (no âmbito de uma área específica de conhecimento).	Brasil (2022a).
Especialidade	Trata-se do 4º nível da área de conhecimento do CNPq. É a caracterização temática da atividade de ensino e pesquisa. Ela pode ser incorporada em distintas grandes áreas, áreas e subáreas.	Fundação (2022)
Examinador	Trata-se de integrante da comissão examinadora de defesa de qualificação, dissertação ou tese. É o membro que recebe o convite para participar de uma banca de defesa de qualificação, dissertação ou tese.	Brasil (2019), Brasil (2020b), Brasil (2022b), e Morosini <i>et al.</i> (2006)
Examinador Externo	Trata-se de integrante da comissão examinadora de defesa de qualificação, dissertação ou tese externos ao programa de pós-graduação e à UFES, isto é, deve ser escolhido, preferencialmente, em outra unidade da federação/Instituição.	Brasil (2019), Brasil (2020b) e Brasil (2022b)

Examinador Interno	Trata-se de um integrante da comissão examinadora de defesa de qualificação, dissertação ou tese interno ao programa de pós-graduação e à UFES.	Brasil (2019), Brasil (2020b) e Brasil (2022b)
Grande Área	Diz respeito ao 1º nível da área de conhecimento do CPNq. Trata-se da aglomeração de diversas áreas do conhecimento devido a afinidade de seus objetos, métodos cognitivos e recursos instrumentais representando contextos sociopolíticos específicos.	Fundação (2022)
Instituição de Ensino Superior	Diz respeito à Instituição que fornece um ensino superior aceito pela autoridade competente e faz parte da integração do sistema de ensino superior.	Morosini <i>et al.</i> (2006)
Local	Trata-se do lugar físico ou virtual de registro do Trabalho de Conclusão de Curso. Quanto ao local físico, refere-se ao nome da cidade de apresentação do trabalho acadêmico. Neste caso, refere-se ao Trabalho de Conclusão de Curso de pós-graduação stricto sensu (dissertação ou tese). Quando se refere ao local virtual, diz respeito ao nome do caminho ou diretório de um computador ou no âmbito Web.	Universidade Federal do Espírito Santo (2015) e Especialista do domínio
Local Físico	Trata-se do local físico (concreto) dos documentos acadêmicos e institucionais.	Universidade Federal do Espírito Santo (2023a) e especialista de domínio
Local Virtual	Trata-se do caminho virtual dos documentos acadêmicos e institucionais. É o caminho no computador ou internet que pode tratar um lugar virtual como se fosse um lugar físico.	Especialista do negócio, Sawaya (1999) e Universidade Federal do Espírito Santo (2023a)

Membro	Integrante da comissão examinadora de defesa de qualificação, dissertação ou tese interno ou externo ao programa de pós-graduação e à Instituição que participa.	Brasil (2019), Brasil (2020b) e Brasil (2022b)
Mestrado	Curso de pós-graduação <i>stricto sensu</i> que visa ampliar e aperfeiçoar a competência didática, científica, cultural e profissional dos indivíduos graduados.	Brasil (2022a)
Não Bolsista	Discente que desenvolve pesquisa de maneira voluntária.	Especialista do domínio.
Nato Digital	São documentos produzidos originalmente em formato digital, isto é, já nascem em formato digital. É elaborado originalmente em meio eletrônico como documento, contrato, diploma, publicação, material, conteúdo, etc.	Academia Brasileira (2022), Brasil (2020a) e Universidade Federal do Rio Grande do Sul (2023).
Orientador	É membro e presidente da Banca Examinadora. Caracteriza-se pelo processo de acompanhamento, tutoria, direcionamento e apoio ao estudante orientando em uma série de aspectos pesquisados, que são decisivos na formação dos futuros orientadores e pesquisadores.	Brasil (2022b) e Morosini <i>et al.</i> (2006)
Organização de Ensino	Ente que trabalha com educação de ensino superior. Ele pode ser uma rede de ensino (quando existe) ou uma instituição de ensino superior isolada (quando a rede de ensino não existe). Esse ente é caracterizado por oferecer e gerir o curso de pós-graduação.	Especialista de domínio
Palavra-Chave	Termo que resume as ideias ou temas centrais de uma obra. Trata-se da palavra que identifica um elemento particular ou o seu conteúdo.	Priberam (2022)
Patrocinador	Trata-se da entidade que patrocina a pesquisa.	Especialista do domínio.

Pessoa	Criatura humana ou indivíduo humano enquanto sujeito de direitos e deveres.	Priberam (2022)
Pessoa Institucionalizada	Refere-se a um indivíduo humano associado a uma Instituição de ensino.	Especialista de domínio
Pós-Graduação	Titulações dos docentes que possuem formação superior, tais como: especialização, mestrado, doutorado ou podendo não dispor título de pós-graduação.	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2013)
Pré-Requisito para Aprovação	Trata-se de cada pré-requisito para aprovação no curso stricto sensu.	Especialista do domínio
Produto	Produto é o resultado técnico-científico (projeto experimental ou produto técnico) de um curso de Programa de Pós-Graduação Stricto Sensu (também conhecido como mestrado ou doutorado profissional). É o produto técnico/tecnológico ou trabalho final de pós-graduação como resultado de um curso profissional de Pós-Graduação junto a tese quando mestrado, ou anexo ao doutorado quando se trata de uma tese.	Especialista de domínio
Programa de Pós-Graduação	É uma programação de curso de pós-graduação que são criados quando constatados da sua viabilidade, relevância e número adequado de professores pesquisadores com titulação e produção científica suficientes para dar sustentação à criação de curso de mestrado e/ou doutorado em determinada área de conhecimento, bem como a sua adequação aos critérios da Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior - Capes para apresentação e propostas de cursos novos. Programa de Pós-Graduação pode ser gerido por uma única Instituição de Ensino Superior ou por Rede de Ensino. Assim, compreende programa de mestrado, doutorado, cursos	Brasil (1996), Brasil (2021), Brasil (2022), Brasil (2022a) e Especialista de domínio

	de especialização, aperfeiçoamento, dentre outros, e são abertos a candidatos diplomados em cursos de graduação e que atendam os requisitos das instituições de ensino. Compreende programas de mestrado e doutorado.	
Rede de Ensino	É um coletivo de Instituição de Ensino Superior.	Especialista do domínio
Subárea	É o fracionamento da área do conhecimento definida com base no objeto de estudo e procedimentos metodológicos aceitos e amplamente usados. Diz respeito ao 3º nível da área de conhecimento do CNPq.	Fundação (2022)
Tese	Colaboração acadêmica original para o conhecimento e apresentada para obter o grau acadêmico de doutor e dos títulos universitários livre-docente e professor titular. Constitui trabalho de pesquisa original e individual que forneça contribuição importante para a área de conhecimento na qual está inserida.	Universidade Federal do Espírito Santo (2015), Brasil (2022b)
Tipo de Classificação da Categoria Acadêmica	Trata-se do tipo de classificação da categoria acadêmica: Especial, Privada comunitária, Privada confessional, Privada com fins lucrativos, Privada - Particular em sentido estrito, Privada sem fins lucrativos, Pública Estadual, Pública Federal, Pública Municipal.	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2021)
Tipo de Idioma	Diz respeito aos tipos de idioma (linguagem ou língua falada por um povo ou nação).	Ethnologue Languages of the World (2022) e Fundação (2023)
Tipo de Modalidade de Participação	Diz respeito ao tipo de modalidade de participação em uma IES: Presencial, a distância e semi-presencial.	Brasil (2004b) e Brasil (2018b)
Tipo de Classificação da	Trata-se do tipo de classificação da organização acadêmica: Centro Federal de Educação Tecnológica,	Instituto Nacional de Estudos e Pesquisas

Organização Acadêmica	Centro Universitário, Faculdade, Instituto Federal de Educação, Ciência e Tecnologia, Universidade.	Educacionais Anísio Teixeira (2021)
Tipo de PDF	Diz respeito ao tipo de classificação do arquivo PDF em: PDF tradicional, PDF/A para arquivamento, PDF XFA estático e PDF XFA dinâmico. Os tipos de PDF possíveis para um arquivo são: PDF/A-1a, PDF/A-1b, PDF/A-2, PDF/A-2b, PDF/A-2u, PDF/A-3a, PDF/A-3b, PDF/A-3u, PDF/A-4f, PDF/A-4e, PDF_A1a, PDF_A1b, PDF_A21, PDF_A2b, PDF_A2u, PDF_A3a, PDF_A3b, PDF_A3u, PDF_A4f e PDF_A4e.	Portable Document Format (2023)
Trabalho de Conclusão de Curso	Deve ser um trabalho acadêmico de finalização do curso, sendo necessariamente, pelo menos, na modalidade de uma monografia, ou outros trabalhos técnicos associados quando previstos. São trabalhos que retratam o resultado de estudo ou pesquisa sobre um tema, sejam eles exigidos por disciplina, módulo, estudo independente, curso e programa.	Brasil (2022a), Especialista de domínio, Universidade Federal do Espírito Santo (2015)

Fonte: Realizado pelo próprio autor.